

C55.602:T75

REMOTE SENSING

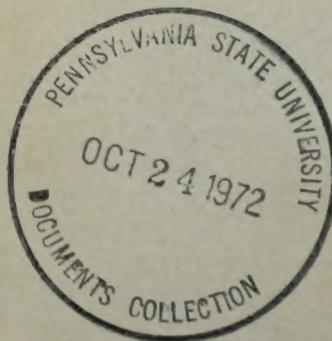
OF THE

TROPOSPHERE

U.S. Department of Commerce
National Oceanic and Atmospheric Administration



University of Colorado





U.S. DEPARTMENT OF COMMERCE

Peter G. Peterson, Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Robert M. White, Administrator

ENVIRONMENTAL RESEARCH LABORATORIES

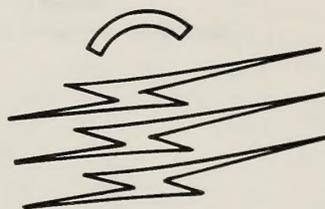
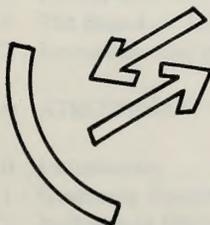
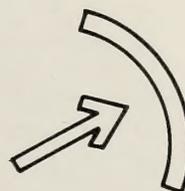
Wilmot N. Hess, Director

UNIVERSITY OF COLORADO

Frederick T. Thieme, President

Remote Sensing of the Troposphere

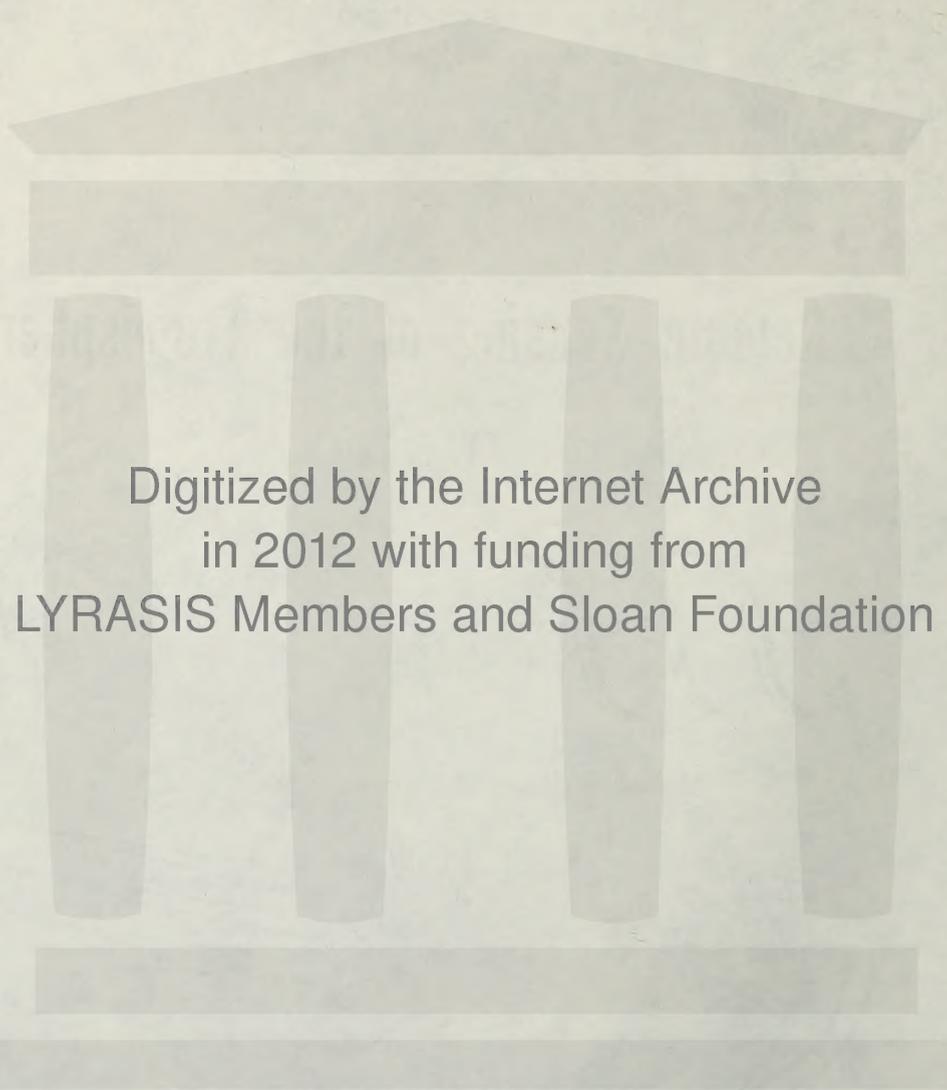
V. E. Derr



WAVE PROPAGATION LABORATORY, ERL
ELECTRICAL ENGINEERING DEPARTMENT, CU
BOULDER, COLORADO

Issued August 15, 1972

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402



Digitized by the Internet Archive
in 2012 with funding from
LYRASIS Members and Sloan Foundation

<http://archive.org/details/remotesensingoft00derr>

CONTENTS

Page

Preface

Editor's Guide

INTRODUCTION, R. W. Knecht 1-1

SECTION I, THE LOWER ATMOSPHERE

CHAPTER 1 THE ATMOSPHERIC ENVIRONMENT, R. L. Grossman

1.0	Introduction	1-1
1.1	Gaseous Constituents	1-1
1.2	The Equation of State	1-1
1.3	Effect of Water Vapor on Air Density	1-2
1.4	Aerosols and Hydrometeors	1-3
1.5	The Vertical Structure of the Atmosphere: Stability and Mixing	1-5
1.6	The Horizontal Structure of the Atmosphere: General Circulation Motions and Scales of Motion	1-11
1.7	Averaging in the Atmosphere	1-14
1.8	Basic Equations of Motion in the Atmosphere	1-19
1.9	Remote Sensing and the Atmospheric Environment	1-23

CHAPTER 2 FUNDAMENTAL PROBLEMS OF ATMOSPHERIC SCIENCE, E. R. Reiter

2.0	Introduction	2-1
2.1	Scales of Motion	2-1
2.2	Primitive Equations	2-3
2.3	The Interaction of Atmospheric Motion Systems	2-7
2.4	Unsolved Problems	2-13
2.5	Conclusions	2-24

CHAPTER 3 ATMOSPHERIC DYNAMICS, W. H. Hooke

3.0	Introduction	3-1
3.1	The Equations of Motion	3-2
3.2	Quasi-Equilibrium Approximations to the Equations of Motion	3-3
3.3	Small-Amplitude Perturbation Approximations to the Equations of Motion	3-7
3.4	Other Approximations to the Equations of Motion	3-8
3.5	Atmospheric Stability	3-9
3.6	The Boundary Layer	3-12
3.7	Remote Sensing in Studies of Atmospheric Dynamics	3-13

CHAPTER 4 ATMOSPHERIC TURBULENCE, W. C. Meecham

4.0	Introduction	4-1
4.1	Governing Equations	4-2
4.2	Background Information on Fluid Mechanical and Related Properties	4-5
4.3	Special Characteristics of the Real Atmospheric Boundary Layer	4-17

CHAPTER 5 ATMOSPHERIC WAVES IN OUTLINE, C. O. Hines

5.0	Introduction	5-1
5.1	Governing Equations for the Elementary Case	5-1

5.2	Elementary Plane-Wave Solutions	5-3
5.3	Complications of Nonlinearity	5-5
5.4	Complications of Instabilities	5-5
5.5	Complications of Molecular Dissipation	5-6
5.6	Complications of Temperature Structure	5-6
5.7	Complications of Wind Structure	5-7
5.8	Complications of Earth Curvature and Rotation Tides	5-7
5.9	The Semidiurnal Tide	5-8
5.10	The Diurnal Tide	5-9
CHAPTER 6	THE ATMOSPHERIC BOUNDARY LAYER, J. A. Businger	
6.0	Introduction	6-1
6.1	Governing Equations	6-2
6.2	The Neutral Atmospheric Boundary Layer	6-5
6.3	The Diabatic Surface Layer	6-11
6.4	The Unstable Boundary Layer	6-21
6.5	The Stable Boundary Layer	6-34
6.6	Effects of Change in Terrain	6-38
CHAPTER 7	ATMOSPHERIC GRAVITY WAVES IN THE PLANETARY BOUNDARY LAYER, W. H. Hooke	
7.0	Introduction	7-1
7.1	Mean Temperature and Wind Structure In and Above the Stable Planetary Boundary Layer	7-2
7.2	Wave Propagation in Boundary-Layer Wind and Temperature Structure	7-3
7.3	Wave Transport of Energy and Momentum in the Boundary Layer	7-8
7.4	Wave Interactions with other Boundary Layer Motions	7-10
7.5	Orographic Effects	7-12
7.6	Summary	7-14
SECTION II, WAVE PROPAGATION AND INTERACTION		
CHAPTER 8	SCATTERING, PROPAGATION AND REFRACTION OF ELECTROMAGNETIC AND ACOUSTIC WAVES, R. H. Ott	
8.0	Introduction	8-1
8.1	Maxwell's Equations	8-1
8.2	Vector and Scalar Potentials	8-3
8.3	Free-Space Green's Function and the Fresnel-Kirchhoff Diffraction Formula	8-4
8.4	Geometrical Optics and Refraction	8-7
8.5	Tropospheric Propagation	8-13
8.6	Scattering by Spheres	8-15
CHAPTER 9	THE SPECTRA OF MOLECULES OF THE EARTH'S TROPOSPHERE, V. E. Derr	
9.0	Introduction	9-1
9.1	The Emission and Absorption of Radiation by Molecules	9-1
9.2	Spectra of Diatomic Molecules	9-5
9.3	The Spectra of Polyatomic Molecules	9-9
9.4	Linewidth as a Function of Temperature and Pressure	9-13
9.5	Atmospheric Emission	9-16
9.6	Methods of Calculating Atmospheric Attenuation	9-17
9.7	Spectra of Atmospheric Constituents	9-20
9.8	Concluding Remarks	9-27

CHAPTER 10 ATOMIC, MOLECULAR, PARTICULATE, AND COLLECTIVE GENERALIZED SCATTERING, R. L. Schwiesow

10.0 Introduction	10-1
10.1 Review of Atomic and Molecular Theory	10-1
10.2 Two Step (Single Photon) Scattering	10-7
10.3 Single Step (Two Photon) Scattering	10-13
10.4 Classical Analysis of Scattering	10-18
10.5 Collective Scattering Effects	10-22
10.6 Summary	10-25

CHAPTER 11 PROPAGATION AND SCATTERING IN RANDOM MEDIA, S. F. Clifford

11.0 Introduction	11-1
11.1 Scattering of Electromagnetic Waves	11-1
11.2 Scattering of Acoustic Waves	11-11
11.3 Line-of-Sight Propagation Effects	11-15

SECTION III, REMOTE SENSING

CHAPTER 12 REMOTE SENSING OF SEA STATE BY RADAR, D. E. Barrick

12.1 Description of the Sea Surface	12-1
12.2 MF/HF Radar Scatter from the Sea	12-8
12.3 Second-Order HF Sea Echo	12-26
12.4 UHF Indirect Bragg Scatter Using Two Frequencies	12-31
12.5 Sea State Effects on a Microwave Radar Altimeter Pulse	12-34

CHAPTER 13 ATMOSPHERIC MOTION BY DOPPLER RADAR, D. A. Wilson and L. J. Miller

13.0 Introduction	13-1
13.1 Basic Radar Principles	13-1
13.2 Doppler Radar Applications	13-9
13.3 Conclusions	13-25

CHAPTER 14 STUDIES OF THE CLEAR ATMOSPHERE USING HIGH POWER RADAR, K. R. Hardy

14.0 Introduction	14-1
14.1 Theory	14-2
14.2 Experimental Verification of the Theory	14-8
14.3 Clear Air Dot Echoes	14-11
14.4 Clear Air Convection	14-13
14.5 Clear Air Radar Patterns in Shear Zones	14-18
14.6 Radar Investigations of Clear Air Turbulence	14-25
14.7 Summary	14-29

CHAPTER 15 RADIOMETRY, J. B. Snider and E. R. Westwater

15.0 Introduction	15-1
15.1 Radiative Transfer	15-1
15.2 Measurement Techniques	15-6
15.3 Applications	15-23

CHAPTER 16 INVERSION TECHNIQUES, E. R. Westwater and O. N. Strand	
16.0	Introduction 16-1
16.1	Physical Examples of Inverse Problems 16-1
16.2	The Fredholm Integral Equation of the First Kind: An Ill-Posed Problem 16-4
16.3	Inversion Techniques 16-6
16.4	Examples of Profile Retrieval by Indirect Sensing 16-9
16.5	Conclusions and Final Remarks 16-11
16.6	Acknowledgements 16-11
CHAPTER 17 ATMOSPHERICS AND SEVERE STORMS, W. L. Taylor	
17.0	Introduction 17-1
17.1	Thunderstorm Electricity 17-1
17.2	Characteristics of Atmospherics 17-2
17.3	Sensing Severe Storm Activity 17-6
17.4	Electricity in the Non-Violent Atmosphere 17-14
17.5	Summary 17-17
CHAPTER 18 TEMPERATURE AND WIND STRUCTURE STUDIES BY ACOUSTIC ECHO-SOUNDING, F. F. Hall	
18.0	Introduction 18-1
18.1	Propagation and Attenuation of Sound in the Atmosphere 18-1
18.2	Scattering of Sound Waves 18-3
18.3	Acoustic Echo Sounder Design 18-6
18.4	Structure Observed in Returned Echo 18-10
18.5	Wind Studies by Acoustic Doppler Methods 18-17
18.6	Conclusions 18-23
CHAPTER 19 PROSPECTS FOR ACOUSTIC ECHO SOUNDING, C. G. Little	
19.1	The Physics of Acoustic Echo Sounding of the Atmosphere 19-1
19.2	Atmospheric Information Potentially Available by Acoustic Echo Sounding 19-2
19.3	Present Measurement Capabilities of Acoustic Echo Sounding 19-2
19.4	Methods for the Expansion of Current Acoustic Echo Sounding Capabilities 19-3
19.5	Future Atmospheric Studies 19-6
CHAPTER 20 APPLICATION OF FM-CW RADAR AND ACOUSTIC ECHO-SOUNDER TECHNIQUES TO BOUNDARY LAYER AND CAT STUDIES, B. R. Bean	
20.0	Introduction 20-1
20.1	The FM-CW Radar 20-2
20.2	Comparison of Tower and Radar Returns 20-3
20.3	Other Examples of Radar Returns (Insects, Convective Elements, Inversions and Waves) 20-5
20.4	Comparison of Acoustic Sounder and Tower Data 20-8
20.5	Observations of Helmholtz Waves in the Lower Atmosphere with an Acoustic Sounder 20-11
20.6	A Note of Caution 20-15
20.7	Conclusions 20-18
CHAPTER 21 PASSIVE SENSING OF NATURAL ACOUSTIC-GRAVITY WAVES AT THE EARTH'S SURFACE, T. M. Georges and J. M. Young	
21.0	Introduction 21-1

21.1	Some Definitions	21-1
21.2	Sensing Techniques	21-2
21.3	Identifying Atmospheric Wave Motions	21-5
21.4	The Atmospheric-Wave Environment	21-9
21.5	Acoustic-Gravity Waves from Convective Storms: A Case Study	21-11
21.6	Manmade and Controlled Sources	21-18
21.7	Observations in Studies of Atmospheric Dynamics	21-19
21.8	Conclusions	21-20

CHAPTER 22 REMOTE SENSING OF THE OCEAN, K. J. Hanson

22.0	Introduction	22-1
22.1	Infrared Detection – Sea Surface Temperature	22-1
22.2	Passive Microwave Detection	22-23
22.3	Ocean Color Detection – Chlorophyll	22-38

CHAPTER 23 ATMOSPHERIC REMOTE SENSING WITH LASER RADAR, R. G. Strauch and A. Cohen

23.0	Introduction	23-1
23.1	Optical Interactions in the Lower Atmosphere	23-1
23.2	Laser Radar for Atmospheric Backscatter	23-5
23.3	Results of Meteorological Lidar Experiments	23-9
23.4	Measurement of Tropospheric Aerosols by Laser Ranging	23-19
23.5	Future Directions for Laser-Probing of the Atmosphere	23-32

CHAPTER 24 REMOTE SPECTRAL SENSING OF POLLUTANTS, R. L. Schwiesow

24.0	Introduction	24-1
24.1	Measurement of Ambient Air Quality	24-1
24.2	Pollutant Mapping	24-7
24.3	Summary	24-12
24.A	Appendix of Infrared Spectral Data	24-14

CHAPTER 25 REMOTE SENSING BY LINE-OF-SIGHT PROPAGATION, R. S. Lawrence

25.1	Review of Selected Aspects of Propagation through Turbulence	25-1
25.2	Remote Measurement of Turbulence	25-4
25.3	Remote Measurement of Wind	25-7
25.4	Remote Measurement of Temperature	25-11
25.5	Measurement of Vertical Temperature Gradient	25-12

CHAPTER 26 REMOTE SENSING FROM SATELLITES, H. W. Yates

26.1	The Satellite as a Platform for Observation	26-1
26.2	Current Operational Satellites and Research Program	26-13
26.3	Trends in Research and Development	26-30
26.4	Compilation of Electronic and Mechanical Imaging Devices for Spacecraft Applications	26-36

SECTION IV, SPECIAL TOPICS

CHAPTER 27 REACTION RATES OF ATMOSPHERIC CONSTITUENTS, E. E. Ferguson

27.0	Introduction	27-1
27.1	Reaction Rate Theory	27-1
27.2	Categories of Reactions	27-5
27.3	Specific Tropospheric Reactions	27-10

CHAPTER 28 DATA GATHERING AND PROCESSING ASPECTS OF REMOTE TROPOSPHERIC SENSING SYSTEMS, S. W. Maley

28.0	Introduction	28-1
28.1	Data Gathering Systems	28-1
28.2	Data Analysis	28-6
	Appendix A – Sampling Theory	28-7
	Appendix B – Minimum Values of Transmission Capacities	28-12
	Appendix C – Minimum Length Data Gathering Systems	28-14
	Appendix D – Description of Random Processes	28-17

CHAPTER 29 CHANNEL MODELING AND SYSTEM OPTIMIZATION FOR REMOTE SENSING: A CANONICAL APPROACH, D. Middleton

29.0	Introduction	29-1
29.1	A Concise Survey of Some Elements of Statistical Communication Theory (SCT)	29-4
29.2	Approaches to the Channel Model: Sources and Fields	29-12
29.3	Reception	29-17
29.4	Examples: Fields and Waveforms	29-19
29.5	Statistical Formulations	29-23
29.6	Some Statistical Results: Covariance of Received Scattered and Ambient Fields	29-26
29.7	Statistical Data Analysis (SDA)	29-30
29.8	Optimum Reception	29-36
29.9	Space-Time Sampling Theory	29-40
29.10	Concluding Remarks	29-42

SECTION V, SUMMARY

CHAPTER 30 STATUS OF REMOTE SENSING OF THE TROPOSPHERE, C. G. Little

30.1	The Role of Remote Sensing in the Troposphere	30-1
30.2	Categorization of the Status of Remote Sensing	30-2
30.3	Acoustic Remote Sensing of the Troposphere	30-3
30.4	Radio Remote Sensing of the Troposphere	30-3
30.5	Optical Remote Sensing of the Troposphere	30-4
30.6	Comparison of Status of Acoustic, Radio and Optical Remote Sensing	30-4
30.7	Estimates of Accuracy of Remote Sensing Measurements	30-5
30.8	Pattern Recognition in Remote Sensing	30-5

Acknowledgements

It is impossible to adequately acknowledge those who did the immense amount of labor on this book and on the course. First, I must express my gratitude to S. W. Maley, Electrical Engineering Department of the University of Colorado, my co-leader in the course, and always a great aid and comfort throughout the trials of producing a book with multitudinous authors and low budgets. Also at the University of Colorado, D. Reese and L. Christiansen gave their time and effort in publicizing the course and arranging housing for students.

R. M. White, the Administrator of NOAA, W. N. Hess, Director of the Environmental Research Laboratories, and C. Gordon Little, Director of the Wave Propagation Laboratory, have been constant in their encouragement, and I am very grateful for their support. R. S. Rhodes, Assistant Director of the Wave Propagation Laboratory, has given wise counsel, particularly in financial and personnel matters. In the early stages, F. S. Settle displayed great initiative in organizing the details, and she designed the basic format of the book. My secretary during the most active phase of preparation has been M. K. Britton. Without her initiative, devotion to excellence, and her cheerful, diplomatic guidance of 30 manuscripts meandering between authors, typists, editors, proofreaders, referees, compositors and layout personnel, this book could not have been completed before the course opened. Publication Services of ERL, headed by M. E. Nason, continued to be responsive in editing, composition, and layout in spite of perpetually missed schedules. J. R. Lebsack, particularly, at great personal sacrifice, has singlehandedly carried the major load in assuring the quality of the layout, the uniformity of the format, and, with great patience and skill, pursued the tedious task of locating and removing typographical errors. In addition I thank our student aids, K. M. Yoshida, C. J. DeHaven, and R. B. Slusher for their patience and care in handling figures, drafting and layout.

I can only thank collectively the more than forty reviewers and the many typists who contributed to the final production. And last, I express my gratitude to the authors for their striving always for technical excellence and clarity, and their tolerance of a demanding editor.

Vernon E. Derr
National Oceanic and Atmospheric Administration
Boulder, Colorado

Preface

This volume arose from the lectures in a course, "Remote Sensing of the Troposphere", given at the University of Colorado, Boulder, during June 19-30, 1972. This course, jointly sponsored by the University of Colorado and the National Oceanic and Atmospheric Administration, presented the basic scientific background necessary for studying the environment – primarily the lower atmosphere – by remote probing methods. Specific systems for remote measurement were discussed, and problems of atmospheric physics, for which these systems yield significant new data, were studied.

Intensive lecture courses, utilizing the services of many scientists, too often lead to a disjointed collection of independent notes. To give this collection a more lasting usefulness, fundamental physical and mathematical tools were reviewed, an index added, and sufficient references provided to guide the reader in more detailed research. Because of the haste required to make the book available to students at registration, it cannot have the quality of a thoroughly prepared text; nevertheless, it is a useful introduction to the important new field of remote sensing of the environment.

Remote measurement is a significant new tool in the study of structure and dynamics of the atmosphere and in applications to weather forecasting, weather modification, pollution studies, severe storm warning, and transportation safety because it combines accurate data with excellent spatial and temporal resolution, in essentially real time, over much larger volumes of the atmosphere than possible by other means. Remote sensing methods have been in use for many years in the form of weather radar and other techniques. However, the concerted effort to make full use of the electromagnetic spectrum and, in addition, acoustic probing of the atmosphere, have recently led to many developments in Doppler radar, FM/CW radar, laser systems, and acoustic echo-sounders. These, combined with new uses of radiometers, supplemented by improved in-situ instruments, and provided with airborne vehicles as well as based on the earth's surface, have brought us very close to a measurement capability by whose means we may obtain data on all important parameters of the lower atmosphere.

These lectures were designed for scientists interested in applying new remote measurement methods to the study of the environment; furthermore they may also benefit those expanding their interests into atmospheric science. Established meteorologists and researchers in related fields will find them relevant to their increasing need for improved atmospheric data. Although terrestrial atmospheric studies are emphasized, the methods discussed are more generally applicable to planetary atmospheres, and to remote probing of the ocean, and to some extent the solid earth.

The text, after an introduction, has five major sections. The first of these, Chapters 1 – 7, deal with the constitution, structure, and dynamics of the troposphere. Chapters 8 – 11, the second section, deal with electromagnetic and acoustic energy, and their propagation through, and interaction with the atmosphere. Here fundamental principles and mathematical tools are presented. The third section, Chapters 12 – 26, is devoted to detailed discussion of the methods of remote sensing. No attempt has been made to include photographic techniques since these are covered in many other publications. Special topics discussed in the fourth section, Chapters 27 – 29, concern the interaction of pollutants with the atmosphere, data handling, and the mathematical theory of communication as applied to the optimization of remote sensing. Chapter 30 is a summary of the present status of tropospheric remote sensing, a realistic evaluation of present capabilities, and a reasoned prognosis of future directions.

Editor's Guide to the Text

Readers are usually quite properly bored with prefaces. However, it is important that the user of this book should be aware of its arrangement in order to use it efficiently.

First, he should be aware that symbols are uniform only within each chapter. Definitions are given in each chapter and summarized in a List of Symbols at the end of each chapter.

The chapters, sections, and subsections are numbered by a decimal system throughout; thus 9.4.2 is the ninth chapter, fourth section, second subsection. Equations are numbered consecutively throughout each chapter, invariably and uniquely signified by a colon. For example, (9:18) is the 18th equation of Chapter nine. Figures and tables are also numbered consecutively, but signified by a prefixed letter. Thus (F9.2) and (T9.4) denote figure 2 and table 4 of Chapter nine, respectively. Pages in the text are numbered by chapter number and page within the chapter. The index refers to chapters and sections, not to pages.

In a text with a wide range of subject matter, the desire to conform entirely to the International System of Units (SI)[†], recommended by the National Bureau of Standards, is almost certain to be thwarted by the naturalness or familiarity of units conventional in some areas of science or technology. In this text, SI units have been employed except when scientific communication may be impaired by too strict adherence. Exceptions may occur when units are long established, or if the ready translation to SI units is not convenient, or if intuitive grasp of a physical situation is made more difficult by unfamiliar units. Where confusion may arise, the units are quoted in conventional and SI units.

V. E. Derr
Boulder, Colorado
March 1972

[†] See Page, C. H. and P. Vigoreaux, *The International System of Units*, National Bureau of Standards, Special Publication 330, or Mechtly, E. A., 1966: *The International System of Units, Physical Constants and Conversion Factors*, NASA SP-7012.

Introduction THE ENVIRONMENTAL RESEARCH LABORATORIES AND REMOTE SENSING

R. W. Knecht

Environmental Research Laboratories
National Oceanic and Atmospheric Administration

It is a pleasure to be able to give the opening introductory lecture in this course on remote sensing of the troposphere. In my presentation I will give an overview of the work of NOAA's Environmental Research Laboratories (ERL) with special emphasis on our remote sensing activities. This approach will make clear the critical importance of remote sensing techniques to the accomplishment of the ERL and NOAA missions.

The Environmental Research Laboratories as the name implies, are intended to be the basic research arm of NOAA. Our mission is to provide the necessary new scientific understanding and technical tools to allow NOAA's service branches such as the National Weather Service and National Ocean Survey to do a better job of providing the required forecasts, warnings, and descriptions for a more intelligent and prudent use of our environment. Because of urgent needs in connection with earthquakes, tsunamis, tornadoes, hurricanes and other severe weather phenomena, droughts, and astronaut radiation hazards, to mention a few, ERL's work necessarily spans the wide range of environmental disciplines from solid earth geophysics through the oceanic and atmospheric sciences to space research.

ERL's parent organization, the National Oceanic and Atmospheric Administration (NOAA), was formed in October 1970 by President Nixon as a part of Presidential Reorganization Plan No. 4. In forming NOAA, the President sought to bring together much of the oceanic and atmospheric environmental work in the civilian side of government. In addition to the Department of Commerce's Environmental Science Services Administration (ESSA), the Bureau of Commercial Fisheries and the laboratories of the Bureau of Sports Fisheries were brought into NOAA from the Department of the Interior as was its Marine Mining Technology Center, the Data Buoy Program from the Coast Guard, the Sea Grant Program from the National Science Foundation, and the National Oceanographic Instrumentation Center from the U. S. Navy. As of January 1972 after a certain amount of internal reorganization, NOAA consisted of the following six Major Line Components: the National Weather Service, the National Environmental Satellite Service, the National Ocean Survey, the National Marine Fisheries Service, the Environmental Data Service, and the Environmental Research Laboratories. NOAA's missions include monitoring and prediction of environmental conditions on both a long and short term basis, and, where appropriate, environmental modification and control. NOAA also has a role in exploration, development and conservation of marine resources.

What is ERL's part in all this? We perform an integrated program of research and research services toward understanding the geophysical environment, toward protecting the environment, and toward improving NOAA's services.

Five of our 11 laboratories are located in Boulder: the Aeronomy Laboratory, Atmospheric Physics and Chemistry Laboratory, Earth Sciences Laboratories, Space Environment Laboratory, and Wave Propagation Laboratory. The Program Manager for Weather Modification also makes his home in Boulder, although the units under his direct supervision—the Research Flight Facility and the Experimental Meteorology Laboratory—are located in Miami, Florida.

Six other principal units are the Air Resources Laboratories in Silver Spring, Maryland; the Atlantic Oceanographic and Meteorological Laboratories, Miami, Florida; the Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; the Marine Minerals Technology Center, Tiburon, California; the National Severe Storms Laboratory, Norman, Oklahoma; and the Pacific Oceanographic Laboratory, Seattle, Washington.

I shall now discuss the work of each of our laboratories, emphasizing those activities which are pertinent to this course. Because of time limitations, only a brief account of the work of each of the units will be possible. Participants interested in obtaining additional information on any phase of the work of a particular laboratory are encouraged to contact the laboratory director.

A total of about 40% of the work of ERL is carried on in the atmospheric sciences. The work of seven of our laboratories is devoted in part or entirely to research in this area. I'll begin with a discussion of the work of the National Hurricane Research Laboratory which is a part of the Atlantic Oceanographic and Meteorological Laboratories in Miami. NHRL's chief objectives are to understand and beneficially modify hurricanes, the latter project conducted jointly with the Navy and Air Force. It is important to recognize that there is no intention here to dissipate hurricanes; this could trigger climate effects which we are in no position to forecast. On the other hand, a small reduction in peak wind velocity could bring about a tremendous reduction in damage and injuries. And a minor change in direction could steer a hurricane away from a highly populated coastline.

The technique used in hurricane modification involves the seeding of supercooled clouds with silver iodide. The hypothesis is that the addition of these artificial freezing nuclei will transform the water to ice, thus releasing the latent heat of fusion and causing the clouds to grow. When this is done in the hurricane's eye wall or rainband clouds, according to the hypothesis, it induces pressure gradient changes that could reduce the intensity of the storm. Successive eye wall seedings of Hurricane Debbie in 1969 apparently produced wind speed reductions of 30% and 15%, but rainband seeding in Hurricane Ginger in 1971 was inconclusive.

Project Stormfury, which is the name this hurricane work has gone under for several years, employs airborne radar, satellite photos from vehicles in geostationary orbit, and infrared soundings from polar-orbiting weather satellites. These techniques can be viewed as conventional, since they now have been incorporated into the operational mode of the National Weather Service.

The Experimental Meteorology Laboratory, also in Miami, has done some of the basic research that is utilized in Project Stormfury. This group has developed a dynamic seeding technique that is a remarkably potent rain producer in tropical cumulus clouds over South Florida. What EML does is drop silver iodide flares into the tops of supercooled clouds when the tops are around the 20,000-25,000-foot level. As in the hurricane seeding, the water is converted into ice. The released heat of fusion adds buoyancy to the air and makes the cloud grow faster in both the vertical and horizontal dimensions. More air is sucked in, more moisture is processed, and the cloud continues to grow more than it would have if it had not been artificially seeded. The laboratory estimates that seeded single clouds produce 3.3 times more rain than similar unseeded clouds.

However not all clouds can be productively seeded. They must meet certain temperature, altitude, and size conditions. If there is not enough natural convective activity or conditions are too disturbed, seeding would be pointless. In fact, the observations indicate that seeding may actually reduce precipitation on a naturally rainy day.

Their success in single-cloud seeding research has encouraged EML to launch an experiment aimed at promoting cloud mergers, and, consequently, much more rainfall. Mergers occur naturally when clouds undergo significant horizontal expansion; more natural seeding occurs and the result is considerable rainfall over a wider area. In 1970, the laboratory did its first tentative work in this area and it looked promising. That is to say, it appeared that an artificially merged system might produce an order of magnitude more rain than similar but unmerged clouds. In the only case of merger apparently promoted through silver iodide seeding, two merged clouds produced 8,800 acre-feet of rain while two similar clouds managed only 400 acre-feet.

In 1971, a severe drought in South Florida prompted the Governor to ask NOAA to allow the laboratory to use the dynamic seeding technique to help relieve the situation. The project was conducted in April and May of 1971, seedable clouds appearing only during the latter month. EML believes mergers were produced on at least five days. Using conservative estimates, they computed the additional rainfall produced by seeding at 100,000 acre-feet, between \$5 and \$10 million worth of water depending on whether the price used is based on municipal or irrigation water. The benefit-cost ratio then was either 32 or 64. The program did not break the drought, the region's water deficit being about 3,000,000 acre-feet, but it did help the situation some. The next step may be drought prevention on a routine basis.

The prime rainfall measuring instrument in the cumulus seeding work has been the University of Miami's specialized 10-centimeter radar. The radar displays rainfall at three levels shown graphically as contours. In real time, the radar operators guide the project aircraft to seedable clouds.

Remote sensing is the backbone of much of the Atmospheric Physics and Chemistry Laboratory's weather modification research. The Laboratory's interests range from suppressing lightning for launches at Cape Kennedy and modifying costly fogs at the Panama Canal to experimenting with snowfall redistribution and hail suppression.

Working on the problem of induced lightning, APCL researchers have fired rockets tipped with steel balls into thunderstorms to bleed off lightning before the field intensifies enough to create a natural lightning discharge. Remote sensing techniques are used here both from the ground and from project aircraft to assess the cloud electric field as it builds and to determine when and where to launch the triggering rocket.

In the laboratory's work with the National Hail Research Experiment in northeastern Colorado (a project managed by the National Center for Atmospheric Research) infrared sensing has been used from aircraft to map hail swaths, and should be able to measure the depth of the hail on the ground. Remote sensing is also used in the National Snowfall Redistribution Project, in which the laboratory is assessing the feasibility of redistributing the massive snowfalls coming off Lake Erie in the early winter. In this effort we are over-seeding the storms in order to reduce the size of the falling snowflakes so that the wind will spread the snowfall over a larger area reducing the fall along the shore belt.

Infrared sensors are also used to study the injection of water vapor into the stratosphere by thunderstorms and high altitude jets. In connection with the Apollo/lightning work, ERL has studied the ionized tail of the Saturn rocket through thermal photography.

The laboratory has pioneered in the use of lidar to study high altitude pollution. Suspended particulates in the upper atmosphere have been of increasing concern, and APCL has found that the density of such particulates over the North Atlantic Ocean has doubled since the Carnegie cruises of the early 1900's. A thorough study and mapping of particulate levels over the globe will yield much information on pollution dispersal, retention, and the overall state of our atmosphere. Lidar is a near-perfect tool for such studies.

The National Severe Storms Laboratory uses conventional remote sensing observational techniques to develop better ways of detecting and providing advance warnings of severe weather. Their annual spring observational program in Oklahoma during the storm season has added substantially to the weather scientists' knowledge of thunderstorms and tornadoes.

But one area that has eluded analysis is the dynamics of severe storms. This has been due to the lack of feasible means of studying the interior processes of storms. Research aircraft generally do not penetrate thunderstorms because of the danger of destructive turbulence and the possibilities of hailstone damage. While the use of radiosondes is not impossible, it would be a logistical nightmare to chase thunderstorms along roads and across fields, in an attempt to release instrumented balloons into rapidly moving storm cells.

Conventional radar in the past 20 years or so has enormously increased our knowledge of the structure of storms. It can locate storms, determine their general movement, and map out areas of precipitation within the clouds, but it tells us little of the dynamic nature of storms.

NSSL is now using Doppler radar to measure the radial velocity of windblown precipitation particles inside the clouds. By combining radial velocities obtained with two or three Doppler radars probing the same storm, we are able to get a three-dimensional, quantitative, dynamic view of storm interiors. Parallel development of means of processing the massive amounts of data streaming out of the radars is making the system practical. For the first time in meteorological research, we can study updrafts, downdrafts, convergence, divergence, and vortices in a developing storm by actually observing them.

The new system—we expect to have two Dopplers in operation by 1973—and methodology may enable forecasters to provide earlier and more accurate tornado warnings and predictions. (See the Wave Propagation Laboratory section for tornado detection via electro-magnetic signatures.)

The laboratory's radar expertise has also been used in such areas as water management and identification of hazardous flying conditions. Working with National Weather Service hydrologists, our meteorologists provided continuous, radar-derived estimates of rainfall rate and its geographical distribution. The hydrologists used the digitized radar data, in the form of 14,000 intensity points on a 40,000-square-mile area, to determine the probable amount and location of runoff. Then they adjusted their surface water management systems to conserve the maximum amount of rainwater from the storm under surveillance.

Introduction

NSSL radar specialists, together with their counterparts at the Royal Aircraft Establishment, have related radar signatures and turbulence from data recorded on more than 500 flights in and around storms. They are also able to forecast the arrival time and intensity of gust fronts associated with moving storms. Based on this work, detailed information on the use of radar in avoiding storm turbulence hazards has been published in a Federal Aviation Administration circular to pilots.

The Air Resources Laboratories work on problems of air pollution, environmental quality, and long term climatic change. This is our largest group, with headquarters in Silver Spring, Maryland, and branches in Idaho Falls, Idaho; Las Vegas, Nevada; Raleigh, North Carolina; Oak Ridge, Tennessee; and Mauna Loa Observatory, Hawaii.

The meteorological problems associated with the growing uses of nuclear energy are manifold. Air Resources Laboratories provide meteorological support to the Atomic Energy Commission for work associated with nuclear weapons, nuclear rockets, industrial nuclear explosives, nuclear fallout, and nuclear reactors. They develop models of mesoscale meteorological circulations, transport, dispersion, diffusion, and deposition to provide evaluations of radiological hazards in the geographical areas concerned, areas that range from local to global. They engage in extensive analyses of meteorological and radiological data after nuclear events. The laboratories expect soon to have in operation six new portable, computerized PACER wind-finding radars to replace the cumbersome and obsolete M-33 radars. This is a major step toward the laboratories' long-range goal of a completely automated meteorological data system, supported by the Environmental Protection Agency.

ARL is carrying out a variety of urban air pollution studies. The studies cover quantification of air pollution potential forecasts, transport and dispersion of pollution plumes from tall (800-foot) stacks, the "urban heat island" and its effects on pollution, and long-term geophysical aspects of air pollution. For example, recent radiation measurements in Cincinnati and other locations have shown a marked decrease in ultraviolet (0.59 micron). Some of the smoke-plume work is done with lidar equipment.

The research program of the Geophysical Fluid Dynamics Laboratory is directed toward fundamental understanding of large-scale atmospheric and oceanic circulations and their interactions. They are working on numerical models of the world's fluid envelope. Development of valid theories and workable numerical models would put in our grasp two major capabilities:

- 1) Long-range ocean/atmosphere prediction and improved short-term predictions.
- 2) Techniques for discovering and testing future large-scale weather and climate modification approaches.

Even with today's incredible array of environmental sensors, it is still difficult to make an accurate prediction beyond two or three days, in part because of lack of computing power and in part because only about 20% of the earth is under sufficiently detailed weather observation to permit high-resolution meteorological analysis. And good data are often lacking in that 10%. The tropical oceans, where much of the weather is generated, are hardly observed at all. GFDL is using these minimal, but real, data to build numerical models that have already made global and regional forecasts up to 10 days in advance, albeit with limited accuracy. (One great advance in meteorological observational technique is the satellite infrared spectrometer, which is providing extremely valuable temperature global data.)

The main computer at the laboratory's Princeton facility is a UNIVAC 1108. An IBM 360/91 is used for specialized problems. A new, considerably more powerful 64-track machine, Texas Instruments' Advanced Scientific Computer, expected to be delivered in 1973, will not only enhance the research already mentioned, but will also be used to determine what meteorological characteristics are predictable beyond several weeks, a month, a season, and a year. These studies could provide information on the ultimate limit of predictability of the weather.

Weather and climate modification predictions will be less restricted, since they do not require day-by-day time resolution. The laboratory will be in a position to develop a numerical model that could run through, perhaps, a 100-year simulation on the climatic effects of increasing carbon dioxide and particulates in the atmosphere.

Moving to our space research program, the Space Environment Laboratory has developed a technique for forecasting the occurrence and location of solar flares. The technique used in these forecasts is replete with remote sensing of various kinds of electromagnetic radiation and energetic particles from ground and space platforms. The most critical sensor in the array is one of the oldest types of remote sensing instruments, a small astronomical telescope set up to photograph the sun in hydrogen-alpha light.

Careful analysis of the markings on the solar disk is used to infer magnetic fields on the H-alpha surface, the characteristics of which reveal flare potential.

Space environment forecasts, including geomagnetic storms, are useful to the National Aeronautics and Space Administration, the Air Force, long-distance telecommunicators, and power companies, among others. The relationship of solar flares and geomagnetic storms to power operations is particularly interesting. In the northern part of the country and in Labrador, power companies have gone back in their records to look at power outages that they could not explain at the time. They have found a close correlation with magnetic storms. Toronto, for example, was blacked out for the better part of a day because of a magnetic storm that induced surges in the power lines. Now, in many cases, we can warn them in time to take preventive action.

SEL is also developing a solar-terrestrial environment model to accurately describe events on the sun's surface, how they propagate through space, whether they will reach the earth environment, and, if so, how they will affect human activity. The laboratory hopes to have a serviceable computerized model by 1980.

The space environment group has a number of other activities involving remote sensing. They have found, for example, using radio sensing gear, that thunderstorms transmit great sound pulses into the ionosphere. Echoes from the ionosphere indicate that the upward surge of atmospheric molecules associated with the sound pulse is amplified from about 15 feet at thunderstorm level to 15,000 feet in the ionosphere.

The laboratory also hopes to fly a small space-weather modification experiment aboard a geostationary satellite. When activated, the device would inject a small blob of plasma into the magnetosphere and hopefully precipitate a flow of natural electrons along the earth's magnetic field and down into the northern and southern ionospheres producing aurora and other short-lived geophysical events.

The Aeronomy Laboratory does research on the physical and chemical processes of the ionosphere and, most recently, of the troposphere. The major portion of their work in the past has been laboratory experiments and observational studies of ion reaction processes, some of it *in situ* and some through remote sensing. They have been extremely productive.

For example, in the late 1960's an Air Force Cambridge group showed that at 80 kilometers altitude the most common positive ion was something that appeared to be $H_5O_2^+$, two water molecules and a proton. At first the observers thought their spectrometer was sick. But, laboratory work by our people showed that such a "water cluster" ion was theoretically possible.

The emphasis now is on neutral chemistry. Using infrared lasers and techniques developed in the ion reaction rate studies, AL is working on the basic chemistry of air pollution in terms of its chemical products and reaction rates. The laboratory has worked out a theory of atmospheric turbulence—based on theoretical studies in plasma physics—which is being used in pollution dispersion research. They are also developing a model of the chemistry and dynamics of the stratosphere.

The laboratory is also involved in rocket work. In 1970, they completed the first successful measurement of atomic oxygen in the ionosphere. What they did was expose silver foil to the outside so that the ambient atomic oxygen oxidized the silver thus changing the electrical resistance of the foil. Measurement of the change in electrical resistance of the silver showed the amount of atomic oxygen present.

AL used to operate a large backscatter radar facility at Jicamarca, Peru, which is now Peruvian control. It is one of the best such radars partly because of its proximity to the equator. Our people still use the high-powered facility to make indirect measurements of electron densities and temperatures, ionospheric winds, and electric fields.

We feel that the most exciting thing happening now in ionospheric physics is the artificial production of short-lived ionospheric disturbances. Together with the Institute for Telecommunications Sciences (which was formerly a part of ERL but is now with the Office of Telecommunications), the Aeronomy Laboratory participates in the use of an extremely powerful radio transmitter in Platteville, Colorado. The vertical-looking antenna array, fed two megawatts of power, radiates at between five and ten megahertz. Focusing gives the beam an effective radiated power of 100 megawatts. It heats the ionospheric electrons, producing a disturbed volume 50 to 100 miles in diameter and electrons motion along the

geomagnetic tubes of force. Observations are made chiefly with ionosondes and airglow photometers. Among the major early findings were unexpected attenuation of radio reflectivity from the heated electrons and generation of the Spread-F phenomenon.

The ionospheric modification work is also expected to benefit the field of plasma physics. The natural ionospheric plasma has some significant advantages over laboratory plasmas, such as those used in nuclear fusion research. Chief among these is the ionosphere's unbounded nature, as opposed to the experiment-disrupting walls of laboratory containment devices. This natural outdoor plasma can now be simultaneously modified and observed from the ground.

The Wave Propagation Laboratory devotes itself almost entirely to the development and application of new methods of remote sensing for geophysical research, improved weather services, and the study of environmental problems. The names of the program areas serve to indicate the breadth of the laboratory's activities: optical propagation, atmospheric spectroscopy, environmental radiometry, micrometeorology, meteorological Doppler radar, geoacoustics research, and atmospheric acoustics.

We expect that, in due course, by using such remote sensors as Doppler radar, acoustic sounders, and lidar, one will be able to do a much better job than that now being done by the standard weather station. From one point on the ground vertical profiles of most of the important weather parameters can be obtained continuously in time and space.

WPL's special studies section conducts several exploratory research projects not assigned to the programs. In one such project, the spectra of ocean waves and mean-surface velocity of the waves by means of radar sea-scatter have been obtained. In another the feasibility of making direct measurements by remote sensing techniques of the fluxes of heat, momentum, and water vapor is being explored.

The optical propagation group is developing lasers to increase the accuracy of distance-measuring two orders of magnitude over the simple single laser technique, which is limited to 1 part in 10^7 . Eventually distance measurements might be further improved by adding a radio frequency to the system.

Another scheme employing a laser has measured integrated wind velocity across a 10-mile path between two mesas. The beam traveled across Boulder from north to south and back to a pair of detectors. The time lag in the appearance of a given scintillation pattern in the two detectors is indicative of the wind velocity transverse to the optical path. This technique appears to have application in certain air pollution prediction problems.

The atmospheric spectroscopy program is engaged in studies of identification and measurement of the gaseous constituents of the polluted atmosphere. They have measured infrared absorption spectra of SO_2 and NH_3 , and have calculated absorption spectra of H_2O , O_3 , CO_2 , CH_4 , N_2O , CO , and HCl . Raman lidar is being used in investigations of SO_2 , water vapor, and temperature fluctuations. The lidar technique is also being developed for wind sensing.

WPL was the first to use a dual-Doppler radar system to obtain, analyze, and display two-dimensional wind field data from inside a convective storm. The work resembles that being carried out at the National Severe Storms Laboratory. One of the major differences is that WPL uses portable three-centimeter radars while the severe storms group uses a more sensitive, but more cumbersome, ten-centimeter instrument.

The environmental radiometry program is doing what I consider some of ERL's most exciting work, the search for electromagnetic signatures of tornadoes. They have observed greatly enhanced, distinctive atmospheric electrical discharges when tornadoes and funnel clouds were reported within 60 miles. It may be that we have a reliable tornado warning system in the making. The radiometry people are also developing methods for measuring liquid water and ice content of clouds and storms.

The micrometeorology program investigates atmospheric microstructure and turbulence for use in measurements of vertical fluxes of heat, momentum, and mass, and interpretation of electromagnetic wave propagation. Specifically, they study temperature, humidity, refractive index, wind, and other parameters as a function of height, time, and atmospheric stability. They are also using an FM, continuous-wave radar to study the dynamics of clear air in the boundary layer.

The overall objective of the geoacoustics research program is to find out what role acoustic-gravity waves play in atmospheric dynamics. They study the interactions of acoustic-gravity waves and other geophysical phenomena, construct theoretical models of wave generation and propagation, and coordinate technical support for a global network of microbarograph observatories. They have been able to associate infrasonic waves with such events as thunderstorms, solar flares, volcanic eruptions, re-entering space vehicles, earthquakes, and tornadoes.

WPL's atmospheric acoustics program makes use of the extreme sensitivity of sound waves to atmospheric temperature fluctuations. Echoes can be obtained from distances of 3,000 feet whenever the air temperature fluctuations exceed .001 degree Centigrade. The acoustic sounding technique is being refined for studies of thermal convection, temperature inversions, internal gravity waves, and temperature fluctuations in the boundary layer. By permitting continuous monitoring of boundary layer structure and processes to an extent never before possible, the technique is helping provide a totally new picture of the dynamics of the lower atmosphere.

Acoustic sounding also is being rapidly developed for practical application. The program scientists have found 1) that they can see the internal structure of the atmosphere in dramatic detail up to 5,000 feet, 2) that it is possible to measure wind profiled and temperature inversions for use in air pollution monitoring and forecasting, and 3) that it should be possible to detect low-level windshear and turbulence at airports to help in safety and efficiency of takeoffs and landings. Doppler techniques and the use of three echo sounders simultaneous are now being used for three-dimensional studies of the structure and dynamics of the boundary layer.

Our Earth Sciences Laboratories work mainly in seismology and geomagnetism. They have a wide variety of experiments underway to measure strain, earth movement, and changes in magnetic and gravity fields. This is an effort to obtain a comprehensive understanding of the rate of strain accumulation, and what signals, if any, precede a major earthquake shock. Studies of continental drift, interaction between the earth's rotation, magnetic fields, and seismic activity, and other geophysical phenomena are being carried out to help reduce the amount of uncertainty in man's view of the workings of his planet.

One of the recent areas of interest is plate tectonics, the study of the shape, motions, and interactions of the great plates making up the earth's crust. One of the major causes of earthquakes along the North American-Pacific margin is the interaction of crustal plate edges moving at different speeds along the coastal region. Land west of the Baja California split is propagating northward, and it does look as if a section of the State of California is going to drift away from the mainland one of these ages. The Pacific plate is "diving" into the Aleutian Trench, which it created, and in the process is rubbing against and upthrusting the Aleutian island block, producing islands and earthquakes at that interface.

For years, the San Francisco branch of ESL has studied earth creep, one of the surface manifestations of plate interaction. They have measured offset in street curbs, fences, and lines of trees in orchards in order to determine the average annual rate of creep, which is from .5 to 2.5 centimeters depending on location. They emplace a 10-15-foot steel rod in a trench across a fault, anchor one end, and place the other end against an indicator. Reading the indicator once a week, they obtain a greater time resolution of creep activity, which has been found to vary in rate. Commencement of rate change sometimes propagates fairly fast along a fault. Lack of creep in a fault region may indicate a geological hang-up somewhere below, and in fact may herald a sudden readjustment or earthquake.

The seismological component of the laboratory operates some 20 observatories in the U. S. and its possessions, and helps operate and maintain about the same number of seismological stations owned by universities, research institutes and other federal agencies. The geomagnetic group operates 14 standard magnetic observatories in the U. S., Puerto Rico, Guam, and Antarctica to obtain records of short-and-long-term changes in intensity and direction of the earth's field.

ESL's geophysicists are working on mathematical and physical models of the sources of the earth's magnetic field and its changes with time.

A detailed search into geomagnetic data archives led to a harmonic analysis of the earth's field for the 1725-1965 period. Paleomagnetic studies are underway compiling a detailed history of reversals in the earth's field and interpreting the implications of the present decreasing trend in field strength.

The laboratory is doing a theoretical investigation of the hydrodynamics of the earth's core and its relationship to the observed magnetic field. The historical study of field reversals has suggested the possibility of determining fluid motion in the upper 200-300 kilometers of the core.

ESL workers have identified the galactic component of the earth's field. In developing what is probably the most sensitive magnetometer in existence, another group may have an instrument applicable to earthquake prediction studies. Finally, a small exploratory research effort on the correlations of geyser activity, gravitational tides, and earthquake occurrence is being carried on.

The oceanographic part of the Atlantic Oceanographic and Meteorological Laboratories involves studies of the structure and motions of the ocean, characteristics of the ocean basin, and sea-air interaction. Remote sensing is also coming into its own in these fields.

AOML has devoted considerable effort to the study of the Gulf Stream, the Loop Current in the Gulf of Mexico, the Antilles Current, the Florida Current, and their inter-relationships. Work done on the Gulf Stream, for example, is toward delineating its meanders and documenting its structure, its mean flow, and the modulation of this flow. A program in estuarine dynamics has as its objective the development of data to enhance tide and tidal current predictions and flushing characteristics of estuaries, information that will be required in designing ocean waste disposal systems.

AOML marine geophysicists and geologists have completed the first standard crustal section across an entire ocean basin, from the United States to Africa a continuation of a section across the continental U. S. and some distance into the Pacific Ocean. In addition, they have discovered massive geological domes, the first of their kind in the deep ocean, off northwest Africa. These formations may be salt domes, which would suggest the presence of oil. An important result emerging from this work is a theoretical understanding of the geological processes involved in the deposition and placement of petroleum and minerals.

One AOML geophysicist is the co-originator of the seafloor spreading theory, which has had a large role in the virtual rebirth of marine geology and geophysics in the past ten years. Following the statement of the seafloor spreading theory, other workers developed the concept of plate tectonics. Together, these unifying hypotheses nicely related continental drifts, the dynamics of the earth's crust, and many seismological observations. One of the fascinating scientific outputs of this new thinking is a series of maps that identifies the tracks of the continents as they are believed to have moved over the earth's surface during the past 200-million years and how they are expected to move in the next 50-million years.

The sea-air interaction program is contributing toward improvement of our capability to predict the marine environment and the earth's weather. The sea-air interface is of critical importance since the atmosphere derives much of its energy from the sun-warmed sea, while the wind imparts momentum to the waves of the ocean and affects evaporation. Knowledge of these complex interactive mechanisms and of the exchanges of mass and energy across the interface is essential to a global environmental prediction system.

One of the most serious problems in sea-air interaction research is a deficiency in reliable and repeatable measurements. Although this group is working on the improvement of conventional *in situ* measurements, they are coming to depend more and more on remote sensing instrumentation. Airborne lasers are used to observe wave conditions and wave growth. They have employed passive sensors to find that microwave emissions from the ocean surface vary according to wind and sea state although the physical mechanisms are poorly understood. Active microwave systems have been used to discover that there is a sea-state dependent signature. The potential of employing sensing systems of this type from satellites is now being actively studied.

Remote sensing of other parameters of interest to physical oceanographers is also being pursued. For example, identification of different water masses by color differences appears possible where surface temperature variations are obscured by summer solar heating in sub-tropical waters. These data have biological and fisheries implications. It has also been demonstrated that tides are observable from a satellite equipped with a radar altimeter, which will be orbited abroad the GOES-C. Another satellite-borne instrument, an infrared temperature sensor, coupled with the color observations should provide information on currents, upwellings, and locations of "hot spots" which could contribute to tropical storm intensification.

Another technique AOML has begun to employ is the manned deep submersible. A series of investigations has been in progress off the east coast of the U. S. in waters down to 2,000 meters. The vehicles are the ALVIN, belonging to the Woods Hole Oceanographic Institution, and a Perry PC-8. One of the regions of interest is the Mid-Atlantic Bight stretching from New York to Virginia. The dumping of sewage and solid wastes in this high-population-density belt has created a major pollution problem. The manned submersibles are used in studying dynamic processes related to waste disposal in five submarine

canyons bisecting the bight: the Hudson, Wilmington, Baltimore, Washington, and Norfolk. The investigators emplace bottom current meters and collect bottom and suspended sediment samples.

The Pacific Oceanographic Laboratories, has done a great deal of analytical work with SEAMAP data (SEAMAP is an extensive series of geophysical surveys in the North Pacific Ocean). POL researchers are also investigating internal ocean waves and have developed a model describing propagation of the waves. Their work also includes studies of the physical and chemical properties of deep water in the North Pacific, discovery and description of a countercurrent in the Tasman Sea, and discovery of anomalous heating in the Panama Basin. As a continuation of the transcontinental and transatlantic crustal section, POL plans also to traverse the Pacific, making gravity and magnetic measurements, seismic profiles, and sediment samples.

POL has a unique group stationed in Hawaii that concentrates on tsunami problems. Their objectives are to learn more about the generation, propagation, and on-shore run-up mechanisms of seismic sea waves (tsunamis). And they are helping to improve NOAA's Pacific tsunami warning system. In cooperative research with the University of Hawaii, they have developed several types of deep-sea tsunami gauges, which are sensitive enough to detect a 1-centimeter tsunami wave or tidal change from the deep ocean bottom without interference from surface choppiness.

Finally, we have the Marine Minerals Technology Center whose major concern at the present time is offshore mining. MMTC is developing techniques for bringing materials from the ocean bottom to a surface vessel through pipes and for drilling beneath the bottom. The possibility of bringing up mineral rich (cobalt, nickel, and copper) manganese deposits up from 10,000-15,000-foot depths is another of their objectives. Large sections of the ocean floor are literally covered with manganese nodules and pavements.

MMTC has started a several year program off Massachusetts aimed at determining the environmental effects of offshore dredging and mineral extraction. They are running a series of aquarium studies on the effects of various kinds and degrees of turbidity on fish and shellfish. Representatives of MMTC have observed environmental disturbances of sand and gravel dredging operations in the United Kingdom, where 20 percent of those construction materials come from more than three miles offshore.

The center is a relatively small operation at present, but with a growing national need for minerals that are becoming more difficult and expensive to find and extract from the ground, we expect the MMTC to grow considerably in the coming decade.

In conclusion let me simply say that I hope that my presentation has given you a better impression of the work of the Environmental Research Laboratories of NOAA and especially of the critical importance we assign to remote sensing. I sincerely hope that you find the course that is about to begin profitable and that the visitors among you enjoy your stay in Boulder.

Chapter 1 THE ATMOSPHERIC ENVIRONMENT

Robert L. Grossman

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

A brief presentation of atmospheric composition, basic equations and definitions used in the atmospheric sciences, atmospheric general circulation, atmospheric energy budget, and turbulence theory is given. These are broadly related to remote sensing. The purpose of this chapter is to provide background material for later, more detailed presentations.

1.0 Introduction

The atmosphere is the gaseous envelope of the earth. It extends from the surface of the earth to several earth radii where it merges with the interplanetary medium. Most of the atmosphere's mass is below a height of 10 km. Compared with the radius of the earth (6371 km), the bulk of the atmosphere is contained in a relatively thin envelope. However, when this collection of viscous gases, water vapor, and floating particulate matter is acted upon by an equator-to-pole air temperature gradient, the evaporation and condensation of water, the rotation of the earth, and surface friction, it becomes a very complex physical system. In the midlatitudes our main appreciation of this physical system manifests itself in the passage of weather systems. While weather systems are relatively short term, rarely lasting more than seven days in any given location, they have a very large impact upon our way of life. The longer term effects of the atmospheric system are manifest in what we experience as climate.

One of the continuing goals of atmospheric scientists is to be able to predict the motions of the atmospheric fluid over a wide range of *scales*. These scales include small scale turbulence (1 cm to 1 km), mesoscale systems such as thunderstorms, tornadoes, and squall lines (~1 km to 100 km), synoptic scale characteristics such as weather systems (100 km to several thousand kilometers) as well as climatic trends which occur over periods of hundreds of years and have a scale of tens of thousands kilometers. It is implied in this discussion that time scales and length scales are closely associated in the atmosphere, a fact which will be further explored in 1.7.4 and 2.1.

The purpose of this chapter is to introduce the general characteristics of the atmospheric environment and to broadly associate remote sensing with the continuing investigation of the atmospheric system.

1.1 Gaseous Constituents

Table (1.1) shows the chemical composition of the atmosphere. Almost 99% of the mixture consists of nitrogen and oxygen resulting in a mean molecular weight of $28.96 \text{ g-mole}^{-1}$ at sea level. The mean molecular weight of the atmosphere is shown in (F1.1) as a function of height. We see that the molecular weight is constant to about 90 km, indicating a constant relative proportion of the gases in the mixture. This region is called the *homosphere* and is characterized by fluid motion-scales, chaotic in nature, which are much larger than the mean free path of the individual molecular motions. Above this level the gas mixture is affected by molecular diffusion processes, since the low density results in a mean free path which is larger than the fluid motion scales and individual molecular motion predominates. Thus the proportionality of each gas begins to vary with height so that only the lightest gases are found at successively higher altitudes. This region is called the *heterosphere*.

1.2 The Equation of State

In the lower 50 km of the homosphere, the gaseous mixture can be treated as an ideal gas with very little error. This allows the application of the equation of state for ideal gas to the bulk of the atmosphere. The equation is

$$p = R_d \rho T \quad (1:1)$$

where p is the atmospheric *pressure* in newtons $- m^{-2}$, ρ is *density* in $g - m^{-3}$, T is *absolute temperature* in degrees kelvin (for example, 273K) and R_d is the *specific gas constant for dry air* equal to 2.8704×10^{-1} joules $g^{-1} K^{-1}$. While atmospheric scientists follow SI units when dealing with ρ and T , it is convenient for them to introduce a unit for p known as the *millibar*. A millibar (mb) is equal to $10^{-4} N - m^{-2}$.

The most commonly measured state variables in atmospheric science are pressure and temperature, since the measurement of density is very difficult. Certain remote sensing techniques are able to measure density and temperature; at present these techniques are not economical. The importance of (1:1) is that it allows manipulation of many equations describing atmospheric physics into forms which are easily handled by commonly measured variables. This will become evident in subsequent chapters.

Table 1.1 Normal Composition of Clean, Dry Atmospheric Air Near Sea Level

Constituent gas and formula	Content, percent by volume	Content variable relative to its normal	Molecular weight*
Nitrogen (N ₂)	78.084	—	28.0134
Oxygen (O ₂)	20.9476	—	31.9988
Argon (Ar)	0.934	—	39.948
Carbon dioxide (CO ₂)	0.0314	†	44.00995
Neon (Ne)	0.001818	—	20.183
Helium (He)	0.000524	—	4.0026
Krypton (Kr)	0.000114	—	83.80
Xenon (Xe)	0.0000087	—	131.30
Hydrogen (H ₂)	0.00005	?	2.01594
Methane (CH ₄)	0.0002	†	16.04303
Nitrous oxide (N ₂ O)	0.00005	—	44.0128
Ozone (O ₂)	Summer: 0 to 0.000007	†	47.9982
	Winter: 0 to 0.000002	†	47.9982
Sulfur dioxide (SO ₂)	0 to 0.0001	†	64.0628
Nitrogen dioxide (NO ₂)	0 to 0.000002	†	46.0055
Ammonia (NH ₃)	0 to trace	†	17.03061
Carbon monoxide (CO)	0 to trace	†	28.01055
Iodine (I ₂)	0 to 0.000001	†	253.8088

* The content of the gases marked with an asterisk may undergo significant variations from time to time or from place to place relative to the normal indicated for those gases.
From the U.S. Standard Atmosphere, 1962

1.3 Effect of Water Vapor on Air Density

Water vapor is considered a separate atmospheric constituent and thus is not listed in (T1.1) as part of the gaseous mixture. Compared with nitrogen, the percentage of water vapor present in the atmosphere is very small. At sea level in the most humid tropical environments, the *absolute humidity*, ρ_v , rarely reaches $25 g/m^3$ which is to be compared with the typical density of dry air, ρ , of $1000 g/m^3$.

However, water vapor is the most variable and most important of atmospheric constituents. Its variability is due to the fact that it is not easily mixed on a global scale but is closely related to its sources (tropical oceans) and its sinks (midlatitude continents and oceans). The importance of water vapor in atmospheric physics is manifest in its ability to change phase from gas to liquid to solid within atmospheric pressure and temperature ranges, producing *hydrometeors* such as rain, snow, and hail. Furthermore, certain remote sensors utilize a sensitivity of the radio refractive index, N , to water vapor. This will be explored in the appendix.

Even though the water vapor in a given parcel† of air is generally between 0.1 and 2.5% of the total mass, it has a significant effect upon the air density. Therefore the equation of state, (1:1), must reflect the presence of water vapor in the air. The reciprocal of the mean molecular weight of moist air, \bar{m} , can be written

$$\frac{1}{\bar{m}} = \frac{1}{M_d + M_v} \left[\frac{M_d}{m_d} + \frac{M_v}{m_v} \right] \tag{1:2}$$

where M is the mass of each constituent in the mixture and m is the molecular weight of the constituent. The v subscript denotes water vapor and the d subscript denotes dry air. Equation (1:2) can be reduced to

$$\frac{1}{\bar{m}} = \frac{1}{m_d} \left[\frac{1 + w/\epsilon}{1 + w} \right] \tag{1:3}$$

where $\epsilon = \frac{m_v}{m_d} \approx 1.607$ and $w = \rho_v/\rho_d$. The variable w is known as the mixing ratio of water vapor and is often given in units of grams per kilogram.

Returning to (1:1) we have

$$p\alpha = R_d T = \frac{R^*}{m_d} T \tag{1:4}$$

where α is the specific volume of the gas (i.e., the reciprocal of density) and R^* is the universal gas constant. To include the effect of water vapor, all that is necessary is to substitute \bar{m} for m_d to give

$$p\alpha = \frac{R^*}{\bar{m}} T = \frac{R^*}{m_d} \frac{1 + w/\epsilon}{1 + w} T . \tag{1:5}$$

If the terms in parentheses appearing in (1:5) are grouped with T, (1:5) can be written

$$p\alpha \cong 1.607 R_d T_v . \tag{1:6}$$

T_v is called the virtual temperature and is expressed by

$$T_v = T \left(\frac{1 + w/\epsilon}{1 + w} \right) . \tag{1:7}$$

Physically, T_v can be interpreted as the temperature a sample of dry air at a given pressure would have to achieve to have the same density as a parcel with a mixing ratio of w at the same pressure.

Equation (1:5) shows that the effect of water vapor is to cause a moist parcel of air to be *less dense* than a parcel of dry air at the same temperature and pressure.

1.3.1 Moisture Variables.

Table (1.2) lists some of the more commonly used moisture variables. With the use of specific humidity the equation for virtual temperature can be approximated by

$$T_v = T(1 + 0.61q) . \tag{1:8}$$

1.4 Aerosols and Hydrometeors

There are two further classes of constituents which make up the atmospheric environment: aerosols and hydrometeors. *Aerosols* are particulate matter suspended or slowly falling in the gaseous mixture of the atmosphere. While aerosols can be either liquid or solid in phase, they are differentiated from hydrometeors chiefly in terms of size and content. *Hydrometeors* are the solid or liquid phase forms of water in the gaseous atmosphere. Hydrometeors are usually much larger than aerosols. Examples of aerosols are smokes, haze, clouds, and fogs whose droplet size is below 1 micron in radius, and dispersed, finely divided soils.

† A parcel is defined as “an imaginary body of air to which may be assigned any or all of the basic dynamic and thermo-dynamic properties of atmospheric air” (Glossary of Meteorology). Further, parcel properties are uniform within the parcel and its motions do not induce motions in its atmospheric environment.

Hydrometeors are generally considered as the end products of precipitation processes in the atmosphere, this particle class includes mists, rain, freezing rain, ice pellets, snow, hail, blowing snow (i.e., lifted from the earth's surface), blowing spray as well as those forms of clouds and fogs which have droplet radii greater than 1 micron.

There is a very complex relationship between aerosols and hydrometeors which is a subject of the study of cloud microphysics. The initial stages of the formation of hydrometeors requires two conditions; the first is that a given volume of moist air be very slightly supersaturated and the second is the presence of small particles known as *hygroscopic nuclei* or Aitken nuclei which act to initiate the condensation of the supersaturated vapor forming the nascent hydrometeor.

Figure (1.2) presents a distribution of aerosols in a continental air mass. The Aitken nuclei can be considered the initial nuclei present in the atmosphere. Their source is ultimately the earth's surface.† Coagulation processes transform a small percentage of the Aitken nuclei into nuclei with larger radii by causing several small particles to stick to one another. These particles are called coagulation nuclei. Dispersion nuclei are those large radius particles which are borne aloft by strong surface mixing. These nuclei are generally very finely divided soils whose radius are smaller than 0.1 micron. Maritime nuclei are usually very large and are thought to be a result of rapid evaporation of sea spray and breaking bubbles on the sea surface which leave a small solid residue aloft composed of sea salts.

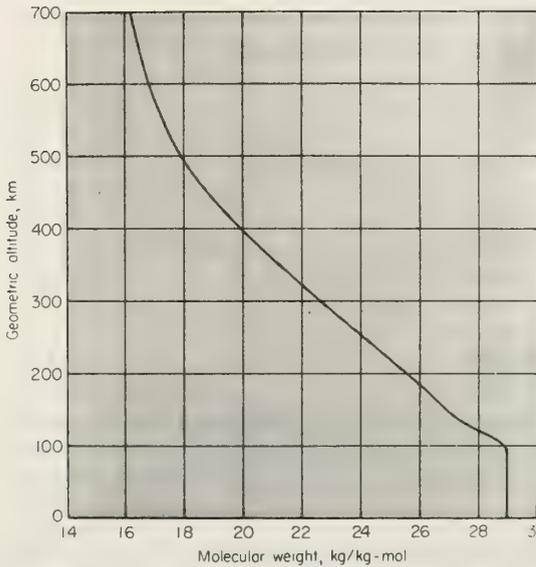


Figure 1.1 Mean molecular weight of dry air as a function of geometric altitude (from U.S. Standard Atmosphere, 1962).

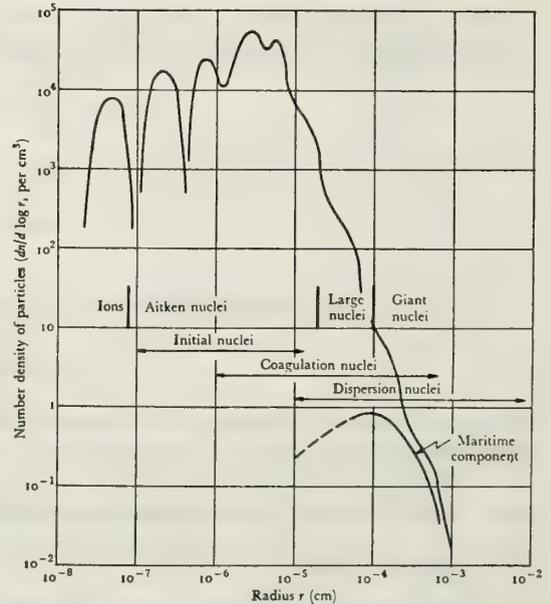


Figure 1.2 Average size-distribution of nuclei in continental air masses (after Junge, 1952).

The sources of aerosols in the atmosphere are generally classed into gas-phase reactions, combustion products, and dispersion of solutions. Gas-phase reactions are chemical reactions of various gases in the atmosphere, often very minor constituents such as nitrous oxide, terpenes, and hydrocarbons, which produce solid particles. Combustion products, usually the result of inefficient burning, are a large source of aerosol content. However, on a global scale natural sources of combustion such as forest fires and volcanism are larger than man-made combustion products. Man-made combustion products can be very crucial locally and may even be responsible for modifying rainfall patterns in local areas (Changnon et al., 1971).

Aerosols and hydrometeors are very important to techniques of remote sensing whether acoustic, radio, or optical. While the effect of these additional constituents of the atmosphere on remote sensing techniques is very complex and will be covered in detail in later chapters, some general comments about the nature of their influence can be simply stated. Aerosols and hydrometeors can be looked at from two points of

† Some workers hypothesize that meteor showers produce significant amounts of initial nuclei (Lapaz, 1958).

view by the scientists using a remote sensor. Firstly, they may be regarded as *environmental noise*: that is, elements which tend to confuse the signal received by the sensor and which originate outside the sensor electronic system. In this context, aerosols and hydrometeors are a nuisance and often require corrections to the raw data to reduce the effects of their presence in the received signals. On the other hand, aerosols and hydrometeors can be utilized in remote sensing techniques as *tracers*. In this case their effects on the received signal are enhanced so that motions, distributions, and chemical compositions of the aerosols or hydrometeors being “interrogated” by the sensor can be analyzed.

Both of the points of view come together in a mutual effort to determine exactly the effect of the media upon the propagation of acoustic, radio, and optical waves. It will be seen in subsequent chapters that a great deal of experimental and theoretical effort has been spent in determining the absorption and scattering characteristics of aerosols and hydrometeors.

1.5 The Vertical Structure of the Atmosphere: Stability and Mixing

In 1.1 a discussion of the vertical structure of the atmosphere was presented which was based upon chemical considerations. This type of vertical structure classification gave an idea of the vertical extent over which (1:1) or (1:6) could be used with R_d constant. However, rockets fitted with temperature measuring devices as well as similarly instrumented balloons and high-flying aircraft, revealed a statistically consistent vertical variation of temperature. This led to a classification of the vertical structure of the atmosphere based upon the variation of temperature with height.

Before discussing this classification, a theoretical presentation of the vertical variation of pressure and temperature would be helpful, since it provides physical insight into atmospheric motions. In the case of the vertical pressure gradient a useful approximation will be presented. The vertical temperature gradient on the other hand, provides valuable criteria for vertical stability in the atmosphere.

1.5.1 Variation of Pressure with Height: the Hydrostatic Equation.

The vertical variation of pressure with height is an approximation of the vertical equation of atmospheric motion (Section 1.9, (1:42)). The following assumptions apply to a parcel of air:

1. No vertical acceleration of the parcel.
2. If the parcel is in motion, frictional or stress forces are small.
3. Forces due to the rotation rate of the earth are negligible.

With these assumptions, a force diagram can be drawn for the parcel as in (F1.3).

Since the sum of the forces must be equal to zero,

$$-(p_o + dp)dx dy + p_o dx dy - g \rho dx dy dz = 0. \tag{1:9}$$

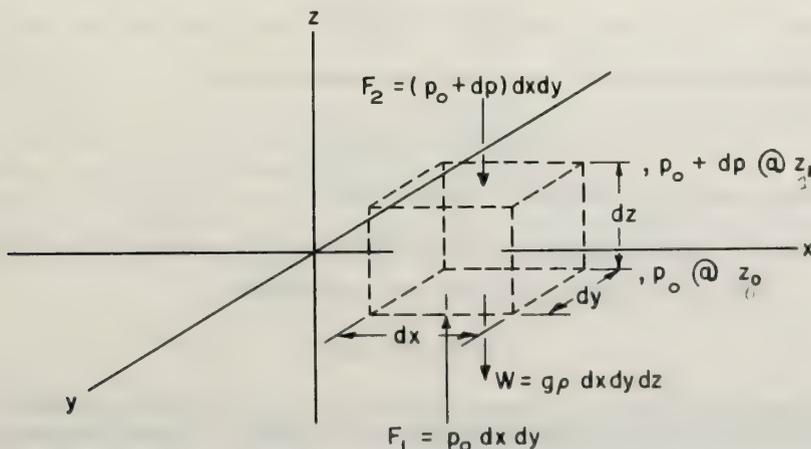


Figure 1.3 Free body diagram of an atmospheric parcel under the influence of pressure and gravity.

Table 1.2 Commonly Used Moisture Variables

Variable	Equation	Dimensions	Definitions
Absolute humidity	$\rho_v = w\rho_d$	$g\ m^{-3}$	$\rho_d \equiv$ dry air density
Mixing ratio	$w = \rho_v/\rho_d$ $w = \epsilon \frac{e}{p - e}$ $w \cong \epsilon \frac{e}{p}$	$g\ kg^{-1}$	$e \equiv$ partial pressure of water vapor (usually in millibars).
Relative humidity	$r = w/w_s \times 100$	per cent	$w_s \equiv$ mixing ratio at saturation. <i>Saturation</i> is when the parcel can hold no more water vapor without some condensation.
Specific humidity	$q = \epsilon \frac{e}{p}$ $q \cong w$	$g\ kg^{-1}$	

Here p_0 is the atmospheric pressure at level z_0 , dp is the difference in pressure between levels z_0 and z_1 , and g is the acceleration due to gravity. Equation (1:9) reduces to

$$\frac{dp}{dz} = -\rho g \quad (1:10)$$

This is a very important relationship in the study of the atmosphere and is called the *hydrostatic equation*. By use of (1:6) this relationship becomes

$$\frac{dp}{p} = -\frac{g}{R_d \bar{T}_v} dz \quad (1:11)$$

If a mean, virtual temperature, \bar{T}_v , is assumed for a layer of the atmosphere from Z_0 to Z , (1:11) can be integrated exactly to give

$$\frac{p}{p_0} = \exp \left\{ \frac{-g\Delta z}{R_d \bar{T}_v} \right\} \quad (1:12)$$

Taking p_0 as the surface pressure (1:12) can be written in its familiar form,

$$\frac{p}{p_0} = \exp \left\{ \frac{-gz}{R_d \bar{T}_v} \right\} \quad (1:13)$$

which indicates that pressure decreases logarithmically with height.

1.5.2 The Vertical Variation of Temperature: Potential Temperature, the Adiabatic Lapse Rate and Stability.

The first law of thermodynamics is functionally described by

$$dQ = c_p dT - \alpha dp \quad (1:14)$$

Here c_p is the specific heat of dry air at constant pressure and dQ is an increment of *sensible heat*. We can derive an important atmospheric variable called the potential temperature by assuming that (1:14) describes an adiabatic process. An *adiabatic process* is one in which no sensible heat is exchanged into or out of the thermodynamic system under consideration. In this case the thermodynamic system is an atmospheric parcel.

With this assumption (1:14) becomes

$$c_p dT - \alpha dp = 0 \quad (1:15)$$

Use of (1:1) allows

$$\frac{dT}{T} - \frac{R_d}{c_p} \frac{dp}{p} = 0 \quad (1:16)$$

which upon integration yields

$$\frac{T}{T_0} = \left(\frac{p}{p_0} \right)^{R_d/c_p} \quad (1:17)$$

If P_0 is taken to be 1000 millibars and the notation for T_0 at that pressure is changed to θ we have,

$$\theta = T \left(\frac{p}{1000} \right)^{R_d/c_p} \quad (1:18)$$

which is the defining equation for the *potential temperature*, θ . The potential temperature is a *conservative quantity* (i.e., does not change with either time or displacement) for adiabatic processes in the atmosphere and for this reason is often utilized in tracing atmospheric motions. Physically, potential temperature is the temperature a parcel of dry air at pressure, p , would have if it were brought adiabatically to a pressure of 1000 millibars. Since *entropy*, dQ/T , is invariant in an adiabatic process, these processes and the atmospheric motions associated with them are known as *isentropic processes and motions*.

A fundamental measure in atmospheric science is the *temperature lapse rate*, or the vertical variation of temperature, dT/dz . It is an important concept in the description of atmospheric stability. From (1:16) it is apparent that a dry adiabatic process has a distinct lapse rate associated with it. Combining (1:11) and (1:16) gives, upon assuming no water vapor,

$$\frac{dT}{T} + \frac{R_d}{c_p} \frac{g}{R_d T} dz = 0 \quad (1:19)$$

which becomes

$$\Gamma_d \equiv - \frac{dT}{dz} = + \frac{g}{c_p} = 9.76^\circ \text{K/km} \quad (1:20)$$

where Γ_d is known as the dry adiabatic lapse rate. Be aware of the notational convention presented here which holds for any discussion of temperature lapse rate in atmospheric science. Since, in general, temperature in the lower atmosphere decreases with increasing altitude, the following convention is used

$$\gamma \equiv - dT/dz$$

where γ is defined as the lapse rate of temperature. Therefore negative γ indicates increasing temperature with increasing altitude.

Two important points should be made with reference to (1:20). The first is that due to the adiabatic assumption, an atmospheric layer with $\gamma = \Gamma_d$ has no sensible heat flux into or out of the layer. The second point is that since θ must be constant in an adiabatic process, layers which have $\gamma = \Gamma_d$ also have $d\theta/dz = 0$. It is often stated that layers with the above properties are *well mixed layers*.

We now turn to a discussion of atmospheric stability which will point out the importance of (1:20). Consider a system which consists of an air parcel and its surrounding environment. Environmental properties will be denoted by a prime ('). Assuming the environmental air to have no vertical accelerations (1:10) becomes

$$-\alpha' \frac{dp'}{dz} - g = 0 \quad (1:21)$$

However, the vertical motion of the air parcel will be the subject of the present discussion so the description of its motion is

$$-\alpha \frac{dp}{dz} - g = \frac{d^2 z}{dt^2} \quad (1:22)$$

A further assumption is that in a displacement of the parcel vertically from its equilibrium condition, the pressure of the parcel *immediately* adjusts to the pressure of the environment. Thus (1:21) can be written as

$$-\alpha' \frac{dp}{dz} - g = 0 \quad (1:23)$$

Eliminating dp/dz from (1:22) and (1:23) results in

$$g \left(\frac{\alpha - \alpha'}{\alpha'} \right) = \frac{d^2 z}{dt^2} \quad (1:24)$$

Use of (1:1) in (1:24) gives

$$\frac{d^2 z}{dt^2} = g \left(\frac{T - T'}{T'} \right) \quad (1:25)$$

Exploring small displacements of the parcel about an equilibrium position, z_0 , allows a Taylor series expansion of (1:25). Applying a few minor approximations, (1:25) becomes a second order differential equation with constant coefficients

$$\frac{d^2 z}{dt^2} + \frac{g}{T_0} (\Gamma_d - \gamma) z = 0 \quad (1:26)$$

which is the fundamental equation of static stability in a dry atmosphere. Since $d\theta/dz = \Gamma_d - \gamma$ (an exercise left to the reader), examination of (1:26) leads to the following *criteria of stability* in a dry atmosphere (summarized in T1.3 below).

The frequency of oscillation for the stable case can be obtained from the characteristic of (1:26) and is

$$f_s = \sqrt{\frac{g}{T} (\Gamma_d - \gamma)} \quad (1:27)$$

which is also known as the *Brunt-Väisälä frequency*.

With this background the vertical temperature structure of the atmosphere, which is presented in (F1.4), can be discussed with some physical insight. However, it is necessary to slightly expand upon the concept of vertical mixing in the atmosphere. There are three basic classes of physical phenomena which effect vertical transport of scalar or passive atmospheric properties such as temperature and water vapor. These are *convective mixing* in which the atmosphere is mixed due to gravity acting upon vertical density anomalies, *mechanical mixing* in which inertial forces rather than gravitational forces are dominant, and *radiative transfer*

Table 1.3 Criteria for Vertical Stability in a Dry Atmosphere.

Potential Temperature Gradient	Temperature Gradient, γ	Remarks on Equation (1:26)
I. $d\theta/dz < 0$	$\gamma > \Gamma_d$ (super adiabatic lapse rate)	The solution is <i>unstable</i> . Vertical acceleration of the parcel in the direction of displacement from z_0 will result.
II. $d\theta/dz > 0$	$\gamma < \Gamma_d$ (sub adiabatic lapse rate)	The solution is <i>stable</i> . The parcel will oscillate, in the absence of friction, about z_0 with a frequency, f_s .
III. $d\theta/dz = 0$	$\gamma = \Gamma_d$ (adiabatic lapse rate)	The solution is <i>neutral</i> . No acceleration of the parcel will result upon a displacement from z_0 .

processes. An example of convective mixing would be vertical transports which are the result of a large thunderstorm or, on a global scale, the Hadley cell (see 1.7.2). An example of mechanical mixing would be vertical transports due to flow with large vertical gradients of horizontal wind speed or waves caused by flow over mountains (lee waves) which become unstable and break. These mixing processes are active over nearly the entire range of atmospheric motion scales (see 1.7.4 for discussion of motion scales).

Radiative transfer processes in the atmosphere are very complex. These processes are dependent upon minor atmospheric constituents such as water vapor, CO₂, O₃ aerosols and hydrometeors. Radiative transfer processes tend to produce non-neutral lapse rates (*adiabatic lapse rates*), either stable or unstable, depending upon the vertical structure of the constituents which dominate the radiative process. In contrast mechanical mixing and convective mixing tend to bring the lapse rate to the neutral condition.

From (F1.4) we note that the *troposphere* is very slightly stable with a mean $\gamma = 6.5^\circ\text{K}/\text{km}$. Since a $\gamma = 9.76^\circ\text{K}/\text{km}$ indicates a very well mixed layer, the troposphere lapse rate denotes a large amount of convective and mechanical mixing in this layer of the atmosphere. The *tropopause* indicates the end of the troposphere and the beginning of the *stratosphere*; it is an example of a *substantial surface* or *surface of discontinuity* (note the high value of d^2T/dz^2). There is very little vertical mixing across a substantial surface; therefore, one can conclude that stratospheric air has quite different properties than tropospheric air even though both are in the homosphere. The tropopause is not continuously present over the entire globe so that some stratospheric air does reach the earth's surface. This is particularly true in the region of jet streams. The stratosphere often has two distinct regions within it: an isothermal region (i.e., temperature in constant with height) and a region where $\gamma < 0$ (i.e., temperature is increasing with height). While both regions are stable according to (T1.3), the lower region lapse rate is indicative of more mixing motion than in the upper region. Radiative transfer processes probably dominate the vertical temperature variation in the upper, more stable region.

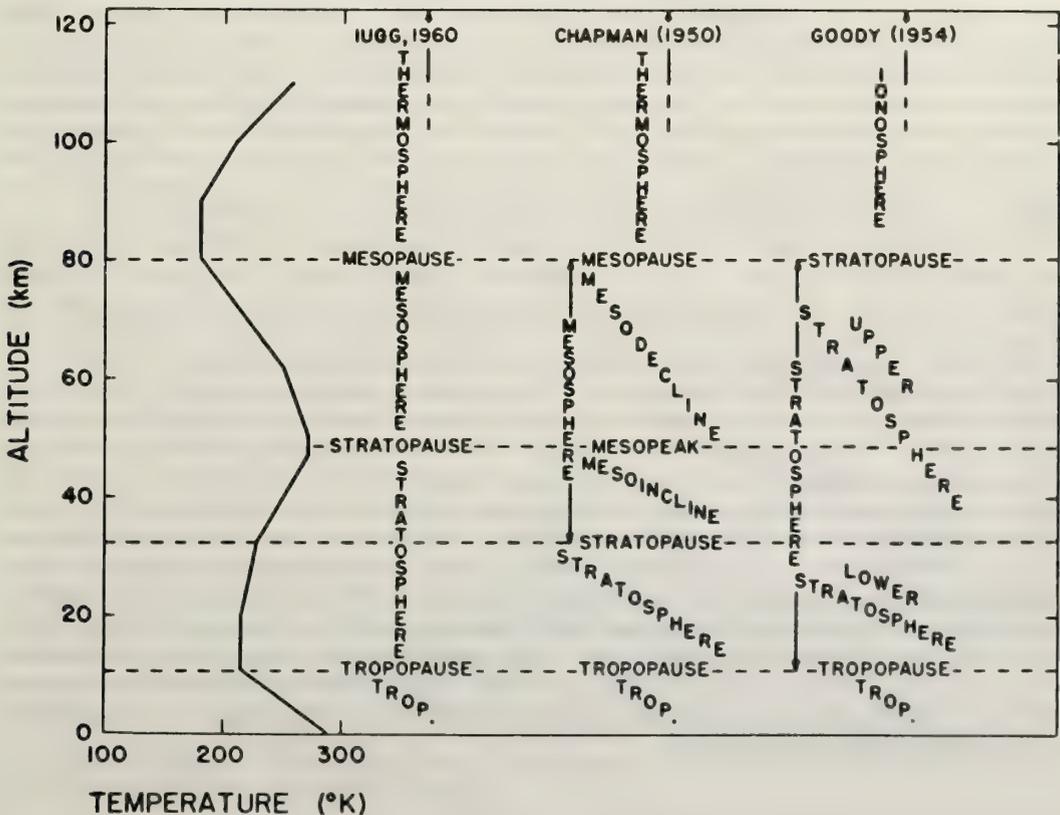


Figure 1.4 Vertical distribution of temperature up to 110 km according to the U.S. Standard Atmosphere, 1962, and common systems of nomenclature (from Craig, 1965).

The *mesosphere* lapse rate approaches that of the troposphere indicating that once again some kind of vertical mixing process is present. In the *thermosphere*, molecular diffusion becomes a dominant process (see 1.2) and the layering of the molecules forces stability which is reflected in the lapse rate.

1.5.3 Importance of the Troposphere and Planetary Boundary Layer: Parameterization of Sub-Grid Scale Phenomena

This course will emphasize the physical processes of the troposphere. Since we live in the troposphere, knowledge of its characteristics are undoubtedly of great benefit to mankind. Even though there is evidence of tropospheric-stratospheric coupling and some evidence of extra-terrestrial influences upon tropospheric motions (generally in the form of slight variations of incoming solar radiation and tides in the atmosphere), the troposphere is very nearly a closed atmospheric system.

Thus, characteristics of this layer can be quantitatively deduced from measurements of variables within the layer itself. This quantitative deduction is made possible by simulating the general circulation of the atmosphere (see 1.6 and Chapter 2) using basic or primitive equations which define: a) mean atmospheric motions, b) thermodynamic properties including phase change of water, c) conservation of mass, with the restriction upon the equations that the total system conserve energy.

A further advantage of this kind of atmospheric study is that the primitive equations (see 1:42, 1:53 and Chapter 2) are time dependent. Therefore, motion fields and hydrometeor fields can be predicted, or forecast. Since weather is defined as the short term variation (minutes to months) of the state of the atmosphere, the use of these equations permits advance notice of weather variations, or *weather forecasting*. The use of this ability to mankind is obvious.

However, the use of these primitive equations is severely hampered by inadequate knowledge of a) the initial conditions defining the atmospheric system and b) the boundary conditions acting upon the system. This course will concentrate upon the latter of these two serious problems. Boundary conditions which affect the equations are active over a layer of the atmosphere known as the *planetary boundary layer* or *Ekman layer*. Normally encompassing the lower 3 km of tropospheric air, it is this layer that is coupled to the surface of the earth through surface friction, or drag.† Equally important, however, is the fact that through this layer pass the initial energy and momentum sources and sinks which characterize the general circulation. Knowledge of the spatial and temporal variation of boundary layer exchange processes and the methods of including such processes in the primitive equations are the major stumbling blocks to mastery of weather forecasting, especially predictions over relatively long time periods (i.e., about five days). To quote Mason (1971):

“... in the numerical simulation of ... larger scale motions, many smaller scale phenomena have to be parameterized preferably in terms of observable, large scale dependent variables. These phenomena include: molecular scale processes, such as radiative transfer as a function of the larger scale distributions of CO₂, O₂, water vapor, cloud, and aerosol particles; microscale processes such as turbulent exchanges of heat, momentum in the planetary boundary layer, internal turbulent diffusion and dissipation, and convective transports. . . .”

The chief manifestation of the problem is given by the first part of the above quotation. The use of the primitive equations is simplified by calculating solutions to the equations at specified grid points which are distributed vertically and horizontally over the globe. It is obvious that in such a scheme the number of grid points is important. Further, the spacing between the grid points determines the number of points used in the calculation of the various horizontal fields of motion, temperature, and moisture at different vertical levels of the atmosphere. At present both the observational system, which provides the initial conditions, and the numerical calculation system, which provides the forecast, are based upon grid points whose distance interval is much larger than motions present in the planetary boundary layer. Because of this spacing the effects of these motions are lost to the calculation even though they are physically very important! Thus one of the principal tasks facing atmospheric scientists today is the parameterization of sub-grid scale phenomena in terms of grid scale variables. This is particularly true of those workers involved with the planetary boundary layer. Many of the remote sensing techniques and most of the results presented in this course are directed to quantifying such parameterization schemes.

† Note: Above the Ekman layer the vertical momentum flux is very small, thus neglected in (1:42) & (1:53). Air motion in this region of the troposphere is *quasi-geostrophic* (see 1.8.2 & (1:62)).

1.6 The Horizontal Structure of the Atmosphere: General Circulation Notions and Scales of Motion

1.6.1 Basic Notions Regarding the Global Atmospheric Circulation System

In this section, the horizontal structure of the atmosphere is examined. Atmospheric motion is three-dimensional and time-dependent. Vertical motions are linked to horizontal motions through the *equation of continuity* or conservation of mass equation,

$$\frac{1}{\rho} \frac{d\rho}{dt} = \nabla \cdot \mathbf{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \tag{1:28}$$

Here $\nabla = \frac{\partial}{\partial x} \hat{i} + \frac{\partial}{\partial y} \hat{j} + \frac{\partial}{\partial z} \hat{k}$, the del operator, and $\mathbf{V} = u\hat{i} + v\hat{j} + w\hat{k}$, the total wind vector (see 1.8.2 for discussion of coordinate system used). If we assume incompressibility†, $d\rho/dt = 0$ and, upon assuming $w = 0$ at $z = 0$, (1:28) becomes

$$w_z = - \int_0^z \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \delta z \tag{1:29}$$

Equation (1:29) states that the vertical wind at level z is proportional to the integral of the divergence‡ of the horizontal wind field to level z . The relationship between the various scales of horizontal and vertical motion and their temporal variability is extremely complex; however, the following general statement can be made: the life cycle of various atmospheric weather phenomena is related to their horizontal scale size. The larger the phenomena, the longer it exists. Thus, as an example, hurricanes last longer than individual thunderstorms while both are products of intense convection. However, certain mean motions are evident on a climatic scale, that is for time periods extending over hundreds of years. For these motions the time-motion scale dependence is small. What we actually see as variations about climatic means are principally due to weather scale motions. In the following discussion of the general circulation we shall see that it is characterized by both weather and climatic scale motions, (1:29), and an additional concept known as the conservation of angular momentum. The conservation of angular momentum is simply stated by

$$\rho_1 \omega_1 r_1 = \rho_2 \omega_2 r_2 \tag{1:30}$$

Here ω is the angular velocity of an air parcel and r is its perpendicular distance from the earth’s axis (greatest at the equator, decreasing to zero at the pole); the subscripts denote a parcel at two different positions in the atmosphere.

1.6.2 A Qualitative Description of the General Circulation of the Atmosphere

This discussion will be limited to the northern hemisphere with the understanding that the mechanisms discussed are broadly symmetrical about the latitude 5°N also known as the *heat equator*. Certain differences do exist, chiefly because the predominance of the earth’s land mass in the northern hemisphere. These differences are covered in advanced texts on meteorology and general circulation theory.

Energy in the form of latent and sensible heat†† is exported from approximately 30°N southward within the planetary boundary layer by the world’s most constant wind system, the trade wind regime. This air carries angular momentum with it as well, since it is coupled to the rotating earth via surface friction. The air reaches a region of convergence closely associated with the heat equator known as the Intertropical Convergence Zone (ITCZ or ITC). Here according to (1:29) this very moist, momentum laden boundary layer air is lifted to form very large convective clouds which often aggregate to form cloud systems. This intense and widespread convective activity “pumps” much of the energy and momentum gained in the trade regime to

† Incompressibility in the atmosphere is reasonable except in considering motion fields in deep convective clouds.

‡ The mathematical quantity $(\partial u/\partial x + \partial v/\partial y)$ is known in atmospheric science as horizontal divergence of the wind field when positive and horizontal convergence when negative.

†† Latent heat in the sense used here is heat absorbed in the process of evaporation of water. When the water vapor is later condensed, latent heat is released to warm the atmosphere. Sensible heat is that heat energy exchange which directly warms or cools the atmosphere (see 1.6.2).

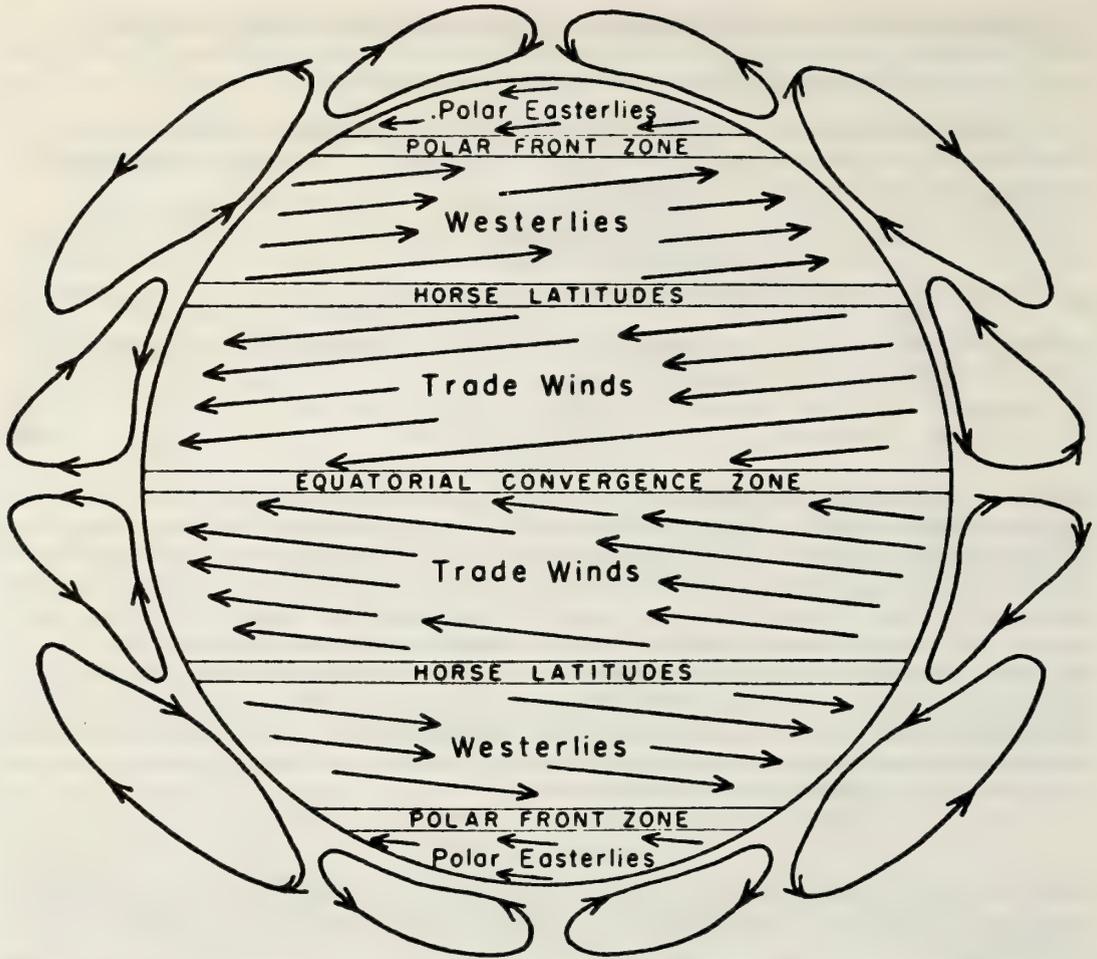


Figure 1.5a Schematic representation of the general circulation, modified after Bergeron and Rossby (from Byers, 1959).

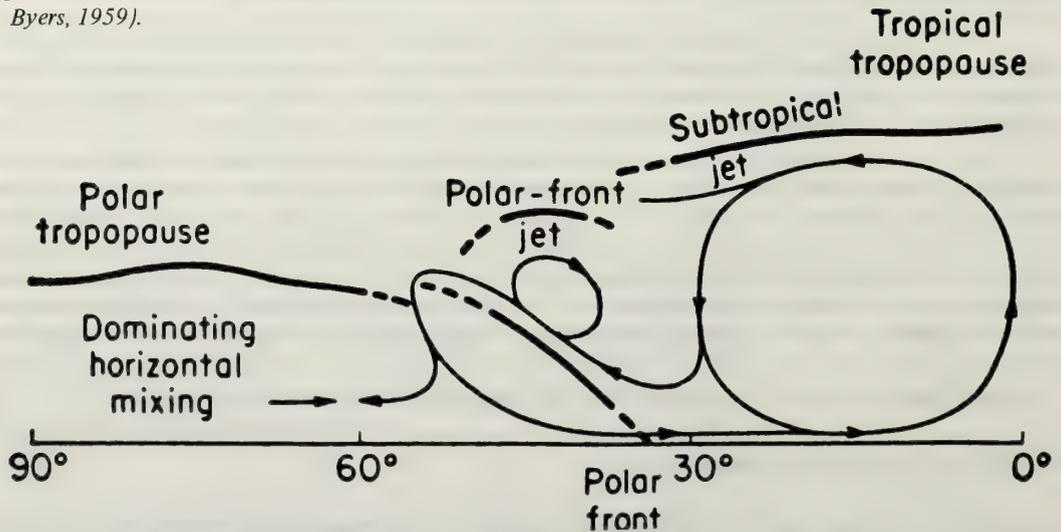


Figure 1.5b Schematic of a meridional circulation scheme after Palmén (from Byers, 1959).

levels near the tropopause. At this level through mechanisms not yet fully understood this energy and momentum is dispersed from the clouds and begins to move northward under the influence of a climatic scale motion system known as a *meridional cell*. The existence of this cell, elaborated in (F1.5a and b), was postulated by the English scientist, Hadley, in the 1700's! Thus, this circulation is often called a *Hadley Cell*.

The air moves northward until it reaches about 30°N once again. The momentum gained by the air in its journey to the ITCZ is now reflected by an increase in horizontal wind speed in accordance with (1:30), since r is becoming smaller. The conservation of angular momentum mechanism is coupled with another mechanism affecting horizontal wind speed which is dependent upon horizontal temperature gradients known as the thermal wind. With warm air to the south and cold air to the north, this mechanism tends to increase the westerly component of the horizontal wind, u . The influence of these two mechanisms at 30°N results in a very concentrated band of high speed winds known as a *jet stream*, which occurs at or near tropopause level. The atmospheric jet stream system (its corollary in the ocean is the Gulf Stream and the Kuro Siwo Current systems) is very complex and not yet fully understood. However, it does seem to transfer heat, kinetic energy, and momentum into the midlatitudes. This transfer is most obvious to us when it comes in the form of weather systems.

Large scale, or synoptic weather systems (see F1.9 for definition of atmospheric scales) are with few exceptions controlled by the horizontal wind field in the near tropopause region. These weather systems are considered sinks of kinetic energy and momentum in the general circulation; furthermore, they transfer sensible heat northward in the northern hemisphere. The important sink of kinetic energy is affected by boundary layer processes which bring high momentum, high energy air to the earth's surface where the momentum and energy are dissipated by friction forces. Thus we see through this qualitative explanation that the trade regime is an energy source and the midlatitude westerlies are an energy sink for the general circulation. Furthermore, boundary layer processes are crucial both to providing the energy and dissipating energy, though source and sink are separated by thousands of kilometers.

1.6.3 Latitudinal Irradiance and the General Circulation

The ultimate source of energy for the general circulation is solar radiation. Viewed from space the earth is very nearly in thermal equilibrium with incoming short wave radiation from the sun (with wavelength, λ , between 0.4μ and 0.7μ) balancing outgoing longwave, or infrared radiation, ($0.8\mu < \lambda < 30\mu$) from the earth. For our purposes of discussion, we can regard the solar energy that reaches the outer limits of the atmosphere as relatively constant.

The amount of solar radiation which eventually reaches the surface of the earth, if averaged over a period of a year, is a strong function of latitude. This is mainly the result of the tilt of the earth's axis away

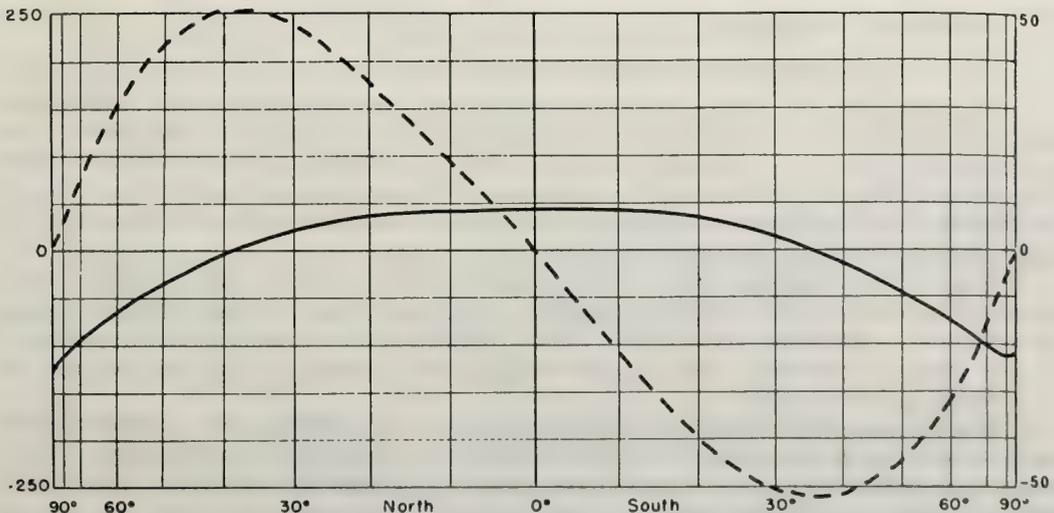


Figure 1.6 Excess of absorbed solar radiation over outgoing infrared radiation (solid curve), in Watts - m^{-2} (left scale); and northward transport of energy by the atmosphere and oceans required for balance (dashed curve) in units of 10^{14} Watts (right scale) (from Lorenz, 1967).

from the earth-orbital plane, with secondary effects contributed by surface reflectivity and atmospheric transmissivity of the incoming radiation. The outgoing infrared radiation, if averaged over a yearly period, is very nearly constant with latitude. Therefore, the excess radiation absorbed by the earth's surface (primarily in the oceanic regions) also has a heavy latitudinal dependence which is shown in (F1.6).

From (F1.6) we see an excess of solar radiation generally between 30°N and 30°S with a deficit from about 50°N and 45°S to the respective poles. This energy gradient is reflected mainly by temperature excess (above the mean earth temperature) in the trade wind regions and temperature deficit from the midlatitude regions to the poles. The general circulation of the atmosphere (and to some extent the oceans) is nature's way of reducing this gradient of energy. The solid line in (F1.6) shows the amount of energy transport necessary to balance the deficit at the poles.

The discussion given in 1.6.2 shows that the net energy excess received by the earth in the trade wind regions is transformed into many different forms. Figure (1.7) provides insight into the total energy balance of the atmosphere and shows the various forms the incoming energy from the sun acquires. In (F1.7) those blocks which are directly applicable to the discussion of the general circulation are denoted by Roman numerals. Subtracting II from I gives a global estimate of the net energy excess, which is equal to the amount of energy in III. Comparison of III to I shows that only 37% of the available incoming solar radiation is actually received at the earth and is utilized in driving the hydrological cycles and thermodynamic circulation systems. However, most of that available energy is returned to the reservoir from which it came in the form of precipitation (hydrometeors) and sensible heat transfer; therefore, approximately 1% of the available energy in I is actually utilized in the kinetic energy of the flow (i.e., the world wind systems) shown in IV. This kinetic energy is dissipated at the earth's surface by boundary layer processes (i.e., V in F1.7) and returned to the reservoir of internal energy as heat. Thus, the atmosphere "heat engine" is very inefficient if we consider the motion of the fluid as its primary manifestation.

1.6.4 Horizontal Scales of Motion in the Atmosphere

The general circulation acts to transfer energy and momentum across latitude circles. The major portion of this energy transfer is in the form of "eddies" or perturbations upon some mean flow condition (see 1.8.4 and 1:35). These "eddies" seem to be present in the atmosphere in a wide, apparently continuous spectrum of "sizes" or horizontal scales. These horizontal scales range from large planetary waves (see F1.8), to weather systems which are revealed as cloud spirals in satellite and manned space flight photographs, on to such phenomena as thunderstorms, and finally ending viscous dissipation at millimeter eddy size. These phenomena, their scales of motion and the capability of some "state of the art" remote sensing techniques discussed in this course are presented in (F1.9).

1.7 Averaging in the Atmosphere

1.7.1 Meteorological Approximations and Basic Coordinate Systems

As stated before, the atmosphere is a three-dimensional, time dependent system. However, some atmospheric phenomena are considered as two-dimensional and with the help of radial symmetry even one-dimensional (as in F1.9). This approximation is effected by either neglecting one of the horizontal dimensions and observing a "vertical slice"; or alternatively the vertical dimension can be neglected and the horizontal structure investigated at a given level. Both methods are called a *quasi-two-dimensional* (n.b. "quasi" means "almost") approximation.

A fully three-dimensional look at the atmosphere during which the time dependence of the atmospheric parameters can be investigated is rare. Very often some kind of averaging is either voluntarily performed or is a result of the data acquisition system. Averaging in the atmosphere is usually either spatial (i.e., over space) or temporal (i.e., over time); however, often the investigator is faced with a mixture of the two or a temporal-spatial average.

Most atmospheric sensing systems, including many remote sensors, are fixed in an Eulerian coordinate system, in which the flow moves past the sensor. In this coordinate system the total derivative with respect to time, or *substantial derivative* is estimated by looking at the partial derivative with respect to time plus the *advective terms* as in (1:31) for the scalar, ϕ .

$$\frac{d\phi}{dt} = \frac{\partial\phi}{\partial t} + \mathbf{V} \cdot \nabla\phi \quad (1:31)$$

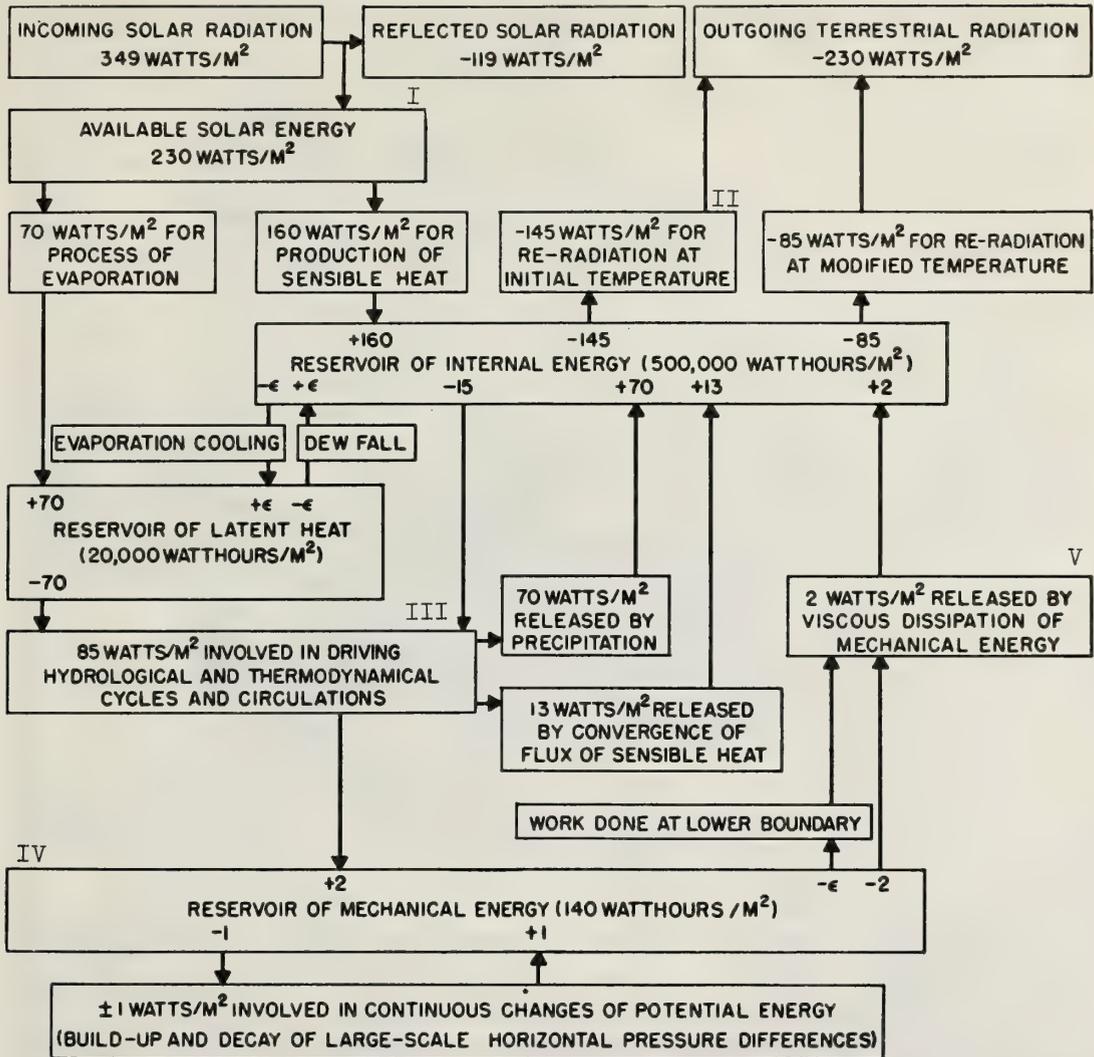


Figure 1.7 The global mean energy cycles of the atmosphere. A solar constant of $1395 \text{ Watts} - \text{m}^{-2}$ is assumed. ϵ denotes an average rate of less than $.5 \text{ Watts} - \text{m}^{-2}$. Accuracy of the derived values is approximately 10% (from Lettau & Haugen, 1957).

Here $\partial\phi/\partial t$ is known as the *local change* of ϕ , \mathbf{V} is the total wind vector as in (1:43), and $\nabla\phi$ is the gradient of ϕ .

In another approach, a given volume of air is “followed” by the sensor and the variation over a given period of time of a parameter, say ϕ , is noted, evaluating the substantial derivative of ϕ directly. This “moving” coordinate system is known as a *Lagrangian coordinate system*.

1.7.2 Ensemble Averaging, Time Averaging, Ergodicity

The motion of the atmosphere and the variation of atmospheric parameters, if viewed from instant to instant, is random. Processes which are random in nature are called *stochastic processes*. These processes are best physically described by investigating the statistical parameters associated with these phenomena such as the mean of a given quantity, its variance, or higher moments like its skewness, or kurtosis (defined in a later chapter). Often this methodology reveals an order not readily apparent in the appraisal of a set of instantaneous values. The appreciation of atmospheric general circulation properties, as an example, is the result of such investigation.

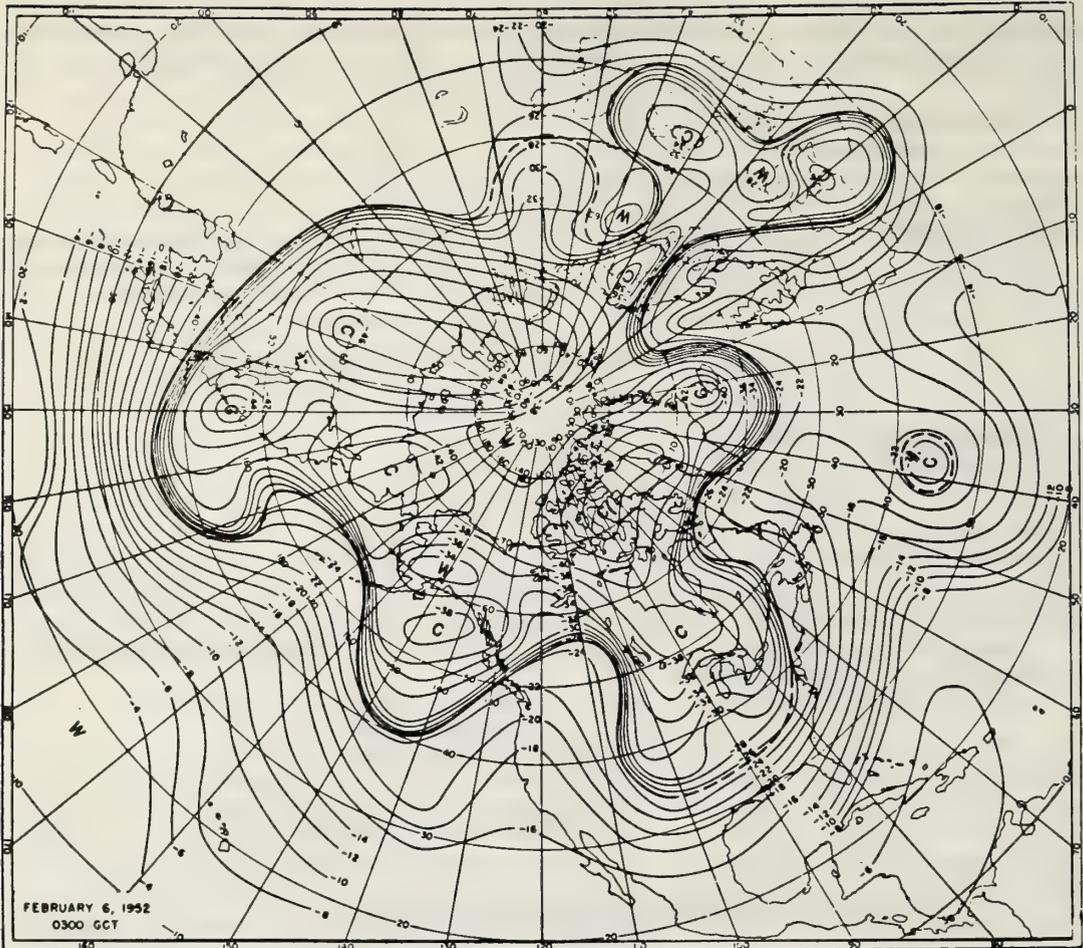


Figure 1.8 Temperature distribution ($^{\circ}\text{C}$) at 500 mb pressure level, 6 Feb 52 (after Bradbury and Palmén, 1953). The dark line shows structures of planetary wave pattern. (Figure taken from Riehl, 1962.)

Evaluation of statistical parameters in a random field is difficult and in actual experience is made only with some fundamental assumptions about the nature of the phenomena. In previous discussions, our attention has been drawn to the properties of a fictional, but useful concept, that of a parcel of air. For our present discussion, we must note that the atmosphere is made up of a vast population of these parcels. In order to investigate the three-dimensional characteristics of the atmosphere, this population must be divided into *subsets* or *sub-populations* as in the discussion of quasi-two-dimensionality in 1.7.1. Order comes out of apparent chaos when the statistical parameters of these subsets are investigated. *population*

Let us take, for example, a subset of parcels which are at a level z , in the atmosphere. In order to obtain the exact statistical parameters, the subset must be infinite in extent. This is our first qualification, in the atmosphere the sample is large, but finite.

In the Lagrangian coordinate system, we will have to average the various moments of the subset population over the entire population, this is known as an *ensemble average*. To perform this mathematically satisfying operation, an infinite number of sensors will be needed which can be interrogated at any given instant. In the Eulerian coordinate system, we will have to have an infinite number of atmospheres with similar populations of parcels at z , which can be sampled at a point (x, y, z) at any instant.

Clearly both schemes are impossible, which brings us to the *principle of ergodic stationarity* or *ergodicity*. The principle of ergodicity states that if a stochastic process is independent of time (i.e., *stationary*), the ensemble average in the Lagrangian sense is equal to an infinite time average in the Eulerian sense. This principle assumes that over an infinite period all the parcels will flow past the Eulerian sensor at (x, y, z) .

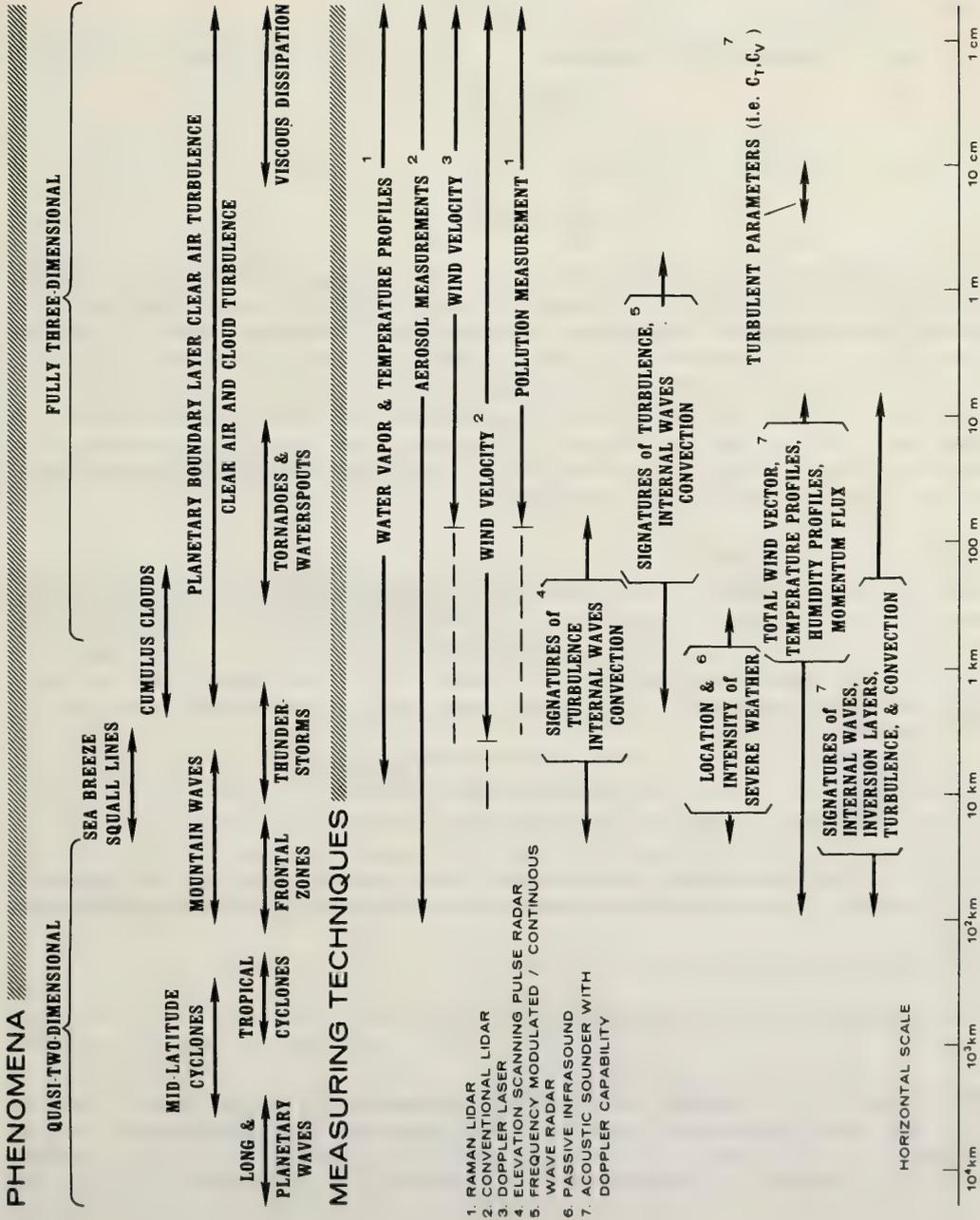


Figure 1.9 Atmospheric scales of motion and remote sensing techniques (scales and phenomena provided by D. Lilly, NCAR).

This principle is to the experimenter's advantage, for with few exceptions the atmosphere is sensed in a Eulerian framework. Ergodicity does allow experimenters to quantitatively investigate atmospheric parameters from the Eulerian sensors and draw conclusions about the entire subset, but not without the two important qualifications: 1) the subset is finite and 2) the stochastic process must be stationary. This argument is presented in concise mathematical form in chapter 11.1.2.

1.7.3 The Taylor Hypothesis

If stationarity is applied to (1:31) one obtains,

$$\frac{\partial \phi}{\partial x} = -\mathbf{V} \cdot \nabla \phi . \quad (1:32)$$

Considering only the x-component (1:37) becomes

$$\frac{\partial \phi}{\partial t} = -\frac{1}{u} \frac{\partial \phi}{\partial t} . \quad (1:33)$$

If a constant value for u is assumed, then by observing the local time change of ϕ , its horizontal gradient is evaluated. Further, if some initial value of ϕ is assigned at a given time, then further time variations of ϕ are equivalent to spatial variations of ϕ . This principle is called the *Taylor hypothesis* and in boundary layer work is often referred to as *frozen turbulence*. As an example of how this is utilized we take a vertically pointing acoustic sounder. This device senses fluctuations of temperature and wind over a vertical distance above the sounder and plots these fluctuations as a function of height and time. By multiplying this time scale by an appropriate mean wind, the time scale becomes a distance scale, thus converting a time series into a quasi-two-dimensional vertical "slice" of the atmosphere. This is a common practice in atmospheric science. However, it should not be applied to non-stationary processes.

1.7.4 The Reynolds Averaging Method.

One of the basic observations in atmospheric physics is the correspondence between the size scale of a phenomena and its lifetime. Small scale phenomena such as dust devils have lifetimes of a few minutes while large scale weather systems have lifetimes often as long as a week. This observational principle coupled with (1:32) links spatial averaging to time averaging at an Eulerian sensor. By taking progressively longer and longer averages of a single time series, signatures of progressively larger scale phenomena become apparent while the smaller scale phenomena are averaged out.

One of the most useful mathematical tools available to the atmospheric scientist is Reynolds Averaging. It has been used to simplify the Navier-Stokes Equation, the basic fluid motion equation used in mathematically modelling atmospheric and oceanic flow (discussed further in 1.8), by separating the equation into a mean equation of motion and an eddy, or perturbation equation of motion (see 1.8.2, 1:53). An outgrowth of this simplification scheme was the concept of eddy correlation, used in the determination of vertical and horizontal fluxes of atmospheric variables such as temperature, water vapor, and momentum (see 1.8.3).

The average value of a variable, ϕ , is

$$\bar{\phi} = \frac{1}{T} \int_0^T \phi(t) dt . \quad (1:34)$$

In the above integral, T is the averaging time or averaging period and $\phi(t)$ is a random function. If the total value of ϕ at any instant can be described as the sum of this average value, $\bar{\phi}$, and some departure from this average, ϕ' , then

$$\phi = \bar{\phi} + \phi' . \quad (1:35)$$

Sir Osborne Reynolds, who derived this averaging method, proposed the following rules which apply to variables described by (1:35).

$$\left. \begin{aligned}
 \text{I. } \overline{\phi + \epsilon} &= \bar{\phi} + \bar{\epsilon} \\
 \text{II. } \overline{c\phi} &= c\bar{\phi}; c \text{ is a constant} \\
 \text{III. } \overline{\phi\epsilon} &= \bar{\phi} \bar{\epsilon} \\
 \text{IV. } \overline{\lim \phi} &= \lim \bar{\phi}
 \end{aligned} \right\} \quad (1:36)$$

Equation (1:36.IV) allows the following operation,

$$\frac{\partial \bar{\phi}}{\partial t} = \overline{\frac{\partial \phi}{\partial t}}$$

Applying (1:35) and (1:36.I) where $\epsilon = \phi'$, results in

$$\bar{\phi}' = 0 \quad (1:37)$$

This is a basic requirement for the use of Reynolds Averaging and is often quite difficult to achieve experimentally. The main problems the experimentalist encounters when trying to satisfy (1:37) are non-stationary conditions and adequate averaging time when having to electronically or digitally filter the data (Lester, 1970).

1.8 Basic Equations of Motion in the Atmosphere

The second Law of Motion proposed by Newton is stated,

$$\frac{d(\rho V \mathbf{u})}{dt} = \mathbf{F} \quad (1:38)$$

where \mathbf{u} is the velocity of the fluid parcel, V is the parcel volume, \mathbf{F} is the resultant force. When applied to a three-dimensional coordinate system (1:38) becomes,

$$\frac{d(\rho V \mathbf{V})}{dt} = \Sigma \mathbf{F} \quad (1:39)$$

where $\mathbf{V} = u\hat{i} + v\hat{j} + w\hat{k}$ (see 1.8.1) is the fluid velocity and \mathbf{F} is a force vector. Assuming there are no sources or sinks of ρ results in

$$\frac{d\mathbf{V}}{dt} = \frac{\Sigma \mathbf{F}}{\rho V} \quad (1:40)$$

for a Lagrangian coordinate system. For an Eulerian coordinate system (1:40) it is

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} = \frac{\Sigma \mathbf{F}}{\rho V} \quad (1:41)$$

Expanding \mathbf{F} into the physical quantities responsible for the forces acting upon the fluid (see T1.4) and taking (1:41) per unit volume of air one obtains

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} = -\nabla \phi - \frac{1}{\rho} \nabla p + \frac{1}{\rho} \mathbf{Q} + 2\Omega \times \mathbf{V} \quad (1:42)$$

This is the *basic*, or *primitive equation of motion* in vector form and is also known as the *Navier-Stokes Equation*.† Strictly speaking ∇p and \mathbf{Q} combine to form the deformation force; however, they are separated since ∇p is due to external deformation forces and \mathbf{Q} is due to internal deformation forces.

† In this form (1:42) is for total motion; that is, mean motion plus perturbation or eddy motion. The apparently arbitrary division into mean and eddy motion is deceiving, since the division is often made by deciding which size scale of motion is to be studied or predicted. (Recall the discussion of subgrid scale parameterization in 1.5.3)

1.8.1 The Tangent Plane Coordinate System

There are several coordinate systems applicable to the earth's surface for which (1:42) can be written. Among them are spherical coordinates (the most general), cylindrical coordinates (useful in hurricane research), and plane coordinates. Our discussion of (1:42) with respect to the tangent plane coordinate system (F1.10). The velocity vector in this system (see 1:28, 1:39) is

$$\mathbf{V} = u\hat{i} + v\hat{j} + w\hat{k} \tag{1:43}$$

where

+ u is the eastward wind component (i.e. wind blowing to east)

+ v is the northward wind component

+ w is the upward vertical wind component.

For very large atmospheric motions such as planetary waves, a correction for the earth's curvature must be included when using this system. However, for most of the motions discussed in subsequent chapters, the tangent plane coordinate system will suffice.

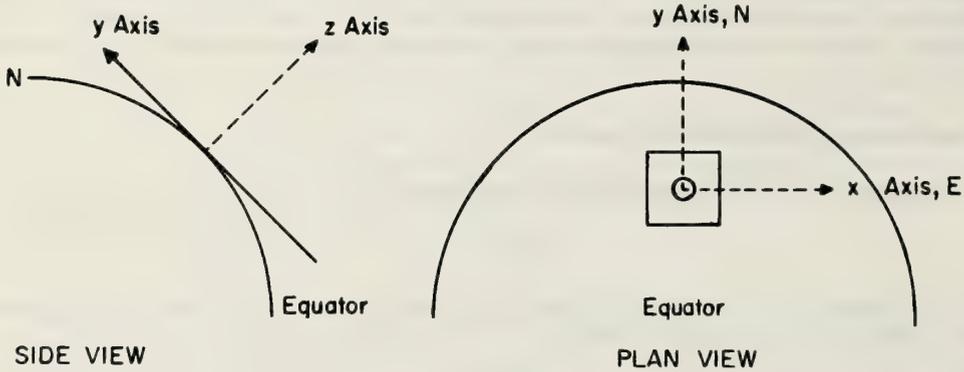


Figure 1.10 Schematic representation of tangent plane coordinate system

1.8.2 The Reynolds' Stress: Introduction to Turbulence, Eddy Correlation, Use of Profiles, Ageostrophic Wind

From the abstract thought of Newton, combined with countless verifications, it was shown that the stress or internal deformation of a fluid was proportional to the gradient of the fluid speed,

$$\tau = \mu \frac{\partial u}{\partial r} \tag{1:44}$$

where τ is the stress, a force per unit area; μ is a constant of proportionality known as the *dynamic coefficient of molecular viscosity* and $\partial U/\partial r$ is the gradient of the fluid speed in the direction or *fluid shear*. There are two conditions in a fluid for zero stress. The first is that $\mu = 0$; fluids with this property are often called ideal, frictionless, or inviscid fluids. The second is that $\partial U/\partial r = 0$. The coefficient μ is a physical property of a fluid.

This stress or friction force enters into (1:42) as \mathbf{Q} . Often (1:42) is made non-dimensional by forming ratios of the forces active in the equation. In this way, coefficients of various non-dimensional variables show their relative importance to the Navier-Stokes Equation. The coefficient of non-dimensional \mathbf{Q} when (1:42) is non-dimensionalized by dividing by the left hand side of the equation ($d\mathbf{V}/dt$ is also known as the inertial force) is called the *Reynolds Number*, R_e .

$$R_e = \frac{\rho U r}{\mu} \tag{1:45}$$

Thus the Reynolds Number is the ratio of the inertial forces (excluding Coriolis force) to the viscous forces in the fluid flow regime.

For the purposes of our discussion there are three critical values of R_e :

I. $R_e \leq 10$ (approximately) in which case the flow is *laminar* and dependent upon fluid viscosity, as in syrup.

II. $50 < R_e < 10^3$ in which case the motion is *transitional*, characterized by regular vortex patterns.

III. $R_e > 10^4$ in which case the motion becomes chaotic and irregular, or *turbulent*, and dependent upon eddy viscosity (see below).

The Reynolds Number for the atmosphere is generally above 10^4 . In this case frictional forces become important in (1:42); however, they are only vaguely linked to the molecular properties of the fluid as in Q . Viscosity does play a part in fluid motion with high Reynolds Number but its relationship to the flow is far more complex than in (1:44).

In an attempt to simplify (1:42), atmospheric scientists proposed a relationship analogous to (1:44),

$$\tau = \mu_e \frac{\partial u}{\partial r} \tag{1:46}$$

In this equation, μ_e is the *eddy viscosity* of the fluid due to the turbulent nature of its motion. This stress is about six orders of magnitude larger for turbulent flow than for viscous flow, $R_e < 10$, with the same values of shear.

The force per unit mass exerted by all stresses is

$$\frac{1}{\rho} Q = \frac{1}{\rho} \left\{ Q_x \hat{i} + Q_y \hat{j} + Q_z \hat{k} \right\} \tag{1:47}$$

where Q is the component of Q on a given coordinate axis. For example, the x component† of (1:47) is

$$\frac{1}{\rho} Q_x = \frac{1}{\rho} \left\{ \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} \right\} \tag{1:48}$$

where the various stresses are shown in (F1.11). Where from consideration of viscous flow theory,

$$\tau_{xx} = \mu \left(\frac{2u}{2x} + \frac{2v}{2x} - \frac{2}{3} (\Delta \cdot v) \right) \tag{1:49}$$

which holds for viscous or laminar flows.

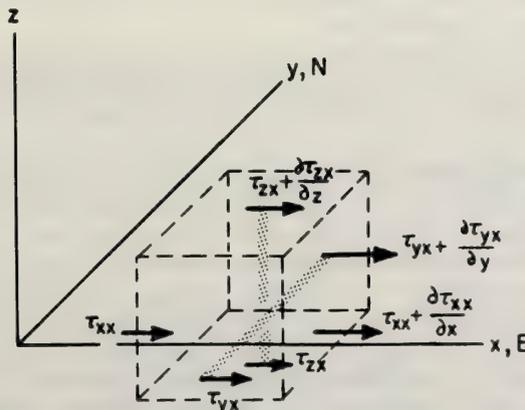


Figure 1.11 Schematic representation of x components of stress dyad on a unit volume

† In some work the tangent plane system is rotated so that the x axis is the direction of the mean wind velocity.

For turbulent flow \mathbf{Q} can be neglected but friction forces due to turbulence cannot. To illustrate this, we take the x-component of (1:42) on a geopotential surface (i.e., $\nabla\phi = 0$),

$$\frac{du}{dt} + \rho \frac{\partial u}{\partial t} + \rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial y} + \rho w \frac{\partial u}{\partial z} = \rho f v - \frac{\partial p}{\partial x}. \quad (1:50)$$

Here $f = 2\Omega \sin\beta$ is the Coriolis parameter with Ω , the scalar angular velocity of the earth and β the latitude. Considering the atmosphere incompressible and using (1:28),

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u u}{\partial x} + \frac{\partial \rho u v}{\partial y} + \frac{\partial \rho u w}{\partial z} = \rho f v - \frac{\partial p}{\partial x} \quad (1:51)$$

Applying the Reynolds Averaging Rules, (1:36), to (1:51) and noting that terms like

$$\overline{vu} = \bar{v} \bar{u} + \overline{v'u'} \quad (1:52)$$

results in

$$\frac{\partial \rho \bar{u}}{\partial t} + \frac{\partial \rho \bar{u} \bar{u}}{\partial x} + \frac{\partial \rho \bar{u} \bar{v}}{\partial y} + \frac{\partial \rho \bar{u} \bar{w}}{\partial z} = \bar{v} f - \frac{\partial \bar{p}}{\partial x} - \frac{\partial \rho \overline{u'u'}}{\partial x} - \frac{\partial \rho \overline{u'v'}}{\partial y} - \frac{\partial \rho \overline{u'w'}}{\partial z}. \quad (1:53)$$

The second term on the right hand side of (1:52) is defined as the *covariance* of v' and u' , or *eddy correlation* of v' and u' . In integral form this is

$$\text{Cov}(v'u') \equiv \overline{v'u'} \equiv \frac{1}{T} \int_0^T u(t)v(t) dt \quad (1:54)$$

where, in general $u(t)$ and $v(t)$ are both random functions. Subtracting the average value of (1:28) from (1:53) we obtain the Navier-Stokes equation for the x-component of the mean flow

$$\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} + \bar{w} \frac{\partial \bar{u}}{\partial z} = \bar{v} f - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x} + \frac{1}{\rho} \left\{ \frac{\partial \rho \overline{u'u'}}{\partial x} + \frac{\partial \rho \overline{u'v'}}{\partial y} + \frac{\partial \rho \overline{u'w'}}{\partial z} \right\}. \quad (1:55)$$

Equation (1:55) is analogous to the x-component of (1:42) for motion on a geopotential surface. The last term in brackets is analogous to $\frac{1}{\rho} \mathbf{Q}$ in (1:42) with the important qualification that in (1:42), \mathbf{Q} accounts for stresses due to the molecular viscosity of the fluid while in (1:55), the bracketed term accounts for stresses due to the turbulence acting upon the mean flow.

The stresses expressed in (1:55) are known as the *Reynolds stresses*. In order to simplify (1:55) mathematically, the analogy expressed by (1:46) is introduced into (1:55) to obtain,

$$\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} + \bar{w} \frac{\partial \bar{u}}{\partial z} = \bar{v} f - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x} + \frac{1}{\rho} \left\{ \frac{\partial}{\partial x} \left[\mu_e \frac{\partial \bar{u}}{\partial x} \right] + \frac{\partial}{\partial y} \left[\mu_e \frac{\partial \bar{u}}{\partial y} \right] + \frac{\partial}{\partial z} \left[\mu_e \frac{\partial \bar{u}}{\partial z} \right] \right\}. \quad (1:56)$$

From (1:55) and (1:46) the following identities are presented

$$\tau_{zx} = -\rho \overline{u'w'} = -\mu_e \frac{\partial \bar{u}}{\partial z}, \quad (1:57)$$

$$-\frac{\tau_{zx}}{\rho} = u_*^2 = \overline{u'w'} = K_M \frac{\partial \bar{u}}{\partial z}, \quad (1:58)$$

$$\mu_e = \rho K_M. \quad (1:59)$$

In (1:57), (1:58), (1:59) above, u_* is the friction velocity and K_M is the *austausch* or mixing coefficient for momentum. The y and z components of the Navier-Stokes Equation can be treated in a similar manner to yield equations like (1:55) and (1:56).

Some consequences of applying (1:56) to the atmosphere can now be explored. In most atmospheric motions we find that $\partial\bar{u}/\partial z \gg \partial\bar{u}/\partial x$, $\partial\bar{u}/\partial y$ and that $\partial^2 u/\partial z^2 \gg \partial\mu/\partial z$ so that (1:56) can be approximated by

$$\frac{d\bar{u}}{dt} \cong f\bar{v} - \frac{1}{\rho} \frac{\partial\bar{p}}{\partial x} + K_M \frac{\partial^2 u}{\partial z^2} \quad (1:60)$$

From (1:60) we draw the important conclusion that stress is proportional to the curvature of the *horizontal wind profile* (i.e., the variation of u with height).

If we consider the special case of no stress and no acceleration in the flow (1:60) becomes

$$f\bar{v} = \frac{1}{\rho} \frac{\partial\bar{p}}{\partial x} \quad (1:61)$$

This is a very useful relationship known as the *geostrophic wind equation*, which for \bar{u} is

$$f\bar{u} = - \frac{1}{\rho} \frac{\partial\bar{p}}{\partial y} \quad (1:62)$$

A consequence of this type of flow is that fluid motion must be parallel to lines of equal pressure, or isobars. If the turbulent stresses are added to (1:61) and (1:62), a change in the flow pattern will result. The resultant winds are called *ageostrophic winds*. The fluid motion is directed across the isobars to lower pressure in order to balance the decelerating influence of the stresses. Over a large area this could cause an area of convergence and applying (1:29) large scale upward motion would result. In the Ekman layer the winds are ageostrophic.

A further justification for the use of (1:57) and (1:58) which relate the Reynolds' stresses to the shear of the mean wind is purely economical. It is easier and cheaper to measure a mean wind speed than to measure a quantity such as $\overline{u'w'}$ or $\overline{u'v'}$. Further from the discussion in 1.5.3 the effect of boundary layer processes in numerical simulation of the atmosphere must be related to commonly measured variables. Therefore much of the boundary layer research reported in subsequent chapters will ultimately reside in equations such as (1:57), especially in the functional form of the variation of μ_e and related profiles of wind, temperature, and water vapor with time, space and stability.

1.9 Remote Sensing and the Atmospheric Environment

In this chapter, we have endeavored to introduce the very basic notions and definitions which will aid in understanding the more detailed material to be presented in subsequent chapters. We shall now relate some broad technical relationships which exist between remote sensing and the material presented in the previous sections.

Except for satellite applications, (F1.9) shows that the state of the art reaches, with few exceptions, from the microscale to mesoscale phenomena. Satellites are able to resolve synoptic scale to mesoscale phenomena, so that remote sensing in general is able to sense all scales of atmospheric phenomena. However, (F1.9) shows that different remote sensors have the capability to sense only particular signatures of a physical element such as wind speed, temperature, turbulent characteristics, chemical composition, etc., leaving gaps in the ability to measure all characteristics of atmospheric phenomena on all scales by remote sensing.

In fact, remote sensing is far too general an expression to discuss without including what is being sensed. With this in mind (T1.5) is presented which gives the general roles of remote sensing techniques to the various atmospheric parameters at different scales. It can be safely said, with the exception of certain satellite measurements, such as vertical temperature sounding, that remote sensing will not replace *in situ* measurements at this point in time because of the complexity in interpreting remotely sensed signals and because of the expense generally incurred in obtaining remotely sensed data. However, remote sensing is a very important complement to *in situ* measurements, an eloquent example being given in the following discussions of clear air radars (Chapters 14, 18, 19, 20).

The Atmospheric Environment

Table 1.4 Physical Origin of Forces Acting upon an Atmospheric Volume

Notation in (1:47)	Physical Force
$\nabla\phi$	gradient of <i>geopotential</i> ; $\phi = gz$; a force due to gravity
∇p	a force due to the gradient of pressure, p , external deformation force.
Q	stress forces due to the viscosity of the fluid; internal deformation force.
$2\Omega \times V$	a "fictitious" inertial force due to relative motion of the fluid on the earth which in rotating with an angular velocity of Ω , known as the <i>Coriolis force</i> .

Table 1.5 Roles of Remote Sensing Techniques in Describing the Atmospheric Environment (Number Corresponds to Remote Sensors Listed Below)

Parameter	Scale		
	Synoptic	Meso	Micro
Wind	2,9	2,3,7	2,3,7
Temperature	8	7,1,8	7,1
Moisture		7,1	7,1
Pollutants	10	10	1
Signatures of turbulent phenomena		4,7,5	4,7,5
Turbulent parameters			7
Signatures of weather phenomena	6,9	9	

Remote Sensors

1. Raman lidar	6. Passive infrasound
2. Conventional lidar	7. Acoustic sounder with doppler capability
3. Doppler laser	8. Satellite mounted infrared & microwave radiometers
4. Elevation scanning pulse radar	9. Satellite mounted cloud photography
5. Frequency modulated/continuous wave radar	10. Satellite mounted infrared & microwave spectrometers and multiband photography

Table A1

N (N units)	T (°K)	ρ (mb)	e (mb)	a (°K ⁻¹)	b (mb ⁻¹)	c (mb ⁻¹)
319	288.2	1013	10.2	-1.27	4.5	0.27

1-10 References

- Bean, B. R., and E. J. Dutton, 1968: *Radio Meteorology* Dover Publications, New York, N.Y. , 435 pp.
- Bradbury, D. and E. Palmén (1953), On the existence of a polar front zone at the 500 mb level. *Bull. Amer. Met. Soc.*, *34*, 56-63.
- Byers, R. B., 1959: *General Meteorology* McGraw-Hill, New York .
- Changnon, S., F. A. Huff, R. G. Simon (1971), METROMEX: an investigation of inadvertent weather modification, *Bull. Am. Met. Soc.*, *52*, No. 10, 958-967.
- Craig, R. A., (1965), *The Upper Atmosphere: Meteorology and Physics*, International Geophysics Series, Vol. 8, T. van Meighem, ed. Academic Press, New York .
- Junge, C., (1952), Die konstitution des atmosphärischen aerosols, *Ann. Met. (Beiheft)*.
- Lapaz, L., (1958), Effects of Meteors on the Earth. *Adv. in Geophy.*, *4*, 272-273.
- Lester, P. F. (1970), Some physical and statistical aspects of clear air turbulence, Colorado State University, Dept. Atmospheric Science, Atmospheric Science Paper No. 163, 154 pp.
- Lettau, H. H., and D. A. Haugen, (1957), Empirical Models of Thermal Structure in the Lower Atmosphere and Soil, Section 1, Ch. 2 of Handbook of Geophysics for Air Force Designers, Geophysics Research Directorate, USAF.
- Lorenz, E., (1967), The Nature and Theory of the General Circulation of the Atmosphere. World Meteorological Organization Monograph.
- Mason, B. J. (1971), Future developments in meteorology: an outlook to the year 2000, *Quarterly J. of the Royal Met. Soc.*, *96*, No. 409, 349-368.
- Riehl, H., (1962), Jet Streams of the Atmosphere, Tech. Paper No. 32, Dept. Atmos. Sc., Colo. St. Univ.
- U.S. Standard Atmosphere*, (1962) U.S. Government Printing Office, Washington, D.C.

**APPENDIX
THE RADIO REFRACTIVE INDEX OF AIR**

Electromagnetic waves traveling through the atmospheric medium are affected by the medium's refractive index. The refractive index is a function of pressure, temperature, and water vapor as well as other gaseous constituents of the atmosphere. In that part of the electromagnetic spectrum above 30 megahertz refractive index is referred to as the *radio refractive index* and is approximated (Bean and Dutton, 1968) by

$$N = 77.7 \frac{p}{T} + 3.73 \times 10^5 \frac{e}{T} \quad (\text{A:1})$$

For convenience

$$N = (n - 1) \times 10^6 \quad (\text{A:2})$$

where n is the actual value of refractive index.

Applied to remote sensing, variations in the radio refractive index are both a nuisance and a blessing in a manner similar to the discussion of aerosols and hydrometers in 1.4. For some systems (e.g., microwave distance measuring equipment) radio refractive index fluctuations are environmental noise which must be eliminated in the processed signal. For other systems, such as the FM-CW radar and other clear air radars, refractive index fluctuations are indicative of atmospheric structure and are often utilized to trace phenomena much larger than the scale of the fluctuations themselves. Since

$$N = N(T, p, e) \quad (\text{A:3})$$

we can write

$$dN = \frac{\partial N}{\partial T} dT + \frac{\partial N}{\partial p} dp + \frac{\partial N}{\partial e} de \quad (\text{A:4})$$

Thus for some average atmospheric condition the root mean square value of (A:4) can be written in finite difference form as

$$N = \left[(a\Delta T)^2 + (b\Delta p)^2 + (c\Delta e)^2 \right]^{1/2} \quad (\text{A:5})$$

Table A1 shows the values for a , b , and c for U.S. Standard Atmosphere surface conditions.

One sees from (TA1) that ΔN is quite sensitive to Δe and least sensitive to Δp . Thus in moist regions (which includes all but the most arid regions of the lower troposphere) fluctuations in N are caused by fluctuations in water vapor content. In relatively dry regions, such as the middle and upper troposphere, N fluctuations are indicative of temperature variations.

FURTHER READING

Hess, S. L., 1959: *Introduction to Theoretical Meteorology*, (H. Holt, N. Y.) 362 pp.

List of Symbols

F	force vector	r	relative humidity; a generalized length in the x,y,z coordinate system
K_m	austausch coefficient for momentum flux	t	time
N	radio refractive index of air	u	x component of wind velocity
Q	sensible heat	w_s	saturation mixing ratio
Q	internal deformation force; stress force	v	y component of wind velocity
R*	universal gas constant	w	ratio of ρ_v/ρ_d ; z component of wind velocity
R_d	specific gas constant for dry air	u_*	friction velocity
R_e	Reynolds Number	Γ_d	dry adiabatic temperature lapse rate
T	absolute temperature; averaging time	Ω	rotation rate of the earth
T_v	virtual temperature	a	specific volume
V	wind velocity	γ	temperature lapse rate
V	volume	ϵ	ratio of m_v to m_d , scalar variable
c_p	specific heat for dry air	θ	potential temperature
e	partial pressure of water vapor	λ	wave length
f	coriolis parameter	μ	micron; dynamic coefficient of molecular viscosity
f_s	Brunt-Väisälä frequency	μ_e	eddy viscosity
g	acceleration due to gravity	ρ	air density
\hat{i}	unit vector in x direction	ρ_d	dry air density
\hat{j}	unit vector in y direction	ρ_v	absolute humidity or water vapor density
\hat{k}	unit vector in z direction	ϕ	a scalar variable; geopotential
\bar{m}	mean molecular weight of moist air	τ	stress dyad
m_d	mean molecular weight of dry air		
m_v	mean molecular weight of water vapor		
p	pressure		
q	specific humidity		

Chapter 2 FUNDAMENTAL PROBLEMS OF ATMOSPHERIC SCIENCE

Elmar R. Reiter

Department of Atmospheric Science
Colorado State University

One of the truly fundamental problems of meteorology is the (non-linear) interaction between various scales of motion. Examples of such interactions and some of the difficulties that present themselves in their assessment are outlined briefly.

2.0 Introduction

A chapter with such an all-encompassing title, by necessity, has to be biased heavily by my own sense of priorities. The Committee on Atmospheric Science, CAS (1971) recently summarized its evaluation of existing problems, emphasizing the areas of weather prediction, air quality control, weather modification, and the dissemination of information. In the present discussion I will attempt to point out those shortcomings in our concepts of the atmosphere which, in my opinion, are fundamental to the wide and complex field of atmospheric science and its tasks as envisioned in the above-mentioned CAS report.

2.1 Scales of Motion

Figure (2.1), in a schematic three-dimensional presentation, shows a variety of atmospheric phenomena ranging from tornadoes to climatic changes, in their time- and space-scale relationships. Also indicated in this diagram are the various theoretical approaches and prediction methods by which these phenomena appear tractable with the present "state of the art". In reading the text that accompanies the original publication of (F2.1), one senses the mixed emotions by which present forecasting capabilities were evaluated by the scientific team assembled in CAS. The difficulties in ascribing quantitative values of predictability to various scales of atmospheric motion must have prevented the members of CAS from filling in the space provided for such an evaluation in (F2.1).

The fundamental problem of atmospheric science becomes quite obvious from (F2.1): It is that the atmosphere is a continuum in time and space, that contains motion systems, or eddies, in the frequency and wave number spaces. A representation of atmospheric motions in frequency space provided, for instance, by spectrum analysis of wind speeds at jet-stream level over a single station, reveals a spectrum of kinetic energy (F2.2 and T2.1) that shows several prominent wave bands and considerable variability with time and location of measurement, but nevertheless is *uninterrupted* and continuous throughout the frequency range accessible to detailed measurements (Vinnichenko 1970, and Vinnichenko and Dutton 1969). Yet, as shown in (F2.1), we are forced to apply simplified mathematical models, which we call "theory", to relatively narrow bands within the spectrum of atmospheric motions.

In viewing (F2.2) we have to admit that none of these "bands" of atmospheric eddy motions stands by itself. Each "band" interacts with neighboring bands. It is these "non-linear interactions" which present, in my opinion, *the* fundamental problem in the understanding of the atmosphere. As an example of these interactions, we may consider the release of latent heat in the convective "hot" towers of cumulonimbi. Even though this heat is supplied to the atmosphere within the clouds which usually, in the tropics, are arranged in clusters, the main benefit to the atmosphere in terms of a gradual warming of the environment comes in neighboring regions with compensating descending motions and adiabatic heating. The convection inside the clouds, even the cloud clusters themselves, may be considered a mesoscale phenomenon. The adiabatic sinking, hence the warming of the environment, however occurs on a larger, possibly synoptic, scale. The energy input into the atmosphere, thus, takes place on a smaller space scale than the scale of flow patterns which are benefitting from this input, and which are exporting sensible heat from the area under consideration. Such small-scale convective motions play an important role in the heating of the free atmosphere in low latitudes. Since it is this heating effect, together with the cooling at high latitudes, that drives the general circulation of the atmosphere, we have to conclude that the meso-scale convective patterns in the tropics interact even with the planetary-scale motions.

The large-scale divergence and convergence patterns embedded in the planetary flow, on the other hand, trigger the ascending and descending motions that give rise to clouds with "hot-tower" convection

and to cloud-free areas with sinking air currents and adiabatic warming. It is, therefore, not only the meso-scale that acts upon the synoptic and planetary scales of motion, but the latter scales *inter-act* again with the smaller scales of atmospheric flow.

Various attempts are underway to put numerical values on the magnitude of these interactions. In the inertial subrange of turbulence, if it exists in the free atmosphere, the interaction between eddies of different frequencies is viewed as a simple "cascade" of energy from large to small eddies. We assume that the rate of energy dissipation, ϵ , describes this cascading process of energy between low and high frequencies within the inertial subrange, according to

$$S(k) = a\epsilon^{2/3}k^{-5/3} \tag{2.1}$$

where k is the wave number, a is a constant, and S is the energy density of the one-dimensional spectrum. At frequencies lower than those characteristic of the inertial subrange, even a crude estimate of non-linear interaction becomes very difficult. Attempts to parameterize the "subgrid-scale flux" of energy by convective motions in numerical prediction schemes, for instance, are geared to circumvent, however inadequately, the problem of such non-linear interactions between scales of motion.

Many of the approximations, that make the "theories" indicated in (F2.1) mathematically tractable, are based upon the assumption that effects of other scales of motion may either be neglected (like the Coriolis term in small-scale phenomena) or can be fit into a suitable mold (e.g., $\nu\partial^2u_i/\partial x_i^2$ for a friction term in the equations of motions, that supposedly accounts for "sub-grid scale" effects. ν is the kinematic viscosity of air, u_i is the flow component in the coordinate direction x_i).

Charney (1948), by dimensional analysis, tested the validity of certain of these assumptions to which meteorologists, especially in the theoretical fields, felt themselves attracted. Our present concept of "scales of atmospheric motion" has matured, however, beyond the magnitude estimates of certain terms in the primitive equations. In order to accommodate more recent concepts of atmospheric behavior, we will define *scales of motion* as synonymous to certain bands of finite width in the continuous spectrum of

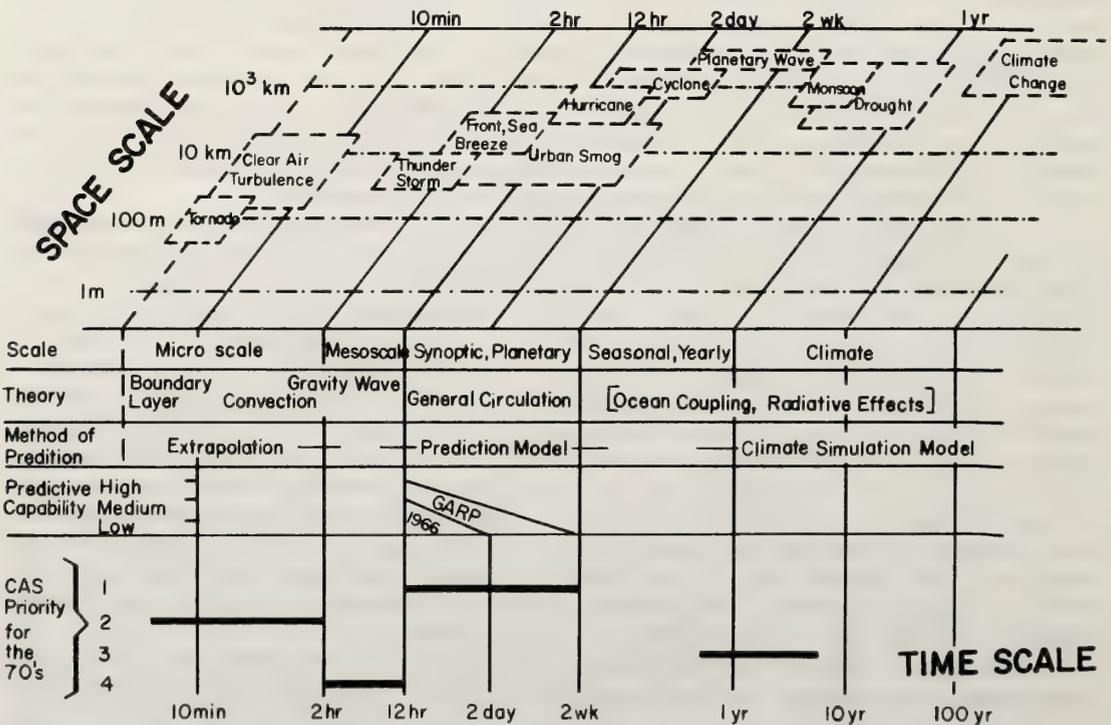


Figure 2.1 Time and space scales of atmospheric phenomena. (Committee on Atmospheric Sciences, 1971),

motions. We will play along with the scale suggestions plotted along the time axis of (F2.1) without adhering too strictly to them.

2.2 Primitive Equations

It is not an easy task, to accommodate the concept of scales of atmospheric motion in the framework of mathematical equations to which we have accustomed ourselves. In the following, I will attempt a crude thumb-nail sketch of what should be considered, without trying to cover all details, important though they may be.

The equations of motion in tensor notation may be written as [see e.g. Hinze (1959), Lumley and Panofsky (1964)]

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \nu D_{ji} - \frac{2}{3} \frac{\partial}{\partial x_i} \nu \theta - g \delta_{3i} - 2\epsilon_{ijk} \Omega_j u_k \quad (2:2)$$

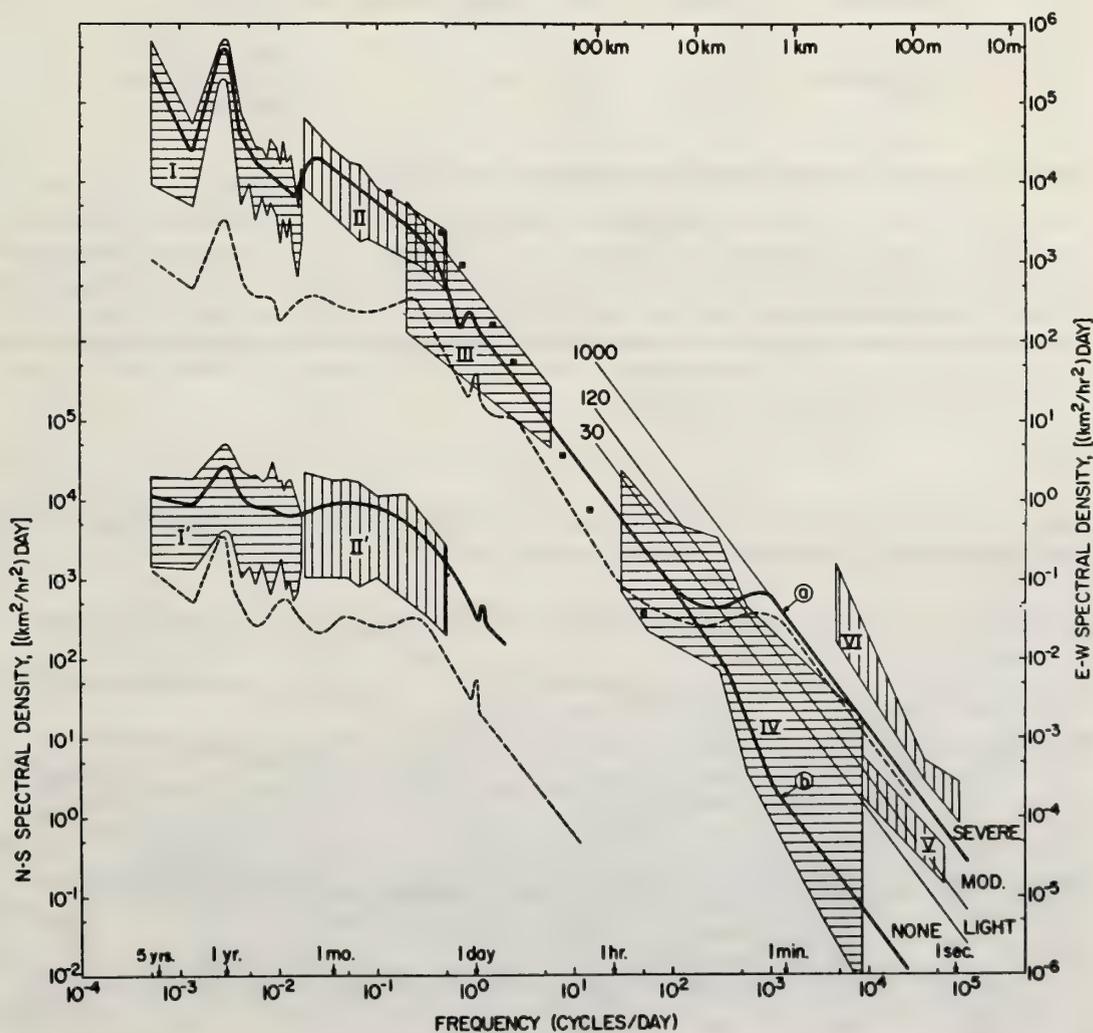


Figure 2.2 Average spectra of E-W and N-S wind components (solid lines) in the free atmosphere (see Table 2.1). Dashed lines are corresponding spectra near the ground (after Van der Hoven [1957] and Oort and Taylor [1967]). Solid squares are Mantis' (1963) Lagrangian data (from Vinnichenko and Dutton, 1969).

Table 2.1 Sources and Characteristics of Spectra Included in Figure 2.2.
(From Vinnichenko and Dutton, 1969).

Box Number	Source	Altitudes, km	Period Covered	Location	Data Characteristics	Number of Spectra in the Box
I, I'	Vinnichenko [1970]	3, 6, 10, 12, 14, 16, 18, 20	1952-1968	Washington, D. C., U.S.A.	Monthly averaged rawinsonde winds	8
II, II'	Chiu [1959]	3, 6, 10, 12, 14, 16, 20	1/1/53-4/30/54	Belmar, N. J., U.S.A.	Daily rawinsonde winds	7
III	Vinnichenko et al. [1968]	3, 5, 7, 9, 12, 15, 18	7/66-1/67	Kharkov, Ukraine, USSR	2-hour-interval rawinsonde winds during each month	14
IV*	Vinnichenko et al. [1968]	1, 3, 5, 7, 9	7/66-1/67	Kharkov, Ukraine, USSR	Doppler navigator and hot-wire anemometer, IL-18 aircraft	60
V*	Mather [1968]	Jet stream level	2/68	Denver, Colorado, U.S.A.	Doppler navigator pressure and vanes probe, T-33 aircraft	7
VI*	Steiner and Rhyne [1964]	5-13 Inside cb clouds	1960-1961	U.S.A.	Accelerometer and flow vane probe, T-33 aircraft	5

* These data represent longitudinal spectra with respect to a heading of aircraft.

The symbols used in this equation are explained in detail in 2.5.1. The reader should not be dismayed by the many subscripts but should keep in mind that $i = 1, 2, 3$ marks the x, y, z components, respectively, in a Cartesian coordinate system. The symbols j and k are counted through 1, 2, 3 with each value of i . The Einstein summation convention has been adopted whereby terms with repeated indices are summed. Thus $u_j(\partial u_i / \partial x_j)$ becomes $u(\partial u / \partial x) + v(\partial u / \partial y) + w(\partial u / \partial z)$. The definitions of the Kronecker delta, δ_{ij} , and of the alternating tensor, ϵ_{ijk} , are explained in 2.5.1.

We now define averages of a quantity, A , by

$$\bar{A} = \frac{1}{T} \int_{-T/2}^{T/2} A(t + \tau) d\tau \tag{2:3}$$

where T is the time period of averaging, t is the point in time for which this average holds, and $d\tau$ is the time interval between data points used in the averaging process.

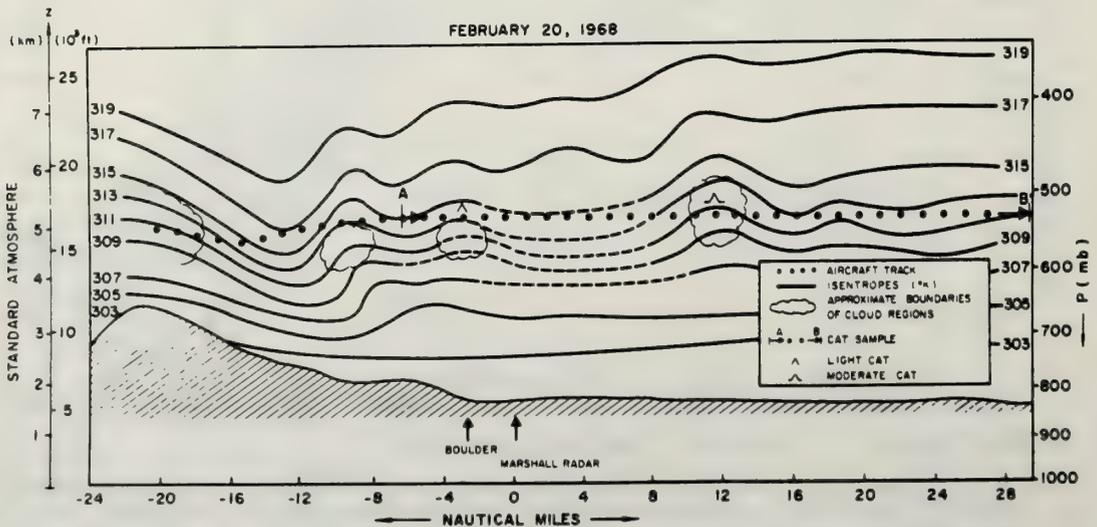


Figure 2.3 Vertical cross-section along aircraft track (after Lester, 1970).

Departures from the mean, A' , are defined by

$$A = \bar{A} + A' \tag{2:4}$$

With this definition we are introducing *volens volens* the aforementioned scale concept, because the magnitude of the variance of A , defined as $(A')^2$, will depend on the length of T , but also on the place, t , in the history of the atmosphere. We can demonstrate easily that also \bar{A} depends on T and t .

This dependence generates a practical problem, which is illustrated in the following example. For the large quantity of data accumulated in fine-scale turbulence measurements by aircraft, it is most convenient to obtain average values of \bar{u}_i by application of a mathematical or electronic low-pass filter (Graham, 1963; Martin, 1963; Lester, 1970; Stankov, 1970). High-pass filtering yields values for u'_i . Since u_i is a function of time, especially under the non-stationary and strongly intermittent conditions under which clear-air turbulence (CAT) occurs in the atmosphere, not all Reynolds averaging rules are generally satisfied. We may accept

$$\bar{\bar{u}}_i = \bar{u}_i \tag{2:5}$$

$$\overline{u_i^2} = \bar{u}_i^2 \tag{2:6}$$

Overbars denote averages of the quantity over which they extend. Two overbars stand for repeated averaging processes. "Primes" symbolize departures from a mean value.

The validity of
$$\overline{u'_i} = 0 \tag{2:7}$$

depends on the quality of the filters used in the aforementioned data reduction scheme. Even if the error made by the assumption that condition (2:7) is met is small, the magnitude of

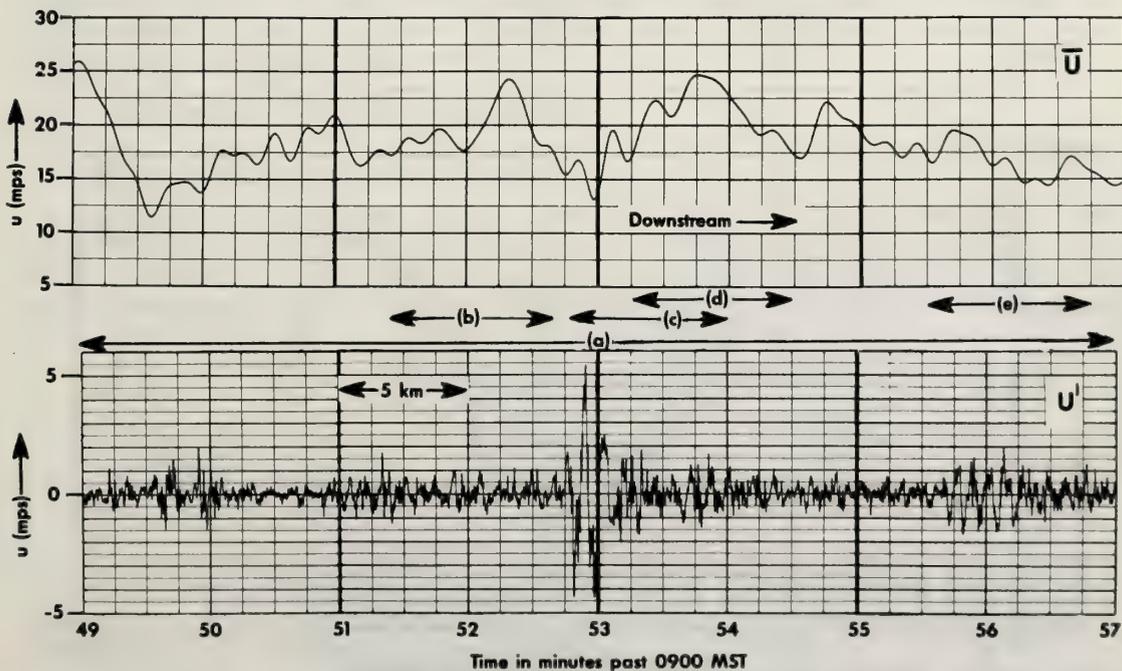


Figure 2.4 Top: Velocity component along aircraft track after low-pass filtering. Bottom: Velocity component along aircraft track after high-pass filtering. Time intervals identified by letters pertain to spectra shown in (F2.8) (after Lester, 1970).

$$\overline{u_i' u_i'} \neq 0 \tag{2:8}$$

may be appreciable if $\overline{u_i'}(t)$ varies strongly within the averaging interval T . This has been demonstrated by Lester (1970) in a number of examples.

Figure (2.3) shows the lee-wave situation over Boulder, Colorado, on 20 February 1968. This gravity-wave phenomenon is produced in a thermally stable atmosphere by shearing flow over corrugated terrain. Flight measurements along the indicated airplane track are given in (F2.4). The upper portion of this diagram contains the low-pass portion of the wind fluctuations parallel to the aircraft track; the lower part of the diagram gives the high-pass portion. Note the strong CAT associated with a gravity wave in the lee of the mountains. Figure (2.5) compares \overline{u} with $\overline{\overline{u}}$ and $\overline{u^2}$ with $\overline{\overline{u^2}}$. Obviously, conditions (2:5) and (2:6) are well satisfied. According to (F2.6), $\overline{u'} \neq 0$. The magnitude of this term may be reduced, however, by choosing a mathematical filter with a narrower transition zone. Figure(2.7) demonstrates the inequality in (2:8). In expanding (2:2) by using the perturbation quantities given by (2:4), we will have to retain, therefore, terms such as (2:8). Doing this, the equation for the mean motion, obtained by low-pass filtering of the perturbation equation containing the aforementioned substitutions, may be written as

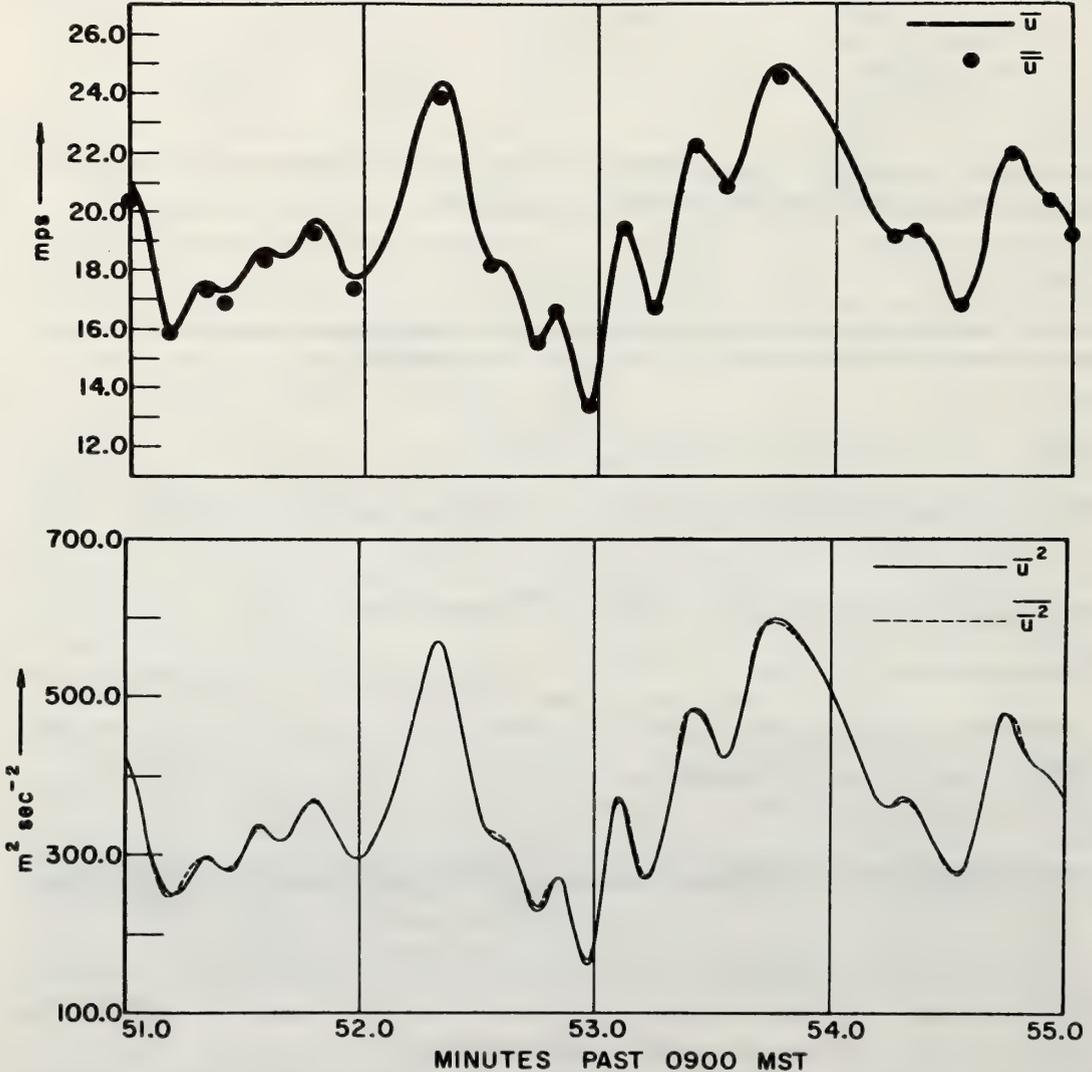


Figure 2.5 Comparison between \overline{u} and $\overline{\overline{u}}$ and $\overline{u^2}$ and $\overline{\overline{u^2}}$ for a case in which the high-pass filter was taken as the difference between all pass and low-pass filters. (after Lester, 1970).

$$\begin{aligned} \frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} + [\bar{u}'_j \frac{\partial \bar{u}_i}{\partial x_j}] + [\bar{u}_j \frac{\partial \bar{u}'_i}{\partial x_j}] = \\ - \frac{1}{\bar{\rho}} \frac{\partial \bar{p}}{\partial x_i} - 2\epsilon_{ijk} \bar{\Omega}_j \bar{u}'_k - g\delta_{3i} + [\frac{\bar{T}'}{\bar{T}} g\delta_{3i}] + \nu \frac{\partial^2 \bar{u}_i}{\partial x_j \partial x_j} + \\ + \frac{1}{3} \nu \frac{\partial}{\partial x_i} \frac{\partial \bar{u}_j}{\partial x_j} - \overline{u'_j \frac{\partial u'_i}{\partial x_j}} \end{aligned} \quad (2:9)$$

In this equation we have applied the Boussinesq (1903) approximation which neglects density fluctuations except where they appear with the gravitational force in the form of buoyancy (Dutton and Fichtl, 1969; Ogura and Phillips, 1962). We have, furthermore, substituted

$$\rho' = \left(\frac{\partial \rho}{\partial T}\right)_p T' = -\frac{\rho}{T} T' \quad (2:10)$$

The continuity equation with the Boussinesq approximation may be written as

$$\frac{\partial \bar{u}_i}{\partial x_i} = -\frac{\bar{u}_i}{\bar{\rho}} \frac{\partial \bar{\rho}}{\partial x_i} \cong 0 \quad (2:11)$$

Terms in (2:9) which are set in brackets will vanish if Reynolds averaging rules were strictly applicable. So will the term $2\epsilon_{ijk} \bar{\Omega}_j \bar{u}'_k$, which we have omitted altogether.

It bears on the trivial to mention that many of the terms in (2:9) such as the last one, are impossible to measure with present hardware to the required degree of accuracy, even over confined regions and during limited time periods. Equations stating the conservation of mass and of entropy, together with (2:9), complete the system of equations that describes the mean state of a turbulent atmosphere (see e.g., Landau and Lifshitz, 1959; Lumley and Panofsky, 1964). The terms of these additional equations are equally difficult to balance with presently available measurements, if one wishes to do adequate justice to the variability of atmospheric parameters with x, y, z, t, and k.

2.3 The Interaction of Atmospheric Motion Systems

2.3.1 The Energy Equation of Turbulent Motion

Subtraction of (2:9) describing the mean motion from a form of (2:2) that has been expanded by a substitution of perturbation quantities, thus describing the total motions, yields the momentum equation for the perturbation motion *only*. Multiplying the latter by u'_i yields, after averaging over time, T, the kinetic energy equation for the perturbation motion (Lumley and Panofsky, 1964; Dutton, 1969, Lester, 1970):

$$\begin{aligned} \frac{\partial \bar{e}_{ii}}{\partial t} + \bar{u}_j \frac{\partial \bar{e}_{ii}}{\partial x_j} + \overline{u'_i u'_j} \frac{\partial \bar{u}_i}{\partial x_j} - [\overline{u'_i u'_j} \frac{\partial u''_i}{\partial x_j}] + \overline{u'_i u'_j} \frac{\partial u'_i}{\partial x_j} \\ - \overline{u'_i u'_j} \frac{\partial u'_i}{\partial x_j} = -\frac{1}{\bar{\rho}} \overline{u'_i} \frac{\partial \bar{p}'}{\partial x_i} + \frac{\overline{u'_i T'}}{\bar{T}} g\delta_{3i} + \nu \overline{u'_i} \frac{\partial^2 u'_i}{\partial x_j \partial x_j} \end{aligned} \quad (2:12)$$

In this equation, $\bar{e}_{ii} = 1/2 [\overline{u_1'^2} + \overline{u_2'^2} + \overline{u_3'^2}]$ which is the kinetic energy of three dimensional perturbation motions. Again, terms in brackets will vanish if Reynold's averaging rules are strictly applicable. This is not likely to be the case if high-pass low-pass filtering techniques are applied to an inhomogeneous set of data, such as depicted, for instance, in (F2.4).

With (2:12) we are caught on the horns of the same dilemma as with (2:9): Many of the terms are difficult, if not impossible, to evaluate with the measurement technology applied in the past. Simplifying assumptions had to be introduced, some of them of questionable validity. Dutton (1969) considered the

energy budget of horizontally homogeneous turbulence, where $\partial \bar{e}_{ii} / \partial x_1 = \partial \bar{e}_{ii} / \partial x_2 = 0$. Lester (1970) demonstrated, however, that under conditions of severe clear-air turbulence (CAT) such an assumption is invalid. This becomes evident from the decrease of spectral densities as one proceeds downstream from the region of strong turbulence shown in (F2.4). The one-dimensional spectra for the longitudinal (u_1) component of motion are given in (F2.8) for the various time intervals indicated by letters in (F2.4). We have to assume that the three-dimensional spectra of \bar{e}_{ii} follow a similar, decreasing tendency downstream of the region of severe CAT as shown in (F2.8).

The example given here pertains to CAT. We must suspect a similar, nonhomogeneous behavior of atmospheric motions at the meso-scale. Evidence for this comes from the locally confined nature of squall lines and other convective systems which, according to (F2.1) are prominent members of this "frequency band" of meso-scale atmospheric motions.

2.3.2 Interaction Between Scales of Motion

Following Dutton (1969) we may substitute in (2:12) spectra and cospectra as functions of wave number, k (cycles per meter). We will symbolize spectrum functions in the wave-number domain by $\langle \rangle$. Thus, we obtain

$$\begin{aligned} \frac{\partial \langle \bar{e}_{ii} \rangle}{\partial t} = & - \bar{u}_j \frac{\partial \langle \bar{e}_{ii} \rangle}{\partial x_j} - \langle u'_i u'_j \rangle \frac{\partial \bar{u}_i}{\partial x_j} - \langle u'_j \frac{\partial \bar{e}_{ii}}{\partial x_j} \rangle - \\ & - \frac{1}{\rho} \langle u'_j \frac{\partial p'}{\partial x_i} \rangle + \frac{g}{T} \langle u'_3 T' \rangle - \epsilon(k) + R_{NL}(k) \end{aligned} \quad (2:13)$$

where

$$\epsilon = - \nu \overline{u'_i \frac{\partial^2 u'_i}{\partial x_j \partial x_j}} = \int_0^\infty \epsilon(k) dk \quad (2:14)$$

is the rate of dissipation of kinetic energy by viscous forces and $R_{NL}(k)$ represents the non-linear interactions between various wave numbers. The latter term vanishes if (2:13) is integrated over the total wave number space, thus again yielding (2:12). The first term on the right-hand side is ignored by Dutton

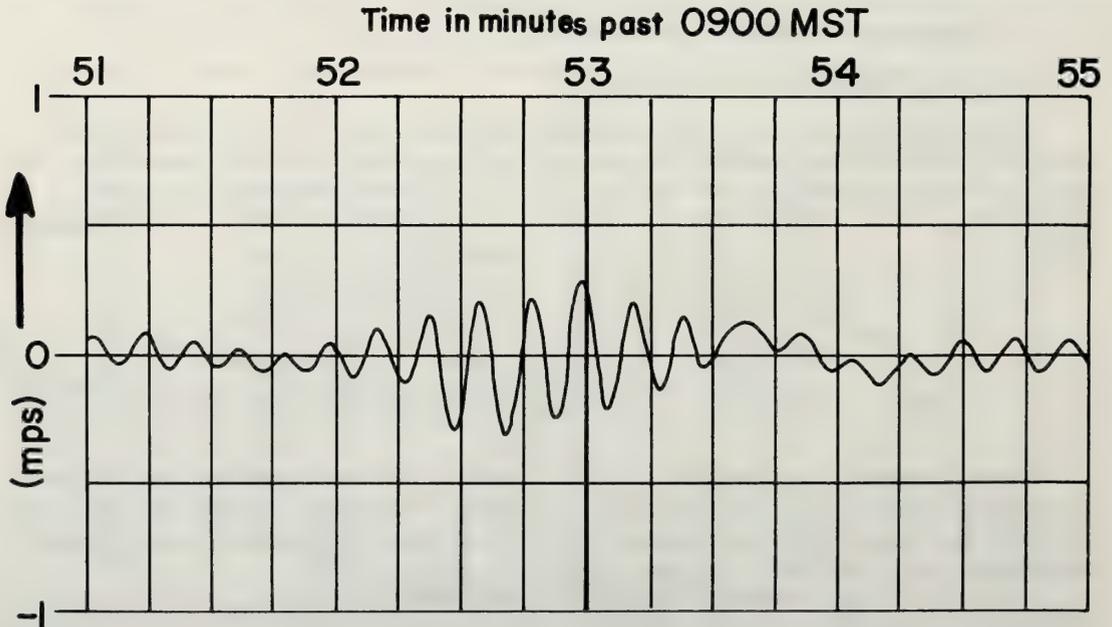


Figure 2.6 Values for $\overline{u'}$ for same filter techniques as used in (F2.5) (after Lester, 1970).

(1969). According to (F2.8), however, it may contribute in an important way to (2:13) under conditions of severe CAT, and in other cases of non-homogeneous fields of atmospheric perturbation motions.

An important quantity which is difficult to measure is $R_{NL}(k)$. Vinnichenko (1970) surmised that the non-linear interactions between the meso-scale and small-scale turbulence are rather small over wide areas of the globe and during long periods of time when no CAT is observed. Under such conditions, spectra of atmospheric motions would drop off at high frequencies, as indicated by the solid line "b" in (F2.2). Figure (2.9) gives a "close-up" view of such spectral behavior of atmospheric motions (spectrum No. 1). Apparently with no CAT present the dissipation of kinetic energy in the atmosphere is at a minimum level, representing a "well-oiled" machinery. Only relatively small amounts of energy "cascade" from the meso- to the micro-scale.

Under conditions of light to moderate CAT [spectrum No. 2 in (F2.9)] a more or less continuous energy transfer seems to take place between meso- and micro-scale and the non-linear interaction term, R_{NL} , becomes significant. With severe CAT [spectrum No. 3 in (F2.9)] energy is "dumped" in excessive amounts into the micro-scale. The value R_{NL} assumes the role of an instability mechanism, such as, for instance, visible in Kelvin-Helmholtz waves. We will deal with such possible mechanisms in 2.4.

Figure (2.2) indicates various regions between small and large scales of atmospheric motions, in which $R_{NL}(k)$ may have interesting implications on energy transitions in wave-number space. I should caution again, however, that (F2.2) contains one-dimensional spectra, whereas the full impact of the term $R_{NL}(k)$ in (2:13) pertains to the three-dimensional wave-number space.

It has been mentioned in 2.1 that the interaction between "bands" of atmospheric motions in the wave-number or frequency spaces presents one of the, if not *the*, fundamental question marks that plague modern meteorology. On the planetary scale of atmospheric motions attempts have been made to estimate energy transfers between wave numbers (see e.g. Steinberg and Wiin-Nielson, 1971; for additional references see Reiter, 1969). The GARP effort addresses itself to the same problem of the meso-scale. The emphasis in recent CAT and mountain-wave investigations has moved forward the same complex problem with respect to the micro-scale of motions. All these efforts, future as well as present, could be termed highly successful, if the spectrum range shown in (F2.2) and implied in (2:13) could be divided into a few discrete bands of finite width, and R_{NL} could be estimated in terms of interactions between these bands, rather than as a continuous function of k .

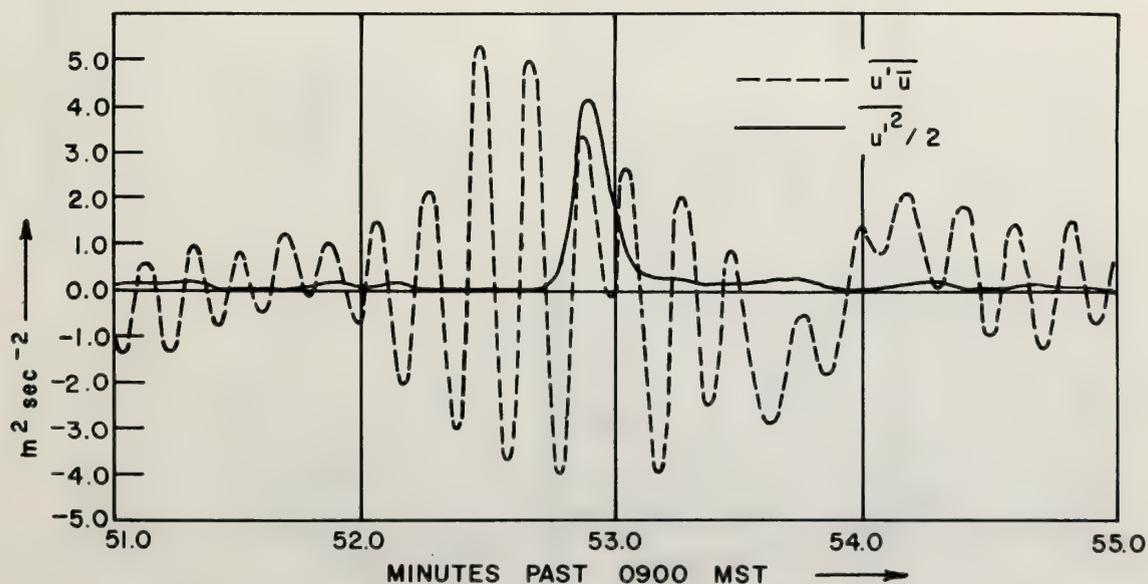


Figure 2.7 Values of $\overline{u'u}$ (dashed line) and $\overline{u'^2}/2$ (solid line) for same filter techniques as used in (F2.5) (after Lester, 1970).

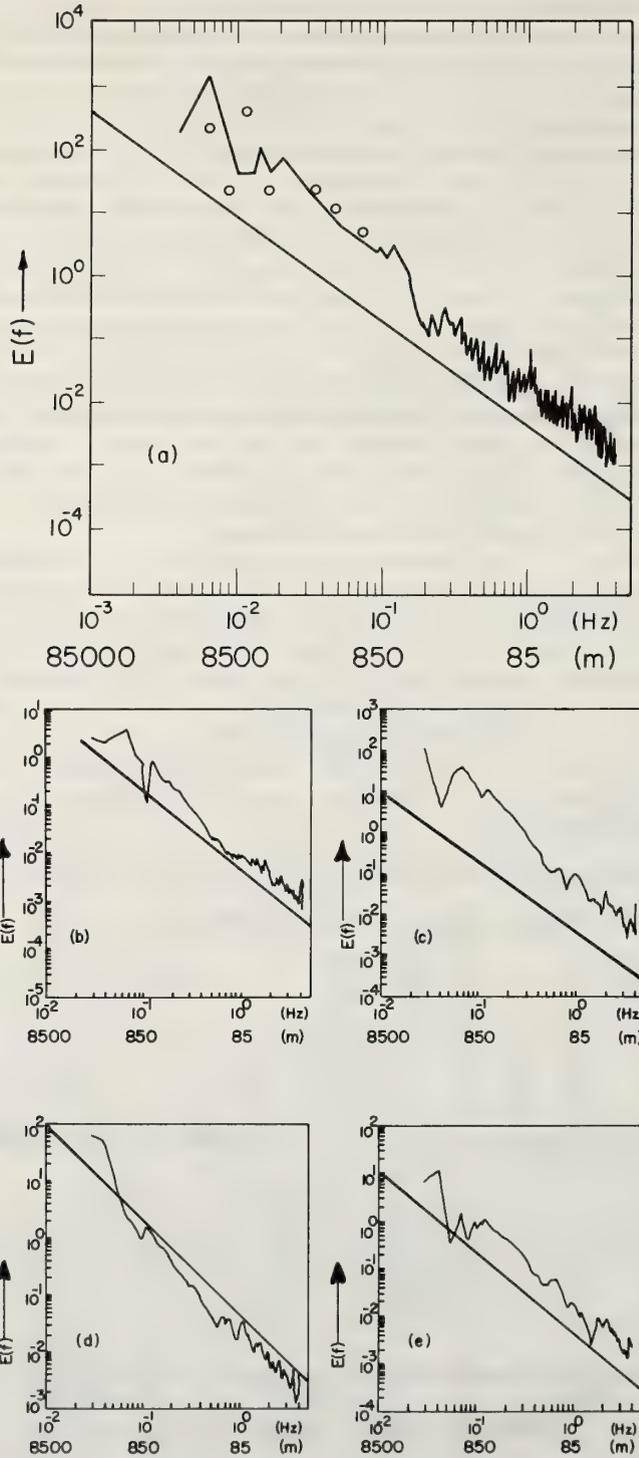


Figure 2.8 Energy spectra for time intervals indicated by letters in (F2.4). Dots represent spectral estimates based on longitudinal component of Doppler winds. $-5/3$ line is entered at same coordinate position for easy reference (from Lester, 1970).

If we wish to attain this "limited" goal of establishing values R_{NL} between wave bands, we will have to achieve a *quantitative* knowledge of instability processes in the atmosphere, ranging from the growth of large-scale baroclinic waves to the breakdown of internal, shallow shearing layers into turbulence. It will not suffice to state that a certain structural configuration of the atmosphere "tends" to be unstable, as for instance diagnosed by Richardson's number (see Richardson, 1920). We will have to know *quantitatively* the degree of instability, its relation to a finite or infinitesimal perturbation that prompts the breakdown of flow, and the amount of energy release into a higher-frequency range of the spectrum of eddy motions by such a breakdown process. Obviously we are a far cry from achieving this goal. In my opinion, a heavy input from field measurement programs will be required to arrive at empirical estimates of these energy releases. Remote sensing of atmospheric structure and flow should play an important role in

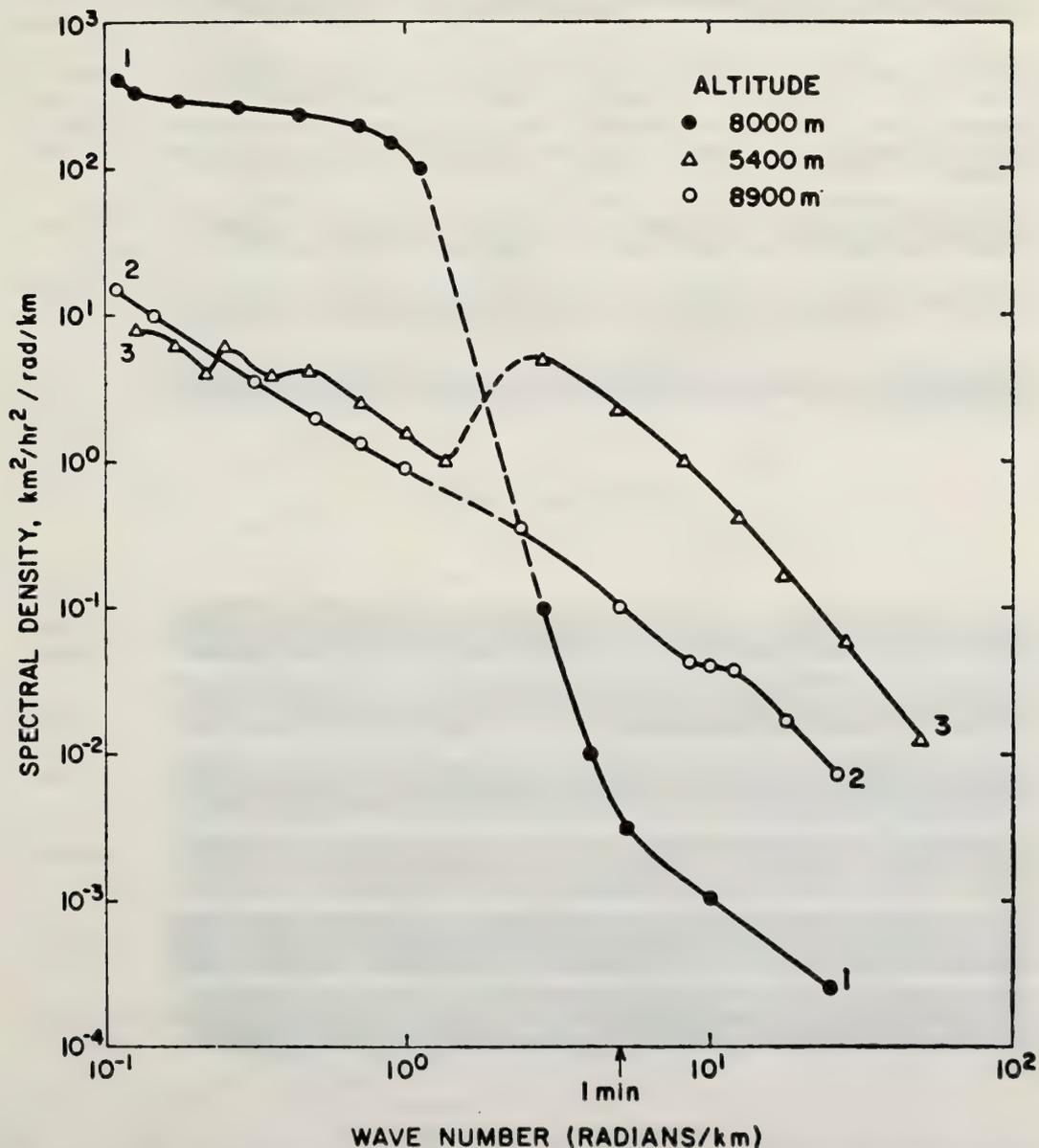


Figure 2.9 Three types of transitions between meso- and micro-scale spectra in the free atmosphere (from Vinnichenko, 1970).

such measurement programs. Quantitative estimates of energy releases, in turn, may serve as input for more sophisticated general circulation and dynamic prediction models.

The statements in the preceding paragraphs implied that the interaction term $R_{NL}(k)$ transfers energy from low to high wave numbers only. Research in the domain of planetary and cyclone-scale wave numbers reveals that this is not necessarily so (for references see Reiter, 1969). There are occasions where cyclone waves “feed” the longer planetary waves. We could envision similar transfers of energy towards lower wave numbers or longer waves to operate, at least occasionally, at the meso-scale. The release of latent heat in a squall line—a typical meso-scale phenomenon—could, for instance, benefit the kinetic energy associated with a cyclone wave.

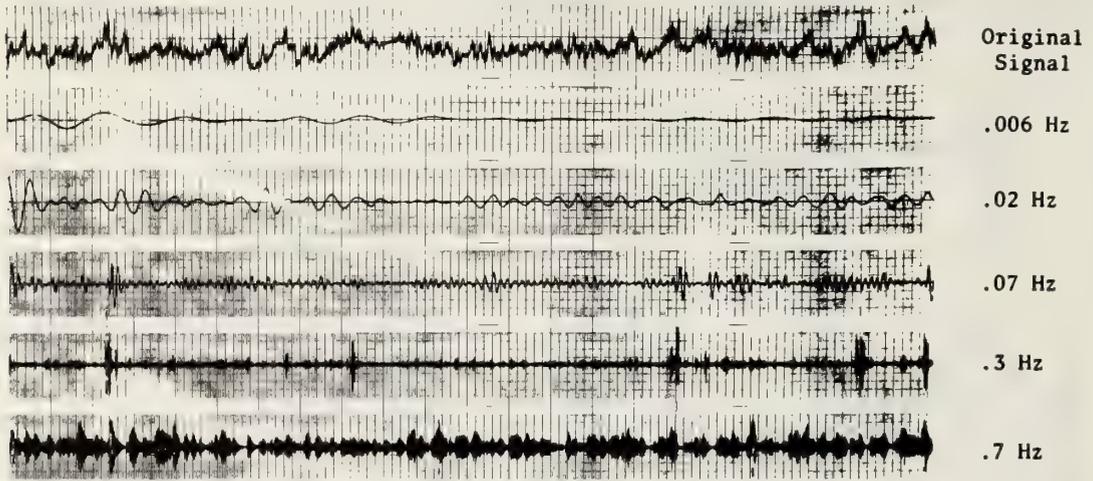


Chart speed 1 cm/sec

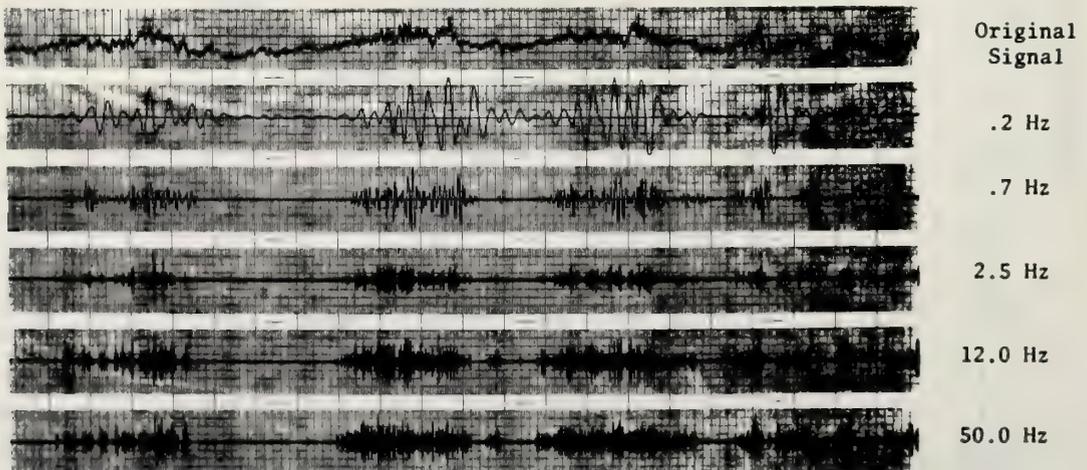


Chart speed .2 cm/sec

Figure 2.10 Original turbulent signal of wind speeds 20 feet above ground, and electronically filtered signals at various frequencies (after Stankov, 1970).

Even if we assume that the energy transfer in the domain of small-scale turbulence, e.g. in CAT, occurs exclusively in a cascading mechanism from larger eddies towards smaller eddies, we will have to grant the atmospheric circulation at the meso-scale and at even larger scales the possibility of moving kinetic energy “up” or “down” the spectrum of wave lengths and eddy sizes. At which wave number this schizophrenic behavior of the atmosphere begins, and what prompts a specific behavior, we do not understand at this time.

To illustrate the point in question let us, again, consider a convective cloud system: Within its “hot towers” a certain amount of potential energy is converted into kinetic energy. Part of this energy will be dissipated into small-scale turbulence within the rising air and between the up- and down-draft regions and their environment. Part of the released energy will be exported into larger-scale circulation features, such as jet streams.

2.4 Unsolved Problems

The problems outlined in the foregoing discussion have an immediate bearing upon a number of research areas which are presently under investigation. I will attempt to point out a few such “dangling loose ends” without raising the claim of providing a complete shopping list of open questions in atmospheric science.

2.4.1 Small Scale

Turbulence theories presently in vogue concern themselves with “equilibrium” conditions. This implies that, among other things, the spectral densities in a given frequency band are not time or space

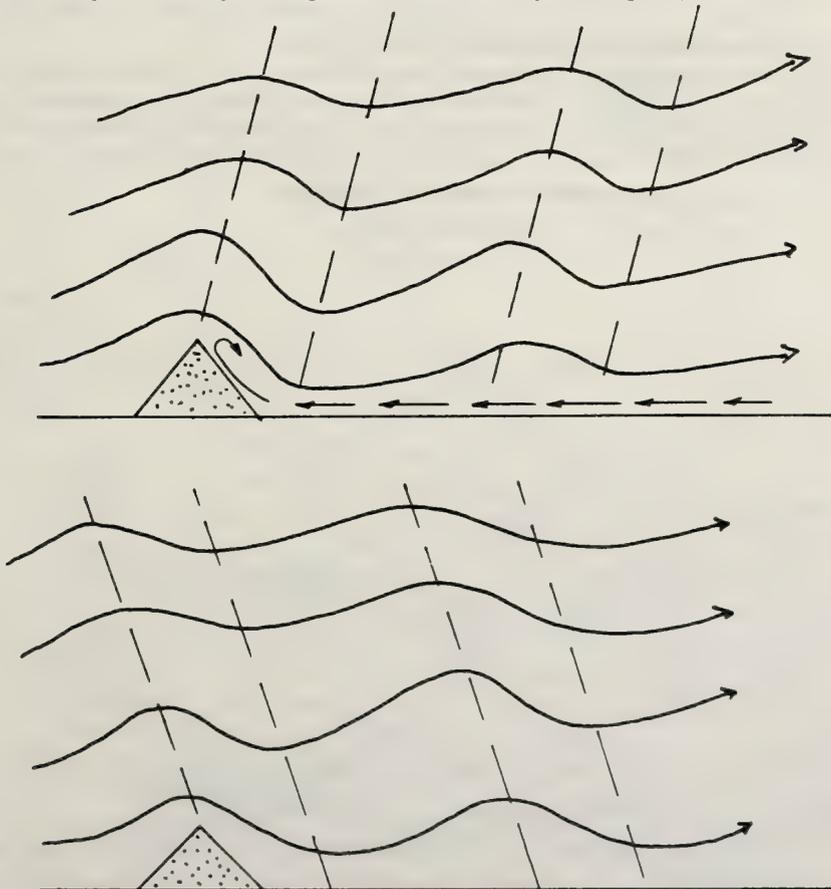


Figure 2.11 Schematic sketch of lee-waves with forward (a) and (b) backward leaning wave fronts.

dependent. Equation (2:1) satisfies these conditions for a one-dimensional spectrum. How do we approach cases, however, in which bursts of strong turbulence are embedded in a background of weaker turbulence? Figure (2.8) and its comparison with (F2.4) already indicated some of the consequences: Obviously a spectrum analysis of a relatively long time record that extends over individual turbulent bursts and over the more quiescent portions of the record as well (F2.8a) will not do justice to the kinetic energy within, and the behavior of, the confined patches of strong turbulence. These patches, in most instances, quite obviously are not in an equilibrium state of turbulence. Under equilibrium condition they would receive as much energy through $R_{NL}(k_L)$ from lower-frequency eddy phenomena as they pass on through $R_{NL}(k_H)$ to higher-frequency eddies and to viscous dissipation. In the formative stages of such a turbulent patch, we should expect that $R_{NL}(k_L) > R_{NL}(k_H)$, so that kinetic energy is allowed to build up over a finite range of the turbulent spectrum. The opposite inequality would hold during the dissipating stage of the patch (F2.8).

In order to understand the time-history of $S(k)$ in such a turbulence patch it will be necessary to estimate the "normal" rate of dissipation, ϵ , of the equilibrium background turbulence into which this patch feeds its excess energy. In addition we will have to know the source strength and its time variation (possibly along a stream line) that provides the excess energy, and the release of excess energy to the equilibrium environment.

As the energy source of non-equilibrium patches of CAT we may identify in most cases the vertical wind shear which may lead to unstable Kelvin-Helmholtz waves (see e.g., Phillips 1967; Hardy, Glover, and Ottersten, 1969; Bretherton, 1969; Woods, 1969; Scorer, 1969; Thorpe, 1969). These are gravity waves which amplify until their crests degenerate into sharp peaks that finally bend over and disintegrate into small-scale turbulence. With turbulence under near-neutral conditions bouyant forces may provide an energy source (see 2:12). Figure (2.10) shows that intermittent turbulent patches not only are characteristic for CAT, but also for turbulence in the planetary boundary layer.

The rate at which excess energy is released from the patch to the "equilibrium turbulence" of the environment most likely is a function of ϵ of this environmental turbulence. Sophisticated measurement programs, preferably using remote sensing techniques that do not disturb the turbulent patch during its development and decay, will be needed to specify such a functional relationship. A solution to this

WIND SPEED — MAGNY-LES-HAMEAUX
11-12 MARCH 1968

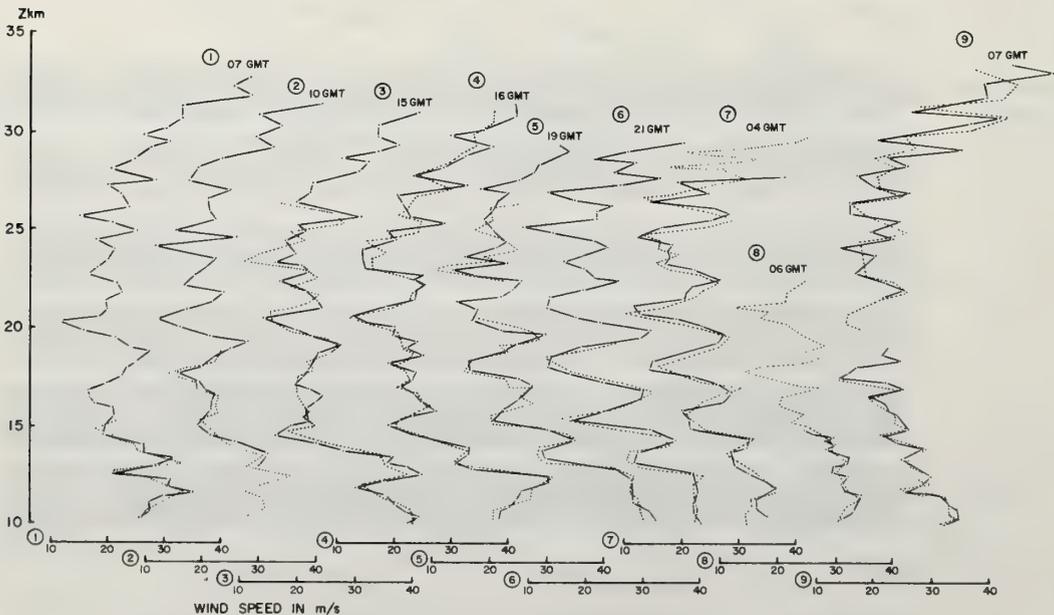


Figure 2.12 Vertical wind profiles (m/sec) at Magny-les-Hameaux, France, 11-12 March, 1968 (after Barbé et al., 1971).

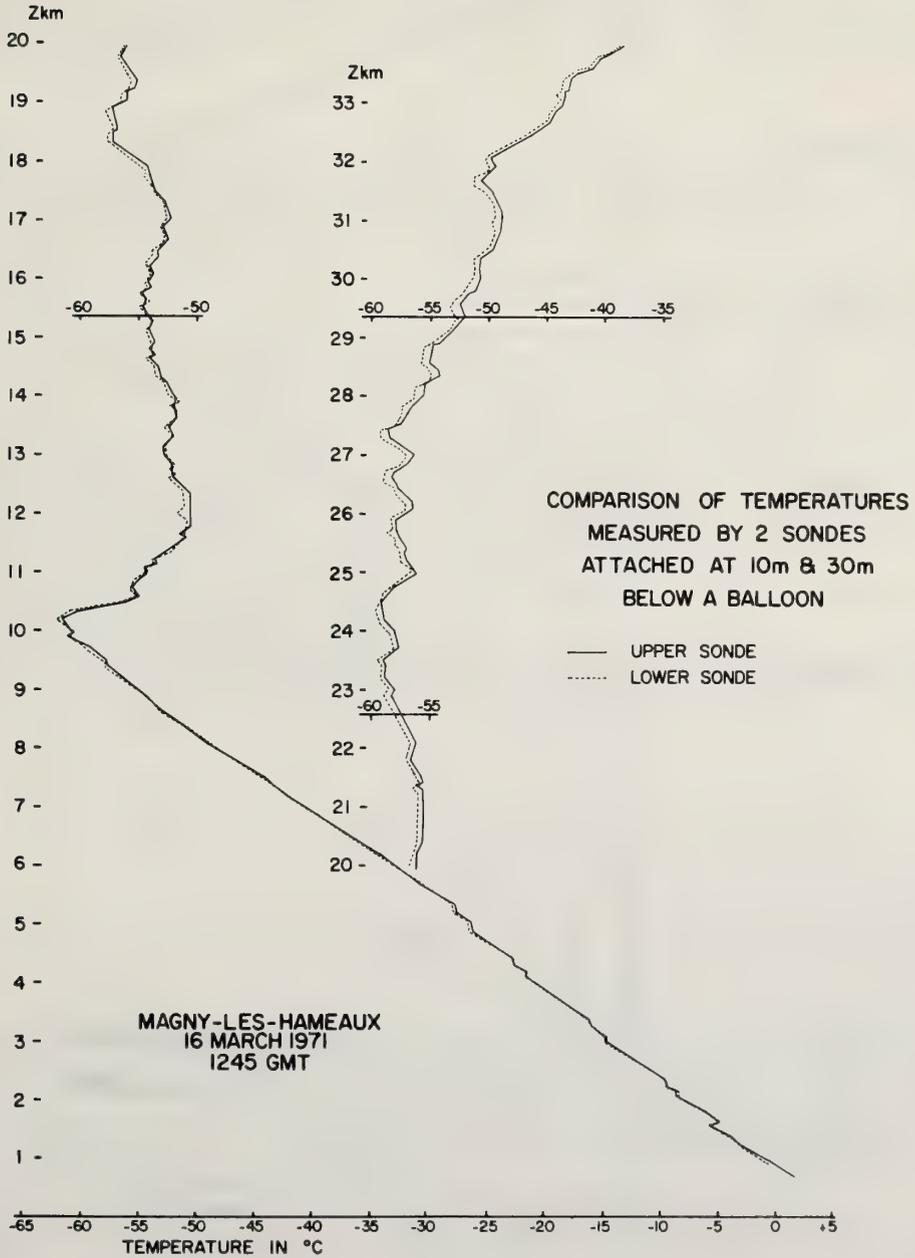


Figure 2:13 Vertical temperature profiles measured by two sondes at 10 m and 30 m distance from the same balloon, at 1245 GMT, 16 March 1971, Magny-les-Hameaux, France (after Barbé et al., 1971).

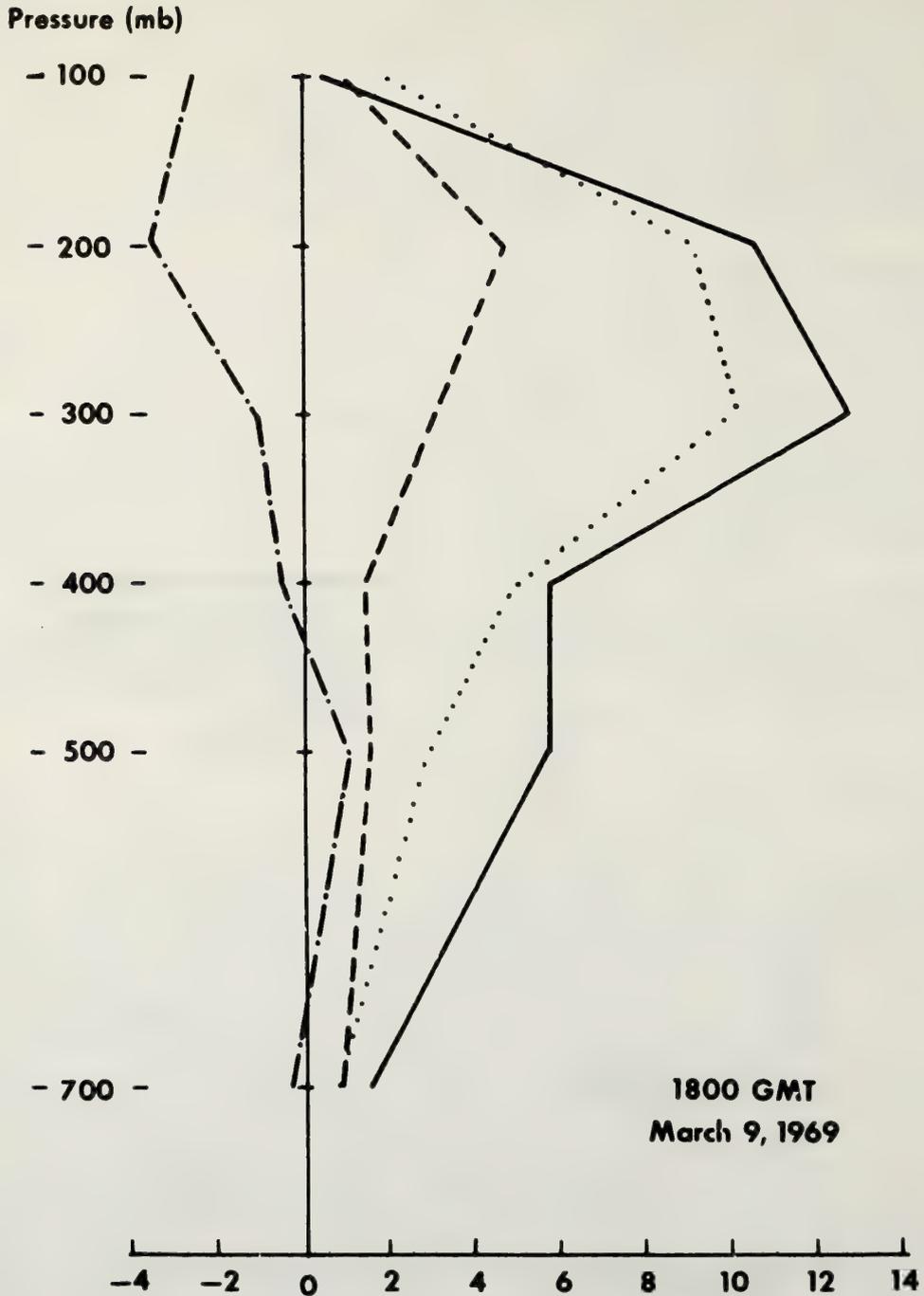


Figure 2.14 Individual terms in the zonal stress gradient computation over an area covering the southwestern United States. The dotted line is the inertial term du/dt ; dashed line is $g\partial z/\partial x$; dot-dashed line is the Coriolis term $(-fv)$; solid line is F_x , or $-\rho\partial\tau_x/\partial p$. Units are 10^4 sec^{-2} (10^2 cm sec^{-2}) (after Wooldridge, 1970).

problem will greatly aid in the assessment of the danger to small aircraft from wing-tip vortices shed by large jumbo jets.

I should emphasize again, as I have done before, that a one-dimensional spectrum will not suffice to estimate the non-equilibrium behavior of turbulent patches. The time history of eddy kinetic energy will have to be explored in the three-dimensional wave-number space. Such treatment will also shed additional light on the significance of a spectral "gap", its presence or absence (Bretherton, et al., 1969), and its influence on the magnitude of the non-linear interaction term $R_{NL}(k_G)$ where k_G signifies the wave-band characteristic of the "gap" region.

2.4.2 Mesoscale

Considerable efforts are presently under way to explore the convective processes in and above the planetary boundary layer. It is by such processes that, especially in the tropics, the bulk of latent and sensible heat is released into the upper troposphere. "Hot tower" convection, thus, becomes an important driving mechanism of the general circulation of the atmosphere.

There are, however, more implications that lend emphasis to convective processes, especially within the planetary boundary layer. It is within this layer that the angular momentum exchange between earth and atmosphere takes place. The "Ekman spiral" that crudely characterizes flow conditions in the planetary boundary layer also provides the mass flow from high to low pressure regions. Frictional forces in the planetary boundary layer and their diurnal variation appear to be an important factor in generating local wind systems, such as the low-level jet stream over the Great Plains (for a summary see Reiter, 1972).

In spite of the obvious importance of the flow processes within the planetary boundary layer, or "Ekman layer," in the maintenance of the general circulation, in air pollution transport, etc., this layer of the atmosphere remains relatively poorly explored and understood. The Ekman (1905) theory itself predicts a departure of the surface wind by a 45° angle from the direction of the geostrophic wind. This is a gross overestimate compared to real conditions. The theory is totally inadequate at the equator, where the wind direction jumps from 45° to the *left* of the geostrophic direction in the northern hemisphere, to a 45° departure to the *right* in the southern hemisphere.

Grisseier and Jacobsen (1970) suggest a remedy for these shortcomings by averaging the equations of motions twice, with two different high-pass filters. One filter would consider small-scale mechanical turbulence as it enters the (constant) eddy viscosity coefficient K in the Ekman spiral

$$\begin{aligned}\bar{u} &= \bar{u}_g (1 - e^{-az} \cos az) \\ \bar{v} &= \bar{u}_g e^{-az} \sin az \\ a &= \sqrt{f/2K}\end{aligned}\tag{2:15}$$

The second high-pass filter takes into account the convective-scale motions in the planetary boundary layer. With this, Grisseier and Jacobsen (1970) arrive at an increased vertical exchange of momentum in the boundary layer, that increases the wind speed at the earth's surface over that which would be obtained from (2:15), and at the same time reduces the angle of departure of the surface wind vector from the geostrophic wind direction. Also a continuous shift of wind direction is obtained near the equator.

If we define the eddy viscosity coefficient as

$$K = -\frac{\overline{u'w'}}{\frac{\partial \bar{u}}{\partial z}} = \varrho^2 \frac{\partial \bar{u}}{\partial z},\tag{2:16}$$

the dependence of K on the scale of atmospheric motions is obvious. Specifically, as suggested in (2:13), we may replace $\overline{u'w'}$ by the cospectrum $\langle u'w' \rangle$. The basic Ekman theory assumes that this cospectrum function does not contribute towards the value K beyond a certain cut-off wave number of atmospheric

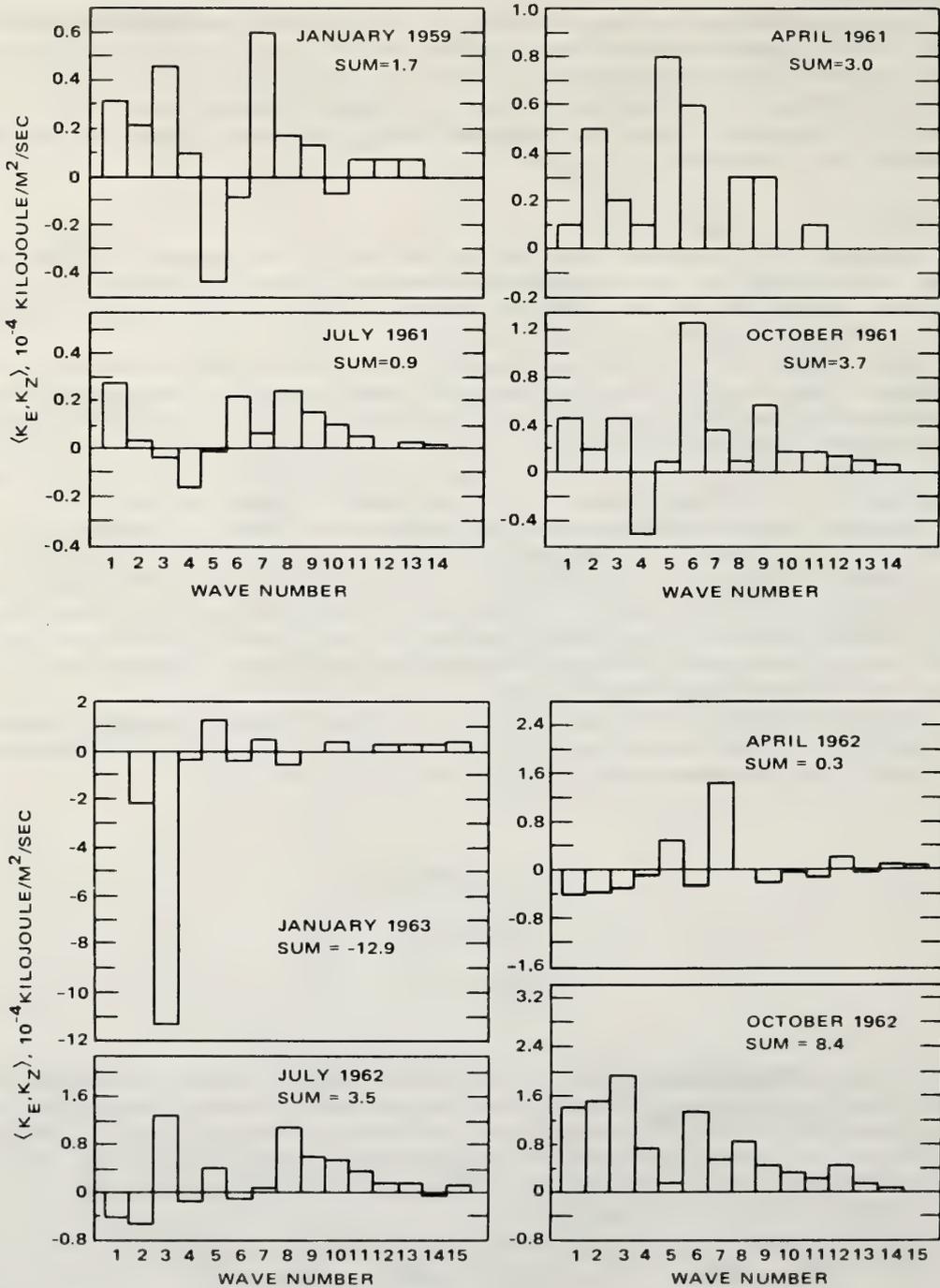


Figure 2.15 Energy exchange from eddy to zonal kinetic energy as a function of wave number for the months indicated. Upper four diagrams based on 850- and 500-mb data and lower four diagrams on 850-, 700-, 500-, 300-, and 200-mb data (from A. Winn-Nielsen, J. A. Brown, and M. Drake, 1964).

motions that describes the boundary in wave-number space of "small-scale" turbulence. This is a serious shortcoming of the "K-theory" that not only governs the interpretation of vertical wind profiles in terms of turbulence effects, but also the diffusion theories which fall back upon the Fickian diffusion equation. Panchev (1968) considered the dependence of K on the scale of atmospheric motions with scales still larger than the meso-scale and arrived at a function dependence of K on the spectrum of the *horizontal* wind (for a summary see Reiter, 1970).

It should be pointed out that $\partial\bar{u}/\partial z$ in the denominator of (2:16) also varies with the length of the time interval, T, of averaging as well as with the time coordinate, t, as may be gathered from (F2.4). This variation will be especially significant if high-pass and low-pass filtering techniques have been employed in arriving at u' , w' , and \bar{u} , respectively.

As has been shown by Lettau (1967), Bonner and Paegle (1970) and others (see Reiter, 1972), the behavior of the eddy viscosity coefficient in the planetary boundary layer, especially its diurnal variation, is instrumental in generating local wind systems, such as the low-level jet stream over the gently sloping Great Plains. A better understanding of K as a function of atmospheric scales of motion would greatly enhance our capabilities of modelling such local circulation phenomena.

Our discussion of meso-scale problems to this point emphasized the convective processes, mainly in the planetary boundary layer. These processes have two characteristic features in common with the small-scale turbulence which we discussed in the preceding chapter: They are more or less random in nature, and they transport mass properties *along* their gradient. (The heat transport into the upper troposphere *against* the mean vertical gradient of potential temperature is brought about by the release of latent heat during condensation of water vapor).

In addition to such convective motions, we have a whole spectrum of organized wave phenomena, from short gravity waves with wave lengths of tens of meters, to gravity inertia waves of several hundred kilometers length. We are disregarding in this discussion horizontally two-dimensional waves, such as planetary Rossby waves. Our concern rests with those wave modes that have significant vertical amplitudes.

It can be shown that such waves may transport sizeable amounts of momentum and kinetic energy vertically through the atmosphere. Such transports, in essence, may be studied by perturbation equations, such as (2:9) or (2:12). A typical example of gravity waves occurs in the lee of mountains (see e.g. Scorer, 1949). Figures (2.3) and (2.4), top, reveal the existence of a family of such waves in the potential isotherm pattern and in \bar{u} , respectively. As may be seen in this diagram, the waves generated in a stably stratified atmosphere flowing over corrugated terrain are neither very regular in their appearance, nor can we assume *a priori* that the wave pattern is stationary in time—even though current lee-wave theories give this impression. The complexities of the underlying surface which forces the wave motion, together with the complexity of atmospheric structure in the temperature and velocity fields, subject to time and space variations, give rise to the observed inhomogeneities that depart from the relatively simple patterns obtained from theoretical modelling approaches.

Because of these inhomogeneities in the observed gravity wave structure, we should not expect such waves to appear as singular "spikes" in spectra obtained, for instance, from aircraft traverses of a lee-wave train. Rather flat "humps" in the representation of one-dimensional spectra should be found instead. Thus, at first glance, a spectrum obtained from flights through a layer with convective motions may look rather similar to one that was obtained under gravity-wave activity. Nevertheless, we have to allow for very distinctive differences in the transport mechanisms by waves as opposed to those by convective motions.

We have argued previously that the momentum exchange in convective motions may be measured by the co-spectra $\langle u'w' \rangle$ (see 2:13) which may be used in the calculation of the eddy viscosity coefficient, (2.16). The same eddy exchange or "austausch" concept, that we have applied to momentum, may be advanced to mass characteristics, such as heat, specific humidity, hence radio refractive index, chemical composition, etc. Thus we arrive at definitions of eddy exchange coefficients for heat, K_H , and for neutral admixtures, K_N , which are numerically of the same order of magnitude as the eddy viscosity coefficient, or the eddy exchange coefficient of momentum, K_M (Businger, et al., 1970; for additional references see Reiter, 1972).

Equation (2:16) also gives numerical results for wave motions, especially if these waves are structured so that they will transport momentum (and energy) vertically. Momentum is generally transported upwards in a system of gravity or lee-waves, if the wave fronts lean forward in the vertical (F2.11); similarly, momentum is carried downward, if the wave fronts lean back. Laminar wave flow will,

however, *not* transport mass and its associated characteristics (such as heat, chemical admixtures, etc.) normal to the flow direction. Thus one should expect large discrepancies between K_M , K_H , and K_N .

The *real* atmosphere appears to be neither truly “undulant” nor truly “turbulent,” except under very ideal conditions. Even with well-developed gravity waves, such as they are often observed in the lee of mountains, random turbulent eddies with their specific transport characteristics of kinetic energy, momentum, and mass properties are super-imposed upon a quasi-laminar “undulance” that does not contribute to the vertical mass transport. Unfortunately, one-dimensional spectral presentation of wind data obtained, for instance, from aircraft measurements, does not present a strict separation between “undulance” and “turbulence.” More sophisticated data analysis techniques will have to be advanced before such a distinction can be made.

The unscrambling of undulance and turbulence is further complicated by the interaction between these two modes of flow: Wave action in a stable stratification may generate super-critical wind shears, giving rise to a breakdown of the wave flow into turbulence (Phillips, 1967). The Kelvin-Helmholtz and Görtler instabilities are typical examples in question.

Such a breakdown of waves into turbulence may occur relatively frequently in the atmosphere. Figure(2.12)shows a detailed vertical wind profiles obtained in the stratosphere by accurate radar soundings (Barbé, 1971). Temperature profiles with alternating stable and adiabatic regions indicate the presence of turbulent processes (F2.13). It has been argued by Weinstein, Reiter and Scoggins (1966) that the oscillating wind profiles, such as the ones shown in (F2.12), are a manifestation of gravity-inertia waves. The turbulent exchange of heat, which generates adiabatic layers within an originally stable region, on the other hand, signals the past (“fossil”) or present action of turbulence in that region.

Our present knowledge of gravity-inertia waves in the atmosphere is much too rudimentary to allow an assessment of all possible consequences on atmospheric structure and dynamics. Present measurement

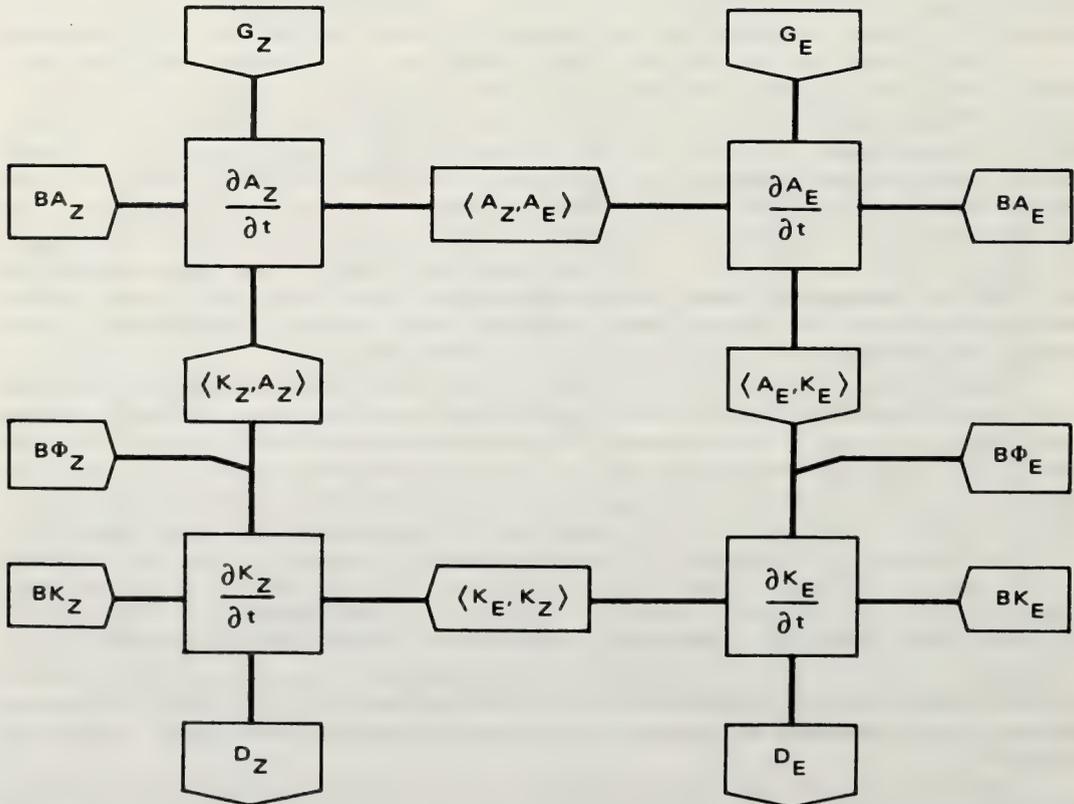


Figure 2.16 Schematic diagram of energy balance. Pentagons are pointed in the direction of energy transfer for positive values of these processes (adapted from Muench, 1965).

techniques are inadequate to measure the wave lengths, phase velocities, and wave amplitudes of these waves. Since we have to suspect that phase velocity is a function of wave length over a wide range of such waves, we have to be prepared to deal with group velocity effects as well. This point may be illustrated by considering a typical lee-wave case. There we find "standing" lee waves with a wave length of the order of 10 km and a phase velocity of zero, but superimposed upon these are shorter gravity waves within shallow shearing layers, visible as cloud "ripples", that appear to travel with the mean wind of the layer. It may well be that certain cases of severe CAT encountered by the "Concorde" prototypes over the Atlantic, seemingly in the "middle of nowhere", were caused by a fortuitous superposition of various gravity-inertia wave modes which, by a group velocity effect, may have generated locally subcritical Richardson numbers.

A detailed knowledge of both the spectrum and the behavior of gravity and gravity-inertia waves in the atmosphere is more important than only as an academic exercise. Wooldridge (1970) has shown that the frictional term in the equations of motion, which accounts for sub-grid scale momentum transports, may be of sizeable magnitude in lee-wave situations (F2.14). Similar conclusions were reached by Lilly (1971). It appears that in the case of sharp westerly (chinook) winds at the earth's surface to the lee of the Rocky Mountains, there is a direct momentum transport from jet stream level to the ground. One part of this momentum is transported by "undulance" in backward-leaning waves (F2.11), part by turbulence. Both modes of transport appear to be directed downward. With the easterlies at the earth's surface, "undulance" moves westerly momentum upward in forward-leaning waves. (Present lee-wave theories are unable to handle such a flow stratification of westerlies on top of low-level easterlies.) Small-scale turbulence, manifest by CAT near the jet-stream, would still transport momentum downward in the direction of the momentum gradient. Unfortunately, at this point, we do not have direct measurements of the cospectra of $\langle u'w' \rangle$ in a forward-leaning wave case that would substantiate the upward flux of momentum at relatively low frequencies, and the downward flux at high frequencies. Wooldridge's (1970) momentum balance estimates suggest that during periods of easterlies at the earth's surface, and with widespread wave activity (as observed from Apollo VI photographs) the upward transport by "undulance" exceeds the downward transport by turbulence.

From the foregoing, it becomes quite obvious that the parametrization of "frictional" processes on the subgrid scale by using values of K_M in conjunction with an "austausch" hypothesis for momentum fluxes may lead to considerable misjudgements. Wave activity may very well entail kinetic energy and momentum fluxes which are directed *against* the mean gradient, thus giving rise to "negative viscosity" (Starr, 1968). It appears to me that the long-range forecasting goals which we have set for ourselves, and to which efforts such as GATE and GARP are major contributions, will be impossible to attain if we fail to achieve a detailed understanding of the spectrum of waves in the atmosphere.

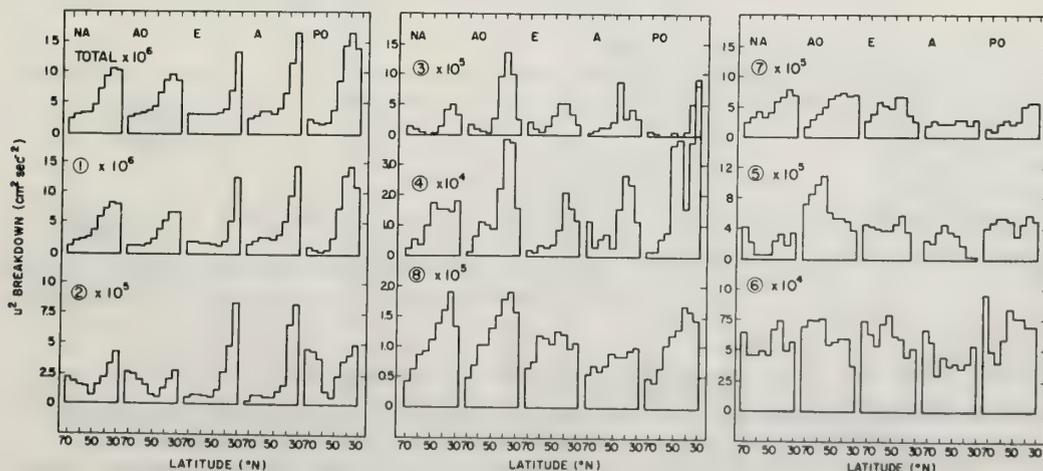


Figure 2.17 The breakdown of the u^2 component of specific kinetic energy into baroclinic, barotropic, transient, standing, eddy, and mean meridional components according to (2:17). The terms are plotted for January, 1964 by latitude and for geographic sectors indicated in Table 2.2.

2.4.3 Large-Scale Motions

The problem of non-linear interaction between various frequency or wave-number domains, which has been described earlier as being *the* fundamental problem that plagues modern meteorology could be stated in slightly different terms: Dynamic meteorology gives us certain criteria that predict certain instabilities in large-scale flow, e.g., dynamic instability with excessive anticyclonic shear, baroclinic instability, etc. These criteria are a long cry from an actual assessment of the amount of kinetic energy that is released into a specific wave-number (or into a band of wave-numbers within the spectrum of large-scale motions), and of the rate at which this release occurs. We should know, furthermore, whether this release is accomplished by a re-distribution of kinetic energy within the spectrum of motions, or by releases of potential energy within the same wave-number domain, or by a combination of both processes.

We have to give credit to a number of noble efforts that have tried to deal with the energy transfer between different planetary wave numbers (see Reiter, 1969, Chapter 4; Steinberg and Wiin-Nielsen, 1971). These efforts are based mainly upon a statistical treatment of horizontal wind components in a latitude band of one hemisphere and on a given pressure surface. To the best of my knowledge, we have not yet advanced far enough to “marry” such statistical treatments to the dynamic principles by which the atmosphere works. Again it is the poorly understood non-linear processes which make atmospheric perturbations grow at certain rates that defy our efforts of formalistic attack. The advent of a new generation of ultra-fast computers may help us in bridging these gaps in our knowledge.

The problem of “understanding”—to the point of predictability—the large-scale circulation of the atmosphere is compounded by the many degrees of freedom which the atmosphere has in executing the seemingly simple task of transferring heat and angular momentum between equator and pole, thereby maintaining a climatic balance.

As an example, (F2.15) shows the energy transformation from eddy to zonal kinetic energy as a function of wave number. It is quite obvious from these data, that during the same months of different years, the energetics of the atmosphere may follow drastically different cycles. The relatively “simple” pattern of energy transfers between zonal and eddy modes (subscripts Z and E) and between available potential (A) and kinetic energy (K) given in (F2.16) as originally designed by Lorenze (1955) and Muench (1965) does not suffice to describe the complicated working of the general circulation. Even in this “simple” pattern of (F2.16) it is difficult to supply reliable numerical values from the existing radiosonde networks of the northern hemisphere for all boxes. For the southern hemisphere such numerical estimates, at the present, appear to be a hopeless task. Wooldridge and Reiter (1970) have shown, as have a number

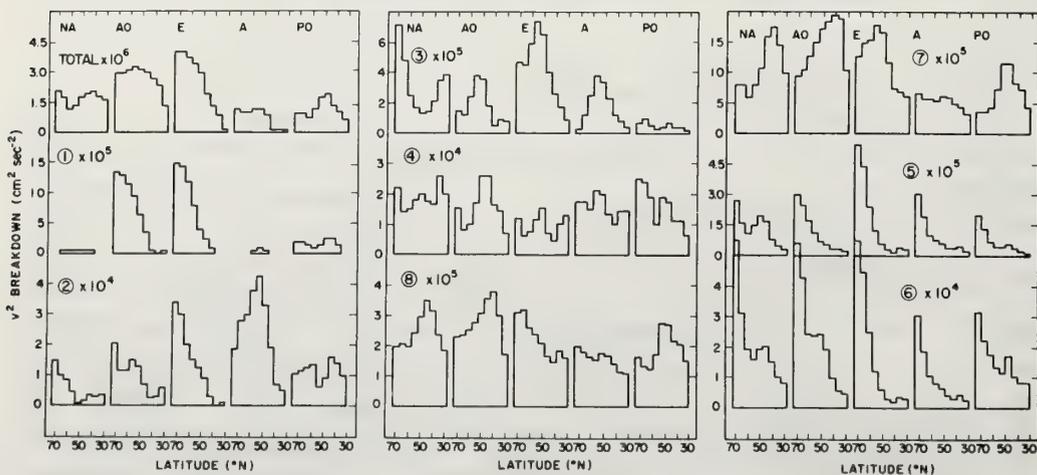


Figure 2.18 The breakdown of the v^2 component of specific kinetic energy into baroclinic, barotropic, transient, standing, eddy, and mean meridional components according to (2:17). The terms are plotted for January 1964 by latitude and for the geographic sectors indicated in Table 2.2.

of other authors (see, e.g., Reiter, 1969a), that the circulation of the southern hemisphere behaves quite differently from the circulation of the northern hemisphere during equivalent seasons. We cannot simply apply our findings from the northern hemisphere on a global basis.

The energy cycle shown schematically in (F2.16) may be subdivided by introducing a more sophisticated multivariate treatment of energy. Wiin-Nielsen (1962) suggested a division of kinetic energy into a barotropic contribution from vertically averaged mean winds, and a baroclinic contribution that contains the effects of vertical wind shears measured by the local departures of winds from the layer-mean wind (for details see Reiter, 1969, also Reiter and Glasser, 1969; Reiter, 1971).

Equation (2:17) gives a breakdown of kinetic energy in pressure, longitude, and time coordinates:

total kinetic energy of u_i component, averaged over $p, \lambda,$ and t	$[u_i^2] (p, \lambda, t)$
1 barotropic, standing, mean meridional	$[u_i]^2 (p, \lambda, t)$
2 baroclinic, standing, mean meridional	$+ [([u_i] (p))]^2 (\lambda, t) (p)$
3 barotropic, standing, meridional eddy	$+ [([u_i] (p)) (\lambda)]^2 (t) (\lambda)$
4 baroclinic, standing, meridional eddy	$+ [([u_i] (p, \lambda))]^2 (t) (p, \lambda)$
5 barotropic, transient, mean meridional	$+ [([u_i] (p, \lambda))]^2 (t) (t)$
6 baroclinic, transient, mean meridional	$+ [([u_i] (p)) (\lambda)]^2 (t) (p, t)$
7 barotropic, transient, meridional eddy	$+ [([u_i] (p)) (\lambda, t)]^2 (\lambda, t)$
8 baroclinic, transient, meridional eddy.	$+ [(u_i)^2 (p, \lambda, t)] (p, \lambda, t)$

(2:17)

Brackets indicate averages with respect to the coordinates given by the parenthesized subscripts. Parentheses symbolize departures from the average formed along the coordinate given by the subscripts (Reiter, 1969a, b). According to this table, we may identify 8 different contributions towards the total kinetic energy, each carrying a specific meaning. Figures(2.17)and(2.18)show a breakdown of u^2 and v^2 for January, 1964 into these eight terms, for 9 latitude bands for the 5 longitude sectors defined in Table(2.2)

Table 2.2 Longitudinal Sectors Used in (F2.17) and (2.18)

North America	NA	130°W- 60°W
Atlantic Ocean	AO	80°W- 10°W
Europe	E	5°E- 75°E
Asia	A	75°E-145°E
Pacific Ocean	PO	145°E-145°W

As if (2:17) were not complicated enough, we may expand each term that contains departures from a time average in frequency space by spectrum analysis, considering that

$$[(u_i)^2 (t)] (t) = \int_0^{\infty} [(u_i)^2 (t)] (t) S(f) df \quad (2:18)$$

A similar expansion could be worked out in wave-number space for terms that contain departure from a longitudinal average. Vinnichenko (1970) suggested that

$$S(k) = \frac{C_g}{2\pi} S(f) \quad (2:19)$$

where C_g is the group velocity of large-scale waves or eddies whose phase velocities depend on wave length or eddy size.

Figure (2:15) indicated that the behavior of the general circulation in the wave-number domain may alter drastically from year to year within the same month and the same season. By virtue of (2:19) we have to suspect the same variations for frequency space. We are far from an understanding of why this is so.

Figure(2.16) indicates rather inadequately the various paths of energy transformation in the atmosphere. If we could ever predict, from a set of initial conditions, the path that the atmosphere is about to adopt in its energy transfer and transformations for a coming month or a coming season, we would have advanced long-range forecasting by a significant step.

2.5 Conclusions

The foregoing discussion hardly presents a complete shopping list of current problems in meteorology, let alone those in atmospheric science. One could easily fill a book with questions that are begging for answers. Instead of providing a lengthy catalog of possible Ph.D. dissertations, I have attempted to focus attention on a relatively coherent problem that permeates a wide area of concern in present meteorology; i.e., the lack of understanding of the interactions between different scales of motion. It is this lack of knowledge that curtails the success of even the most sophisticated forecasting and numerical modelling schemes of atmospheric circulation patterns on every scale which have been designed to date. This also hinders to a considerable extent the reliability with which we can estimate the impact of planned and inadvertent weather modification on local and global scales.

It is quite clear to me that a good number of questions raised in the foregoing chapter cannot be dealt with easily. The study of scales of atmospheric motion and their interaction calls for sophisticated measurements and data analysis techniques. One might justifiably state, that *the* problem of atmospheric science, even more fundamental than the set of problems outlined on the preceding pages, is to cope with the demand for more adequate data on our environment.

If we look at the bulk of data presently available to the meteorological researcher, we realize immediately that solutions to the challenging problems that lie ahead are bogged down by pedestrian approaches. Our techniques of collecting, assimilating, massaging, and comparing data have not progressed much since the "horse-and-buggy" days of atmospheric science. It is true that technology has provided us with sophisticated tools, from satellites to ground-based remote sensors. High-speed computers are available to digest information almost as rapidly as it can be generated. The breakdown occurs when one comes to the point where the human mind has to "make sense" of the avalanche of factual information that pours in from sensors and computers. We have progressed very little in designing a digestive apparatus that would help to assemble information beyond the present computer output stage, and that would make the multitude of information bits more palatable to the human mind. Weather display, for instance, still uses the wall-paper approach of charts and graphs, similar to those used forty years ago. The fact that these graphs are generated by a computer rather than by a technician with a green eye-shade alters little of the obsolescence of the approach.

If we aim to accomplish certain of the goals hinted at in the preceding pages we will have to take much bolder steps in the direction of data assembly and display techniques. Electronic technology should not be called upon only to deliver millions of information bits from a satellite. It should also help us to make *immediate* use of *all* of this information and prevent its storage in acres of gray filing cabinets marked "For Future Reference."

2.6 References

- Barbé, G. D., (1971), Structure de la stratosphere moyenne et elevee. Detection de la turbulence en ciel clair. Application aux vols envisages pour les avions de transport supersoniques. Paper presented at the Joint Conference on Aeronautical Meteorology, 24-26 May, 1971, Paris.
- Bonner, W. D. and J. Paegle, (1970), Diurnal variations in boundary layer winds over the south-central United States in summer, *Monthly Weather Review*, 98,10, 735-744.
- Boussinesq, J., 1903: *Theorie analytique de chaleur*, 2, Gauthier-Villars, Paris.

- Bretherton, F. P., (1969), Waves and turbulence in stably stratified fluids. *Radio Science* 4, 12, 1279-1287.
- Bretherton, F. P., (Editor), G. Bull, R. S. Lindzen, Y-H. Pao, E. R. Reiter, and H. Wilhelmsson (1969), The spectral gap. *Radio Science*, 4, 12, 1361-1363.
- Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley (1970), Flux-profile relationships in the atmospheric surface layer. Air Force Cambridge Research Laboratories (unpublished manuscript).
- Chiu, Wan-cheng (1959), The wind and temperature spectra of the upper troposphere and lower stratosphere over North America. SR4, Project 435, Contract AF 19(604)-1755, New York University.
- Committee on Atmospheric Science (1971), *The Atmospheric Sciences and Man's Needs; Priorities for the Future*. National Academy of Sciences, Washington, D.C., 88 pp.
- Dutton, J. A. (1969), An energy budget for a layer of stratospheric CAT. *Radio Science*, 4, 12, 1137-1142.
- Dutton, J. A. and G. H. Fichtl (1969), Approximate equations of motion for gasses and liquids. *J. Atmos. Sci.* 26,2, 241-254.
- Ekman, V. W. (1905), On the influence of the earth's rotation on ocean currents. *Arkiv Matematik*, 2, 11, 1-51.
- Graham, R. J. (1963), Determination and analysis of numerical smoothing weights. NASA Technical Report R-179, 28 pp.
- Griesseier, H., and I. Jacobsen (1970), Versuch der Einbeziehung konvektiver Prozesse in die Theorie der Ekman-Schicht. *Archiv Meteorol. Geophys. Bioklim. Series A*, 19, 3, 283-298.
- Hardy, K. R., K. M. Glover, and H. Ottersten, 1969: Radar investigations of atmospheric structure and CAT in the 3 and 20 km region. In: *Clear Air Turbulence and Its Detection*, Y-H. Pao and A. Goldberg, editors, (Plenum Press, New York) 402-416.
- Landau, L. D., and E. M. Lifshitz, 1959: *Fluid Mechanics, Vol. 6, Course of Theoretical Physics*, English Edition. Pergamon Press.
- Lester, P. F. 1970: Some physical and statistical aspects of clear air turbulence. Colorado State University, *Atmos. Sci. Paper No. 165*, 154 pp.
- Lettau, H. (1967), Small to large-scale features of boundary layer structure over mountain slopes. In: *Proceedings of the Symposium on Mountain Meteorology*, Fort Collins, Colorado, June 26, 1967. Atmospheric Science Paper, No. 122, 1-74, Colorado State University.
- Lilly, D. D. (1971), Turbulence in the troposphere and stratosphere. Paper presented at the FAA Symposium on Turbulence, 22-24 March, 1971, Washington, D.C.
- Lorenz, E. N. (1955), Available potential energy and the maintenance of the general circulation. *Tellus*, 7, 2, 157-167.
- Lumley, J. L., and H. A. Panofsky (1964), *The Structure of Atmospheric Turbulence*, Interscience Publishers, J. Wiley & Sons, New York, 239 pp.
- Mantis, H. (1963), The structure of winds of the upper troposphere at mesoscale. *J. Atmos. Sci.* 20, 2, 94-106.
- Martin, M. A. (1963), Digital filters for data processing. GE Technical Information Series, 625D484.

- Mather, G. K. (1968), Measurements of clear-air turbulence using an instrumented T-33 jet aircraft. ICAS Sixth Congress, ICAS paper No. 68-36.
- Meunch, H.S. (1965), On the dynamics of the winter-time stratospheric circulation. *J. Atmos. Sci.* 22, 4, 349-360.
- Ogura, Y., and N. A. Phillips (1962), Scale analysis of deep and shallow convection in the atmosphere. *J. Atmos. Sci.*, 19, 2, 173-179.
- Oort, A. H. and A. Taylor (1969), On the kinetic energy spectrum near the ground. *Monthly Weather Review*, 97, 9, 623-636.
- Panchev, S. (1968), Coefficient of horizontal macroturbulent exchange in the atmosphere. *J. Atmos. Sci.* 25, 5, 933-935.
- Phillips, O. M., 1967: The generation of clear-air turbulence by the degradation of internal waves. *Proceedings of the International Colloquium*, Moscow, June 15-22, 1965, Publishing House, Nauka, 130-138.
- Reiter, E. R. (1969a), Atmospheric transport processes, Part 1: Energy transfers and transformations. U.S. Atomic Energy Commission, Division of Technical Information, TID-24868, 253 pp.
- Reiter, E. R. (1969b), Mean and eddy motions in the atmosphere. *Monthly Weather Review*, 97, 3, 200-204.
- Reiter, E. R. (1970), Atmospheric transport processes, Part 2: Chemical tracers, U.S. Atomic Energy Commission, Division of Technical Information, TID-25314, 382 pp.
- Reiter, E. R. (1971), A multivariate treatment of kinetic energy. Report C00-1340-25 under Contract AT(11-1)-1340 with USAEC, 18 pp.
- Reiter, E. R. (1972), Atmospheric transport processes, Part 3: Hydrodynamic tracers. U.S. Atomic Energy Commission, Division of Technical Information. (In press).
- Reiter, E. R. and M. E. Glasser (1969), Eddy processes in momentum transport and kinetic energy distribution. Colorado State University. Atmospheric Science Paper No. 149, 35 pp.
- Richardson, L. F. (1920), The supply of energy from and to atmospheric eddies, *Proc. Roy. Soc. London*, A, 97, 356-373.
- Scorer, R. S. (1949), Theory of waves in the lee of mountains. *Quart. J. Roy. Meteor. Soc.* 75, 12, 41-56.
- Scorer, R. S. (1969), Billow mechanics. *Radio Science*, 4, 12, 1299-1308.
- Stankov, B. (1970), Some characteristics of turbulence in the lower 200 feet of the atmosphere. Colorado State University, College of Engineering, Research Report and Thesis. CER70-71BS14, 60 pp.
- Starr, V. P., 1968: *Physics of Negative Viscosity Phenomena*. McGraw-Hill Book Company, New York, 256 pp.
- Steinberg, H. L. and A. C. Wiin-Nielsen (1971), On power laws and non-linear cascades in large scale atmospheric flow. Technical Report 002630-4-T, Dept. of Meteorol. and Oceanography, The University of Michigan, 143 pp.
- Steiner, R. and R. Rhyne (1964), Atmospheric turbulence and airplance response in convective-type clouds, *J. Aircraft*, 1, 1, 13-17.

- Thorpe, S. A. (1969), Experiments on the stability of stratified shear flows. *Radio Science*, *4*, 12, 1327-1331.
- Van der Hoven, J. (1957), Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. *J. Meteorol.* *14*, 2, 160-164.
- Vinnichenko, N. K. (1970), The kinetic energy spectrum in the free atmosphere—1 second to 5 years. *Tellus* *22*, 2, 158-166.
- Vinnichenko, N. K. and J. A. Dutton (1969), Empirical studies of atmospheric structure and spectra in the free atmosphere. *Radio Science*, *4*, 12, 1115-1126.
- Vinnichenko, N. K., N. Z. Pinus, S. M. Shmeter and G. N. Shur (1968), *Turbulence in the Free Atmosphere* (in Russian), Gidrometizdat, Leningrad.
- Weinstein, A. I., E. R. Reiter, and J. R. Scoggins (1966), Mesoscale structure of 11-20km Winds. *J. Applied Meteorol.*, *5*, 1, 49-57.
- Winn-Nielsen, A. (1962), On transformation of kinetic energy between the vertical shear flow and the atmosphere. *Monthly Weather Review*, *90*, 8, 311-323.
- Winn-Nielsen, A., J. A. Brown and M. Drake (1964), Further studies of energy exchange between the zonal flow and the eddies. *Tellus*, *16*, 2, 168-180.
- Wooldridge, G. L. (1970), Vertical momentum transport over mountainous terrain. Colorado State University, Atmospheric Science Paper, No. 164, 91 pp.
- Wooldridge, G., and E. R. Reiter (1970), Large scale atmospheric circulation characteristics as evident from GHOST balloon data. *J. Atmos. Sci.* *27*, 2, 183-194.
- Woods, J. D. (1969), On Richardson's number as a criterion for laminar-turbulent-laminar transition in the ocean and atmosphere. *Radio Science*, *4*, 12, 1289-1298.

List of Symbols

CAS	Committee on Atmospheric Science
$S(k)$	Spectrum function in wave-number space
a	universal constant
ϵ	rate of energy dissipation
K	wave number
ν	coefficient of kinematic viscosity
u_i	velocity component along coordinate x_i
x_i	coordinate
t	time
i	coordinate counting index
j	coordinate counting index
k	coordinate counting index
ρ	density
p	pressure
D_{ji}	$= \partial u_j / \partial x_i + \partial u_i / \partial x_j$ Deformation tensor
$= 1/2 D_{kk}$	$= \partial u_k / \partial x_k$ velocity divergence
g	acceleration of gravity
δ_{ji}	kronecker delta, $=1$ for $j=i$, $=0$ for $j \neq i$
ϵ_{ijk}	Alternating tensor
	+1 for all different indices and even permutations;
	-1 for all different indices and odd permutations;
	0 if at least two indices are equal.
Ω_j	component of earth's rotational velocity along coordinate x_j
\bar{A}	mean value of A
T	time interval of averaging
τ	time increment of data sampling
A'	departure of A from \bar{A}
T'	departure from mean temperature
\bar{T}	mean temperature
CAT	clear air turbulence
$\overline{e_{ij}}$	turbulent kinetic energy
R_{NL}	non-linear interaction term
k_G	wave number characteristic of the spectral gap
K	eddy viscosity coefficient
$a =$	$\sqrt{f/2K}$
\bar{u}_g	geostrophic wind
ℓ	mixing length
$S(f)$	spectrum function in frequency space
c_g	group velocity
A^g	available potential energy
K	kinetic energy

Primes symbolize departure from a mean value.

Chapter 3 ATMOSPHERIC DYNAMICS

William H. Hooke*

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Equations describing the whole of atmospheric dynamics are presented, and various equilibrium and linearized perturbation approximations to these equations, describing certain aspects of the dynamics, are developed in turn. Some of the more common instabilities resulting in energy transfer among motions of different scales are then described. The place of the atmospheric boundary layer in atmospheric dynamics is discussed qualitatively. The advantages of remote sensing techniques in studies of atmospheric dynamics are stressed.

3.0 Introduction

Nature has seen fit for the earth's atmosphere to move in a rather subtle and intricate way, developing a variety of circulation systems and motions on all spatial and temporal scales. On the largest of these scales—the so-called planetary and synoptic scales—this atmosphere develops more or less zonally symmetric, gross circulations such as the Hadley cells, spanning many degrees of latitude and persisting throughout the year, seasonally varying wind systems such as the monsoons, and cyclonic and anticyclonic circulations occurring on spatial scales the order of thousands of kilometers and persisting for many days. On the intermediate or mesoscale, there occur jet streams, squall lines, cumulus convection and other processes tens of kilometers in spatial extent and enduring for several hours. And at the large wavenumber and wavefrequency or microscale end of the spectrum are individual turbulent eddies as small as a millimeter in dimension and persisting without significant deformation only for fractions of a second. The quantitative study of all these motions, their separate characteristics, their interactions with one another, and their ultimate effects, is the subject of atmospheric dynamics.

The present chapter presents a thumbnail sketch of this subject. We begin by writing the conservation equations for atmospheric momentum, energy, and mass, which together with the equation of state model the whole of this dynamics (refer to 3.1). The very generality of these equations greatly limits their practical utility, and we therefore treat a number of the more common approximations and idealizations used in application to specific meteorological problems. In the main, the idealizations that we shall consider fall naturally into one of two classes. The first are the so-called quasi-equilibrium approximations (to be discussed in 3.2), describing large-scale motions driven more or less directly by differential heating of the atmosphere and the resulting pressure gradients, while the second are the so-called perturbation approximations (to be discussed in 3.3), used to describe synoptic, meso-, and micro-scale departures of small amplitude from the quasi-equilibrium states. Discussion of a few other important meteorological approximations, used in both the quasi-equilibrium and perturbation applications (e.g., incompressibility, adiabaticity) is deferred until 3.4. The stability of the "equilibrium" states relative to the perturbations is of great interest from the meteorological standpoint, since instabilities of the larger-scale motions implement energy exchanges between these motions and motions on smaller scales: this topic is treated rather briefly in 3.5. The atmosphere and its motion over the earth's surface also provide to the fluid dynamicist a wide variety of boundary-layer fluid-flow problems unique in their scope and complexity, which we consider qualitatively later in 3.6. The chapter concludes with a short discussion of the importance of remote sensing to future studies of atmospheric dynamics (3.7).

Unfortunately, it is generally true that condensed treatments of broad subjects must omit more material than they can include, and this chapter is no exception. In particular, we give short shrift to the major role of radiative transfer and water vapor phase transformations in energizing the atmosphere's dynamics. The reader interested in these particular topics is referred to the standard meteorological texts such as Haltiner and Martin (1957) and Hess (1959), as well as to more advanced and specialized books by Kondratyev (1969) and Dufour and Defay (1963) and the references therein.

*Lecturer, Department of Astro-Geophysics, University of Colorado; Fellow, Cooperative Institute for Research in the Environmental Sciences (CIRES), a joint institute of the University of Colorado and NOAA.

3.1 The Equations of Motion

We take as our starting point the equations governing motion within the earth's atmosphere, which we write in the vector form

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u} \cdot \nabla) \mathbf{u} + 2\rho \underline{\underline{\Omega}} \times \mathbf{u} = -\nabla p + \rho \mathbf{g} + \mu [\nabla^2 \mathbf{u} + \nabla(\nabla \cdot \mathbf{u})/3] \quad (3:1)$$

$$\frac{\partial e}{\partial t} + (\mathbf{u} \cdot \nabla) e = -p \frac{\partial \rho^{-1}}{\partial t} - p(\mathbf{u} \cdot \nabla) \rho^{-1} + \Phi + \frac{1}{\rho} \nabla \cdot (\mathbf{K}_T \nabla T) + J \quad (3:2)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (3:3)$$

$$p = \rho RT \quad (3:4)$$

in a coordinate system rotating with the earth. Here ρ is the density, \mathbf{u} is the velocity, $\underline{\underline{\Omega}}$ is the earth's angular rotation vector, p is the pressure, \mathbf{g} is the acceleration of gravity, μ is the coefficient of viscosity, e is the internal energy per unit mass, Φ is the rate of dissipation of mechanical energy per unit mass due to shear viscosity, \mathbf{K}_T is the thermal conductivity, J is the rate of heat input per unit mass from the radiative processes and water phase transformations, R is the specific gas constant for the earth's atmosphere, and T is the temperature.

The first three of these equations govern conservation of momentum, internal and other forms of energy, and mass respectively; because they express time derivatives of the system variables in terms of the values of those variables at a given instant, they can be used to predict future states of the system and are therefore referred to as *prognostic*. The last equation, (3:4), is an equation of state; it is *diagnostic*, in that it contains no time derivatives. These equations, and the auxiliary equations relating e , Φ , and J to the fundamental variables p , ρ , T , and \mathbf{u} , are derived and presented in a variety of vector and tensor forms in any of the standard texts on fluid mechanics and meteorology (e.g., Landau and Lifshitz, 1959; Batchelor, 1967; Haltiner and Martin, 1957; Hess, 1959).

Two features of these equations require immediate comment. First, they are nonlinear; they contain terms of higher than first order in the dependent variables. These nonlinearities play a major role in coupling motions on various spatial and temporal scales, and complicate both the actual dynamics and its mathematical description. Second, each of the equations contains a great many terms, each in turn resulting from a different physical process. In (3:1), for example, the $\rho \partial \mathbf{u} / \partial t$ and $\rho(\mathbf{u} \cdot \nabla) \mathbf{u}$ terms represent inertial forces acting on the air parcels of interest. In addition, the air parcels experience, in general, pressure-gradient forces ($-\nabla p$), buoyancy forces ($\rho \mathbf{g}$), and frictional forces ($\mu[\nabla^2 \mathbf{u} + \nabla(\nabla \cdot \mathbf{u})/3]$). It is important to note that many terms in (3:1) and (3:2) depend in different ways upon spatial and temporal derivatives of the velocity, pressure and density fields. This means that the character of any particular type of atmospheric motion is strongly influenced by its spatial and temporal scales. For example, oscillatory motions with periods much shorter than a day are relatively uninfluenced by the earth's rotation, since the inertial term in (3:1) is much larger than the Coriolis term; the parcel motion is then very nearly in the direction of the other forces acting on it. For oscillatory motions of large spatial extent and periods of a day or more, this is no longer true; the Coriolis force now dominates, and parcel motion tends to be at right angles to the impressed forces. This feature of the fluid dynamical equations has led to the practice of nondimensionalizing them, so that dimensionless parameters appear in front of each of the terms, giving some indication of the relative importance of each term depending upon the scaling of the particular problem.

Even in the rather general form given above, the equations of motion have been somewhat simplified. In the momentum conservation equation (3:1), for example, a centripetal force term resulting from the rotation of the coordinate axes has been incorporated in the ρg term, following standard meteorological practice. It is assumed that the coefficient of viscosity is constant, in addition, bulk viscosity is ignored. In (3:4) it is assumed that the atmosphere behaves as a single, ideal gas; in fact, the details of atmospheric composition, possibility of energy exchange among the different gaseous species, etc., are consistently ignored throughout this system of equations.

Nevertheless, despite these idealizations, this system, when used in conjunction with the appropriate equations of radiative transfer and the equations describing the energy changes associated with changes in state of atmospheric water (these are needed to specify J) and the necessary boundary and initial conditions, is competent to describe atmospheric motions on all the spatial and temporal scales of interest to us.

Indeed, that is our problem, since if we wish to describe the dynamics of even a single cumulonimbus, for example, we find that solution of the equations in their fullest generality requires more initial information than we can provide, and the equations, even if solved, would provide us more information than we could assimilate (e.g., the location, dimensions, lifetimes, and intensity of every turbulent eddy within the cloud as well as the macroscopic behavior of the cloud as a whole). The history of atmospheric dynamics has thus been a search for methods of further simplifying these equations and the boundary and initial conditions (e.g., by parameterizing the micrometeorological processes) in ways drastically reducing the amount of input required and output obtained while minimizing the inevitable distortion of the essential character of the problems at hand.

As implied above, the useful simplifications of (3:1) through (3:4) differ from application to application. For example, in surface-layer studies, the rotation of the earth is often ignored, while viscous forces, local surface irregularities, etc., play an essential role; in synoptic meteorology, on the other hand, the earth's rotation is fundamental, with viscous effects on occasion almost entirely negligible. It is instructive to quickly examine in sequence some of the approximations more commonly used.

Broadly speaking, we find that the approximations fall into two classes. The first are those approximations which describe quasi-steady or "equilibrium" motions of one sort or another. The desired simplification is achieved by ignoring the partial derivative $\partial/\partial t$ or the prognostic character of the equations. These approximations are considered next in 3.2. The second are those approximations that assume the motions of interest represent perturbations from a known equilibrium state. Here a simplification is achieved by writing the dependent variables, u_i say, in the form

$$u_i = u_{i0} + u'_i, \tag{3:5}$$

where the zero subscript denotes the equilibrium, unperturbed, or average value of u_i , and the prime superscript denotes the departure from the equilibrium value. The departure is then either assumed small, so that terms of second and higher order in the primed quantities are neglected and the nonlinearities in (3:1) through (3:3) are eliminated, or use is made of the fact that the average of $u'_i = 0$. These approximations may apply to synoptic, meso-, and micro-scale perturbations, and are considered in 3.3.

3.2 Quasi-Equilibrium Approximations to the Equations of Motion

3.2.1 The Hydrostatic Approximation

The first approximation we shall consider is one of the best (and most universal) in all of meteorology; it often applies to the perturbation motions as well as those of the equilibrium type. Consider the vertical component of the vector equation (3:1). There are very few cases in which the vertical acceleration of even small volumes of the atmosphere is at all comparable to the acceleration of gravity, which is $\sim 10 \text{ m/s}^2$ (convection is an important exception). In addition, even if we assume a wind speed u of the order of 100 m/s, an extremely large value, we find that the Coriolis acceleration $2\tilde{\Omega} \times u$ is at most $\sim 1.5 \times 10^{-2} \text{ m/s}^2$, so that it is also negligible. Similar order-of-magnitude estimates show that the viscous term

of this equation is small compared to the gravitational term. Thus the large gravitational term of (3:1) can only be balanced by a correspondingly large vertical pressure gradient of average pressure, i.e., we must have the so-called hydrostatic equation

$$\frac{\partial p}{\partial z} = -\rho g . \quad (3:6)$$

When combined with (3:4) and integrated, the hydrostatic equation yields

$$p(h) = p_0 \exp(-\int_0^h dz/H); \quad \rho(h) = \rho_0 \left(\frac{T_0}{T}\right) \exp(-\int_0^h dz/H) \quad (3:7)$$

where $H=RT/g$ is the atmospheric scale height (~ 8 km in the earth's troposphere) and the zero subscripts refer to surface values; pressure and density therefore decrease roughly exponentially with increasing height (see Chapter 1). It is essential to note that (3:6) does not imply that vertical motions do not occur; rather it implies that vertical motions are dictated by the small departures from this approximation.

3.2.2 The Geostrophic Wind and the Thermal Wind Equation

Turning now to the horizontal components of (3:1), we very often find situations in synoptic or global meteorology in which the inertial terms and viscous terms of these equations are small compared to the pressure gradient terms (produced, for example, by a difference in temperature between two air masses). In this event, the pressure gradient will accelerate the air, and the Coriolis force (which acts at right angles to the velocity) will increasingly deflect it, until such time as the two forces are in equilibrium and the eastward or x-directed velocity component u and the northward or y-directed velocity component v are given by

$$fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} , \quad (3:8)$$

$$fv = \frac{1}{\rho} \frac{\partial p}{\partial x} , \quad (3:9)$$

where $f \equiv 2\Omega \sin \phi$ is the so-called Coriolis parameter, and ϕ is the latitude. It is an essential (and unique) characteristic of this type of motion that when equilibrium is achieved, the steady-state wind is *parallel* to the isobars. At temperate and high latitudes, the wind motion is often geostrophic or nearly so; at low latitudes, where f is small, this type of balance is not seen.

Equations (3:8) and (3:9) can be combined in various ways with the hydrostatic equation (3:6) to yield equations expressing wind variations with height in terms of horizontal temperature gradients; one finds, for example, that the zonal wind u varies with height z according to the so-called thermal wind equation

$$\frac{\partial}{\partial z} \left(\frac{u}{T} \right) = - \frac{g}{f T^2} \frac{\partial T}{\partial y} \quad (3:10)$$

(often the variation of T with z is ignored in comparison with the variation of u). This equation is often used in upper atmospheric work to derive horizontal temperature gradients from vertical wind profiles (e.g., Murgatroyd, 1965).

3.2.3 Gradient Flow

A geostrophic flow must, by definition, have zero curvature; when the actual flow is sufficiently curved centripetal accelerations become important, and (3:8) and (3:9) must be replaced by more general forms which now include an inertial term:

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = fv - \frac{1}{\rho} \frac{\partial p}{\partial x} \tag{3:11}$$

$$u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} = -fu - \frac{1}{\rho} \frac{\partial p}{\partial y} \tag{3:12}$$

These are the equations describing the so-called gradient wind. In practice, it generally turns out to be difficult to detect the difference between the gradient and geostrophic winds on weather maps except when the wind speeds are quite high.

3.2.4 Cyclostrophic Flow

In mesoscale and microscale meteorological systems having very high rotational velocities and nearly circular symmetry, such as hurricane centers, tornadoes, water spouts, and dust devils, the centripetal accelerations may become so large that Coriolis forces are negligible, so that the rotational velocity, v_θ , say, is then given to good approximation by the equation

$$\frac{v_\theta^2}{r} = \frac{1}{\rho} \frac{\partial p}{\partial r} \tag{3:13}$$

where r is the radius from the center of the disturbance.

3.2.5 The Effect of Friction on the Geostrophic Wind

Near the earth's surface, frictional forces become large, and the winds can no longer remain geostrophic. For many purposes, the frictional forces can be assumed proportional to the wind speed, with constant of proportionality k , say; then the equations of motion become

$$0 = fv - ku - \frac{1}{\rho} \frac{\partial p}{\partial x} \tag{3:14}$$

$$0 = -fu - kv - \frac{1}{\rho} \frac{\partial p}{\partial y} \tag{3:15}$$

and we find

$$u = - \frac{k}{k^2 + f^2} \frac{1}{\rho} \frac{\partial p}{\partial x} - \frac{f}{k^2 + f^2} \frac{1}{\rho} \frac{\partial p}{\partial y} \tag{3:16}$$

$$v = + \frac{f}{k^2 + f^2} \frac{1}{\rho} \frac{\partial p}{\partial x} - \frac{k}{k^2 + f^2} \frac{1}{\rho} \frac{\partial p}{\partial y} \tag{3:17}$$

replacing (3:8) and (3:9). Thus the component of motion parallel to the isobars is reduced from its geostrophic value ($k = 0$), and a steady component perpendicular to the isobars from high to low pressure is introduced.

To obtain the *height profile* of the wind in this lowest kilometer or so of the air, the frictional forces must be assumed proportional not to the wind speed but to its second derivative with height. Equations (3:14) and (3:15) are then replaced by

$$0 = -\frac{\partial p}{\partial x} + \rho f v + \rho K \frac{\partial^2 u}{\partial z^2} \quad (3:18)$$

$$0 = -\frac{\partial p}{\partial y} - \rho f u + \rho K \frac{\partial^2 v}{\partial z^2} \quad (3:19)$$

where ρK is a so-called eddy exchange coefficient, usually assumed independent of height; integration of these equations with respect to height then yields the wind profile known as the Ekman spiral (see Chapter 6). It might be noted in passing that the terms $\rho K \partial^2 u / \partial z^2$ and $\rho K \partial^2 v / \partial z^2$ have the same form as the contribution one would obtain from the viscous term of (3:1) by assuming an incompressible flow ($\nabla \cdot \mathbf{u} = 0$) and variations in the vertical that are much more rapid than in the horizontal. However, the eddy exchange coefficient K arises from the macroscopic velocity fluctuations associated with turbulence, and is many orders of magnitude greater than the coefficient of molecular viscosity $\nu \equiv \mu / \rho_0$. In actual fact the Ekman spiral is not often seen in nature, in part because the coefficient K is rarely constant with height (see, for example, Lettau and Dabberdt, 1970), and in part because wind profiles of the Ekman type appear to be unstable (Brown, 1970).

3.2.6 Inertial Oscillations

In closing this section, we consider motions in which pressure and frictional forces are negligible, yielding an approximate equilibrium between inertial and Coriolis forces:

$$\frac{du}{dt} = f v \quad (3:20)$$

$$\frac{dv}{dt} = -f u \quad (3:21)$$

These equations integrate immediately to

$$u = f y + c_1 \quad (3:22)$$

and

$$v = -f x + c_2 \quad (3:23)$$

where c_1 and c_2 are constant of integration. Because the Coriolis force acts at right angles to the particle velocity and can do no work on it, $u^2 + v^2 \equiv c^2$, a constant, and this together with (3:22) and (3:23) yields

$$\left(x - \frac{c_2}{f}\right)^2 + \left(y + \frac{c_1}{f}\right)^2 = \frac{c^2}{f^2}, \quad (3:24)$$

which is the equation of a circle of radius c/f . In such circumstances, the air parcels execute so-called inertial oscillations, with period of $\pi/\Omega \sin \phi$ or half a pendulum day (the time required for a Foucault pendulum to execute a complete revolution). Inertial oscillations are rarely observed in the earth's atmosphere, mainly because the assumptions which lead to their derivation (e.g., neglect of viscous effects) are rarely satisfied and because oscillations of this type radiate their energy away from the site; nevertheless they may account for phenomena such as the presence at times of a low-level nocturnal jet over the midwestern United States (Blackadar, 1957).

3.3 Small-Amplitude Perturbation Approximations to the Equations of Motion

3.3.1 Rossby Waves

The first perturbation motion we shall consider owes its existence to the variation of the Coriolis parameter with latitude, and is of great importance in synoptic meteorology. If we take the curl of (3:1) and assume a purely horizontal, non-divergent, frictionless flow and an auto-barotropic atmosphere (ρ a time-independent function of p only), then we find, as shown in meteorological texts, that

$$\frac{\partial \zeta}{\partial t} + u \frac{\partial \zeta}{\partial x} + v \frac{\partial \zeta}{\partial y} + v \frac{\partial f}{\partial y} = 0 \quad (3:25)$$

where $\zeta \equiv \partial v / \partial x - \partial u / \partial y$. If we further assume that u and v are functions of x , the eastward-directed coordinate axis, only, we then have

$$\frac{\partial^2 v}{\partial x \partial t} + u \frac{\partial^2 v}{\partial x^2} + \beta v = 0 \quad (3:26)$$

$$\text{where } \beta \equiv \frac{\partial f}{\partial y} = 2\Omega \cos \phi / R_e \quad (R_e \text{ the radius of the earth})$$

Specializing to wave solutions propagating west-east with a constant speed c , assuming a constant unperturbed eastward flow U_0 , and linearizing, we finally obtain

$$\frac{\partial^2 v}{\partial x^2} + \frac{\beta}{U_0 - c} v = 0 \quad (3:27)$$

$$\lambda^2 = \frac{4\pi^2 (U_0 - c)}{\beta} \quad (3:28)$$

The original analysis is due to Rossby (1939); it has proved extremely successful in the study of long waves ($\lambda \sim 3000$ km) in the temperate-latitude westerlies.

3.3.2 The Atmospheric Tides

Like the ocean, the earth's atmosphere exhibits tidal oscillations, which are discussed in more detail in Chapter 5 (Sections 5.9 through 5.11). The earth's rotation and sphericity are essential to their description. These motions play a major role in the dynamics of the earth's upper atmosphere, but in the troposphere, they are barely detectable, being revealed in temperate-latitude pressure data only after averaging over many days. Solar and lunar gravitational forcing plays some role in their generation, but the predominant forcing is thermal. Lindzen and Chapman (1970) give a rather complete discussion of these motions.

3.3.3 Atmospheric Gravity Waves and Acoustic Waves

Atmospheric gravity waves, the subject of Chapters 5 and 7 and of incidental mention in several others, are not discussed extensively here. Suffice it to say their greatest importance may lie in meso- and micro-scale dynamics; their wavelengths may be as small as a few hundred meters and the wave periods may be as short as a few minutes. The wave phase velocities range from just below the so-called "speed of sound" C , ~ 340 m/s in the troposphere, to arbitrarily small values. For the large-scale waves, the hydrostatic approximation is still valid, but for waves of the shortest periods this equation must be replaced by the more general form

$$\rho \frac{\partial w}{\partial t} = -\nabla p + \rho \underline{g} \quad (3:29)$$

where w is the vertical velocity (see Chapter 5). The earth's rotation and sphericity play a negligible role in their description.

Acoustic waves have periods even shorter than the atmospheric gravity waves, and phase velocities equal to C ; atmospheric compressibility plays a major role in these motions (see Section 3.4). These properties of acoustic waves prohibit their effective coupling with other atmospheric motions, and they do not appear to play a major dynamical role. They are generated, however, in association with certain atmospheric events (such as severe weather, for example), and their ability to propagate undistorted for long distances from their sources makes them interesting as remote indicators of these events (Chapter 21).

3.3.4 Turbulence

The theory of atmospheric turbulence is considered in detail in Chapter 4 of this volume, in books (e.g., Lumley and Panofsky, 1964), as well as in countless journal articles. It is mentioned here merely to emphasize that the standard method of treatment is to consider the turbulent velocity, temperature, density, and pressure fluctuations to be perturbations on the mean background flow. Here, however, because of the nonlinear character of turbulence, the mathematical description must retain terms of higher than first order in the perturbation quantities. To render the problem tractable, additional simplifications are introduced, for example, considerable use of averaging in time and space, assumptions of homogeneity and stationarity, and a variety of closure assumptions. Thus, for example, we may define a time average of the wind over a time interval T centered at t as

$$\bar{\underline{u}} \equiv \frac{1}{T} \int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T} \underline{u} dt \quad , \quad (3:30)$$

expressing the instantaneous wind as

$$\underline{u} \equiv \bar{\underline{u}} + \underline{u}' \quad , \quad \text{so that } \bar{\underline{u}'} \equiv 0 \quad . \quad (3:31)$$

Writing expansions analogous to (3:31) for the other dependent variables, we may then obtain new forms of the equations of motion. We find, for example, assuming that the motion is incompressible (see Section 3.4), that the equation of motion for the *mean* velocity $\bar{\underline{u}}$ is

$$\bar{\rho} \frac{\partial \bar{\underline{u}}}{\partial t} + \bar{\rho} (\bar{\underline{u}} \cdot \bar{\nabla}) \bar{\underline{u}} + 2\bar{\rho} \bar{\Omega} \times \bar{\underline{u}} = -\bar{\nabla} \bar{p} + \bar{\rho} \bar{\underline{g}} + \mu \nabla^2 \bar{\underline{u}} + \frac{\partial}{\partial x} (-\overline{\rho u' u'}) + \frac{\partial}{\partial y} (-\overline{\rho v' u'}) + \frac{\partial}{\partial z} (-\overline{\rho w' u'}) \quad , \quad (3:32)$$

where $\underline{u}' = (u', v', w')$. This expression differs from (3:1) by the addition of the second-order perturbation terms $\partial/\partial x(-\overline{\rho u' u'})$, $\partial/\partial y(-\overline{\rho v' u'})$ and $\partial/\partial z(-\overline{\rho w' u'})$ which are the so-called Reynolds or eddy stresses and represent forces experienced by the mean flow as the result of the turbulence present. The effects of these terms are often approximated by using an effective eddy viscosity as done in Section 3.2.5. Similar second-order terms appear in the mean energy conservation equation (Chapters 1 and 5).

3.4 Other Approximations to the Equations of Motion

Thus far in our discussion we have confined our attention to approximations to the equation of momentum conservation (3:1). There are, however, important approximations that can also be made to the equations of energy and mass conservation, and we now consider these.

The most important approximation which is usually made to (3:2), the equation of energy conservation, is that the motion under consideration is adiabatic, i.e., that parcel motion occurs without the introduction or removal of heat from the parcel. In this case (3:2) simplifies to

$$\frac{\partial p}{\partial t} + (\bar{\underline{u}} \cdot \bar{\nabla}) p = \frac{\gamma p}{\rho} \left[\frac{\partial \rho}{\partial t} + (\bar{\underline{u}} \cdot \bar{\nabla}) \rho \right] \quad , \quad (3:33)$$

(compare 5:3), or equivalently, as shown in meteorological texts, to

$$\theta \equiv T \left(\frac{1000}{p} \right)^{R/c_p} = \text{const} \quad (3:34)$$

where c_p is the specific heat at constant pressure for dry air, p is the pressure in mb, and θ is the so-called potential temperature. This approximation is often valid in description of the perturbation motions, since these are of such short duration that heat exchange through radiative transfer or by conduction across the parcel boundaries of interest is negligible. It is very often used in atmospheric gravity wave studies, for example (see Chapter 5).

In the mass conservation equation (3:3), it can often be assumed without appreciable error that the motion is incompressible, i.e., that

$$\nabla \cdot \underline{\underline{u}} = 0 \quad (3:35)$$

This approximation is standard in treatment of all the equilibrium motions, and it applies to certain of the perturbation motions to good approximation as well. In fact, it is generally regarded as sufficiently accurate to permit its use in reduction of certain remote sensing data. For example in the two-Doppler radar technique for measuring convective motions, it is used to supply the third velocity component (see Chapter 13, 13:53). It does not, however, apply to all perturbation motions; in particular, it applies to gravity waves (see 5:29, for example) only in certain limits, such as the so-called Boussinesq approximation, in which the atmosphere is assumed incompressible except for the buoyancy term in the equation of motion.

In closing this last section on approximations to the equations of motion, we might note that nowhere have we discussed the approximations and exact forms of the equations used in numerical modeling of the earth's atmospheric circulation. The need in such modeling to minimize the importance of small numerical errors, and to reduce computation time, while simultaneously including all physical processes of importance, has led to a search for alternative forms of the basic equations couched in terms of different variables. For a discussion of these points, as well as a more general treatment of the subjects of atmospheric circulation, the reader is referred to the texts by Thompson (1959), Lorenz (1967), and Palmén and Newton (1969) and the references therein.

3.5 Atmospheric Stability

Of great importance in meteorology is the subject of the stability of the atmospheric state of motion relative to the small perturbations in that state. If the atmospheric state under consideration is stable, then this state will continue to prevail until the external forcing agents driving the motion themselves change. If the atmospheric state under consideration is unstable with respect to small perturbations of a given type, then the naturally occurring infinitesimal fluctuations of this type will grow exponentially, at least initially, until nonlinear or dissipative effects take over, and so the state of the system will be changed. In the new state of motion resulting from the original unstable state, some of the energy and momentum of the original flow are distributed among other motions having different spatial and temporal scales; thus, instability provides an important mechanism for energy exchange between motions. If the new state is itself unstable with respect to some other perturbations, then that state too will change, and so forth; if the succeeding states continue to be unstable, we approach a turbulent condition.

A number of methods are available to the theorist for treating atmospheric stability problems. These include: the parcel method, which estimates the forces acting on a parcel displaced from its initial or equilibrium position in an attempt to determine whether the parcel accelerates or decelerates relative to this position; normal mode analysis, in which the frequencies are allowed to be complex; treatment of the initial value problem using Laplace transforms; and numerical solution of the initial value problem. In what follows, where we make any attempt to describe the instabilities quantitatively we use the parcel approach, not because it is the best or the most general, but simply because, where applicable, it is the most intuitive.

3.5.1 Static Stability and Instability

Consider the atmosphere to be at rest, and consider further a parcel of air at rest in this atmosphere. We now subject this parcel to a slow vertical displacement, say upwards for the sake of definiteness, during which the parcel expands as a result of the ambient atmospheric stratification (see Section 3.2.1), but in such a way that the parcel pressure and the pressure external to it remain equal. In its new position, the parcel is no longer in equilibrium; it accelerates, with its acceleration given by

$$\frac{dw}{dt} = -g - \frac{1}{\rho'} \frac{\partial p}{\partial z} \quad (3:36)$$

where ρ' is the parcel density. It has been assumed here that $p' = p$, where p' is the parcel pressure, and p is the pressure of the environment. Here, and in what follows, the primed quantities refer to the environment. From (3:6) and (3:4) it follows that

$$\frac{dw}{dt} = g \frac{(T' - T)}{T} \quad (3:37)$$

If we assume that the parcel motion is adiabatic, its temperature change with height will be $-g/c_p$ and if we define $dw/dt = d^2z/dt^2$, where z is the parcel displacement, then we can write

$$\frac{d^2z}{dt^2} = g \left(-\frac{1}{T} \frac{\partial T}{\partial z} - \frac{g}{c_p T} \right) z = -g \frac{d(\ln\theta)}{dz} z, \quad (3:38)$$

where the last equality follows from (3:34) and (3:4). When $d(\ln\theta)/dz > 0$, i.e., when the temperature lapse rate (defined to be the negative of the temperature gradient) is less than the dry adiabatic lapse rate g/c_p ($\approx 1^\circ\text{C}/100\text{m}$), the atmosphere is statically stable. Equation (3:38) then describes an oscillatory parcel motion, with oscillation frequency equal to the Brunt-Väisälä frequency (see Chapters 5 and 7). When $d(\ln\theta)/dz = 0$, the atmosphere is convectively unstable; parcel acceleration is exponential. When $d(\ln\theta)/dz < 0$, we have a state of neutral stability. All three states are observed in the earth's atmosphere. For example, in the atmospheric boundary layer at night, the ground radiates its heat into space, the lowermost layers of the earth's atmosphere rapidly cool by conduction to the earth, and $d(\ln\theta)/dz > 0$. During the day, solar insolation rapidly heats the earth's surface, the air near the surface also heats up rapidly, and $d(\ln\theta)/dz < 0$. At intermediate times, a state of neutral stability is achieved (for more discussion, see Chapter 6). We see therefore, that the adiabatic lapse rate is an important dividing line in meteorology. Where the temperature lapse rate exceeds the adiabatic value, we have convection (there are modifications to this picture when we consider a moist atmosphere in which changes of phase occur, but space does not permit their consideration here; see any meteorological text). Where the temperature lapse rate falls below this value, convection is inhibited. Convection, of course, is an important type of atmospheric motion, since it may dominate in the daytime boundary layer and since it is responsible for cumulus cloud formation.

We note parenthetically that for extreme temperature gradients, i.e., for $\partial T/\partial z = -g/R \approx 3.4^\circ\text{C}/100\text{m}$ or less, the atmospheric density is actually constant or increasing with height. Such a condition obviously leads to overturning and instability, but these conditions are almost never achieved, except possibly in layers immediately adjacent to the ground.

3.5.2 Dynamic Stability and Instability

Study of static stability is not very satisfying, since if the atmosphere is statically unstable, it will be set into motion, and we must then determine whether it is dynamically unstable. It turns out, for example, that viscosity and thermal conduction inhibit the convective instability described in the previous subsection. This more general dynamical problem has been treated for the case of a fluid confined between two horizontal plates separated by a distance h (the problem of "Benard convection") and it is found that free convection occurs when (e.g., Lin, 1955)

$$\text{Ra} \equiv -\frac{\Gamma g \alpha h^4}{\kappa \nu} > 1708, \quad (3:39)$$

where Ra is the so-called Rayleigh number, Γ is the temperature gradient, α is the coefficient of expansion, κ is the thermal diffusivity, and ν is the viscosity.

Another important type of dynamical instability arises in shear flows in stably stratified fluids. Correct application of the parcel argument in this case appears to be a bit subtle, but an appropriate formulation appears to have been given recently by Hines (1971). His analysis in its most general form applies to tilted shear flows, which he finds to be inherently unstable (and this finding should certainly inspire more research on this source of atmospheric instability) but we shall confine our attention here to the more limited case of a vertical shear in a horizontal flow. We thus consider the initial interchange of two parcels of air in a shear flow separated by a distance δ along the z-axis. The first parcel is taken to have unit volume, and the second volume $(p_1/p_2)^{1/\gamma}$ where γ is the ratio of specific heats and p_1 and p_2 are the respective pressures at the first and second locations. These two parcels can then be interchanged without disturbing the surrounding medium. Hines finds that the net increase in gravitational potential energy resulting from the interchange is

$$G = \rho g [d(1n\theta)/dz] \delta^2 \cdot \quad (3:40)$$

In a statically stable atmosphere, G is positive, so that the virtual interchange will be energetically possible only if an amount of kinetic energy greater than G is available for conversion. Hines calculates the maximum amount of kinetic energy available from the interchange to be

$$K_E = \rho \left[\frac{\partial u}{\partial z} \right]^2 \delta^2 / 4 \quad (3:41)$$

which together with (3:40) implies that the atmosphere will be dynamically stable unless

$$\text{Ri} \equiv g \frac{d(1n\theta)/dz}{(du/dz)^2} \leq 1/4, \quad (3:42)$$

where Ri is the so-called Richardson's number of the flow. More general analyses (Miles, 1961; Howard, 1961; Chimonas, 1970) show that condition (3:42) is indeed a necessary condition for instability to occur. Note that Ri is a ratio between the stabilizing influence of the thermal stratification and the destabilizing influence of wind shear; when $\text{Ri} < 0$, the atmosphere is *statically* unstable (see the previous subsection). The Kelvin-Helmholtz waves generated by such dynamically unstable flows have been studied observationally and theoretically for a great many years (e.g., Haurwitz, 1941; Reiter, 1963), and they are mentioned in other chapters throughout this volume.

So far we have treated the stability of the atmosphere with respect to vertical displacements. We now wish to consider the stability of motions relative to displacements in the horizontal as well as the vertical. One such instability of particular importance in synoptic meteorology is the baroclinic instability. Space does not permit a full description of it here, but a very clear treatment may be found in Pettersen (1956). The instability arises because the state of the temperate-latitude troposphere is markedly baroclinic, that is to say, the surfaces of constant pressure and constant density are not parallel. The potential energy of such a region decreases if the colder air sinks and the warmer air rises. The internal energy decreases as well, and thus energy is available for the growth of disturbances. Parcel arguments show that this energy is not readily available for Rossby waves of very short or very long wavelengths, but that it is available to perturbations in an intermediate band of wavelengths (~ 2400 km, for vertical shears $7 \times 10^{-4} \text{ s}^{-1}$, mean lapse rate $\sim 6.5^\circ\text{C}/\text{km}$ and latitude $\sim 45^\circ$). Comparison of the wavelengths observed at temperate latitudes with those predicted support the notion that baroclinic instability is responsible for the growth of disturbances in the temperate-latitude westerlies.

Another instability of interest is the so-called barotropic instability, which arises from certain vorticity distributions in, for example, the zonal westerly wind patterns seen at temperate latitudes. In particular, if the vorticity vanishes at a point of maximum shear in the background flow, waves moving with velocities less than the background wind at this point but greater than the minimum westerly wind velocity may be amplified (see Kuo, 1949).

In closing this section, we might mention the inertial instabilities which arise in a rotating fluid mass when the velocity distribution is such that kinetic energy of the rotation is available to kinetic energy of the perturbations. Parcel arguments show that this occurs if the absolute angular momentum of the rotation decreases outward from the rotational axis, i.e.,

$$r \frac{\partial \omega_a}{\partial r} + 2\omega_a < 0 \quad , \quad (3:43)$$

where ω_a is the absolute angular rotational velocity and r is the radial distance from the rotational axis. In applying this criterion to motions in the earth's atmosphere, we usually find that synoptic motions at high and temperate latitudes are stable in this respect, but at low latitudes, systems having a high rotational velocity, such as hurricanes, may not be.

3.6 The Boundary Layer

No discussion of atmospheric dynamics would be complete without some consideration of the role played by the atmospheric boundary layer. As its name implies, this is the region of the earth's atmosphere in direct contact with the surface of the earth and the oceans. It is such an interesting region, both by virtue of the diversity of the physical processes internal to it and by its immediacy to man and his activities that it receives considerable attention throughout this volume; Chapter 6 describes it in detail and Chapter 7 is even devoted solely to atmospheric gravity wave propagation in it. For this reason, we shall not attempt to give a complete description of its structure and dynamics here, but confine ourselves to a few general remarks. We begin by commenting on the fundamental differences between the atmospheric boundary layer and boundary layers normally encountered by the fluid dynamicist. First, the boundary, the boundary layer, and the fluid flow above it are *geophysical* in size and character; this increase in the scale of the problem results in stratification of the fluid by virtue of the earth's gravitational field, and in unique motions of the fluid around mesoscale obstacles such as mountains (and cities), again influenced by the earth's gravity and rotation (see, e.g., the papers contained in Reiter and Rasmussen, 1967). Second, the boundary is an active boundary, not a passive one. In most fluid dynamical problems, it suffices merely to consider the viscous drag exerted by the boundaries on the flow of interest; the flow itself is usually considered to have some external origin (this is accomplished in airfoil problems merely by changing reference frames, for example). In the atmospheric case, however, the earth's boundary acts as a potent driving force for the fluid flow itself. In part it accomplishes this by converting shortwave solar radiation, to which the atmosphere is rather transparent, to long-wave radiation to which the atmosphere is rather opaque; in this way it becomes the immediate source of the radiative energy which the atmosphere absorbs. In addition, the earth's boundary is largely liquid; in consequence it supplies a great deal of energy to the atmosphere in the form of water vapor, with its associated latent heat of condensation available for driving convective motions, hurricanes, etc. Finally, the liquid nature of the boundary also implies a more complicated air-boundary interaction, since over the water the atmosphere in motion drives wind waves as well as ocean currents, with important implications for air-earth momentum and energy exchange.

Because the atmospheric boundary layer is unique in so many respects, the great body of boundary-layer knowledge gained from the aeronautical and laboratory applications cannot be merely carried over to serve in this new application. The underlying philosophy of most approaches to boundary-layer problems, however, can be so applied. Specifically, this consists in the use of similarity theory—isolating those dimensionless quantities or ratios needed to completely specify the boundary layer structure apart from some scaling factors.

Though but one region of the atmosphere, the boundary layer exhibits a considerable vertical structure. The lowermost layer lies essentially within or just above the roughness elements composing the

surface. Here the turbulent momentum flux of the higher levels is converted to viscous stress (resulting from *molecular* viscosity) over aerodynamically smooth surfaces and to pressure forces over aerodynamically rough terrain. Turbulent fluxes of other quantities such as heat and water vapor give way to molecular diffusion. Only the inertia and molecular viscous terms are retained in the equations of motion for this layer. Above this interfacial layer lies the constant-flux layer, in which fluxes of momentum, heat, and water vapor, while turbulent, do not vary appreciably with height. This layer typically extends a few tens of meters above the earth's surface. Above the constant flux layer lies the so-called planetary boundary layer, which is the layer of transition between the constant flux layer and the free atmosphere (see Section 3.2.5). The height of this layer is somewhat variable (and indeed, somewhat arbitrary), but for practical purposes it may be assumed to be the order of 1 km. In addition to this rather gross structure, the boundary layer exhibits an infinite variety of intricate substructure and microdynamics such as wave motions, turbulence, and secondary flows of other kinds which are rather transitory. While there is considerable interest in the microscale dynamics per se, there is also widespread interest in learning how to parametrize its dynamical effects in ways suitable for use in general circulation models.

3.7 Remote Sensing in Studies of Atmospheric Dynamics

We close this chapter with a few comments on the role of remote sensing in future studies of atmospheric dynamics. To begin with, the major advantage of remote sensing over *in situ* measurements from the standpoint of the synoptic meteorologist is the possibility of obtaining data from inaccessible atmospheric regions (e.g., at great altitudes, over the oceans, and the earth's poles). Both present and contemplated numerical models of this dynamics require initial input data in the form of height profiles of wind, temperature, and humidity at points the order of 100 km apart over the earth's entire surface; remote sensing appears to provide the only means by which data acquisition on this scale is economically feasible. Secondly, remote sensors typically acquire *continuous* spatial and temporal data on the parameters they measure. Such data (witness the acoustic-sounder and FM-CW radar data shown elsewhere in this volume, for instance) not only provide the extensive coverage required for operational purposes, such as thunderstorm and tornado tracking, but also provide the theorist with unique insights and inspiration in his task. To achieve the same coverage using *in situ* instruments would require extensive and intricate three-dimensional networks of these instruments as well as sophisticated data processing systems. In fact, remote sensing devices have already shown that to achieve the necessary coverage in some micrometeorological applications, the *in situ* instruments would have to be so densely packed as to destroy the very boundary-layer flow they were attempting to study. By contrast, remote sensing instruments do not significantly perturb the medium.

Finally, the remote sensing instruments readily provide volume and temporal averages often more meaningful than the point values obtained by conventional sensors.

3.8 References

- Batchelor, G. K., 1967: *An Introduction to Fluid Mechanics*, Cambridge University Press.
- Blackadar, A. K. (1957) Boundary layer wind maxima and their significance for the growth of nocturnal inversions, *Bull. Am. Meteorol. Soc.* 38, 283-290.
- Brown, R. A. (1970) A secondary flow model for the planetary boundary layer, *J. Atmos. Sci.* 27, 742-757.
- Chimonas, G. (1970) The extension of the Miles-Howard theorem to compressible fluids, *J. Fluid Mech.* 43, 833-836.
- Dufour, L., and R. Defay, 1963: *Thermodynamics of Clouds*, Academic Press, New York and London.
- Haltiner, G. J., and F. L. Martin, 1957: *Dynamical and Physical Meteorology*, McGraw-Hill, New York.
- Haurwitz, B., 1941: *Dynamic Meteorology*, McGraw-Hill, New York.
- Hess, S. L., 1959: *Introduction to Theoretical Meteorology*, Holt, Rinehart, and Winston, New York.

- Hines, C. O., 1971: Generalizations of the Richardson criterion for the onset of atmospheric turbulence, in Quart. J. Roy. Meteorol. Soc. 97, 429-439.
- Howard, L. N. (1961): Note on a paper of John Miles, J. Fluid Mech. 10, 509-512.
- Kondratyev, K. Ya. (1969): *Radiation in the Atmosphere*, Academic Press, New York and London.
- Kuo, H-I (1949): Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere, J. Meteorol. 6, 105-122.
- Landau, L. D., and E. M. Lifshitz, 1959: *Fluid Mechanics*, Pergamon Press, New York.
- Letau, H. H., and W. F. Dabberdt (1970): Variangular wind spirals, Boundary Layer Meteorol. 1, 64-79.
- Lin, C. C., 1955: *The Theory of Hydrodynamic Stability*, Cambridge University Press, Cambridge, England.
- Lindzen, R. S., and S. Chapman (1969), Atmospheric tides, Space Sci. Rev. 10 3-188.
- Lorenz, E. N. (1967): *The Nature and Theory of the General Circulation of the Atmosphere*, World Meteorological Organization.
- Lumley, J. L., and H. A. Panofsky, 1964: *The Structure of Atmospheric Turbulence*, John Wiley and Sons, New York.
- Miles, J. W. (1969): On the stability of heterogeneous shear flows, J. Fluid Mech. 10, 496-508.
- Murgatroyd, R. J. (1965): Winds in the mesosphere and lower thermosphere, Proc. Roy. Soc. A 288, 575-589.
- Palmén, E., and C. W. Newton, 1969: *Atmospheric Circulation Systems*, Academic Press, New York.
- Pettersen, S., 1956: *Weather Analysis and Forecasting*, Vol. I, McGraw-Hill, New York.
- Reiter, E. R., 1963: *Jet Stream Meteorology*, University of Chicago Press, Chicago.
- Reiter, E. R., and J. L. Rasmussen (1968), Ed., Proc. Symposium on Mountain Meteorology, 26 June 1967, Atmos. Sci. Paper No. 122, Dept. Atmos. Science, Colorado State University, Fort Collins, Colorado.
- Rosby, C. G. (1939): Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action, J. Marine Res. 2, 38-55.
- Thompson, P. D., 1959: *Numerical Weather Analysis and Prediction*, MacMillan, New York.

List of Symbols

C	the speed of sound	u_i	a general dependent variable (a zero subscription on such a variable indicates an unperturbed equilibrium, average or surface value, while a prime superscript denotes a departure from such an equilibrium value. A superbar indicates a time average).
c	Rosby-wave phase velocity	v	the meridional wind component, directed parallel to the y-axis
c_1, c_2	constants of integration	w	the vertical wind component, directed parallel to the z-axis
c_p	specific heat of dry air at constant pressure	x	the eastward-directed axis of a right-handed coordinate system
e	internal energy per unit mass	y	the northward-directed axis of a right-handed coordinate system
f	the Coriolis parameter, = $2\Omega \sin \phi$	z	the vertically-directed axis of a right-handed coordinate system; height, sometimes used as a dummy variable
G	net increase in gravitational potential energy resulting from a virtual parcel interchange	α	expansion coefficient
g	the acceleration of gravity	β	$\partial f / \partial y$
H	the atmospheric scale height	Γ	$\partial T / \partial z$
h	height	γ	ratio of specific heats
J	rate of heat input per unit mass from radiative processes, water phase transformations	ζ	$\partial v / \partial x - \partial u / \partial y$
K	eddy diffusivity (ρK is an eddy exchange coefficient)	Θ	potential temperature; as a subscript, denoting an angular coordinate
K_E	maximum amount of kinetic energy available from a parcel interchange	κ	thermal diffusivity
K_T	thermal conductivity	λ	wavelength
k	frictional coefficient	μ	coefficient of viscosity
p	pressure	ν	μ / ρ (molecular viscosity)
R	specific gas constant for the earth's atmosphere	ρ	density
R_a	Rayleigh number	Φ	rate of dissipation of mechanical energy per unit mass due to shear viscosity
R_i	Richardson's number	ϕ	latitude
R_e	radius of the earth	Ω	earth's angular rotational velocity
r	a radial coordinate	ω_a	absolute angular rotational velocity
T	temperature		
t	time		
U_0	an unperturbed zonal flow upon which a Rossby wave is superposed		
u	a vector velocity		
u	the zonal wind component, directed parallel to the x-axis		

Chapter 4. ATMOSPHERIC TURBULENCE

William C. Meecham

University of California at Los Angeles

The equations governing fluid motion are briefly derived and discussed. After some simplification, for atmospheric turbulence problems, the equations are averaged. Some background information and definitions for turbulent fluid mechanical properties are presented. Step by step the consequences of statistical stationarity, homogeneity, and isotropy are examined. Experimental data for correlation functions and spectra are presented. A new modified discussion of the cascade theory of turbulence is given and its consequences analyzed. The relationship between the Taylor microscale and the viscous dissipation length is discussed and reasons for the difference between the two analyzed. After presentation of atmospheric thermal effects and earth surface roughness effects, semi-empirical formulas for the determination of structure constants for atmospheric temperature fluctuations are presented in tabular form.

4.0 Introduction

Turbulent phenomena comprise one of the great unsolved areas of science and engineering. The equations governing fluid motions are well known and have been known for many years. One might wonder, in such a circumstance, why turbulent phenomena are not by now completely understood. The reason that this is so, as will be seen in succeeding sections, is that the governing equations are strongly nonlinear. As a consequence, the great body of mathematics developed chiefly for linear differential equations is not useful. Fluid dynamicists have dealt with these difficulties chiefly by specializing flows to idealized situations involving boundary layers, wakes, etc., and by approximating turbulent behavior with not always satisfactory empirical laws and rules. If one attempts to solve the equations of motion for turbulent fluids, one finds extremely complicated solutions (functions of position and of time) corresponding to the observed complexities of turbulent flow. Indeed the solutions, either as calculated when this is possible (rarely), or the experimentally observed field variables are so complicated that it is not likely that detailed information concerning them would be useful even if it were available. The approach to these difficulties has been largely one of taking statistical averages of the physical quantities of interest. These averaged quantities have a much simpler dependence upon the position coordinates and upon the time than do the individual fluctuating members from which the averages are constructed. This is the procedure followed in this chapter as will be seen below.

Upon considering possible atmospheric turbulence effects, one is all but overwhelmed by the tremendous variability of the large number of pertinent variables, not to mention the great complexity of the equations governing the processes. For those characteristics of the atmosphere having the greatest influence on propagating radiation one has to consider the large scale effects of the variation of weather and of geographic location (for example, latitude and proximity to mountains, lakes, deserts and the like). Short term variations can be caused by winds and gusts, solar heating (as modified by clouds); they are affected by the character of the surface (local roughness effects). The list of variables could be greatly extended.

Often we need only certain of the characteristics of these fluctuating quantities. But even our limited requirement presents great difficulties of definition, interpretation, and prediction.

In 4.1 we sketch the derivation of the equations governing fluid motion and discuss those equations. After some simplifications suited to the atmospheric turbulence problems at hand we write various field quantities in terms of their averages and their fluctuations from those averages. The field quantities written in these terms are then substituted in the equations governing fluid flow and the resulting equations are in turn averaged. The results of this process appear in (4:19) and (4:20). In 4.2 some background information on fluid mechanical properties is presented. In particular the variations of the index of refraction with the field variables is discussed (the index is of particular importance for optical propagation in the atmosphere). Then the cross correlations of fluid velocity, of temperature, and of velocity with temperature are presented. The time average for obtaining these correlations is defined. We then discuss step-by-step the consequences of statistical stationarity, statistical homogeneity, and statistical isotropy. Furthermore the consequences are discussed assuming, as one often can, that the fluid flow is essentially incompressible. In the same section

experimental data for correlation functions are presented. Finally the structure functions involving velocity and temperature fluctuations are defined. Then in 4.2.3 the energy spectra are defined and discussed and some experimental results presented. In 4.2.4 a new, modified discussion of the cascade theory of turbulence is presented and its consequences analyzed for their bearing upon correlation functions and energy spectra. Experimental data supporting the so called five-thirds law for the velocity energy spectrum are presented. The relationship between the Taylor micro-scale and the viscous dissipation length is discussed and reasons for the difference between the two analyzed. In 4.3 we examine the effects of "real" atmospheric boundary layers. In particular the effect of the earth's surface-roughness is examined largely through the use of empirical and order of magnitude relations. Some data are presented for the effects of various rough surfaces. In 4.3.2 temperature effects in the atmospheric boundary layer are examined and the customary physical quantities for such boundary layers, i.e., the Richardson's number and the Obhukov length are defined. A table is given for the available empirical formulas for the determination of structure constants for temperature fluctuations. The reasoning in support of such relations is discussed.

4.1 Governing Equations

The equations of fluid mechanics which are relevant to our atmospheric problems are those involving the motion of viscous, compressible, Newtonian fluid (this last referring to a fluid in which the viscous stress is proportional to the rate of shear strain) in a uniform gravitational field and non-rotating system (Coriolis force effects can be treated separately). They are given in standard references (e.g., Landau and Lifshitz, 1959). It should be remarked at the outset that the equations, though initially quite complicated, can usually be considerably simplified for atmospheric turbulence problems. A brief discussion of the derivation of the equations will be helpful. In deriving the equations a useful concept is that of the substantial derivative. Consider a small element of mass within the fluid at some position at a given initial time. The time rate of change of some property of that element of mass, following it as it flows, can be obtained by differentiation. Suppose that the property is given the name F . The rate of change of that property following the flow (the substantial derivative or convected derivative) is found to be†

$$\frac{D}{Dt} F \equiv \frac{\partial F}{\partial t} + u_i^* F_{,i} \quad (4:1)$$

Here u^* (with components u_i^*) is the local value of the fluid velocity. We adopt the notation which will be convenient here: $F_{,i} = \partial F / \partial x_i$, where x_i are the Cartesian coordinates of position with x_3 vertical; the coordinates will sometimes be designated x, y, z . Further, we adopt the summation convention; that is, we should sum from 1 to 3 over subscripts which are repeated within a single term.††

We can make use of the substantial or convective derivative given in (4:1) to obtain the governing equations. First of all, consider the equation of continuity. The quantity of mass of a given element of fluid must be conserved as the element flows: its substantial derivative must vanish. Let $F = \rho^* \delta V$, where δV is a small volume of fluid at a given instant and ρ^* is its density. Then, since the mass must be conserved, we know

$$\frac{D}{Dt} (\rho^* \delta V) = 0 . \quad (4:2)$$

By simple manipulation, one can show that,

$$(\partial V)^{-1} \frac{D}{Dt} (\delta V) = u_{i,i}^* . \quad (4:3)$$

† In terms of derivatives this is $(DF/Dt) = \partial F/\partial t + u_i^* \partial F/\partial x_i$. It is worth commenting here that we use the equations in their Eulerian form, that is where the derivatives are taken in a fixed frame of reference. The Lagrangian form has derivatives taken in a frame moving locally with the fluid (again "following the flow", i.e., for the time derivative just D/Dt). The difference is most marked in a time integration; the Lagrangian equations would be integrated along flow lines, a conceptually complicated procedure.

†† An alternative discussion is: With $F = F(x,y,z,t)$ then ΔF equals the change in F due to change in time at a fixed point in the fluid plus the change in F which would occur for F not a function of time, but varying with position. This equals $\Delta F = \Delta t(\partial F/\partial t) + \Delta x(\partial F/\partial x) + \Delta y(\partial F/\partial y) + \Delta z(\partial F/\partial z)$, which implies (4:1).

Substituting in (4:2), we have for the equation of continuity†

$$\frac{\partial \rho^*}{\partial t} + (\rho^* u_i^*)_{,i} = 0. \quad (4:4)$$

Consider Newton's Law, $f = (d/dt)mV$. Again using (4:1) for the time rate of change of a quantity following the flow, where now for F we use the i th component of the momentum given by $\rho^* \delta V u_i^*$, we have the relation

$$\frac{1}{\delta V} \frac{D}{Dt} (\rho^* u_i^* \delta V) = G_i, \quad (4:5)$$

where G_i represents the force per unit volume exerted upon the element of fluid††. For our purposes there are three forces acting within the fluid: the force due to pressure changes within the fluid, the force due to viscous stress within the fluid, and the force of gravity.

The mathematical derivation of expressions for the three forces can be found in, for example, Landau and Lifshitz (1959). The pressure force per unit volume is the negative of the pressure gradient (there is no such pressure force in a constant pressure field). The viscous force per volume, allowing for variable viscosity at this stage, is the divergence of the viscous stress. The Newtonian viscous stress, details aside, is the most general symmetric tensor of second-rank which one can form from first-order derivatives of the velocity. The gravitational force per unit volume is in the negative x_3 -direction and of magnitude $\rho^* g$. We substitute the three forces for the right-hand side of (4:5). In the left-hand side make use of the continuity equation given in (4.4). One finds for the Navier-Stokes equations governing fluid flow†††

$$\rho^* \left(\frac{\partial u_i^*}{\partial t} + u_{i,j}^* u_j^* \right) = \left[- \left(p^* + \frac{2}{3} \mu u_{j,j}^* \right) \delta_{ik} + \mu \left(u_{i,k}^* + u_{k,i}^* \right) \right]_{,k} - \rho^* g \delta_{3i}. \quad (4:6)$$

Here g is the acceleration due to gravity, assumed constant. The μ is the fluid viscosity and p^* is the fluid pressure. The δ_{ij} is Kronecker's Delta; it is unity if the subscripts are equal, and otherwise zero.

Consider now the entropy of the moving fluid. The entropy changes in general during the fluid motion. One obtains for the equation governing time rate of change of the entropy following the fluid flow (Landau and Lifshitz, 1959)

$$\rho^* T^* (DS/Dt) = u_{i,j}^* \left[- \frac{2}{3} \mu u_{m,m}^* \delta_{ij} + \mu \left(u_{i,j}^* + u_{j,i}^* \right) \right] + (k T^*_{,i})_{,i}. \quad (4:7)$$

Here T^* is the temperature, S is the entropy per mass, and k is the thermal conductivity. The entropy changes for two reasons. The action of the viscous stresses generates heat which can be shown to lead to the first term on the right side. Further there is heat conduction which also changes the entropy; the net conduction is the divergence of the thermal heat flux giving the second term.

The equation of state for the atmosphere is to an excellent approximation that for an ideal gas,

$$p^* = R \rho^* T^* \quad (4:8)$$

where R is the gas constant per mass, which, for an ideal gas, leads to the relation

$$R = c_p - c_v, \quad (4:9)$$

where c_p and c_v are the specific heats of air per unit mass at constant pressure and constant volume. (We shall neglect at this stage effects of variable moisture content).

† In terms of derivatives this equation is $\frac{\partial \rho^*}{\partial t} + \frac{\partial (\rho^* u_i^*)}{\partial x_i} = 0$.

†† The forces on a given fluid element give the time rate of change of the momentum of that fluid element: we need the *substantial* derivative of the momentum for a proper statement of Newton's law.

††† In the notation of vector analysis, this is $\rho^* \left(\frac{\partial \mathbf{u}^*}{\partial t} + \mathbf{u}^* \cdot \nabla \mathbf{u}^* \right) = -\nabla p^* - \frac{2}{3} \nabla (\mu \nabla \cdot \mathbf{u}^*) + \frac{\partial}{\partial x_k} \mu \frac{\partial \mathbf{u}^*}{\partial x_k} + \frac{\partial}{\partial x_k} \mu \nabla u_k^* - \rho^* g \mathbf{k}$ with \mathbf{k} the unit vector in the x_3 -direction.

We now define certain distributions $T_0(x_3)$, $p_0(x_3)$ and $\rho_0(x_3)$, for convenience. It will be seen that the introduction of the quantities leads to equations governing the remaining thermodynamic variable variation which are not explicitly dependent on g . These quantities are the variable distributions which lead to a neutrally stable atmosphere: such an atmosphere at rest is neutrally stable to small displacements. These characteristics are discussed in detail in 4.3.2. For now we take them to be defined quantities which are convenient in the discussion. Their governing equations are

$$T_{0,i} = -\frac{g}{c_p} \delta_{3i} = -\frac{(1 - \frac{1}{\gamma})}{R} g \delta_{3i}, \quad (4:10)$$

the second equation following from the thermodynamic relation (4:9) and where $\gamma \equiv c_p/c_v$ and

$$-\frac{1}{\rho_0} p_{0,i} = g \delta_{3i} \quad (4:11)$$

and

$$p_0 = R \rho_0 T_0. \quad (4:12)$$

Consider now the dependence of density changes upon pressure and temperature. There are pressure fluctuations in the atmospheric boundary layer (caused by velocity fluctuations) of order $(10^{-6}) \sigma_u^2$ where σ_u is the RMS (gust) velocity fluctuation, say 3 meters per second. Temperature fluctuations there are of order one degree Kelvin. Hence, in cgs, $(\sigma_p/\bar{p}) \sim (\frac{1}{2} \times 10^{-3}) \times (3 \times 10^2)^2 (10^6) \sim \frac{1}{2} \times 10^{-4}$; $(\sigma_T/\bar{T}) \sim \frac{1}{3} \times 10^{-2}$ where the over bars indicate time averages, discussed below. Consequently from (4:8) for the small relative fluctuations which we have, we can neglect the effect of pressure fluctuations on the density fluctuations.

As discussed in connection with (4:10), the introduction of p_0 , T_0 and ρ_0 (independent of x_1 , x_2 and t_0) leads to equations for $p^* - p_0$, $T^* - T_0$ and $\rho^* - \rho_0$ which are independent of g . Now the difficult nonlinear character of the equations makes simplification essential. This is usually accomplished by asking for less information. Instead of searching for the complicated functions representing the pressure etc., we ask rather for their (much simpler) averages. We are lead to define the following variables:

$$\begin{aligned} p^* &= p_0 + \bar{p} + p \\ T^* &= T_0 + \bar{T} + T \\ \rho^* &= \rho_0 - \rho_0(\bar{T} + T)/T_0 \end{aligned} \quad (4:13)$$

and

$$\mathbf{u}^* = \bar{\mathbf{u}} + \mathbf{u}, \quad (4:14)$$

Here in the third equation, we make use of the equation of state and have supposed that the relative deviation of the temperature from the stable temperature T_0 is small. The overbar indicates time averaged quantities. Note here $\bar{p}^* = p_0 + \bar{p}$ etc., i.e., \bar{p} is the difference between the average pressure and the time-independent (neutrally stable) p_0 . Instantaneous fluctuations are given by p , T , and u . Note that the quantities p , T , and u are *fluctuations* (in keeping with fluid mechanics usage) not the total pressure etc. The time average of such fluctuations by definition vanishes. We suppose, in this discussion, that fluctuating quantities of chief interest to us are such that their statistical characteristics do not change over moderate time. The process is then said to be statistically stationary. It is, of course, true that over periods of an hour or so average characteristics of atmospheric turbulence can and often do change significantly. However, it is usually sufficient to average over time somewhat shorter than this; over such shorter times we suppose that the process is, as stated, stationary. Of course the averaging time must be large compared with characteristic time of the fluctuations. Finally we also suppose that the relative deviation of the pressure from the stable atmospheric pressure is small.

Taking the ratio of (4:11) to (4:12), we find $(dp_0/dx_3) = -p_0 g/RT_0$. Thus a characteristic length for the atmosphere (the so-called scale height of the atmosphere) is seen to be

$$H \equiv \frac{RT_0}{g} = c_0^2/\gamma g \tag{4:15}$$

where c_0 is the speed of sound for the temperature T_0 . The H is the vertical distance over which the pressure changes by a fractional amount. For phenomena with characteristic lengths considerably less than H, we find that the equation of continuity (4:4) reduces to a simple incompressibility condition, which in conjunction with the defining equation for the velocity (4:14) leads to

$$\bar{u}_{i,i} = u_{i,i} = 0 \tag{4:16}$$

One readily obtains the thermodynamic relation for entropy changes (Lumley and Panofsky, 1964),

$$dS = \frac{c_p}{T^*} dT^* - R \frac{dp^*}{p^*} \tag{4:17}$$

If now we consider changes in entropy following the flow, we have the relation

$$\frac{DS}{Dt} = \frac{c_p}{T^*} \frac{DT^*}{Dt} - R \frac{Dp^*}{Dt} \tag{4:18}$$

where (4:18) is to be substituted for the entropy change in (4:7). In that equation we can ordinarily neglect viscous heat production in the energy balance, thus neglecting the first term on the right side of (4:7). Further, in atmospheric problems we can usually neglect changes in the viscosity and thermal conductivity, at least at the scale lengths of interest in this discussion. Using the various approximations and definitions, we substitute in (4:6) and (4:7). In discussions in this report, we shall primarily be interested in average quantities and accordingly we average the result of the substitutions in (4:6) and (4:7), recalling that the averages of the fluctuation quantities p , T , and u all vanish. One finds

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_{i,j} \bar{u}_j = -(\bar{u}_i \bar{u}_j)_{,j} - \frac{1}{\rho_0} \bar{p}_{,i} + \nu \bar{u}_{i,jj} + \frac{g}{T_0} \bar{T} \delta_{i3} \tag{4:19}$$

$$\frac{\partial \bar{T}}{\partial t} + \bar{T}_{,i} \bar{u}_i = -(\bar{T} \bar{u}_i)_{,i} + k \bar{T}_{,ij} \tag{4:20}$$

Here it is convenient to define the kinematic viscosity $\nu = \mu/\rho_0$. The first term on the right side of (4:19) represents an effective force per unit mass caused by the divergence of the fluctuating momentum flux. The first term on the right side of (4:20) gives the effective heat flux caused by convective effects. These quantities are particularly important in the discussion which follows.

4.2 Background Information On Fluid Mechanical And Related Properties

4.2.1 Discussion of Index of Refraction

When we are particularly interested in optical frequency propagation, the main effect of the atmosphere comes from changes in the refractive index. The refractive index depends primarily upon the density of the atmosphere. We find in standard references (Allen, 1964) that

$$\Delta n \equiv n - 1 = 82.6 \times 10^{-6} p^*/T^* \tag{4:21}$$

with n the refractive index, p^* the pressure in milli-bars and T^* the temperature in degrees Kelvin. We represent RMS fluctuations of quantities by σ with a subscript indicating the variable involved and find

$$[\sigma_n/(n - 1)]^2 = (\sigma_p/\bar{p}^*)^2 + (\sigma_T/\bar{T}^*)^2 \tag{4:22}$$

valid for small relative fluctuations and neglecting any (slight) statistical dependence of the pressure on temperature fluctuations. In discussion following (4:12) we found we could neglect the effect of pressure fluctuations so,

$$\sigma_n \cong + \left[82.6 \times 10^{-6} \frac{\bar{P}^*}{\bar{T}^*} \right] \frac{\sigma_T}{\bar{T}^*} . \quad (4:23)$$

4.2.2 Simplifying Statistical Assumptions

We suppose, as discussed earlier, that the wind velocity is given by

$$\mathbf{u} + \bar{\mathbf{u}} \quad (4:24)$$

where $\bar{\mathbf{u}}$ is the average wind speed and \mathbf{u} is the instantaneous departure from that average. We are also interested in temperature fluctuations. We have defined the temperature of the air at a given position and time to be

$$T_0 + \bar{T} + T \quad (4:25)$$

where $T_0 + \bar{T}$ is the temperature average over a time – as discussed above – which is shorter than the time within which the whole statistical structure of the atmospheric turbulence changes, and large compared with the time in which the short-term temperature fluctuations occur. The quantity T is the instantaneous fluctuation, as defined earlier.

We consider the cross correlation of quantities measured at two different positions. It is convenient in what follows to define the three Cartesian components, u_i , of the velocity fluctuation vector \mathbf{u} . Consider the cross correlation of the velocity fluctuations measured at two points, \mathbf{r}' and \mathbf{r}'' . This correlation is defined by the time average, represented by $\langle \rangle$ (we use this and overbar interchangeably).

$$Q_{ij}(\mathbf{r}', \mathbf{r}'', t) \equiv \left\langle u_i(\mathbf{r}', t') u_j(\mathbf{r}'', t' + t) \right\rangle = \lim_{T \text{ large}} \frac{1}{T} \int_0^T u_i(\mathbf{r}', t') u_j(\mathbf{r}'', t' + t) dt' . \quad (4:26)$$

Our assumption that the process is statistically stationary means that the correlation is approximately independent of which segment, T , of velocity fluctuation we use in the determination of the average (correlation function). Other cross correlations are important to us; the cross correlation of temperature fluctuations is:

$$S(\mathbf{r}', \mathbf{r}'', t) \equiv \left\langle T(\mathbf{r}', t') T(\mathbf{r}'', t' + t) \right\rangle , \quad (4:27)$$

with the average defined as the time average in a way analogous to that in (4:26). The often-important transport of the temperature fluctuations within the turbulent atmosphere is controlled by the cross correlation of the velocity with the temperature, defined by

$$T_i(\mathbf{r}', \mathbf{r}'', t) \equiv \left\langle T(\mathbf{r}', t') u_i(\mathbf{r}'', t' + t) \right\rangle \quad (4:28)$$

The quantities (4:26) and (4:28) were discussed in connection with (4:19) and (4:20). The time average is analogous to that of (4:26). When the two points of measurement, \mathbf{r}' and \mathbf{r}'' are separated by a distance considerably greater than some length L_0 , the fluctuations become statistically independent of one

another and it follows that the correlations given in (4:26) – (4:28) go to zero (this is so because by the definitions (4:24) and (4:25) the average of the fluctuations themselves must vanish). This distance L_O is called the correlation length, or outer scale. It is of great importance in discussing propagation characteristics of the atmosphere and in many other connections. Similarly, if the time delay, t , between the measurements, even though the two points are closer than L_O , exceeds a time characteristic of the turbulence T_O , then again the fluctuating quantities become statistically independent of one another and the cross correlations given in (4:26) – (4:28) go to zero. This time, T_O , is called the correlation time for the turbulent process. It is typically dependent upon the quantities already defined. In most turbulent processes when measurements are made moving with the mean flow,

$$T_O = \frac{L_O}{\sigma_u} . \tag{4:29}$$

Thus the characteristic time (correlation time) for the turbulent process is of the order of the correlation length, L_O (often called the “size” of the large eddies) divided by the velocity fluctuation. If the turbulence were not being constantly regenerated by the wind shear forces in the atmosphere, the whole turbulent process would decay after several such times T_O leaving no fluctuations. This is the qualitative behavior of turbulence observed in laboratory wind tunnel experiments where the turbulence is not regenerated within the wind tunnel. The discussion here deals with a situation where there is no mean wind (or if there is, as suggested, one must move with the mean flow.) If, as is usual, there is a wind and our measurement point is fixed, there are time fluctuations caused by the transport of the fluid past the measurement point. These effects are discussed below.

Often, atmosphere turbulence shows little statistical change when we move from one position to another. Such behavior is called *statistical homogeneity*. It is, of course, not rigorously true in any application. In the atmosphere, the situation is complicated by the fact that the large scale characteristics change when we move vertically a distance L_O , the scale of the turbulence. Nevertheless, for many purposes, the assumption of statistical homogeneity is a useful approximation. For such homogeneous turbulence, the correlations (4:26) – (4:28) reduce to functions of time and space functions of the displacement vector

$$\mathbf{r} \equiv \mathbf{r}'' - \mathbf{r}' \tag{4:30}$$

between the two points at which measurements are made. It is obvious that this reduces the number of independent variables in the problem. In the remainder of this section, we restrict the discussion to those properties that can be found, at least approximately, by assuming statistical homogeneity.

Often the average quantities defined here, in particular the smaller-scale characteristics, are approximately independent of the orientation of the displacement factor \mathbf{r} . The idealization that the turbulence has this property is called statistical isotropy. This is a further, most useful, simplifying assumption for the discussion of atmospheric turbulence. In conjunction with the previously discussed simplifying assumptions, statistical isotropy leads to the following reductions of complexity in the cross-correlations (4:26) – (4:28). The reader is referred to Batchelor (1953) for a more extended discussion of the symmetry characteristics presented here:

$$Q_{ij} = Q_1(r,t)r_i r_j + Q_2(r,t)\delta_{ij} ; \tag{4:31}$$

$$S = S_1(r,t) ; \tag{4:32}$$

$$T = T_1(r,t)r_i . \tag{4:33}$$

The coefficient functions Q_1 , Q_2 , S_1 , and T_1 are called “scalar generators” and depend only on the *length* of the displacement vector \mathbf{r} and on the time delay between the measurements. The simple rule for forming all such statistically isotropic correlations is to make up all relevant combinations of the vector components r_i and of the Kronecker delta symbol δ_{ij} , as seen in the examples here.

Often in atmospheric turbulent processes, the fluid flow is essentially incompressible see (4:34). As a result of this incompressible condition, the correlations (4:26) can be further simplified

$$U_i, i = 0 \tag{4:34}$$

$$Q_{ij} = - \frac{1}{2r} \frac{\partial Q(r,t)}{\partial r} r_i r_j + \left(Q + \frac{r}{2} \frac{\partial Q}{\partial r} \right) \delta_{ij} \tag{4:35}$$

$$T_i = 0 ; \tag{4:36}$$

the correlation, S , is unchanged by the incompressibility assumption. The function, Q , is a longitudinal velocity correlation: it is the velocity correlation wherein one measures the components in the direction of the displacement vector \mathbf{r} (between the measurement points). Some typical measurements of the function $Q(r,0)$ are shown in (F4.1).

The measurements shown in (F4.1) were made in a wind tunnel. The turbulence is generated by a square-mesh grid of size M . The mean flow speed is U . The Reynolds number, UM/ν , (where ν is the ratio of the viscosity to the density) is given various values in the experiment by varying the flow speed. All measurements are made at a given distance downstream. As is seen from the figure, the dimensionless plot collapses the data at all separations except the smallest ones. The longitudinal correlation, $Q(r,0)$, typically shows very large curvature at the origin, where the measurements depend upon the value of the viscosity. The half-width of the correlation function is on the order of the large scale, (here the mesh size M) and is relatively independent of the viscosity – or the Reynolds number – and thus as we shall see later, independent of what is called the inner scale.

The remarkable fact that the cross-correlation of the temperature and velocity fluctuation vanishes under these assumptions, as seen in (4:36), deserves special comment. This vanishing of the cross-correlation would occur for the cross-correlation of any scalar with the velocity: it is called *statistical orthogonality*. The null result is akin to the vanishing, say, of the cross-correlation of the sine with the cosine function. In a way it corresponds to the two fluctuating quantities being “90° out of phase” with one another. It should be emphasized that (4:36) does *not* imply that the temperature and velocity fluctuations are statistically independent of one another. The cross-correlation of the temperature fluctuation with the square of the velocity fluctuation, for instance, would not vanish in general.

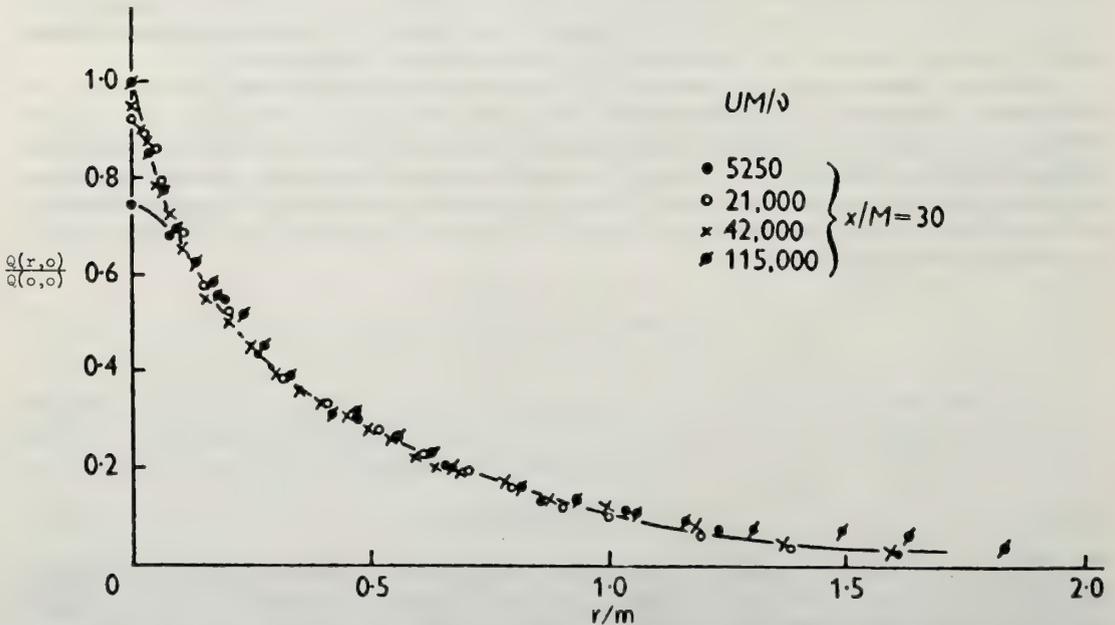


Figure 4.1 Adjusted correlation functions at different Reynolds numbers (Stewart and Townsend, 1951).

The above discussion is based, as indicated, upon idealized assumptions concerning the symmetry of the turbulent structure involved in temperature and velocity fluctuations. First, it is assumed that the turbulence is statistically stationary. Furthermore, we have supposed that the turbulent process is statistically homogeneous and statistically isotropic as well. This is usually found to be the case for the smaller-scaled structure; but for the large scales, that is, for the large eddies, such simplifying assumptions cannot be justified. Velocity and temperature differences are approximately homogeneous and isotropic if r is small compared with L_0 . Thus, consider "structure functions" for these fluctuating quantities. For example, define a structure function for the velocity fluctuation as in (4:37). (Structure functions are usually defined for simultaneous measurements. When time is a variable, it will be implicit in what follows.)

Thus,
$$D_{ij}(\mathbf{r}', \mathbf{r}'') \equiv \left\langle [u_i(\mathbf{r}') - u_i(\mathbf{r}'')] [u_j(\mathbf{r}') - u_j(\mathbf{r}'')] \right\rangle. \quad (4:37)$$

For $|\mathbf{r}'' - \mathbf{r}'| \equiv |\mathbf{r}| \ll L_0$ this correlation is often assumed to be statistically homogeneous and isotropic, that is, to depend only on r : The smaller scale structure within the turbulence is of more universal form than is the large-scale. For such small scales, we can write D_{ij} in a form similar to (4:35) (with $t = 0$ for the simultaneous measurements under discussion). The structure function notation follows Tatarskii (1959). Define the longitudinal structure function, $D_{rr}(r)$, which is the value of (4:37) when the measurement at both points is of the component of the velocity fluctuation in the direction \mathbf{r} ; in detail, $D_{rr}(r) \equiv D_{11}(\hat{\mathbf{r}})$.

In a way analogous to that for the velocity structure function, one can define a structure function for temperature when small-scale effects are statistically homogeneous and isotropic:

$$D_T(r) = \left\langle [T(\mathbf{r}') - T(\mathbf{r}' + \mathbf{r})]^2 \right\rangle \simeq 2S_1(0) - 2S_1(r) \quad (4:38)$$

with the function S_1 defined in (4:27) and (4:32) for simultaneous measurements. When needed, one could similarly define the cross correlation between temperature fluctuations and velocity fluctuations.

There is a further practical consideration which should be mentioned in connection with the measurement of these correlations. Often the turbulence is being transported with a mean velocity, that of the prevailing wind at the altitude in question. This characteristic can be utilized to simplify the measurement process. Let us consider a simple case. Suppose that the horizontal wind speed in the x -direction is \bar{u} ; then if we measure the temperature fluctuation, T , in time at one point and take the cross-correlation of this temperature fluctuation with the temperature fluctuation at a later time, t , at the same point r , we obtain

$$\left\langle T(\mathbf{r}', t') T(\mathbf{r}' - \hat{\mathbf{u}}t, t' + t) \right\rangle, \quad (4:39)$$

with $\hat{\mathbf{u}}$ the unit vector in the wind direction. As discussed above, the characteristic length for the spatial dependence is L_0 ; the characteristic time for the time dependence is L_0/σ_u . It follows that there will be significant changes in the correlation due to the influence of time on the spatial dependence when t is of order L_0/\bar{u} . Similarly, there will be significant changes due to the time delay in the time dependence when t is of order L_0/σ_u . Since σ_u/\bar{u} is typically of order 5 percent, it is seen that the time-dependence effect in the spatial variable position is much more important than that in the time variable. If we neglect the effect of the time delay in the time variable, we, in effect, assume that the turbulence shows no time dependence; it is in steady flow merely being transported with the speed \bar{u} . This is called the Taylor hypothesis or a frozen-flow hypothesis. As we see from this discussion, errors incurred through the use of such a hypothesis amount typically to a few percent. Corresponding correlations utilizing structure function definitions are easily devised by making measurements at two space points. By adjusting the time delay and the separation of the points one may obtain space-time correlations in situations like that described here where there is a mean wind transporting the turbulence.

Finally it is interesting to note that in most turbulence applications the fluctuating quantities ($u_1, u_2, u_3; T, \rho, p$ etc) are approximately Gaussian. This behavior leads to approximate theoretical formulations from which some useful results can be obtained (see e.g. Meecham and Jeng, 1968).

4.2.3 Energy Spectra

Consider the triple Fourier transform of the velocity correlation function,

$$\Phi_{ij}(\mathbf{k}, t) \equiv (2\pi)^{-3} \iiint e^{-i\mathbf{k}\cdot\mathbf{r}} Q_{ij}(\mathbf{r}, t) \, dx dy dz \tag{4:40}$$

where $r = (x, y, z)$. The wave number k is equal to $2\pi/\lambda$, where λ is the size of the “eddy” under examination in the Fourier transform. The turbulent kinetic energy per unit mass is one half the average of the square of the velocity fluctuation. The kinetic energy per unit mass is given by

$$\text{K.E.} = \frac{1}{2} Q_{ij}(0,0) \equiv \int_0^\infty E(k) \, dk \tag{4:41}$$

where we employ the summation convention on the repeated index j – i.e., (4:41) means $(1/2)(Q_{11}+Q_{22}+Q_{33})$. From (4:40) the energy spectrum function $E(k)$ is seen to be related to the Fourier transform by

$$E(k) = 2\pi k^2 \Phi_{ii}(k,0) . \tag{4:42}$$

We have supposed here that the turbulence is statistically isotropic, in which case the spectrum function E , is dependent only on the magnitude of the vector \mathbf{k} . The quantity $E(k)dk$ is seen from its definition to give the energy contained by eddies within the wave number range between k and $k + dk$.

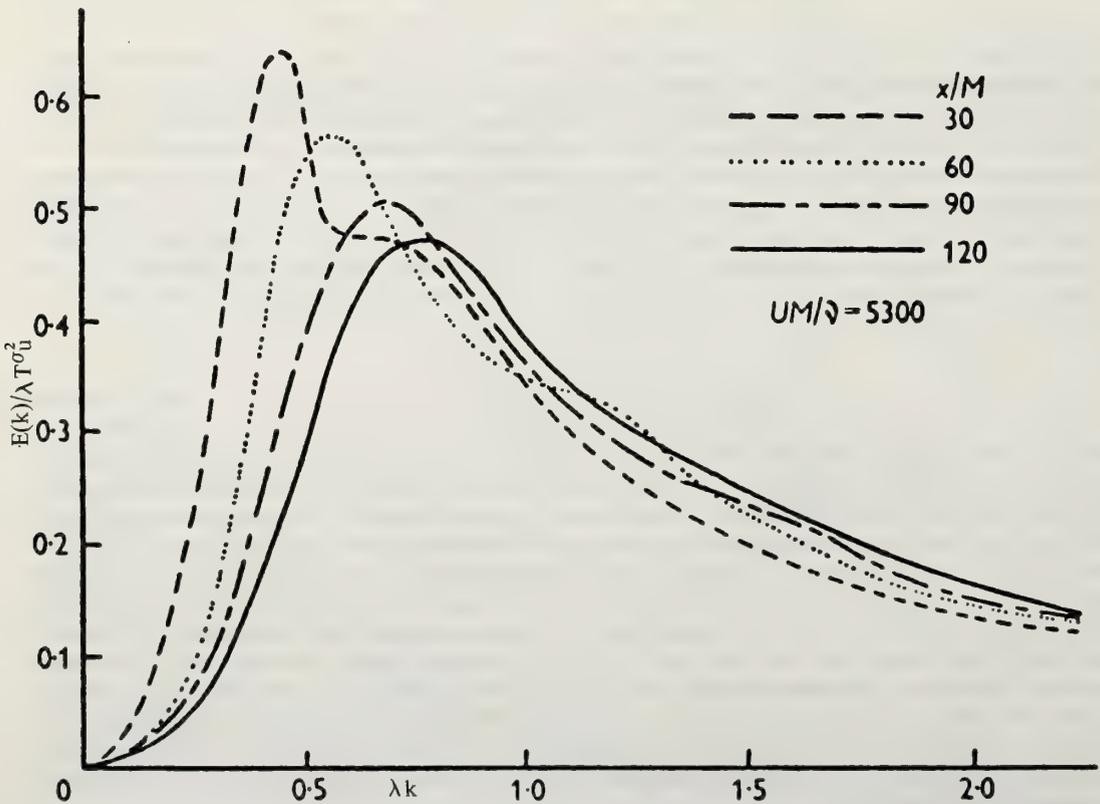


Figure 4.2 Spectrum function at different distances, x , downstream in the turbulence-generating grid in wind tunnel experiments (Steward and Townsend, 1951).

Similarly, we can define the energy spectrum for the turbulent temperature fluctuations using the autocorrelation function S_1 defined in (4:32). We write the Fourier transform of that function

$$\Phi_T(k) = \frac{1}{\pi} \int_0^\infty e^{-ikx} S_1(x,0) dx . \tag{4:43}$$

The mean square temperature fluctuation, which plays a role here akin to kinetic energy above, can be seen to be given by

$$S_1(0,0) = \int_0^\infty \Phi_T(k) dk \equiv \sigma_T^2 . \tag{4:44}$$

Again here we have the physical significance that in the mean square fluctuation the quantity $\Phi_T(k)dk$ gives the “amount of the fluctuation” arising from eddies with wave numbers in the range between k and $k + dk$. (For the cross-correlation defined in (4:28) there is an energy spectrum analogous to E and to Φ_T .) It can be shown that these energy spectra are real, positive functions. Results of wind tunnel measurements of the spectrum function $E(k)$ are shown in (F4.2). The quantity λ_T is the so-called Taylor microscale, or dissipation length. It is defined by (Batchelor, 1953)

$$\lambda_T = \left[15\nu\sigma_u^2/\epsilon \right]^{1/2} = \sigma_u \left[\frac{-d^2 Q(r,0)}{dr^2} \right]_{r \rightarrow 0}^{-1/2} \tag{4:45}$$

with ϵ the time rate of decay (or supply) of kinetic energy per unit mass. The length λ_T is seen to be proportional to the radius of curvature of the correlation function at the origin (F4.1). The length λ_T , though useful in data treatment, is known *not* to be the scale at which dissipation occurs (F4.4 and discussion below in this connection). We return to this point in the next section, following the discussion of the cascade theory. In atmospheric turbulence λ_T is typically measured to be a foot or two (Jones et al., 1969). From the Navier-Stokes equations governing fluid flow, we know that the energy dissipation within approximately incompressible turbulent fluid flow is accounted for entirely by the viscosity. From these equations it can be shown that the energy dissipation, ϵ , is given by

$$\epsilon = 2\nu \int_0^\infty k^2 E(k) dk . \tag{4:46}$$

It follows immediately from (4:40) – (4:42) that the quantity ϵ is proportional to the second derivative of the longitudinal correlation function at the origin: $Q''(0,0)$.

4.2.4 The Cascade Theory of Turbulence

There is a simplifying hypothesis concerning the nature of the energy transfer within turbulence which is useful in connection with atmospheric problems. Some years ago Kolmogorov (1941) proposed the following view of the turbulent energy transfer process (F4.3). A quantity of power per mass, ϵ , is supplied to the turbulence, usually being injected in the largest turbulence scale, of order L_0 (through the non-linearity of the fluid flow process). This energy is passed, again because of the non-linearity of the equations of motion, down to those eddies which dissipate the energy. (If the equations were linear, there would be no energy transfer between different size eddies.) The dissipating eddies are very small, with size of order ℓ_0 , the so-called inner scale. These eddies dissipate the energy as heat. Kolmogorov makes an assumption concerning the way in which the non-linearities of the fluid flow process pass energy from the very large eddies to the very small ones: One might reasonably suppose that this quantity of energy per second, ϵ , is passed from the largest to the next smaller eddies; and in turn from those to yet smaller eddies and so on, down to the smallest (dissipation) eddies, of size ℓ_0 . This (assumed) transfer is called a cascade. There is, as yet, only indirect experimental evidence that this is the mode of energy transfer. Nevertheless, if we suppose that this is the process, there is a considerable simplification in the turbulence description. Suppose that we want to find the energy spectrum $E(k)$. The parameters in the problem consist of ϵ , ℓ_0 , L_0 , and σ_u . The quantities ϵ , L_0 , and σ_u are known to be connected by the following empirical relation (Batchelor, 1953)

$$\epsilon \sim A \frac{\sigma_u^3}{L_0} . \tag{4:47}$$

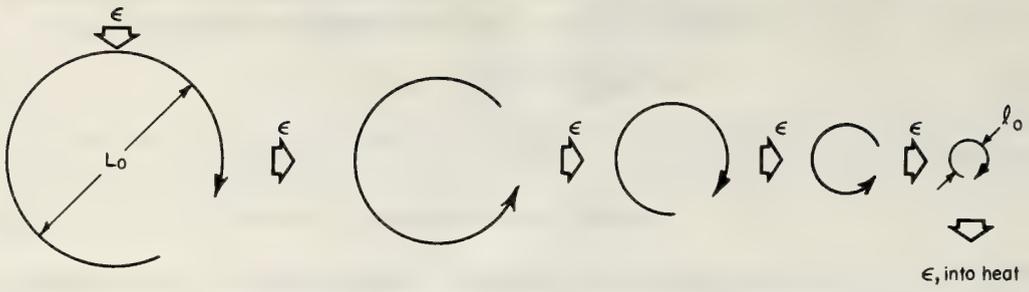


Figure 4.3 Sketch of the hierarchy of eddies for energy cascade.

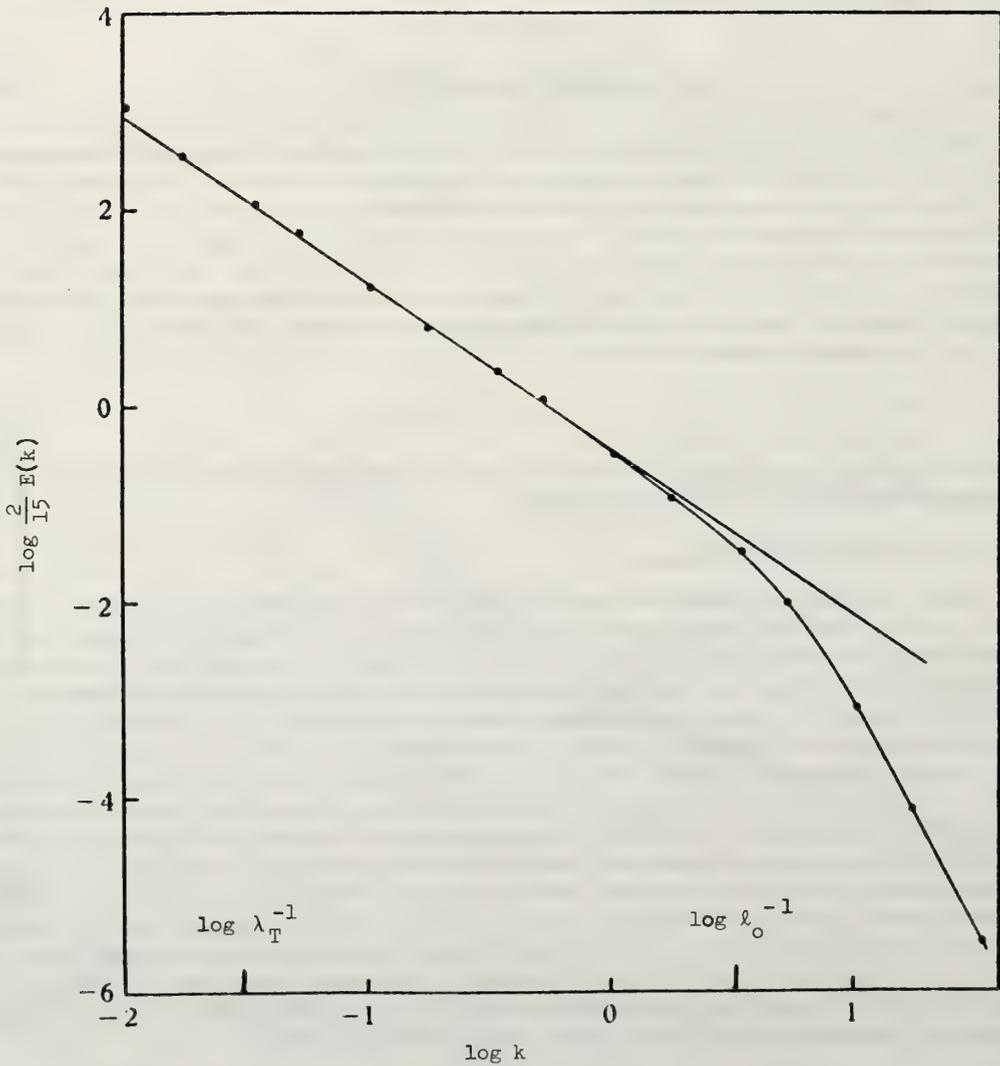


Figure 4.4 A logarithmic plot of the spectrum for one run. The straight line has a slope of $-5/3$; k is in cm^{-1} (Grant et al., 1962).

If one believes that the rate of loss of energy from the large eddies is independent of the small-scale viscosity dependent effects then (4:47) follows by dimensional analysis.† Put another way, we might suppose that the transfer of energy from larger to smaller eddies is accomplished by the nonlinear inertial effects, with viscosity playing a role only in the energy dissipation (into heat) by the smallest eddies. Thus we can eliminate one of the parameters appearing in (4:47) let that one be σ_u . If we examine the energy spectrum in the intermediate range

$$\ell_0 \ll 2\pi k^{-1} \ll L_0 \tag{4:48}$$

we may reasonably suppose that the function $E(k)$ is independent of ℓ_0 (the inner scale) and L_0 (the outer scale). Of course, in order that (4:48) be possible, it is necessary that L_0 be extremely large compared with ℓ_0 . Allowing an order of magnitude for the range of variation of $2\pi k^{-1}$ we see from (4:48) that we need

with

$$\frac{L_0}{\ell_0} = Re'^{3/4} > 10^3$$

$$Re' \equiv \frac{\sigma_u L_0}{\nu}, \tag{4:49}$$

using the size of ℓ_0 given in (4:51) below. Such large Reynolds's numbers are indeed attained in lower-atmospheric turbulence. Using dimensional analysis, it immediately follows from the above reasoning that the energy spectrum is given by

$$E(k) = a_1 \epsilon^{2/3} k^{-5/3}, \frac{2\pi}{L_0} \ll k \ll \frac{2\pi}{\ell_0} \tag{4:50}$$

where here and below the a_n are dimensionless constants of order unity.

Before discussing this energy spectrum function, consider the inner scale, ℓ_0 . Kolmogorov suggested (as seems plausible from the nature of the physical process) that the inner scale is independent of the outer scale, L_0 , and thus depends only upon ϵ and upon the kinematic viscosity, ν . In such a case one obtains from the use of dimensional analysis the following result for the inner scale:

$$\ell_0 = a_2 (\nu^3 / \epsilon)^{1/4}. \tag{4:51}$$

In the atmosphere, ℓ_0 is typically of order of millimeters (Monson et al., 1969). Similarly the *average* velocity change associated with the dissipation is found by dimensional analysis,

$$(\overline{\Delta u^2})^{1/2} \sim (\nu \epsilon)^{1/4}. \tag{4:52}$$

The results of some important measurements of $E(k)$ by Grant et al. (1962), are shown in (F4.4). These experiments were performed in British Columbia in a tidal channel near Vancouver. The results are plotted log-log, to bring out the proposed algebraic dependence of the energy spectrum function upon the wave number see (4:50). It is seen that the spectrum in these experiments had the suggested form over eight octaves of k . The quantity, ϵ , was obtained by integration as in (4:46). The energy spectrum, $E(k)$, was measured directly, in effect using its definition. It is noted that these experiments show that the inner scale, the scale at which dissipation occurs, is of order ℓ_0 , as defined in (4:51). It is approximately at ℓ_0 that the spectrum shows an abrupt reduction, not at the Taylor dissipation length λ_T . The reason that this is so can be seen in the following way (we adopt the view that the cascade description of the turbulent process is valid). The dissipation may be supposed to consist of velocity jumps $\overline{\Delta u^2}^{1/2}$ (4:52) with a volume average of order $(\nu \epsilon)^{1/4}$, where the characteristic distance in which the jump occurs is of order $(\nu^3 / \epsilon)^{1/4}$ (4:51). These dissipation regions are assumed to occupy a fraction α of the volume of the turbulence; of course, α would be near unity

† Dimensional analysis is of particular importance in fluid mechanics. Such analysis is performed as follows: Suppose we have parameters and variables, of importance in a particular problem, called x_1, x_2, \dots, x_n . Form all possible independent dimensionless quantities from them, called x_1, x_2, \dots, x_m . Then any physical law for the problem can be written in the form $\Pi(x_1, x_2, \dots, x_m) = 0$ where Π is some function. For example, consider the period of a simple pendulum. We may reasonably suppose that the important parameters are L , the length of the pendulum; M , its mass and g . We have only one possible dimensionless variable $g\tau^2/L$. Then solving $\Pi(g\tau^2/L) = 0$ for τ we have $\tau = \bar{a}\sqrt{L/g}$ where \bar{a} is dimensionless. Constants like \bar{a} are usually of order unity (between 0.1 and 10) for reasons known only to God. For this problem $\bar{a} = 2\pi$ of course.

if the dissipation regions completely fill the volume. From this model of the dissipation process we can estimate the Taylor microscale from (4:45). We obtain using the Kolmogorov estimates

$$\lambda_T \sim \sigma_u \frac{\ell_0}{(\overline{\Delta u^2})^{1/2}} \sim \sigma_u \left(\frac{\nu^3}{\epsilon} \right)^{1/4} \frac{1}{(\nu\epsilon)^{1/4}} \quad (4:53)$$

or,

$$\frac{\lambda_T}{\ell_0} \sim \frac{\sigma_u}{(\nu\epsilon)^{1/4}}. \quad (4:54)$$

Use (4:47) in (4:54) to obtain the result for the ratio of the Taylor dissipation length to the Kolmogorov dissipation length:

$$\frac{\lambda_T}{\ell_0} \sim (R_\epsilon')^{1/4}. \quad (4:55)$$

Evidently this ratio is typically quite large in atmospheric work, where the fluctuation Reynolds number may be very large indeed. From the Kolmogorov cascade theory we obtained an estimate for $(\overline{\Delta u^2})^{1/2}$ which is the characteristic size of the small-scale velocity fluctuations associated with the viscous dissipation process. The dissipation could be spread somewhat uniformly throughout the turbulent flow. Recent work dealing with the intermittancy of turbulent flow suggests that in fact the dissipation may occur in slip regions. In such a case the individual velocity changes, when they occur, are larger than the *average* represented by $(\overline{\Delta u^2})^{1/2}$. It is of interest to see how the *average* quantity is related to the typical velocity jump involved in the dissipation process. Call the typical velocity jump Δu ; then utilizing the definition that the dissipation occurs within a fraction α of the total turbulent volume, we see

$$\Delta u \sim \alpha^{-1} (\nu\epsilon)^{1/4}. \quad (4:56)$$

If one adopts the customary view that the turbulence is quite intermittent, one may expect to find small values of α , and consequently, to find that the typical velocity jump, Δu , involved in the dissipation process is relatively large, but relatively infrequently encountered. Many other experiments, both in the ocean and in the atmosphere, have verified the Kolmogorov dependence of the energy spectrum function.

The behavior of the energy spectrum function reflects in a corresponding simplification of the correlation function for small separations of the measuring points. Using dimensional analysis, one obtains a result

$$Q(r,0) = \sigma_u^2 - a_3(\epsilon r)^{2/3}. \quad (4:57)$$

The velocity structure function defined earlier takes the form

$$D_{rr}(r) = C(\epsilon r)^{2/3}, \ell_0 \ll r \ll L_0. \quad (4:58)$$

Similarly the temperature-fluctuation structure function within the same range of r is,

$$D_T(r) = C_T^2(r)^{2/3}, \ell_0 \ll r \ll L_0. \quad (4:59)$$

We are often concerned with changes in the index of refraction. We define a structure function for such changes,

$$D_n(r) = \left\langle [n(\mathbf{r}') - n(\mathbf{r}' + \mathbf{r})]^2 \right\rangle, \quad (4:59a)$$

where we again suppose that small scale effects are homogeneous and isotropic. Then employ the definition, (4:27), the result in (4:23), the Kolmogorov form for the temperature structure function to find,

$$D_n(r) = C_n^2 r^{2/3}, \ell_0 \ll r \ll L_0. \quad (4:60)$$

We can estimate the constants in (4:58) – (4:60) as follows. It is reasonable to extrapolate the Kolmogorov forms to $r \sim L_0$. In that range, the structure functions are on the order of the mean square fluctuations. Hence we have from (4:58) using (4:47)

$$C \sim 1 \tag{4:61}$$

and similarly

$$C_T^2 \sim \sigma_T^2 L_0^{-2/3}; C_n^2 \sim \sigma_n^2 L_0^{-2/3} . \tag{4:62}$$

Of course σ_n and σ_T are related by (4:23). Putting in typical values for the physical quantities at altitudes in the atmospheric boundary layer, e.g., we find

$$C_n^2 = 8.7 \times 10^{-13} C_T^2 , \tag{4:63}$$

with temperature measured in degrees Centigrade.

The energy spectra can also be obtained using the reasoning of this section. For shorter wavelengths, but not so short that we are in the viscous range, the spectrum (4:42) can be written

$$E(k) = b_1 \epsilon^{2/3} k^{-5/3}, \frac{2\pi}{L_0} \ll k \ll \frac{2\pi}{\ell_0} , \tag{4:64}$$

recognizing that E can depend only on ϵ , and k in this range. Here b_n are also dimensionless and of order unity. Similarly, the temperature energy spectrum can be written

$$\Phi_T(k) = b_2 \sigma_T^2 L_0^{-2/3} k^{-5/3} \tag{4:65}$$

these spectra can alternatively be obtained by (effectively) taking the generalized Fourier transform of the correlations.

4.3 Special Characteristics of the Real Atmospheric Boundary Layer

This discussion of turbulence has certain simplifying implicit idealizations. The real atmosphere, and in particular, the atmospheric boundary layer near the earth's surface shows irregular deviations from these idealizations which must be taken into account in the discussion of temperature and velocity fluctuations. For example, we must be concerned with surface roughness. Such roughness can vary from the very slight effects of sand particles at the surface of a flat desert (and small ripples on the surface of the ocean, say, on a very quiet day) through effects due to surface vegetation (and moderate ocean surface waves) all the way up to the large-scale "surface roughness" of large buildings, hills and mountains. Additionally, we must be concerned with the effect of temperature gradients in the atmospheric layers of interest to us. We shall, in the present section, give a brief summary discussion of these real-atmosphere boundary layer effects. For details the reader should see, e.g., Priestley, (1959).

4.3.1 Surface Roughness Effects

Consider first the effect of surface roughness, assuming a neutral atmosphere (one in which air displaced vertically retains the same density as its surroundings). We take the temperature effects into account below. If one examines the turbulent boundary layer over an extremely flat surface (e.g., a machined plate), it is found that there is a region very close to the plate in which the flow is strongly dependent on the fluid kinematic viscosity, ν . This region is called the laminar sublayer. The thickness of this layer is typically less than a centimeter. Furthermore, if one examines flow over a surface with smooth protuberances, it is found that the fluid flow very close to such objects is likewise dependent on the kinematic viscosity, but again the characteristic lengths for such sublayers are extremely small. We shall not be interested here in these viscous sublayers. As one moves some distance away from a real surface, it is found that an important length in the discussion is z_0 , the "roughness length," typically of order (though somewhat smaller than) the root mean square displacement of the surface from its average plane; or more simply, z_0 is of order one-tenth the

dimension of the typical projection on the surface in question. In addition to this length, there is another important parameter with dimension of velocity. In this connection, it is customary to characterize the dynamic effect of the rough surface on the flow through the use of a "friction velocity," usually given the symbol u_* . This friction velocity is, in turn, defined in terms of the average surface stress, τ , through the relation

$$u_* = \left(\frac{\tau}{\rho}\right)^{1/2}; \tag{4:66}$$

u_* is typically between 0.05u and 0.5u for most real surfaces, the lower values corresponding to very smooth surfaces (snow, desert) and the upper values to surfaces with obstructions like trees and houses. The velocity fluctuation near the rough surface is of order u_* .

From experiments, one finds a logarithmic dependence of the mean velocity on the distance from the average plane of the rough surface (out of the viscous sublayer)

$$\bar{u} = \frac{u_*}{k_K} \ln \frac{z}{z_0}, \tag{4:67}$$

where k_K , the so-called Karman constant, is approximately 0.4. The quantity z_0 can be obtained experimentally, and is found to be, as suggested above, of order (but somewhat smaller than) the average height of surface roughness. Typical values for z_0 obtained experimentally for the indicated types of surfaces are given in (T4.1); V_2 is the wind speed.

TABLE 4.1 Roughness Parameters of Various Surfaces

Type of Surface	z_0 (cm.)
Smooth mud flats	0.001
Smooth snow on short grass	0.005
Desert (Pakistan)	0.03
Snow surface, natural prairie	0.10
Mown grass:	
1.5 cm.	0.2
3.0 cm.	0.7
4.5 cm. $\left\{ \begin{array}{l} v_2 = 2 \text{ m sec}^{-1} \\ v_2 = 6-8 \text{ m sec}^{-1} \end{array} \right.$	2.4
	1.7
	9.0
Long grass, 60-70 cm. $\left\{ \begin{array}{l} v_2 = 1.5 \text{ m sec}^{-1} \\ v_2 = 1.5 \text{ m sec}^{-1} \\ v_2 = 6.2 \text{ m sec}^{-1} \end{array} \right.$	6.1
	3.7

As discussed in the preceding review of turbulence theory, a quantity of great interest to us is the outer scale (sometimes called the scale of the energy containing eddies) given the symbol L_0 . For height, z , above the surface, considerably greater than the roughness length, z_0 , we can often assert that the only length scale pertinent to the discussion is the quantity, z , itself. (We rule out the kinematic viscosity ν and z_0 as pertinent parameters for such heights.) It is, in fact, found experimentally that the outer scale is of order the height above the surface in the circumstances described here (though there is some contrary evidence perhaps ascribable to gross topographic characteristics.)

There is another length which at much greater altitude becomes important; it is the scale-height of the atmosphere found in (4:15) to be

$$H = RT_0/g. \tag{4:68}$$

This scale height is approximately 10 km. The temperature fluctuation effects of greatest interest to us in this report usually occur at heights considerably less than this scale height.

4.3.2 Temperature Effects in the Atmospheric Boundary Layer

In this section, we discuss the effects of temperature gradients on the atmospheric boundary layer. A qualitative discussion may be helpful first. Evidently, if the temperature of the air near the surface of the earth is considerably lower than the temperature in regions just above the surface of the earth, one may expect that the temperature gradient has a stabilizing effect. On the other hand, if the average temperature of the atmosphere decreases as we increase height above the surface of the earth, and if it decreases sufficiently rapidly, we can expect the air layer to be unstable – that is, we expect the hotter air near the surface of the earth to rise. As it rises, it would see ambient temperatures progressively lower than the temperature of the parcel of air in question, thus causing, one may assume, extensive thermal instability and mixing. There is, as well, the complicating effect of adiabatic temperature changes. Suppose the atmosphere has a negative temperature gradient, and we consider a parcel of air beginning near the earth’s surface at the higher temperature. As it rises the ambient pressure reduces and, of course, so also does the pressure of the parcel of air. This pressure reduction causes an adiabatic reduction in temperature.

Consider the value of the density of the parcel of air after it rises, compared with the ambient density of the surrounding atmosphere. If the density of the rising air is less than the ambient density, the atmospheric boundary layer will be unstable, and conversely. If the density of the rising parcel of air just equals that of the ambient density, the atmosphere is said to be neutrally stable. In this section, we discuss these simple thermal effects and then consider the effect of wind shear. Evidently if the temperature reduces less rapidly than the critical value of the temperature gradient, the temperature gradient has a stabilizing effect on the atmosphere. If there is a wind shear present, as is typically the case, we must balance the de-stabilizing effects of the wind shear against the stabilizing thermal gradient. In the remainder of the section we put these qualitative ideas in quantitative form, insofar as is possible with the present state of knowledge.

Begin the quantitative discussion by considering the behavior of the atmosphere without wind shear. We have the ideal gas law relating the thermodynamic quantities of interest

$$p^* = \rho^* RT^* \tag{4:69}$$

In this discussion, we shall neglect fluctuations in the functions p^* , ρ^* , T^* , or (with the same effect) consider a time-steady flow.

In addition, we have the hydrostatic pressure relationship between the pressure and density

$$dp^* = -\rho^* g dz \tag{4:70}$$

where dp^* is the change in the ambient pressure within the atmosphere in the distance dz (here z is the vertical height above the earth’s surface). Combining these equations, we obtain for the relative change in pressure within the height, dz ,

$$\frac{dp^*}{p^*} = -\frac{g}{R} \frac{dz}{T^*} \tag{4:71}$$

These equations could be integrated to obtain the atmospheric pressure as a function of the (as yet arbitrary) temperature, function of z . From (4:69) and (4:71) we obtain for the relative change in the ambient density within a height dz

$$(d\rho^*_{amb}/\rho^*) = -(gdz/RT^*) - dT^*/T^* \tag{4:72}$$

Consider a parcel of air rising a distance, dz , sufficiently rapidly so that the thermodynamic process is adiabatic; its pressure is assumed to be always the ambient value of the pressure at its altitude. The relative adiabatic change in density of such a parcel of air is

$$(d\rho_{ad}/\rho^*) = dp^*/\gamma p^* \tag{4:73}$$

For a neutrally stable atmosphere, $d\rho_{ad} = d\rho_{amb}$. From (4:71) and (4:73) we obtain the result

$$\frac{d\bar{\theta}}{dz} \equiv \frac{d\bar{T}}{dz} + \Gamma \begin{cases} < 0, \text{ unstable} \\ > 0, \text{ stable} \end{cases} \tag{4:74}$$

where

$$\Gamma \equiv \frac{g}{R} \left(1 - \frac{1}{\gamma} \right) = - \frac{dT_0}{dz} \quad (4:75)$$

with γ the ratio of the specific heat at constant pressure to the specific heat at constant volume. For the atmosphere we find $\gamma = 1.4$. The quantity Γ is thus about 0.01°C per meter.

Consider now the atmosphere with a wind shear, $d\bar{u}/dz$. It is conventional to form a dimensionless number from the quantity given in (4:74) and this wind shear: the Richardson's number which is,

$$Ri \equiv \frac{g}{\theta} \frac{(d\bar{\theta}/dz)}{(d\bar{u}/dz)^2} \quad (4:76)$$

Evidently, if the Richardson's number is negative, one may suppose that the atmosphere is unstable, for it must be thermally unstable from (4:74) and the wind shear merely decreases its instability. If the Richardson's number is large and positive one may assume that the atmosphere is stable, for it is thermally stable and the wind shear is relatively small. Furthermore, if the Richardson's number is small and positive, one may assume that the atmosphere is unstable due to the relatively large wind shear. The dimensionless number (4:76) is by no means uniquely determined. As a result, a number of other dimensionless quantities have been used in the literature. The chief difficulty arises with the precise way in which one might introduce the wind shear. The discussion is sometimes presented in an energy form, (Priestley, 1959). That is, one might hope to form a dimensionless number from the ratio of the "rate of consumption of energy by the buoyancy forces to the rate of its production by wind shear." It is perfectly possible to write down quantitatively the former rate. However, the rate of production by wind shear cannot be simply described, though of course it does depend in some way upon the wind shear itself.

In connection with the Richardson's number, a scaling length for the height z (Priestley, 1959) has recently become increasingly popular. This length, the Obhukov length, is given by

$$L = -u_*^3 c_p \bar{\rho} / (k_K g H) \quad (4:77)$$

where c_p is the specific heat per mass at constant pressure and H the sensible heat flux given by

$$H = c_p \rho \langle Tu_3 \rangle \quad (4:78)$$

where Tu_3 is the average of the product of the temperature fluctuation and the z component of the velocity fluctuation. In the notation of (4:28) it is $T_3(z, z, 0)$.

Except for extremely small temperature gradients, one may usually neglect Γ so that $\theta \approx T$. If we suppose the typical temperature fluctuation near the ground is of order a defined quantity T_* , (4:78) may be written

$$H = c_p \rho T_* u_* \quad (4:79)$$

In the lower atmosphere, $c_p \rho \approx 1.25$ milliwatt-sec/cm³- $^\circ\text{C}$; temperature fluctuations are typically of order a fraction of a degree Centigrade; $u_* \sim 0.1 \bar{u}$; hence, for 10 m/sec mean wind, we find H is approximately 20-30 milliwatt/cm² under unstable conditions. In stable conditions (for instance on a windless night) H is -2 to -3 mW cm⁻² (Webb, 1969).

Because of the extreme variability of these quantities, it is difficult to give rules for their "typical" behavior. Webb (1969) and others propose the relations for σ_T given in (T4.2) in $^\circ\text{C}$. Here H has units of mW/cm² and z is in cm. The entries for σ_T are obtained as follows: First T_* has been defined to be of order σ_T . The average temperature profile is found experimentally to be best fitted by $z^{-1/3}$. As one might expect, σ_T is found to be proportional to this average in Region 2, giving that entry in (T4.2) when we use the appropriate scaling length (the Obhukov length), L . To find the values for the spectra, first we have from (4:44)

$$\sigma_T^2 = a_8 \int_{k_0}^{\infty} \Phi_T(k) dk \quad (4:80)$$

Table 4.2 Atmospheric Temperature Fluctuation Characteristics

Conditions	Region	Temperature Standard Deviation	Structure Constant	Temperatures, assuming Turbulence–Energy–Cascade
Unstable (e.g., day with wind)	In layer near the ground usually less than a meter; called Region 1	$\sigma_T = 3T_*$	$C_T^2 = 10a_{10} T_*^2 z^{-2/3}$	$\Phi_T = a_4 T_*^2 z^{-2/3} k^{-5/3}$
	z greater than that layer's thickness but less than L ; called Region 2	$\sigma_T = 0.9 T_*(z/ L)^{1/3}$ $= (1.5 \times 10^{-1}) H^{2/3} z^{-1/3}$ [from (4:79)]	$C_T^2 = a_{11} T_*^2 L ^{-2/3} z^{-4/3}$ or $= 10^{-2} a_{12} H^{4/3} z^{-4/3}$	$\Phi_T = a_5 T_*^2 L ^{2/3} z^{-4/3} k^{-5/3}$ or $= a_6 \times 10^{-2} H/z ^{4/3} k^{-5/3}$ [using (4:78)]
Stable (e.g., windless night)	$(z/ L) < 1$	$\sigma_T = 3T_*$	$C_T^2 = 10a_{13} T_*^2 z^{-2/3}$	$\Phi_T = a_7 T_*^2 z^{-2/3} k^{-5/3}$
Stable and Unstable	$(z/ L) > 1$	σ_T exponentially decreasing with scale of order L		

where we suppose $\Phi_T = \beta k^{-5/3}$ the cascade form for the spectrum (4:65). The needed cut-off is of order $2\pi/L_O$ with L_O the outer scale. We further suppose, as discussed earlier, that

$$L_O = a_9 z . \tag{4:81}$$

Substituting and integrating (4:80) we find the entries shown in the last column of the table. Use (4:81) and the relation (4:62) to find the entries for C_T^2 .

A note on the management of the universal constants arising from dimensional analysis may be of interest. Quantities like a_n are ordinarily said to be of ‘‘order unity,’’ i.e., $0.1 < |a_n| < 10$. Clearly the number 10 is not itself universal (most of us have ten fingers). The ‘‘universe’s number’’ is probably something other than 10. In the author’s experience, one seems to arrive at universal constants nearer to one if, whenever possible, numbers are adjusted to be within $\sqrt{10}$ of unity. An example may help. Consider C_T^2 in region 1. Equation (4:62) may be written

$$C_T^2 = a_{14} \sigma_T^2 L_O^{-2/3} \tag{4:82}$$

and use (4:81) and (T4.2)

$$C_T^2 = 9a_{14} a_9^{-2/3} T_*^2 z^{-2/3} . \tag{4:83}$$

The proposal is this: instead of bunching $9a_{14} a_9^{-2/3}$ to form a new constant of order unity, write

$$9a_{14} a_9^{-2/3} = 10 \times 0.9 \times a_{14} a_9^{-2/3} \tag{4:84}$$

$$\equiv 10a_{10} .$$

The surmise is that the universal constant in C_T^2 is of order 10 rather than of order unity. Experiments would have to be performed to check the surmise. (It of course depends on the correctness of the factor 3 in (T4.2).) This method of treating the constants was used throughout this discussion.

The relation for the lower atmosphere

$$C_n^2 = 8.7 \times 10^{-13} C_T^2 \tag{4:85}$$

is useful in connection with Table (4.2)

4.4 References

- Allen, C. W., 1964: *Astrophysical Quantities*. University of London, Athlon Press, London.
- Batchelor, G. K., 1953: *Homogeneous Turbulence*. Cambridge University Press, Cambridge, England.
- Grant, H. L., Stewart, R. W., and Moilliet, A. (1962), Turbulence Spectra. *Journal of Fluid Mechanics*, 12, 241-268.
- Jones, G. W., Jones, J. W., and Monson, K. R. (1969), Interim Analysis of Low Altitude Atmospheric Turbulence. The Boeing Company, LOW-LOCAT Data, Technical Report ASD-TR-69-7, 69.
- Kolmogorov, A. N. (1941), The Local Structure of Turbulence in Incompressible Viscous Flow for Very Large Reynolds Numbers, *C. R. Academy of Sciences, URSS*, 30, 301.
- Lumley, J. L., and Panofsky, H. A. (1964): *The Structure of Atmospheric Turbulence*. John Wiley and Sons, New York.
- Landau, L. D., and Lifshitz, E. M., 1959: Course in Theoretical Physics. *Fluid Mechanics*, 6, English ed., Pergamon Press.
- Meecham, W. C., and Jeng, D. T. (1968), Use of the Wiener-Hermite Expansion for Nearly Normal Turbulence. *Journal of Fluid Mechanics*, 32, 225.
- Monson, K. R., Jones, G. W., and Mielke, R. H., et al. (1969), The Boeing Company, "Low Altitude Atmospheric Turbulence LOW-LOCAT, Phase III Interim Report," Vol. I Data Acquisition and Analysis, Technical Report AFFDL-TR-69-63, 242.
- Obhukov, A. M. (1946), Turbulence in an Atmosphere of Nonhomogeneous Temperature, *Transactions of the Institute for Theoretical Geophysics, USSR* 1, 95.
- Priestly, C. H. B., 1959: *Turbulent Transfer in the Lower Atmosphere*, University of Chicago Press, Chicago.
- Stewart, R. W., and Townsend, A. A. (1951), Similarity and Self-Preservation in Isotropic Turbulence. *Philosophical Transactions A.*, 243, 359.
- Tatarski, V. L., 1959: *The Theory of Fluctuation Effects for Wave Propagation in a Turbulent Atmosphere*. Academy of Science, USSR Press.
- Webb, E. K. (1969), The Temperature of the Lower Atmosphere. Proceedings of REF-EDM Conference held in November 1968, Proceedings published August.

List of Symbols

a_n	dimensionless constants of order unity	Q	longitudinal velocity correlation
A	dimensionless constant in empirical relation for ϵ	Q_1, Q_2	scalar generators for the velocity correlation
b_n	dimensionless constant in the temperature spectrum	Q_{ij}	second order velocity correlation
c_p	fluid specific heat per mass at constant pressure	r	$r'' - r'$
c_v	fluid specific heat per mass at constant volume	r', r''	position vectors for the measurement of correlations
C	velocity structure constant	R	gas constant per mass
C_n	index of refraction structure constant	Re'	fluctuation Reynolds' number
C_T	temperature structure constant	Ri	Richardson's number
D_n	optical index of refraction structure function	S	second order temperature correlation
D_T	temperature structure function	S_1	scalar generator for the temperature correlation
D_{ij}	velocity structure function	S_E	fluid entropy per mass
D_{rr}	longitudinal velocity structure function	t	time; also time delay in correlations
$E(k)$	velocity energy spectrum function	T	averaging time
g	acceleration due to gravity	T	instantaneous temperature fluctuation
G	force per volume exerted on fluid	T_1	scalar generator for the cross correlation of temperature with velocity
H	scale height of the atmosphere	$T_0(z)$	neutrally stable temperature distribution
k	thermal conductivity	T_*	temperature fluctuation near the ground
k	wave vector in Fourier transform	\bar{T}^*	$T_0 + \bar{T}$
k_K	Karman constant	T_i	cross correlation of the temperature with velocity
\hat{k}	unit vector in the vertical (x_3) direction	τ_0	turbulence correlation time
ℓ_0	Kolmogorov dissipation length	\bar{T}	time average of $T^* - T_0$
L	Obhukov length	u	instantaneous velocity fluctuation
L_0	correlation length (outer scale)	\bar{u}	time average of the velocity
M	grid spacing in wind tunnel experiments	u_*	friction velocity
n	optical index of refraction	u_i^*	i th component of total fluid velocity
p	instantaneous pressure fluctuation	u^*	instaneous fluid velocity
$p_0(z)$	neutrally stable pressure distribution	U	flow speed in wind tunnel experiments
\bar{p}	time average of $p^* - p_0$	x_i	i th Cartesian component of position
p^*	total fluid (static) pressure	z_0	surface roughness length
		α	fraction of fluid volume occupied by dissipation regions

ν	μ/ρ_0
Π	Buckingham pi function
ρ^*	total fluid density
$\rho_0(z)$	neutrally stable density distribution
σ_n	RMS value of the index of refraction fluctuation
σ_p	RMS pressure fluctuation
σ_T	RMS temperature fluctuation
σ_u	RMS velocity fluctuation
Φ_T	temperature energy spectrum function
Φ_{ij}	Fourier transform of velocity correlation
x_n	dimensionless variables
γ	c_p/c_v
Γ	gradient of T_0
δV	moving element of fluid volume
Δn	$n - 1$
Δu	typical velocity change in one dissipation region
$\overline{\Delta u^2}$	mean square of velocity change in dissipation regions
ϵ	rate of turbulent kinetic energy decay per unit mass
$\bar{\theta}$	average potential temperature
λ	"eddy size"
λ_T	Taylor dissipation length
μ	fluid viscosity

Chapter 5 ATMOSPHERIC WAVES IN OUTLINE

C. O. Hines

University of Toronto

Toronto 5, Canada

Atmospheric gravity waves are defined, and their properties are deduced for the elementary case of plane waves in an idealized atmosphere. Various complications that arise in the real atmosphere are discussed in turn, and their consequences are described qualitatively.

5.0 Introduction

Atmospheric gravity waves are oscillations of the atmosphere whose nature is strongly affected by the action of gravity. They occur with a wide range of periods and wave-lengths, which is not amenable to precise specification but which we may take as 10 min to 24 hours, and 100 m to 1000 km, respectively, in order to fix ideas temporarily. In a sense, they include a part of the family of tidal oscillations (of period 24, 12, 8, . . . hours) though the term 'gravity wave' is often employed in a more restricted sense which will become apparent. (See Eckart, 1960; Hines, 1960, 1963; Tolstoy, 1963.)

5.1 Governing Equations for the Elementary Case

The elementary properties of gravity waves can be established best by considering the case defined by, (i) the equation of continuous mass conservation,

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \underline{\underline{U}} + \underline{\underline{U}} \cdot \nabla \rho = 0, \quad (5:1)$$

(ii) the force equation,

$$\rho \left(\frac{\partial \underline{\underline{U}}}{\partial t} + [\underline{\underline{U}} \cdot \nabla] \underline{\underline{U}} \right) = -\nabla p + \rho \underline{\underline{g}} \quad (5:2)$$

in which pressure gradients and gravity are the only operative forces causing acceleration, and

(iii) an assumption of adiabaticity in the oscillation:

$$\rho \left(\frac{\partial p}{\partial t} + \underline{\underline{U}} \cdot \nabla p \right) = \gamma p \left(\frac{\partial \rho}{\partial t} + \underline{\underline{U}} \cdot \nabla \rho \right). \quad (5:3)$$

Here ρ , $\underline{\underline{U}}$, p and γ are atmospheric parameters, the density, velocity, pressure and specific-heat ratio, respectively. The latter is taken to be a property of the gas only, invariant even in the presence of the waves ($\gamma = 1.4$ for air). The others are taken to be perturbed (by primed amounts) from certain 'background' values (subscripted 'zero'): $\rho = \rho_0 + \rho'$, $\underline{\underline{U}} = \underline{\underline{U}}_0 + \underline{\underline{U}}'$, $p = p_0 + p'$, and for the present we shall assume that the background wind velocity $\underline{\underline{U}}_0$ vanishes. We may substitute these forms into (5:1) - (5:3), and expand all products that appear. We then find that the equations contain some terms with no perturbation (i.e., primed) parameters, some terms with a single perturbation parameter, some with two, and some with three. We now 'linearize' the equations by ignoring all terms that have more than one perturbation parameter. This is done on the grounds that the perturbations are presumed to be small, and products of perturbation quantities negligible—an assumption that can be justified (or not) *a posteriori* in application to any particular system of waves. The result is:

$$\frac{\partial \rho'}{\partial t} + \rho_0 \nabla \cdot \underline{\underline{U}}' + \underline{\underline{U}}' \cdot \nabla \rho_0 = 0 \quad (5:4)$$

$$\rho_0 \frac{\partial \underline{\underline{U}}'}{\partial t} + \nabla p' - \rho_0' \underline{\underline{g}} = -\nabla p_0 + \rho_0 \underline{\underline{g}} \quad (5:5)$$

$$\rho_0 \left(\frac{\partial p'}{\partial t} + \underline{U}' \cdot \nabla p_0 \right) = \gamma p_0 \left(\frac{\partial \rho'}{\partial t} + \underline{U}' \cdot \nabla \rho_0 \right) . \quad (5:6)$$

We shall be searching for wavelike solutions for the perturbation quantities, and all terms on the left of (5:5) will, therefore, be variable with time. We may conceive of a background state, which contributes the terms on the right of (5:5), which does not vary with time (or, at any rate, varies on a much longer time scale). The equality of the two sides of (5:5) can then be obtained continuously as time varies, only if each side separately vanishes. Thus,

$$\nabla p_0 = \rho_0 \underline{g} \quad (5:7)$$

for example; this is the hydrostatic relation for the background atmosphere. We will further simplify our system by assuming that the background state is one of constant temperature, in which p_0/ρ_0 must be a constant; say

$$p_0/\rho_0 = C^2/\gamma \quad (5:8)$$

where C is a constant which will turn out to be the 'speed of sound' appropriate to the background temperature. If we now choose Cartesian axes (x,y,z) such that the z axis is directed upward, opposite to \underline{g} , (5:7) and (5:8) combine to yield

$$\frac{\partial p_0}{\partial z} = \frac{dp_0}{dz} = -\rho_0 g = -p_0/(C^2/\gamma g) = -p_0/H \quad (5:9)$$

where H, the 'scale height of the atmosphere', is defined by the equation itself, and is seen to be a constant under present assumptions. This permits immediate integration of (5:9). With substitution in (5:8) as well, we find

$$p_0 = p_g \exp \cdot (z/H) ; \rho_0 = \rho_g \exp \cdot (z/H) \quad (5:10)$$

where p_g and ρ_g are certain constant ('ground-level') values of p_0 and ρ_0 respectively.

When (5:7) is inserted in (5:5), and when the equations in (5:10) are employed to express $\nabla \rho_0$ and ∇p_0 in (5:4) and (5:6), the set (5:4) - (5:6) reduces to

$$\frac{\partial \rho'}{\partial t} + \rho_0 \nabla \cdot \underline{U}' - \rho_0 U'_z/H = 0 , \quad (5:11)$$

$$\rho_0 \frac{\partial \underline{U}'}{\partial t} + \nabla p' - \rho' \underline{g} = 0 , \quad (5:12)$$

and

$$\rho_0 \left(\frac{\partial p'}{\partial t} - p_0 U'_z/H \right) = \gamma p_0 \left(\frac{\partial \rho'}{\partial t} - \rho_0 U'_z/H \right) . \quad (5:13)$$

These constitute the governing equations for the wave system in present circumstances, though their subsequent reduction makes use also of the definition (5:8) for C^2 and that in (5:9) for H. (See Eckart, 1960; Hines, 1960; Tolstoy, 1963.)

5.2 Elementary Plane-Wave Solutions

One can proceed in various ways to search for solutions to (5:11) - (5:13), and those solutions may be of various forms (corresponding to plane waves, cylindrical waves, etc.) depending on the purpose at hand. For our purposes, it is best to discuss plane waves that propagate in a direction perpendicular to the y axis, say, for which $\partial/\partial y = U'_y = 0$. It may be confirmed that the following are solutions of the type sought:

$$\frac{\rho'}{\rho_0 R} = \frac{p'}{p_0 P} = \frac{U'_x}{X} = \frac{U'_z}{Z} = A \exp\left(\frac{z}{2H}\right) \exp\left[i\left(\omega t - k_x x - k_z z\right)\right], \quad (5:14)$$

where A is an arbitrary amplitude factor, ω is an arbitrary (radian) frequency, k_x and k_z are the horizontal and vertical components of some wave vector \underline{k} whose direction specifies the direction of phase propagation, and R, P, X and Z are certain constants which, when multiplied into the right-hand side of (5:14) yield the instantaneous amplitude of the partial density perturbation (p'/p_0), the partial pressure perturbation (ρ'/ρ_0), the horizontal wind perturbation U'_x and the vertical wind perturbation U'_z , respectively. All of these perturbations tend to increase in amplitude with increase of height, or equivalently with decrease of gas density, as the factor $\exp(z/2H)$ in (5:14) reveals. The rate of increase offsets the density decrease in such a way as to maintain the flux of wave energy independent of height (provided k_z contains no imaginary part). The amplitude increase is of little consequence in tropospheric applications since z itself does not vary by as much as $2H$ in the troposphere; but it leads to great enhancements in waves that propagate up to heights of 80 km or more, and makes such waves a major component of the dynamical system there.

The use of complex solutions is, of course, an artificial expedient: the physical solutions are the real parts of the expressions given by (5:14), which are themselves automatically solutions of (5:11) - (5:13). These are solutions, rather, under the prescription

$$k_z^2 = \left(\omega_g^2 / \omega^2\right) k_x^2 - k_x^2 - \frac{1}{4H^2} + \frac{\omega^2}{C^2}, \quad (5:15)$$

where

$$\omega_g^2 \equiv (\gamma - 1) g^2 / C^2, \quad (5:16)$$

and under the further prescriptions

$$R \equiv \omega^2 k_z + i(\gamma - 1)g k_x^2 - i\gamma g \omega^2 / 2C^2, \quad (5:17)$$

$$P \equiv \gamma \omega^2 [k_z - i(1 - \gamma/2)g/C^2], \quad (5:18)$$

$$X \equiv \omega k_x C^2 [k_z - i(1 - \gamma/2)g/C^2], \quad (5:19)$$

$$Z \equiv \omega [\omega^2 - k_x^2 C^2]. \quad (5:20)$$

All of these derive simply from insertion of (5:14) in (5:11) - (5:13), except that the definition of any one of (5:17) - (5:20) is an arbitrary choice.

Examination of (5:15) shows that, between the frequency ω_g and the frequency

$$\omega_a \equiv \gamma g / 2C \quad (5:21)$$

(which must be somewhat greater), one cannot have both k_x and k_z real: phases cannot be propagated in any direction, without an exponential attenuation occurring in some other direction. At frequencies greatly exceeding ω_a , (5:15) approximates to the simple sound-wave relation $k_x^2 + k_z^2 = \omega^2/C^2$, and the waves may be termed 'acoustic'; ω_a is the acoustic cut-off frequency. At frequencies below ω_g , the effects of gravity are important and the waves may be termed gravity waves; ω_g is the gravity-wave cut-off frequency.

Important approximations arise in the gravity-wave case when $|k_z| \gg 1/2H$, which holds for much of the observed spectrum in the troposphere. This would require that the 'vertical wavelength' $\lambda_z (=2\pi k_z^{-1})$ be less than 10 km, say. Then

$$k_z^2 \approx \left(\frac{\omega_g^2 - \omega^2}{\omega^2} \right) k_x^2, \quad (5:22)$$

$$R \approx i(\gamma - 1)g k_x^2, \quad (5:23)$$

$$P \approx \gamma\omega^2 k_z, \quad (5:24)$$

$$X \approx \omega k_x k_z C^2, \quad (5:25)$$

$$Z \approx -\omega k_x^2 C^2. \quad (5:26)$$

From (5:22) we see that the angle of ascent of the phases, α , is given by

$$\tan \alpha = k_z/k_x \approx \left(\frac{\omega_g^2 - \omega^2}{\omega^2} \right)^{1/2}. \quad (5:27)$$

If, further, $\omega \ll \omega_g$, then

$$\tan \alpha \approx \pm \omega_g/\omega = \pm \tau/\tau_g \quad (5:28)$$

where τ is the wave period ($=2\pi/\omega$) and $\tau_g = 2\pi/\omega_g \ll \tau$. In these circumstances, $|\tan \alpha|$ is large and the phases propagate nearly vertically, upwards or downwards; phase planes are nearly horizontal; spatial variations occur predominantly in the vertical direction, and only to a lesser extent horizontally. From (5:25) and (5:26), we see that

$$X/Z = -k_z/k_x \quad (5:29)$$

which implies that the wind vector \underline{U} is perpendicular to the wave vector \underline{k} : the winds are shearing winds, nearly horizontal, and reversing in direction vertically every half wavelength.

There is abundant evidence to the effect that gravity waves occur often in the troposphere, though usually not in the form of well-defined isolated systems suitable for analysis. For this reason, and because of observational difficulties in the past, gravity waves have been studied in detail in the troposphere only rarely.

The growth of amplitude with height makes for easier detection in the upper atmosphere, given suitable means of observation. For example, wave-induced winds are revealed at heights of 80 km and above through the distortion of meteor trails and rocket-released vapor trails. The characteristics of these winds are in accordance with the theoretical expectations described by the foregoing formulas, insofar as the latter are thought to be valid. The time scale of the winds (corresponding to τ) is long in comparison with τ_g ($\sim 5-10$ min at the relevant heights); the winds vary rapidly in the vertical with equivalent $\lambda_z \sim 1-30$ km, typically) and only slowly in the horizontal, and they are primarily horizontal. Further, the exponential growth with height predicted in (5:14) is observed in general, with departures that can be explained by complications to be discussed. The characteristics of many moving ionospheric irregularities are similarly in accord with the theoretical characteristics, although for them the 'movement' of the irregularity is often the movement of the phase surfaces (with horizontal and vertical 'trace' speeds ω/k_x and ω/k_z respectively) rather than the movement represented by the wind vector \underline{U}' .

Because of the anisotropy introduced by gravity, wave energy does not generally move in the same direction as do the phases. Instead, it moves with the group velocity whose x and z components are $\partial\omega/\partial k_x$ and $\partial\omega/\partial k_z$, respectively, the differentiations being performed subject to (5:15) being maintained intact. Whenever $|k_z| \gg 1/2H$, as before, the direction of this group velocity is essentially perpendicular to the direction of phase propagation, the horizontal components of the two being in the same direction, but the vertical directions being opposed. Thus, waves whose energy propagates obliquely upwards are subject to phase progression obliquely downward.

One may expect that most sources of observed upper atmospheric gravity waves will originate in the energy-bearing regions of lower levels. Energy would then be propagating upward and phases downward, in the region of observation, and this is in accord with the general run of observations (See Eckart, 1960; Hines, 1960; Tolstoy, 1963.)

5.3 Complications of Nonlinearity

The foregoing simple discussion must break down at some height even with the simplest of atmospheric characteristics assumed, for the exponential growth contained in (5:14) must ultimately lead to amplitudes of such a large magnitude as to invalidate the conversion of the basic equations, (5:1) - (5:3), into the perturbation equations, (5:4) - (5:6). The nonlinear terms (in primed quantities) can no longer be neglected, once they become comparable to the linear terms: e.g., when $U' \cdot \Delta\rho'$ becomes comparable to $\partial\rho'/\partial t$, which is when $\underline{U}' \cdot \underline{k} \approx \omega$ if $|k| \gg 1/2H$. As a first approximation, one might guess that this occurs, say, when $U'_x \sim \omega/k_x$. Observed values for ω/k_x are typically 10-100 m/s, while wind speeds U'_x are of similar amplitude at and immediately above meteor levels, so nonlinear effects must be anticipated. The situation is not quite this serious, because of the previously noted tendency for \underline{U}' and \underline{k} to be mutually perpendicular in a given wave, so that $\underline{U}' \cdot \underline{k}$ tends to be much less than $U'k$. When more than one wave is present, however, the perpendicularity condition no longer obtains: the \underline{U}' of one wave need not be nearly perpendicular to the \underline{k} of a second wave, and the condition for nonlinear effects to be important is more readily met by the wave system. When nonlinearities are important, energy transfer from one wave to another, the formation of new waves from mutual interactions of old waves, and the 'breaking' of waves, must all be anticipated. (See Einaudi, 1970; Hines, 1960.)

5.4 Complications of Instabilities

Quite apart from nonlinear effects in the wave equations, the linear 'perturbation' theory gives rise to temperature gradients (with fractional variations of temperature given by $T'/T_0 = p'/p_0 - \rho'/\rho_0$) which must become superadiabatic at some levels as the exponential growth works its effects. The atmosphere becomes convectively unstable in these conditions, and turbulence may be expected to develop. This turbulence will act to diffuse the energy and momentum from its organized distribution in the wave system to a disorganized distribution, and so converts wave energy first into turbulent energy and then into heat. The situation may be more serious, in fact, for turbulence might develop even without the vertical temperature gradient becoming superadiabatic: a 'slantwise' superadiabaticity is always possible, leading always to the possibility of instability, and wind shears in the wave system provide a further source of energy for driving turbulence. These effects act to leech energy from the wave systems to heights of 100 km or so, whereupon molecular effects appear to become more significant as a dissipative mechanism. (See Hines, 1960; 1971; Hodges, 1967.)

5.5 Complications of Molecular Dissipation

Viscosity and thermal conduction are inherent properties of the atmosphere, and their importance increases with height as the mean free path increases. At some height, they must come to be significant in the propagation of waves, if the waves have not already been removed by nonlinear or instability (or other) processes.

Properly, one should take molecular viscosity into account by adding a force density $\mu(\nabla^2 \mathbf{U} + \nabla(\nabla \cdot \mathbf{U})) / 3$ on the right-hand side of (5:2) and also a corresponding primed quantity on the right-hand side of (5:12), μ being the molecular viscosity. Analytic solutions can no longer be obtained if one does this, however, for the form of the equation is now altered; whereas other terms in (5:12) have the height variation $\exp(-z/2H)$, this term has the height variation $\exp(+z/2H)$ since μ is virtually height-independent. No adjustment to the form of solution can be found that would permit accurate analytic corrections of a simple type. An elementary calculation, which is supported by more detailed analyses, does however permit an evaluation to be made of the circumstances under which viscosity becomes serious: we take it to be serious when the simple solution (5:14), on insertion into the viscous force density term, yields a magnitude for that term that is comparable to the inertial force density. This yields the order-of-magnitude result (for $|k_z| \gg 1/2H$) $\omega \rho_0 U' \approx \mu k^2 U'$, or

$$\omega/k^2 \approx \mu/\rho_0 = \eta_M \quad (5:30)$$

where η_M is the molecular kinematic viscosity. (If the atmosphere is turbulent, the eddy viscosity η_E replaces η_M in this criterion for important alterations to the wave system.) Equation (5:30) may be combined with (5:15), with $k_z \gg 1/2H$ assumed, to show that the largest k_z that is consistent with (5:30) is to be found at $\omega^2 = \omega_g^2/3$, when $k_z^2 \approx 0.3 \omega_g/\eta_M$. (The precise numerical factor here depends upon the precise conditions assumed for viscosity becoming "important".) The corresponding vertical wavelengths, $\lambda_z \approx 2\pi\sqrt{3\eta_M/\omega_g}$, are of the order 1 km at a height of 100 km and increase upward. These wavelengths are indeed found to be 'cut-off' wavelengths in the spectrum of observed zig-zag profiles of the wind at these heights, as revealed by rocket-released vapor trails; the height separations of successive "zigs" or "zags" almost invariably exceed these λ_z 's.

Thermal conduction may be introduced properly only by rejecting the adiabatic (5:3), and inserting in its place the perfect-gas law plus an equation for heat transfer (including, in principle at least, a term that represents heat generation from viscous losses). Again, no analytic solutions are available in these circumstances. Because of the intimate relationship between thermal conduction and viscosity, however, both being a consequence of molecular transport processes, one might anticipate that thermal conduction would become important only when viscosity becomes important, and that its role as a dissipative mechanism would be quite analogous. These suspicions are borne out by detailed analysis. (See Hines, 1960; Midgley and Liemohn, 1966; Pitteway and Hines, 1963.)

5.6 Complications of Temperature Structure

The background temperature of the real atmosphere is not a constant, but rather varies substantially on a scale of a few kilometers in the vertical. One can infer the qualitative consequences of this by considering an elementary case in which one isothermal half-space is superimposed above another, of different temperature, there being a wave incident upon the interface from below. Certain interfacial conditions must be met: there must be no pressure discontinuity, and the vertical displacement must be the same on both sides of the interface, right at the interface itself. These conditions can be met continuously in time and in space along the interface only if a 'transmitted' wave occurs in the upper half-space, with ω and k_x unchanged from the values obtaining in the incident wave. But (5:15) then implies that k_z must be different in the upper half-space to make up for the differences in H and C : the transmitted wave is 'refracted', to propagate at some appropriate new inclination.

Continuity of pressure and vertical displacement demand more, for they impose two conditions on the amplitudes of the wave systems. These conditions cannot be met simply by an appropriate change of A in (5:14), on going from the lower half-space to the upper, for that provides only a single degree of freedom (namely, the ratio of the two A 's). Instead, it is necessary to add yet another wave, a 'reflected' wave, propagating its energy away from the interface in the lower half-space.

The actual temperature profiles, being smoothly varying, give rise to continuous processes of refraction and internal reflection, with resultant interference effects that can be complex but must be treated in some detail to be treated at all adequately. About the only general statement that can be made is this: a given temperature structure will act to transmit or reflect waves with various ω 's and k_x 's in varying degree, and so will act as a selective filter between sources in one region of the atmosphere and effects in another.

Beyond this, perhaps the simplest general point to be made is that the k_z deduced from (5:15) may be real for some heights and imaginary for others. The waves are 'evanescent' when k_z is imaginary, and there is a strong tendency for the vertical flow of energy to be inhibited in such regions. Whether on this account or on others, strong reflection may occur and the wave energy may be ducted between two heights of strong reflection (one of which may be the ground). Some energy may, nevertheless, escape above the duct region, and there the exponential growth with height could render its effects important despite its intrinsic weakness. This ducting effect may well account for an observed characteristic of many traveling ionospheric disturbances, their maintenance of strength over long distances of propagation: the main reservoir of energy may well lie in a duct below 100-150 km, and may be maintained with little loss despite the small leakage upward, while the little that does leak upward may be observed in the ionosphere and may appear to sustain no loss because it is continually being replenished from below. (See Friedman, 1966; Hines, 1960.)

5.7 Complications of Wind Structure

The real atmosphere supports background winds, which have been ignored so far in the present analysis except for their definition as \underline{U}_0 . Examination of (5:1) - (5:3) will reveal that these winds would alter the perturbation equations (5:11) - (5:13) only by the addition of a $(\underline{U}_0 \cdot \nabla)$ operation in conjunction with each $\partial/\partial t$ operation. In application to waves of the form (5:14), this would result in each appearance of an ω being replaced by the appearance of an 'intrinsic frequency':

$$\tilde{\omega} \equiv \omega - \underline{U}_0 \cdot (\underline{k} + i \hat{z}/2H) \tag{5:31}$$

upon insertion of (5:14) into (5:11) - (5:13), where \hat{z} is a unit vector in the z direction. If, as is usually assumed and as is indeed the case in practice, the vertical component of \underline{U}_0 may be taken to vanish, (5:31) reduces to the elementary form of a Doppler-shifted frequency,

$$\tilde{\omega} = \omega - \underline{U}_0 \cdot \underline{k} = \omega - U_{0x} k_x \tag{5:32}$$

It is this frequency that must now be entered into formulas, (5:15), (5:17) - (5:20), (5:22) - (5:28).

As in the case of temperature structure, wind structure imposes refraction and reflection, and opens the possibility of wave ducts. It serves, more generally, as a selective filter once again. The filter in this case has the further interesting property of being anisotropic: since the value of $\tilde{\omega}$ specified by (5:31) is dependent on the component of \underline{U}_0 in the direction of wave propagation, the filtering effect will vary with that direction. In this respect, winds add a dimension to the filtering effect of temperature structure.

From (5:32) it will be apparent that $\tilde{\omega}$ reduces to zero whenever the horizontal trace speed ω/k_x matches the background wind speed in the direction of propagation, U_{0x} . Any level at which this occurs in the atmosphere is termed a 'critical level', and very important processes occur at such levels, not all of them understood. It seems likely that, for all practical purposes, the energy of a wave is completely destroyed when it reaches such a level if the atmosphere at that level is, itself, sufficiently stable; the wave energy is converted to heat and to kinetic energy of the background flow. (See Bretherton, 1966; Hines and Reddy, 1967; Pierce, 1965.)

5.8 Complications of Earth Curvature and Rotation; Tides

The analysis to this point has taken gravity to lie in a single Cartesian direction, thereby ignoring the sphericity of the earth, and has made no provision for the inclusion of Coriolis and centrifugal forces associated with the earth's rotation. We now consider the complications introduced by these factors and so go beyond the topic of 'gravity waves' as defined in its more restrictive sense by the subject matter of the preceding sections. More specifically, we touch upon the tidal oscillations of the atmosphere, in particular, the tides with periods 12 hours and 24 hours.

At such long periods, the horizontal component of the Coriolis force density, $-2\rho\mathbf{U}\times\boldsymbol{\Omega}$, must be included on the right-hand side of (5.2) if the coordinate system is taken to rotate with the earth. The vertical component of this force, and the centrifugal force, are both ignored in tidal theory; and indeed, vertical accelerations are taken to vanish, which is consistent with the generally small value of ω that characterizes tides and with the small vertical velocities (relative to horizontal velocities) that might be inferred from a simple extrapolation of gravity-wave theory to such small ω 's.

The sphericity of the earth is handled conveniently with the aid of spherical coordinates, r, θ, ϕ , say, but it poses a further problem when contrasted with the circumstances of elementary gravity waves: the solutions must be compatible with the periodicity of 2π in the longitudinal coordinate ϕ , and they must be continuous at the poles $\theta=0, \pi$. The search for valid solutions then becomes a problem in eigenfunctions and eigenvalues.

The eigenfunctions contain longitudinal and latitudinal factors. The former are of the simple form $\exp(im\phi)$ where $m=1, 2, \dots$ corresponding to $1, 2, \dots$ wavelengths round the equator, and so to tides of period 24, 12 \dots hours. The latter are known as Hough functions, which we may denote as $\theta_{m,n}$. The subscript m is identical to the m of the associated longitudinal function. The subscript n is a measure of the degree of complication in the north-south structure of the particular Hough function; for $m \geq 2$, $n-m$ equals the number of nodal surfaces that lie between the north and south poles. The Hough functions, multiplied by their associated $\exp(im\phi)$ functions, are intimately related to spherical harmonics and may be expanded in terms of them. For each eigenfunction there is a corresponding eigenvalue, denoted $h_{m,n}$ which happens to have the dimensions of a length; it is known as the 'depth of the equivalent ocean'.

Given a certain 'm,n' mode of horizontal structure, the vertical structure of the tidal oscillation is associated with a vertical wave number k_z given by

$$k_z^2 = \frac{1}{Hh} \left[\frac{\gamma - 1}{\gamma} + \frac{dH}{dr} \right] - \frac{1}{4H^2} \quad (5.33)$$

This is analogous to (5.15) subject to certain caveats: the ω^2/C^2 term of (5.15) disappears because, with $\omega \ll \omega_a$, it is much smaller in magnitude than the $1/4H^2$ term; similarly, the term $-k_x^2$ is ignored relative to the $(\omega_g^2/\omega^2)k_x^2$ term; the $(\gamma-1)$ of ω_g^2 , as defined by (5.16), is replaced by $(\gamma-1) + \gamma dH/dr$, which serves to take some account of height variations of temperature; and the parameter h now plays the role previously played by ω^2/gk_x^2 (or, as one might say, the horizontal trace speed ω/k_x has been replaced by \sqrt{gh} , which is well known as the speed of a long wave in an ocean of depth h). (See Hough, 1897, 1898; Siebert, 1961; Wilkes, 1949.)

5.9 The Semidiurnal Tide

The semidiurnal tide consists of a superposition of oscillations of the '2,n' type. It is excited primarily by the absorption of solar radiation by ozone, at heights of 25-55 km, say. The latitudinal distribution of the resultant heating most nearly resembles the $\theta_{2,2}$ function, and it leads then to a strong component of the '2,2' type. For this component, $h_{2,2}=7.9$ km. It happens that the k_z which now results from (5.33) is imaginary in the mesosphere, at heights of 60-85 km roughly, and the energy of the '2,2' mode is strongly reflected there. Some transmission does occur, of a strength sufficient to render this mode an important one at meteor heights and above, but it has not the dominance at those heights that it enjoys at lower levels.

The '2,4' and '2,6' modes are excited less strongly, but the h 's for them are smaller and the k_z 's nowhere become imaginary. Their energy reaches meteor heights and overlying levels without suffering serious reflection, in consequence, and they themselves are relatively important in observations made at such heights. (See Butler and Small, 1963; Hines, 1963, 1968.)

5.10 The Diurnal Tide

The diurnal tide consists of a similar superposition. The '1,1' mode behaves in quite a different fashion from the '2,2' mode, however, for it is confined primarily to latitudes below 50° and, within those latitudes, contains a node on each side of the equator. For it, $h_{1,1}=0.7$ km and this leads to vertical wavelengths ($=2\pi/k_z$) of about 25 km. Thus the '1,1' mode reverses in sign, both in latitude and in height, within dimensions over which the ozone heating remains strong and of a single sign; it is excited relatively inefficiently in consequence. In fact, absorption of water vapor low in the atmosphere, over a height range small in comparison with the vertical wavelength, appears to be the dominant mechanism of excitation for this mode.

Despite its relatively modest means of excitation, the '1,1' mode is free to propagate through the mesosphere without serious reflection; it reaches meteor heights in considerable strength, giving rise to winds of the order 40 m/s in its low-latitude belt, and then rapidly dissipating itself through turbulent and molecular loss processes.

A second family of diurnal tidal modes exists for which h is intrinsically negative; it is generally denoted by means of negative values for n . These modes of oscillation are of a different category from those so far discussed, being not of the general 'gravity wave' class. By virtue of their negative h 's, their k_z 's (which are still given by (5.33)) are intrinsically imaginary: energy deposited at some altitude in their excitation tends to remain at that altitude, without suffering dispersal via vertical propagation. The '1,-1' mode extends from pole to pole without a node, and closely matches the latitudinal variation of the ozone heating function. It is excited strongly in consequence, but appears strong only in the vicinity of the ozone layer. Higher order 'negative' modes are also excited, primarily at higher latitudes (thus complementing the low-latitude bias of the 'positive' modes); but again, their energy remains near the level of excitation. The negative modes are of little concern at meteor heights, but local sources at somewhat greater heights may make them relevant there. Indeed, there is some evidence that the '1,-1' mode, excited near heights of 100-130 km, may be the dominant source of the dynamo winds that lead to diurnal variations of ionospheric current systems and so of the ground-level magnetic variations that these systems produce. The '1,-1' mode is also excited in the troposphere by insolation absorption by water vapor and gives rise to much of the tidal oscillation that is observable at ground level. (See Hines, 1968; Kato, 1966; Lindzen, 1967.)

5.11 REFERENCES

- Bretherton, F. P. (1966) The propagation of groups of internal gravity waves in shear flow, *Quart. J. Roy. Met. Soc.* 92, 466.
- Butler, S. T., and D. A. Small (1963) The excitation of atmospheric oscillations, *Proc. Roy. Soc.* A274, 91.
- Eckart, C., 1960: *Hydrodynamics of Oceans and Atmospheres* Pergamon Press, New York, N.Y.
- Einaudi, F. (1970) Shock formation in acoustic-gravity waves, *J. Geophys. Res.* 75, 193.
- Friedman, J. P. (1966) Propagation of internal gravity waves in a thermally stratified atmosphere, *J. Geophys. Res.* 71, 1033.
- Hines, C. O. (1960) Internal atmospheric gravity waves at ionospheric heights, *Can. J. Phys.* 38, 1441.
- Hines, C. O. (1963) The upper atmosphere in motion, *Quart. J. Roy. Met. Soc.* 89, 1.
- Hines, C. O. (1968) Tidal oscillations, shorter-period gravity waves and shear waves, *Met. Mono.* 9, 114.
- Hines, C. O. (1971) Generalizations of the Richardson criterion for the onset of atmospheric turbulence, *Quart. J. Roy. Met. Soc.* 97, 429.

- Hines, C. O. and C. A. Reddy (1967) On the propagation of atmospheric gravity waves through regions of wind shear, *Can. J. Phys.* *72*, 1015.
- Hodges, R. R. (1967) Generation of turbulence in the upper atmosphere by internal gravity waves, *J. Geophys. Res.* *72*, 3455.
- Hough, S. S. (1897, 1898) On the application of harmonic analysis to the dynamical theory of the tides, *Phil. Trans. Roy. Soc.* *A189*, 201, and *A191*, 139.
- Kato, S. (1966) Diurnal atmospheric oscillation, *J. Geophys. Res.* *71*, 3201 and 3211.
- Lindzen, R. S. (1967) Thermally driven diurnal tide in the atmosphere, *Quart. J. Roy. Met. Soc.* *93*, 18.
- Midgley, J. E., and H. B. Liemohn (1966) Gravity waves in a realistic atmosphere, *J. Geophys. Res.* *71*, 3729.
- Pierce, A. D. (1965) Propagation of acoustic-gravity waves in a temperature and wind-stratified atmosphere, *J. Acoust. Soc. Am.* *37*, 218.
- Pitteway, M. L. V., and C. O. Hines (1963) The viscous damping of atmospheric gravity waves, *Can. J. Phys.*, *41*, 1935.
- Siebert, M. (1961) Atmospheric tides, *Adv. in Geophys.* *7*, 105.
- Tolstoy, I. (1963) The theory of waves in stratified fluids including the effects of gravity and rotation, *Rev. Mod. Phys.* *35*, 207.
- Wilkes, M. V., 1949: *Oscillations of the Earth's Atmosphere*. Cambridge Univ. Press, Cambridge, England.

List of Symbols

A	wave amplitude factor; a constant	Z	constant relating A to the wave-associated vertical velocity perturbation
C	“speed of sound” ($\equiv [\gamma p_0 / \rho_0]^{1/2}$)	α	$\tan^{-1} (k_z / k_x)$
g	acceleration of gravity (when used as a subscript on p and ρ , it refers to ground level values of these quantities)	γ	atmospheric specific heat ratio
H	atmospheric scale height	η_M	molecular kinetic viscosity ($\equiv \mu / \rho_0$)
$h_{m,n}$	equivalent depth corresponding to $\theta_{m,n}$	θ	colatitude
\tilde{k}	angular wave vector (horizontal and vertical components k_x and k_z)	$\theta_{m,n}$	Hough function; the m is identical to the m of the associated longitudinal function. m-n is the number of nodal surfaces between the poles.
P	constant relating A to the wave-associated pressure perturbation	λ_z	vertical wavelength ($\equiv 2\pi / k_z$)
p	pressure	μ	molecular viscosity
R	constant relating A to the wave-associated density perturbation	ρ	density
r	radial coordinate, origin at the center of the earth	τ	wave period ($\equiv 2\pi / \omega$)
t	time	τ_g	Brunt period ($\equiv 2\pi / \omega_g$)
\tilde{U}	vector wind velocity	ϕ	longitude
X	constant relating A to the wave-associated horizontal velocity perturbation	$\tilde{\Omega}$	earth’s angular rotation vector
x	horizontal coordinate, taken to be in the direction of wave propagation	ω	angular wave frequency
y	horizontal coordinate, taken to be propagation to the direction of wave	ω_a	acoustic cutoff frequency ($\equiv \gamma g / 2C$)
z	vertical coordinate	ω_g	gravity-wave cutoff frequency ($\equiv (\gamma - 1)^{1/2} g / C$)
\hat{z}	a unit vector in the vertical direction	$\tilde{\omega}$	“intrinsic” angular wave frequency ($\equiv \omega - \tilde{U}_0 \cdot \tilde{k}$)

Throughout the text a zero subscript denotes an unperturbed or “background” value of the subscripted quantity; a prime superscript denotes a wave-associated perturbation.

Chapter 6 THE ATMOSPHERIC BOUNDARY LAYER

J. A. Businger
University of Washington
Seattle, Washington

After a condensed introduction of the basic equations governing the boundary layer a description is given of the flux profile relations in the surface layer, the structure of the steady state neutral boundary layer, free convection, the critical Ri number, secondary flows such as rolls, plumes and dustdevils and the effects of change of terrain.

6.1 Introduction

The atmospheric boundary layer may be defined as that portion of the atmosphere where the direct effect of the surface is noticeable. In most cases, particularly those where the momentum and heat exchange at the surface are significant this definition does not give any serious problems. When we have a strong wind, significant momentum is being transferred to the surface and a well developed “*wind spiral*” can be identified which describes the transition from the wind near the surface to the “*geostrophic*” wind at a height of about one kilometer. When the wind is geostrophic a balance is reached between the pressure gradient force and the Coriolis force. Thus as soon as this is the case no more vertical transport of momentum takes place and therefore we are at a level beyond which the effect of the surface is negligible. Similarly when there is an upward heat flux at the surface the boundary layer is well defined. In this case the effect of the wind stress or momentum flux is usually secondary to the effect of the heat flux. The boundary layer is now defined as the height where the “*inversion*” occurs. The warm air near the surface rises to a height where the ambient “*potential temperature*” is the same or higher than the potential temperature of the rising air. At approximately this point an inversion is formed which nicely defines the top of the boundary layer. These two cases of the boundary layer will be discussed in detail in the following sections.

However, the boundary layer is not always as well defined. The following major exceptions to our definition may be noted.

1. *The equatorial boundary layer.* The Coriolis force is negligible and no equilibrium with the pressure force is obtained. It is unclear what the height of the boundary layer is because a geostrophic wind is not found. Fortunately in many cases a well defined inversion layer exists, the so-called tropical inversion which in itself is not fully understood but which suitably may be considered the top of the boundary layer.
2. *The stable boundary layer.* When the heat flux is downward the boundary layer is stable and the turbulent exchanges are suppressed and consequently the coupling between adjacent layers is weakened. The stable boundary layer therefore is often complex including inertial waves, gravity waves, patches of turbulence, etc. The coupling with the surface is weak and it is very difficult to tell where its influence stops. An extreme condition of this kind is often found in the arctic in winter. The winds may be very light and the temperature may not vary appreciably from the surface up to well into the stratosphere. Not only the boundary layer seems to have disappeared but the entire *troposphere* as well. In fact, in a broader more indirect sense the troposphere like the boundary layer reflects the effect of the surface on the atmosphere. The tropopause may be compared with the inversion layer above the unstable boundary layer.
3. The boundary layer is also poorly defined in case the atmosphere is unstable over a deep layer so that convective clouds and showers develop. In this case, much and sometimes all of the troposphere seems to be absorbed by the boundary layer. However, the interaction of the atmosphere with the surface is both direct and indirect, and an understanding of meso scale convective processes and cloud dynamics is required. In this case the boundary layer is often defined as the layer extending from the surface to the base of the clouds. This concept is quite useful when the cloud development is limited and no precipitation occurs.

This chapter deals mainly with what we know of the boundary layer. Therefore seemingly an out of proportion large fraction of it deals with steady state horizontally uniform conditions. Occasionally a

reluctant excursion has been made in the area of non-steady state or non horizontal uniformity. This aspect of boundary layer research is essentially still wide open and although some observations exist, their interpretation is largely speculative.

6.2 Governing Equations

The reading of this section can best be started with 6.2.3, because the results of this subsection are used in 6.2.1 and 6.2.2. The reasons why the equations of motion and heat conduction equations have been mentioned first is to give proper emphasis on the major equations. Subsection 6.2.3 may be considered an appendix to 6.2.1 and 6.2.2. Furthermore it has been assumed that the reader is familiar with the technique of splitting the variables into mean and fluctuating quantities and consequently taking averages of products etc., see e.g., (6:17), and Chapter 4.

The following discussion is not fully rigorous. The major assumptions are discussed. Terms that are neglected without discussion are assumed to be unimportant. For a more detailed discussion of the basic equation see Lumley and Panofsky (1965).

6.2.1 Equations of Motion

Neglecting the viscous terms the equations of motion for the atmosphere may be written in the form

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial}{\partial x_j} (\rho u_i u_j) = - \frac{\partial p}{\partial x_i} - 2 \rho \vec{\Omega} \times \vec{v} - \delta_{i3} \rho g \quad (6:1)$$

In the boundary layer often the assumption of horizontal uniformity is made, and after taking averages the equations (6:1) may be written in component form

$$\frac{\partial \bar{\rho u}}{\partial t} + \frac{\partial}{\partial z} \bar{\rho u w} = - \frac{\partial \bar{p}}{\partial x} + f \bar{\rho v} \quad (6:2)$$

$$\frac{\partial \bar{\rho v}}{\partial t} + \frac{\partial}{\partial z} \bar{\rho v w} = - \frac{\partial \bar{p}}{\partial y} - f \bar{\rho u} \quad (6:3)$$

$$\frac{\partial \bar{\rho w}}{\partial t} + \frac{\partial}{\partial z} \bar{\rho w^2} = - \frac{\partial \bar{p}}{\partial z} - \bar{\rho} g \quad (6:4)$$

For all practical purposes the terms on the left hand side of equation (6:4) are negligible in comparison to those on the right hand side. The hydrostatic approximation

$$- \frac{\partial \bar{p}}{\partial z} = \bar{\rho} g \quad (6:5)$$

is therefore valid.

The second term on the left hand side of (6:2) and (6:3) expresses the turbulent vertical momentum transport; this term is the reason this set of equations is not closed. To make the equations tractable an eddy viscosity K_m in analogy with molecular kinematic viscosity, ν , is introduced so that

$$\bar{\rho} \overline{u' w'} \equiv - \bar{\rho} K_m \frac{\partial \bar{u}}{\partial z} \quad (6:6)$$

If $\overline{\rho u w}$ is expanded into mean and turbulent components, and after using $\overline{\rho w} = 0$, (6:17), it is easy to show that

$$\overline{\rho u w} \cong \overline{\rho} \overline{u' w'}$$

is a very good approximation.

By making reasonable assumptions about the functional relationship of K_m with height it is possible to integrate (6:6), see 6.3.1 and 6.4.2.

Furthermore $\overline{\rho' u'} \ll \overline{\rho u}$ and $\overline{\rho' v'} \ll \overline{\rho v}$. Equations (6:2) and (6:3) may therefore be simplified to

$$\frac{\partial \bar{u}}{\partial t} = f(\bar{v} - v_g) + \frac{\partial}{\partial z} K_m \frac{\partial \bar{u}}{\partial z} \tag{6:7}$$

$$\frac{\partial \bar{v}}{\partial t} = -f(\bar{u} - u_g) + \frac{\partial}{\partial z} K_m \frac{\partial \bar{v}}{\partial z} \tag{6:8}$$

where $\partial p / \partial x = -f v_g$ and $\partial p / \partial y = f u_g$; u_g and v_g are the components of the geostrophic wind.

Equations (6:7) and (6:8) may be combined into one equation by introducing the variable,

$$W \equiv (u + iv) e^{ift} ,$$

with the result†

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial z} K_m \frac{\partial W}{\partial z} + if V_g e^{ift} \tag{6:9}$$

where $V_g \equiv i(\rho f)^{-1}(\partial p / \partial x + i \partial p / \partial y)$. Equation (6:9) is the diffusion equation with a source term $P(z, t) = if V_g e^{ift}$, and a variable diffusion coefficient $K_m(z, t)$. This equation has a general solution when $K_m = \text{const}$, which has been discussed by Sobolev (1964) and applied by Ching and Businger (1968). This equation is particularly useful for analytical studies of the boundary layer.

6.2.2 The heat conduction (enthalpy) equation

The first law of thermodynamics applied to a parcel of air in the atmosphere leads to the following expression for enthalpy ($c_p T$), which is commonly known as the heat conduction equation

$$\frac{\partial}{\partial t} (c_p \rho T) + \frac{\partial (c_p \rho T u_i)}{\partial x_i} = R_n \tag{6:10}$$

where R_n is the rate of radiative absorption per unit volume. The term expressing molecular conduction has been neglected. Again if we assume horizontal uniformity and take the average we obtain

$$c_p \bar{\rho} \frac{\partial \bar{T}}{\partial t} + c_p \bar{\rho} \frac{\partial}{\partial z} (\overline{w' T'}) = -\frac{\partial}{\partial z} F_n \tag{6:11}$$

The term $\overline{w' T'}$ which represents the eddy correlation between temperature and vertical velocity may be simplified by the introduction of an eddy heat transfer coefficient K_h (eddy thermal diffusivity) so that

$$\overline{w' T'} \equiv -K_h \frac{\partial \bar{\theta}}{\partial z} \tag{6:12}$$

The eddy enthalpy flux or eddy heat flux F_h is defined by $F_h = c_p \bar{\rho} \overline{wT}$, therefore also

$$F_h = -c_p \bar{\rho} K_h \partial \bar{\theta} / \partial z. \quad (6:13)$$

Equation (6:11) may be integrated over the height h_0 at which the heat flux is zero which results in

$$F_h(0) = \int_0^{h_0} c_p \bar{\rho} \frac{\partial T}{\partial t} dz + F_n(0) - F_n(h_0) \quad (6:14)$$

This equation and some of its consequences will be discussed in (6.5.2).

6.2.3 Equation of continuity and equation of state

These two equations have already been used implicitly in various forms in formulating the previous equations. However, for handy reference we shall mention them here explicitly.

The continuity equation which expresses the conservation of mass may be written as

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i} (\rho u_i) = 0 \quad (6:15a)$$

or

$$\frac{\partial u_i}{\partial x_i} = -\frac{1}{\rho} \frac{d\rho}{dt} \quad (6:15b)$$

The right hand side reflects the relative change in density of a parcel of air following the flow, which is much smaller than the local changes in velocity. Therefore a good approximation is

$$\frac{\partial u_i}{\partial x_i} = 0 \quad (6:16)$$

This is the continuity equation for an incompressible flow. It may be used in conjunction with (6:7), (6:8), (6:9), (6:11), and (6:12) but should be avoided in manipulating (6:2), (6:3), (6:4) and (6:10).

If we assume horizontal uniformity and steady state and average (6:15a) we find

$$\frac{\partial}{\partial z} \bar{\rho} \bar{w} = 0$$

in addition, since $w = 0$ for $z = 0$ we must have $\bar{\rho} \bar{w} = 0$ or $\bar{w} = -\frac{1}{\rho} \overline{wT}$

$$(6:17)$$

a relation that has been used to go from equations (6:2) and (6:3) to (6:7) and (6:8) respectively, and from (6:10) to (6:11).

The equation of state for an ideal gas which applies to the atmosphere is

$$p = R\rho T \quad (6:18)$$

Because pressure fluctuations are relatively much smaller than temperature and density fluctuations, i.e.,

$$\frac{p'}{p} \ll \frac{\rho'}{\rho} \text{ and } \frac{T'}{T}$$

we have as a good approximation that

$$\frac{\rho'}{\rho} = -\frac{T'}{T} \tag{6:19}$$

a very useful simple relation. Equations (6:10) and (6:19) are often referred to as the Boussinesq approximation.

6.2.4 The Turbulent Energy Equation

A basic problem in the study of turbulence is the generation and dissipation of turbulent kinetic energy. An expression for the kinetic energy balance may be obtained by averaging each component of (6:1) and then subtracting these average equations from (6:1). The result is that expressions are found for the turbulent components of the velocity (u' , v' , w'). Scalar multiplication of each of these equations by the corresponding turbulent components of velocity and averaging yields three equations for the three components of kinetic turbulent energy. Summing these last three equations results in the turbulent kinetic energy equation. For details of this equation the reader is referred to chapter 4. In this chapter we are primarily interested in investigating the case of quasi steady state and horizontal uniformity. The equation then simplifies to

$$\frac{\partial \bar{e}}{\partial t} = -\overline{u'w'} \frac{\partial \bar{u}}{\partial z} + \frac{g}{T} \overline{w'T'} - \frac{\partial}{\partial z} \overline{w'(e+p'/\rho)} - \epsilon \tag{6:20}$$

The first term on the right hand side represents the shear production of turbulence; it expresses a transfer from the kinetic energy of the mean flow to turbulent flow. The second term expresses the rate of work done by the buoyant forces. This term is positive in the unstable boundary layer (upward heat flux) and negative in the stable boundary layer. The third term is a combination of divergence of turbulent kinetic energy and rate of work done by the pressure fluctuations. This term does not generate or destroy turbulent kinetic energy but redistributes it. The integral of this term over the entire boundary layer vanishes. The last term represents the viscous dissipation in short hand notation.

Although this equation is simply derived from the equations of motion it provides valuable physical insight and will be referred to in following sections.

6.3 The Neutral Atmospheric Boundary Layer

Although the neutral boundary layer is a rather special case which rarely occurs in reality and in itself is rather uninteresting, we believe that its structure is better understood than the non-neutral cases and therefore justifies a detailed discussion seemingly out of proportion with its importance.

6.3.1 Structure of the Surface-layer

The surface layer is usually defined as the lowest part of the boundary layer where the fluxes may be considered independent of height. This is approximately the case in the lowest 10 to 20 m. In this layer the interaction with the surface is very strong and adjustment to the surface conditions is relatively rapid. A quasi steady state may therefore often be assumed.

The momentum which is transferred down through the layer to the surface exerts a drag force per unit area on the surface which is called the surface stress τ . According to Newton's Third Law we have

$$\rho \overline{u'w'} = -\tau \tag{6:21}$$

It is customary to introduce the friction velocity $u_* \equiv (\tau/\rho)^{1/2}$ as a convenient scaling velocity. Equation (6:6) may now be written for the surface layer in the form

$$K_m \frac{\partial \bar{u}}{\partial z} = u_*^2 \quad (6:22)$$

where u_* is independent of height. In order to obtain the wind profile from this equation it is necessary to make an assumption about K_m . This can be done by developing Prandtl's mixing length model (1932) or by simple dimensional analysis. K_m has the dimensions of a velocity times a length. The natural choice of a velocity scale is u_* and of a length scale is the height z .† We therefore assume

$$K_m \propto u_* z \quad (6:23)$$

or

$$K_m = k u_* z$$

where k is a constant of proportionality, the von Karman constant. Equation (6:23) combined with (6:22) yields

$$\frac{\partial \bar{u}}{\partial z} = \frac{u_*}{kz}$$

This expression indicates an infinite shear at the surface $z = 0$ which, of course, is unrealistic. In order to keep the shear finite a surface roughness z_0 is introduced such that

$$\frac{\partial \bar{u}}{\partial z} = \frac{u_*}{k(z+z_0)} \quad (6:24)$$

Integration yields

$$\frac{\bar{u}}{u_*} = \frac{1}{K} \ln \frac{z+z_0}{z_0} \quad (6:25)$$

which is the well-known logarithmic wind profile.

There is considerable experimental evidence that verifies this profile. Therefore although the assumptions leading to it are somewhat vague they are justified a posteriori by the observations.

The constants k , the von Karman constant, and z_0 are not given by theory but have to be determined by experiment. Their precise physical meaning is not fully understood.

The von Karman constant, k , was derived theoretically by Hamel (1943) but close examination of this "theory" left the scientific community unconvinced. The numerical value of this constant is also still in doubt. Wind tunnel experiments suggest that $k \cong 0.4$. Experiments in the atmosphere show a rather large scatter. So far, the best experiments carried out by the Boundary Layer Branch of the AFCRL in Kansas yielded $k = 0.35$ (Businger et al., 1971) which is in agreement with an extrapolation by Tennekes (1968) of wind tunnel data to very large Reynolds numbers. More careful experimental work is needed to settle this matter.

The roughness length z_0 , is indeed related to the structure of the surface, but in a rather complex way. It may be useful to think of this length as the size of the smallest turbulent eddies, because it is

†The similarity argument leading to equation (6:23) can be used (actually by-passing (6:22)) directly on $\partial \bar{u}/\partial z$. By stating $\partial \bar{u}/\partial z = f(u_*, z)$ the obvious combination which is dimensionally correct is $\partial \bar{u}/\partial z \propto u_*/z$ and this leads similarly to (6:24).

assumed that the flow is turbulent right at the surface. The eddy viscosity at the surface $K_m(z=0) = k u_* z_0$ is consistent with this notion.

6.3.1.1 Aerodynamic Smooth Flow†

When the surface is relatively smooth and z_0 becomes small the value of the eddy viscosity at the surface, $k u_* z_0$, may become of the same magnitude as the kinematic viscosity, ν . In that case (6:22) is no longer correct near the surface and we should have

$$(K_m + \nu) \frac{\partial \bar{u}}{\partial z} = u_*^2$$

When $k u_* z_0 \ll \nu$ the flow at the surface is entirely determined by viscous shear and turbulent transfer is negligible. This condition is called “aerodynamic smooth flow.” Nikuradse (1933) carried out extensive wind tunnel experiments in order to define the transition from smooth to rough flow. He found that for

$$\frac{u_* z_0}{\nu} < 0.13$$

the flow is smooth and that for

$$\frac{u_* z_0}{\nu} > 2.5$$

the flow is fully-rough; and a transition regime exists in between those two values. $u_* z_0/\nu$ may be considered to be the Reynolds number of the smallest eddies in the flow. Nikuradse who related the roughness length z_0 , to the actual size of the roughness elements, 1_r , of the surfaces he experimented with. He found the empirical relation

$$z_0 = 1_r/30$$

For a smooth flow z_0 is to be replaced by u_*/ν in (6:24) so that the shear $(\partial u/\partial z)_0$ near the surface has the finite value

$$\left(\frac{\partial u}{\partial z}\right)_0 = \frac{u_*^2}{\nu}$$

Integration of (6:24) with this condition yields the wind profile over a smooth surface

$$\frac{\bar{u}}{u_*} = \frac{1}{K} \ln \left(\frac{ku_* z}{\nu} + 1 \right)$$

However, this is not quite correct because there is a transition from laminar to turbulent flow starting from the surface upward and (6:24) is strictly valid for the fully turbulent regime. Nikuradse found that a good approximation for the flow over a smooth surface is given by

$$\frac{\bar{u}}{u_*} = \frac{1}{K} \ln \frac{u_* z}{\nu} + 5.5$$

for $z \gg \nu/u_*$.

The earth’s surface is rarely aerodynamically smooth but it does occur, notably with light winds over water, snow or ice surfaces. Some recent measurements taken in Utah indicate that smooth flow also does occur over the Bonneville Salt Flats. In this context it should be noted that water surfaces are, in general, quite smooth even in the presence of wind driven waves.

†For a more complete discussion the reader is referred to Sutton’s (1953) text on Micrometeorology.

6.3.2 Steady State Similarity

The results obtained for the surface layer may be used in the description of the entire boundary layer as the asymptotic condition to which the wind profile in the "outer" layer must adjust itself. The procedure to be followed now is to formulate a "velocity defect law" (Townsend, 1956) which holds in the region where the boundary layer approaches the free stream velocity, i.e., the geostrophic wind. Subsequently the two profiles will be matched in an area of overlap. This technique was first used in this context by Csanady (1967) and somewhat refined by Blackadar and Tennekes (1968).

As we have seen in the surface layer the scaling velocity is u_* and the scaling length is z_0 which in similarity form may be expressed as

$$\frac{u}{u_*} = f_1 \left(\frac{z}{z_0} \right) \quad (6:25a)$$

and (6:25) is the explicit form for this relation. This means that when we know z_0 and u_* we don't need to know what the geostrophic wind u_g nor what the Coriolis parameter f is. In the outer part of the boundary layer the influence of z_0 is no longer felt but instead a new length scale is needed. Such a scale is formed by u_*/f which may be considered proportional to the height of the boundary layer. The similarity profile is now of the form

$$\frac{u}{u_*} = \frac{u_g}{u_*} + f_2 \left(\frac{zf}{u_*} \right) \quad (6:26)$$

where f_2 is a complex function to allow for the change in direction from u_* at the surface to u_g at the top of the boundary layer.

We assume now that there is an overlap region where equations (6:25) and (6:26) are the same. Since the form of f_1 is given by (6:25) the logarithmic wind profile f_2 must be logarithmic as well in this region. Therefore we assume

$$f_2 = \frac{1}{k} \ln \frac{zf}{u_*} + A \quad (6:27)$$

where A is complex, i.e., $A = A_r + i A_i$. If we substitute (6:27) in (6:26) and subtract (6:26) from (6:25) we obtain

$$\frac{1}{k} \ln \frac{u_*}{z_0 f} - A = + \frac{u_g}{u_*} \quad (6:28)$$

If the angle between τ_0 and u_g is α this equation may be split into real and imaginary parts,

$$\frac{u_g}{u_*} \cos \alpha = \frac{1}{k} \ln \frac{u_*}{f z_0} - A_r \quad (6:29)$$

and

$$\frac{u_g}{u_*} \sin \alpha = -A_i$$

The details of the profile are known when A_r and A_i are determined. Csanady (1967) quotes as reasonable estimates

$$A_r = 3.8 \text{ and } A_i = 10.7 \quad (6:30)$$

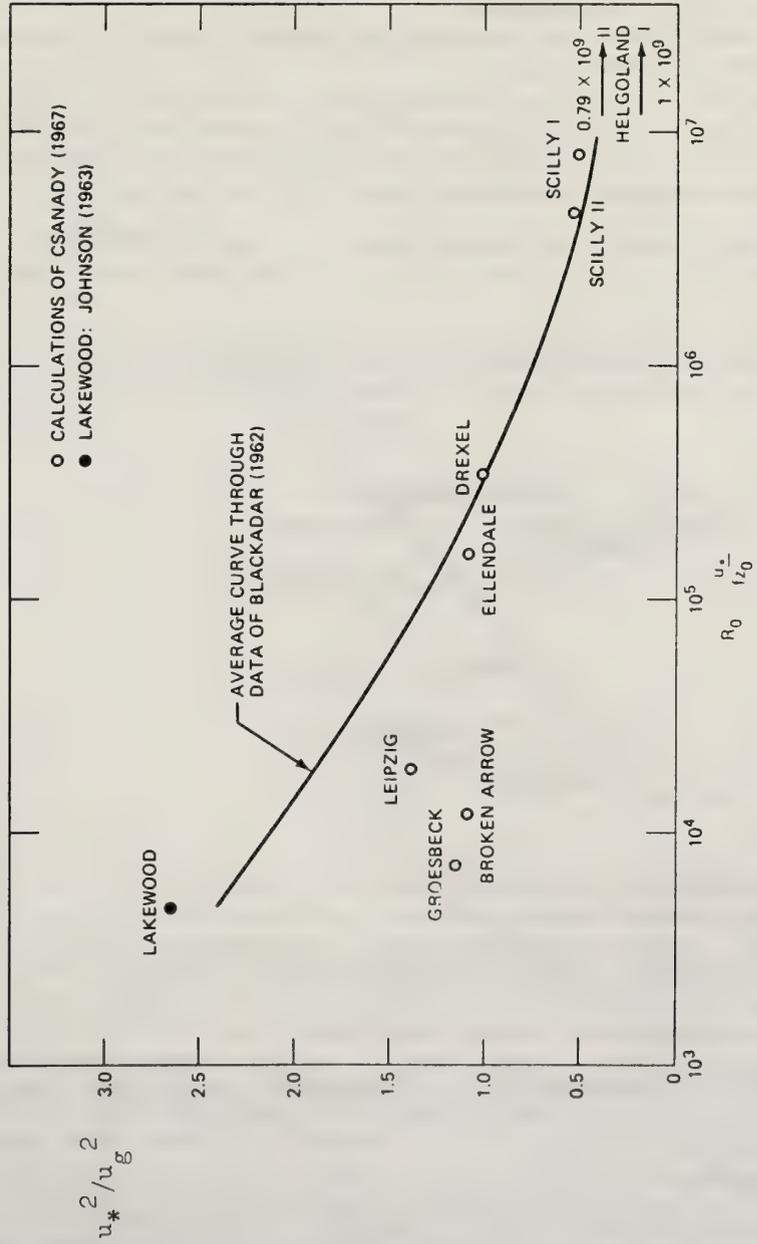


Figure 6.1 Geostrophic drag coefficient as function of surface Rossby number. The points are from Csanady (1967) and the curve is from the average curves of Blackadar (1962).

So far the structure of the neutral boundary layer has been expressed in terms of u_* , f and z_0 . However, from a macroscopic point of view it is desirable to express it in terms of u_g , f and z_0 or in terms of the surface Rossby number, $Ro = u_g/fz_0$, because the geostrophic wind u_g may be considered a boundary condition for the boundary layer which may be derived from the large scale pressure gradient. The stress or friction velocity u_* is usually not readily available and therefore can be given in the dimensionless form u_*/u_g as a function of Ro .

In (F6.1) u_*^2/u_g^2 , which is the geostrophic drag coefficient, is given as a function of Ro , and in (F6.2) $\sin \alpha$ is given as a function of u_*/u_g .

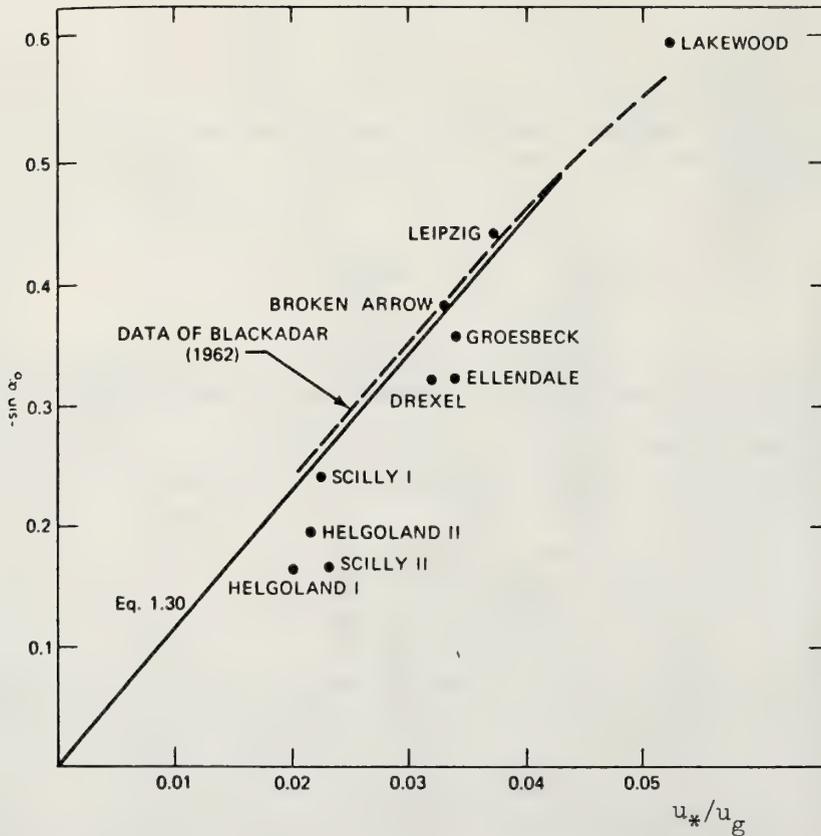


Figure 6.2 Angle of ground stress against isobars, as function of geostrophic drag coefficient (from Csanady, 1967). The dotted line has been constructed from the average curves given by Blackadar (1962).

6.3.3 Ekman Instability†

When a steady state is assumed in (6:7) and (6:8), we are left with a simple balance between Coriolis and viscous forces. Under the additional assumption of constant K , they can be solved analytically, producing one of the rare closed form solutions to the primitive equations.

†Sections 6.3.3 and 6.5.5 have been contributed by R. A. Brown.

$$\begin{aligned}\bar{u} &= u_g + e^{-z/\delta} (v_g \sin z/\delta - u_g \cos z/\delta) \\ \bar{v} &= v_g - e^{-z/\delta} (u_g \sin z/\delta + v_g \cos z/\delta)\end{aligned}\tag{6:31}$$

$$\delta = \sqrt{2K/f}$$

The path to this solution was paved with many assumptions which are subject to question in the event that the final solution fails to be verified observationally. This proved to be the case, as the analytic solution, the Ekman spiral, is rarely if ever observed in the atmospheric boundary layer. Thus, the major thrust of planetary boundary layer modelling was directed toward semi-empirical description, building from the surface outward.

However, it is interesting to return to the Ekman solution, and to check to see if it truly is a solution for the Coriolis/viscous balance. This can be done by checking to see if the solution is stable to infinitesimal perturbations.

When simple harmonic wave perturbations are added to the mean velocities in (6:1) and the resulting equations are linearized, non-dimensionalized, the mean terms from (6:7) and (6:8) are subtracted out and Coriolis and stratification effects are neglected, a classical perturbation equation, the Orr-Sommerfeld equation results

$$(V_E - c)(\phi'' - \beta^2\phi) - V''\phi = \frac{1}{i\beta Re} (\phi'''' - 2\beta^2\phi'' + \beta^4\phi)\tag{6:32a}$$

where

$$V = V_E + v ; Re = \frac{v_g\delta}{K_m} ; \delta = \left(\frac{2K_m}{f}\right)^{1/2}$$

$$v = \psi_z ; w = -\psi_y$$

$$\psi = \phi(z) e^{i\beta(y-ct)}\tag{6:32b}$$

$$\phi' \equiv \frac{\partial\phi}{\partial z}$$

This is an equation for the vertical structure of the stream function ϕ for two-dimensional perturbation oriented at arbitrary angles to the geostrophic flow. Since the time dependent behavior of the perturbation is assumed to be $\propto e^{ict}$, any solution satisfying the boundary conditions and having a negative imaginary component of the complex eigenvalue c will indicate growing perturbations, hence unstable conditions. The existence of such solutions will depend upon the parameters β , the wave number ($2\pi/\text{wavelength}$), the Reynolds/Rosby number, $v_g(K_m f)^{-1/2}$ and the vertical velocity profile V which depends upon the orientation of the perturbation in the mean flow.

Solutions to this equation indicate exponentially growing perturbations when $V'' = 0$ somewhere in the velocity profile, e.g., (Lilly, 1966), (Brown, 1970). Since the inflection point is a characteristic of the Ekman layer turning velocity solution, it is inherently unstable and cannot be expected to appear. The critical minimum Re for the neutral layer is about 50 (about 10 for unstable stratification). Typical atmospheric values are about 1000. Some consequences of this instability are discussed in 6.5.5.

6.4 The Diabatic Surface Layer

6.4.1 Formulation of a Stability Parameter

The unstable surface layer is different from the neutral surface layer because the turbulent structure is affected by the presence of a heat flux. This is clearly demonstrated by (6:20) where an important production term appears which is proportional to the heat flux. Richardson (1920) recognized the importance of (6:20), and introduced a stability parameter by dividing the second term by the first term

on the right hand side and assuming $K_h = K_m$ resulting in the dimensionless number known as the Richardson number

$$Ri \equiv \frac{g}{\theta} \frac{\overline{\partial\theta/\partial z}}{(\overline{\partial u/\partial z})^2} \quad (6:33)$$

This number indicates the relative importance of the two production terms and as such serves as a useful stability parameter.

As we shall see in 6.4.3 the assumption that $K_h = K_m$ is not quite valid, therefore it is better to leave the two terms in the form as they appear in (6:20). The ratio then is called the flux-Richardson number, Rf

$$Rf = -\frac{g}{T} \frac{\overline{w'T'}}{\overline{u'w'}} \partial\bar{u}/\partial z \quad (6:34)$$

This number has the drawback that it contains a mixture of covariances and mean profile information, which makes it more difficult to determine than Ri.

The search for a suitable stability parameter may also be carried out using similarity arguments. The dimensionless quantities are then constructed in such a way that they represent simple variables such as a dimensionless velocity, temperature and height.

Obukhov (1946)† was the first to analyze the boundary layer this way. He formulated the height where the buoyant energy production is equal to the shear production.

If in (6:34) $\partial\bar{u}/\partial z$ is replaced by $u_*/k(z+z_0)$, as was given in (6:24), a dimensionless height, ζ , is obtained

$$\zeta = -\frac{g}{T} \frac{\overline{w'T'} k(z+z_0)}{u_*^3} \quad (6:35)$$

It is clear that the quantity

$$-\frac{T}{g} \frac{u_*^3}{\overline{w'T'}} k = L \quad (6:36)$$

has the dimension of a length L , (the Obukhov length) and contains only quantities that are approximately constant throughout the surface layer. Also it should be noted that L contains only fluxes in addition to the constants. It may therefore be considered the quantity that characterizes the structure of the surface layer.

6.4.2 Flux-profile Relationships

A central objective of micrometeorological research is to establish fluxes from a knowledge of the mean profiles. Many efforts in this direction have been undertaken in the last 25 years or so, e.g., (Obukhov, 1946), (Lettau, 1949), (Monin and Obukhov, 1954), (Businger, 1955), (Swinbank, 1964), etc. Although theoretical efforts to obtain analytical solutions have not been successful it has been possible to use the similarity description to obtain semi-empirical relations.

A fairly complete description has been given recently by Businger et al. (1971) using observations taken in Kansas by the Boundary Layer Branch of AFCRL.

The wind shear was expressed in the dimensionless form $\phi_h \equiv z/\theta_* (\partial\bar{\theta}/\partial z)$, where $\theta_* \equiv -\overline{w'T'}/u_*$

†A translation of Obukhov's original paper appeared in *Boundary Layer Meteorology* 1971, p. 7-29.

Both ϕ_m and ϕ_h were expressed as functions of the dimensionless height ζ . The results are given in (F6.3) and (F6.4) respectively.

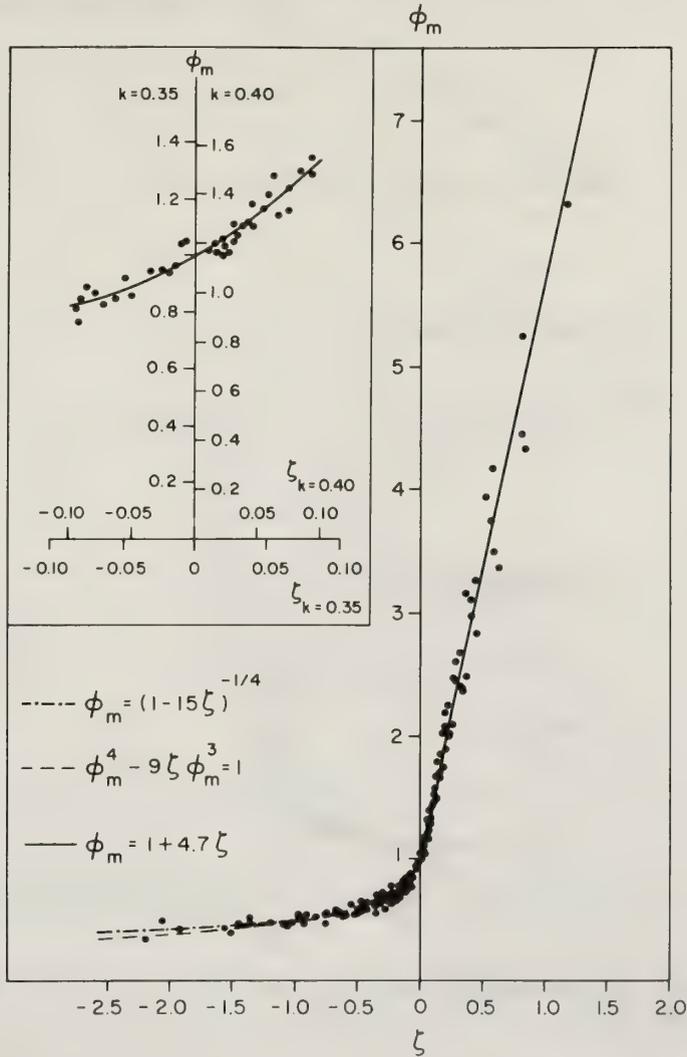


Figure 6.3 Comparison of dimensionless wind shear observations with interpolation formulas (from Businger et al., 1971).

The points fit the following analytical relations quite well

$$\phi_m = (1 - 15 \zeta)^{-1/4} \quad \text{for } \zeta < 0 \quad (6:37a)$$

$$\phi_m = 1 + 4.7 \zeta \quad \text{for } \zeta > 0 \quad (6:37b)$$

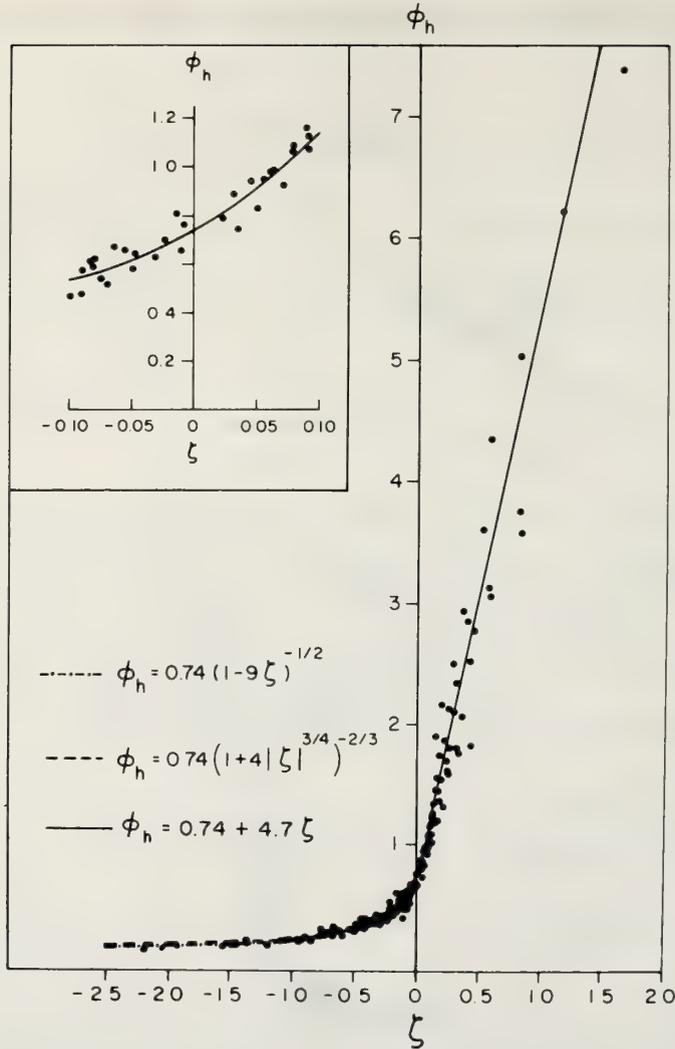


Figure 6.4 Comparison of dimensionless temperature gradient observations with interpolation formulas (from Businger et al., 1971).

and

$$\phi_h = 0.74(1-9\zeta)^{-1/2} \quad \text{for } \zeta < 0 \tag{6:38a}$$

$$\phi_h = 0.74 + 4.7\zeta \quad \text{for } \zeta < 0 \tag{6:38b}$$

When $\zeta = 0$; $\phi_m = 1$ which leads to the logarithmic profile, see (6:24) for $z \gg z_0$. The fact that for $\zeta = 0$; $\phi_h = 0.74$ indicates that $K_h/K_m \equiv \alpha \cong 1.35$, because $\phi_m/\phi_h = \alpha$. This is a larger ratio than usually

was assumed for neutral conditions, but it seems in close agreement with laboratory measurements (Hinze, 1959).

Equations (6:37) and (6:38) may be integrated to give the following explicit expressions for the wind and temperature profiles (Paulson, 1970)

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \left(\ln \frac{z}{z_0} - \psi_1 \right) \quad \text{for } \zeta < 0 \quad (6:39a)$$

where
$$\psi_1 = 2 \ln [(1+x)/2] + \ln [(1+x^2)/2] - 2 \tan^{-1} x + \pi/2$$

and
$$x = (1-15\zeta)^{1/4} = \phi_m^{-1}$$

$$\frac{\bar{u}}{u_*} = \frac{1}{k} \left(\ln \frac{z}{z_0} + 4.7 \zeta \right) \quad \text{for } \zeta > 0 \quad (6:39b)$$

and

$$\frac{\bar{\theta} - \theta_0}{\theta_*} = 0.74 \left(\ln \frac{z}{z_0} - \psi_2 \right) \quad \text{for } \zeta < 0 \quad (6:40a)$$

where
$$\psi_2 = \ln [(1+y)/2]$$

and
$$y = (1-9\zeta)^{1/2} = 0.74 \phi_h^{-1}$$

and θ_0 is the extrapolated temperature for $z = 0$. This is not necessarily the actual surface temperature.

$$\frac{\bar{\theta} - \theta_0}{\theta_*} = 7.4 \ln \frac{z}{z_0} + 4.7 \zeta \quad \text{for } \zeta > 0 \quad (6:40b)$$

From these equations it is clear that the diabatic profiles depend on two height scales z_0 and L . In most cases $z \gg z_0$ and the effect of z_0 is not felt when the gradients are considered, i.e., (6:37) and (6:38), but in the integrated form z_0 cannot be ignored. The structure of the surface layer is therefore defined by both z_0 and L . For full similarity conditions of the surface layer the ratio z_0/L should be the same (Businger, 1955).

Equations (6:39) and (6:40) may be used to compute the fluxes when observations of the profiles are available. By least square fitting of the smoothed profiles of \bar{u} and $\bar{\theta}$ to (6:39) and (6:40) u_* , θ_* and z_0 may be obtained and consequently also L . The heat flux can then simply be obtained from $F_h = -c_p \rho u_* \theta_*$ and the stress from $\tau = \rho u_*^2$.

6.4.3 The Functions $Ri(\zeta)$ and $\alpha(\zeta)$.

The relations $\phi_m(\zeta)$ and $\phi_h(\zeta)$ define the structure of the diabatic surface layer, so $Ri(\zeta)$ and $\alpha(\zeta)$ may be derived from (6:37) and (6:38). Nevertheless it is useful to show these relations explicitly because of the additional physical insight they provide.

Using the identities

$$\text{Ri} = \zeta/\alpha\phi_m = \zeta\phi_h/\phi_m^2 \quad (6:41)$$

and

$$\alpha = \phi_m/\phi_h \quad (6:42)$$

combined with (6:37) and (6:38) yield

$$\text{Ri} = \frac{0.74 \zeta (1-15\zeta)^{1/4}}{(1-9\zeta)^{1/2}} \quad \text{for } \zeta < 0 \quad (6:43a)$$

$$\text{Ri} = \frac{0.74 \zeta + 4.7 \zeta^2}{(1+4.7 \zeta)^2} \quad \text{for } \zeta > 0 \quad (6:43b)$$

and

$$\alpha = \frac{1.35 (1-9\zeta)^{1/2}}{(1-15\zeta)^{1.4}} \quad \text{for } \zeta < 0 \quad (6:44a)$$

$$\alpha = \frac{1.0 + 4.7\zeta}{0.74 + 4.7\zeta} \quad \text{for } \zeta > 0 \quad (6:44b)$$

These functions have been displayed in (F6.5) and (F6.6). The fit with the data is reasonable as should be expected because these are the same data as have been used in (F6.3) and (F6.4), from which the fits for ϕ_m and ϕ_h were obtained.

From (F6.5) it is apparent that in the unstable range $\text{Ri} \cong \zeta$, a condition which was independently assumed by Pandolfo (1966), Businger (1966), and Dyer (unpublished), and which makes the description of the surface layer rather simple. This approximation is not so good near the neutral point, because for $\zeta \rightarrow 0$, $\text{Ri} \cong 0.74\zeta$; however, at $\zeta = -0.1$ the difference between $\text{Ri} = \zeta$ and (6:43a) is only 15% and this difference decreases with decreasing values of ζ to about 4% for $\zeta \rightarrow -\infty$.

In the stable range it is apparent that Ri reaches a limit for $\zeta \rightarrow \infty$. Equation (6:43b) indicates that $\text{Ri} \cong 0.21$ for $\zeta \rightarrow \infty$. Webb (1970) obtained similarly a limit of about 0.20. If, as conditions become more stable, the heat flux and shear stress approach zero at the same rate, then L approaches zero because it contains the shear stress in the numerator to a higher power than the heat flux in the denominator. Thus as the turbulence diminishes and the flow becomes laminar $\zeta \rightarrow \infty$ and we can interpret the limiting value of Ri as the critical Ri -number. More experimental data are needed to verify the critical value mentioned here.

Equations (6:44a) and (6:44b) indicate that $\alpha > 1$ for all values of ζ . The data show rather a substantial scatter but mainly support this conclusion. In 6.4.2 it was already pointed out that $\alpha = 1.35$ for $\zeta = 0$.

In the unstable range α increases rapidly with decreasing ζ indicating that the transfer of heat becomes much more efficient than the transfer of momentum under unstable conditions.

6.4.4 Free Convection

Many efforts have been made to describe the boundary layer under the conditions of an upward heat flux without a mean wind, corresponding to the asymptotic case of $L \rightarrow 0$ and commonly known as

“free convection”. Although this condition is seemingly easy to describe using similarity arguments, observational results agree only in part with the predictions.

In free convection u_* vanishes and the governing variables reduce to $\overline{w'T'}$, z , and g/T . From these quantities a free convection scaling velocity and scaling temperature may be defined as follows

$$u_f = \left(\frac{gz}{T} \overline{w'T'} \right)^{1/3}$$

and

(6.45)

$$\theta_f = \left\{ \frac{T}{gz} (\overline{w'T'})^2 \right\}^{1/3}$$

Wyngaard et al (1971) argue that while the general four-variable similarity theory allows the prediction of a flow property to within an unknown *function* of ζ , the free convection prediction is to within an unknown *constant*. This may be demonstrated using, e.g., $\phi_h(\zeta)$.

In $\phi_h = z/\theta_* \partial\theta/\partial z$ the θ_* should be replaced with θ_f and the similarity argument requires now that

$$\frac{z}{\theta_f} \frac{\partial\theta}{\partial z} = a_1 \tag{6.46}$$

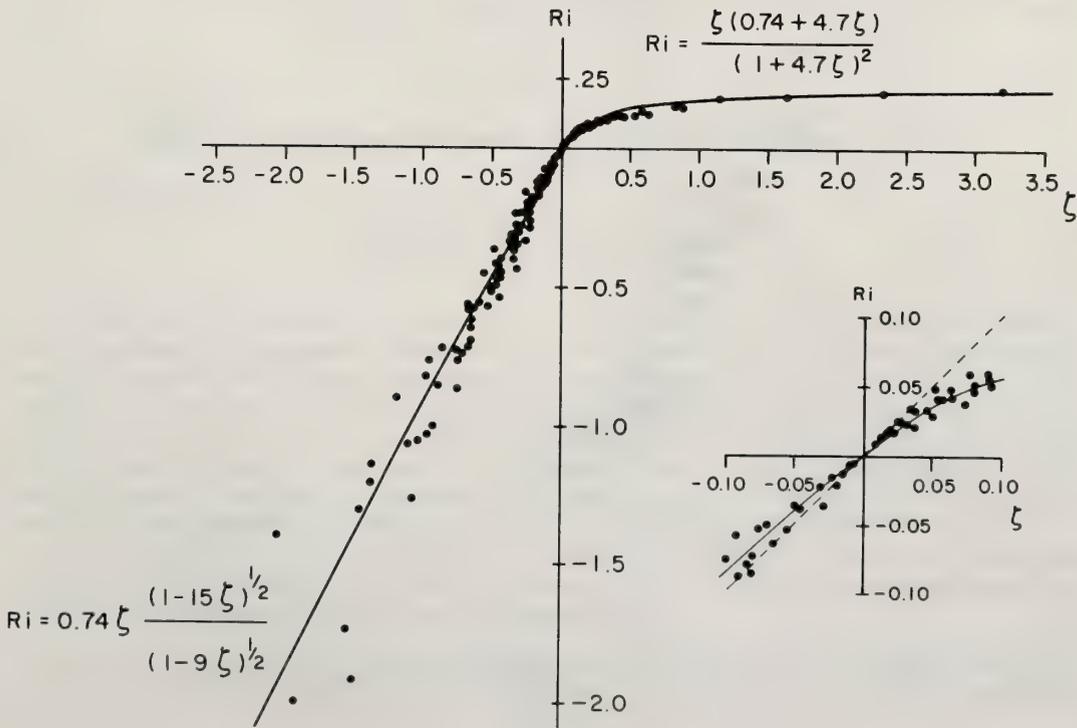


Figure 6.5 The dependence of Richardson number on stability (from Businger et al., 1971).

where a_1 is a constant. By rewriting (6:46) in terms of u_* and θ_* , using (6:45) it is found that

$$\frac{z}{\theta_*} \frac{\partial \theta}{\partial z} = -a_1 (-\zeta)^{-1/3} \quad (6:47)$$

which says that θ_h should approach a $-1/3$ power at large negative ζ . This is the same prediction that was originally made by Prandtl (1932), and later reformulated by Obukhov (1946) and independently by Priestly (1954). However, for large $(-\zeta)$ values (6:38a) approaches

$$\phi_h \propto (-\zeta)^{-1/2} \quad (6:48)$$

and this seems to be supported by the observations (see F6.7). Also Dyer's (1965) analysis of the Kerang and Hay data support this result. Elliott (1966) pointed out that proportionality (6:48) implies from the definitions of θ_h and ζ

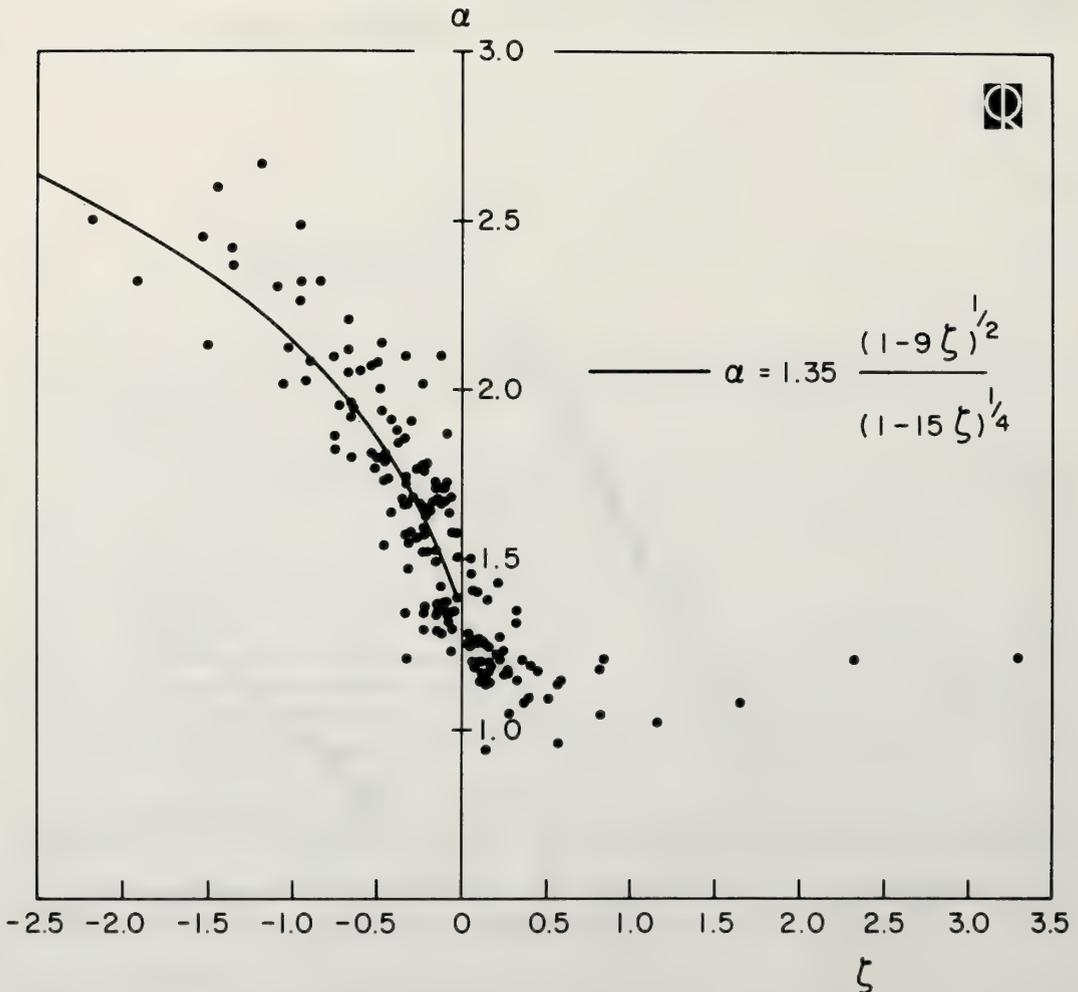


Figure 6.6 The dependence of the ratio of eddy diffusivities on stability (from Businger et al., 1971).

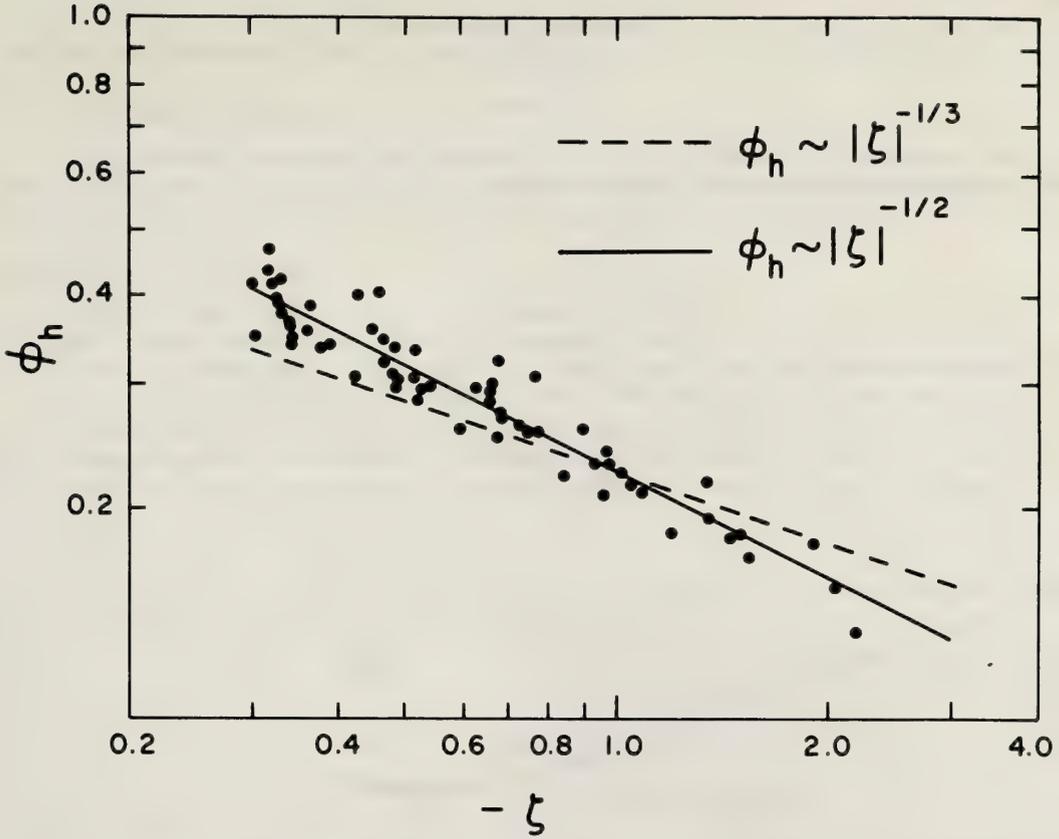


Figure 6.7 The dimensionless temperature gradient under very unstable conditions.

$$\frac{\partial \theta}{\partial z} \propto (\overline{w'\theta'})^{1/2} \left(\frac{g}{\theta}\right)^{1/2} u_*^{1/2} z^{-3/2} \tag{6:49}$$

This means that the effect of u_* is still felt at large values of $(-\zeta)$, which leads to a conflict when $u_* \rightarrow 0$. Businger (1972) argues that when $\bar{u} = 0$ and consequently $u_* = 0$ there are still substantial horizontal motions near the surface which have been introduced by the large scale convection. Close to the surface over a relatively short period compared to the large scale convection but a relatively long period compared to the time it takes to develop a local wind profile the structure of the profiles must be approximately the same as in the case when there is a large scale mean wind. Consequently there is locally shear production of turbulence and locally a friction velocity w_* may be defined. It is this w_* which then enters as a scaling parameter in (6:49). This point will be discussed further in the next section.

On the other hand the free convection argument which led to (6:46) and was not confirmed by observations, when applied to the variances of vertical velocity σ_w^2 , and temperature fluctuations, σ_θ^2 , describes the observations remarkably well. The predictions are

$$\frac{\sigma_w}{u_f} = a_2 \tag{6:50}$$

and

$$\frac{\sigma_\theta}{\theta_f} = a_3 \tag{6:51}$$

which equations in terms of u_* and θ_* may be rewritten as

$$\frac{\sigma_w}{u_*} = a_2 (-\zeta)^{1/3} \tag{6:52}$$

and

$$\frac{\sigma_\theta}{\theta_*} = a_3 (-\zeta)^{-1/3} \tag{6:53}$$

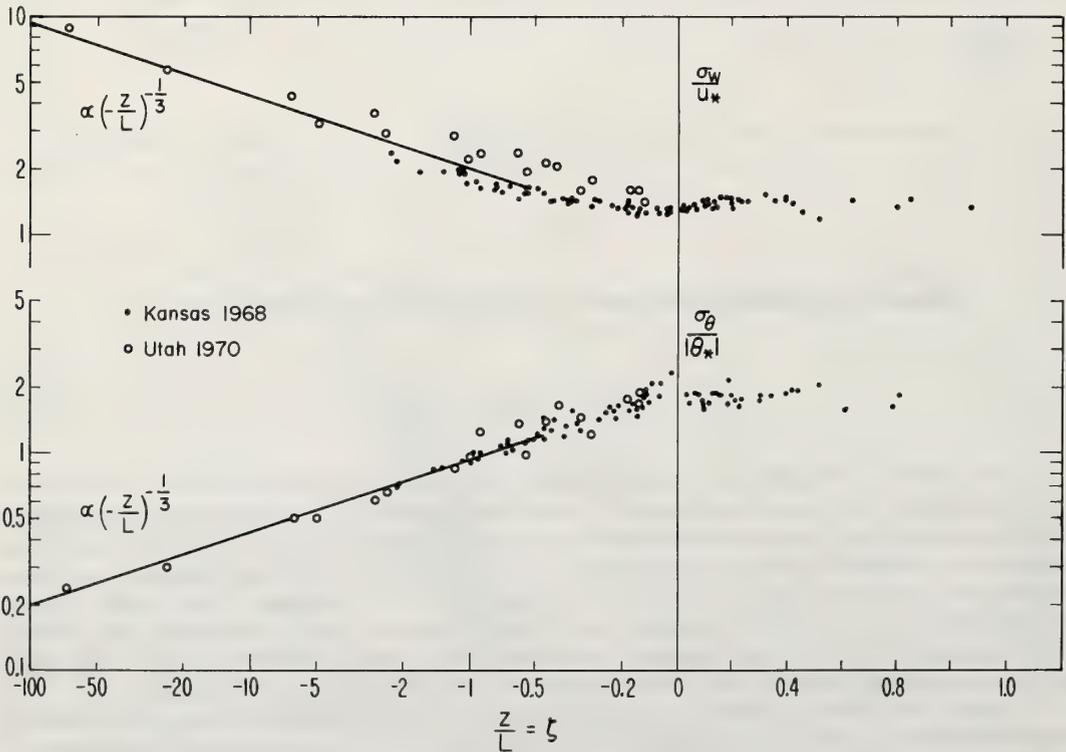


Figure 6.8 σ_w/u_* and $\sigma_\theta/|\theta_*|$ versus ζ . The Kansas data for σ_w/u_* are from Haugen et al., (1971); and for $\sigma_\theta/|\theta_*|$ from Wyngaard et al. (1971). The Utah data have been taken over the Bonneville Salt Flats, (unpublished).

Figure (6.8) shows how well these relations fit the data for large $(-\zeta)$. In fact, (6:53) is already valid for $\zeta \cong 0.1$. From this figure it is seen that $a_2 \cong 2$ and $a_1 \cong 0.9$.

In general it is found that average quantities containing only w' and θ' follow the free convection similarity quite well, whereas mean profiles or quantities that also include u' do not follow this similarity.

6.5 The unstable boundary layer.

In the preceding sections we have discussed the asymptotic conditions for the unstable boundary layer, i.e., the structure near the surface as a function of stability, and some characteristics of the entire boundary layer under neutral conditions. Now we shall attempt to fill in the open spaces.

6.5.1 Characteristic observations

Although several aspects of the unstable boundary layer are not understood, observations provide for a relatively simple description of the average quasi steady condition.

- a. Above the surface layer as soon as $(-\zeta) \gg 1$ the gradient of potential temperature vanishes. Thus the bulk of the boundary layer (typically from 100-1000 m) has an almost uniform potential temperature. Both (6:47) and (6:48) indicate an asymptotic approach to a uniform potential temperature with increasing height. A consequence of this is that the concept of an eddy transfer coefficient for heat, K_h becomes useless, see (6:12), because the heat flux itself is still finite.

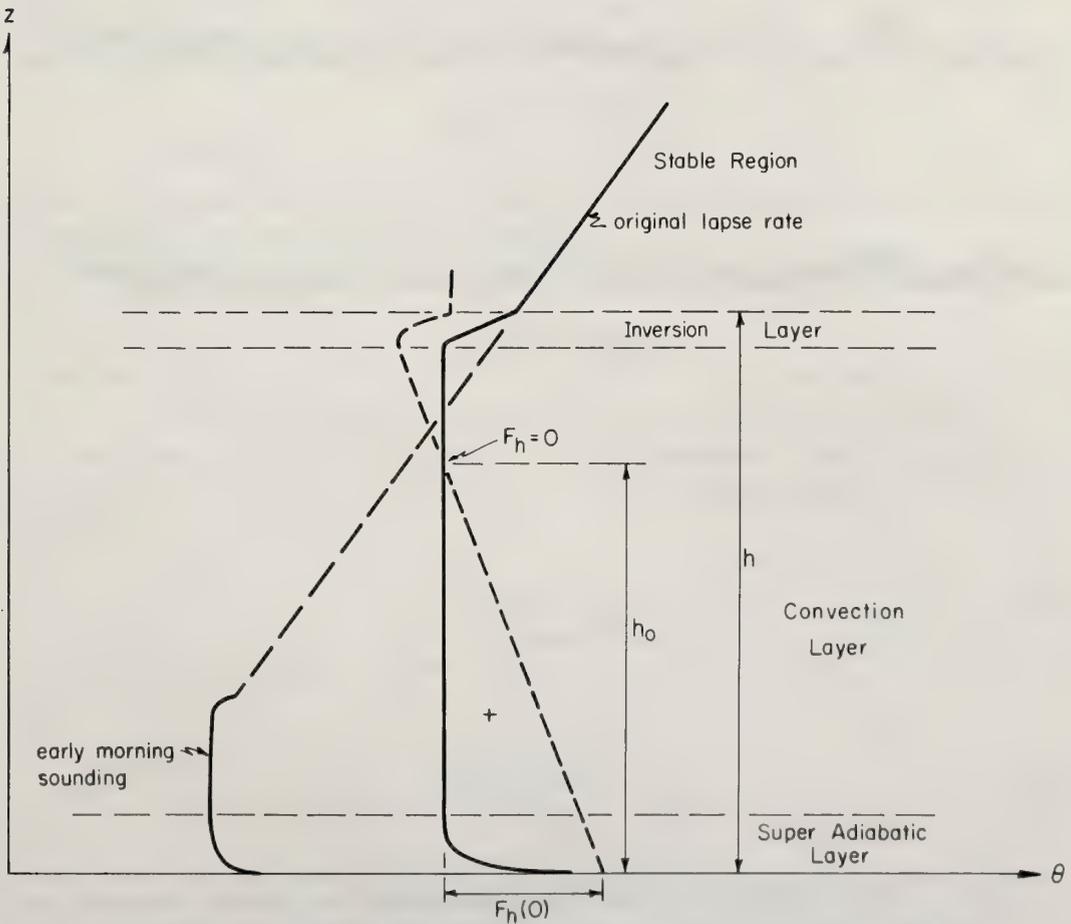


Figure 6.9 Schematic diagram of the unstable boundary layer.

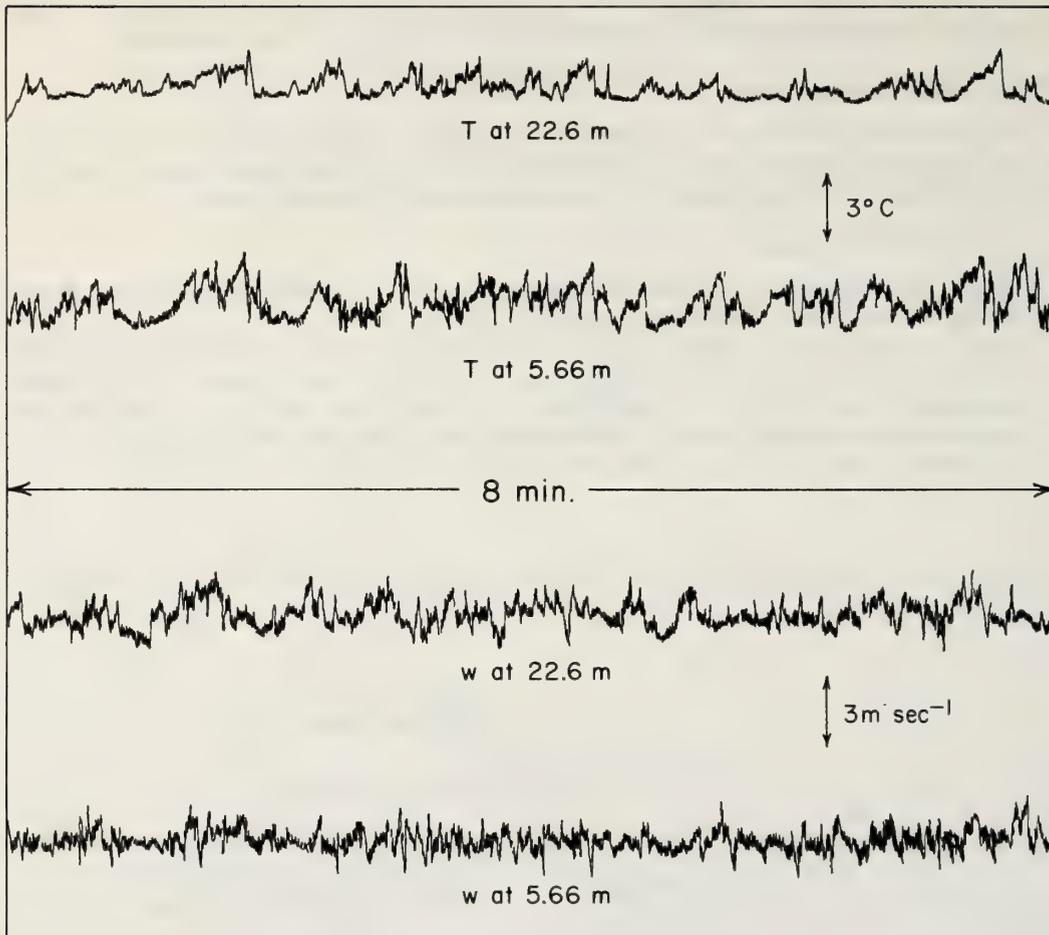


Figure 6.10 Sample of temperature on vertical wind component at two heights.

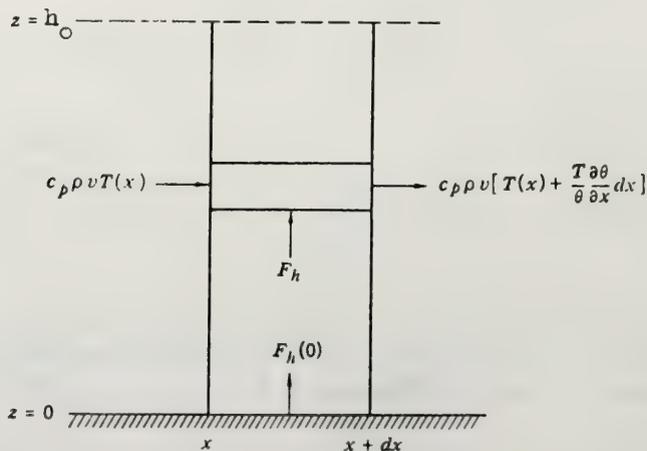


Figure 6.11 Distribution of vertical heat flux in a column of air of height h_0 , illustrating (6:54).

- b. The top of the boundary layer is usually marked by an inversion. This inversion is caused by the turbulent mixing of the air from below with the warmer air above. This results in a downward heat flux in the upper part of the boundary layer. A simplified discussion of this aspect has been given by Ball (1960). The height of the inversion layer depends on several factors such as the total amount of heat that entered the atmosphere at the earth's surface during the day, the lapse rate of the undisturbed air above the inversion, large scale convergence or divergence, etc. Figure (6.9) displays schematically the average temperature distribution of the unstable boundary layer.
- c. The temperature fluctuations have a characteristic structure, as demonstrated in (F6.10). It is clear that with increasing height the quiescent periods become longer and the total variance decreases as was already discussed in 6.4.4. The vertical velocity structure is quite different. The fluctuations appear to be much closer to a normal distribution and the variance increases with height in the surface layer, as also was shown in 6.4.4, and reaches a maximum at about 200-300 m. The vertical coherence of the fluctuations is a remarkable feature of the unstable boundary layer indicating plume like structures. This aspect will be discussed in more detail in 6.5.3.

6.5.2 The Vertical Heat Flux

The heat flux is the dominant factor in the formation of the unstable boundary layer. The heat which enters at the surface is distributed over a large segment of the boundary layer, changing the temperature and generating turbulent kinetic energy through the buoyancy force, see 6.5.3. If the divergence of net radiation is negligible, (6:10) may be integrated and the heat flux may be expressed by the following equations, see (F6.11)

$$F_h(0) = \int_0^{h_0} c_p \rho \frac{dT}{dz} dz \quad (6:54)$$

where h_0 is the height to which the heat flux extends. This height is somewhat below the height of the inversion which is considered the top of the boundary layer.

Under conditions of horizontal uniformity (6:54) reduces to

$$F_h(0) = \int_0^{h_0} c_p \rho \frac{\partial T}{\partial t} dz \quad (6:55)$$

Since over the bulk of the boundary θ and $\partial\theta/\partial t$ are independent of height and because $\partial\theta/\partial t = \theta/T \partial T/\partial t \cong \partial T/\partial t$, also $\partial T/\partial t$ is approximately independent of height. Consequently (6:55) may be integrated to

$$F_h(0) \cong c_p \rho \frac{\partial T}{\partial t} h_0 \quad (6:56)$$

and

$$F_h(z) \cong F_h(0) \left(1 - \frac{z}{h_0}\right) \quad (6:57)$$

These approximate equations may have practical use, because it enables a fairly detailed description of the heat flux throughout the boundary layer when $\partial T/\partial t$ and h_0 are known. Although in the lowest 30-100 m θ decreases with height, $\partial T/\partial t$ may still be approximately independent of height. In any case the error introduced in this lowest layer is usually negligible.

In (F6.12) a series of profiles from the Cedar Hill Tower (Kaimal, 1966) are given illustrating the development of the unstable boundary layer in the early morning. The sensible heat flux at the surface computed with (6:56) is about 8mWcm^{-2} for the period 0731 to 0827 CST and about 16mWcm^{-2} for the period from 0827 to 0926 CST.

6.5.3 Turbulent Energy Distribution and Scaling Parameters

In order to obtain a better physical insight in the unstable boundary layer it is useful to consider the equation for turbulent kinetic energy (6:20) more closely. Assuming steady state and multiplying by kz/u_*^3 this equation may be written after using (6:57) as

$$\frac{\overline{u'w'}}{u_*^2} \phi_m - \zeta \left(1 - \frac{z}{h_0}\right) - \frac{kz}{u_*^3} \frac{\partial}{\partial z} \overline{w' \left(e + \frac{p'}{\rho}\right)} = \phi_\epsilon \tag{6:58}$$

where $\phi_\epsilon = kze/u_*^3$. When this equation is integrated over the entire boundary layer the third term vanishes (see 6.2.4). By assuming that $\overline{u'w'}$ behaves in a similar way with height as $w'T$, a point which needs further consideration; and using (6:37a) the average production of the first term over the entire boundary layer is approximately

$$\frac{1}{h_0} \int_0^{h_0} \frac{\overline{u'w'}}{u_*^2} \phi_m dz \cong \frac{16}{21} \left(\frac{-L}{15h_0}\right)^{1/4}$$

for $h_0/L \gg 1$.

The second term, on the other hand, averaged in the same way yields

$$\frac{1}{h_0} \int_0^{h_0} -\zeta \left(1 - \frac{z}{h_0}\right) dz = \frac{h_0}{-6L}$$

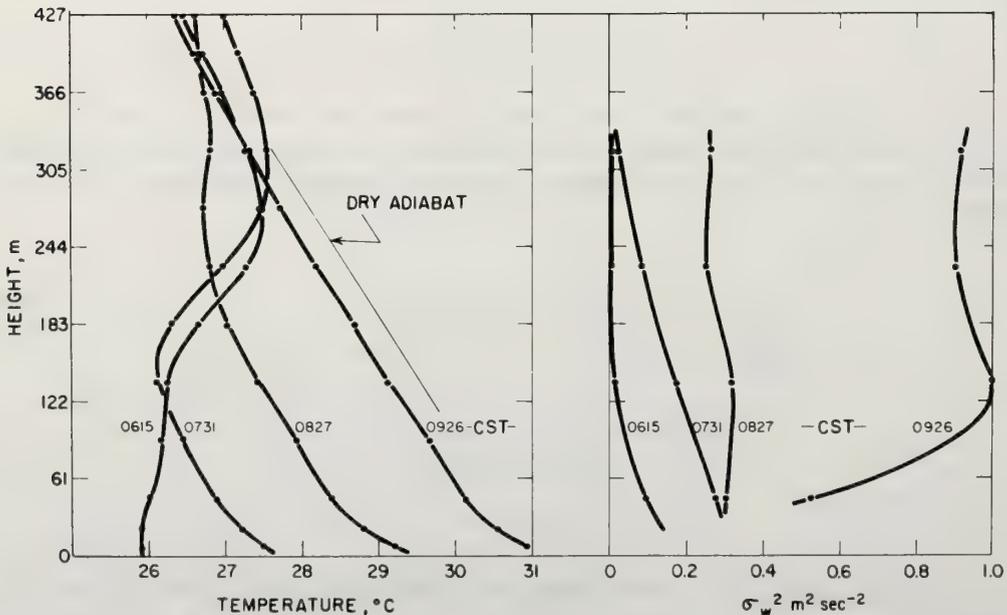


Figure 6.12 Profiles of temperature and σ_w (variance of wind as observed on Cedar Hill Tower in the morning of 6 August 1963 (Kaimal, 1966).

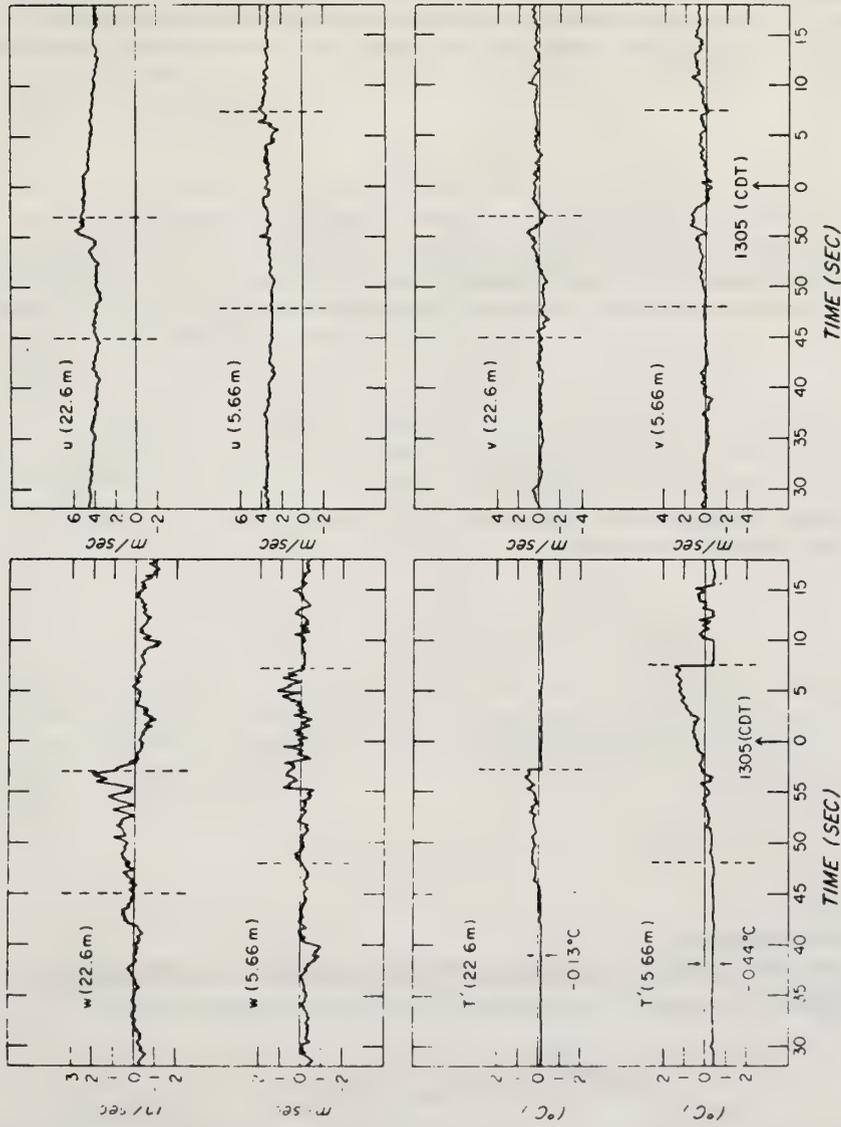


Figure 6.13 Traces of u , v , w , and T during passage of the convective plume (from Kaimal and Businger, 1970).

It is clear that for $h_0 \gg L$ the average buoyant energy production is much larger than the average shear production. In fact, in most cases the shear production can be neglected if the total turbulent kinetic energy in the boundary layer is considered. Because typically $|L| < 100$ m and $h_0 < 1000$ m therefore $h_0/|L| < 10$. Only in the cases that $|L| > 100$ m and becomes of the same order of magnitude as h_0 does the shear production play a significant role. But these conditions in the surface layer are considered to be near neutral. This illustrates the fact that the neutral boundary layer, as discussed in (6.3), rarely occurs.

The transition region from the neutral case to the well developed unstable case, $|L| < 100$ m, is very complex and little is known about its detailed structure. Although the well developed unstable case is not yet fully understood, it is considerably simpler and some general comments can be made. Equation (6:58) when averaged over the boundary layer reduces to

$$\frac{h_0}{-6L} = \langle \phi_\epsilon \rangle \quad (6:59)$$

where $\langle \phi_\epsilon \rangle \equiv 1/h_0 \int_0^{h_0} \phi_\epsilon dz$ the average dissipation of turbulent energy.

The average dissipation $\langle \epsilon \rangle$ over the boundary layer is defined by the characteristic turbulent velocity and the scale of turbulence. (In the neutral surface layer we have $\epsilon = u_*^3/kz$). So a reasonable assumption is

$$\langle \epsilon \rangle = a_3 \frac{\langle \sigma_w^3 \rangle}{h_0} \quad (6:60)$$

From laboratory measurements of Deardorff and Willis (1967), Tennekes (1970) estimated $a_3 \cong 2.4$. By introducing (6:60) into (6:59) we obtain

$$\langle \sigma_w \rangle = a_4 \left\{ \frac{g}{T} h_0 \frac{F_h(0)}{c_p \rho} \right\}^{1/3} = a_4 u_{f0} \quad (6:61)$$

a result consistent with (6:52). Or in dimensionless form

$$\frac{\langle \sigma_w \rangle}{u_*} = a_4 \left(\frac{h_0}{-L} \right)^{1/3} \quad (6:62)$$

The constant a_4 is of order unity.

Furthermore if the assumption is made that the average heat flux is proportional to the product of $\langle \sigma_w \rangle$ and $\langle \sigma_\theta \rangle$, i.e.,

$$\langle \overline{w'T'} \rangle = 0.2 \langle \sigma_w \rangle \langle \sigma_\theta \rangle \quad (6:63)$$

where the coefficient 0.2 is based on data given by Tsvang and Volkov (1967) and by Telford and Warner (1964), an expression similar to (6:53) is obtained, using (6:57)

$$\langle \sigma_\theta \rangle = a_5 \left(\frac{h_0}{-L} \right)^{-1/3} \quad (6:64)$$

where a_5 again is of order unity.

The relevant scaling parameters in the bulk of the unstable boundary layer are apparently the same as in the free convection case. Tennekes (1970), therefore suggests that as soon as $\zeta \gg 1$ to call the condition "local free convection." The following scaling parameters may be identified:

- z_0 the roughness length
- L the Obukhov length, a stability parameter
- h_0 the height where the heat flux is zero
- h the height of the inversion layer, or top of the boundary layer ($>h_0$)
- u_* the friction velocity
- u_{f0} the convection velocity
- θ_* scaling temperature in the surface layer
- θ_{f0} convective temperature scale

It is of interest to note that the height of the neutral boundary layer, which may be characterized by $h_n = u_*/f$ does not appear in the unstable boundary layer. The reason is that the structure is entirely dominated by the buoyant energy production and therefore the heat flux. It seems reasonable to assume that the momentum flux adjusts itself to the height imposed by the heat flux. This suggestion has been made by Deardorff (1970a, b) and Businger (1971). On the other hand Tennekes (1970) sees a real dilemma in the two conflicting height scales, i.e., h_0 and h_n .

In (6.4.4) a local friction velocity w_* was introduced to account for the fact that even when the large scale average value of $u_* = 0$ there is shear production of turbulence near the surface caused by the larger scale convection in the boundary layer. It is tempting to speculate that a relationship between w_* and u_{f0} exists analogous to the relation between u_* and u_g in the neutral case. This means that the surface roughness z_0 should play a role in the constant of proportionality as well as the height of the boundary layer, h . These considerations lead to the postulate that

$$w_* = a_6 u_{f0} \ell_n \frac{z_0}{h} \tag{6:65}$$

The constant a_6 is presumably a universal constant. The effect of the Coriolis force is assumed to be negligible which may not be entirely correct.

The shear production of the turbulent energy near the surface will now be proportional to w_*^3/z in analogy with the neutral case (u_*^3/kz). And the ratio of buoyant production over shear production may be written (6:65)

$$\frac{g}{T} \frac{\overline{w'\theta'z}}{w_*^3} = a_7 \left(\ell_n \frac{h}{z_0} \right)^3 \frac{z}{h} \tag{6:66}$$

The interesting result is that this ratio is not directly dependent on the heat flux. It also suggests there is a lower limit to the Obukhov length provided this length is generalized to include w_* , so that

$$-\frac{z}{L} = \frac{\text{buoyant production}}{\text{shear production}}$$

In this case

$$L_{\min} = -a_8 h \left(\ell_n \frac{h}{z_0} \right)^{-3} \tag{6:67}$$

where $a_8 = (ka_7)^{-1}$. Equation (6:67) suggests that the L_{\min} will be larger over rough terrain than over smooth terrain under otherwise similar conditions.

The behavior of ϕ_h with height is now consistent with (6:48) and (6:49) provided u_* is replaced by w_* .

6.5.4 Plumes and Dustdevils

In the unstable boundary layer various convective structures appear some of which with a remarkably orderly structure. This is rather intriguing because they appear when the various forms of turbulent energy input are well developed and one would expect the most chaotic structure. Nevertheless plumes and dustdevils have been identified for many years, and dustdevils in particular appear to be well organized and relatively stable systems.

Convective plumes are usually identified by their saw-tooth appearance in temperature recordings obtained from a stationary sensor, (F6.10). Invariably, they exhibit vertical continuity and a tilt in the downwind direction. They are most clearly defined when the wind is light and the heat flux is relatively large. The dustdevil is probably several orders of magnitude less frequent than the plume, but is a more spectacular and energetic system. During the experiments in Kansas in 1967 and 1968 only one dustdevil passed through the instrumented tower and could be clearly identified as such. Kaimal and Businger (1970) have given a description of this case as well as an analysis of a typical plume. The data on which they based their plume and dustdevil analysis are given in (F6.13) and (F6.14) respectively.

a. The plume

The most striking feature of the plume is the sharp drop in temperature at the upwind edge. This microfront may be less than 1 cm thick which is the scale where viscous dissipation and molecular conduction becomes dominant. The maintenance of such a sharp front in the turbulent environment requires very active stretching along the front and strong convergence perpendicular to it. The w component shows a marked increase going from the lower to the upper level which is consistent with the stretching. Also w is predominantly positive within the plume (as delineated by the temperature trace, (F6.13)) and negative outside it. This characteristic is more pronounced at the upper than at the lower level. The u component remains low within much of the plume but increases suddenly just ahead of the front. An increase in u would be expected from the horizontal convergence associated with vertical stretching, although at the upper level it occurs as much as 2 seconds ahead of the front, which means that a strong burst of upward momentum transport occurs just before the passage of the front. This is frequently the case but not the rule for the average plume. The v component is relatively inactive, indicating that there is no rotation associated with the plume. However, many plumes that were examined showed a small direction shear at the front, which probably results from concentration of vorticity due to the stretching along the front.

A remarkable feature also is the uniform temperature before and after the plume. By comparing this temperature at the two levels it turns out that the temperature is adiabatic to within the accuracy of the profile measurements. This indicates that very little warm air is mixed into the quiescent region.

The data presented in (F6.13) allows us to speculate on the structure of the plume. Figure (6.15) illustrates why, although the horizontal wind increases with height, the plume may maintain a constant tilt angle β during its horizontal movement with the wind. Also significant entrainment must take place because the vertical transport at the upper level is about 2.5 times that of the lower level.

b. The dustdevil

The dustdevil shows a strong updraft around the vortex and a downdraft within the core. This appears to be a general characteristic of well-developed dust devils and other concentrated vortices in the atmosphere. The region of the strongest downward velocity is identified as the center of the dustdevil (indicated by arrow in (F6.14)). From the figure it is also clear that the region of downdraft corresponds with the strongest shear in the v component.

The dustdevil has many striking differences with respect to the plume:

1. Since the area surrounding the dust devil is quite turbulent, the limits of the dust devil are not as clearly defined as in the plume. At 5.66m the transition from upward to downward motion may be considered the outside boundary of the dust devil, but no such transition occurs at 22.6 m within the 25 sec of observations presented here. Our data show that the w component at 22.6 m

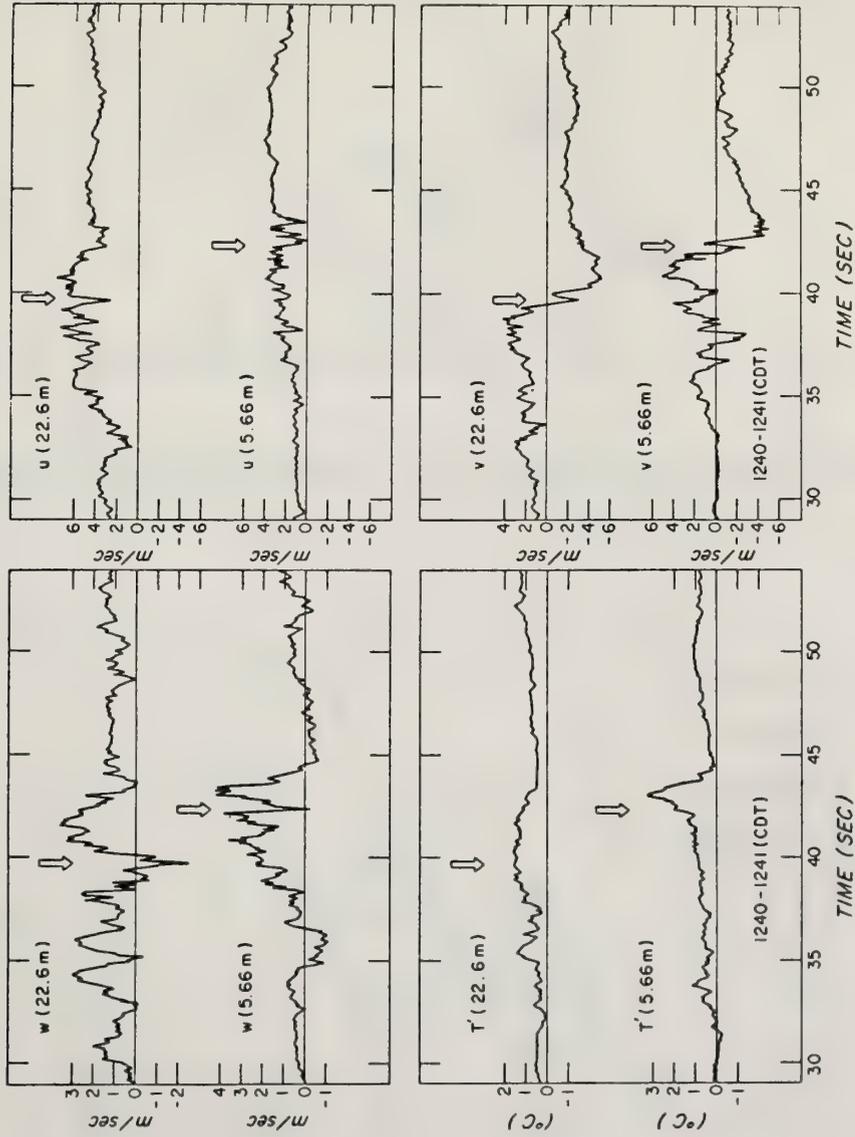


Figure 6.14 Traces of u , v , w , and T during passage of the dust devil (from Kaimal and Businger, 1970).

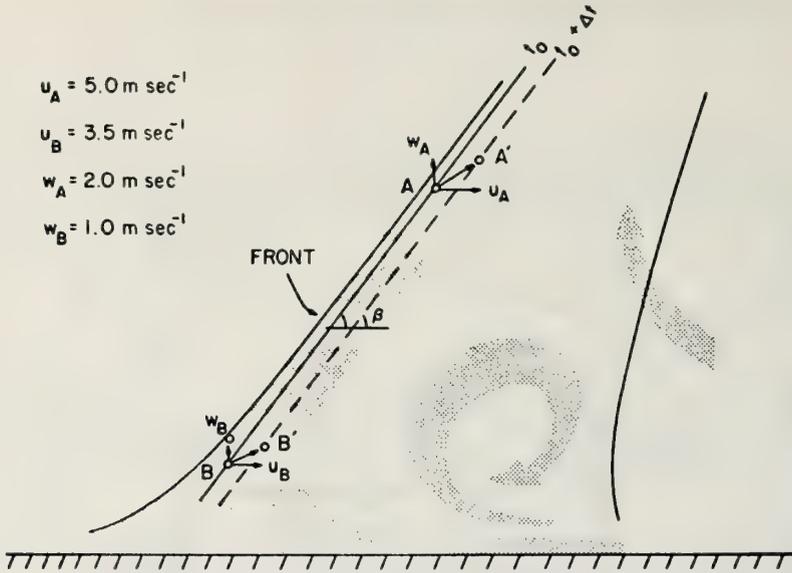


Figure 6.15 Two-dimensional model of the convective plume (from Kaimal and Businger, 1970).

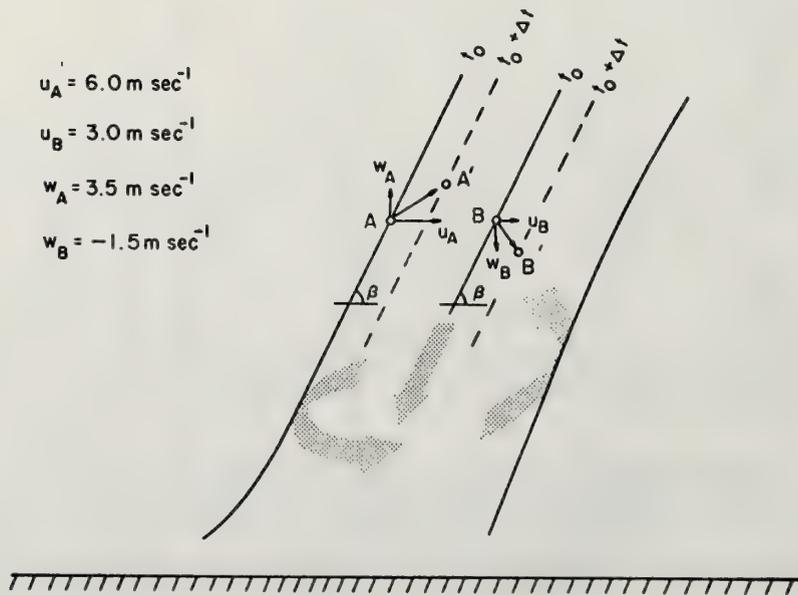


Figure 6.16 Two-dimensional model of the dust devil (from Kaimal and Businger, 1970).

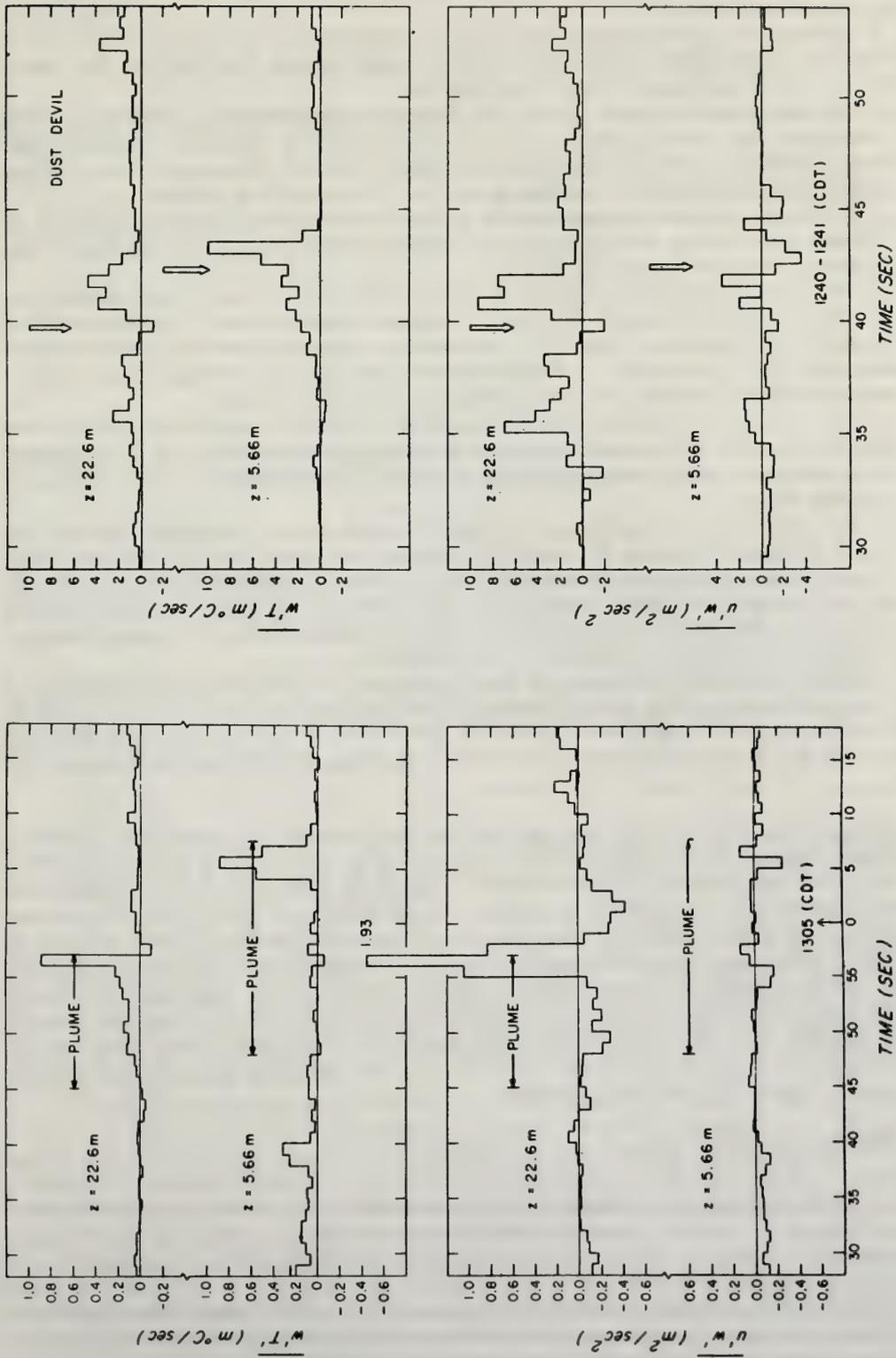


Figure 6.17 Variation of heat and momentum fluxes for the plume (1-sec averages) and the dust devil (0.5 sec averages) (from Kaimal and Businger, 1970).

was predominantly positive from about 8 sec preceding this record to about 28 sec following it. The same comments apply also to T' .

2. Ignoring the central downdraft, which is more pronounced at 22.6 m than at 5.66 m, the w component in the vortex shows roughly the same magnitude at both levels, implying that there is very little vertical stretching between 5.66 and 22.6 m.

3. The temperature trace does not show the characteristic sawtooth shape of the plume. There is a large blunt peak toward the rear of the vortex at 5.66 m which is absent at 22.6 m. Temperature excess is about the same at both levels; this, together with the absence of vertical stretching, suggests that entrainment is not nearly as important in the dust devil as in the plume.

4. It appears as though u increased suddenly at both levels immediately prior to the arrival of the dust devil. The sharp dip in the region of the downdraft is merely a consequence of the tilt in the dust devil, as is demonstrated by (F6.16).

5. The v component, which is a measure of the tangential velocity in the dust devil, indicates that the rotation is counterclockwise. This is of no particular significance since it has been established (Sinclair, 1969), (Carroll and Ryan, 1970) that there is no directional preference in the rotation of dust devils. But it is important to note that both the diameter of the strongest rotation and the tangential velocities associated with it remain essentially constant with height.

6. The fluxes of heat and momentum computed from the observations given in (F6.13) and (F6.14) indicate that the dustdevil transfers about 10 times as much as the plume. Both the plume and dustdevil show strong upward momentum flux at the 22.6 m level near the center, see (F6.17), a puzzling feature.

7. The plume which is basically a non-rotating system appears to be transported at some mean velocity close to the ground. The speed of translation u_t as determined from (F6.16) is about 2 m sec^{-1} which corresponds to the mean horizontal windspeed at some level below 0.5 m. The dust devil, on the other hand appears to move at a speed higher than the mean wind speed observed at 32 m (F6.16). This suggests that the dust devil extends vertically well into the boundary layer and is propelled at the mean wind speed near its center of gravity.

It should be emphasized that whereas the plume example that has been given here is a fairly typical one, the dust devil may be rather atypical because it is the only one that presented itself for analysis. However, comparison with the wind data on dust devils presented by Ryan and Carroll (1970) shows that the Kansas dust devil is fairly normal with a maximum vertical velocity and diameter somewhat larger than average.

The juxtaposition of the plume and the dust devil leads naturally to a speculation concerning a possible relation between the two. The stretching at the back of the plume tends to concentrate vorticity, which occasionally may conceivably be large enough to transform the plume into a dust devil. Many plume observations indicate some concentration of vorticity near the temperature discontinuity. One case was found in the 1968 Kansas observations where the vorticity was quite significant, almost suggesting an intermediate stage between plume and dust devil. However, the transformation of a plume into a dust devil must be a rare event because the number of plumes present at any time is much larger than the number of dust devils. Bergman (1969) showed that a vortex is most unstable when it is being formed, i.e., when the tangential velocity is small with respect to the vertical velocity. It is likely therefore that in most cases the vortex instability will destroy the plume and only in the rare event when the plume is able to concentrate vorticity rapidly enough will it turn into a dust devil.

6.5.5 Rolls (Secondary Flows)

The inherent instability of the Ekman profile, as discussed in 6.3.3, is difficult to incorporate into a flow model, because it effectively prevents the Ekman profile (6:31) from developing in the first place. Nevertheless, it is tempting to start with (6:31) and to carry out a finite perturbation analysis. This has been done by Brown (1970).

The growing perturbation as sketched in 6.3.3 has been included in an energy balance equation in order to estimate how much energy is available for its growth. The pertinent energy equation, similar to (6:20) when averaged over a volume element which includes one wavelength of the secondary flow may be

written

$$\frac{\partial \bar{e}_2}{\partial t} = \frac{1}{h\lambda} \int_0^h \int_0^\lambda \left\{ -(uw) \frac{\partial U}{\partial z} + vw \frac{\partial V}{\partial z} \right\} + \frac{g}{T} wT - \left(v \frac{\partial p}{\partial y} + w \frac{\partial p}{\partial z} \right) \} dy dz \quad (6:68)$$

The left hand side represents the average change in secondary flow energy. The first term on the right hand side represents the interchange of energy with the mean flow; the second term is the rate of work done by buoyancy due to the secondary flow and the last term represents a redistribution of energy within the boundary layer, the average value of which when integrated over the boundary layer vanishes, see 6.2.4. The dissipation term has been neglected in (6:68). This is justifiable because the scale of the secondary flow is very much larger than the smallest eddies where viscous dissipation becomes important.

For the neutral case the requirement of equilibrium i.e., $\partial \bar{e}_2 / \partial t = 0$ may be expressed after a horizontal integration as

$$\int_0^h \overline{vw} \frac{\partial U}{\partial z} = 0 \quad (6:69)$$

where the overbar indicates an average in the y-direction. This condition constrains the amplitude of the perturbation. The amplitude distribution of the perturbation also determines the magnitude of the secondary flow as given by (6:32b). The net effect of the momentum transfer from the mean flow to the secondary flow and vice versa is that an extra stress term appears in the steady state form of the boundary layer (6:7) and (6:8) with constant K. If the coordinate system is chosen so that the roll axis is in the x direction this stress term appears only in the equation for the v component. A closed solution exists for the thus modified (6:7) and (6:8) giving the modified mean flow profiles. It is of interest to note that the energy is transferred from the mean flow to the rolls in the lower part of the boundary layer and vice versa in the upper part of the boundary layer. This is also apparent by comparing the modified hodograph with the original Ekman spiral (F6.18). The maximum growth rate is found for $\epsilon = 18^\circ$, i.e., for a roll axis 18° to the left of the geostrophic wind. Figure (6.19) illustrates the pattern of secondary flows that is obtained with this analysis.

In the above analysis it is assumed that

- a. The shape of the perturbation is relatively insensitive to the resulting changes in the mean velocity profile, and
- b. The final modified profile is relatively stable to perturbations, although it still exhibits inflection points. There is little or no energy available for the growth of these perturbations. Observational evidence seems to justify this assumption.

The analysis for the thermally stratified case requires two modifications, (a) The perturbation equation (6:32) must now include terms that deal with the stratification. This is accomplished by introducing a bulk Richardson number ($Ri = g(\partial \bar{T} / \partial z) \{ \bar{T} (v_g / \delta)^2 \}^{-1}$), characteristic for the boundary layer. By doing so a 6th order perturbation equation results. (b) In the energy equation (6:68) the buoyancy term becomes important. However, this term is difficult to evaluate because the horizontal heat flux distribution has not been specified. A full analysis of the heat conduction equation in conjunction with the secondary flow model will be required.

Brown (1972) analyzed the modified Orr-Sommerfeld equation for a range of stabilities. The angle ϵ increases for stable stratification and decreases for unstable stratification, reaching $\epsilon = 0$ for moderately negative bulk Richardson numbers. The dynamic inflection point instability is modified somewhat, and the possibility of convective instabilities enters. The energy available for the secondary flow decreases with stability and increases with instability. This makes the roll phenomenon a much more common event in the unstable boundary layer justifying its discussion in this section.

Haugen et al. (1971) recently reported observations of short periods of both strong heat flux and strong momentum flux in the surface, which they call bursts. The intermittent nature of this phenomenon suggests a possible relation to the rolls discussed above. If this is so the heat flux contribution to (6:68) is not negligible. In fact we would expect this contribution to be positive enhancing the energy of the rolls.

The Atmospheric Boundary Layer

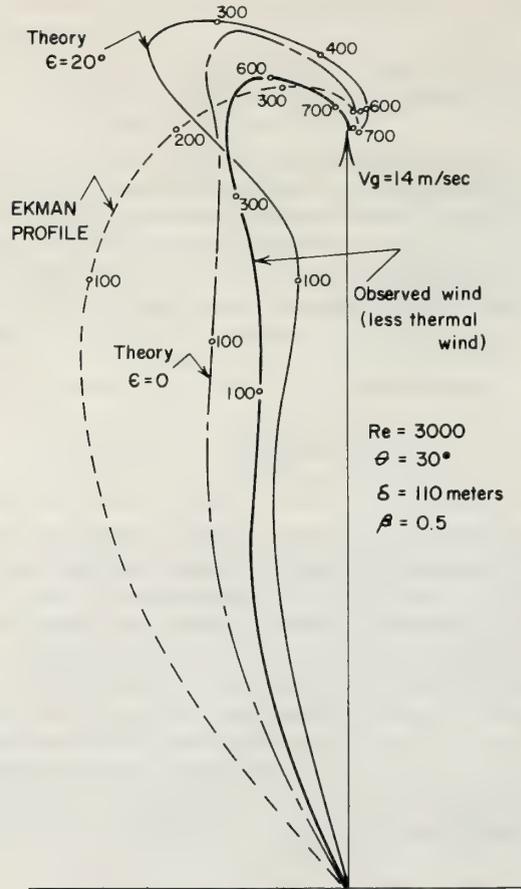


Figure 6.18 Mean wind hodographs for Jacksonville, April 4, 1968; the Ekman profile, and the theoretical profile including secondary flows (after Brown, 1970).

6.6 The Stable Boundary Layer

The structure of the stable boundary layer is markedly different from the unstable boundary layer. In fact, the neutral boundary seems to be in the middle of an almost discontinuous transition from stable to unstable or vice versa. In (F6.20) this characteristic is illustrated by a plot of the observed geostrophic drag coefficients versus stability. The very different character of the two stability ranges is also apparent in the discussion of the diabatic surface layer 6.4. However, the stability concepts introduced in 6.4 are still valid in both ranges.

6.6.1 The Critical Richardson Number

In his original study Richardson (1920) asked the question under what stability conditions would there be just no turbulence. Assuming $K_h = K_m$ and neglecting viscous dissipation he argued that $Ri_{CR} = 1$. Or, if we do not assume $K_h = K_m$ an equivalent statement is $R_{fCR} = 1$. This means that the shear production of turbulent kinetic energy is entirely compensated for by a negative buoyant production (6.20). The limit of $R_{fCR} = 1$ is an absolute upper limit. Since viscous dissipation plays an important role as a turbulent energy sink it is to be expected that the true critical value is less. From the experimental data in (F6.5) a value of $Ri_{CR} \cong 0.2$ is suggested, and because $K_h/K_m \cong 1.2$ for large ζ , $R_{fCR} \cong 0.25$. These values are much smaller than the value Richardson originally suggested.

Richardson's original approach and the experimental data presented are relevant to the transition from turbulent to laminar, and because the production terms contain the fluxes R_f is the relevant parameter to consider.

Following Richardson several investigations have been made to refine his criterion notably by Ellison (1957), Townsend (1958), and more recently by Beals (1970) and Arya (1972). These semi empirical theories all yield values of $R_{fCR} < 1$. The most recent result by Arya suggests a range for R_{fCR} from 0.15-0.25, a result which is not entirely independent of the above mentioned observed value.

The opposite problem of a transition from laminar to turbulent is in a sense simpler because a well defined laminar shear flow is analyzed with respect to infinitesimal perturbations. When the perturbations grow the flow is considered unstable and a breakdown into turbulence is expected. This analysis leads to a Ri_{CR} because the fluxes are purely molecular and consequently negligible, whereas the gradients are well defined. Taylor (1931) and Goldstein (1931) analyzed the stability of the flow this way and found $Ri_{CR} = 0.25$. The analysis was refined by Miles (1961) with the same result.

A somewhat different but very simple analysis, using the technique of exchange of parcels has been given by Ludlam (1967) with again the result that $Ri_{CR} = 0.25$. Businger (1969) extended Ludlam's analysis and showed that Ri_{CR} has a range of 0.25-1, depending on how much of the stable layer would participate in the exchange of parcels. Experimental evidence presented by Woods (1969) supports this notion. It may be argued therefore that the criterion for finite perturbations is less stringent than for infinitesimal perturbations. In a recent study Hines (1971) generalized Ludlam's approach to variable fields of u and θ and also finds less stringent criteria for the onset of turbulence.

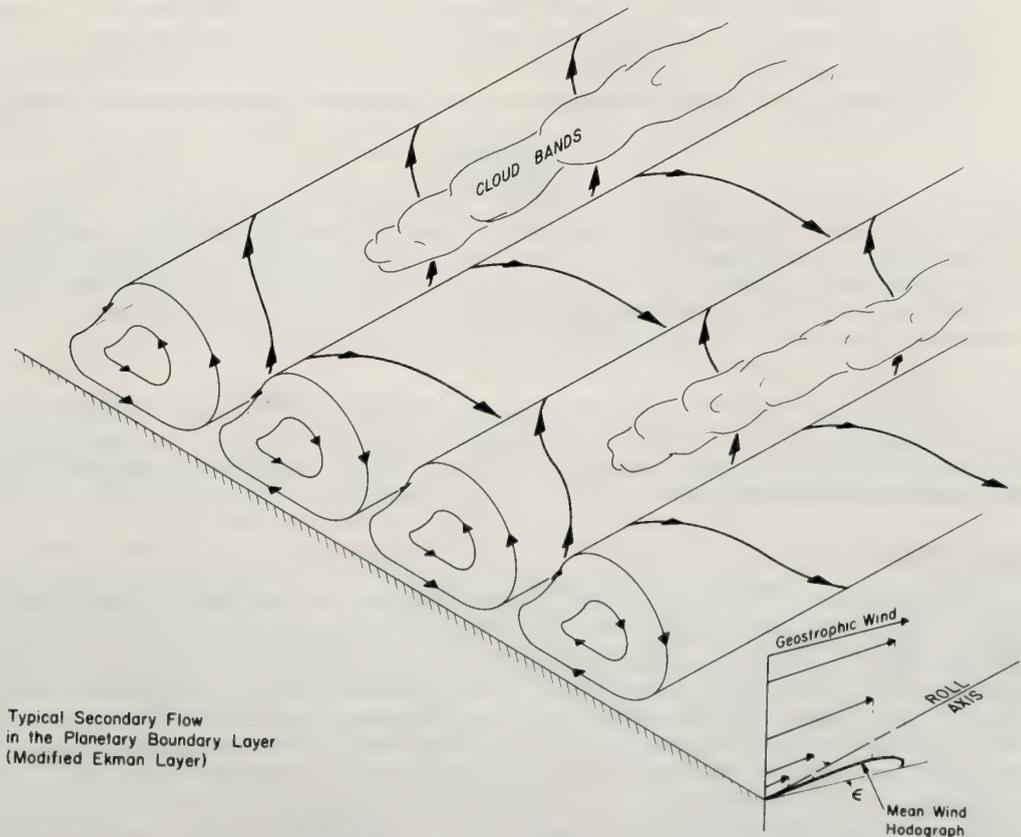


Figure 6.19 Typical secondary flow in the planetary boundary layer (after Brown, 1972).

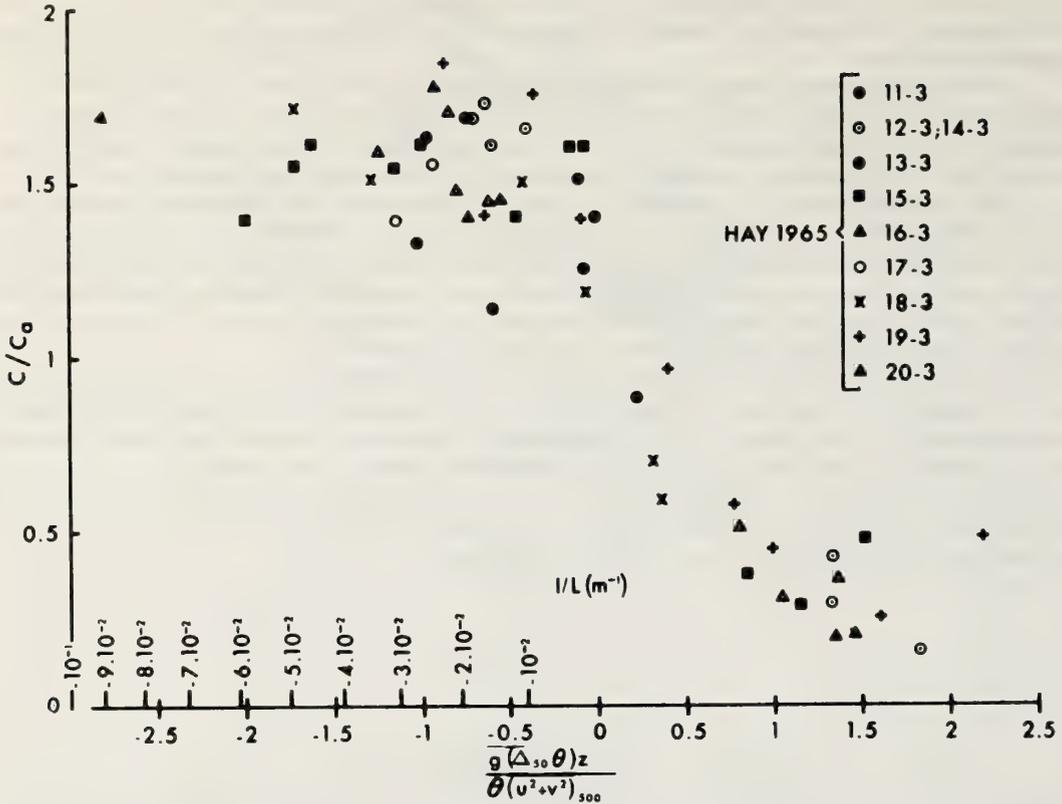


Figure 6.20 The effect of stability on the geostrophic drag coefficient. The ratio $C/C_a = u_*/u_{*a}$ is the friction velocity over the adiabatic friction velocity when the geostrophic wind is constant. In this case not the geostrophic wind, but the wind 500m above the surface is observed. The term $\Delta_{50}\theta$ designates the potential temperature difference between the surface and a level 50mb from the surface (which is at approximately 500m) (data from R. H. Clarke, 1970).

It is clear that the last word about Ri_{cr} and R_{fcr} has not been said. Careful experimental work is needed to determine more precisely these values. However, there is no doubt that a transition exists from laminar to turbulent and vice versa depending on the stability. This fact has profound consequences for the structure of the stable boundary layer.

6.6.2 Consequences of Ri_{cr}

A qualitative discussion of the sequence of events that occur when the boundary layer changes from unstable to stable may help to illustrate the complexities that are inherent to the stable boundary layer. Sometime during the afternoon of a clear day the heatflux changes sign from upward to downward and the lowest part of the boundary layer becomes stable. Shortly before the entire boundary layer was still in turbulent motion because there was both buoyant and shear energy production. The upward heat flux continues for some time in the upper part of the boundary layer after the lower part has become stable. The flux Richardson number is always small near the surface (except when it is completely calm) because the windshear is large. So, following R_f as a function of height it will increase from the surface on up; initially approximately proportional to the height then gradually less and eventually it reaches a maximum. Because the upper part of the boundary layer is still unstable R_f should be rather small there or even negative. The point is that if R_{fcr} is reached sometime after the transition from unstable to stable it will be reached first where the maximum value occurs at some height above, but relatively close to, the surface. As soon as this happens the turbulence will be dampened and a laminar layer will tend to form. This layer is

an effective barrier for all the fluxes. The transfer of momentum and heat from higher layers will be blocked. Under the laminar layer the transfer of momentum will continue down to the surface until the available momentum is depleted or R_f has become larger than critical. The result is that the wind diminishes and a period of calm sets in. This is a well known phenomena under fair weather conditions. The temperature near the surface drops dramatically because the outgoing net radiation is not compensated for by the downward heat flux. In the mean-time above the laminar layer momentum is still transferred downward whereas little heat is transferred. Consequently the momentum increases in the upper part of the laminar layer because it cannot pass through this layer. A strong windshear builds up and since there is not a similar effect of the heatflux R_i must decrease, and eventually below $R_{i_{cr}}$. This means that the laminar layer will gradually be eaten away by turbulence from above. Eventually the turbulence reaches the ground associated with a burst of momentum and heat.

After this the entire sequence of events may repeat itself. When the momentum is transferred to the ground the windshear will decrease and again $R_{f_{cr}}$ may be reached at a certain height above the ground, etc. It is therefore unlikely that long periods of steady state occur in the stable boundary layer. Also the layer of constant flux, i.e., the surface layer may become quite small or disappear altogether.

Although $R_{f_{cr}}$ may not be reached for sometime in the upper part of the boundary layer the production terms for turbulent kinetic energy become quite small and dissipation continues with the result that the whole level of turbulence decreases. Also the turbulent transfer decreases, and, where a layer has become laminar, even vanishes. Consequently, the strong linking of layers in the vertical, that existed in the unstable boundary layer to the extent that the whole upper part of the layer became practically uniform in potential temperature and momentum, has disappeared. Each layer will now be able to search for its geostrophic balance. But because the layer is initially considerably out of geostrophic balance, inertial oscillations may develop. This phenomenon is also related to the so-called nocturnal jet (Buajitti and Blackadar, 1957).

6.6.3 Interaction of Turbulence with Gravity Waves

One of the reasons for the sharp contrast between the stable and unstable boundary layers is the fact that the Brunt-Väisälä frequency, N is real in a stable layer and imaginary in an unstable one.

$$N \equiv \left(\frac{g}{\theta} \frac{\partial \theta}{\partial z} \right)^{1/2} \quad (6:70)$$

This frequency comes about as soon as the buoyancy force is a restoring force†, and the waves associated with it are called gravity waves. An extensive literature exists concerning these waves and for a detailed discussion the reader is referred to Chapter 7.

When N becomes real turbulent energy may excite gravity waves and the cascading process is slowed down. This may be related to the interesting way σ_w/u_* behaves as a function of stability, see (F6.8). In the unstable regime σ_w/u_* follows (6:52) quite well; near neutral apparently a minimum is reached, and a sudden rise occurs as soon as the stable regime is entered. Intuitively one would think that negative buoyancy would suppress the vertical component of turbulence more strongly than the other components. What actually happens is that the correlation between u and w is more strongly affected than w itself. This suggests that some of the variance of w must be attributed to gravity waves. It is difficult to formulate this in a quantitative form.

†A simple way of visualizing why N is the appropriate frequency may be given with the following consideration. Assume a parcel of air in a uniformly stable layer being displaced over a distance z from its equilibrium position. The restoring buoyancy force then is equal to the acceleration of the parcel, i.e., $g/\theta (\partial \theta / \partial z) z = dw/dt = d^2z/dt^2$. If no other forces act on the parcel this differential equation leads to a simple oscillation $z = z_0 \sin Nt$ where N is the frequency of oscillation.

6.7 Effects of change in terrain

So far the emphasis of our discussion has been on situations of steady state and horizontal uniformity. In reality horizontally uniform surfaces are quite rare over land so much so that it took serious efforts to find sites sufficiently representative of this. The well known sites near Hay and Kerang in Australia as well as the site near Liberal, Kansas are good examples of such efforts. On the other hand, the sea surface is usually relatively uniform, validating much of the previous discussions.

The consideration of non uniform surfaces opens up a host of complexities, only a few of which have been investigated into some depth. Most investigations have dealt with the effect of change in surface conditions such as a change in terrain from a smooth to a rough surface or from a cold to a warm surface or vice versa. In these cases it is assumed, and we shall adopt this assumption, that the wind direction is perpendicular to the surface discontinuity.

6.7.1 The internal boundary (neutral case).

The effect of a change in terrain is a modification of the boundary layer. When the air moves over a discontinuity in the surface characteristics its effect will be propagated downwind and upward. The transition zone between the original unmodified air and the modified air is called the internal boundary, and is illustrated by (F6.21). The formulation of the steady state problem near the surface may be considered two dimensional and the equations of motion (6:1) may be simplified to

$$\frac{\partial}{\partial x} \overline{\rho u^2} + \frac{\partial}{\partial z} \overline{\rho u w} = -\frac{\partial p}{\partial x} \tag{6:71}$$

$$\frac{\partial}{\partial x} \overline{\rho u w} + \frac{\partial}{\partial z} \overline{\rho w^2} = -\frac{\partial p}{\partial z} - \bar{\rho} g \tag{6:72}$$

Under neutral conditions and by comparing the order of magnitude of the various terms (6:71) reduces to

$$\bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{w} \frac{\partial \bar{u}}{\partial z} = -\partial \frac{\overline{u'w'}}{\partial z} = \frac{1}{\rho} \frac{\partial \tau}{\partial z} \tag{6:73}$$

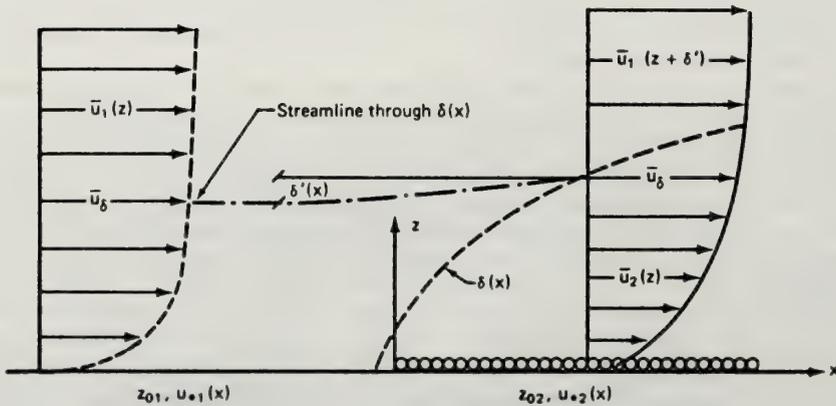


Figure 6.21 Schematic representation of the development of an internal boundary layer: flow from smooth to rough (after Plate, 1971).

and (6:72) reduces to (6:5), the hydrostatic approximation. The internal boundary layer problem then is determined by (6:73), the equation of continuity (6:16) in the form $\partial u/\partial x + \partial w/\partial z = 0$, and boundary conditions.

Equation (6:73) may formally be integrated to the height of the internal boundary layer δ , using the continuity equation, with the result

$$\frac{\partial}{\partial x} \int_0^\delta \bar{u} (\bar{u} - u_\delta) dz = \frac{1}{\rho} (\tau_\delta - \tau_2) = \frac{\tau_\delta}{\rho} - u_{*2}^2 \tag{6:74}$$

a result similar to the one obtained by von Karman (1921) and Pohlhausen (1921) for the laminar case. The index 2 indicates the flow downstream of the discontinuity.

In order to integrate (6:74) the shape of the wind profile needs to be known as well as the height of the internal boundary. From (F6.21) it is seen that $\bar{u}_\delta = \bar{u}_1 (\delta - \delta')$, where the index 1 indicates the flow upstream of the discontinuity. Because the boundary layer upstream of the discontinuity is assumed to be a constant stress layer it follows that

$$\tau_\delta = \tau_1 = \rho u_{*1}^2$$

Elliott (1958) assumed that δ' is negligible and that the profile before and after the discontinuity is logarithmic, i.e.,

$$\frac{\bar{u}_1}{u_{*1}} = \frac{1}{k} \ln \frac{z}{z_{01}} \quad \text{and} \quad \frac{\bar{u}_2}{u_{*2}} = \frac{1}{k} \ln \frac{z}{z_{02}}$$

but \bar{u}_δ is the same in both regimes, thus

$$u_{*1} \ln \frac{\delta}{z_{01}} = u_{*2} \ln \frac{\delta}{z_{02}} \tag{6:75}$$

Equations (6:74) and (6:75) represent two equations for the two unknown u_{*2} and δ and can be solved. However, the solution is a rather complicated algebraic expression which, except near $x = 0$, may be approximated rather closely by

$$\frac{\delta}{z_{02}} = a \left(\frac{x}{z_{02}} \right)^{0.8} \tag{6:76}$$

$$a \cong 0.75 - 0.03 \ln \frac{z_{02}}{z_{01}}$$

Although Elliott's theory is rather crude observations fit remarkably well. Many improved theories have been proposed since Elliott's (1958) pioneering paper, but relatively little improvement has resulted. Notably Panofsky and Townsend's (1964) effort, although suggesting a reasonable modification of Elliott's theory improved the fit with observations very little if any. Plate and Hidy (1967) considered the small displacement δ' , and introduced a correction of the \bar{u}_2 profile. Their results are slightly better than Elliott's, but lack of space prevents us from fully discussing this method.

A somewhat different approach to determine the height of the internal boundary has been given by Kazanski and Monin (1957). They suggest that the height of the internal boundary will be related to a vertical propagation velocity w_b which they assume to be proportional to σ_w . Monin (1958) suggested because σ_w is related to u_* that

$$w_b = u_{*2} f(\zeta) \tag{6:77}$$

where $f(\zeta)$ is a function of stability. In the neutral case $f(0)$ is a constant, because σ_w/u_* is a constant (about 1.3).

The height δ of the internal boundary is now simply determined by the equation

$$\frac{d\delta}{dt} = w_b = u_{*2} f(0).$$

but

$$dt = \frac{dx}{u_\delta}, \text{ and } \bar{u}_\delta = \frac{u_{*2}}{k} \ln \frac{\delta}{z_{02}}, \text{ so}$$

$$\frac{d\delta}{dx} = k \frac{f(0)}{\ln \frac{\delta}{z_{02}}}$$

which may be integrated to

$$x = \frac{1}{kf(0)} \delta \left(\ln \frac{\delta}{z_{02}} - 1 \right) \quad (6:78)$$

an equation which was derived by Miyake (1965). A good approximation of this equation is

$$\delta/z_{02} = 0.54 \left(\frac{x}{z_{02}} \right)^{0.84} \quad (6:79)$$

where for the constant $1/kf(0) = 1.45$ has been used. Equation (6:79) fits observations fairly well, see (F6.22). The fact that with this approach an empirical constant $f(0) \cong 1.7$ has been used eliminates in this case the need to use (6:74).

More detailed studies including the distribution of shear and momentum transfer within the internal boundary layer have been carried out by Bradshaw, Ferris and Atwell (1967) and independently by Peterson (1969). The numerical results of the horizontal stress distribution obtained by Peterson agree quite well with the observations of Bradley (1968). Figure (6.23) shows the fit between Peterson's theory and Bradley's data. An interesting point to note here is how the value of u_{*2}^2/u_{*1}^2 compares with the one expected for the change of geostrophic drag (F6.1). The change in roughness length was about two orders of magnitude. If we compare Bradley's measurements with the change in drag coefficient to be expected from the change in surface Ro number then it is clear that there is a fair amount of agreement already for a fetch of about 100 m or so.

6.7.2 The Internal Boundary Layer (Diabatic Case)

The generalization of the theory of the internal boundary to the diabatic case may be done with either Elliott's or Kazanski and Monin's approach. However, it is essential that the heat flux is known on both sides of the discontinuity, and consequently the Obukhov lengths L_1 and L_2 can be determined.

The generalization of Elliott's theory then proceeds by using the diabatic windprofile (6:39) in (6:74) and (6:75) leading to a rather complex set of equations. Although no basic difficulties are expected in carrying this out, so far it has not been reported in the literature.

Kazanski and Monin's approach actually has been designed to include the effect of stability. The function $f(\zeta)$ must now be formulated. Because $f(\zeta) = w_b/\sigma_w \cdot \sigma_w/u_*$, and w_b/σ_w is assumed to be constant (about 1.3), we need to know how σ_w/u_* varies with stability, which for strong instabilities has been given by (6:52). A more complete formulation for this ratio in the unstable range has been given by Businger (1971)

$$\frac{\sigma_w}{u_*} = \{ 1.6 + (-2\zeta)^{2/3} \}^{1/2} \quad (6:80)$$

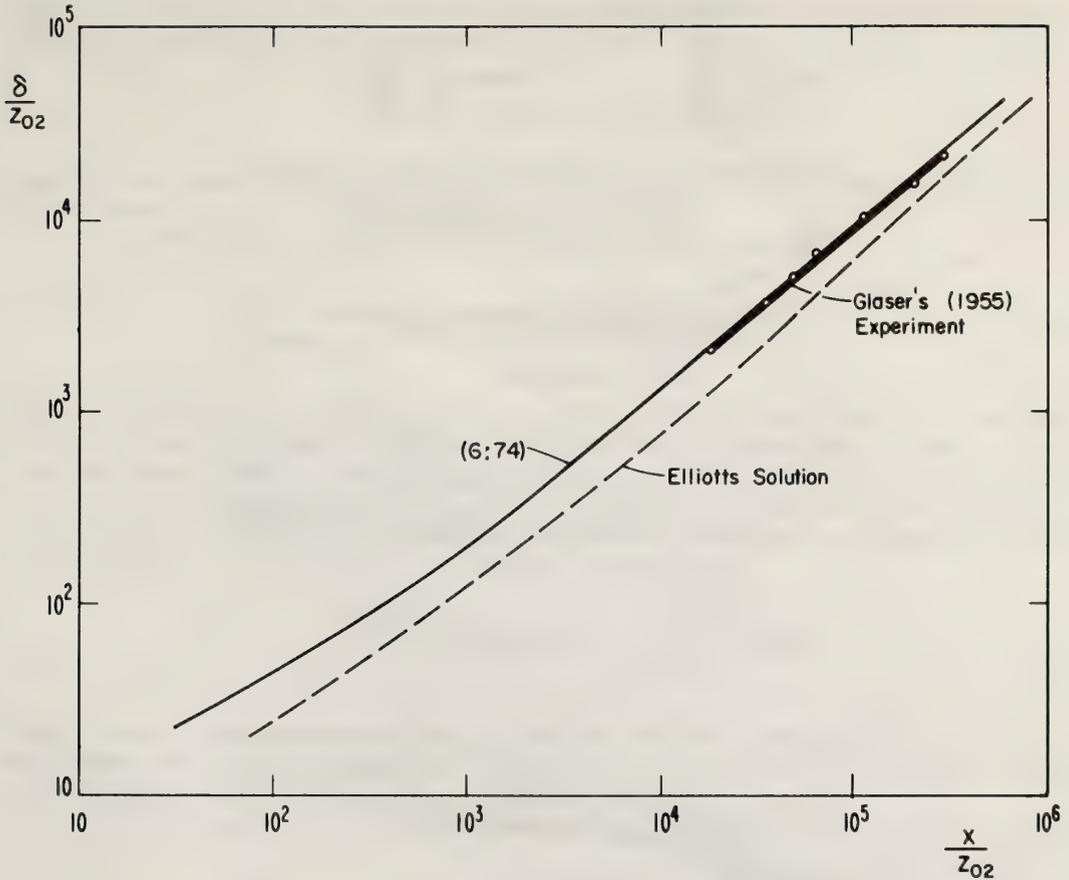


Figure 6.22 Height of internal boundary layer under neutral conditions (after Miyake, 1965).

Thus

$$f(\zeta) = 1.3 \{1.6 + (-2\zeta)^{2/3}\}^{1/2}$$

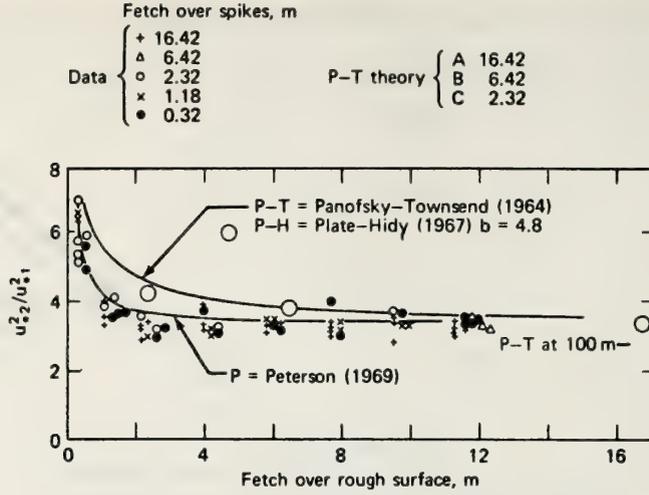
and because

$$\frac{d\delta}{dx} = \frac{k u_* z_2 f(\zeta)}{u_\delta(\zeta)} \tag{6:81}$$

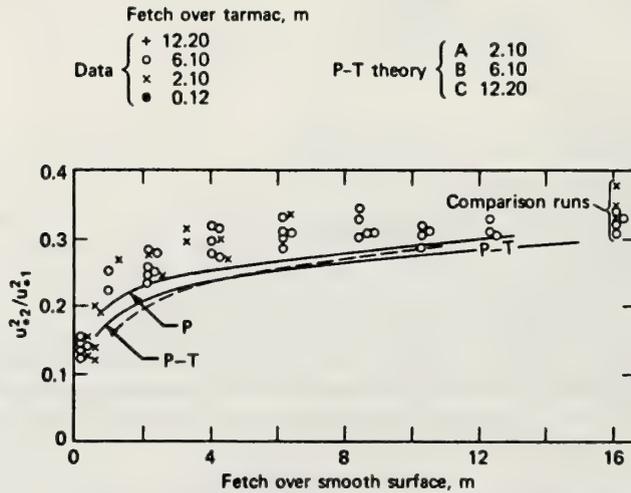
σ may be found by integrating this expression after substitution of (6:39). This has been done numerically and is displayed in (F6.24).

6.7.3 Air Modification

So far we have considered only the growth of the internal boundary in the surface layer. It is clear that new complexities are encountered when the entire boundary layer is considered. The problem then becomes essentially three-dimensional and simple analytical solutions are out of the question.



(a) Variation of surface shearing stress downwind of smooth-rough transition ($z_0 = 0.002$ cm to $z_0 = 0.25$ cm).



(b) Variation of surface shearing stress downwind of rough-smooth transition ($z_0 = 0.25$ cm to $z_0 = 0.002$ cm).

Figure 6.23 Variation of surface shearing stress downwind of smooth-rough transition ($z_0 = 0.0002$ cm to $z_0 = 0.25$ cm).

However, one exception of a simple case exists in the problem of air modification, when cold continental air flows over a warm sea or lake surface. Assumptions made in this case are steady state and uniform surface temperature. Equation (6:54) reduces to

$$F_h(0) = \int_0^{h_0} c_p \rho v \frac{\partial T}{\partial x} dz \tag{6:82}$$

and since θ , and $\partial\theta/\partial x$ are uniform with height, this equation may be integrated approximately to

$$F_h(0) \cong c_p \bar{\rho} v \frac{T}{\theta} \frac{\partial \theta}{\partial x} h_0 \tag{6:83}$$

where the bar denotes the average over the boundary layer. Furthermore by extrapolating (6:40a) to $-\xi \gg 1$ and rearranging the terms the heat flux may be expressed by

$$F_h(0) = c_p \rho u_* (\theta_o - \theta) \left\{ \ln \left(\frac{4}{9} - \frac{L}{z_o} \right) \right\}^{-1} \tag{6:84}$$

where θ is the potential temperature of the bulk of the boundary layer and θ_o the potential temperature of the surface.

By equating (6:83) and (6:84) a differential equation in $\theta(x)$ is obtained

$$\frac{\partial \theta}{\partial x} h_o = b (\theta_o - \theta) \tag{6:85}$$

where $b \equiv \rho u_* / \rho v (\theta/T) \left\{ \ln (-4/9 [L/z_o]) \right\}^{-1}$ a coefficient which varies slowly with $(\theta_o - \theta)$ and therefore with x . The height h_o depends on the temperature distribution of the original cold air before it reaches the ocean and on the change in potential temperature $\theta(x)$ which occurs with increasing distance, x , to the shore. This last difference is expressed by $\theta(x) - \theta(o)$ which must be distinguished from $\theta(x) - \theta_o$, the potential temperature difference between the boundary layer and the surface. Because the potential temperature must be continuous at h_o , this height may be expressed by

$$h_o = \frac{\theta(x) - \theta(o)}{\Gamma - \gamma} \frac{T}{\theta} \tag{6:86}$$

where Γ represents the adiabatic lapse rate and γ the lapse rate of the original cold air. Note that h_o here has the same role as the height of the internal boundary in the previous sections. The basic difference is that the thermal structure entirely determines the height of the modified layer.

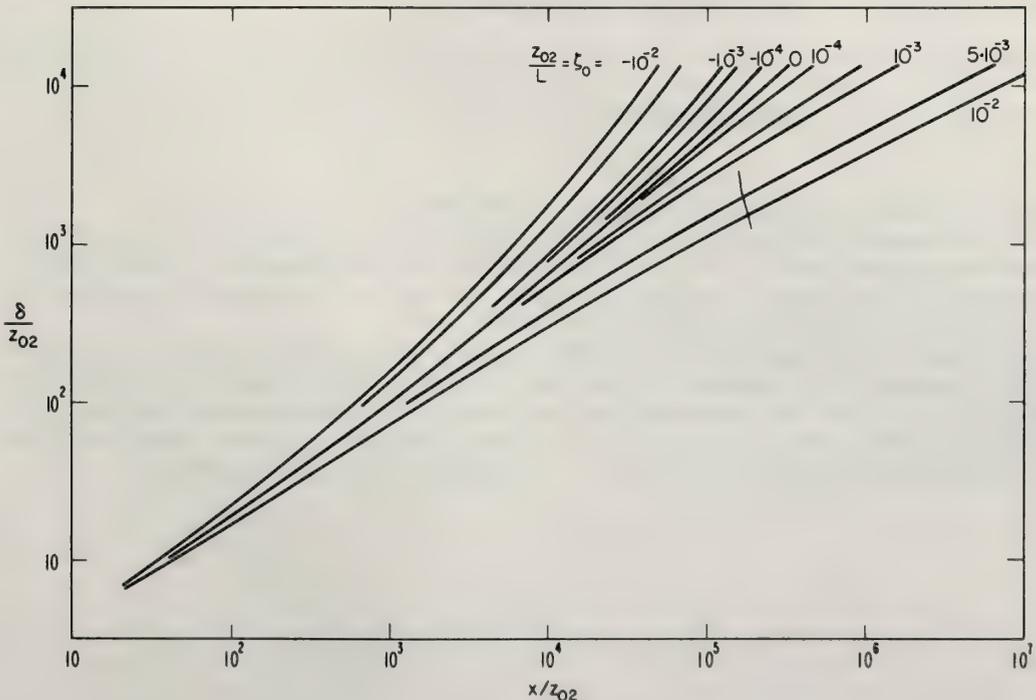


Figure 6.24 Height of internal boundary layer under diabatic conditions.

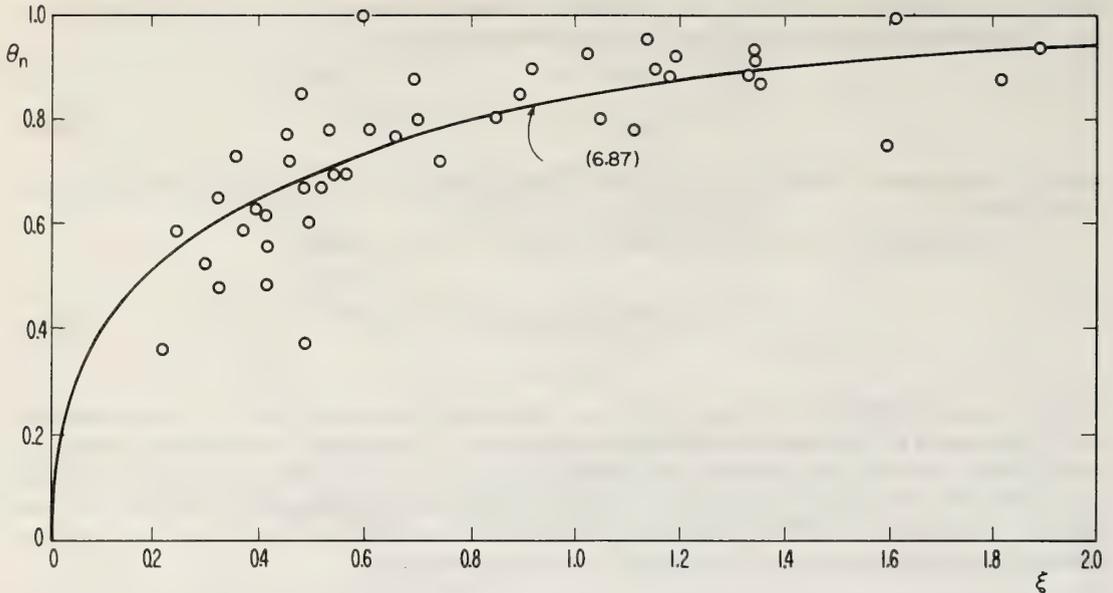


Figure 6.25 Air modification of cold air over a warm surface. (Businger, 1954; data from Burke, 1945).

Upon substituting (6:86) into (6:85) and by assuming that b is a constant (6:85) may be integrated to

$$\theta_n + \ln(1 - \theta_n) = \xi \quad (6:87)$$

where $\theta_n \equiv \frac{\theta(x) - \theta(o)}{\theta_o - \theta(o)}$, a normalized potential temperature

and $\xi \equiv \frac{b(\Gamma - \gamma)}{[\theta(o) - \theta_o]} \frac{\theta}{T} x$, a normalized distance.

Figure (6.25) illustrates how well (6:87) predicts actual observations.

A similar analysis can be given for the change in water vapor content when the cold air moves over the warm water surface, however, with the additional difficulty that the mixing ratio or specific humidity does not have to be continuous at the height h_o . Usually there will be a discontinuity of humidity at this height going from relatively moist modified air to dry undisturbed air.

It is not too difficult to extend the discussion to more complex situations including a more realistic temperature sounding of the undisturbed air, variations in the sea surface temperature and the variation of coefficient b with distance. The experiments that are planned during the International Field Year of the Great Lakes and which will be carried out over Lake Ontario this year may provide valuable data on the type of air modification discussed here, and enable us to develop a useful numerical scheme.

For a complete theory of air modification and the formation of an internal boundary layer, we must be able to predict the effects of simultaneous changes of heat flux and roughness, as well as account for the diurnal cycle of heating, etc. Also we must be able to incorporate the complexities of the terrain. An illustration of a fairly complicated situation is given in (F6.26) in which air modification of a cold lake breeze flowing over Chicago is illustrated by a series of successive temperature soundings. (From Carson and Nelson, 1969).

For a summary of recent efforts to develop numerical models of the internal boundary layer the reader is referred to Plate (1971).

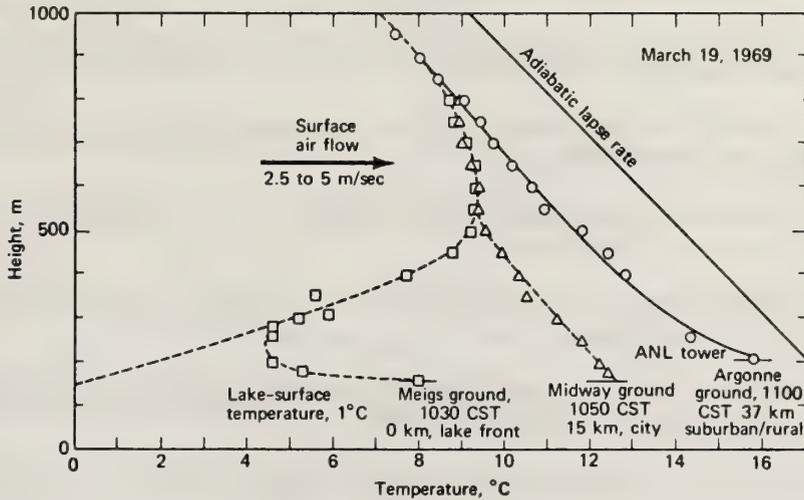


Figure 6.26 Example of a thermal internal boundary layer: vertical temperature profiles in Chicago during an onshore wind from Lake Michigan, Mar. 19, 1969 (from Carson and Nelson, 1969). Δ , Midway Airport. O , Argonne National Laboratory. \square , Meigs Airport (after Plate, 1971).

6.8 References

- Arya, S. P. S. (1972), The condition for the maintenance of turbulence in stratified flows, *Quart. J. Roy. Met. Soc.*, *98*.
- Ball, F. K. (1960), Control of inversion height by surface heating, *Q. J. Roy. Met. Soc.*, *86*, 983-994.
- Beals, G. A. (1970), An investigation of the critical Richardson number for turbulence in the lower atmosphere, Ph.D. thesis, Univ. of Washington, 134 pp.
- Bergman, K. H. (1969), On the dynamic stability of convective atmospheric vortices, Ph.D. thesis, University of Washington.
- Blackadar, A. K. (1962), The vertical distribution of wind and turbulent exchange in a neutral atmosphere, *J. Geophys. Res.*, *67*, 3095-3102.
- Blackadar, A. K., and H. Tennekes (1968), Asymptotic similarity in neutral barotropic planetary boundary layers, *J. Atmos. Sci.*, *25*, 1015-1020.
- Bradley, E. F. (1968), A micrometeorological study of velocity profiles and surface drag in the region modified by a change in roughness, *Quart. J. Roy. Met. Soc.*, *94*, 361-379.
- Bradshaw, P., D. H. Ferris, and N. P. Atwell (1967), Calculation of boundary layer development using the turbulent energy equation, *J. Fluid Mech.*, *28*, 593-616.
- Brown, R. A. (1970), A secondary flow model for the planetary boundary layer, *J. Atmos. Sci.*, *27*, 742-757.
- Brown, R. A. (1971), On the Inflection point instability of a stratified Ekman boundary layer, NCAR MS 71-117.
- Brown, R. A., and F. Lee (1972), A shooting method application to the stability problem for a stratified rotating boundary layer. NCAR MS 71-164.
- Buajitti, K., and A. K. Blackadar (1957), Theoretical studies of diurnal wind structure variations in the planetary boundary layer. *Quart. J. Roy. Meteor. Soc.*, *83*, 486-500.
- Burke, C. J. (1945), Transformation of polar continental air to polar maritime air. *J. Meteor.* *2*, 94-112.
- Businger, J. A. (1954), Some aspects of the influence of the earth's surface on the atmosphere, *Med. Verh. Kon. Ned. Meteor. Inst.*, No. 61, 78 pp.
- Businger, J. A. (1955), On the structure of the atmospheric surface layer, *J. Meteor.* *12*, 553-561.
- Businger, J. A. (1966), Transfer of momentum and heat in the planetary boundary layer, *Proc. Symp. Arctic Heat Budget and Atmospheric Circulation*, The RAND Corporation, Santa Monica, 305-322.
- Businger, J. A. (1969), Note on the critical Richardson number(s), *Quart. J. Roy. Met. Soc.* *95*, 653-654.
- Businger, J. A. (1971), Comments on "Free convection in the turbulent Ekman layer of the Atmosphere", *J. Atmos. Sci.* *28*, 298-299.
- Businger, J. A. (1972), "A note on free convection" to be published in *Boundary Layer Meteorology*.

- Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley (1971), Flux profile relationships in the atmospheric surface layer, *J. Atmos. Sci.*, *28*, 181-189.
- Carroll, J. J., and J. A. Ryan (1970), Atmospheric vorticity and dust devil rotation, *J. Geophys. Res.*, *75*, 5179-5184.
- Carson, J. E. and D. M. Nelson (1969), Chicago aircraft sounding program, in Radiological Physics Division Annual Report, July 1968-June 1969, USAEC Report ANL-7615, pp. 135-142, Argonne Natl. Lab.
- Ching, J. C. and J. A. Businger, (1968), The response of the planetary boundary layer to time varying pressure gradient force, *J. Atmos. Sci.*, *25*, 1021-1025.
- Clarke, R. H. (1970), Observational studies in the atmospheric boundary layer, *Quart. J. Roy. Met. Soc.*, *96*, 91-114.
- Csanady, G. T. (1967), On the resistance law of a turbulent Ekman layer, *J. Atmos. Sc.* *24*, 467-471.
- Deardorff, J. W. (1971), Investigation of neutral and unstable planetary boundary layers, NCAR Manuscript No. 71-105.
- Deardorff, J. W. (1970), Preliminary results from numerical integrations of the unstable planetary boundary layer, *J. Atmos. Sci.*, *27*, 1209-1211.
- Deardorff, J. W. (1970), Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection, *J. Atmos. Sci.*, *8*, 1211-1213.
- Deardorff, J. W., G. E. Willis, and D. K. Lilly (1969), Laboratory investigation of non-steady penetrative convection, *J. Fluid Mech.*, *35*, 7-31.
- Deardorff, J. and G. Willis (1967), Investigations of turbulent convection between horizontal plates, *J. Fluid Mech.*, *28*, 675-704.
- Dyer, A. J. (1965), The flux gradient relation for turbulent heat transfer in the lower atmosphere, *Quart. J. Roy. Meteor. Soc.*, *91*, 151-157.
- Elliott, W. P. (1958), The growth of the atmospheric internal boundary layer, *Trans. Amer. Geophys. Union*, *39*, 1048-1054.
- Elliott, W. P. (1966), Daytime temperature profiles, *J. Atmos. Sc.*, *23*, 678-681.
- Ellison, T. H. (1957), Turbulent transport of heat and momentum from an infinite rough plate, *J. Fluid Mech.*, *2*, 456-466.
- Frisch, A. S. and J. A. Businger (1970), A study of convective elements in the atmospheric boundary layer, to be published in *Boundary Layer Meteorology*.
- Glaser, A., (1955), The temperature above an airport runway on a hot day, A and M College of Texas, Dept. of Ocean., Reports No. 3 and 4.
- Goldstein, S., (1931), On the stability of superposed streams of fluids of different densities, *Proc. Roy. Soc. A* *132*, 524-548.
- Hamel, G. (1943), Streifenmethode und Ähnlichkeits betrachtungen zur turbulente Bewegung, *Abhlg. Preus. Akad. Wiss., Math naturwiss. Klasse no. 8*, 24 pp.

- Haugen, D. A., J. C. Kaimal, and E. F. Bradley, An experimental study of Reynolds stress and heat flux in the atmospheric surface layer. *Quart. J. Roy. Meteor. Soc.*, *97*, 168-180.
- Hines, C. O. (1971), Generalizations of the Richardson criterion for the onset of atmospheric turbulence, *Quart. J. Roy. Meteor. Soc.* *97*, 429-439.
- Hinze, J. O., 1959: *Turbulence*, McGraw-Hill, 586 pp.
- Kaimal, J. C. (1966), An analysis of sonic anemometer measurements from the Cedar Hill tower. Environmental Res. Paper No. 215, AFCRL-66-542.
- Kaimal, J. C. and J. A. Businger (1970), Case studies of a convective plume and a dust devil, *J. Appl. Met.*, *9*, 612-620.
- von Karman, Th. (1921), Über laminare and turbulente Reibung, *ZAMM*, *1*, 233.
- Kazanski, A. B., and A. S. Monin (1957), The form of smoke jet, *Izv. Acad. Nank. USSR (ser. geofiz.)* no. 8, 1020-1033.
- Lettau, H., 1959: Wind profile, surface stress, and geostrophic drag coefficients in the atmospheric surface layer, *Advances in Geophysics*, Vol. 6, Academic Press, New York, pp. 241-257.
- Lettau, H. (1962), Theoretical wind spirals in the boundary layer of a barotropic atmosphere, *Beitr. Phys. Atm.*, *35*, No. 3-4.
- Lilly, D. K. (1966), On the stability of Ekman boundary flow, *J. Atmos. Sci.*, *23*, 481-494.
- Ludlam, F. H. (1967), Characteristics of billow clouds and their relation to clear-air turbulence, *Quart. J. Roy. Met. Soc.*, *93*, 419-435.
- Lumley, J. L. and H. A. Panofsky, 1964: *The structure of Atmospheric Turbulence*, Wiley Interscience, New York, 239 pp.
- Miles, J. W. (1961), On the stability of heterogeneous shear flows, *J. Fluid Mech.*, *11*, 284-290.
- Miyake, M. (1965), Transformation of the atmospheric boundary layer over inhomogeneous surfaces, Scientific Report, Dept. of Atmos. Sci., Univ. of Wash.
- Monin, A. S. (1959), Smoke propagation in the surface layer of the atmosphere, *Advances in Geophysics*, Vol. 6, Academic Press, New York, 331-341.
- Nikuradse, J. (1933), Strömungsgesetze in rauhen Röhren, *Forschungsheft* 361.
- Obukhov, A. M. (1946), Turbulence in an atmosphere with a non-uniform temperature, *Trudy Inst. Teoret. Geofis. AN SSSR* no. 1. (Translation in: *Boundary Layer Meteorology*, *2*, 7-29).
- Panofsky, H. A., and A. A. Townsend (1964), Change of terrain roughness and the wind profile, *Quart. J. Roy. Meteorol. Soc.*, *90*, 147-155.
- Pandolfo, J. (1966), Wind and temperature profiles for constant flux boundary layers in lapse conditions with a variable eddy conductivity to eddy viscosity ratio, *J. Atmos. Sc.* *23*, 495-502.
- Paulson, C. A. (1970), The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer, *J. Appl. Meteor.*, *9*, 857-861.

- Peterson, E. W. (1969), Modification of mean flow and turbulent energy by change in surface roughness under conditions of neutral stability, *Quart. J. Roy. Meteorol. Soc.*, *90*, 561-576.
- Plate, E. J., 1971: *Aerodynamic characteristics of atmospheric boundary layers*, U. S. Atomic Energy Comm., Div. of Tech. Inf. 25465, U. S. Dept. of Commerce, Springfield, Va., 22151, 190 pp.
- Plate, E. J. and G. M. Hidy (1967), Laboratory study of air flowing over a smooth surface onto small water waves, *J. Geophys. Res.* *72*, 4627-4641.
- Pohlhausen, K. (1921), Zur näherungsweise Integration der Differentialgleichungen der laminaren Reibungsschicht, *ZAMM*, *1* 252.
- Prandtl, L. (1932), Meteorologische Anwendungen der Strömungslehre, *Beitr. Phys. frei. Atm.*, *19*, 188-202.)2.
- Priestley, C. H. B. (1954), Convection from a large horizontal surface, *Australian J. Phys.* *6*, 279-290.
- Richardson, L. F. (1920), The supply of energy to and from atmospheric eddies, *Proc. Roy. Soc. A* *97*, 354-373.
- Ryan, J. A. and J. J. Carroll (1970), Dust devil wind velocities: Mature state, *J. Geophys. Res.*, *75*, 531-542.
- Schlichting, H., 1955: *Boundary layer theory*, McGraw-Hill, New York, 535 pp.
- Sinclair, P. C. (1969), General characteristics of dust devils, *J. Appl. Meteorol.*, *8*, 32-45.
- Sobolev, S. L. (1964) *Partial differential equations of mathematical physics*. Eng. Transl. Oxford, Pergamon Press, or Reading, Mass., Addison-Wesley, 427 pp.
- Sutton, O. B., 1953: *Micrometeorology*, McGraw-Hill, 333 pp.
- Swinbank, W. C. (1964), The exponential wind profile, *Quart. J. Roy. Met. Soc.*, *90*, 119-135.
- Swinbank, W. C., and A. J. Dyer, (1968), *Micrometeorological Expeditions, 1962-1964*, Div. of Meteor. Phys. Tech. Paper No. 17, CSIRO, Australia.
- Taylor, G. I. (1931), Effect of variation in density on the stability of supercooled streams of fluids, *Proc. Roy. Soc.*, *A 132*, 499-523.
- Telford, J. W., and J. Warner (1964), Fluxes of heat and vapor in the lower atmosphere derived from aircraft observations, *J. Atmos. Sci.*, *21*, 539-548.
- Tennekes, H. (1968), Outline of a second-order theory of turbulent pipe flow, *AIAA Journal*, *6*, 1735-1740.
- Tennekes, H. (1970), Free convection in the turbulent Ekman layer of the atmosphere, *J. Atmos. Sci.* *27*, 1027-1034.
- Townsend, A. A., 1956: *The structure of turbulent shear flow*, Cambridge University Press, 315 pp.
- Townsend, A. A. (1958), Turbulent flow in a stably stratified atmosphere, *J. Fluid Mech.*, *3*, 361-372.
- Tsvang, L. R. and Y. A. Volkov (1967), *Iz. Atmos. Oceanic Phys.*, *3*, 790-792.

- Webb, E. K. (1970), Profile relationships: The log-linear range, and extension to strong stability, *Quart. J. Roy. Meteorol. Soc.* 96, 67-80.
- Woods, J. D. (1969), On Richardson's number as a criterion for laminar-turbulent-laminar transition in the ocean and the atmosphere, *Radio Sci.*, 4, 1289-1298.
- Wyngaard, J. C., O. R. Cote and Y. Izumi (1971), Local free convection, similarity, and the budgets of shear stress and heat flux, *J. Atmos. Sci.*, 28, 1171-1182.

List of Symbols

$a_1 - a_8$	constants	u_*	friction velocity
c	phase speed	u, v, w	components of velocity in x, y, z directions respectively
c_p	specific heat at constant pressure	$u_f \equiv \left(\frac{T}{gz} \overline{w'T'} \right)^{1/3}$	
e	$\frac{1}{2} (u'^2 + v'^2 + w'^2)$	u_{f0}	convection velocity
e_2	kinetic energy of secondary flow $= \frac{1}{2} (v^2 + w^2)$	\vec{v}	velocity vector
f	Coriolis parameter	V	mean horizontal windprofile in boundary layer
F_h	heat flux	$W \equiv (u + iv)e^{ift}$	
F_n	net radiation flux	w_*	free convection friction velocity
g	acceleration due to gravity	z_0	roughness length
h	height of boundary layer	$\alpha \equiv K_h/K_m$	
$h_n \equiv u_*/f$		β	parameter in Orr-Sommerfeld eq.
h_0	height of the level where the heat flux is zero	δ	scaling height of Ekman spiral; height of internal boundary layer
k	von Karman constant	ϵ	rate of dissipation of turbulent energy
K_h	eddy thermal diffusivity	$\zeta \equiv - \frac{g}{T} \frac{\overline{w'T'} k(z + z_0)}{u_*^3}$	
K_m	eddy viscosity (turbulent transfer coefficient)	θ	potential temperature
$L \equiv - \frac{T}{g} \frac{u_*^3}{k w'T'}$		$\theta_* \equiv - \frac{\overline{w'T'}}{u_*}$	
ℓ_2	size of roughness elements	$\theta_f \equiv \left\{ \frac{T}{gz} \overline{(w'T')^2} \right\}^{-1/3}$	
$N \equiv \left(\frac{g}{\theta} \frac{\partial \theta}{\partial z} \right)^{1/2}$	Brunt-Väissälä frequency	λ	roll wavelength
p	pressure	ν	kinematic viscosity
R	gas constant	ρ	density
Re	Turbulent Reynolds number $v_g \delta / K_m = v_g \left(\frac{2}{fK} \right)^{1/2}$	τ	shear stress
$Ri \equiv \frac{g}{\theta} \frac{\partial \theta / \partial z}{(\partial \bar{u} / \partial z)^2}$		ϕ	perturbation stream function
$R_f \equiv - \frac{g}{T} \frac{\overline{w'T'}}{u'w' \partial \bar{u} / \partial z}$		$\phi_\epsilon \equiv \frac{kz\epsilon}{u_*^3}$	
R_n	rate of radiative absorption per unit volume	$\phi_h \equiv \frac{k}{\theta_*} \frac{\partial \bar{\theta}}{\partial z}$	
Ro	Rosby number	$\phi_m \equiv \frac{kz}{u_*} \frac{\partial \bar{u}}{\partial z}$	
T	temperature	$\vec{\Omega}$	the angular velocity of the earth's rotation

Chapter 7 ATMOSPHERIC GRAVITY WAVES IN THE PLANETARY BOUNDARY LAYER

William H. Hooke[†]

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Theory and observation of atmospheric gravity waves in the planetary boundary layer, estimates of their energy and momentum transport, as well as their interactions with each other and other atmospheric motions, are briefly reviewed.

7.0 Introduction

The two previous chapters have introduced, in a rather general way, the subject of atmospheric gravity waves and described in broad terms the structure and dynamics of the planetary boundary layer. Chapter 5 has provided a statement of what an atmospheric gravity wave is, described the propagation of plane waves of this class, and considered complications in this description introduced by nonlinear and dissipative processes and by the background temperature and wind structure of the atmosphere in which the waves propagate. Chapter 6 has described this structure as it is found in the planetary boundary layer, enumerated the important boundary layer processes, and outlined the theoretical framework used in their description. The material of these two chapters, as well as data recently gathered by the remote sensing techniques described elsewhere in this volume, indicates that atmospheric gravity waves are a major feature of the planetary boundary layer as well as the free atmosphere, and that the wind and temperature structure of the boundary layer and the atmosphere immediately above it modify the wave propagation. For example, the waves of largest amplitude at ground level are often those whose energy is trapped or ducted there by this structure. The "critical level" phenomenon discussed in Chapter 5, involving wave propagation to a level where the background wind velocity equals the wave phase speed, may also be an important aspect of wave propagation in the boundary layer.

It should be remembered, however, that the latter phenomenon results in momentum and energy exchange between the wave and the background flow, and thus it must be considered a potentially important boundary-layer process as well. In addition, the boundary layer is turbulent, this turbulence giving rise to an eddy viscosity which acts in a fashion analogous to molecular viscosity (and much more effectively, at tropospheric levels) to dissipate the waves. As the waves dissipate, their energy and momentum must be converted into heat and into kinetic energy and momentum of the mean boundary-layer flow. (It is possible, in fact, though by no means confirmed, that such momentum deposition by the waves may contribute in part to the wind profile commonly known as the nocturnal jet, and thought to be the result of an inertial oscillation [e.g., Blackadar, 1957].) Near the surface, where the Richardson number is small, the turbulence may interact with the waves in such a way as to lead to energy exchanges between the turbulence, the waves and the background wind which result in explosive growth in both wave amplitudes and turbulence intensities at the expense of the mean flow (Chimonas, 1972).

In addition to playing a role in momentum and energy transport processes internal to the boundary layer, atmospheric gravity waves provide a means for energy and momentum exchange between the boundary layer and the free atmosphere above it. Lee waves, a form of atmospheric gravity wave generated by air flow over mountains, have been shown to exert a powerful drag on that flow, contributing a significant fraction of the vertical momentum transport required for global momentum balance (e.g., Bretherton, 1969; Lilly, 1971). Turbulence in the unstable daytime boundary layer generates gravity waves in the inversion layer above it through the agency of penetrative convection (Townsend, 1966, 1968; Gossard et al., 1971), the process permitting a certain vertical energy transport from the boundary layer.

[†]Lecturer, Department of Astro-Geophysics, University of Colorado; Fellow, Cooperative Institute for Research in the Environmental Sciences (CIRES), a joint institute of the University of Colorado and NOAA.

Thus we acknowledge the existence of atmospheric gravity waves in the planetary boundary layer and their modification by its structure, but we also emphasize that the waves in turn play a part in the momentum and energy transport processes fundamental to boundary layer physics, perhaps determining to a substantial degree the nature of the background in which they propagate. The goal of the present chapter, then is to synthesize and enlarge upon the material of the previous two in order to describe the nature of atmospheric gravity wave propagation in the planetary boundary layer and to elucidate the specific role these waves play in boundary-layer dynamics. Because this subject is in its infancy, however, such an attempt must necessarily fail; where it does, we must for the moment content ourselves with defining the major areas of uncertainty and indicating experiments or theoretical developments that would improve our understanding.

7.1 Mean Temperature and Wind Structure in and above the Stable Planetary Boundary Layer

We take as our starting point a description of the mean or background temperature and wind structure in the stable planetary boundary layer and in the free atmosphere immediately above it, since this structure strongly influences the nature of the wave propagation. Figure(7.1) shows the temperature and wind structure of the boundary layer itself, described in terms of appropriate dimensionless variables, as determined at temperate latitudes in experimental studies by Clarke (1970). The temperature profile results from radiational cooling of the earth's surface at night in conjunction with heat conduction to the surface both from the soil and from the lowermost atmospheric layers, a process described in detail in meteorological texts. The normalized potential temperature, $\bar{\theta}$, increases monotonically with increasing height, implying static stability (Chapter 3), but the stability itself decreases with increasing height. The normalized wind speed \bar{u} also increases with height. The actual wind speed u is of course highly variable, but is typically of the order of 10m/s to 20m/s at a height of 1 km. The veering of the wind, as represented by \bar{v} , is appreciable and must be taken into account in experimental studies, but the vector nature of the dynamical problem will not be considered further here. Strong solar heating during the daytime typically destroys the stability of the boundary layer (as explained in Chapter 6), and may produce a layer characterized by strong turbulent convection. During the day, this convection pushes the top of the boundary layer to heights of a kilometer or so, where it is often capped by an elevated inversion layer at a height of something less than 2 km. The inversion is typically sufficiently stable to withstand further erosion by the turbulence of the unstable, daytime boundary layer (Charnock and Ellison, 1967; Sheppard, 1970). At night, the stable planetary boundary layer is usually only a few hundred meters thick. Between the top of the boundary layer (or the elevated inversion immediately above it) and the tropopause, the troposphere is, on the average, only very slightly stable, and wind speed generally increases with increasing height (to speeds as high as 100m/s or more in the vicinity of jet streams).

One function of the temperature and wind profiles of special interest from the dynamical standpoint (see 7.4) is the so-called gradient Richardson number, defined as

$$Ri \equiv g \frac{d(\bar{\rho}\bar{\theta})/dz}{(d\bar{U}/dz)^2 + (d\bar{V}/dz)^2} \quad (7:1)$$

where z is the height, g is the acceleration of gravity, $\bar{\theta}$ is the mean potential temperature, and \bar{U} is the mean wind speed. Alternatively, we may write

$$Ri = \frac{N^2}{(d\bar{U}/dz)^2 + (d\bar{V}/dz)^2} \quad (7:2)$$

where $N^2 = g d(\bar{\rho}\bar{\theta})/dz$ is the square of the so-called Brunt-Väisälä frequency, a generalization of ω_g as defined in the isothermal case by (5:16). Note that N^2 , so defined, may be negative. For $Ri < 0$, the atmosphere is statically unstable, permitting thermal convection. There is general theoretical and observational agreement that for $0 < Ri \ll 1$, the atmosphere is dynamically unstable, encouraging the exponential growth of wave motions including those of the gravity-wave type, but statements more specific than this have proven difficult to verify. Thus while Miles (1961), Howard (1961), and Chimonas (1970) have shown theoretically that horizontal flows with vertical shear are stable provided that $Ri \geq 1/4$ everywhere, Hines (1971) has suggested that this criterion is too restrictive when applied to non-horizontal flows, and as yet no one has been able to prove (or disprove) that in general the condition that $0 < Ri < 1/4$ somewhere in the flow is *sufficient* for dynamical instability. On the observational side, difficulties in measuring the required parameters, as well as debate over just what spatial and temporal averages we should use in calculating Richardson numbers, have

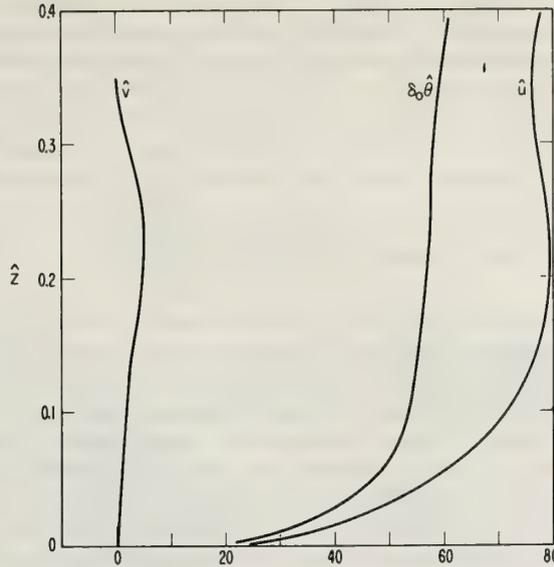


Figure 7.1 Mean stable boundary-layer profiles of $\hat{u}(=\bar{U}/u_*)$, $\hat{v}(=\bar{V}/u_*)$, and $\hat{\theta}(=(\theta - \theta_0)/T_*)$ in terms of non-dimensional height fz/u_* . Here \bar{U} and \bar{V} are velocity components parallel to rectangular coordinate axes x and y (z vertical), θ is the potential temperature, θ_0 is the potential temperature at the roughness height z_0 , u_* is the friction velocity, and T_* is a standard scaling temperature (see Chapter 6) (from Clarke, 1970).

prevented our learning much more than the minimal information that regions of turbulence are typically regions of small Ri. Clarke (1970) has computed the gradient Richardson number over z intervals of 0.05 (here $z \equiv fz/u_*$, where f is the coriolis parameter, z is the true height and u_* is the friction velocity) using the mean profiles shown in(F7.1), and concludes that in the stable case, $Ri > 1$ at heights above $\hat{z} > 0.05$ (~ 100 m or so, depending on conditions). The Richardson number decreases with decreasing height, however, and it is observed that in the stable surface boundary layer (the bottom few tens of meters), Ri tends to approach and fall below $1/4$ (Webb, 1970; Oke, 1970; Businger et al., 1971). This feature of the atmospheric boundary layer holds profound consequences for the nature of the interaction between waves, turbulence, and the mean flow (see 7.4.4).

7.2 Wave Propagation in Boundary-Layer Wind and Temperature Structure

Studies of wave propagation in the wind and temperature structure of the boundary layer and the atmosphere immediately above have been by no means exhaustive. What work has been done, mainly by E. E. Gossard and coworkers (see references), has concentrated on modeling the marine layer, the overlying inversion layer, and the remainder of the troposphere by means of two- and three-layer models in which background wind velocity and temperature gradient are held constant within each layer. Under such conditions, the linearized equations governing the wave motion in each layer yield solutions of the form $\exp i(kx \pm nz - \omega t)$ for the variables $U \equiv \rho_0^{-1/2}u$, $W \equiv \rho_0^{-1/2}p$, and $P = \rho_0^{-1/2}p$, where u and w are the wave-associated horizontal and vertical velocity perturbations, p is the pressure perturbation, and ρ_0 is the background density. Gossard found that the waves observed at his location (San Diego) had phase velocities small compared to the speed of sound and vertical skin depths or vertical wavelengths small compared to a scale height as would be expected for waves confined to relatively thin height ranges in the neighborhood of tropospheric temperature inversion layers. The propagation equation for the j^{th} layer could then be approximated by

$$n_j^2 = \frac{k^2}{\omega^2} (N_j^2 - \omega^2) \tag{7:3}$$

where N_j is the Brunt-Väisälä frequency of the j^{th} layer (compare 5:27). Certain boundary and interfacial conditions are then used to relate the n_j 's. Specifically, mass and momentum conservation require that parcel displacement and pressure be continuous across each interface. At the ground, which is usually taken to be plane and horizontal, the vertical velocity must vanish; at infinity, the wave energy must vanish for wave trapping. In the cases considered by Gossard at San Diego, the waves observed had phase speeds far in excess of the background wind speeds, so that the effect of the wind shear was negligible; the above continuity conditions thus required (see Gossard and Munk, 1954; Gossard, et al., 1970 for slightly different forms of this equation) that

$$n_2 \cot n_2(h_2 - h_1) = \frac{in_1 n_3 \cot(n_1 h_1) + n_2^2}{n_1 \cot(n_1 h_1) - in_3} \quad (7:4)$$

where h_1 is the depth of the boundary layer, and $h_2 - h_1$ is the thickness of the overlying inversion. Equation (7:4) is valid for both real and imaginary values of n_1 and n_2 ; n_3 must be imaginary for wave trapping.

At times, the elevated inversion is more stable than either the boundary layer below or the troposphere above it; in such instances there will be some waves (with frequencies such that $N_2 > \omega > N_1, N_3$) trapped in the inversion and evanescent above and below it. Gossard and Munk (1954) give dispersion diagrams for such waves (their figure 12) as well as plots of wave-associated velocity and pressure perturbation profiles (their figure 13) for specific values of the governing parameters. At other times, the boundary layer itself will be more stable than either of the two regions above it, and there will be some waves ($N_1 > \omega > N_3$) which will be trapped within the marine layer itself. Gossard, et al. (1970) give a dispersion diagram for a specific such case, reproduced here as (F7.2), based on a specific temperature sounding (F7.3), and obtain plots of wave-associated velocity and pressure perturbations for the fundamental mode (F7.4).

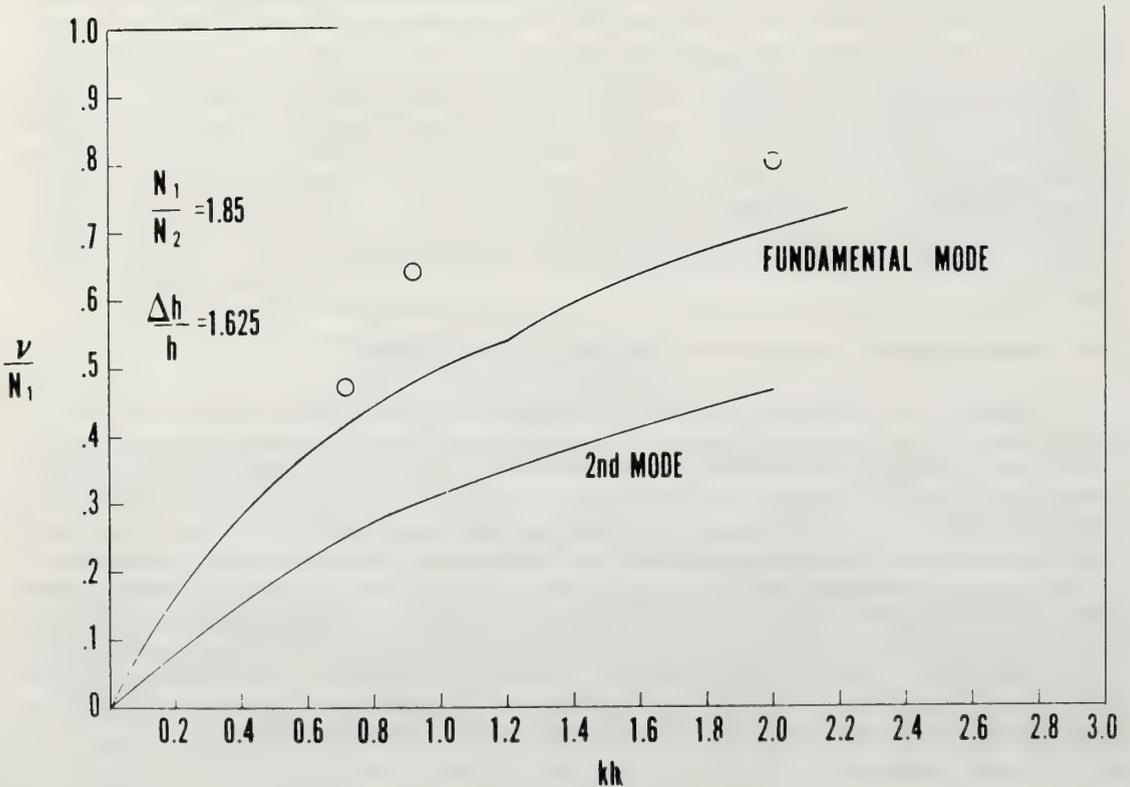


Figure 7.2 Dispersion relation for fundamental and second atmospheric gravity wave modes for the 3-layer atmospheric model shown in (F7.6). Circled points refer to the observed wave velocities for the events in (F7.3, F7.4, and F7.5) (from Gossard et al., 1970).

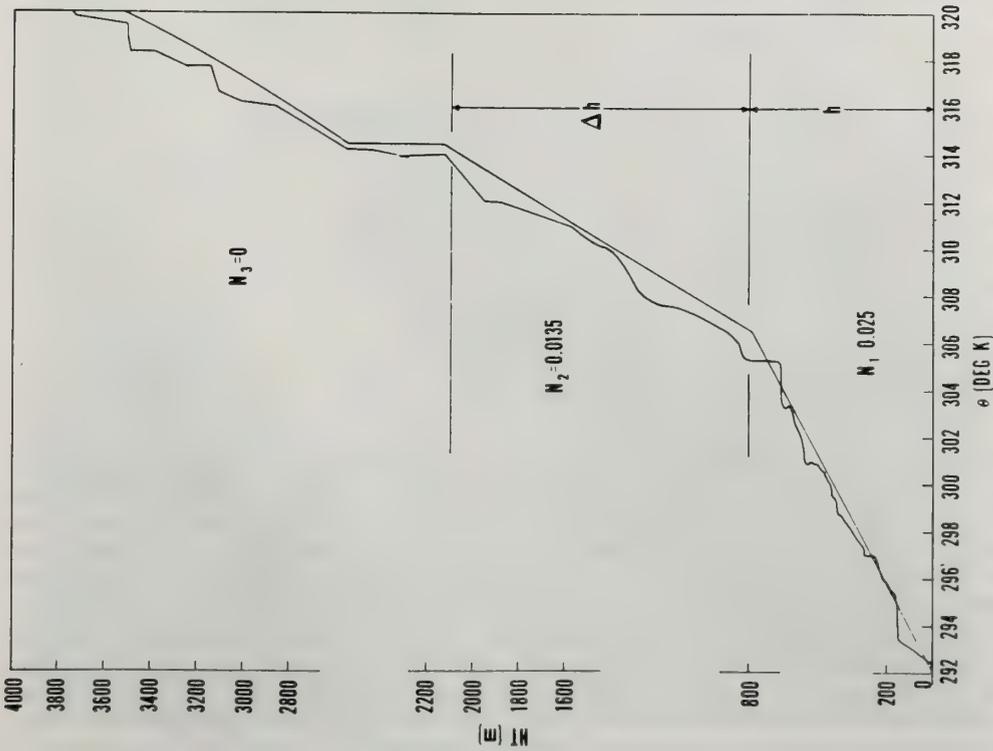


Figure 7.3 Potential temperature sounding for July 11, 1969. N_i are the Brunt-Vaisala frequencies for the three layers indicated by horizontal lines (from Gossard et al., 1970).

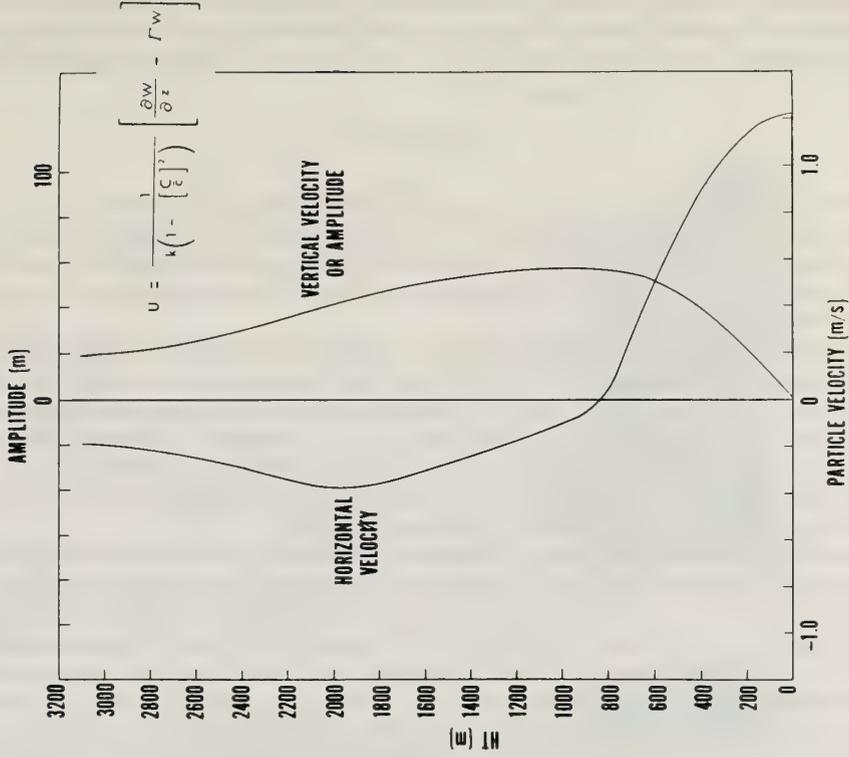


Figure 7.4 Horizontal and vertical parcel velocities calculated from the model of (F7.6) (from Gossard et al., 1970).

Gossard and his colleagues have observed atmospheric gravity waves in the boundary layer for years using a combination of pressure sensors, anemometers, and radiosondes, and, more recently, an FM-CW sounder. To estimate the horizontal phase velocity V_{ph} of the waves observed, they have used the "impedance relation"

$$V_{ph} = p/\rho_0 u, \quad (7:5)$$

which follows immediately from the horizontal component of (5:12). They then compared the observed phase velocities with those computed for trapped modes in two and three layer atmospheric models fitted to radiosonde data; computed and observed phase velocities were in good agreement, suggesting that the waves of largest amplitude observed at and near the earth's surface were typically those whose energy was trapped near inversions in the lower troposphere.

More recently, Gossard et al.,(1970) and Gossard and Richter (1970) have used the FM-CW radar (see Chapter 20) to provide additional insight in comparisons of this type and continue to find good agreement between theory and observation for both wave phase velocity and wave amplitude. In particular, Gossard and Richter (1970) show that the cusped shape of the waveforms thus observed is consistent with non linear theory developed by Hunt (1961) for the propagation of internal waves at an interface just above a solid boundary. Figures (7.5),(7.6), and (7.7) display FM-CW radar records and surface meteorological records for several atmospheric-gravity-wave events studied by Gossard et al. (1970).

For waves of small amplitude or waves occurring in the presence of turbulence, estimating the wave phase velocity from the impedance relation (7:3) is subject to considerable error because of the signal-to-noise problem; it is possible, however, by using instrument arrays of various types, to measure the wave phase speeds more directly. An example of the use of this procedure is shown in (F7.8), where an acoustic-sounder record (see Chapters 18 and 19) is compared with microbarograph records (see Chapter 21) from spaced stations arranged as shown in (F7.9). The white line superimposed on the acoustic-sounder record is the pressure trace from the microbarograph closest to the sounder; it is clear that the sounder and the microbarographs are indeed sensing the same wave phenomenon. The microbarograph records from all four stations are shown above and below the sounder record, where they have been time-shifted to provide maximum correlation during the periods indicated. The wave parameters thus deduced for the 0100-0200 LT and 0300-0330 intervals are shown in Table (7.1). These records and their interpretation are discussed in more detail by Hooke et al., (1972).

Table 7.1 Wave parameters determined from microbarograph data for March 3, 1971.

Wave Parameters	0100-0200 LT	0300-0330 LT
period	525 sec	265
horizontal phase speed	13 m/sec	6.5
horizontal wavelength	6.8 km	1.7
direction of arrival	270° E of N	295°
mean surface wind speed	1 m/sec	<0.2
mean surface wind direction	220° E of N	240°

From the work described above, we may infer that the small-scale temperature structure of the atmosphere is capable of trapping atmospheric gravity waves of small vertical wavelength and small phase velocity. Waves not having the proper phase velocities and wavelengths are by contrast free to propagate their energy into the upper atmosphere. Thus, out of a rather broad spectrum of waves which may be present in the atmosphere, the trapped portion of that spectrum may be expected to dominate observations made at ground level. Nevertheless, a number of questions remain unanswered by such studies. For example, we have learned nothing of the generation mechanisms for the broad spectrum of waves we have assumed to be present in order to account for the observations. Gossard and his coworkers also found that the waves seldom propagate in the direction of the background wind, and suggest that this is because such waves, finding critical levels in the mean flow, have their coherence destroyed by this interaction. Is this surmise correct? What about the waves trapped in the very stable boundary layer itself? To understand their origins and subsequent propagation, is it

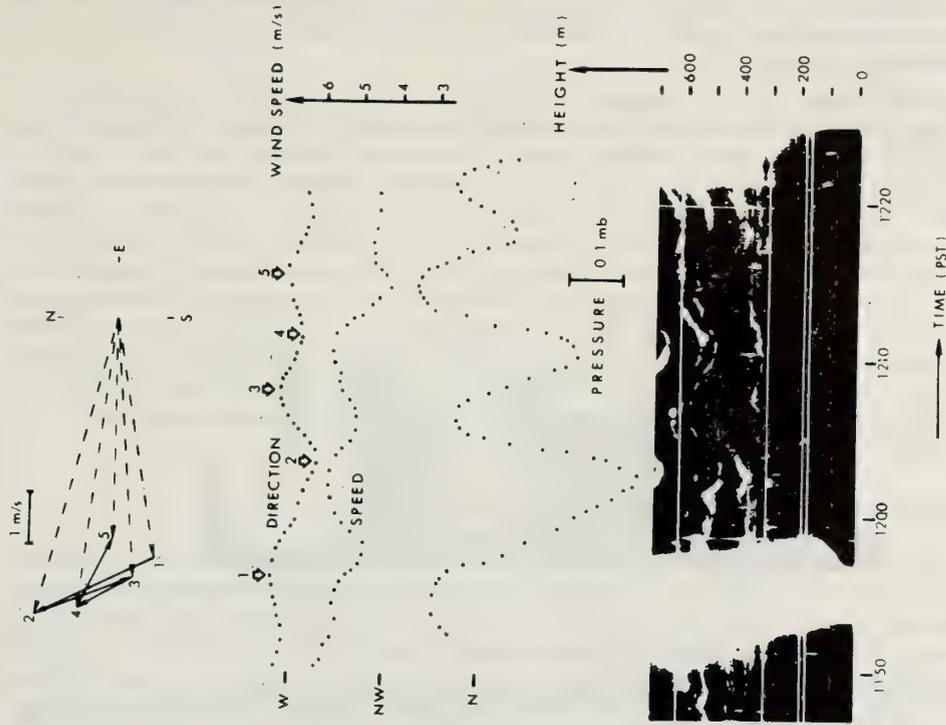


Figure 7.6 Radar sounding and surface meteorological records for July 11, 1969 (from Gossard et al., 1970).

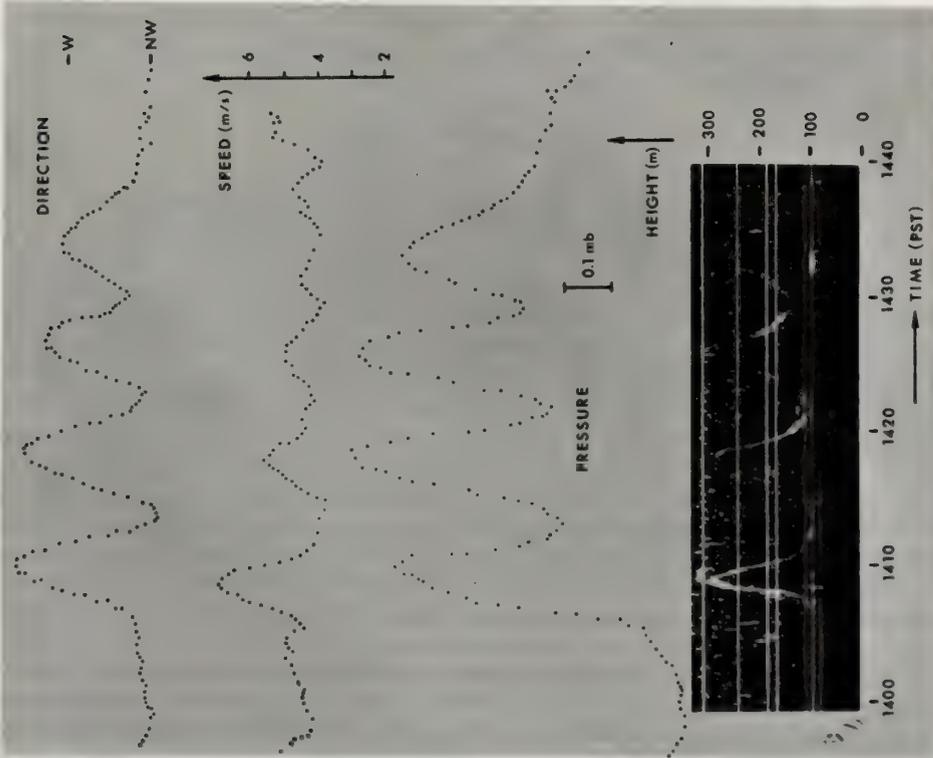


Figure 7.5 Radar sounding and surface meteorological records for November 13, 1969 (from Gossard et al., 1970).

sufficient to use average wind and temperature profiles of the type obtained by Clarke (see above), or is it necessary to consider small-scale, short-lived departures from these average profiles? Do the small-scale inversions seen in boundary layer temperature and wind profiles at low elevations persist long enough to permit small-scale wave trapping?

In closing this section on wave propagation in the boundary layer, we note that the boundary-layer eddy viscosity also affects wave propagation, particularly as it determines the minimum scale sizes for wave propagation (see 5:30 and the pertinent discussion). The momentum eddy exchange coefficient $\eta_e = \mu_e/\rho_o$ is highly variable, but should increase linearly from zero near the ground to between 0.1 and $10\text{m}^2/\text{s}$ (e.g., Monin and Zilitinkevitch, 1967) at heights of several tens of meters and then remain constant or decrease slightly with increasing height up to the top of the planetary boundary layer. For waves of ten-minute period, say, the corresponding minimum permitted wavelengths according to (5:30) then range between 20m and 200m, depending upon the value of η_e considered. These waves, however, do not even propagate for a distance equal to a single wavelength before their energy is almost completely dissipated. If we adopt instead the criterion that the waves must propagate a distance the order of the boundary-layer thickness ($\sim 1\text{ km}$) before their amplitudes fall to e^{-1} of their original values, then we find (using equation 19 of Pitteway and Hines 1963, for example) that the minimum permitted wavelengths are the order of 300m and 1500m respectively for the extreme values of η_e .

7.3 Wave Transport of Energy and Momentum in the Boundary Layer

We begin this section by defining wave energy and momentum, since these concepts were not introduced in Chapter 5. It is found that such definitions are not unambiguous; wave energy and momentum flux are second-order quantities, and there is not a priori reason why they should be expressible simply in terms of products of the first-order quantities of the perturbation theory given in Chapter 5; they may in fact include terms which are the products of second and zeroth order quantities. This problem has occasioned considerable discussion, notably by Eckart (1960), Bretherton (1966), Hines and Reddy (1967) and Jones (1969, 1971), but rather than be drawn into debate, we shall merely repeat the conventional definitions (products of first-order terms), since these have been used in the applications (e.g., Gossard, 1962; Bretherton, 1969; Hines 1970a), and since in certain applications, they do indeed appear to be the correct forms (e.g., Jones, 1971). They are in subscript vector notation:

$$FE_i = pu_i \quad (7:6)$$

and

$$FM_{ij} = \rho_o u_i u_j, \quad (7:7)$$

where FE_i is the i^{th} component of the wave energy flux, FM_{ij} is the transport of the i^{th} component of the wave momentum in the j^{th} direction, p is the perturbation pressure, ρ_o the unperturbed density, and u_i the i^{th} component of the perturbation velocity. Most of the time, of course, we are interested in time-averages of these quantities.

Recently Hines (1970) has used (7:7) to estimate the time-averaged vertical flux of horizontal momentum associated with atmospheric gravity waves observed by Gossard et al. (1970) on August 6, 1969. In this case, he found a downward flux through the 350m level of negative momentum or an upward flux of positive momentum (subject to a possible ambiguity discussed in his paper) $\rho_o \langle u_x u_z \rangle^2 = 0.16\rho_o$ in MKS units. At the same time, the total momentum available in the bottom 350m was $\sim 10^3 \rho_o$ in MKS units, implying that this total would be removed by the wave flux at the rate given above in a time the order of 100 min, and suggesting that at least in this instance, the wave played a fundamental role in the boundary layer momentum balance.

The Doppler capability of an array of acoustic sounders (see Chapter 18) has also been used by staff of the Wave Propagation Laboratory to estimate wave-associated momentum fluxes. Wave motions are particularly amenable to such analysis since their long periods permit considerable averaging, and it appears that fluxes greater than 1 dyne/cm^2 or so should be measureable with such systems. Although the available experimental results have not been fully analyzed, the technique appears quite promising.

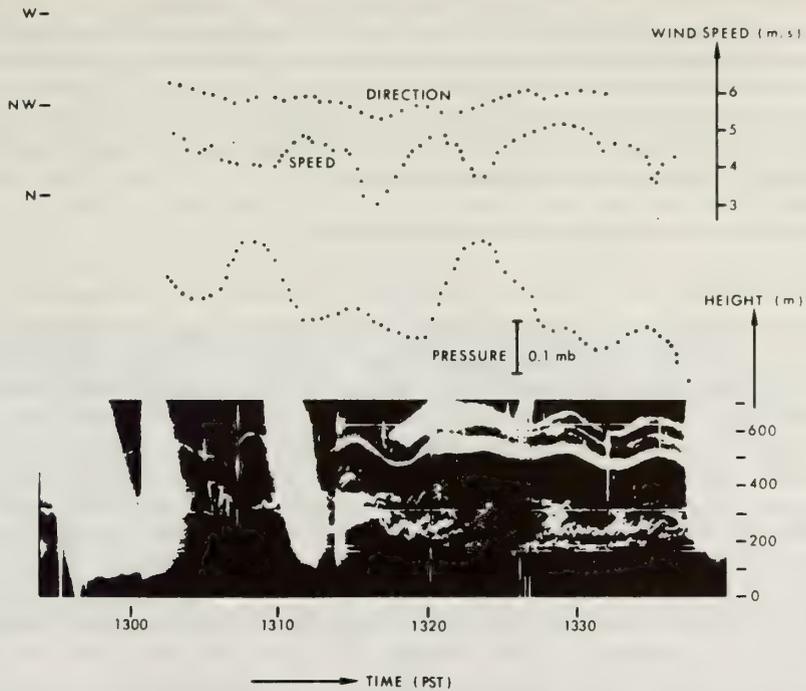


Figure 7.7 Radar sounding and surface meteorological records for July 11, 1969. Vertical light areas are rain showers (from Gossard et al., 1970).

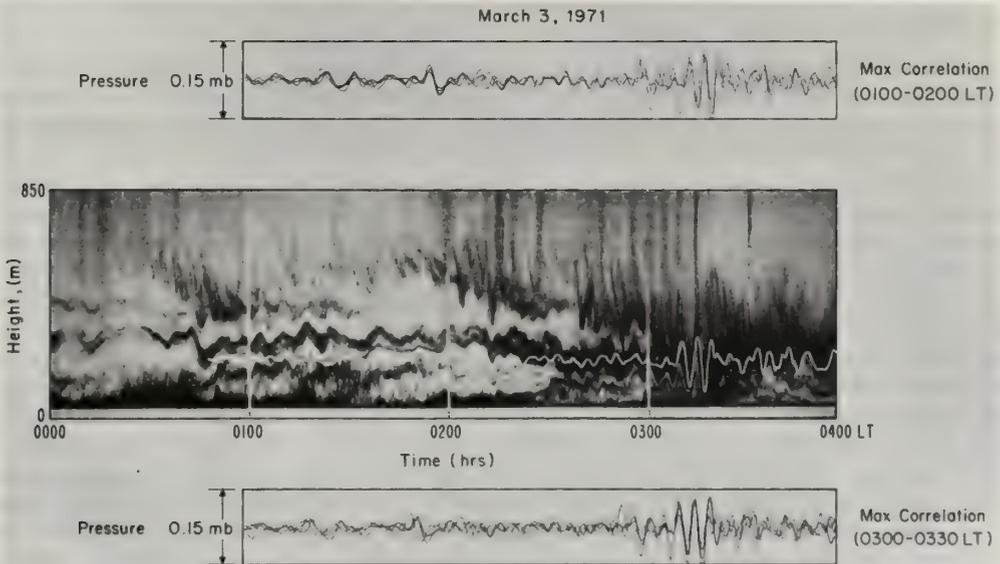


Figure 7.8 Table Mountain microbarograph and acoustic sounder records for the early morning of March 3, 1971. Dark portions of the acoustic sounder record represent regions of strong echo returns. As displayed above the sounder record, the pressure traces have been time shifted relative to one another so as to provide maximum correlation in the 0100-0200 LT time interval. As displayed below it, they have been time shifted to provide maximum correlation during the event occurring at about 0315 LT. The white line superimposed on the sounder record is the pressure trace of the M3 microbarograph (see F7.9) (from Hooke et al., 1971).

Gossard (1962) has used the approach outlined in the previous section to estimate the vertical energy flux produced by the untrapped waves, finding that in some cases, it could amount to as much as 10^3 ergs/cm² sec, or roughly 1/4 of the total rate of turbulent energy dissipation in a unit column of atmosphere (which he estimated to be $\sim 4.5 \times 10^3$ ergs/cm² sec).

Studies of this type are intriguing but hardly definitive, however (for example, Gossard ignored the effects of winds), and more work is needed to obtain representative estimates of the role played by atmospheric gravity waves in energy and momentum transport.

7.4 Wave Interactions with Other Boundary Layer Motions

7.4.1 The Mean Flow

We have discussed one aspect of wave interaction with the mean flow, namely the idea that wave momentum flux into and out of the boundary layer can serve as a major source or sink for the mean boundary-layer momentum. It should be obvious that for this momentum and energy transfer to occur, there must be generation or dissipation, rather than total reflection, of the wave in the boundary layer. The waves can be dissipated in one of several ways, either by the viscous action of pre-existent turbulent eddies, as discussed briefly in 7.2, or by a "critical level" interaction (see 5.8) with the mean flow at some level where the wave phase speed matches the background wind speed and wave amplitudes and associated wind shears increase until nonlinear and viscous effects take over. This latter subject has not been considered in specific application to the boundary layer, but a general discussion of the critical level interaction, and shortcomings in our present understanding of it, may be found in papers by Bretherton (1966, 1969a), Booker and Bretherton (1967), Hines and Reddy (1967), Jones (1967, 1968, 1971), and Jones and Houghton (1971). Generally speaking, our understanding of this interaction is better for large R_i ($> 1/4$) in the background flow than for small R_i ($< 1/4$). For small positive R_i , there is some suggestion that the wave reflection coefficient will be greater than unity, implying wave generation at the critical level (Jones, 1968; Chimonas, 1972, also refers in passing to the possible role of critical levels in boundary-layer wave generation).

7.4.2 Other Waves

Although it has been investigated only superficially, this subject is an important one. In a stably stratified fluid, the solutions to the equations of motion will be gravity waves; the actual character of the motion depends largely upon the wave amplitudes and thus the strength of their nonlinear interactions. It will range from "wavelike", when few waves are present and the interactions are weak, to "pseudo-turbulent", when many waves are present and the interactions are strong. As pointed out in Chapter 5, atmospheric gravity waves self-interact relatively weakly, but the interaction between two or more waves may be relatively strong, particularly in situations where components with wave vectors \mathbf{k}_1 and \mathbf{k}_2 form resonant triads with a component $\mathbf{k}_3 = \mathbf{k}_1 \pm \mathbf{k}_2$; the condition for resonance being $\omega_3 = \omega_1 \pm \omega_2$ satisfying the wave equation. Phillips (1966) indicates that in such an event, energy will tend to be partitioned equally among the three waves in a time the order of $(k_1 k_2 v_1 v_2)^{-1/2}$ where v_1 and v_2 represent velocity amplitudes of the two waves initially present. He further points out that even in the non-resonant case, the two interacting waves can sustain forced oscillations which in turn may interact with other forced modes or with freely propagating wave modes leading to a cascade of interactions possessing many of the characteristics of turbulence; this interaction will be weak so long as $(k_1 k_2 v_1 v_2)^{-1/2} \ll N$ (note that this is equivalent to saying that the Richardson number associated with the wave motions must be large). If we adopt values $k \sim 0.1 \text{ m}^{-1}$, $v \sim 0.2 \text{ m/s}$, however, which may be representative for wave disturbances of the very smallest scale in the lower boundary layer, it appears that this condition will be violated, and in such an event non-resonant interactions may be expected to play some role in the surface boundary layer. This subject requires further investigation. Hasselmann (1966, 1967) and Bretherton (1969b) have also considered aspects of this problem.

7.4.3 Turbulence

To my knowledge, there has been little or no published work in the subject of atmospheric gravity wave scattering and dissipation by turbulence, although these appear to be important subjects in the boundary layer context and should be pursued further. There has been, however, a significant amount of work done on the generation of atmospheric gravity waves by turbulence, most of it in application to the solar atmosphere

and in explanation of the high temperature of the solar corona. The work closely parallels pioneering work by Lighthill (1952, 1954, 1955) on the generation of acoustic waves by turbulence, in which it is assumed that the turbulence is unaffected by the waves it generates and in which it is found that the acoustic power generated is indeed weak and a rapidly increasing function with increasing Mach number characteristic of the turbulent fluctuations. The low generating efficiency and its improvement with increasing Mach number arise because the sound waves, unlike the turbulent eddies, are longitudinal, not rotational, motions and propagate much faster (with the speed of sound, C) than the eddies which produce them; hence the coupling between the two types of motion is minimal. When the calculation is repeated for the analogous case of gravity-wave generation by a turbulent flow in a stably stratified fluid, it is found that the power output is infinite (Stein, 1967). Even though the vertical energy flux is found to be finite, this is a physically unacceptable result, and it is quite obvious that the assumption that the turbulence is unaffected by the waves it generates is unreasonable in this case. At the root of the difficulty is the fact that atmospheric gravity waves, like turbulent eddies, are rotational, and like turbulent eddies, they may have arbitrarily small phase velocities; hence the coupling between the motions is quite efficient (and in some cases, the distinction between the two may be arbitrary). Although this work has been directed toward understanding of the solar atmosphere, and not the atmospheric boundary layer, and although a number of theoretical uncertainties remain, the generation of waves by turbulence, with its concomitant radiation of energy and momentum from the boundary layer, appears to be a promising area for future boundary-layer research.

Naturally occurring turbulent convection in the unstable daytime boundary layer may also generate gravity waves in the overlying inversion through the mechanism of penetrative convection (Townsend, 1966, 1968). This process has been simulated numerically (Deardorff, 1969), and studied in the laboratory (Deardorff, et al., 1969). It also appears to have been observed in nature (Gossard et al., 1971); (F7.10) shows an example.

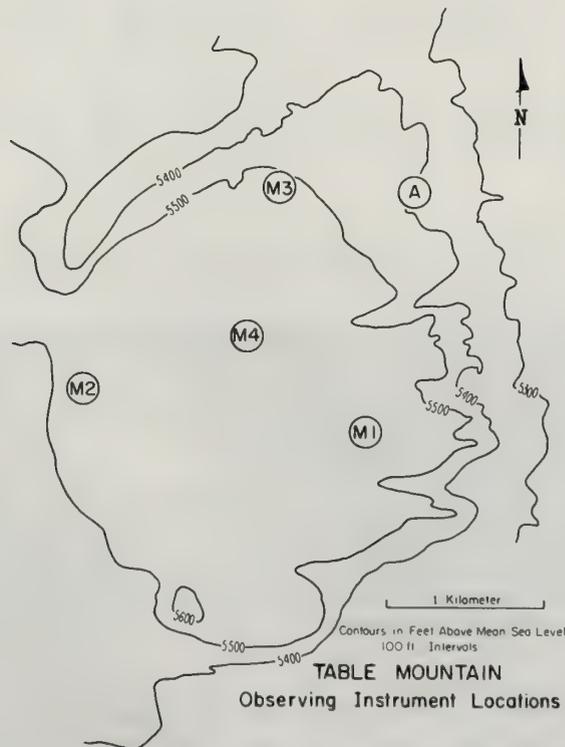


Figure 7.9 Location of the observing instruments on Table Mountain. The circles marked M1, M2, M3, and M4 indicate the positions of the microbarographs, and the circle marked A represents the position of the acoustic sounder. The entire system is located just east of the foothills of the Rocky Mountains (from Hooke et al., 1971).

Just as turbulence may generate waves, waves may produce turbulence when their amplitudes are sufficiently large as to induce convective or Kelvin-Helmholtz instability (see Chapter 3). In the case of convective instability, the wave motion may perturb the temperature profile to such an extent that the lapse rate becomes superadiabatic (Hodges, 1967; Orlanski and Bryan, 1969); in the Kelvin-Helmholtz case the wave motion may drive the Richardson number below 1/4 (Gossard et al., 1971). Obviously in an initially stably stratified background fluid, one expects the Kelvin-Helmholtz instability to set in first, but the situation is dynamic and the relative importance of the two types of instability depends critically upon the growth rates of the instabilities involved.

7.4.4 The Chimonas Instability – Wave/Turbulence/Mean Flow Interaction in the Surface Boundary Layer

Recently G. Chimonas (1972) has treated theoretically the interaction between waves, turbulence, and mean flow in the surface atmospheric boundary layer. We summarize this work briefly as follows. It is inherent in the nature of an atmospheric gravity wave that it perturbs the background wind and temperature fields of the medium in which it propagates. It will consequently perturb the gradient Richardson number of that fluid as defined by (7:1). In the statically stable surface boundary layer, the ambient or background Richardson number is close to a critical value separating dynamically stable and unstable regimes (see the references cited earlier in 7.1). In this case, a perturbing atmospheric gravity wave of sufficient amplitude will alternately drive the Richardson number above and below this critical value at different phases of the wave cycle. At one such phase in each wavelength, the wave then causes the laminar flow to break down into turbulence. The turbulence then acts on the background momentum and temperature gradients to produce enhanced momentum and energy fluxes at this wave phase. These fluxes appear as forcing terms which must be added to the equations of momentum and energy conservation describing the wave motion. Thus (5:2) and (5:3) are generalized to

$$\rho_0 \left(\frac{\partial}{\partial t} + \underline{u} \cdot \underline{\nabla} \right) \underline{u} - \rho' \underline{g} + \underline{\nabla} p = -\underline{\nabla} \cdot \underline{\mathbf{M}} \quad (7:8)$$

$$\rho_0 \left(\frac{\partial}{\partial t} + \underline{u} \cdot \underline{\nabla} \right) p = \gamma p_0 \left(\frac{\partial}{\partial t} + \underline{u} \cdot \underline{\nabla} \right) \rho - (\gamma - 1) \rho^2 \underline{\nabla} \cdot (\underline{\mathbf{H}}) \quad (7:9)$$

where $\underline{\nabla} \cdot \underline{\mathbf{M}}$, the divergence of the momentum flux tensor, and $\underline{\nabla} \cdot \underline{\mathbf{H}}$, the divergence of the heat flux, represent the forcing terms. Chimonas was able to show that because the background momentum and temperature gradients are positive in the stable boundary layer (see 7.1 and the profiles of F7.1), and because of the nature of the wave mode he considered, the fluxes occur in just the right sense to reinforce the original wave. This in turn increases the intensity and spatial extent of the turbulent regions, resulting in an initially rapid growth of wave amplitudes and turbulent intensities. Chimonas' quasilinear analysis does not take explicit account of the concomitant drain in the energy and momentum of the background flow, and because of this and other simplifications it is inadequate to determine the exact form of the growth and the mechanism by which it is ultimately quenched.

7.5 Orographic Effects

Much of the discussion of this chapter and the one preceding tacitly assumes the existence of a horizontal plane lower boundary characterized by roughness lengths the order of cm. Except over the oceans, however, the lower boundary of the earth's atmosphere is not so smooth; instead it is characterized by mesoscale and synoptic-scale undulations or irregularities which act as important obstacles to the atmospheric flow (e.g., Reiter and Rasmussen, 1967). Under appropriate conditions, these obstacles generate a type of atmospheric gravity wave, fixed with respect to the obstacle but propagating relative to the medium, known as a "lee wave". The motion is often rendered visible in the lee of mountains by clouds forming at alternate phases of the wave motion there. These waves may play a major role in vertical momentum transport in the earth's atmosphere (generating momentum fluxes ~ 5 dynes/cm²); the subject has been developed extensively in papers by Lyra (1943), Queney (1947), Scorer (1949), Eliassen and Palm (1960), Blumen (1965), Bretherton (1969a), Lilly (1971) and others.

7.6 Summary

To summarize: atmospheric gravity waves exist in the planetary boundary layer as well as in the free atmosphere above it. Their propagation is substantially modified by the lower atmospheric temperature and wind structure; in particular, it appears that the waves of largest amplitude observed near ground level have their energy trapped there by this structure, although comparison between theory and observation on this point is crude, in the sense that the complex structure of the lowest few thousand meters is typically approximated by simple two-and three-layer models.

Wave-associated momentum and energy exchange may be considerable. In specific cases, waves have been found to supply as much as 1/4 of the total energy dissipated by a unit column of atmosphere. In a several-hour interval, they may be able to supply or withdraw much of the momentum contained in the entire boundary layer. The waves interact with all other boundary-layer motions: they may enhance or diminish the mean flow, and they may energize, or draw energy from, the ambient turbulence. They may, in suitable circumstances, use the turbulence to draw energy from the mean flow and distribute it between wave and turbulence. The waves may also interact strongly with one another. Orographically generated waves may play a major role in global atmospheric momentum transport.

We remain largely ignorant of the true role of atmospheric gravity waves in boundary-layer dynamics. We do not clearly understand how the waves are generated, although it generally seems that dynamical instabilities, critical levels, turbulence, penetrative convection of convectively unstable elements and orographic effects somehow play a role. Although avenues of research for exploring the interactions of waves with other atmospheric motions have been opened, we do not yet know which interactions are important and which may be safely ignored. Although we understand that waves can generate turbulent regions and use these to feed on the energy and momentum in the background temperature and wind structure in cases where the ambient Richardson number $Ri \sim 1/4$, we do not know the extent to which this background condition is met.

Finally, our ignorance of the gravity-wave role in boundary-layer dynamics continues because we lack sufficient data. Even the best and most comprehensive experiments performed to date often provide too little information about the background flow (one radiosonde observation per experiment, for example, instead of one every few minutes or so during the experiment), or too little information about the dynamics of the atmosphere adjacent to but exterior to the boundary layer. Experimenters often complain that when they press wave theorists for just which atmospheric parameters they would like to have measured, the theorists reply, "Everything." But this is not merely a theorist cop-out. It is a simple reflection of the true state of



Figure 7.10 Convection cells in the marine layer generating gravity waves in the inversion above (from Gossard et al., 1971).

affairs; unless the theorist knows the mean background wind speed, for example, he can't even determine the most elemental quantities such as the so-called intrinsic wave frequency (defined by 5:32), which he needs in order to compare observed wave parameters with the pertinent dispersion equation.

Perhaps more importantly, we lack the kind of extensive statistical data taken at large numbers of places and under a variety of conditions that would permit us to determine whether wave phenomena in the boundary layer are merely isolated, rather special "events," or whether they are so pervasive that their momentum and energy transport should be parametrized and included in models of general atmospheric circulation, numerical weather prediction, etc. We can only hope that the newly developed remote sensing techniques described in this volume will prove capable of acquiring the continuous, extensive, and complete data coverage we now require.

7.7 References

- Blackadar, A. K., (1957) Boundary-layer wind maxima and their significance for the growth of nocturnal inversions, *Bull. Am. Meteorol. Soc.* **38**, 283-290.
- Blumen, W., (1965) Momentum Flux by mountain waves in a stratified, rotating atmosphere, *J. Atmos. Sci.* **22**, 529-534.
- Booker, J. R., and F. P. Bretherton, (1967) The critical layer for internal gravity waves in a shear flow, *J. Fluid Mech.* **27**, 513-539.
- Bretherton, F. P., (1966) The propagation of groups of internal gravity waves in shear flow, *Quart. J. Roy. Meteorol. Soc.* **92**, 466-480.
- Bretherton, F. P., (1969a) Momentum transport by gravity waves, *Quart. J. Roy. Meteorol. Soc.* **95**, 213-243.
- Bretherton, F. P., (1969b) Waves and turbulence in stably stratified fluids, *Radio Sci.* **4**, 1279-1287.
- Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, (1971) Flux-profile relationships in the atmospheric surface layer, *J. Atmos. Sci.* **28**, 181-189.
- Charnock, H., and T. H. Ellison, (1967) The boundary layer in relation to large-scale motions of the atmosphere and ocean, *Study Conf. on GARP (ICSU/IUGG; WMO and COSPAR)*.
- Chimonas, G., (1970) The extension of the Miles-Howard theorem to compressible fluids, *J. Fluid Mech.* **43**, 833-836.
- Chimonas, G., (1972) The stability of a coupled wave-turbulence system in a parallel shear flow, *Boundary Layer Meteorol.* in press, 1972.
- Clarke, R. H., (1970) Observational studies in the atmospheric boundary layer, *Quart. J. Roy. Meteorol. Soc.* **96**, 91-114.
- Deardorff, J. W. (1969) Numerical study of heat transport by internal gravity waves above a growing unstable layer, *Phys. Fluids, Supplement II, High Speed Computing in Fluid Dynamics*, 189-194.
- Deardorff, J. W., G. E. Willis, and D. K. Lilly (1969), Laboratory investigation of non-steady penetrative convection, *J. Fluid Mech.* **35**, 7-31.
- Eckart, C. H., 1960: *Hydrodynamics of Oceans and Atmospheres*, Pergamon, Oxford, New York .
- Eliassen, A., and E. Palm, (1960) On the transfer of energy in stationary mountain waves, *GeoF. Publ. Oslo* **22**, 1-23.

- Gossard, E. E., (1962) Vertical flux of energy into the lower ionosphere from internal gravity waves generated in the troposphere, *J. Geophys. Res.* 67, 745-757.
- Gossard, E. E., and W. Munk, (1954) On gravity waves in the atmosphere, *J. Meteorol.* 11, 259-269.
- Gossard, E. E., and J. H. Richter, (1970) The shape of internal waves of finite amplitude from high-resolution radar sounding of the lower atmosphere, *J. Atmos. Sci.* 27, 971-973.
- Gossard, E. E., J. H. Richter, and D. Atlas, (1970) Internal waves in the atmosphere from high resolution radar measurements, *J. Geophys. Res.* 75, 903-913.
- Gossard, E. E., D. R. Jensen, and J. H. Richter, (1971) An analytical study of tropospheric structure as seen by high-resolution radar, *J. Atmos. Sci.* 28, 794-807.
- Hasselmann, K., (1966) Feynman diagrams and interaction rules of wave-wave scattering processes, *Rev. Geophys.* 4, 1-32 (1966).
- Hasselmann, K., (1967) A criterion for nonlinear wave stability, *J. Fluid Mech.* 30, 737-739.
- Hines, C. O., (1970) Comments on paper by E. E. Gossard, J. H. Richter, and D. Atlas, "Internal waves in the atmosphere from high resolution radar measurements," *J. Geophys. Res.* 75, 5956-5959.
- Hines, C. O., (1971) Generalizations of the Richardson criterion for the onset of atmospheric turbulence, *Quart. J. Roy. Meteorol. Soc.*, 97, 429-439 (1971).
- Hines, C. O., and C. A. Reddy, (1967) On the propagation of atmospheric gravity waves through regions of wind shear, *J. Geophys. Res.* 72, 1015-1034.
- Hodges, R. R., Jr., (1967) Generation of turbulence in the upper atmosphere by internal gravity waves, *J. Geophys. Res.* 72, 3455-3458 (1967).
- Hooke, W. H., J. M. Young, and D. W. Beran, (1972) Atmospheric waves observed in the planetary boundary layer using an acoustic sounder and a microbarograph array, *Boundary Layer Meteorol.*, in press.
- Howard, L. N., (1961) Note on a paper of John Miles, *J. Fluid Mech.* 10, 509-512.
- Hunt, J. N., (1961) Interfacial waves of finite amplitude, *La Hoville Blanche* 4, 515-525.
- Jones, W. L., (1967) Propagation of internal gravity waves in fluids with shear flow and rotation, *J. Fluid Mech.* 30, 439-448.
- Jones, W. L., (1968) Reflexion and stability of waves in stably stratified fluids with shear flow: a numerical study, *J. Fluid Mech.* 34, 609-624.
- Jones, W. L., (1969) The transport of energy by internal waves, *Tell us* 21, 177-184.
- Jones, W. L., (1971) Energy-momentum tensor for linearized waves in material media, *Rev. Geophys. and Space Phys.* 9, 917-952.
- Jones, W. L., and D. D. Houghton, (1971) The coupling of momentum between internal gravity waves and mean flow: a numerical study, *J. Atmos. Sci.* 28, 604-608.
- Lighthill, M. J., (1952) On sound generated aerodynamically, *Proc. Roy. Soc. London*, A211, 564-587.

- Lighthill, M. J., (1954) On sound generated aerodynamically. II. Turbulence as a source of sound, Proc. Roy. Soc. London, 222A, 1-31.
- Lighthill, M. J., 1955: *The effect of compressibility on turbulence*, I.U.T.A.M. and I.A.U. Symposium L2, North Holland Publishing Company, Amsterdam, 121-130.
- Lilly D. D. K., (1971) Observations of mountain-induced turbulence, J. Geophys. Res. 76, 6585-6588.
- Lyra, G., (1943) Theorie der stationären Leewellenströmung in Freier Atmosphäre, Z. angew. Math. Mech. 23, 1-28.
- Miles, J. W., (1961) On the stability of heterogeneous shear flows, J. Fluid Mech. 10, 496-508.
- Monin, A. S., and S. S. Zilitinkevitch, (1967) Planetary boundary layer and large-scale atmospheric dynamics, Study Conf. on GARP (ICSU/IUGG/WMO/COSPAR).
- Oke, T. R., (1970) Turbulent transport near the ground in stable conditions, J. Applied Meteorol. 9, 778-786.
- Orlanski, I., and K. Bryan, (1969) Formation of the thermocline step structure by large amplitude internal gravity waves. J. Geophys. Res. 74, 6975-6983.
- Phillips, O. N., 1966: *The Dynamics of the Upper Ocean*, Cambridge University Press, 261 pp.
- Pitteway, M. L. V., and C. O. Hines, (1963) The viscous damping of atmospheric gravity waves, Can. J. Phys. 41, 1935-1948.
- Queney, P., (1947) Theory of perturbations in stratified currents with applications to air flow over mountain barriers, Miscell. Rept. No. 23, 31, Department of Meteorology, University of Chicago.
- Reiter, E. R., and J. L. Rasmussen, (1967) Ed, Proc. Symposium on Mountain Meteorology, 26 June 1967, Atmos. Sci. Paper, No. 122, Dept. Atmos. Sci., Colorado State University, Fort Collins, Colorado.
- Scorer, R. S., (1949) Theory of waves in the lee of mountains, Quart. J. Roy. Meteorol. Soc. 76, 41-56 (1949).
- Sheppard, P. A., (1970) The atmospheric boundary layer in relation to large-scale dynamics, "The Global Circulation of the Atmosphere," Royal Meteorological Society, London, G. A. Corby, Ed.
- Stein, R. F., (1967) Generation of acoustic and gravity waves by turbulence in an isothermal stratified atmosphere, Solar Phys. 2, 385-432.
- Townsend, A. A., (1966) Internal waves produced by a convective layer, J. Fluid Mech. 24, 307-319.
- Townsend, A. A., (1968) Excitation of internal waves in a stably-stratified atmosphere with considerable wind shear, J. Fluid Mech. 32, 145-171.
- Webb, E. K., (1970) Profile relationships: the log-linear range, and extension to strong stability, Quart. J. Roy. Meteorol. Soc. 96, 67-90.

List of Symbols

FE_i	ith component of wave energy flux	u_*	friction velocity
FM_{ij}	flux of ith component of wave momentum in the jth direction	\bar{V}	mean wind velocity directed parallel to y-axis
f	Coriolis parameter	\hat{V}	\bar{V}/u_*
g	acceleration of gravity	V_{ph}	phase velocity
H	turbulent heat flux	v_1, v_2	wave amplitudes
h_i	boundary layer depth	W	$\rho_0^{1/2} w$
$h_2 - h_1$	thickness of overlying inversion	w	wave-associated vertical velocity perturbation
k	wave vector		
k	horizontal wave number	x, y, z	a right-handed Cartesian coordinate system, with z the vertical axis (height)
k_1, k_2	wave numbers		
M	turbulent momentum flux tensor	\hat{z}	fz/u_*
N	Brunt-Väisälä Frequency	z_0	roughness height
n	vertical wave number	γ	ratio of specific heats
P	$\rho_0^{-1/2} p$	η_e	μ_e/ρ_0
p	pressure perturbation	θ	potential temperature
Ri	Richardson number		
T_*	a scaling temperature	θ_0	$\theta(z_0)$
U	$\rho_0^{1/2} u$	$\bar{\theta}$	mean potential temperature
\bar{U}	mean wind speed, directed parallel to x-axis	$\hat{\theta}$	$(\theta - \theta_0)/T_*$
\hat{U}	\bar{U}/u_*	μ	eddy viscosity coefficient
u	a vector velocity	ρ_0	background density
u	wave-associated horizontal velocity perturbation	ω	angular wave frequency
		ω_g	isothermal Brunt-Väisälä frequency

Chapter 8 SCATTERING, PROPAGATION AND REFRACTION OF ELECTROMAGNETIC AND ACOUSTIC WAVES

R. H. Ott

Institute for Telecommunication Sciences
Office of Telecommunications

In this chapter we derive the mathematical methods used in Kirchhoff-Fresnel diffraction, ray-tracing, refraction and Mie scattering by spheres, together with the range of validity of these formulas. These formulas have application in the scattering and propagation problems which arise in remote sensing of the environment.

8.0 Introduction

Remote sensing techniques rely upon information received from the environment in the form of electromagnetic or acoustic radiation. To interpret this information, we must understand the way the radiation reacts with the environment. Thus we are led to the study of electromagnetic theory, which is based upon a fundamental set of equations known as Maxwell's equations. These equations are discussed briefly in 8.1. In the derivation of many formulas used in scattering and propagation problems the introduction of vector and scalar potentials is required. The electromagnetic fields in Maxwell's equations are represented in terms of these potentials. These potentials satisfy the reduced wave equation and this chapter is basically concerned with some of the mathematical methods for solving this equation. The potentials are introduced in 8.2. Since only a very limited number of problems arising in scattering and propagation can be solved exactly, various approximate techniques must be introduced. A very useful approximate scattering formula is the Kirchhoff-Fresnel diffraction formula. This formula introduced in 8.3 is useful for studying the effects of scattering by arbitrarily shaped targets in the air (e.g. birds) or on the ground.

Ray-tracing is a useful tool for analyzing propagation when the refractive index is a function of position. In 8.4 geometrical optics and the concept of ray-tracing (the eiconal and transport equations) are introduced and the range of validity of various approximations such as WKB is discussed. Qualitative results for ray-tracing in inhomogeneous and anisotropic mediums are given.

In 8.5 tropospheric propagation is discussed together with the canonical solutions for tropospheric ducting. In 8.6 classical scattering by acoustically hard and soft spheres is considered, together with the Mie series, as well as some qualitative results for electromagnetic scattering by spheres. Mie scattering is also considered in Chapter 13.

8.1 Maxwell's Equations

The basic laws of electricity and magnetism can be summarized in differential form by four equations (MKS rationalized system of units)

$$\begin{aligned} \text{Coulomb's law:} & \quad \underline{\nabla} \cdot \underline{\underline{D}}^\dagger = \rho \\ \text{Ampère's law:} & \quad \underline{\nabla} \times \underline{\underline{H}} = \partial \underline{\underline{D}} / \partial t + \underline{\underline{J}} \\ \text{Faraday's law:} & \quad \underline{\nabla} \times \underline{\underline{E}} = - \partial \underline{\underline{B}} / \partial t \\ \text{Absence of free} & \\ \text{magnetic poles:} & \quad \underline{\nabla} \cdot \underline{\underline{B}} = 0 \end{aligned} \tag{8:1}$$

†The bar under a symbol denotes a vector quantity.

The quantities in (8:1) are

\underline{E} , the electric intensity (volts per meter)

\underline{H} , the magnetic intensity (amperes per meter)

\underline{D} , the electric flux density (coulomb's per square meter)

\underline{B} , the magnetic flux density (webers per square meter)

\underline{J} , the electric current density (amperes per square meter)

ρ , the electric charge density (coulomb's per cubic meter)

The equations which characterize the medium are

$$\underline{D} = \epsilon \underline{E} \quad (8:2)$$

$$\underline{B} = \mu \underline{H}$$

$$\underline{J} = \sigma \underline{E}$$

and in free space

$$\begin{aligned} \underline{D} &= \epsilon_0 \underline{E} \\ \underline{B} &= \mu_0 \underline{H} \end{aligned} \quad (8:3)$$

where ϵ (farads per meter) is the permittivity, μ (henrys per meter) is the permeability and σ (mhos per meter) is the conductivity. The speed of light in free space is $c = 1/\sqrt{\mu_0 \epsilon_0}$. Together with the Lorenz force equation

$$\underline{F} = \rho(\underline{E} + \underline{u} \times \underline{B}) \quad \text{in Newtons} \quad (8:4)$$

with \underline{u} (meters per second) the velocity of the charged particle, Newton's second law of motion

$$\underline{F} = m \underline{a} \quad (8:5)$$

with m (kilograms) the mass and \underline{a} (meters per sec²) the acceleration of the particle, and Maxwell's equations all macroscopic linear electromagnetic phenomena are described. These equations provide a complete description of the classical dynamics of interacting charged particles and electromagnetic fields.

The boundary conditions which show how the fields behave across boundaries where the medium changes in (8:1) are

$$\underline{e}_n \times (\underline{H}^2 - \underline{H}^1) = \underline{J}_s \quad (8:6)$$

$$\underline{e}_n \times (\underline{E}^2 - \underline{E}^1) = 0$$

$$\underline{e}_n \times (\underline{D}^2 - \underline{D}^1) = \rho_s$$

$$\underline{e}_n \times (\underline{B}^2 - \underline{B}^1) = 0$$

where \underline{e}_n is the unit normal to the surface s in (F8.1). The subscript s in (8:6) refers to currents and charges on the surface s . The equation of continuity based upon conservation of charge is

$$\underline{\nabla} \cdot \underline{J}_v + \frac{\partial \rho_v}{\partial t} = 0 \quad (8:7)$$

where \underline{J}_v is the volume current density (amperes per cubic meter) and ρ_v is the volume charge density (coulombs per cubic meter).

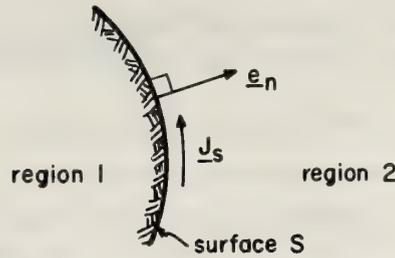


Figure 8.1 Surface currents.

The differentiation of vector quantities in (8:1) through (8:7) follows from the rules of elementary calculus and, for example, the following definitions for gradient, divergence and curl in rectangular coordinates:

$$\begin{aligned}
 \underline{\nabla} f &= \underline{e}_x \frac{\partial f}{\partial x} + \underline{e}_y \frac{\partial f}{\partial y} + \underline{e}_z \frac{\partial f}{\partial z} \\
 \underline{\nabla} \cdot \underline{F} &= \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \\
 \underline{\nabla} \times \underline{F} &= \underline{e}_x \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) + \underline{e}_y \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \\
 &\quad + \underline{e}_z \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right)
 \end{aligned}
 \tag{8:8}$$

where f is a scalar function and \underline{F} is a vector function. Throughout this chapter, unit vectors are denoted by a lower case \underline{e} .

The following vector identities will be used

$$\underline{\nabla} \cdot (\underline{\nabla} \times \underline{F}) = 0$$

$$\underline{\nabla} \cdot \underline{\nabla} f = \underline{\nabla}^2 f \tag{8:9}$$

$$\underline{\nabla} \times \underline{\nabla} f = 0$$

$$\underline{\nabla} \cdot (f \underline{F}) = \underline{F} \cdot (\underline{\nabla} f) + f(\underline{\nabla} \cdot \underline{F})$$

and are easily verified in rectangular coordinates using (8:8).

8.2 Vector and Scalar Potentials

Maxwell's equations consist of a set of coupled first-order partial differential equations relating the various components of electric and magnetic fields. They can be solved as they stand in only very simple situations. It is therefore convenient to introduce potentials, obtaining a smaller number of second-order equations. From Maxwell's equations

$$\underline{\nabla} \cdot \underline{B} = 0$$

and from (8:9) \underline{B} can be represented locally in the form

$$\underline{B} = \underline{\nabla} \times \underline{A} \tag{8:10}$$

where \underline{A} is the vector potential. It follows from (8:9) that \underline{A} may be altered by adding any vector function of the form $\underline{\nabla}g$, where g is a suitable scalar. Putting this into Faraday's law in (8:1) we have

$$\underline{\nabla} \times \left(\underline{E} + \frac{\partial \underline{A}}{\partial t} \right) = 0 \quad (8:11)$$

Again, from (8:9), a scalar function ρ exists such that, locally at least,

$$\underline{E} + \frac{\partial \underline{A}}{\partial t} = -\underline{\nabla}\rho \quad (8:12)$$

where the negative sign is chosen only to improve the final appearance of our equations. We have from (8:10) and (8:12)

$$\begin{aligned} \underline{E} &= -\underline{\nabla}\rho - \frac{\partial \underline{A}}{\partial t} \\ \underline{B} &= \underline{\nabla} \times \underline{A} \end{aligned} \quad (8:13)$$

After a considerable number of algebraic maneuvers using (8:9) it can be shown (e.g., Jones, 1964) in source-free regions where $\underline{J} = 0$ and in "slowly varying" mediums where $(\underline{\nabla}\epsilon)/\epsilon \cong 0$ and $(\underline{\nabla}\mu)/\mu \cong 0$ that \underline{E} and \underline{H} satisfy the Helmholtz wave equation.

$$\underline{\nabla}^2 \underline{F} + \omega^2 \underline{\mu}(\underline{r}) \epsilon(\underline{r}) \underline{F} = 0 \quad (8:14)$$

where \underline{F} represents \underline{E} or \underline{H} . We have assumed harmonic time dependence $e^{i\omega t}$ for \underline{E} and \underline{H} so that for example $\partial/\partial t$ is replaced with $i\omega$. This chapter is basically concerned with some of the mathematical methods for solving (8:14).

8.3 Free-Space Green's Function and the Fresnel-Kirchhoff Diffraction Formula

We stated in the last section that the components ψ of \underline{E} and \underline{H} satisfy the scalar wave equation often referred to as the reduced wave equation

$$\underline{\nabla}^2 \psi + k^2 \psi = -f(\underline{R}) \quad (8:15)$$

where $k = \omega/c$ and ω is the angular frequency. We have introduced a term on the right hand side of (8:15) to represent the source distribution. The function $f(\underline{R})$ could for example represent the currents in a transmitting antenna.

One method for obtaining a solution of the wave equation is based upon the solution of (8:15) when the right-hand-side is a point source. Such a solution is known as a Green's function. The argument of the Green's function depends upon two distinct points: the point (x,y,z) at which ψ is to be evaluated and the variable point (x',y',z') associated with the region occupied by the source distribution $f(\underline{R})$. The two points may be specified by the two position vectors \underline{R} and \underline{R}' , respectively; furthermore

$$\underline{r} = \underline{R} - \underline{R}' \quad (8:16)$$

is the vector between the two points and in the case of rectangular coordinates

$$r = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2} \quad (8:17)$$

Provided we know the solution of (8:15) for a point source at \underline{R} , the solution for our given arbitrary distribution $f(\underline{R})$ follows at once by superposition

$$\psi(\underline{R}) = \int_V G(\underline{R}, \underline{R}') f(\underline{R}') dv' \quad (8:18)$$

in which $G(\underline{R}, \underline{R}')$ is a Green's function and the integration is carried out over the volume occupied by f . To find G , ψ in (8:18) is substituted into (8:15). Then

$$\int_V [\nabla^2 G(\underline{R}, \underline{R}') + k^2 G(\underline{R}, \underline{R}')] f(\underline{R}') dv' = -f(\underline{R}) \quad (8:19)$$

with the term in the brackets having the properties of an impulse function (Schwartz, 1950) so it is convenient to introduce the volume impulse function $\delta(\underline{R}-\underline{R}')$ at this point. The impulse function may be defined by

$$\int_V f(\underline{R}') \delta(\underline{R}-\underline{R}') dv' = f(\underline{R}) \quad (8:20)$$

From the definition of the impulse function and (8:19) it follows that our Green's function satisfies

$$\nabla^2 G(\underline{r}) + k^2 G(\underline{r}) = -\delta(\underline{r}) \quad (8:21)$$

Although (8:21) is also an inhomogeneous, partial differential equation, it has the advantage over (8:15) that the source is concentrated at \underline{R}' . The function $G(\underline{r})$ in (8:21) therefore depends only upon r and not θ or φ because of the spherical symmetry about the point \underline{R}' . The Laplacian operator ∇^2 in spherical coordinates independent of θ and φ is

$$\nabla^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) \quad (8:22)$$

and (8:21) becomes

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial G}{\partial r} \right) + k^2 G = -\delta(\underline{r}) \quad (8:23)$$

Solutions of (8:23) are

$$G = \frac{e^{\pm ikr}}{r} \quad (8:24)$$

as the reader may easily verify by substitution into (8:23). Since we have adopted $e^{i\omega t}$ time convention, the Green's function that represents an outward going wave is

$$G(\underline{r}) = \frac{e^{-ikr}}{r} \quad (8:25)$$

An example of the use of the free-space Green's function occurs in the derivation of the often used Kirchhoff-Fresnel diffraction formula. We start with Green's second identity

$$\int_V (\psi \nabla^2 \omega - \omega \nabla^2 \psi) dv = \int_S (\psi \nabla \omega - \omega \nabla \psi) \cdot \underline{e}_n da' \quad (8:26)$$

where S is a closed surface bounding the volume V and \underline{e}_n denotes a unit normal to S pointing outward to S (see F8.2). This identity is certainly valid when ψ and ω and their first- and second-order partial derivatives are continuous within and on S .

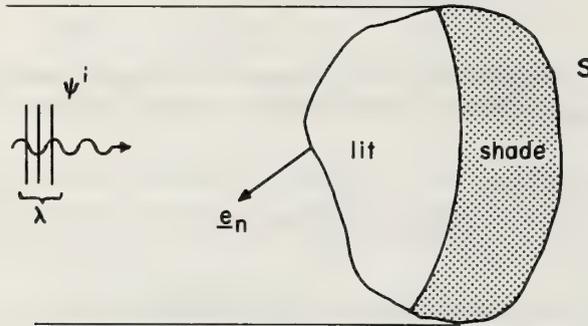


Figure 8.2 Scatterer immersed in an incident plane wave showing illuminated and shaded portion of S .
Multiplying (8:21) by ψ and (8:15) by G and subtracting the resulting two equations gives

$$\psi \nabla^2 G - G \nabla^2 \psi = -\psi \delta(r) + G(r) f(\underline{R}) . \quad (8:27)$$

Substituting (8:27) into (8:26) and performing the volume integration over the singularity at $r = 0$ yields

$$\begin{aligned} \psi(\underline{R}) &= \int_V G(\underline{R}, \underline{R}') f(\underline{R}') dv' \\ &+ \frac{1}{4\pi} \int_S \left\{ \frac{e^{-ikr}}{r} \frac{\partial \psi}{\partial n} - \psi \frac{\partial}{\partial n} \left(\frac{e^{-ikr}}{r} \right) \right\} da'. \end{aligned} \quad (8:28)$$

We see from (8:28) that knowledge of the field and its normal derivative on the surface S are required to compute the field at a point \underline{R} in space. In the case of acoustic problems, the velocity potential ψ may satisfy a sound-hard boundary condition

$$\frac{\partial \psi}{\partial n} = 0, \quad \psi \text{ on } S \quad (8:29)$$

The sound-hard boundary condition corresponds in the electromagnetic scattering case to ψ representing a component of the magnetic field \underline{H} tangential to the surface of the scatterer S . For the case of a sound-soft surface, the potential ψ satisfies the boundary condition

$$\psi = 0, \quad \psi \text{ on } S \quad (8:30)$$

The sound-soft boundary condition corresponds in the electromagnetic scattering case to ψ representing a component of the electric field \underline{E} tangential to the surface S .

A much used approximate solution to the integral equation in (8:28) makes use of the Kirchhoff approximation (sometimes equivalent to the physical optics approximation). The Kirchhoff approximation states

$$\begin{aligned} \psi &= \begin{cases} \psi^i, & \text{illuminated portion of } S \\ 0, & \text{dark portion of } S \end{cases} \\ \frac{\partial \psi}{\partial n} &= \begin{cases} \frac{\partial \psi^i}{\partial n}, & \text{illuminated portion of } S \\ 0, & \text{dark portion of } S \end{cases} \end{aligned} \quad (8:31)$$

where ψ^i is the incident field in F.8.2 and is given by

$$\psi^i = \frac{1}{4\pi} \int_V \frac{e^{-ik|\underline{R}-\underline{R}'|}}{|\underline{R}-\underline{R}'|} f(\underline{R}') dv' \quad (8:32)$$

Using (8:31) the scattered field in (8:28) becomes

$$\psi^s(\underline{R}) = \frac{1}{4\pi} \int_S \left\{ \frac{e^{-ikr}}{r} \frac{\partial \psi^i}{\partial n} - \psi^i \frac{\partial}{\partial n} \left(\frac{e^{-ikr}}{r} \right) \right\} da' \quad (8:33)$$

and the total field is

$$\psi = \psi^i + \psi^s \quad (8:34)$$

Equation (8:33) is sometimes referred to as the Kirchhoff-Fresnel diffraction formula. It is also closely related to the Born approximation. Two criticisms of Kirchhoff's theory are (a) it does not account for a transition region or penumbra between the light and dark portions of S , and (b) the boundary values of ψ and $\partial \psi / \partial n$ as given in (8:31) are incompatible. In spite of these difficulties, the Kirchhoff theory does give results which agree very well with experiment. For additional discussion, the reader is referred to Kottler (1923) and Stratton (1941).

The Kirchhoff-Fresnel diffraction formula in (8:33) is useful in calculating the field scattered by arbitrarily shaped objects at high frequencies. That is, the smallest dimension a of the scatterer must satisfy $ka > 5$.

8.4 Geometrical Optics and Refraction

When the wavelength is so small that significant changes occur over distances large compared with it, we can assume that in local regions the field behaves as if it were in a homogeneous medium. Locally the wavefronts behave as plane waves. This is referred to as the "geometrical optics" approximation. In this approximation attention is concentrated on rays or bundles of rays. The "ray method" has application in the remote sensing field in lidar and laser propagation provided we have simple refraction effects; i.e., the bending of rays near the earth's surface would not be properly treated using a "simple" ray tracing technique because of diffraction effects. In general, "ray tracing" gives an indication of where rays are going but not the amplitude or total phase of the ray at a general point in space (Jones, 1964).

Since the potential ψ in the reduced wave equation may represent the velocity potential in acoustic problems as well as one of the components of the electric field \underline{E} or magnetic field \underline{H} in an electromagnetic problem, the mathematical methods in this section apply to the problems in chapter 21 (for example wind driven acoustic waves).

Consider the components φ of \underline{E} and \underline{H} satisfying the reduced wave equation

$$\nabla^2 \varphi + k_0^2 n^2(\underline{r}) \varphi = 0 \quad (8:35)$$

where $k_0 = \omega/c$ and $n(\underline{r})$ is the index of refraction of the medium. Since the wavelength λ is assumed small or the wavenumber ($k_0 = 2\pi/\lambda$) large we consider solutions of (8:35) of the form

$$\varphi(\underline{r}) \cong e^{-ik_0 \psi(\underline{r})} \sum_{m=0}^{\infty} \frac{\varphi_m(\underline{r})}{(-ik_0)^m}, \quad k_0 \rightarrow \infty \quad (8:36)$$

The proposed solution in (8:36) represents an asymptotic series (Copson, 1965). The first term is called the dominant term and we frequently will write

$$\varphi(\underline{r}) \sim e^{-ik_0 \psi(\underline{r})} \varphi_0(\underline{r})$$

meaning that

$$\varphi(\underline{r}) e^{ik_0 \psi(\underline{r})}$$

tends to $\varphi_0(\underline{r})$ as $k_0 \rightarrow \infty$. Substituting (8:36) into (8:35) and performing all the required differentiations (Lewis et. al., 1967) gives

$$\begin{aligned} e^{-ik_0 \psi(\underline{r})} \{ k_0^2 [n^2 - (\nabla \psi)^2] \sum_{m=0}^{\infty} \frac{\phi_m}{(-ik_0)^m} - ik_0 \nabla^2 \psi \sum_{m=0}^{\infty} \frac{\phi_m}{(-ik_0)^m} \\ - 2ik_0 \underline{\nabla} \psi \cdot \sum_{m=0}^{\infty} \frac{\underline{\nabla} \phi_m}{(-ik_0)^m} + \sum_{m=0}^{\infty} \frac{\nabla^2 \phi_m}{(-ik_0)^m} \} = 0. \end{aligned} \quad (8:37)$$

Equating the coefficients of like powers of k_0 to zero gives

$$k_0^2 : (\nabla \psi)^2 = n^2(\underline{r}), \quad (\text{the eikonal† equation}) \quad (8:38)$$

$$k_0 : (\nabla^2 \psi) \varphi_0(\underline{r}) + 2 \underline{\nabla} \psi \cdot \underline{\nabla} \varphi_0(\underline{r}) = 0 \quad (\text{the transport equation}) \quad (8:39)$$

$$k_0^0 : (\nabla^2 \psi) \varphi_1 + 2 \underline{\nabla} \psi \cdot \underline{\nabla} \varphi_1 + \nabla^2 \varphi_0 = 0 \quad (8:40)$$

$$k_0^{-m+1} : (\nabla^2 \psi) \varphi_m(\underline{r}) + 2 \underline{\nabla} \psi \cdot \underline{\nabla} \varphi_m(\underline{r}) + \nabla^2 \varphi_{m-1}(\underline{r}) = 0, \quad m=2, 3, \dots \quad (8:41)$$

and we have a recursive system.

†The term eikonal comes from the Greek "image"

The solution of (8:38) gives the ray paths; i.e., the ray paths are perpendicular to surfaces of constant phase, $\psi(\mathbf{r}) = \text{constant}$, and may be found once $\psi(\mathbf{r})$ is known.

The first term in (8:36) is the "geometrical optics approximation". The next term is $\varphi_1/(-ik_0)$ and if $\varphi_1/k_0 \ll \varphi_0$ we may use the first term alone.

The relationship of the geometrical optics approximation to the WKB method may be illustrated by considering a one-dimensional case:

$$\frac{d^2\varphi}{dx^2} + k_0^2 n^2(x)\varphi = 0 \tag{8:42}$$

The solution of (8:42) with $n(x) = \text{constant}$ is of the form

$$\varphi = A e^{\pm ik_0 n x} \tag{8:43}$$

This suggests we remove the fast variation in φ with the substitution

$$\varphi = f(x) e^{-ik_0 g(x)} \tag{8:44}$$

Substituting (8:44) into (8:42) gives

$$\begin{aligned} [f''(x) - 2ik_0 f'(x)g'(x) - ik_0 g''f] e^{-ik_0 g(x)} \\ + k_0^2 [n^2(x) - (g')^2] f(x) e^{-ik_0 g(x)} = 0 \end{aligned} \tag{8:45}$$

If the fast variation of φ with x is contained in the term $e^{-ik_0 g(x)}$, then

$$f''(x) \cong 0 \tag{8:46}$$

and (8:44) is a solution to (8:42) if

$$g'(x) = n(x), \text{ or } g = \int n(x) dx \tag{8:47}$$

and

$$\frac{f'(x)}{f(x)} = -\frac{1}{2} \frac{n'(x)}{n(x)} \tag{8:48}$$

or

$$f(x) = \text{const } [n(x)]^{-1/2} \tag{8:49}$$

The general solution to (8:42) is

$$\varphi(x) \cong \frac{1}{\sqrt{n(x)}} \{ c_+ \exp[ik_0 \int n(x) dx] + c_- \exp[-ik_0 \int n(x) dx] \} \tag{8:50}$$

The condition of validity (that f'' be "small") is

$$\frac{if''}{k_0} + (2f'n + n'f) = 0 \tag{8:51}$$

The criterion in (8:51) is equivalent to $\varphi_1/k_0 \ll \varphi_0$ in the geometrical optics approximation. The geometrical optics approximation also requires that $\varphi_0(\underline{r})$ be bounded. Because of (8:51) the WKB method is most useful in cases where the media is smoothly varying while geometrical optics is more generally applicable in cases where n changes discontinuously.

The solution of the eikonal equation will define the ray trajectory. From (8:38)

$$\underline{\nabla}\psi = n \underline{e}_s, \quad (8:52)$$

with \underline{e}_s , a unit vector, defined as

$$\underline{e}_s = \frac{\underline{\nabla}\psi}{|\underline{\nabla}\psi|} \quad (8:53)$$

The relation

$$\psi(\underline{r}) = \text{const.} \quad (8:54)$$

defines an equi-phase surface. The vector $\underline{\nabla}\psi$ is normal to the wave fronts and thus points in the direction of wave propagation. Thus \underline{e}_s is a unit vector along the direction of propagation of the local geometrical optics field. From (F8.3)

$$\underline{e}_s = \frac{d\underline{R}}{ds} = \text{unit tangent vector to ray path at } \underline{R} \quad (8:55)$$

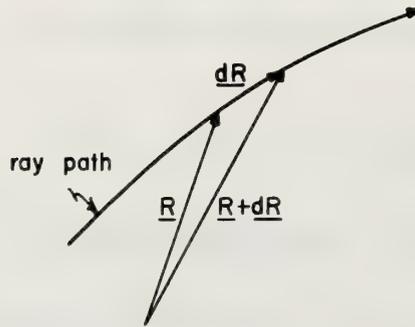


Figure 8.3 Geometry for derivation of a unit tangent vector to ray path.

Then (8:52) becomes

$$\underline{\nabla}\psi = n \frac{d\underline{R}}{ds} \quad (8:56)$$

Differentiating both sides of (8:56) with respect to s ,

$$\frac{d}{ds} \underline{\nabla}\psi = \frac{d}{ds} \left(n \frac{d\underline{R}}{ds} \right) \quad (8:57)$$

or

$$\underline{\nabla} \frac{d\psi}{ds} = \frac{d}{ds} \left(n \frac{d\underline{R}}{ds} \right) \quad (8:58)$$

Now

$$\frac{d\psi}{ds} = \underline{e}_s \cdot \nabla\psi = n \quad , \quad (8:59)$$

and (8:58) becomes

$$\nabla n = \frac{d}{ds} \left(n \frac{d\mathbf{R}}{ds} \right) \quad , \quad (8:60)$$

Multiplying both sides of (8:60) by n and defining

$$d\tau = \frac{ds}{n} \quad (8:61)$$

the ray equation in parametric form becomes

$$\frac{d^2\mathbf{R}}{d\tau^2} = \nabla \left(\frac{n^2}{2} \right) \quad (\text{Ray Eq.}) \quad (8:62)$$

As an example of the use of (8:62), consider a homogeneous medium with $n = \text{constant}$. Then

$$\frac{d^2\mathbf{R}}{d\tau^2} = 0$$

and the solution is

$$\mathbf{R} = \underline{a}_0\tau + \underline{a} \quad ,$$

which is a straight line. The ray paths in a homogeneous medium are straight lines.

At an abrupt change in refractive index we can no longer use the ray equation to determine the refractive index since $\nabla\psi$ is discontinuous. However, we may avoid this difficulty by the following approach. We have, taking the curl of the eikonal equation

$$\nabla \times \nabla\psi = \nabla \times (n \underline{e}_s) = 0 \quad (8:63)$$

Applying Stokes' Theorem to the pillbox in (F8.4) and using

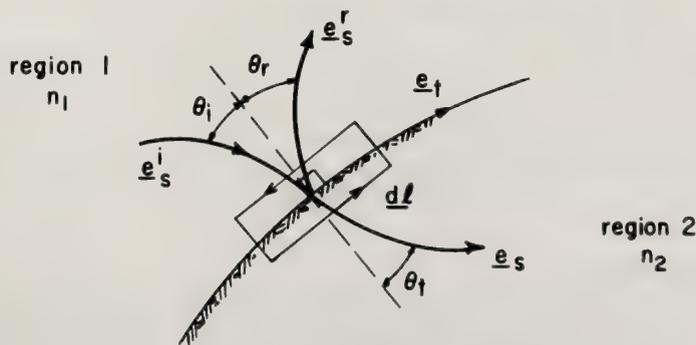


Figure 8.4 An abrupt change in index of refraction.

$$\int_C n \underline{e}_s \cdot d\underline{l} = 0 \quad (8:64)$$

we find

$$n_1(\underline{e}_s \cdot \underline{e}_t) = n_2(\underline{e}_s \cdot \underline{e}_t) \quad (8:65)$$

which becomes for the incident and reflected rays

$$\theta_i = \theta_r \quad (8:66)$$

and for the incident and refracted rays

$$n_1 \sin \theta_i = n_2 \sin \theta_t \quad (\text{Snell's law}) \quad (8:67)$$

which is the basic relation for ray tracing in inhomogeneous media.

Consider next the solution of the transport equation. Multiply (8:39) by φ_0^* , where the star denotes the complex conjugate, to obtain

$$\varphi_0^* \varphi_0 (\nabla^2 \psi) + 2(\nabla \psi \cdot \nabla \varphi_0) \varphi_0^* = 0 \quad (8:68)$$

where ψ is a real function for a lossless medium. Consider the complex conjugate of the transport equation in (8:39)

$$\varphi_0 \varphi_0^* (\nabla^2 \psi) + 2(\nabla \psi \cdot \nabla \varphi_0^*) \varphi_0 = 0 \quad (8:69)$$

and adding (8:68) to (8:69) gives

$$|\varphi_0|^2 \nabla^2 \psi + \nabla \psi \cdot \nabla |\varphi_0|^2 = 0 \quad (8:70)$$

But

$$\nabla \cdot [|\varphi_0|^2 \nabla \psi] = \nabla |\varphi_0|^2 \cdot \nabla \psi + |\varphi_0|^2 \nabla^2 \psi, \quad (8:71)$$

and so

$$\nabla \cdot [|\varphi_0|^2 \nabla \psi] = 0 \quad (8:72)$$

Defining the divergenceless vector

$$\underline{S} = |\varphi_0|^2 \nabla \psi = |\varphi_0|^2 n \underline{e}_s \quad (8:73)$$

and applying Gauss' theorem to the ray tube in(F8.5)gives

$$S_1 dA_1 - S_2 dA_2 = 0$$

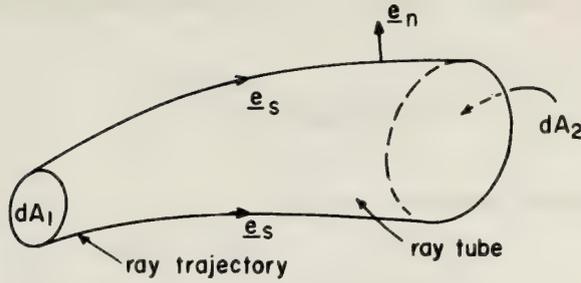


Figure 8.5 Tube of rays used in derivation of (8.76).

or

$$S_2 = S_1 \frac{dA_1}{dA_2} \quad (8:74)$$

with

$$|\underline{S}| = |\varphi_0|^2 n \quad (8:75)$$

From (8:75) and (8:74)

$$|\varphi_0^2| = |\varphi_0^1| \left[\frac{n_1 dA_1}{n_2 dA_2} \right]^{-1/2} \quad (8:76)$$

which gives the magnitude of the field at point 2, φ_0^2 , in terms of the field at point 1, φ_0^1 . The factor under the radical gives the convergence or divergence of the bundle of rays as one traverses the ray path. At a caustic dA_2/dA_1 goes to zero corresponding to a point where the rays cross and the geometrical optics approximation breaks down. To solve the problem at caustics the field is expressed in terms of Airy functions (Lewis, et al., 1967). At edges and shadow boundaries, the geometrical optics field also breaks down and we can use Keller's Geometrical Theory of Diffraction (Keller and Levy, 1959).

In summary, we have shown how to solve the wave equation (8:35) when the index of refraction $n(\underline{r})$ is a slowly varying function of position. The WKB method was compared with the asymptotic expansion solution. The zero order solution of the latter predicts the ray direction, the first order solution gives the ray amplitude provided focusing does not occur. A further result of this section is Snell's law in (8:67) which is the fundamental equation for finding out where rays go in a lossless medium having discontinuities in the refractive index. At a point in the medium where total reflection occurs, the angle of incidence equal to the angle of reflection gives the ray direction.

8.5 Tropospheric Propagation

In this section we investigate the effects of the refractivity structure $n(\underline{r})$ on the field strength. In general the refractivity structure has extremely serious effects at frequencies above 300 MHz (UHF and SHF). In general the field strength is a function of frequency, the geometry of the problem, (i.e., location of transmitter, receiver and various scatterers) together with the refractivity $n(\underline{r})$. In remote sensing applications, we are interested in the inverse problem; i.e., how does the refractivity, $n(\underline{r})$, depend upon field strength? For example, when there is an inversion layer in the atmosphere, there is in general an increase in observed field strength between two antenna sites. This occurrence of an inversion layer may correspond to a trapping of pollutants in the air resulting in smog. The increase in signal strength in this example could be used as a smog alert.

In order to study refractivity effects in the troposphere we again need to study the reduced wave equation

$$\nabla^2 \varphi + k_0^2 n^2(\mathbf{r}) \varphi = 0 \quad (8:77)$$

A one-dimensional form of the reduced wave equation

$$\frac{d^2 \varphi}{dz^2} + k_0^2 n^2(z) \varphi = 0 \quad (8:78)$$

gives the WKB solution (8:50). We can classify the behavior of solutions to (8:78) as follows: a) $n^2(z)$ is real (no absorption of rays) but takes on positive and negative values. A zero in $n^2(z)$ then corresponds to a point where total reflection occurs, b) $n^2(z)$ is real and positive. In this case there will be no downward reflection of the rays, c) $n^2(z)$ complex corresponding to a bending and absorption of the rays. The WKB approximation also applies to problems where partial reflections occur in a continuous layer (Bremmer, 1958).

The classical method for accounting for refractivity effects is to assume an effective earth's radius (Schelleng, et al., 1933). By assuming an effective earth radius the waves travel in straight lines over an earth with a radius of 4/3 the actual earth radius. The ray trajectories in a linear atmosphere where

$$n^2(z) = n_0^2 (1 + az) \quad (8:79)$$

are easily derived using the ray equation (8:62) derived in 8.4 earlier. In (8:79) n_0 is the refractive index of the atmosphere at the surface of the earth. The constant a is a measure of the rate of change of the dielectric constant and is on the order of 10^{-7} m^{-1} . From (8:62) we have

$$\frac{d^2 x}{d\tau^2} = 0 \quad (8:80)$$

$$\frac{d^2 z}{d\tau^2} = \frac{an_0^2}{2}$$

with solutions

$$\begin{aligned} x(\tau) &= a_0 + a_1 \tau \\ z(\tau) &= a_2 + a_3 \tau + \frac{an_0^2}{4} \tau^2 \end{aligned} \quad (8:81)$$

The constants in (8:81) are easily evaluated assuming the transmitter is on the earth so that when $\tau = 0$, x and z are both zero. Then

$$\begin{aligned} x(\tau) &= (\cos \alpha) \tau \\ z(\tau) &= (\sin \alpha) \tau + \frac{an_0^2}{4} \tau^2 \end{aligned} \quad (8:82)$$

where α is the elevation angle measured from the horizontal.

When $n(z)$ has a minimum as shown in (F8.6) a duct is formed. Such ducts can arise in the troposphere from temperature inversions. A complete and comprehensive description of the problem of tropospheric ducting is given in Chapter 12 of Wait (1962).

In summary, the WKB solution and geometrical optics are the most frequently used mathematical tools for investigating refractivity effects in the troposphere.

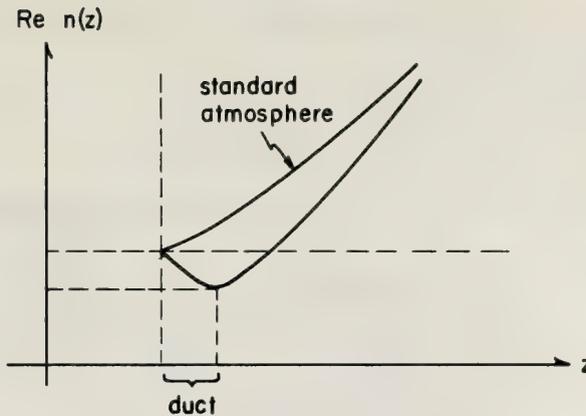


Figure 8.6 Values of $n(z)$ for a standard atmosphere and when a duct is formed.

8.6 Scattering by Spheres

The scattering of electromagnetic and acoustic waves by a sphere has been one of the most widely studied problems in diffraction theory. Logan (1965) traces the early studies of the scattering of plane waves by a sphere. In a paper published in 1908, G. Mie obtained a rigorous solution for the diffraction of a plane-monochromatic wave by a homogeneous sphere of any diameter and composition. Since then because of the numerous papers published on the subject it is impossible to give a complete bibliography in this chapter. Two papers giving a rather large amount of numerical data are: Bistatic scattering by lossy dielectric spheres with surface impedance boundary conditions (Garbacz, 1964) and Electromagnetic scattering by radially inhomogeneous spheres (Wait, 1963).

In this section we show a derivation for the field scattered by an acoustically soft sphere. This serves as an example of the classical separation of variables technique for solving the reduced wave equation (8:15) and comparison of this mathematical method with previous ones gives us a feeling for the various methods. The scattering by a sphere applies to aerosols and particulate matter in the atmosphere. This subject is also treated in Chapters 10 and 13.

8.6.1 An Acoustically Soft Sphere

The boundary condition satisfied by the potential is

$$U = 0 \text{ at } r = a, \text{ (Dirichlet boundary condition)}$$

where we are using a spherical coordinate system (r, θ, ϕ) where origin is at the center of the sphere. Let the incident wave be a plane wave traveling in the positive direction as shown in (F8.7)

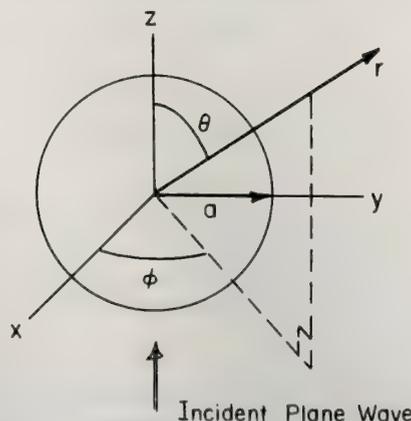


Figure 8.7 Scattering of a plane wave by a sphere.

$$U^i = e^{-ikz} = e^{-ikr \cos \theta} \quad (8:83)$$

which can be written alternatively as (Harrington, 1961)

$$U^i = \sum_{n=0}^{\infty} (n + \frac{1}{2}) e^{in\pi/2} j_n(kr) P_n(\cos \theta) \quad , \quad (8:84)$$

where j_n is a spherical Bessel function of order n , and P_n is a Legendre polynomial of order n ,

$$j_n(x) = \sqrt{\frac{2\pi}{x}} J_{n+\frac{1}{2}}(x) \quad .$$

The scattered field is also of the form

$$U^s = \sum_{n=0}^{\infty} a_n h_n^2(kr) P_n(\cos \theta) \quad , \quad (8:85)$$

where $h_n^2(kr)$ is a spherical Hankel function of the second kind representing an outgoing disturbance from the coordinate origin. The total field from (8:83) and (8:85) is

$$U = \sum_{n=0}^{\infty} \{(n + \frac{1}{2}) e^{in\pi/2} j_n(kr) + a_n h_n^2(kr)\} P_n(\cos \theta) \quad (8:86)$$

Application of the boundary condition yields

$$a_n = -(n + \frac{1}{2}) e^{in\pi/2} \frac{j_n(ka)}{h_n^2(ka)} \quad (8:87)$$

so that (8:86) and (8:87) represent a solution to the problem with no restrictions! For a general treatment of the electromagnetic scattering by spheres together with the Mie coefficients corresponding to (8:87) the reader is referred to Chapter 10, equation (10:89).

When the sphere becomes electrically large; i.e., $ka \geq 5$ where a is the radius of the sphere, we can use the Kirchhoff-Fresnel diffraction formula to compute the scattered field. From (8:28) for an acoustically soft sphere

$$U^s = \frac{1}{4\pi} \int_A \frac{e^{-ikR}}{R} \frac{\partial U^i}{\partial n} da' \quad (8:88)$$

where

$$\frac{\partial U^i}{\partial n} = \underline{e}_n \cdot \nabla U^i = \underline{e}_r \cdot \nabla U^i \quad (8:89)$$

where $\underline{e}_n = \underline{e}_r$ is the outward normal to the surface of the sphere. From (8:83) we find

$$\underline{e}_r \cdot \nabla U^i = -ik \cos \theta e^{-ikr \cos \theta} \quad (8:90)$$

where the geometry of the problem is shown in (F8.8). We have selected the stationary phase point (point where the angle of incidence equals the angle of reflection) so that the z' -axis is along the outward normal and the x' -axis is in the plane of incidence. Substituting (8:90) into (8:88) and taking the slowly

varying factors outside the integrand, we find

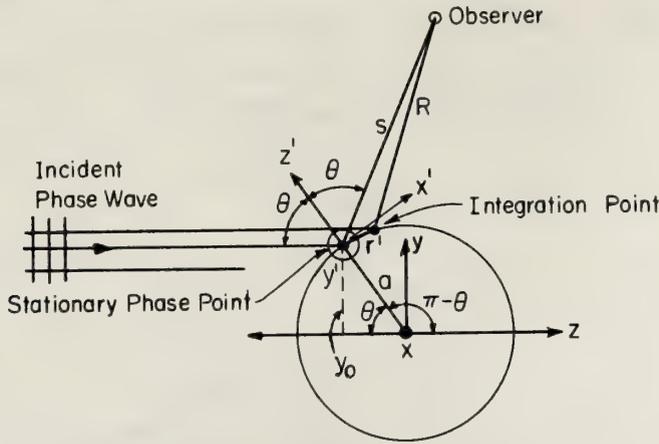


Figure 8.8 Geometry for high frequency scattering by an acoustically soft sphere.

$$U^s = \frac{-ik \cos \theta}{4\pi s} \int_{\text{lit region}} e^{-ik(R+z)} da' \quad (8:91)$$

The equation of the surface near the stationary phase point is

$$z' \cong \frac{1}{2} [z_{xx}(x')^2 + 2z_{xy}x'y' + z_{yy}(y')^2] \quad (8:92)$$

where $z_{xx} = \partial^2 z / \partial x^2$, etc., and where we have assumed that variations in the phase, when the integration point is near the stationary phase point, are described by quadratic terms alone (Jones, 1964). This will be adequate provided the observation point does not lie near a caustic (e.g., the shadow boundary) where cubic terms are required in (8:92). The coordinates of the incident plane wave in the $x'y'z'$ system are

$$\begin{aligned} z + (a^2 - y_0^2)^{1/2} &= x' \sin \theta - z' \cos \theta \quad , \\ y - y_0 &= x' \cos \theta + z' \sin \theta \quad , \end{aligned} \quad (8:93)$$

where we have a translation of origins and a rotation. The distance R from the integration point to the observation point (x_0', y_0', z_0') is

$$R^2 = (x_0' - x')^2 + (y_0' - y')^2 + (z_0' - z')^2 \quad (8:94)$$

where

$$x_0' = s \sin \theta, \quad y_0' = 0, \quad z_0' = s \cos \theta \quad (8:95)$$

Substituting (8:95) and (8:92) into (8:94) and keeping terms in $x', y', (x')^2$ and $(y')^2$ in the binomial expression, we find

$$\begin{aligned} R \cong s - x' \sin \theta - z_{xy}x'y' \cos \theta + \frac{(x')^2}{2} \left(\frac{1}{s} - \frac{\sin^2 \theta}{s} - z_{xx} \cos \theta \right) \\ + \frac{(y')^2}{2} \left(\frac{1}{s} - z_{yy} \cos \theta \right) \quad (8:96) \end{aligned}$$

From (8:93) and (8:96)

$$\begin{aligned}
 R + z &= -\sqrt{a^2 - y_0^2} + s - 2z_{xy}x'y' \cos \theta \\
 &+ \frac{(x')^2}{2} \left[\frac{1}{s} \cos^2 \theta - 2z_{xx} \cos \theta \right] \\
 &+ \frac{(y')^2}{2} \left[\frac{1}{s} - 2z_{yy} \cos \theta \right]
 \end{aligned} \tag{8:97}$$

For a sphere

$$\begin{aligned}
 z_{xx} &= z_{yy} = -1/a \\
 z_{xy} &= 0
 \end{aligned} \tag{8:98}$$

and (8:97) becomes

$$\begin{aligned}
 R + z &= \sqrt{a^2 - y_0^2} + s \frac{(x')^2}{2} \left[\frac{1}{s} \cos^2 \theta + \frac{2}{a} \cos \theta \right] \\
 &+ \frac{(y')^2}{2} \left[\frac{1}{s} + \frac{2}{a} \cos \theta \right]
 \end{aligned} \tag{8:99}$$

Substituting (8:99) into (8:91) and extending the limits of integration on x' and y' to infinity in the positive and negative directions gives

$$\begin{aligned}
 U^s &= -e^{-ik(s-\sqrt{a^2-y_0^2})} \\
 &\left\{ \frac{\cos \theta}{\left(\frac{4s^2}{a^2} + 1\right) \cos \theta + \frac{2s}{a} (1 + \cos^2 \theta)} \right\}^{1/2}
 \end{aligned} \tag{8:100}$$

where the factor under the radical is referred to as a divergence factor and shows how the incoming rays are reflected into a diverging bundle of rays.

Near the light-shadow boundary (penumbra) in (F8.8) special attention is required to evaluate the field. A discussion of the problem is given in Jones (1964) and an example is treated by Ott (1970).

The scattering cross section (echo area) of a perfectly conducting sphere is defined as

$$\sigma = 4\pi R^2 |U^s/U^i|^2 \tag{8:101}$$

where U^s is given for example for an acoustically soft sphere in (8:85). In the limit

$$\begin{aligned}
 \sigma &\rightarrow \pi a^2 \quad (\text{geometrical cross section}) \\
 ka &\rightarrow \infty
 \end{aligned} \tag{8:102}$$

and for small ka

$$\begin{aligned}
 \sigma &\rightarrow \frac{9\lambda^2}{4\pi} (ka)^6 (\text{Rayleigh scattering}) \\
 ka &\rightarrow 0
 \end{aligned} \tag{8:103}$$

which is a good approximation when $a/\lambda < 0.1$. The region between the Rayleigh law and the geometrical cross section is called the resonance region and is shown plotted in (F8.9).

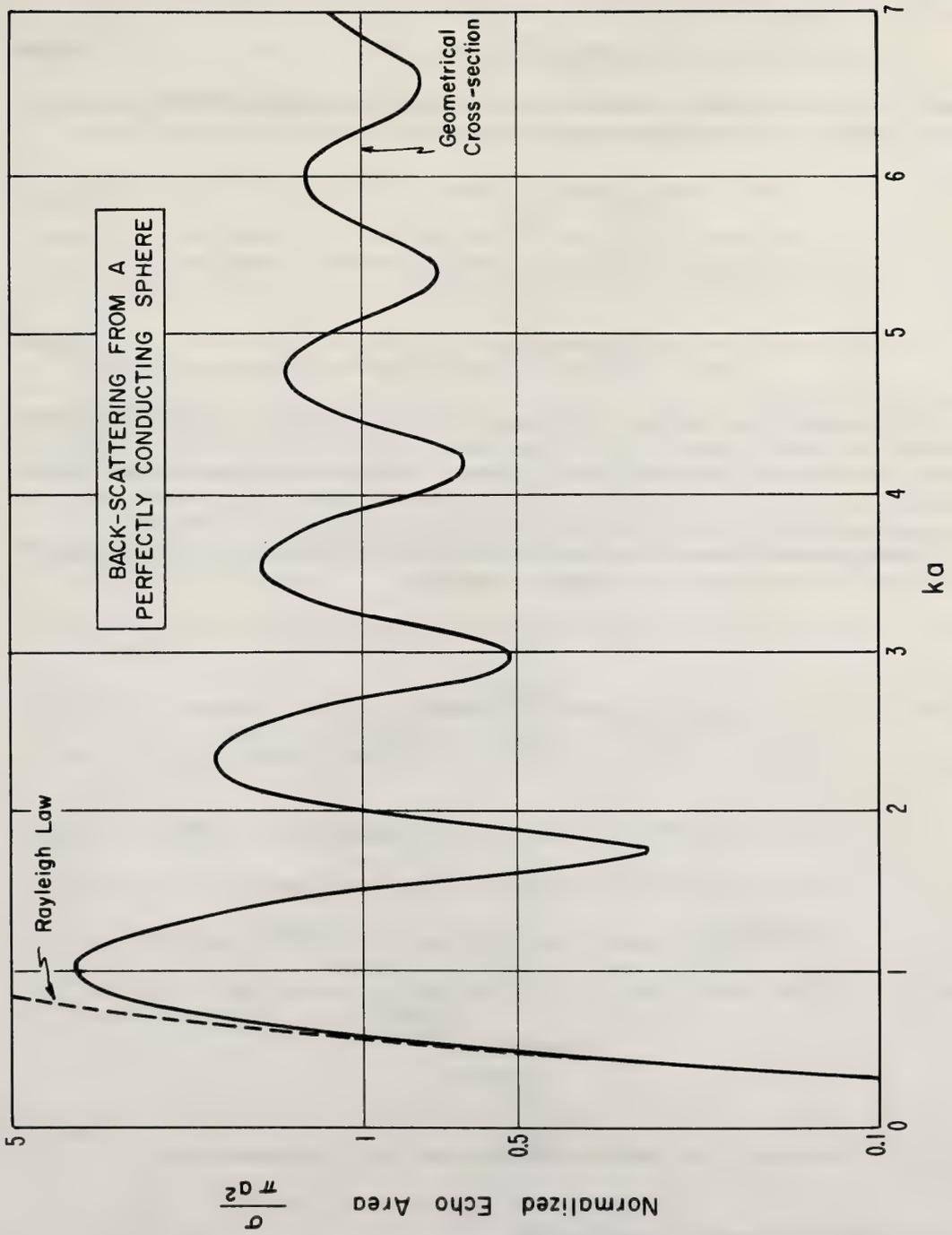


Figure 8.9 Normalized echo area for a perfectly conducting sphere.

8.6.2 Electromagnetic Spheres

The case is now considered where $ka \ll 1$ and $k_s a \ll 1$ (where k_s is the wave number within the sphere) and the source and field points are both far from the sphere. In the case of a small dielectric sphere, it can be shown that (Harrington, 1961) the echo area is given by

$$\sigma = 4\pi a^2 (ka)^4 \left| \frac{\epsilon_s - 1}{\epsilon_s + 2} \right|^2 \tag{8:104}$$

where ϵ_s is the relative dielectric constant of the sphere and $\mu_s = \mu_0$. Making the substitution $a = d/2$, Wilson in chapter 13, equation 13:16 expresses this result as

$$\sigma = \frac{\pi^5 d^6}{\lambda^4} \left| \frac{\epsilon_s - 1}{\epsilon_s + 2} \right|^2 \tag{8:105}$$

The result in (8:104) applies to the case of absorbing spheres with ϵ_s replaced with

$$\epsilon'_s = \epsilon_s - i\sigma/\omega\epsilon_0 \tag{8:106}$$

where σ is the conductivity.

An interesting physical application of Rayleigh scattering is the explanation for why the sky is blue. "The light that reaches us from the sky has been scattered by the air molecules in the atmosphere and possibly by water droplets. The radius of these is less than 0.00002 mm which is small compared with the wavelength (about 0.0006 mm) of visible light. The law of Rayleigh scattering is, therefore, applicable and the shorter wavelengths (blue) are scattered more than the longer wavelengths (red)" (Jones, 1964).

Finally, we wish to discuss the scattering by a small perfectly conducting sphere as a function of θ and φ . In (F8.10) we show the power scattering of a small sphere

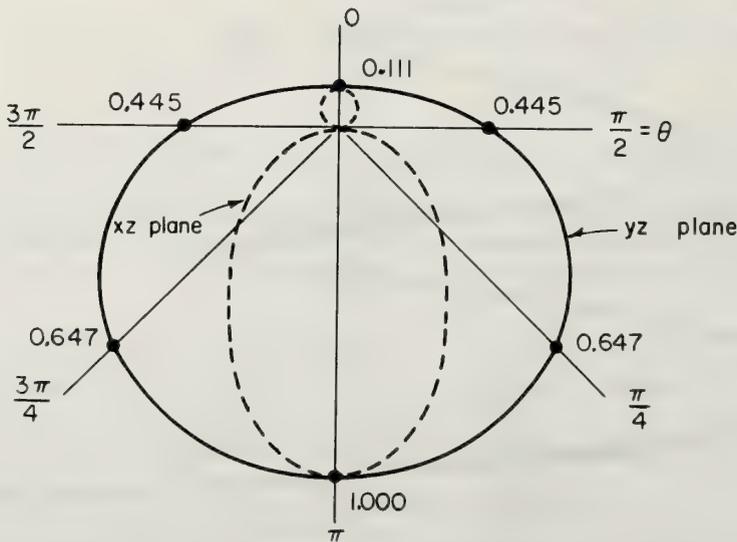


Figure 8.10 Power scattering pattern of a small perfectly conducting sphere.

in the xz plane ($\varphi = 0^\circ$) and the yz plane ($\varphi = \pi/2$).

It is evident from (F8.10) that the backscattering ($\theta = \pi$) is a maximum and is considerably greater than the forward scattering ($\theta = 0^\circ$). When the sphere is constructed of dielectric material the scattered energy is almost uniformly distributed with θ .

8.7 References

- Bremmer, H., 1958: Propagation of Electromagnetic Waves, in *Handbuch Der Physik*, Springer-Verlag, Berlin, Göttingen, Heidelberg, pp. 423-638.
- Copson, E. T., 1965: *Asymptotic Expansions*, Cambridge University Press, Cambridge, England.
- Garbacz, R. J. (1964), Bistatic scattering from a class of loss dielectric spheres with surface impedance boundary conditions, *The Physical Review*, 133, No. 1A, A14-A16, Jan.
- Harrington, R. F., 1961: *Time-Harmonic Electromagnetic Fields*. McGraw-Hill Book Co., Inc., New York.
- Jackson, J. D., 1962: *Classical Electrodynamics*. John Wiley and Sons, Inc., New York.
- Jones, D. S., 1964: *The Theory of Electromagnetism*. MacMillan and Co., New York.
- Keller, J. B., and B. R. Levy (1959), Decay exponents and diffraction coefficients for surface waves on surfaces of non-constant curvature, *IRE Trans. Ant. Prop.* AP-7, S 52-S61.
- Kottler, F. (1923), Electromagnetic theory of bending around dark screens, *Ann. Physik* 71, p. 457.
- Lewis, R. M., Norman Bleistein, and Donald Ludwig (1967), Uniform asymptotic theory of creeping waves, *Comm. Pure Appl. Math.* 20, No. 2, 295-328, May.
- Logan, N. A. (1965), Survey of some early studies of the scattering of plane waves by a sphere, *Proc. of the IEEE*, 53, No. 8, Aug.
- Mie, G (1908), Contributions to the optics of dense media of special colloidal metal solutions, *Ann. Physik* 25, p. 377.
- Ott, R. H. (1970), Alternative method for deriving the Fock currents, *J. Math. Phys.* II, p. 590, No. 2, Feb.
- Schelleng, J. C., C. R. Burrows, and E. B. Ferrell (1933), Ultra-short-wave propagation, *Proc. IRE*, 21, pp. 427-463.
- Schwartz, L., 1950-51: *Théorie des distributions*. Vols. 1 and 2, Hermann and Co., Paris.
- Stratton, J. A., 1941: *Electromagnetic Theory*. McGraw-Hill Book Co., Inc., New York.
- Wait, J. R., 1962: *Electromagnetic Waves in Stratified Media*. Pergamon Press, New York.
- Wait, J. R. (1963), Electromagnetic scattering from a radially inhomogeneous sphere, *Appl. Sci. Res. B* (Netherlands), 10, No. 5-6, pp. 441-450.

List of Symbols

List of Symbols

\underline{E}	The electric intensity	\underline{e}_s	Unit vector along direction of propagation
\underline{H}	The magnetic intensity	U	Scalar potential
\underline{D}	The electric flux density	j_n	Spherical Bessel function
\underline{B}	The magnetic flux density	n	Index of refraction ($n = \sqrt{\epsilon_r \mu_r}$)
\underline{J}	The electric current density	ρ	The electric charge density
\underline{F}	Force	ϵ	Permittivity
\underline{u}	Velocity	μ	Permeability
m	Mass	φ	Scalar potential
\underline{e}_n	Unit normal to surface	ω	Angular frequency
\underline{A}	Vector potential	ψ	Scalar potential
k	Wave number	δ	Dirac delta function
\underline{R}	Position vector	σ	Scattering cross-section
\underline{G}	Green's function	λ	Wave length
S	Surface of integration		
P_n	Legendre polynomial		

Chapter 9 THE SPECTRA OF MOLECULES OF THE EARTH'S TROPOSPHERE

Vernon E. Derr†

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

The spectra of molecules of the earth's troposphere are discussed in a combination of classical and quantum approaches with the objective of gaining insight into spectral characteristics as functions of temperature and pressure. Methods of calculating the attenuation due to the atmosphere as a function of wavelength, temperature, pressure and altitude are given. The spectra of the more abundant natural constituents of the atmosphere are briefly described.

9.0 Introduction

An electromagnetic wave launched into a transmitting medium may change in intensity, phase, direction, or polarization in its passage. These changes are caused by and are characteristic of the medium through which the wave travels; media such as liquid, gases, solids, and aerosols produce a variety of effects. The discussion in this chapter concentrates on the changes caused in electromagnetic transmission by gases, principally those of the earth's atmosphere. The statistical character of atmospheric transmission due to the turbulent motion of the atmosphere is treated in Chapter 11, and aerosols are treated in Chapters 10 and 23. Consequently, this chapter is limited to the interactions of gaseous molecules with electromagnetic radiation, generally in pressure and temperature ranges found in the earth's atmosphere. Because of the brevity required in these notes, they are restricted to only the most elementary treatment of spectra. The objective of the notes will be achieved if the reader begins to obtain an understanding of the relationship between molecular structure and microwave and infrared spectra, and if he perceives the method of calculating attenuation due to atmospheric constituents in practical cases.

When radiation interacts with molecular systems, it can pass through unchanged (transmission), or it can be deviated from its path without change of frequency (Rayleigh scattering), or it can be deviated and changed in frequency (Raman or fluorescent scattering). Also, the energy may disappear from the wave and reappear as collisional energy or heat (collisional relaxation), or remain stored in the internal energy of the molecular system (metastable states). All of these interactions diminish^{††} the amount of energy passing from a source to the detector when a material system is placed in the path. Attenuation of electromagnetic radiation in the atmosphere arises from these molecular effects and from scattering of energy by particles (Mie scattering). The basic spectroscopic processes of emission and absorption of radiation are treated in this chapter. Scattering processes, Raman, and Mie, are treated in Chapters 8 and 10. Mie scattering differs from the other processes because it involves clusters of matter rather than single molecules. It is an important process in the atmosphere and the mathematical treatment has enough in common with other scattering processes to encourage its treatment in the same chapter with molecular scattering processes. Other processes involving particles emitted or absorbed (e.g. the Compton effect) are not treated.

It is not possible, within the limited space available in this chapter to present a detailed account of molecular spectroscopy. However, a summary is given of the basic theory, using classical and quantum concepts freely intermixed in order to present an intuitive understanding, without insisting on rigor. The goal of the chapter is to present the spectral characteristics of the principal gaseous atmospheric constituents under conditions in which they are found in the troposphere.

9.1 The Emission and Absorption of Radiation by Molecules

Experiment has shown that a molecule or atom cannot be composed of electrons behaving in accordance with the Newton equations; therefore it is necessary here to describe such systems quantum mechanically. In order to limit the systems under consideration, only molecules of the earth's atmosphere and their spectra, in a range of frequencies from radio to ultraviolet are discussed. Although all intuitive pictures of

† Adjunct Professor, Dept. of Electrical Engineering, University of Colorado

†† We do not here consider the case where multiple scattering may increase the energy impinging on a detector.

molecules are flawed, it is useful to consider that molecules bound together by complex forces arising from the electrical interactions of electrons and nuclei are like a set of balls (nuclei) held together by springs. This model is used throughout this chapter. Thus, in a state where there is no energy in the system[†], the configuration of the molecule will remain unchanged and it will continue to have the same orientation in space. A microscopic model of this kind would appear to be stimulated to either vibration or rotation by any amount of energy, no matter how small. However, a system of molecular size has only a discrete set of energy states in which it can exist, called quantum states. Electrons bound to the nuclei also may move. Thus, just as in vibration and rotation, molecules may exist also in many different electronic energy states. Quantum systems interact with electromagnetic radiation by absorbing or emitting energy whose frequency is determined by the Bohr rule that the energy change ΔE between energy states is equal to a constant times the frequency of the emitted or absorbed photon, i.e.

$$\Delta E = h\nu \quad (9:1)$$

where h is Planck's constant. The molecule will absorb with high probability only if the frequency ν of the irradiation is close to a possible change in energy of the molecule as determined by $\Delta E = h\nu$. That is to say a molecule in a given energy level can absorb radiation only if the irradiation frequency can closely match the energy difference between the occupied level and another level.

Molecules may have energy in rotational, vibrational and electronic modes of excitation; in addition the electrons and nuclei possess spin energy (i.e. they rotate about an axis through their center of gravity). Generally, large differences in energy are found between these various modes of motion. Interactions may occur between the modes with energy shared or transferred. The spin energy of electrons and nuclei often make small changes in the energy of other modes, producing "fine" structure and "hyperfine" structure in the spectrum (9.4.5). Combinations of rotational, vibrational and electronic modes occur. However, it is frequently possible to consider them separately and study only those energy states which involve vibration and rotation of the molecules, rather than excited electronic states. Usually the lowest energy states of the three modes of excitation, rotational, vibrational and electronic differ greatly and usually the rotational mode is of lowest energy and the electronic mode of highest energy. Approximately, the absorption due to rotation is in the wavelength range from 10^{-1} m to 3×10^{-4} m, the differences between vibrational levels range from 5×10^{-1} m to 5×10^{-7} m, while electronic transitions usually occur at shorter wavelengths. In general, with many exceptions and overlaps, the pure rotational states of molecules have spectra in the microwave, millimeter and submillimeter ranges; the vibrational-rotational states have spectra in the infrared and visible ranges, while electronic spectra occur in the visible and ultraviolet range. At higher frequencies, and thus higher energies (generally in the ultraviolet), ionized and dissociated states occur. Only the vibrational and rotational states will be considered in this chapter, since their application to remote sensing is more immediate.

In summary, in this discussion, we consider the molecules to be capable of rotation and vibration, but with the added quantum-mechanical requirement that the possible energy states for the rotational and vibrational energy are discrete. Since spectra due to electronic motion are not of importance at this time in remote measurement of the lower atmosphere, they are not discussed.

9.1.1 Molecular Energy Levels

The principal features of infrared and microwave spectra can be explained by assuming that the molecules behave as if they were rigid rotators and harmonic oscillators. More subtle spectral features depend on recognizing interactions between rotation and vibration and also electronic and nuclear interactions. However, the classical rigid rotator and the quantum rotator differ in an important characteristic. The solution of the Schrödinger equation^{††} for a rigid rotator shows that it can have only the energy levels $E = h^2 J(J + 1)/8\pi^2 \hat{I}$, $J = 0, 1, 2, \dots$; \hat{I} is the moment of the inertia of the rotator. By comparison with the classical angular momentum, we see that the rotator has an angular momentum of $\sqrt{J(J + 1)}h/2\pi$. In systems which must be described by quantum theory, the angular momentum is allowed to assume only certain specific orientations with respect to an axis. The projection of the angular momentum vector J , in units of $h/2\pi$ is $M = -J, -J + 1,$

[†] Such an oscillator has a "zero-point" energy, i.e. its lowest energy state still has one-half quantum of energy (Herzberg, 1950).

^{††} The Schrödinger equation governs quantum mechanical systems just as Newton's equation governs classical systems.

..., $J - 1, J$. The quantities J and M are integers and are called quantum numbers. The axis is any arbitrary axis; however in actual measurements it is defined by a physical entity such as an electric or magnetic field or the direction of a light beam. Similarly the vibrational and electronic motions of a molecule are quantized, that is, the energy of these modes may not take all values, but only a discrete (perhaps infinite) set of energy values. Similarly, as previously mentioned, the vibrational motion can have only a discrete set of energy values and the energy associated with the electronic motion is also discrete.

Since in quantum theory the emission or absorption of a light quantum occurs as a result of a transition of a system between its allowed quantum states, the frequency of the absorbed or emitted quantum is $\nu = E_n/h - E_m/h$, where E_n and E_m are the energies of the rotational states, n referring to the upper state.

9.1.2 Emission and Absorption Probabilities

In elementary quantum mechanics there are two principle ways of dealing with the problem of absorption and emission of radiation. The first is the use of the radiation theory developed phenomenologically by Einstein and the second is the use of the solution of the time dependent Schrödinger equation to determine transition probabilities of quantum mechanical systems. In this chapter the mathematically simpler phenomenological theory is used. Chapter 10 develops the more sophisticated method necessary to discuss scattering problems.

The Einstein method assumes that a system may exist in a discrete energy state and that the emission or absorption of radiation from the system is accompanied by a decrease or an increase of the energy of the system. We here assume that the system is a molecule. It is further assumed that a radiation field at the molecule may cause the system to absorb energy or to emit energy. Confining our attention to two states we assume that the molecule is irradiated at frequency $\nu_{mn} = (E_n - E_m)/h$. If the molecule is in the lower state, it may go to the upper state by absorbing radiant energy. If it is in the upper state it may be caused to emit radiant energy by the presence of electromagnetic radiation. If no radiation is applied to the molecule it will remain in the lower state, if it is there, and if that is the lowest possible energy state. If it is in any upper state it may undergo a transition to a lower state, emitting radiation, even if no radiation is applied. This process is called spontaneous emission. ("Spontaneous absorption" would be forbidden because of lack of a source of energy.) The processes depending on a radiation field are called induced or stimulated emission, and absorption. A group of isolated excited molecules (that is with internal energy above the ground state) would be observed, according to this theory, to emit radiation with a certain probability, i.e. a certain intensity of radiation would be measured. This is spontaneous emission, if the density of molecules and radiation is sufficiently low to prevent the emitted energy from causing stimulated emission. If this group (or a similar group) is prepared in exactly the same way as before, and illuminated by radiation of a suitable frequency, then the radiation is observed to increase (not counting the applied radiation). The increase in radiation is due to induced radiation. The spontaneous radiation from isolated molecules is incoherent, that is the electromagnetic radiation from each has unrelated, randomly distributed phases. Induced emission has a definite phase and direction relationship to the irradiation; this effect is used in lasers where a strong electromagnetic field is amplified by causing a preponderance of coherent induced emission over the incoherent spontaneous emission. Since the total energy of the radiation emitted in a transition between two states is finite, the radiation must have the characteristics of a "packet" of radiation, usually called a quantum of radiation or a photon.

Clearly, induced emission and absorption should be in some way dependent on the strength of the radiation field. We assume that it is linearly dependent on the energy density of the radiation, that is, the probability of a molecule absorbing a quantum of energy $h\nu_{mn}$ and undergoing a transition in unit time from m to n is

$$P_{mn} = \rho(\nu_{mn})B_{mn} \quad (9:2)$$

where $\rho(\nu_{mn})$ is the radiation density at frequency ν_{mn} and B_{mn} is called the Einstein coefficient of absorption. B_{nm} is the Einstein coefficient of induced emission and in the case of spontaneous emission is replaced by A_{mn} , the Einstein coefficient of spontaneous emission. The radiation density $\rho(\nu_{mn})$ is the time average of the energy density, i.e. $\rho(\nu_{mn}) = (1/2)\epsilon_0 E_0^2$ (joules m^{-3}) where E_0^2 is the time average of the

square of the electric field. This expression is valid for a plane wave. (For isotropic radiation see Panofsky and Phillips (1956)). The values of the coefficients are fundamental in any radiative transition and are independent of whether there is thermal equilibrium. They can be calculated by quantum mechanical methods, but it can be shown by purely statistical considerations (Pauling and Wilson, 1935, p. 301) that they are related by

$$B_{mn} = B_{nm} \quad (9:3)$$

and

$$A_{mn} = \frac{8\pi h\nu_{mn}^3}{c^3} B_{mn} . \quad (9:4)$$

From (9:3) we see that the probabilities of absorption and induced emission are equal. Equation (9:4) shows that spontaneous emission increases rapidly with frequency.

We would like to be able to calculate the values of the Einstein coefficients, in terms of molecular characteristics, for then we could predict the emission and absorption due to the molecule. This calculation is too long to present in this chapter, but the results may be given. (For details see Heitler, 1960 and Pauling and Wilson, 1935.) It can be shown that (Eisberg, 1961) the probability of a spontaneous emission per unit time under plane polarized irradiation is

$$A_{mn} = \frac{64\pi^4 \nu_{mn}^3}{3\epsilon_0 hc^3} |\mu_{mn}|^2 , \quad (9:5)$$

where μ_{mn} is the dipole moment matrix element of the molecule for the transition between the states m and n , and ϵ_0 is the capacitivity[†] (dielectric constant). The absolute value squared of the dipole moment matrix element is defined as the sum of the absolute squares of the x , y , z , components of the dipole matrix element, $|\mu_{mn}|^2 = |\mu_{xmn}|^2 + |\mu_{ymn}|^2 + |\mu_{zmn}|^2$. The components are defined in quantum mechanics as a mixed average over initial and final states of the corresponding component of the electric dipole moment of the molecule, e.g., $\sum_j e x_j$, where e is the electronic charge and x_j the x -th component of the j -th electron. Further

discussion may be found in Eisberg (1961). For the purposes of this chapter, we may consider that the dipole moment matrix element is merely a way of describing the characteristics of the molecule. Alternative descriptions use the related quantities, line strength, and oscillator strength (Green and Wyatt, 1965). Thus the determination of the emission and absorption coefficients for molecules can be reduced to determining the dipole moment matrix element, since from (9:4) we find

$$B_{mn} = \frac{8\pi^3}{3\epsilon_0 h^2} |\mu_{mn}|^2 . \quad (9:5a)$$

9.1.3 Selection Rules

Although a molecule may exist in many states, it cannot, in general, undergo a transition from a given occupied state to any other state. The rules governing the states to which it can go by absorption or emission of a photon are called selection rules. A full treatment of these rules can be found in Herzberg (1945 and 1950). The most important will be summarized here under the frequently made assumption that the interaction of the radiation with the molecule is approximated by an electric dipole interaction. In a molecule, if e_i are the charges of the N particles with coordinates x_i , y_i , z_i , the components of the electric dipole moment μ are $\mu_x = \sum e_i x_i$, etc. The interaction energy of an electric dipole with electromagnetic energy is $\mu \cdot \mathbf{E}$, where \mathbf{E} is the electric vector. By (9:2) and (9:5a) the transition probability from state m to state n under the influence of the radiation is found to be proportional to the $|\mu_{xmn}|^2$ where $\mu_{xmn} = \int \psi_n^* \mu_x \psi_m d\tau$, and ψ_i is the wave function of the i -th unperturbed state. Selection rules (for dipole transitions) are determined by finding those values of n and m for which the matrix elements μ_{mn} are different from zero.

[†] Previously called permittivity.

For diatomic molecules if J is the quantum number of the total angular momentum \mathbf{J} (integer or half integer) and M the quantum number of its projection on an axis, the selection rules require that the changes in J and M in the transition between two levels are $\Delta J = \pm 1$ and $\Delta M = \pm 1, 0$. The rule on M plays a role only in a magnetic or electric field. The vibrational motion also is characterized by a quantum number v . There is no strict selection rule on v , but those transitions for which $\Delta v = \pm 1$ generally have greater intensity (see 9.2.2).

The selection rules discussed above assume that the interaction energy between the electromagnetic field and the molecule depends on the dipole moment. Some molecules may not possess a dipole moment, due to their configuration of charges, but may have higher order electric moments (e.g. electric quadrupole moment), or magnetic moments, which interact with radiation. The selection rules for such transitions are different from those for dipole interactions. Often such higher order transitions are weaker than those involving dipole moments. See Herzberg (1945, 1950).

Further study of selection rules involves the examination of the symmetry properties of the wave functions under reflections at the origin or under exchange of identical nuclei. For details of these rules and application to more complicated molecules, see Herzberg (1950).

9.2 Spectra of Diatomic Molecules

By the use of the Schrödinger equation, the spectra of molecules may be classified, the interactions of various modes of motion may be understood, and to some extent the spectra may be calculated (Pauling and Wilson, 1935; Nielsen, 1959). By the introduction of some experimentally determinable data such as dipole moments, intensities and spectral line frequencies, an essentially complete account may be given of the spectra of atmospheric constituents. The task of determining the spectra is simplified by a most useful approximation due to Born and Oppenheimer; their argument shows that since the nuclei are much more massive than the electrons, the electrons move through many cycles of the motion before the nuclei move appreciably. Hence the electronic motion can be calculated approximately under the assumption of fixed nuclear positions. Thus the Schrödinger equation may be separated into two equations, one describing the electronic motions, the other the motion of the nuclei. In turn the wave equation for the nuclei may be separated approximately into equations for the rotation and vibration of the molecule. As a result of this, the energy of a molecule may, as a first approximation, be expressed as a sum of energies due to the electronic, vibrational, rotational, and interaction energies:

$$E = E_e + E_v + E_r + E_I, \quad (9:6)$$

in which the interaction energy is often small compared to the other terms. Because of these approximations we may treat rotational, vibrational and electronic spectra separately to some extent. However, internal motions of the molecule may occur in which combinations of these three types of motion occur simultaneously, such as rotation-vibration spectra or electronic spectra in which electronic motions are involved with vibrational and rotational motions. For further discussion of these conditions under which the interaction term may be ignored and those under which it is significant, see Nielsen (1959). The translational motion of the molecule enters into the spectra through the Doppler effect (see 9.4.3).

9.2.1 Rotational Spectra of Diatomic Molecules

As mentioned above, the rotation of the molecule as a whole has a total angular momentum $\sqrt{J(J+1)} \hbar/2\pi$ where J can only have the integer values $J = 0, 1, 2, \dots$. The projection of this angular momentum onto any axis (without loss of generality, choose the z -axis) is $M \cdot \hbar/2\pi$ where M can take the integer values from $-J$ to J , that is $-J, -J+1, \dots, 0, 1, \dots, J-1, J$. This is the property of quantization of molecular systems mentioned above. Both the angular momentum and its component are quantized. The axis (z -axis) on which the component of angular momentum is measured is usually defined by an electric or magnetic field, or other physical quantity defining an axis.

From the angular momentum we may, just as in ordinary mechanics, calculate the quantized energy of rotation. It is

$$E_r = \frac{\hbar^2}{8\pi^2 \bar{m} r^2} J(J+1) = \frac{\hbar^2}{8\pi^2 \bar{I}} J(J+1), \quad (9:7)$$

where \bar{m} is the reduced mass of a pair of nuclei, $\bar{m} = m_1 m_2 / (m_1 + m_2)$, and r is the internuclear distance. Clearly it is advantageous to let $B = h^2 / 8\pi^2 \bar{I}$, so that we may write

$$E_r = BJ(J + 1) \quad (9:8)$$

To a first approximation a rotating diatomic molecule may be thought of as a pair of balls connected by a spring. Clearly, as the balls rotate, the spring stretches, increasing the internuclear distance. It can be shown that the energy must be modified to include this effect and can be given approximately by:

$$E_r = BJ(J + 1) - DJ^2(J + 1)^2 \quad (9:9)$$

where D is a small (compared to B) constant. (It is approximately equal to $D = 4B^3 / \omega^2$. The quantity ω is defined in 9.2.2. A typical spectrum of diatomic molecules is shown in (F9.1).

9.2.1.1 Digression on Units

A variety of units have been used traditionally by spectroscopists. They arise from the characteristics of instruments, from convenience, and from choices once convenient and retained through custom. Dispersive spectrographs, using gratings and prisms, most conveniently read the wavelength of spectral lines. Although Angstroms (\AA) are still used as units of wavelength, we will use meters, or powers (possibly negative) of tens of meters. Thus $5000 \text{ \AA} = 5000 \times 10^{-8} \text{ cm} = 500 \text{ nm}$. The now preferred micrometer (μm) is the same as the frequently used micron (μ). However, it should be noted that the photon emitted or absorbed by a molecule has energy equal to a difference of energy levels; its energy is proportional to its frequency. Thus it is convenient to have a unit derivable from wavelength (read from the spectrometer dial) but proportional to frequency (or energy). Such a unit is the wavenumber, given as the reciprocal of the wavelength, $1/\lambda$, where λ is usually given in centimeters. Thus the frequency in much spectroscopic literature is given in cm^{-1} , signified $\tilde{\nu}$. Clearly, since $E = h\nu$ then $E = hc/\lambda$. When the energy is measured in cm^{-1} , the values are called term-values.

Further, microwave spectroscopists measure the position of spectral lines in Hz, MHz, or GHz. It is convenient to remember that 1 cm^{-1} , 30,000 MHz and 1 cm wavelength are equivalent, as are $10,000 \text{ cm}^{-1}$, 10,000 \AA , 1μ , $1 \mu\text{m}$ and 1000 nm. When we wish to specify units, we will let $B_{\tilde{\nu}}$ signify the use of units of cm^{-1} , and B_ν signify the use of frequency units. As an example, for the CO molecule, $B_\nu = 55,344.9 \text{ MHz}$, $D_\nu = 0.174 \text{ MHz}$.

9.2.2 Vibrational Spectra

The vibrational energy levels of diatomic molecules can be determined by the solution of the vibrational part of the separated Schrödinger equation (Pauling and Wilson, 1935; Nielsen, 1959). An analysis of a harmonic oscillator model of a molecule shows that the energy levels are given by $E = \omega_e(v + 1/2)$, where the vibrational frequency ω_e is measured in cm^{-1} . The spectral lines occur then at $\nu = E(v'') - E(v')$ and with the selection rule $\Delta v = \pm 1$, the frequency is $\nu = \omega_e$. The possible values of $v = 0, 1, 2, \dots$

The models serve as a first approximation to the interpretation of infrared spectra, but fail to account for the details observable with high resolution spectrometers. The principal modifications of the theory are the assumptions that the oscillator is anharmonic and that it is a non-rigid rotator. The details of the observed infrared "band" structure of molecular spectra are explained by the formulae arising from these assumptions.

Under these assumptions the energy levels of a diatomic molecule are given by

$$E(J, v) = \omega_e \left(v + \frac{1}{2} \right) - \omega_e x_e \left(v + \frac{1}{2} \right)^2 + J(J + 1)B - DJ^2(J + 1)^2 - a \left(v + \frac{1}{2} \right) J(J + 1) \quad (9:10)$$

where $v = 0, 1, 2, \dots$ and $J = 0, 1, 2, \dots$. The energy levels of a linear triatomic molecule are given by formulas of the type (Herzberg, 1950).

$$E(J, v) = \omega_e \left(v + \frac{1}{2} \right) + x_e \omega_e \left(v + \frac{1}{2} \right)^2 + J(J + 1)B - DJ^2(J + 1)^2 \quad (9:11)$$

Figure (9.1) shows typical energy level structure and spectra of a diatomic molecule. B and D may depend on v .

The energy levels of a vibrating-rotator may be given approximately by $E = \omega_e \left(v + \frac{1}{2} \right) + J(J + 1)B$.

Consider a transition from the state v' (less energy) to state v'' (greater energy), i.e., an absorption spectrum.

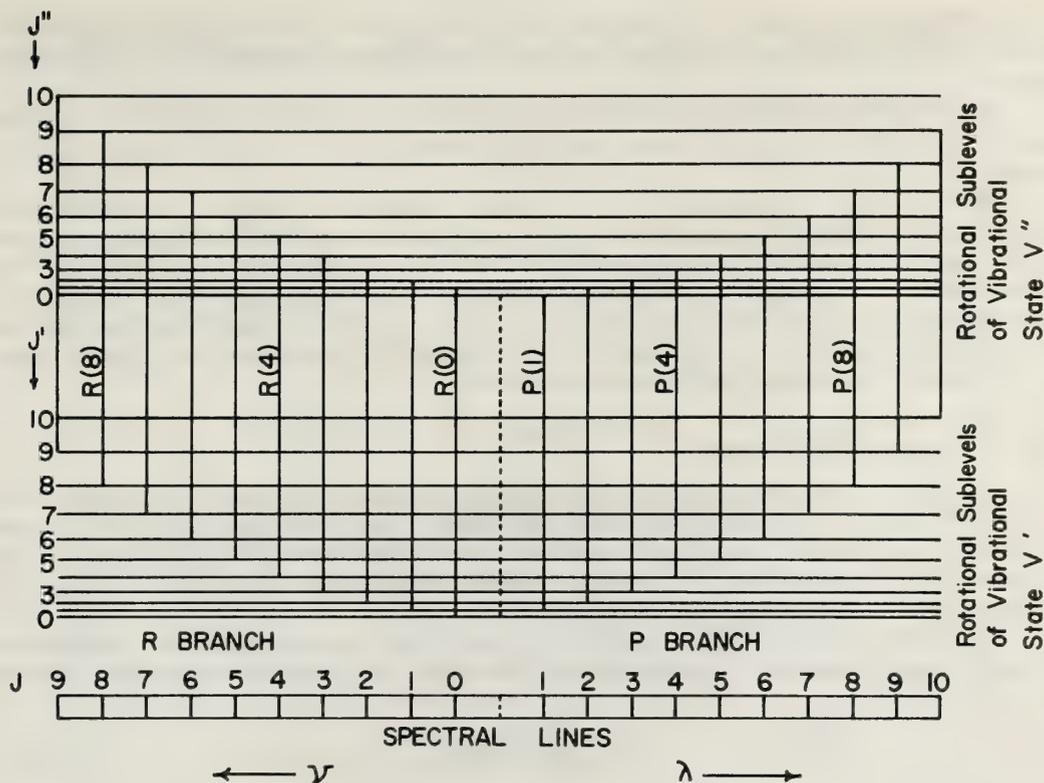


Figure 9.1 Vibrational-rotational transitions and spectra of a diatomic molecule.

Then the frequency of the transition (absorption or emission) is (in cm^{-1}) (B_v is a function of v),

$$\tilde{\nu} = \omega_e(v'' - v') + B_v'' J''(J'' + 1) - B_v' J'(J' + 1) \quad (9:12)$$

The selection rule for v is $v'' - v' = \pm 1, 2, \dots$, and for J , $\Delta J = \pm 1$. ($\Delta v = 0$ is also allowed; this is the pure rotation spectrum.) Let $(v'' - v')\omega_e = \nu_0$, the frequency of the pure vibrational frequency, and examine the two formulas for $\Delta J = 1$, $\Delta J = -1$ separately. We have, respectively,

$$\tilde{\nu}_R = \tilde{\nu}_0 + 2B_v'' + (3B_v'' - B_v')J + (B_v'' - B_v')J^2, \quad J = 0, 1, \dots \quad (9:13)$$

$$\tilde{\nu}_P = \tilde{\nu}_0 - (B_v'' + B_v')J + (B_v'' - B_v')J^2, \quad J = 1, 2, \dots \quad (9:14)$$

In order to simplify the discussion, without losing the main features of the spectrum, we assume that the rotational constants for the upper and lower levels, B'' , B' , are equal. This amounts to an assumption of no interaction between rotational and vibrational motions. We then have, replacing J' by J , and B'' and B' by B ,

$$\tilde{\nu}_R = \tilde{\nu}_0 + 2B + 2BJ, \quad J = 0, 1, \dots \quad (9:13a)$$

and

$$\nu_P = \tilde{\nu}_0 - 2BJ, \quad J = 1, 2, \dots \quad (9:14a)$$

For these special cases (9:13a) gives a series of spectral lines, as J runs over its possible values, of increasing frequency. This set of lines is called the R branch. Equation (9:14a) gives a series of lines of decreasing frequency, known as the P branch (F9.1). For a more detailed discussion of these spectra when the

rotational-vibrational interaction is not neglected, see Herzberg (1967). Typical examples of spectra of diatomic molecules are shown in (F9.2) These are low resolution spectra in which the rotational structure is not seen.

Note that (F9.1) shows no line at $\tilde{\nu} = \tilde{\nu}_0$. Such a line, known as the Q branch, for which $J'' = J'$, seldom occurs in the absorption spectra of diatomic molecules (Herzberg 1950), but is often observed in larger molecules.

Thus, in making a transition from one level (or state) to another, the frequency of the spectral lines is determined by the difference of energy levels as shown above. However, no spectral "line" is perfectly narrow, i.e. it is not exactly monochromatic. The line intensity is a "bell-shaped" curve as a function of frequency. The molecule absorbs or emits strongly at the frequency at the highest point on the curve, but may absorb or emit weakly far from the line center. Thus a material, such as water vapor, with thousands of absorption lines, can have appreciable absorption in a region devoid of line centers, due to the sum of the "tails" of many nearby lines. The subject of the width of spectral lines is discussed in 9.4.

9.2.3 The Boltzmann Distribution Law

Consider a volume containing a gas which has the same temperature throughout. The molecules in this gas, by collision and by radiation, exchange energy. Thus the total energy of the system is distributed in some manner among the molecules. The absorption spectrum of this gas-filled volume is strongly dependent on the distribution of the molecules among the various possible energy states, since the possible molecular transitions, following irradiation, depend on the initial energy states of the molecules.

The Boltzmann distribution law, valid over a very wide range of conditions, states that the number of molecules in a given energy E_n is proportional to the exponential of the energy of that given state in units of kT , i.e.,

$$N_n = G e^{-E_n/kT} \quad (9:15)$$

where k is the Boltzmann constant, and T is Kelvin temperature. G is a constant such that $\sum_{n=1}^{\infty} N_n = 1$. Thus

the number of molecules, in a system in temperature equilibrium, in an energy state n , is a monotonically decreasing function of the energy, when at each energy there exists only one distinguishable state.

Molecules may have many states of the same energy, and these states must be counted separately, since they may each be occupied. When this phenomena, called degeneracy, occurs, we may put the equation explicitly

$$N_n = \frac{N g_n \exp(-E_n/kT)}{\sum_i g_i \exp(-E_i/kT)} \quad (9:16)$$

N is the total number of molecules in the system, and g_i is the number of states having energy E_i . For the lowest vibrational state, in the rotational state J , a diatomic molecule has $g_J = (2J + 1)$ because of the $(2J + 1)$ - fold degeneracy of the rotational states in the absence of an external field, i.e. the $2J + 1$ values of M designated states of equal energy (9.2.1).

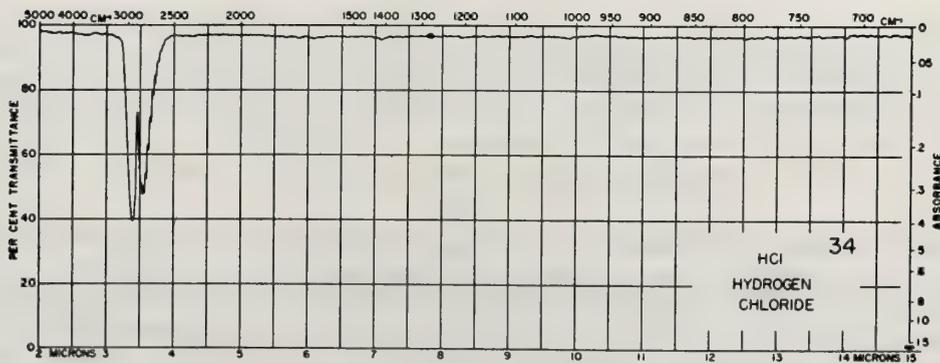
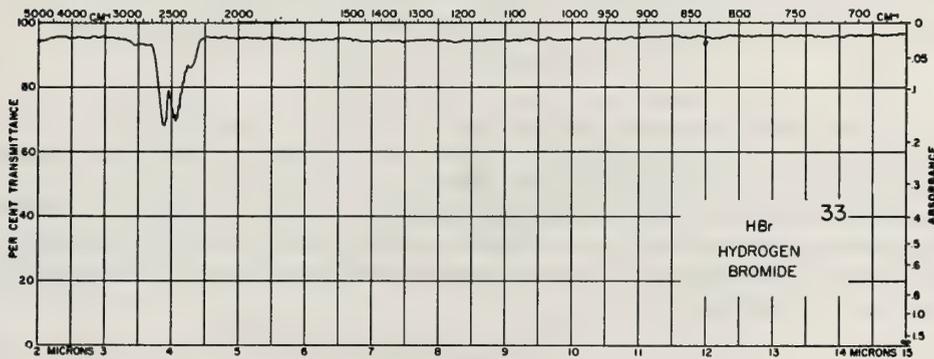
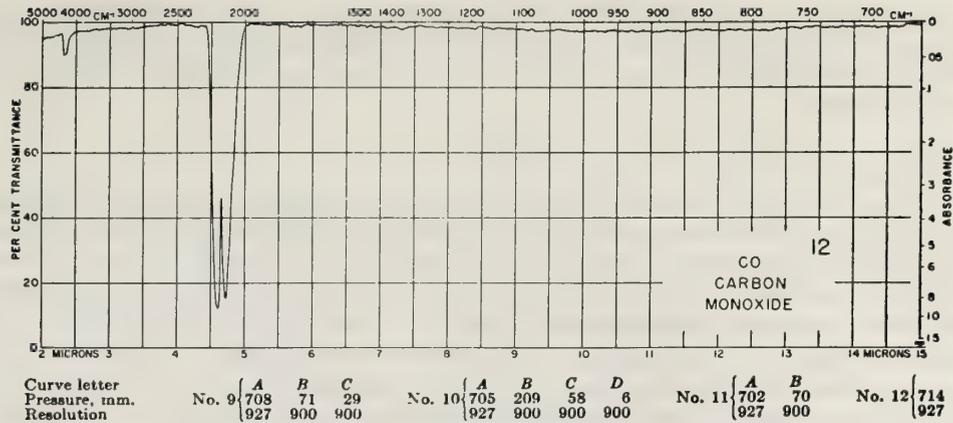
Special cases of this formula are useful (Herzberg, 1950). When there is no degeneracy, e.g. in a vibrational spectrum of a diatomic molecule, then the number of molecules in the v -th vibrational state is

$$N_v = N \exp(-E_v/kT) \quad (9:17)$$

In a rotational state at room temperature, the number of molecules in the J -th rotational state, the degeneracy being $2J + 1$, is

$$N_J = \frac{NB}{kT} (2J + 1) e^{-BJ(J+1)/kT} \quad (9:18)$$

(If E_v is given in cm^{-1} , replace it by $E_v hc$; similarly for B .) The populations given by the above equations determine the quantities N_{ij} of 9.4, below.



Curve letter	No. 32	15 ^{dd}	No. 33	710	No. 34	705	No. 35	A	B
Pressure, mm.		927		900		927		707	56
Resolution		927		900		927		927	900

dd See Table I, *dd*

* Absorbance at about 3.8 μ probably due to impurity.

Figure 9.2 Spectra of typical diatomic molecules (after Pierson, et al., 1956).

9.3 The Spectra of Polyatomic Molecules

When the molecule contains more than two nuclei it has at least the complexity of the famous three body astronomical problem, and further approximations and elegant approaches are needed. The Schrödinger equation can again be separated into a rotational wave equation, a vibrational wave equation and an electronic wave equation, and the total energy is the sum of three separate energies, plus interaction energy, as in (9:6).

In the rotational spectra it is found that the forms of the spectra are quite different for two cases: symmetric-top molecules are those which have two equal principal moments of inertia, asymmetric-top molecules have three unequal moments of inertia. The vibrational spectrum depends on the geometric arrangement of nuclei.

9.3.1 Symmetric-top Molecules – Rotational Spectrum

Two principal moments of inertia can be equal even in molecules without symmetry. No known cases of these occur among the spectra of interest here. The important symmetric tops are those whose moments of inertia are equal, precisely, because of the symmetry of the molecule. In molecules with an axis of three-fold symmetry (i.e. an axis such that rotation by 120° produces the same configuration) or higher symmetry, the two moments of inertia in the plane perpendicular to the axis of symmetry must be equal. We choose I_A to be the moment of inertia along the symmetry axis and A is the corresponding rotational constant. Similar conventions hold for B and C ; here the moments of inertia lie in the plane perpendicular to the symmetry axis. If $I_A < I_B = I_C$, the molecule is called a prolate symmetric top; if $I_A > I_B = I_C$, an oblate symmetric top. These are the shapes of the ellipsoid formed on principal axes with lengths equal to the moments of inertia.

The energy levels of a symmetric top molecule (in the absence of stretching) are given by

$$E(J,K) = BJ(J+1) + (A-B)K^2 \quad (9:19)$$

J is again the total angular momentum quantum number ($J = 0, 1, 2, \dots$) and the square of the projected angular momentum onto an axis, usually taken as the body fixed symmetry axis, is $K^2 h^2 / 4\pi^2$, where K may take the values $-J, -J+1, \dots, J$. Thus J cannot be smaller than K .

For a diagram of the energy levels of symmetric tops see p. 51 of Townes and Schawlow (1955). The allowed frequencies of symmetric top molecules may be determined by the usual difference of energy levels, when the selection rules are known. The selection rules are $\Delta K = 0$, and $\Delta J = \pm 1$ for $K = 0$; $\Delta J = 0, \pm 1$, for $K \neq 0$. Thus the transition frequency is

$$\nu = 2B_v(J+1) \quad (9:20)$$

in the absence of centrifugal stretching and other small effects. (Diatomic molecules are special cases of symmetric tops; hence the spectra of many atmospheric molecules are described by such formulae.)

If the symmetric top is not assumed rigid, the energy levels are better approximated by (Herzberg, 1945)

$$E(J,K) = BJ(J+1) + (A-B)K^2 - DJ^2(J^2+1) \quad (9:21)$$

We note a significant difference between the rotational-vibrational spectra of a simple rotator (9.2.2) and those of a symmetric top molecule. Since $\Delta J = 0$ is permitted for $K \neq 0$, a new series of lines, the Q branch (9.2.2) may appear in the gap between the P and R branches. The only stable diatomic molecule with an *observable* Q branch is nitric oxide (NO), (see F9.3). The reasons for this are discussed in Herzberg (1950). Since K enters as a squared quantity and since for each J there are $J+1$ values of K^2 , then for each value of J there are $J+1$ different energy levels. Because in (9:20) K does not enter the expression for the frequency of symmetric tops there is only one spectral line for each value of J . The transitions are called degenerate in such cases.

When the interaction between centrifugal stretching and rotation are taken into account, the spectral frequencies are given by

$$\nu = 2B(J+1) - 4D_J(J+1)^3 - 2D_{JK}(J+1)K^2 \quad (9:21a)$$

for both prolate and oblate tops. In this case the rotational transition is split into $J+1$ spectral lines, with separation dependent on K^2 .

9.3.2 Asymmetric-top Molecules – Rotational Spectrum

The energy levels of asymmetric-top molecules are considerably more complicated than those of the

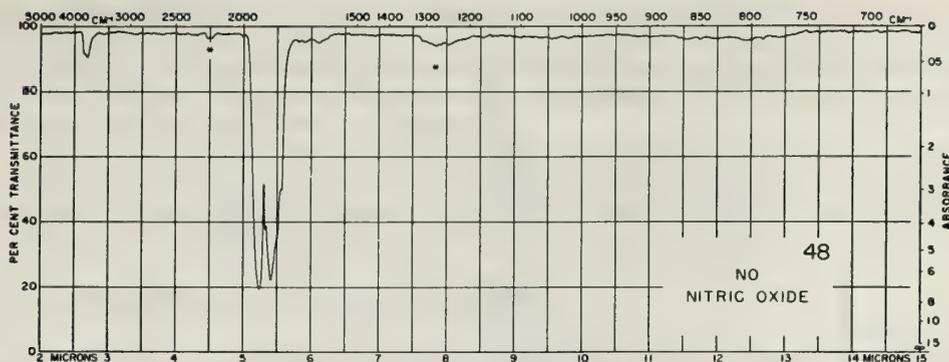


Figure 9.3 Fundamental vibrational band of NO (after Pierson, et al., 1956).

symmetric top and details must be obtained from the references. The rotational energy may be written

$$E_T = \frac{1}{2} (A + C)J(J + 1) + \frac{1}{2} (A - C)E_T(\kappa) \quad (9:22)$$

where E_T is a numeric which can be found in tables given by Townes and Schawlow (1955). E_T is a function of κ ; $\kappa = (2B - A - C)/(A - C)$ is called Ray's asymmetry parameter. It is a measure of the degree of asymmetry of the molecule. τ is $K_{-1} - K_1$, where K_{-1} is the limiting value of the quantum number K for the limiting prolate symmetric-top and K_1 for the limiting case of the oblate symmetric-top. In Townes and Schawlow (1955) can be found a thorough description of the arrangement of energy levels. Consult also Herzberg (1950) and Nielsen (1959).

9.3.3 Polyatomic Molecules – Vibrational Spectra

For simplicity consider the molecule to be represented by a set of balls with springs connecting them, picturing the nuclei bound by forces arising from the electronic and nuclear interactions. This complex arrangement can be simplified by realizing that just as in classical mechanics, any possible motion of the molecule can be represented as a superposition of "normal" vibrations. (The "normal coordinate method" is also used to simplify complex passive electrical circuit analysis.) If the normal coordinates can be determined, all possible vibration modes are determined as linear combinations of the basic set. If the energy of the k -th

vibrational mode is E_k the total energy is $E_v = \sum_{k=1}^{3n} E_k$, where n is the number of atoms in the molecule. Each mode can be described as a harmonic oscillator, hence the total energy can be written (Pauling and Wilson, 1935, Chap. 10)

$$E = \sum_k \left(\nu_k + \frac{1}{2} \right) h \nu_k, \quad (9:23)$$

where $\nu_k = 0, 1, 2, \dots$ and ν_k is the frequency of the k th normal mode of vibration. The molecule H_2O , for example, has three independent vibrational modes which may be specified by values of ν_1, ν_2, ν_3 . In accordance with the quantization discussed in (9.1.1) the normal frequencies are quantized and ν_1, ν_2, ν_3 are their quantum numbers.

The n nuclei of a polyatomic molecule, imagined to be held together by bonds, can have their positions described by giving three Cartesian coordinates for each nucleus. These coordinates, $3n$ in all, are not the most useful for describing the relative motions of the nuclei, which are the vibrational motion. For example, two balls connected by a spring could be described by the translation of the center of mass and the rotation about an axis, and by the distance between the two balls. Only the latter coordinate would be relevant to a study of the vibration. Thus, for vibrational motion we may ignore translation of the center of mass and rotation about 3 independent axes through the center of mass. In general the vibrations are described by $3n - 6$ coordinates.

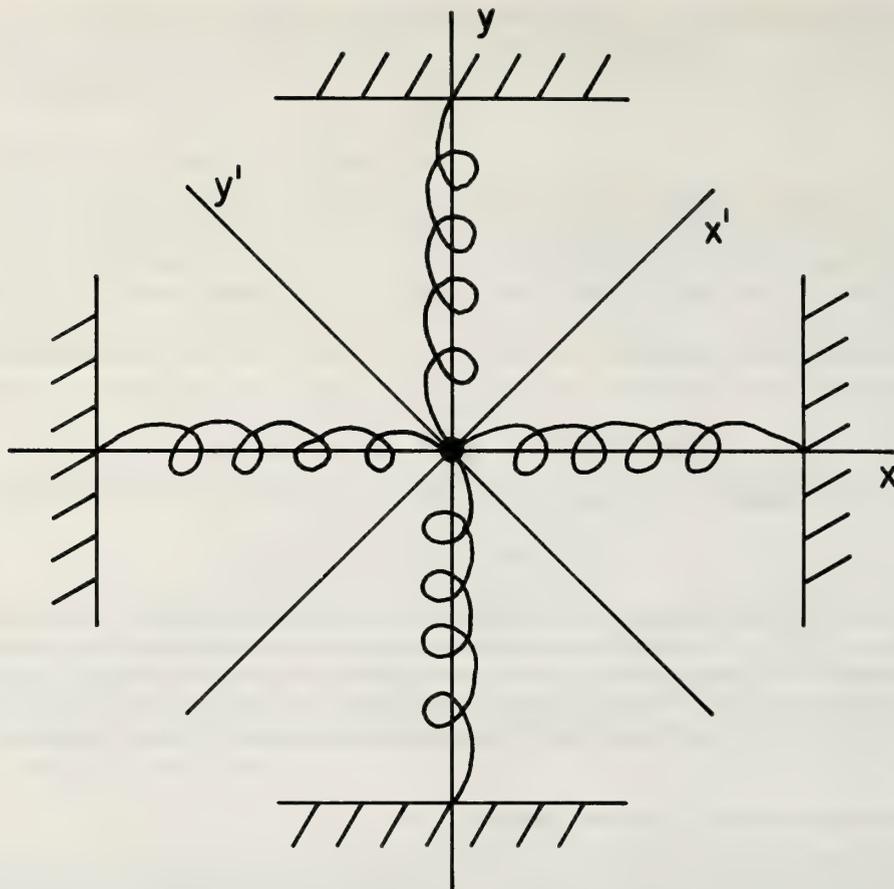


Figure 9.4 Particle, held by springs, in two dimensions, described by two different coordinate systems, $(x - y)$ and $(x' - y')$.

(Linear molecules would have $3n - 5$ such internal degrees of freedom because rotation cannot occur about the linear figure axis.) We may use many different arbitrary coordinate systems to describe vibrations so long as they are independent and $3n - 6$ in number.

As a simple illustration, consider a particle held in two dimensions by springs (F9.4). The supports do not allow translation or rotation. We may see that the selection of the $x'-y'$ coordinate system leads to a complicated description, since the motion along the x' (or y') axis leads to a change of kinetic and potential energy in each spring. If we use an $x-y$ coordinate system, then displacement along each axis is dependent on only one spring. Note that each coordinate system equally well describes the position of the particle and that positions in each coordinate system may be expressed as a linear combination of the other coordinates. The reader may verify that the equations of motion for the particle in the unprimed system are $m\ddot{x} + k_x x = 0$ and $m\ddot{y} + k_y y = 0$, while for the primed coordinates they are $m\ddot{x}' - k_x x' - k_{xy} y' = 0$ and $m\ddot{y}' - k_{xy} x' - k_y y' = 0$, where m is the particle mass, and the k 's are the appropriate spring force constants. If we had started with the "wrong" coordinates and obtained the second set of equations we would be able to simplify them, "uncouple" them, by substituting appropriate coordinates which are a linear combination of the prime coordinates. The mass may carry out complicated motions which are linear combinations of two harmonic motions of (in general) different frequencies.

The more general case of a polyatomic molecule has similar characteristics. Such a system has $3n - 6$ (or $3n - 5$) different normal vibrations. In a normal vibration mode each particle carries out simple harmonic motion and all particles move in phase and have the same frequency of vibration. If the system is randomly

excited, say by random blows to the nuclei, then the resulting complicated motion can be described by a superposition of the normal modes with suitably chosen amplitudes. The amplitudes are quantized as harmonic oscillators, as discussed above. Some of the above discussion must be modified when degeneracy exists. For a discussion of this and other details see Herzberg (1945). Linear vibrations of a linear XYZ molecule (no equal nuclei) and for SO₂ are shown in (F9.5).

Transitions may occur in which one of the vibrational quantum numbers change. If the change is by more than one unit, then resulting band (due to rotational transitions) is called an overtone. If several quantum numbers change during a transition, it produces a combination band. These transitions greatly increase the complexity of the resulting spectrum. For details see Herzberg (1945).

9.4 Linewidth as a Function of Temperature and Pressure

If in the frequency spectrum of any oscillator, classical or quantum, there occurs a region which is relatively narrow, and if the frequency spectrum is a "bell-shaped" function, we may devise a measure of the "width" of this curve. Frequently, in the output of a radio oscillator, or a klystron or magnetron, we may observe such shapes of the frequency spectrum, and also, if the spectrum of an atom or molecule is investigated with sufficient resolution, we may find such shapes. Similarly we find such shapes in the absorption as a function of frequency in physical systems such as L-R-C filters, wave guide filters and the absorption spectra of gases, liquids and solids. We may loosely call such an emission or absorption an absorption line or emission line. However, strictly speaking, in quantum systems, the term line is reserved for the emission or absorption due to a transition between two distinct levels or (somewhat more loosely used) between two sets of unresolved levels. The term "line" arises from the image of a spectrograph slit as seen on a photographic plate. It has been observed experimentally that no spectral line is vanishingly narrow. An observing instrument with sufficient resolving power will always reveal a finite linewidth. The apparent breadth of spectral lines (and bands) of a gas, liquid or solid depends on the pressure, temperature and composition of the sample observed, on the presence of fields, particularly electric and magnetic fields, and on the uniformity of these factors.

9.4.1 The Natural Linewidth

An isolated molecule, not subject to a radiation field, in an upper state with a probability per unit time for spontaneous emission given by (9:5), may go to a lower state with the emission of a photon. If Δt is the lifetime in the state (equal to the reciprocal of the probability per unit time of a transition), the spread of energy $\Delta E (= h\Delta\nu)$ cannot be less than $\Delta E \approx h/\Delta t$, where $\hbar = h/2\pi$ by the Heisenberg uncertainty principle (Eisberg, 1961). But $\Delta t = 1/A_{mn}$. Thus

$$\Delta\nu = \frac{32\pi^3\nu_{mn}^3}{3\epsilon_0hc^3} |\mu_{mn}|^2 \tag{9:24}$$

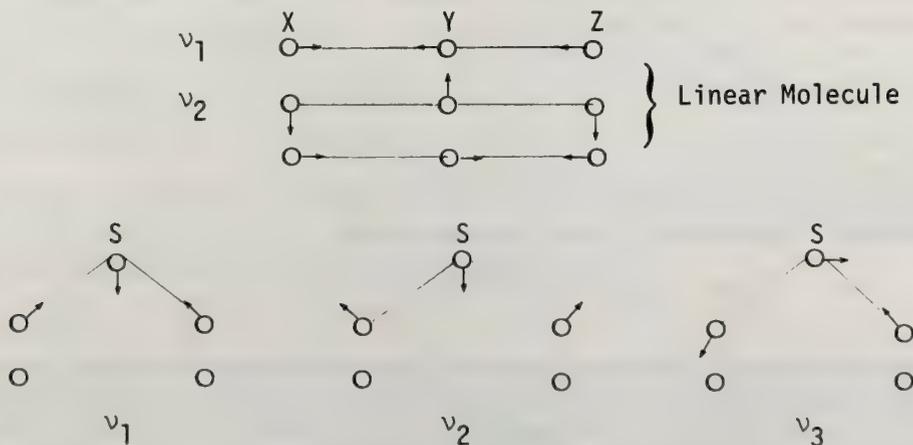


Figure 9.5 Normal vibration modes for linear XYZ molecule and for SO₂.

Since this natural linewidth is proportional to the cube of the transition frequency, it will be seen that it is not important in the microwave or even the infrared region, but becomes increasingly important at visible, ultraviolet and higher frequencies.

9.4.2 Classical Radiator-Linewidth

A radiator damped by radiative reaction, i.e. by loss of energy to the electromagnetic field, produces a field proportional to $E_0 \exp(-i\omega_0 - a/2)t$. We may visualize the radiator as a charge, bound by a harmonic force to a point. The Fourier transform of this field is $E(\omega) = (E_0/2k)/[i(\omega - \omega_0) - a/2]$. Then the radiated intensity as a function of frequency is

$$I = \frac{I_0 a}{2\pi} \frac{1}{(\omega - \omega_0)^2 + a^2/4}, \quad (9:25)$$

with a full frequency width at half intensity of $\Delta\omega \approx a$. Although the form of this line shape for a classical emission and absorption of light is the same as deduced from quantum mechanics, the linewidth has a different frequency dependence (Heitler, 1960). The quantity $1/a$ is the time taken by the energy to fall to $1/e$ of its initial value. It is called the lifetime of the oscillator.

9.4.3 Line Broadening by Collisions and Doppler Effect

Line broadening is caused by collision or interruptions, by Doppler effects, by electric and magnetic fields. Radiative broadening effects (Heitler, 1960) can be significant but will not be treated here.

The Doppler effect, the shift in frequency of radiation from a body in motion, may be calculated by observing that the frequency change is $\Delta\nu = \frac{u}{c} \nu_0$ (ν_0 = center line frequency, u = the component of velocity in the direction of observation), and the probability that u lies between u and $u + du$ is $\sqrt{\frac{\beta}{\pi}} e^{-\beta u^2} du$ ($\beta = \frac{m'}{2RT}$, m' = molecular weight, R = universal gas constant). Then the intensity

$$I(\nu) = D \exp\left[-\beta c^2 (\nu - \nu_0)^2 / \nu_0^2\right] \quad (9:26)$$

where D is a normalizing constant.

Spectral line broadening by interruption caused by collisions of molecules in heat motion is the predominant line width effect at atmospheric pressures. In its simplest form, the theory asserts that the broadening is caused by collisions between molecules, which perturbs the radiator in such a way that its position (x) and velocity (\dot{x}) are distributed in a random way compared with their values before collision. The form of this distribution determines the form of the line width (Breene, 1961). The shape due to Lorentz is obtained by an assumption that the average values of x and \dot{x} after collision are zero and the absorption coefficient has the form for a single absorption line:

$$\hat{k}(\nu) = \frac{8\pi^2 \nu N_{ij}}{3hc\epsilon_0} |\mu_{ij}|^2 \left[\frac{a}{(\nu_{ij} - \nu)^2 + a^2} + \frac{a}{(\nu_{ij} + \nu)^2 + a^2} \right], \quad (9:27)$$

where ν_{ij} = molecular transition frequency, ν = frequency of the incident radiation, N_{ij} = number of molecules in state i minus number in state j (per unit volume). The absorption coefficient is the fractional absorption of radiation, per unit distance of travel. That is, $dI/I = -\hat{k}(\nu)dx$ or $I = I_0 e^{-\hat{k}(\nu)x}$.

By assuming that the distribution of x and \dot{x} is of a Boltzmann type rather than uniform, Van Vleck and Weisskopf (1945) obtained, for the microwave region, the form

$$\hat{k}(\nu) = \frac{8\pi^2 \nu^2 N_{ij}}{3hc\nu_{ij}\epsilon_0} |\mu_{ij}|^2 \left[\frac{a}{(\nu_{ij} - \nu)^2 + a^2} + \frac{a}{(\nu_{ij} + \nu)^2 + a^2} \right]. \quad (9:28)$$

Gross (1955) and Zhevakin and Naumov (1963), using a model of Uhlenbeck and Wang, calculated the distribution of position and velocity from kinetic equations and obtained (for the infrared and optical range)

$$\hat{k}(\nu) = \frac{8\pi^2 \nu_{ij} N_{ij}}{3hc\epsilon_0} |\mu_{ij}|^2 \frac{4\nu^2 a}{(\nu_{ij}^2 - \nu^2)^2 + 4\nu^2 a^2} \quad (9:29)$$

Very few line shapes have been measured with sufficient precision to choose between these shapes. There is some evidence that in the microwave and millimeter regions the Zhevakin-Naumov shape fits best, but in the infrared the Lorentz shape is often chosen for simplicity. With this choice it can be written (for the near infrared) in the form

$$\hat{k}_i(\nu) = \frac{S_i}{\pi} \frac{a_i}{(\nu - \nu_i)^2 + a_i^2} = S_i \cdot b(\nu) \quad (9:30)$$

where S_i is the line intensity (or line strength), a_i the half width at half height, and ν_i is the frequency at peak intensity of the i -th line. The line strength is a measure of the transition probability of a given transition. It is equal to $p_{ij} \cdot 3h\lambda^3/64\pi^4$. When the line shape is introduced, it must be chosen so that $S = \int_{-\infty}^{\infty} \hat{k}(\nu)d\nu$. It is related to the oscillator strength. See Green and Wyatt (1965). Line strengths must often be determined experimentally. The half width is pressure and temperature dependent:

$$a_i(P,T) = a_o \left(\frac{P}{P_o} \right) \left(\frac{T_o}{T} \right)^n \quad (9:31)$$

where P is pressure. The subscripts indicate values at a reference temperature and pressure. The line intensity depends on the temperature only, through the Boltzman distribution factor (Deutschman and Calfee, 1967), and can be expressed as

$$S(T) = S_o \left(\frac{T_o}{T} \right)^m \exp \left[-\frac{E''}{k} \left(\frac{T_o - T}{T_o T} \right) \right], \quad (9:32)$$

where E'' is the upper level. Although the quantities α_o and S_o are theoretically calculable, the difficulties of calculation usually result in the necessity to determine S_o , α_o , n and m experimentally for each line. $P_o = 1$ atm, $T_o = 287.7$ K, $n = 0.62$, and $m = 1.5$ are typical values of these constants for atmospheric water vapor.

The broadening of spectral lines due to the electric fields arising in a plasma containing electrons and ions is not so completely developed that formulae can be given to apply to all cases (Breene, 1961). The mechanism of the broadening is clear although the mathematical difficulties are great. A plasma containing electrons and ions has electric fields due to each of these. The slowly moving ions produce virtually static fields, while the rapidly moving electrons produce electric fields varying a great deal and quickly.

The energy of a molecule in an electric field is dependent on the orientation of its angular momentum vector \mathbf{J} with respect to the field \mathbf{E} . Only a discrete set of orientations of \mathbf{J} to \mathbf{E} may exist, in particular $J + 1$ (or $J + 1/2$) states where J is an integer (or half-integer). The energies of these states are $W = W_o - \bar{\mu}_E E$, where $\bar{\mu}_E$ is the average component of the molecule's electric moment in the field direction, and W_o is the energy without an electric field. The energies of the molecular levels are thus varying rapidly due to the electric field of the ions and electrons and hence the spectral range over which absorption may occur is broadened. The exact spectral range will be determined by the form of the fields (point, dipole, quadrupole) and by the statistics of the relative motion of radiator and perturber. Broadening due to dipoles has the Lorentz form

$$I(\nu) = \frac{S}{\pi} \frac{\delta/2}{(\nu - \nu_o)^2 + (\delta/2)^2}, \quad (9:33)$$

where $I(\nu)$ is the radiated intensity as a function of frequency, S is the line strength, ν_o the resonant frequency and δ , the half width, is proportional to $4.54 \hat{N}\mu$ where \hat{N} is the number of dipoles per c.c. and μ is the dipole moment. For other cases reference may be made to Breene, (1961).

9.4.4 Relaxation

The molecules in a gaseous medium which absorb energy from impinging radiation are put into an excited state. If they were to again fall to the ground state emitting a photon of the same energy through spontaneous emission, they would contribute to the field in a random way, in general not coherent with it and not in the same direction as the incident field. Thus this spontaneous emission represents a loss mechanism in the transmission of energy in a beam from a source to a detector. If the radiation field is sufficiently strong,

induced emission and amplification rather than net absorption will occur, as in laser action, if sufficiently high population of upper levels is achieved (population inversion).

There are other processes involved in the deexcitation or loss of energy of an excited state. Nonradiative deexcitation may occur in which the energy of the molecule, through collision with other molecules, is lost to kinetic energy. In addition, depending on the selection rules, a molecule in an excited state may reach the ground state by several intermediate levels, resulting in the emission of several photons of lower energy than that absorbed, and of course combinations of several processes may occur in a given sample.

If these processes do not operate or operate with insufficient efficiency, the medium may saturate: that is, reach a state where little further absorption may occur since the molecules are not in the lower state from which they start the absorption process. This process can occur at low pressure when the intensity of radiation (I) is so large that the molecules cannot get rid of the absorbed energy fast enough. Under these conditions the molecular energies are no longer distributed as in the Boltzmann manner, $n_0/n_1 = e^{-h\nu/kT}$. The new distribution leads to a different linewidth (Townes and Schawlow, 1955):

$$\hat{k}(\nu) = \frac{8\pi^2 n_0}{3ckT\epsilon_0} |\mu_{ij}|^2 \nu^2 \frac{a}{(\nu - \nu_0)^2 + a^2 + \frac{16\pi^2 t}{3ch} |\mu_{ij}|^2 \nu I a} \quad (9:34)$$

Here n_0 is the number of molecules per unit volume in the ground state i , and $2t$ is the rate of approach to equilibrium of the system, without radiation present, i.e. to the Boltzmann distribution. Thus, under saturation conditions the maximum absorption is decreased and the half width is increased.

9.4.5 Fine and Hyperfine Structure

A single spectral line may be broadened by the processes described above. The line may appear broadened due to unresolved structure in a spectrograph of insufficient resolving power. Two sources of such unresolved structure frequently found in spectra examined with only moderate resolving power are the effects of the electron spin and nuclear spin.

An electron has an intrinsic spin angular momentum S which results in a magnetic moment μ_s . The projection of this magnetic moment on an axis is quantized and can assume only two values, each parallel to the axis, in opposite directions. The quantum number for the spin is s and it has the value $1/2$. The total angular momentum is $S = \sqrt{s(s+1)} h/2\pi$ and the projection $S_z = m_s h/2\pi$ where $m_s = \pm 1/2$ is the quantum number of the projection.

The coupling between the electron magnetic moment and the orbital magnetic moment may produce a splitting of the energy levels for orbital electrons, producing more than one spectral line, lying close together. It is this effect which may produce the broadened appearance of spectra in molecular electronic transitions. The spectral structure is called fine structure.

Hyperfine structure typically three orders of magnitude smaller than fine structure arises from the nuclear magnetic dipole moment μ_i due to nuclear spin angular momentum I_n . Similarly to electron spin $I_n = \sqrt{i(i+1)} h/2\pi$ and $I_{n_z} = m_i h/2\pi$. However, i is not confined to the value $1/2$, and $m_i = -i, -i+1, \dots, +i-1, +i$. The nuclear spin may interact with the other angular momentum of the molecule through its magnetic momentum, splitting energy levels and allowing many close lying spectral lines. For further details see Eisberg (1961).

A further source of splitting and apparent broadening may arise if the sample of molecules has a variety of nuclear isotopes. The spectrum may be changed by mass differences and nuclear spin differences.

9.5 Atmospheric Emission

The atmosphere absorbs and also emits. Calculated emission spectra of the sky have not been successful in predicting the spectral radiance as a function of frequency. The radiation observed from the atmosphere arises from spectral emission excited by the sun, from Rayleigh, Raman and fluorescent scattering, and molecular emission.

Measurements of the spectral radiance of the sky have been made by Bell et al., (1960). The scattering of sunlight is of course only important in the daylight and, because of the dependence of Rayleigh scattering

on the fourth power of the frequency, most important at the higher frequencies, generally higher than green light. The molecular emission of the sky is important at wavelengths longer than $4\mu\text{m}$. Either type of emission is modified by atmospheric absorption bands. In remote probing of the atmosphere by active methods, these emissions form a noise background.

Table (9.1), after Ligda, shows the background noise levels at $\lambda = 0.7\mu\text{m}$. This region is of particular interest because the ruby laser wavelength occurs at 6934 \AA . The ruby laser has been used in laser radar (lidar) form for atmospheric remote sensing and the atmospheric background is of importance in determining its signal-to-noise characteristics.

Table 9.1 Sky Background Noise Levels at 700 nm

Condition	Spectral Radiance ($\text{Wm}^{-2} \text{ ster.}^{-2} \text{ \AA}^{-1}$)
Sun	2.0×10^2
Clouds	$6.0 \times 10^{-3} - 5 \times 10^{-2}$
Moon	3.0×10^{-3}
Clear day sky, blue	$9.0 \times 10^{-4} - 9 \times 10^{-3}$
Dark day sky	$9.0 \times 10^{-6} - 8 \times 10^{-4}$
Twilight	$1.0 \times 10^{-7} - 1.0 \times 10^{-5}$
Moonlit clouds	$8.0 \times 10^{-9} - 8.0 \times 10^{-8}$
Moonlit sky	$1.0 \times 10^{-9} - 1 \times 10^{-8}$
Dark night sky	7.0×10^{-10}

The variation of spectral radiance of the sky, at sea level and at 100,000 feet altitude, as a function of wavelength, is given by Clark (1969).

9.6 Methods of Calculating Atmospheric Attenuation

Spectral absorption by atmospheric gases is a fundamental atmospheric property, important in telecommunications, radar, remote observations by satellite, and the study of atmospheric properties. It is especially of great importance in the study of the heat balance of the earth. Thus the determination of the absorption properties of the atmosphere as a sum of the absorption properties of the separate constituents is basic to many areas of science and engineering.

Unfortunately a simple approach to the problem of determining spectra is not possible. It is impractical to consider a library of sample spectra to consult as needed, since the spectra are strongly dependent on temperature, pressure and on varying constituents such as water, aerosols, and local pollutants. In addition the spectra observed will vary considerably with the resolution of the instrument used to measure them and with the inhomogeneities and length of any path in the atmosphere. Some variation of background would be encountered from night to day because of the scattering of sunlight and molecular emission. In place of such a library, it is possible to compute, for any specified atmospheric situation and any observational mode, the expected spectra, if basic experimental data have been previously obtained.

The calculation of absorption spectra can be a difficult and expensive computer calculation when the number of lines is large and their intensities, line widths and frequency locations are varied and unsystematic. Consequently a variety of methods have been devised to shorten the calculation. Each of these requires some compromise of accuracy.

Elsasser (1942) has used a model of the spectra in which he assumes that the lines within an absorption band have Lorentz shapes, with the same line strength, and are equally spaced. Goody (1964) used a statistical model in which the Lorentz shaped lines were assumed randomly spaced and the line strengths were Poisson distributed. A review of these and other calculation methods is given by Moskalenko and Mirumyants (1970).

However, the solution of atmospheric transmission problems, the determination of the absorptive and emissive properties underlying atmospheric thermodynamics, and the measurement of parameters useful in remote sensing all require an understanding of the "complete detailed spectrum, band for band and line for line." In order to give a complete, reliable answer to the question of, say, the percentage transmission at between 2 and $2.5\mu\text{m}$ over a slant path through the atmosphere it is necessary to have a large amount of data on the line positions, line shapes and widths and line strengths, as well as the parameters of the transmitting

and receiving systems. Not only the lines in this range must be known, but also those nearby, since at higher pressures they will contribute by their collision-caused width. The principal absorbing molecules in this range are H_2O , CO_2 and CH_4 ; but in other ranges, and to some extent in the near infrared, we must include O_2 , N_2 , CO , O_3 , NO_2 , NO , N_2O , HNO_3 , among the normal constituents. The largest permanent dipole moments are possessed by H_2O (1.84×10^{-18} cgs esu (electrical)) and by O_2 (1.854×10^{-20} erg/gauss (magnetic)). Ozone (O_3) is important in the upper atmosphere, although in the lower atmosphere its absorption bands are masked by water vapor and carbon dioxide. Weak lines of CO , NO , NO_2 and N_2O are found at centimeter and millimeter wavelengths. Their low concentration (less than 10 parts per million by volume) and low dipole moments (less than 1.0×10^{-18} cgs esu) do not allow a large contribution. At some wavelengths their effects can be important, as can that of CH_4 . In addition, in some geographical regions, pollutants play a role, and in the upper atmosphere free radicals and molecular ions are important. All of these constituents have regional, seasonal and random variations. The complexity of the problem is illustrated by Figure (9.6), the atmospheric absorption spectrum (Gates, 1960, and Gates and Harrop, 1963), obtained by using the sun as a source, over the range from 1 to $13\mu\text{m}$. Clearly the traditional printed listing of line positions and strengths is not a useful way of collecting or presenting data. Consequently an effort is now under way to collect all the data on the major atmospheric constituents into a format suitable for input into a large computer. This work is being performed at the Wave Propagation Laboratory, NOAA-ERL, Department of Commerce, Boulder, Colorado.

The objective of this work is to provide a source, available to any interested person, of all the data relevant to atmospheric transmission. Questions of atmospheric transmission can be answered by computation in a Fortran program (Deutschman and Calfee, 1967). At present only a small portion of the data is assembled, principally the major normal atmospheric constituents, in the near infrared, from .7 to $15\mu\text{m}$. Some data on atmospheric constituents is available outside the .7- $15\mu\text{m}$ range.

The calculation of the accumulative absorption due to atmospheric gases on a slant path is given in detail by Calfee and Gates (1966) and Deutschman and Calfee (1967). The procedure consists of dividing the path into as many layers as necessary to approximate the physical situation. Each layer is assigned a single pressure, temperature and absorber concentration. The line shape may be chosen in any form when evidence is sufficient to distinguish between the various candidates (9.4), but normally the Lorentz shape is used. The atmospheric transmittance as a function of frequency is $\tau_n(\nu) = \exp[-\hat{k}(\nu)w_n]$, where

$$\hat{k}(\nu) = \sum_i \frac{S_i}{\pi} \frac{\alpha_i}{(\nu - \nu_i)^2 + \alpha_i^2} \quad (\text{see 9.4.3}).$$

w_n is the amount of the absorbing constituent[†] in the n -th layers. The

total transmittance, through i layers, $\tau(\nu) = \prod_{n=1}^i \tau_n(\nu)$. The summation over spectral absorption frequencies ν_i must be extended well beyond the interval of interest because of the cumulative effect of the line tails. The transmittance spectrum must then be degraded by a slit function, its width and shape depending on the system employed in observation or communication. Since S is independent of frequency we see that

$$S = \int_{-\infty}^{\infty} \hat{k}(\nu) d\nu. \quad (9:35)$$

The line strength is a constant for a given temperature and is related to the probabilities of transitions between the initial and final energy levels of the absorbing molecule and to the relative population of these levels (9.4.3).

9.6.1 Effect of Filter Width

The detection of radiation always involves a resolution factor or filter passband. The shape and particularly the width of the filter passband influence the recorded spectrum significantly (Calfee, 1966). Figure (9.7) shows a theoretical, i.e., high resolution, spectrum compared with a spectrum degraded or convoluted by a spectral slit width or filter function. That is, if the spectrum is $I(\nu)$, the degraded spectrum is

$$\bar{I}(\nu) = \int_{\nu-\hat{\nu}/2}^{\nu+\hat{\nu}/2} I(\nu') \Sigma(\nu') d\nu', \quad (9:36)$$

where $\Sigma(\nu')$ is the filter shape and $\hat{\nu}$ is the maximum width of the filter. Such a filter function performs a

[†] For water, the conventional units for w_n are gm/cm^2 ; for other gases the usual units for w_n are atm cm.

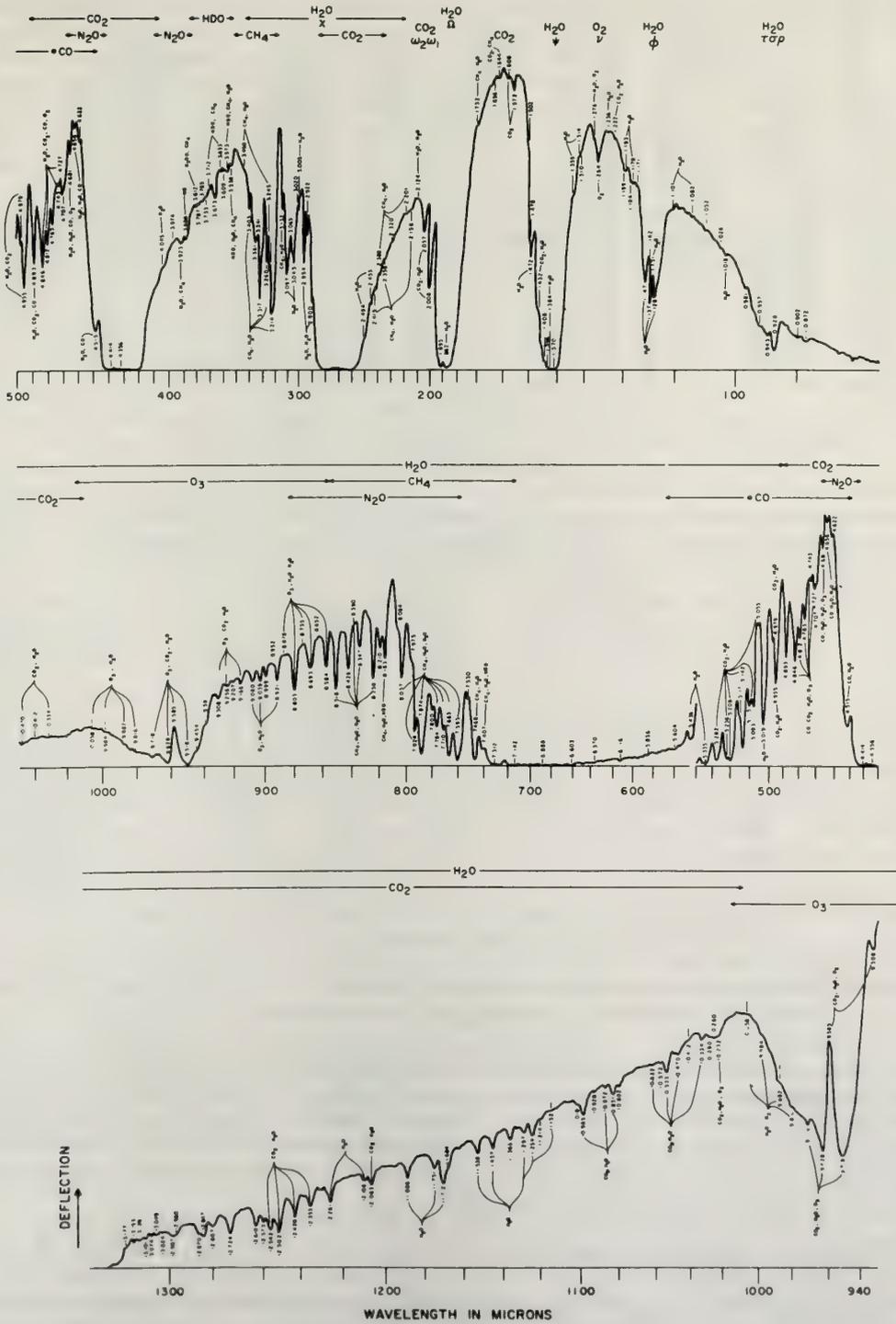


Figure 9.6 Low resolution atmospheric infrared absorption spectrum, 1 to 13 μm , using the sun as a source.

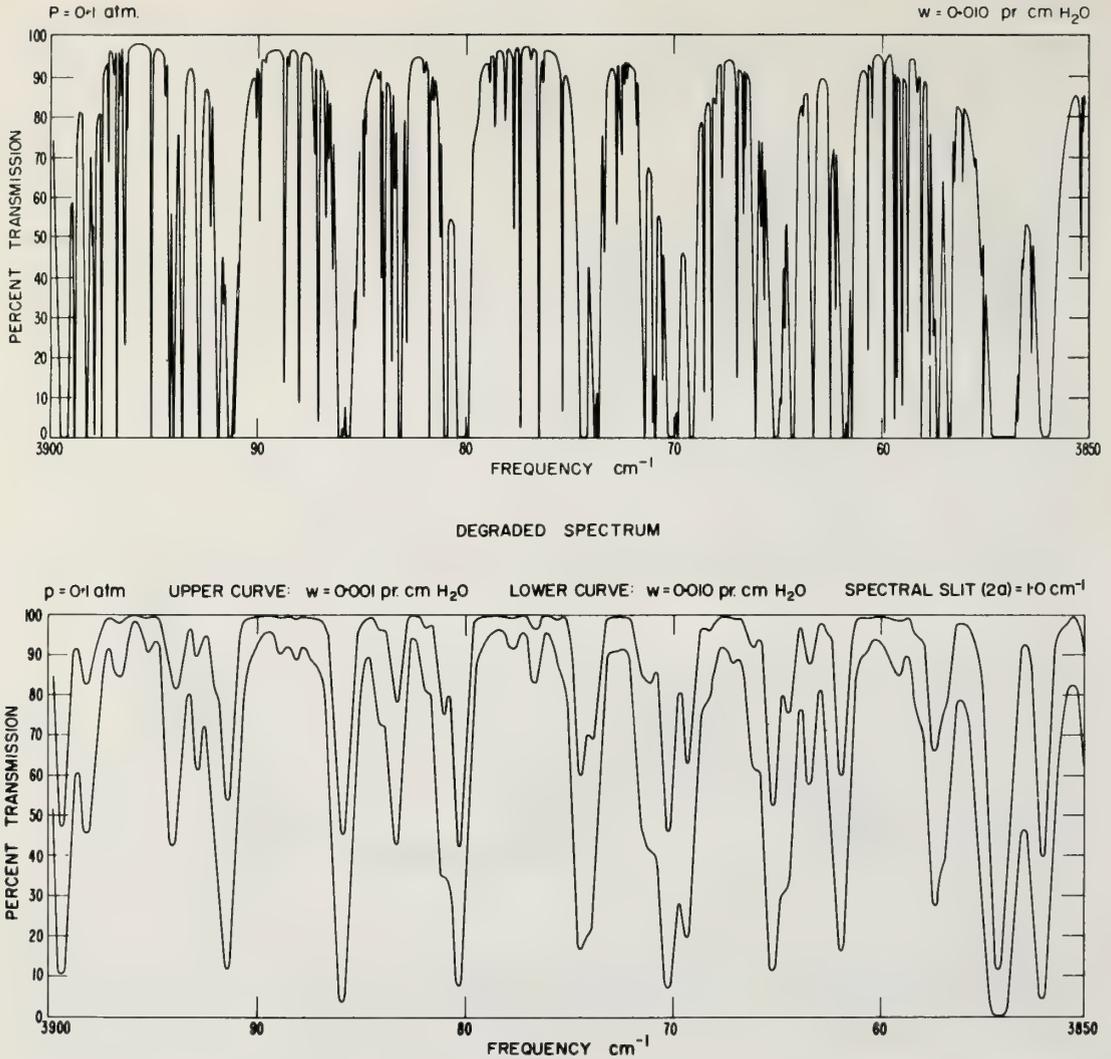


Figure 9.7 Theoretical, high resolution spectrum of H_2O , 3850 to 3900 cm^{-1} , compared with spectrum broadened by instrumental bandwidth (spectrometer slit equivalent to 1 cm^{-1} width).

sliding average of the “true” spectrum. It should be noted how differently the two spectra appear in (F9.6). Given the degraded spectrum, it is possible to compensate mathematically for the filter and to recover the original spectrum if the filter shape is known, and the signal-to-noise ratio is high. Only by the proper use of a slit function in the calculation of a spectrum can a proper comparison be made with experimental results. Also, the spectrum observed with finite resolving power cannot be well compared with computed spectra unless the finite slit width is included in the computation.

9.7 Spectra of Atmospheric Constituents

The objective of this section is to present the absorption[†] spectra of normal atmospheric molecular constituents, principally in the microwave, infrared and visible spectral regions. This restriction is violated when ultraviolet spectra are available and of importance in atmospheric transmission. The frequency of bands

† For emission spectra see references to Herzberg.

and lines, their widths as functions of temperature, pressure and composition and the corresponding transmission as a function of frequency are desirable parameters to list for each molecule. However, such a complete presentation is not possible because of the complexity of many spectra and because many of these quantities are not known. Thus only selected spectra, mostly those of importance in remote sensing, are presented. Sufficient references are given for more complete information.

In contrast to the normal atmospheric molecules, the pollutants form an almost endless list, and their spectra are often more complex and less well recorded or understood. Some of these will be presented in Chapter 24.

In the atmosphere various minor constituents produce weaker absorption bands than water. Water is chiefly confined to the lower atmosphere and since constituents such as CO_2 , CO , N_2O and CH_4 are uniformly distributed, absorption at higher altitudes due to these components become comparatively important since water vapor diminishes. The absorption due to these components can be seen in (F9.6).

9.7.1 N_2

Molecular nitrogen has no permanent dipole moment, hence no microwave or infrared spectra, i.e., no rotational or vibrational spectra. It does have a continuous infrared spectrum due to a collision induced dipole moment and a spectrum arising from electronic transitions, lying in the ultraviolet.

Dipole moments can be induced in molecules during collision processes and are sufficiently strong to produce an observable absorption in long paths through the atmosphere. The absorption due to induced dipole moments in N_2 lies in the region 2400 to 2500 cm^{-1} and in the rotational bands from 300 cm^{-1} to lower frequencies, centered near 100 cm^{-1} (Farmer and Houghton, 1966). At 29 cm^{-1} the absorption due to N_2 accounts for 22% of the total in a vertical path from sea through the atmosphere (Bosomworth and Gush, 1965).

The emission spectrum of the night sky contains blue, violet and near ultraviolet radiation due to the Vegard-Kaplan bands of N_2 . These bands have been observed in gas discharges, but not in absorption (Herzberg, 1959).

9.7.2 O_2

Oxygen, like N_2 , possesses no permanent electric dipole moment, hence exhibits no rotational or vibrational spectrum. However, it has a magnetic moment arising from the interactions of electron spin and the angular momentum of the molecule as a whole. As the result of this unusual situation it does possess a microwave spectrum, near 60 GHz with one transition near 118 GHz. Another series of lines arising from different selection rules, combining transitions between molecular-rotational and electronic spin states, appear in the submillimeter region, extending into the infrared.

Strong O_2 bands occur at 7596 \AA and 12680 \AA . Atmospheric absorption bands are found in the infrared at $1.27\text{ }\mu\text{m}$ and $1.07\text{ }\mu\text{m}$. For some weak O_2 lines near laser wavelengths, see Long (1965).

An interesting phenomenon occurs in atmospheric O_2 which absorbs sound radiation in a moist atmosphere. A near coincidence exists between the lowest vibrational level of oxygen (1556 cm^{-1}) and the infrared vibration of H_2O at 1595 cm^{-1} whose upper state is $v_1 = 0$, $v_2 = 1$, $v_3 = 0$. As a result, the percentage of O_2 molecules, excited by the compression of sound waves into the first vibrational level, which remain in the excited state, may then collide with H_2O molecules and transfer their energy into the 010 vibrational state of H_2O , effectively heating the H_2O by this resonant exchange. The interchange of energy is quite rapid (Henderson, Clark, Lintz, 1965).

9.7.3 H_2O

The most complicated spectrum which we have to deal with in normal atmospheric gases is that of H_2O . This molecule is light, highly anharmonic and has a large centrifugal distortion. It is an asymmetric top (no principle moments of inertial equal) molecule. The complication of its spectrum is such that it is impossible to give a single formula covering all vibrational-rotational spectra from which can be calculated the transition frequencies to high accuracy. The best treatment is to determine tables of energy levels by the use of sum and intercombination rules and expansions over short ranges from which the transition frequencies are determined. Coriolis forces and other perturbations play large roles in some transitions. A comparison of a portion of the CO_2 spectrum with a more complicated H_2O spectrum shown in (F9.8) illustrates the great

difference between the spectrum of a linear molecule and an asymmetric top.

The extremely large attenuation due to atmospheric water vapor, in the infrared and far infrared prevents practical use of these wavelengths for astronomical observation or remote sensing through the lower atmosphere (Derr, 1967; Zhevakin and Naumov, 1963; Gates et al., 1964; Benedict et al., 1952).

In the near infrared, the important bands of H_2O are shown in (F9.6). Detailed data on the $2.7\ \mu\text{m}$ band are given in Gates, et al. (1964) and on the 1.9 and $6.3\ \mu\text{m}$ bands in Benedict and Calfee (1967). The microwave and millimeter wavelength spectrum of water is given by Derr (1967), and by DeLucia et al. (1972). The isotopes of water HDO and D_2O occur. HDO is important in long atmospheric absorption paths. D_2O has been observed in the laboratory.

The water molecule has a much more complicated spectrum than carbon dioxide. It is not linear, but rather of triangular form (see F9.5). The three moments of inertia differ greatly. Such molecules are called asymmetric-top rotators. The asymmetry parameter $\kappa = \frac{2B - A - C}{A - C}$ (where A, B, C are, except for constant multipliers, the inverse moments of inertia), is equal to -0.4377 . The rotational constants are, for the lowest vibrational state (DeLucia et al., 1972), $A = 835,895.31\ \text{MHz}$, $B = 435,156.52\ \text{MHz}$, and $C = 278,278.87\ \text{MHz}$. The general theory of such molecules can be found in Herzberg (1945), and Nielsen (1959).

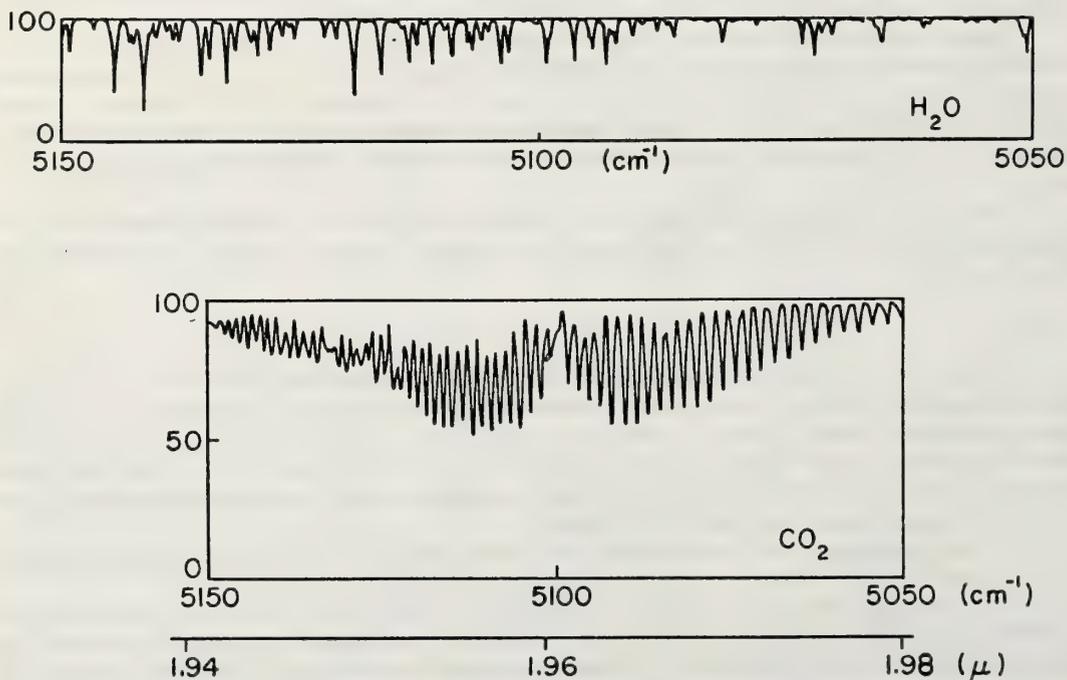


Figure 9.8 Comparison of CO_2 and H_2O spectra.

The rotational constants for water and the observed vibrational states are summarized in Herzberg (1945). Even in a given vibrational state, the constants cannot be used in the usual formulas. They give only an approximation to the calculation of rotation-vibration energies, because of the importance of additional terms in the centrifugal stretching constants and other perturbations.

It is worth emphasizing that the employment of even sixth order terms in the formulas for the rotational transitions in the simpler ground vibrational state is insufficient to give usable accuracy in calculating transitions. Several research groups have performed these very difficult calculations and found that they are not useful for predicting the transition frequencies. Kivelson and Wilson (1952) give an example of an approach which is successful in less asymmetric molecules, but fails in water. A. Clough and W. Benedict, by diagonalizing the Hamiltonian have succeeded in excellent calculational fits of the water spectrum, but their work is not yet published. The basic problem in water is the relating of the energy levels to band constants and the molecular potential function. Difficulties are caused by several types of perturbations which cannot be

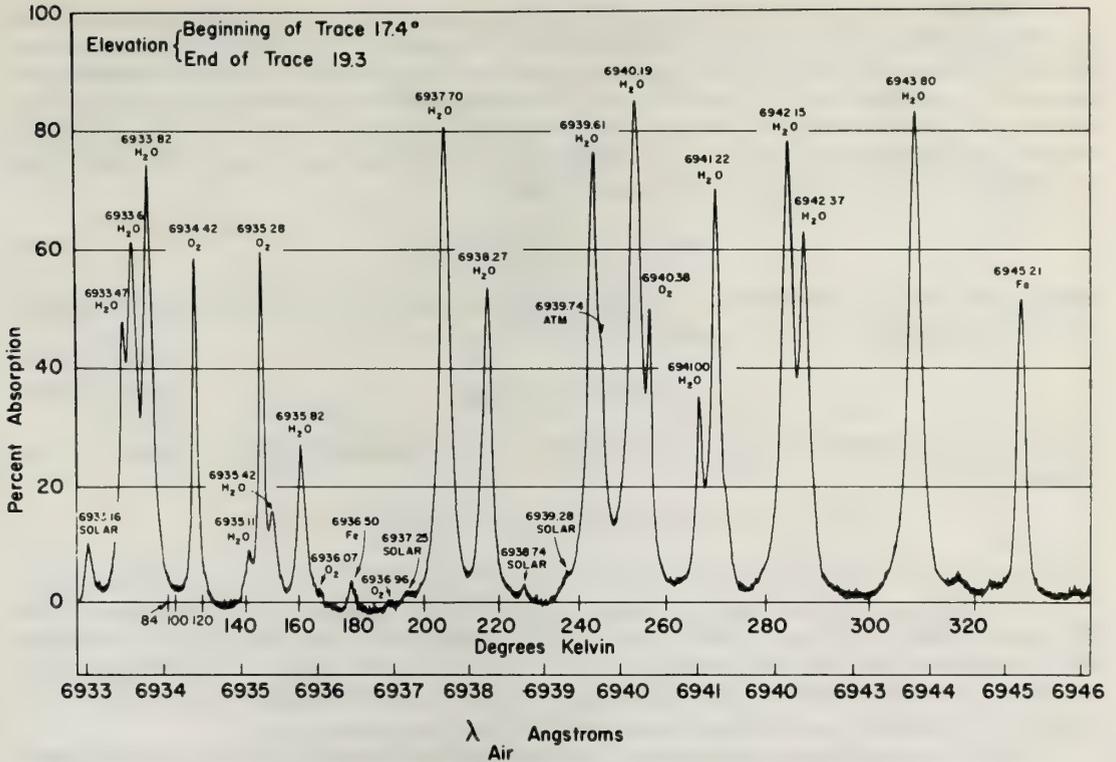


Figure 9.9 Atmospheric absorption spectrum, 6934-6945 Å using the sun as a source (after Long, 1965).

Electric Moment Change	0	C	0	Mode	Frequency (cm ⁻¹)
None	0 →	0	← 0	ν_1	1388
⊥	0	0	0	ν_2	667
	in	out	in		
	← 0	0 →	← 0	ν_3	2349

Figure 9.10 Vibrational modes, CO₂.

neglected in water. (See 9.7.4 for a brief discussion of Fermi resonance.) One important problem arises because the spacing of the rotational levels is not small compared with the vibrational levels. Thus there is an interaction between vibrational and rotational levels which prevents the determination of effective rotational constants. The result is that some lines are shifted and some levels show abnormal intensities.

Although absorption by water vapor in the visible is weak, both it and O₂ have measurable absorption in the region from 540 nm to 1000 nm. Figure(9.10)shows atmospheric absorption the region near the ruby laser line. This spectrum was obtained by Long (1966) through 20 air masses by observing the sun at a very low angle. The long effective path was necessary in order to have observable absorption. Curcio et al. (1964) have compiled information on H₂O and O₂ absorptions over the range 540 nm to 852nm.

Although these absorptions are weak, they have been proposed by various workers as useful for the remote measurement of water vapor content in the earth's atmosphere. The method would be to observe the Rayleigh scattering as modified by atmospheric attenuation on a water absorption line and in a window of the atmospheric absorption, essentially simultaneously. The differential absorption then can be interpreted and related to the water vapor content. The source can be a highly monochromatic ruby laser which can be tuned over a range (by temperature, for example) to include a water line and then a window. The ratio of the on-resonance to off-resonance returns as a function of range is

$$R(z) = C \exp \left[-2 \int_0^z \rho(z') \sigma dz' \right], \quad (9:36)$$

where C is the ratio of on-resonance to off-resonance source intensity and $\rho(z)$ is the density of water vapor and σ is the molecular scattering cross section. The equation may be solved for

$$\rho(z) = \frac{1}{2k} \frac{\partial}{\partial z} \ln R(z). \quad (9:37)$$

Thus water vapor density profiles may be determined in this way by measuring R as a function of z (see Schotland, 1969).

9.7.4 CO₂

Carbon dioxide is a minor but important constituent of the atmosphere. It occupies approximately .033 percent by volume of the normal lower atmosphere. Because of its strong absorption in the infrared it plays an important role in the heat balance of the earth. It is of sufficiently simple structure that its resonant frequencies can be calculated with good accuracy. It is a linear molecule, arranged symmetrically as O-C-O. Since it has no permanent dipole moment it shows no dipole pure rotation spectrum.

The CO₂ molecule has central symmetry and three modes of vibration (F9.10). The motion in general is described by a superposition of normal modes. The ν_1 mode does not absorb infrared radiation (not active in the infrared) and must be observed by Raman spectroscopy. The ν_2 mode is degenerate, since the two modes are identical except for direction. The change in electric moment is perpendicular to the molecular symmetry axis. The change in electric moment for the ν_3 mode is parallel to the symmetry axis.

The frequencies of the CO₂ spectral lines are given by

$$\nu(\nu'J') = \nu(\nu') + J'(J' + 1)B' - J'^2(J' + 1)^2D' \quad (9:38)$$

$$\nu(\nu''J'') = \nu(\nu'') + J''(J'' + 1)B'' - J''^2(J'' + 1)^2D'' \quad (9:39)$$

with the single prime for the upper level, double prime for the lower. Letting $a = B' + B''$, $b = B' - B''$, $c = -2(D' + D'')$, $d = D'' - D'$, we have for the P($\Delta J = -1$) and R($\Delta J = +1$) branches, the transition frequencies

$$\nu(\nu'J' \rightarrow \nu''J'') = \nu_0 + am + (b + d)m^2 + cm^3 + dm^4 \quad (9:40)$$

For the Q branch ($\Delta J = 0$)

$$\nu(\nu'J'' \rightarrow \nu''J'') = \nu_0 + bm + (b + d)m^2 + dm^3 + dm^4 \quad (9:41)$$

For the P branch, $m = -J''$; for the Q branch, $m = J''$; for the R branch, $m = J'' + 1$. The possible values of J depend upon the symmetry character (parity) of the lower state. Only even values of J are allowed for even parity and only odd values are allowed for odd parity (Herzberg, 1945).

It should be noted that a single formula of the type given above can be used in determining the frequencies of the transitions, but only for a given angular momentum condition. For different angular momentum states, the constants must be changed. However, unlike water, within a given vibrational state effective rotational constants can be defined.

A further complication in the CO₂ spectrum occurs because of the "Fermi resonance" which may arise when a binary combination mode and a fundamental mode of vibration (9.3.3) have similar wavenumber

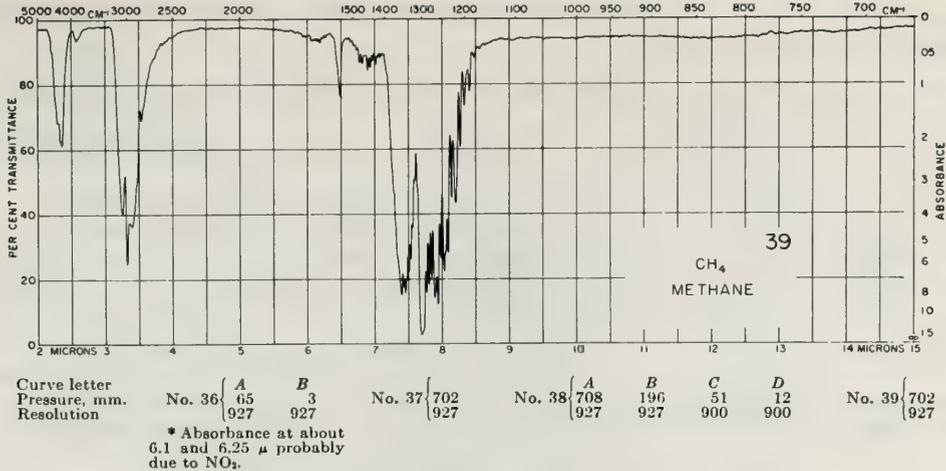


Figure 9.11 Spectrum, 2-15 μm , CH_4 (after Pierson, et al., 1956).

values (see Brittain et al., 1970, for requirements on symmetry). When the interaction occurs, as is the case for the $\nu = 1$ level of the ν_1 band and the $\nu = 2$ level of the ν_2 band in CO_2 , it can lead to a pair of bands with comparable intensity values, in this case near 1388 cm^{-1} and 1285 cm^{-1} . The formulas above must be modified for these cases.

The infrared vibrational spectra are given by Herzberg (1945). The principal bands are at

667.3 cm^{-1}	vs
2349.3	vs
3609	s
3716	s
4860.5	m
4983.5	m
5109	m

where m = medium, s = strong, vs = very strong band strength. Figures (9.6) and (9.8) show some bands of the CO_2 molecule in the infrared. CO_2 has electronic absorption in the visible and near ultraviolet regions.

9.7.5 CH_4 (Methane)

Methane (Herzberg, 1945) has no rotational spectra since it is a spherical top molecule (all three moments of inertia are equal) and possesses no permanent dipole moment. Its infrared vibration spectrum is complicated by Coriolis interactions with overtones of the spectrum. Spectra and lists of lines may be found in Henry, et al. (1970) and Plyer (1960). Figure (9.11) gives a low resolution spectrum of methane in the range 2-15 μm .

The electronic spectrum of CH_4 is insignificant for wavelengths longer than 145.5 nm and consists of continuous absorption bands at shorter wavelengths.

9.7.6 H_2

Because it is a homonuclear molecule, having no dipole moment, molecular hydrogen has no pure rotational spectrum. The vibration-rotation spectrum has bands in the infrared (Herzberg, 1950). Bands occur at visible wavelengths in emission. Electronic transitions occur in the vacuum ultraviolet.

9.7.7 N_2O

Nitrous oxide is a linear polyatomic molecule, in the unsymmetrical form N-N-O. It has a rotational spectrum (Townes and Schawlow, 1955) in the microwave range. Three major infrared bands occur at 588.8,

1285.0, 2223.5 cm^{-1} . Other bands are listed by Herzberg (1945). Figure(9.12)shows the absorption spectrum of N_2O in the 3-16 μm spectral range.

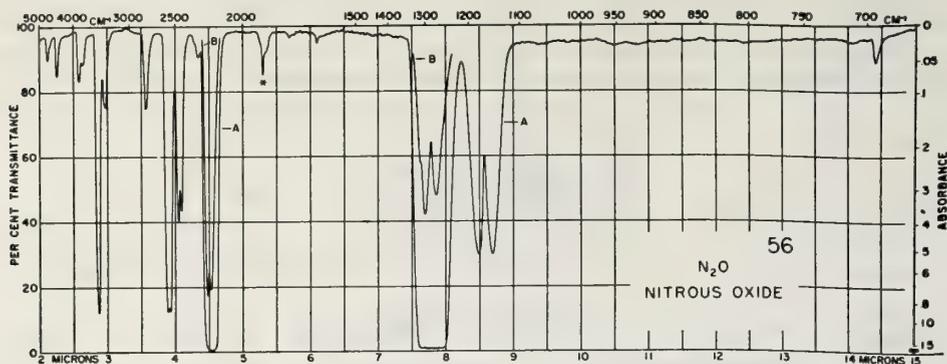


Figure 9.12 Spectrum, 2-15 μm , N_2O (after Person, et al., 1956).

The electronic transitions of N_2O have several broad absorption continua starting at 3065 \AA with maxima near 2900 \AA , 2730 \AA , and 1820 \AA (Herzberg, 1945). There is a window near 2300 \AA .

9.7.8 CO

The microwave rotational spectrum of CO arises from a dipole moment of .10 debyes (Townes and Schawlow, 1955). Its infrared spectrum is given in (F9.2).

The electronic transitions are described by Herzberg (1950).

9.7.9 O_3

The rotational spectrum of ozone is described by Gora (1959), Trambarulo, et al. (1953) and Lichtenstein and Gallagher (1970). A strong complex spectrum starts a lower frequency of 9201 MHz and extends to the infrared region. The rotational constants are $A = 13349.0465$ MHz, $B = 11834.5493$ MHz and $C = 106536.1813$ MHz, but the spectrum is so complicated by internal molecular interactions that the simple rotational spectral formulae (e.g. 9:22) are not very useful.

The infrared spectrum of gaseous ozone has normal vibrational-rotational bands at 710, 1043.3 and 1740 cm^{-1} . Other bands are centered at 2105, 2800 and 3050 cm^{-1} . (See Herzberg, 1945, p. 286.) In the visible faint ozone bands occur at 6020, 5872, and 5750 \AA . The electronic spectrum of O_3 , lying mostly in the ultraviolet region, is of great importance in the earth's atmosphere, since the ozone of the atmosphere absorbs sunlight very strongly. The details of this may be found in references cited by Herzberg (1967, p. 510). The spectrum consists of bands at the following wavelengths: Hartley band (2200 to 3000 \AA , peaked near 2550 \AA); Huggins bands (3000 \AA to 3450 \AA , weak extension to 3740 \AA).

The bands of ozone absorption listed above have a very important consequence for atmospheric remote sensing. Ozone is a very strong absorber of solar radiation in the wavelength region from 230 nm to 300 nm. In the troposphere, for ranges of several kilometers, the absorption by ozone does not prevent laser probing. (More precisely, at 280 nm, the average absorption coefficient in the lowest 6 km of atmosphere is $3 \times 10^{-1} \text{km}^{-1}$, base ten.) However, sunlight is strongly attenuated by ozone predominately occurring between 10 and 40 km. (At 23 km the absorption coefficient is 2.1km^{-1} .) Since, in daylight, the scattered sunlight causes most laser probing devices to be background noise limited, when operating above 300 nm, the elimination of sunlight realized by operation below 300 nm allows a considerable increase in signal-to-noise ratio (See Chapter 23).

9.7.10 HCl

The rotational spectrum of HCl begins in the millimeter region. The dipole moment is 1.18 debye (Townes and Schawlow, 1955). Figure (9.2) shows the fundamental absorption of HCl.

9.7.11 Sources of Information on Atmospheric Spectra

In addition to references already given, the Bibliography contains extensive sources of information on atmospheric spectra.

9.8 Concluding Remarks

It is foreseeable that, in the future, the solutions to problems in atmospheric transmission and emission will be obtained through readily available computer programs. At present this capability exists only for the near infrared. Before it can be completed, the line frequencies, strengths and line width parameters must be determined for all lines of all important atmospheric constituents at all frequencies where telecommunications, radar and atmospheric probing are desirable. In addition, further theoretical and experimental work is needed on line shapes and on absorption in the region between bands. These same methods are applicable to the atmospheres of other planets and to stellar atmospheres.

The author wishes to thank R. F. Calfee and R. L. Schwiesow for extensive aid and discussions. Much of the drudgery of preparing figures was done by Miss K. Yoshida and C. DeHaven, for which I express my great gratitude. Discussions with Dr. Y. Beers have aided very much in accuracy and clarity.

9.9 References

- Benedict, W. S., H. H. Claassen, and J. H. Shaw (1952), Absorption spectrum of water vapor between 4.5 and 13 microns, *J. Res. NBS* 49, 91.
- Benedict, W. S. and R. F. Calfee (1967), Line parameters for the 1.9 and 6.3 micron water vapor bands, U.S. Department of Commerce, ESSA Professional Paper 2.
- Bosomworth, D. P. and H. P. Gush (1965), *Canad. J. Phys.* 43, 729.
- Breene, R. G., 1961: *The shift and shape of spectral lines*, Pergamon Press, Inc., New York.
- Brittain, E. F. H., W. O. George and C. H. J. Wells, 1970: *Introduction to molecular spectroscopy*, Academic Press, New York.
- Calfee, R. F. and D. M. Gates (1966), Calculated slant-path absorption and distribution of atmospheric water vapor, *Appl. Optics* 5, 287.
- Calfee, R. F. (1966), Infrared absorption properties of atmospheric gases, *J. Quant. Spec. Radia. Transfer* 6, 221.
- Calfee, R. F. and W. S. Benedict (1967), Carbon dioxide spectral line positions and intensities calculated for the 2.05 and 2.7 micron regions, U.S. Department of Commerce, NBS Technical Note, 332.
- Clark, W. M. (1969), High altitude daytime sky radiance, *SPIE J.* 7, p. 40.
- Curcio, J. A., L. F. Drummeter and G. L. Knestrick (1964), An atlas of the absorption spectrum of the lower atmosphere from 5400 Å to 8250 Å, *Appl. Opt.* 3, 1401.
- DeLucia, F. C., P. Helminger, R. L. Cook and W. Gordy (1972), Submillimeter microwave spectrum of $H_2^{16}O$, *Phys. Rev. A*, 5, 487.
- Derr, V. E. (1967), Propagation of millimeter and submillimeter waves, Final Report, Contract NAS 12-10, NASA Cambridge, CFSTI, NASA CR-863, Washington, D.C.

- Deutschman, E. M. and R. F. Calfee (1967), Two computer programs to produce theoretical absorption spectra of water vapor and carbon dioxide, U.S. Department of Commerce, NBS Tech. Note 332.
- Eisberg, R. M., 1961: *Fundamentals of Modern Physics*, John Wiley and Sons, Inc., New York.
- Elsasser, W. M., 1942: *Heat transfer by infrared radiation in the atmosphere*, Harvard Met. Studies No. 6 Harvard Univ. Press.
- Falcone, V. J., Jr. (1969), A persistent error in collision-broadened line shape factors, *App. Opt.* 8, p. 2362.
- Farmer, C. B. and J. T. Houghton (1966), Collision-induced absorption in the earth's atmosphere, *Nature* Vol. 209, March 26, 1341-1342.
- Gates, D. M. and W. J. Harrop (1963), Infrared Transmission of the atmosphere to solar radiation, *Appl. Optics* 2, No. 9, 887-898.
- Gates, D. M. (1960), Near infrared atmospheric transmission of solar radiation, *J. Opt. Soc. Amer.* 50, 1299-1304.
- Gates, D. M., R. F. Calfee, D. W. Hansen, and W. S. Benedict (1964), Line parameters and computed spectra for water vapor bands at 2.7μ , U.S. Department of Commerce NBS Mono. 71.
- Goody, R. M., 1964: *Atmospheric Radiation, I. Theoretical Basis*, Oxford University Press, Amen House, London.
- Gora, E. K. (1959), The rotational spectrum of ozones, *J. Mol. Spec.* 3, p. 78.
- Green, A. E. S. and P. J. Wyatt, 1965: *Atomic and Space Physics*, Addison-Wesley Pub. Co., Inc., Reading, Massachusetts.
- Gross, E. P. (1955), Shape of collision-broadened spectral lines, *Phy. Rev.*, Vol. 97, 2, p. 395-403.
- Heitler, W., 1960: *The quantum theory of radiation*, Clarendon Press, Oxford.
- Henderson, M. C., A. V. Clark, and P. R. Lintz (1965), Thermal relaxation in oxygen with H₂O, NDO and D₂O vapors as impurities, *JASA* 37, p. 457.
- Henry, L., M. Husson, R. Andia, A. Valentin (1970), Infrared Absorption Spectrum of Methane from 2884 to 3141 cm^{-1} , *J. Mol. Spec.* 36, 511-520.
- Herzberg, G., 1945: *Infrared and raman spectra of polyatomic molecules*, D. Van Nostrand Company, Inc., Princeton, New Jersey.
- Herzberg, G., 1950: *Spectra of diatomic molecules*, D. Van Nostrand Company, Inc., Princeton, New Jersey.
- Herzberg, G., 1967: *Electronic spectra and electronic structure of polyatomic molecules*, D. Van Nostrand Company, Inc., Princeton, New Jersey.
- Kivelson, D. and E. B. Wilson, Jr. (1952), Approximate treatment of the effect of centrifugal distortion on the rotational energy levels of asymmetric-rotor molecules, *J. Chem. Phys.* 20, 1575.
- Lichtenstein, M. and J. J. Gallagher (1970), Investigation of molecular constituents in the atmosphere, Final Report, AFCRL-71-0006, Contract No. F19628-68-C-0146, AFRCL, Air Force Systems Command.
- Long, R. K. (1965), Atmospheric absorption and laser radiation, Bull. 199, Eng. Exp. Station, Ohio State Univ., Columbus.

- Long, R. K. (1966), Absorption at ruby laser wavelengths for low angle total atmospheric paths, Ohio State Antenna Laboratory Tech. Report 2156-2, 7.
- Moskalenko, N. I. and S. O. Miramyants (1970), Calculation methods of spectral absorption of infrared radiation by atmospheric gases, *Atmos. and Ocean Phys.* 6, 11, pp. 110-1126.
- Oefjen, R. A., E. E. Bell, J. Young, and L. Eisner (1960), Spectral radiance of sky and terrain at wavelengths between 1 and 20 microns, I and II, *JOSA*, 50, 1308-1313.
- Nielson, H. H., 1959: *The vibration-rotation energies of molecules and their spectra in the infrared*, in *Encyclopedia of Physics*, (Ed. S. Flügge), Vol. XXX VII/1, Springer-Verlag, Berlin.
- Panofsky, W. K. H. and M. Phillips, 1956: *Classical electricity and magnetism*, Addison-Wesley Pub. Co., Inc., Reading, Mass.
- Pauling, L. and E. B. Wilson, 1935: *Introduction to quantum mechanics*, McGraw-Hill, New York.
- Plyler, E. K., E. D. Tidwell, L. R. Blaine (1960), Infrared absorption spectrum of Methane from 2470 to 3200 cm^{-1} , *J. Res. NBS* 64A, 201.
- Schotland, R. (1969), Some aspects of remote atmospheric sensing by laser radar, Vol. II of "Atmospheric Exploration by Remote Probes", Final Report of Panel of Remote Atmospheric Probing to NAS, NRC.
- Townes, C. H. and A. L. Schawlow, 1955: *Microwave Spectroscopy*, McGraw-Hill, New York.
- Trambarulo, R., S. N. Ghosh, C. A. Burrus, Jr., and W. Gordy (1953), The molecular structure, dipole moment, and g factor of ozone from its microwave spectrum, *J. Chem. Phys.* 21, p. 581.
- Van Vleck, J. H. and V. F. Weisskopf (1945), On the shape of collision-broadened lines, *Rev. Mod. Phys.* 17, p. 227-236.
- Zhevakin, S. A. and A. P. Naumov (1963), Absorption coefficient of electromagnetic waves by water vapor in the region 2 cm – 10 microns, *Izvestiya Vuzov (Radiofizika)*, 6, 4, 674.

9.10 Bibliography

- Benedict, W. S., A. M. Bass, and E. K. Plyler (1954), Flare-emission spectrum of water vapor in the 1.9 micron region, *J. Res. NBS* 52, No. 3, 161-176.
- Benedict, W. S. (1965), Theoretical studies of high resolution spectra of atmospheric molecules, Final Report, The John Hopkins University, AFCRL Contract AF1 9(604), 6130.
- Bolle, H. J., G. Kuers, F. Moller, and H. Quenzel (1964), Investigation of the infrared emission spectrum of the atmosphere and earth Ludwig-Maximilians-Universitat, Meteorologisches Institut, Munchen, Germany, Part I and II, AFCRL Contract AF 61 (052)-488.
- Bolle, H. J. (1964), Stratospheric water vapor determination by absorption in single lines, *Procs. of International Sym. on Radiation Processes*, Leningrad.
- Bolle, H. J. (1964), Influence of atmospheric absorption and emission on infrared detection range, *Infrared Phys.* 5, 115.

- Burch, D. E., D. Gryvnak, E. B. Singleton, W. L. France, D. Williams (1962), Infrared absorption by carbon dioxide, water vapor and minor atmospheric constituents, AFCRL-62-698, L. G. Hanscom Field, Massachusetts.
- Burch, D. E., D. Gryvnak, R. R. Patty (1964), Absorption by CO₂ between 4500 and 5400 cm⁻¹, Aeronautics, Newport Beach, California, Pub. U-2955, AD-609931.
- Condon, E. U., and G. H. Shortley, 1959: *Theory of atomic spectra*, Cambridge Press.
- Cumming, C., G. R. Hawkins, D. G. McKinnon, J. Rollins, W. R. Stephenson (1965), Quantitative atlas of infrared stratospheric transmission in the 2.7 micron region, Canadian Armament Res. and Dev. Est., CARDE, Technical Report 546/65, Proj. D-46-38-01-19, AD-486695.
- Dobbins, D. L., A. H. LaGrone, M. M. Johnson, & H. R. Arons (1968), Theoretical and experimental study of the absorption of laser signals by atmospheric gases, Report P-29, NSF Grant GA-767, University of Texas, Austin, Texas. (AFCRL-68-0370, AD-676001)
- Erley, D. S. and B. H. Blake (1965 and 1965), Part I: Prism Spectra, Part II: Grating Spectra, Chemical Research Laboratories of the Dow Chemical Company.
- Gryvnak, B. (1966), Absorption by H₂O between 5045-14, 485 cm⁻¹ (0.69-1.98 microns). Aeronautics Pub. U-3704, Newport Beach, California, Contract No. 3560(00).
- Houghton, J. T., N. D. P. Hughes, T. S. Moss, and J. S. Seeley (1961), *An atlas of the Infrared solar spectrum from 1 to 6.5μ observed from a high-altitude aircraft*.
- Howard, J. N. and J. S. Garing (1961), Infrared atmospheric transmission: some source papers on the solar spectrum from 3 to 15 microns. AFCRL, Bedford, Mass., AFCRL 1098.
- Jackson, J. D. (1962), *Classical electrodynamics*, John Wiley and Sons.
- Jager, C., Dordrecht, Holland, and D. Reidel, *The Solar Spectrum*, Proceedings of the Symposium held at the University of Utrecht, 26-31 August 1963.
- Liebe, H., M. C. Thompson, Jr., and T. A. Dillon (1968), Dispersion studies of the 22 GHz water vapor line shape, I. The Lorentzian Behavior, J. Quart. Spec. and Rad. Trans.
- McClatchey, R. A., R. W. Fenn, & J. E. Selby (1971), Optical properties of the atmosphere (Revised). Air Force Systems Command USAF, AFCRL 71-0279.
- Migeotte, M., L. Neven, J. Swenson, and W. S. Benedict (1957), Solar spectrum from 2.18 to 23.7 microns, Part II, Measures and identifications, QC455M3, Spec. Vol. 2, Mem. de la Soc. Roy des Sci. de Liege.
- Minnaert, G. F., W. Mulders, J. Hougast, D. Amsterdam, D. Schnabel, Kampert and Helm (1940), Photometric atlas of the solar spectrum from λ3612 to λ8771 with an appendix from λ3332 to λ3637.
- Mohler, O. C., 1955: A table of solar spectrum wavelengths, 11984 Å to 25578 Å, Univ. of Michigan Press, Ann Arbor, Michigan.
- Mohler, O. C., A. K. Pierce, R. R. McMath and L. Goldberg, 1950: *Photometric atlas of the near infrared solar spectrum, λ8465 to λ25,242*, University of Michigan Press, Ann Arbor, Michigan.
- Moore, S. E., G. F. W. Minnaert, and J. Houtgast (1966), The solar spectrum 2935A to 8770A, National Bureau of Standards, Washington, NBS Monograph 61.

- Panofsky, W. and M. Phillips, 1955: *Classical electricity and magnetism*, Addison-Wesley, Reading, Massachusetts.
- Pierson, R. H., A. N. Fletcher, and E. S. C. Gantz (1956), Catalog of infrared spectra for qualitative analysis of gases, Vol. 28, No. 8, IER, pp. 1218-1239.
- Sakurai, K. and H. P. Broida (1968), On NO₂ fluorescence by 11 argon and krypton lines, Quantum Elec. Conf., Miami, Florida.
- Sadtler Standard Spectra, Sadtler Research Laboratories, Philadelphia, Penn.
- Spectra of Molecules Relevant to Quantitative Analysis of Air Pollutants, Beckman Reprint.
- Swenson, J. W., W. S. Benedict, L. Delbouille, and G. Roland Liege (1970), The Solar Spectrum from λ 7498 to λ 12016; a table of measures and identifications, Societe Royale des Sciences de Liege.
- Thompson, W. I. (1971), Atmospheric transmission handbook, NASA, Report No. DOT-TSC-NASA-1-6.
- Yates, H. W. and J. H. Taylor (1960), Infrared transmission of the atmosphere, U.S. Naval Res. Lab., Washington, D.C., NRL Rep. No. 5453, June 8.

List of Symbols

A	rotational spectrum number	E _O	electric field amplitude
Å	Ångstrom units, 1Å = 10 ⁻⁸ cm	E _T	rotational energy
a, b, c, d	constants for calculating CO ₂ spectra	E _V	vibrational energy
A _{nm}	Einstein coefficient of spontaneous emission	E _T (κ)	numeric energy function for asymmetric top
B	rotational spectrum constant (B _V is a function of v)	G	constant in Boltzmann distribution
B _{mn} , B _{nm}	Einstein coefficient of absorption, induced emission	g, g _J	degree of degeneracy
C	rotational quantum number	h	Planck's constant, 6.62517 × 10 ⁻³⁴ joule sec
c	velocity of light, 2.997930 × 10 ⁸ ms ⁻¹	ħ	h/2π
D, D _J , D _{JK}	rotational centrifugal effect constant	I	radiation intensity
E	energy	i	nuclear spin quantum number
e	electronic charge, 1.60206 × 10 ⁻¹⁹ Coulomb	Î	moment of inertia
E	electric field vector	I _n	nuclear spin angular momentum
E _e	electronic energy of molecule	I(ν)	radiation intensity
E _I	interaction energy of molecule	J	angular momentum quantum number
E _n	energy of state n	K	quantum number of angular momentum, projection on body fixed axis
		k	Boltzmann's constant, 1.38044 × 10 ⁻²³ J K ⁻¹

K_1	K for oblate symmetric top	β	normalizing constant, Doppler broadening equation
K_{-1}	K for prolate symmetric top		
k_x, k_y	spring force constant	Δt	lifetime of a state
$\hat{k}(\nu)$	absorption coefficient	$\Delta\nu$	frequency spread
M	quantum number of angular momentum projection on arbitrary axis	δ	line broadening parameter
		ϵ_0	capacitivity of vacuum, 8.85434×10^{-12} farad m^{-1}
m	constant for calculating CO ₂ spectra	κ	Ray's asymmetry parameter
m	mass	λ	wavelength of radiation
\bar{m}	reduced mass	$\underline{\mu}$	dipole moment
m'	molecular weight	μE	average component of electric moment
m_i	nuclear spin projection quantum number		
		μ_{mn}	dipole moment matrix element
m_s	electron spin projection quantum number	$\underline{\mu}_s$	electron magnetic moment
N	total number of molecules	ν	radiation frequency
\hat{N}	no. dipoles per c.c.	$\tilde{\nu}$	frequency in cm^{-1} units
N_{ij}	difference in occupations of states i, j	ν_0	vibrational frequency in cm^{-1}
		$\rho(\nu_{mn})$	radiation density at frequency ν_{mn}
N_n	number of molecules in energy level E_n	$\rho(Z)$	density of water vapor
		Σ	slit function
P	pressure	σ	molecular scattering cross section
P_{mn}	probability of transition, state m to state n	τ	$K_{-1} - K_1$
R	gas constant 8.31662×10^3 J(kg mole) ⁻¹ K ⁻¹	$\tau_n(\nu)$	transmittance
		Ψ	quantum mechanical state function
$R(Z)$	ratio of on-to off-resonance returns from atmospheric water	ω_e	frequency of vibrational motion
S, S_i	line strength	$\omega_r X_e$	anharmonicity constant, vibrational spectra
s	electron spin quantum number		
s	electron spin angular momentum		
T	Kelvin temperature (K)		
t	one-half rate of approach to equilibrium		
u	velocity component		
v	vibrational quantum number		
W, W_0	energy		
w_n	amount of absorbing constituent		
a	damping constant, halfwidth		

Chapter 10 ATOMIC, MOLECULAR, PARTICULATE, AND COLLECTIVE GENERALIZED SCATTERING

R. L. Schwiesow

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

This chapter discusses scattering theory from three successive points of view. A quantum mechanical treatment of scattering from molecules yields scattering cross sections for resonance, fluorescence, Rayleigh, and Raman scattering. Single particle scattering, or Mie scattering, from dielectrics, is analyzed classically. These results are then applied to collections of scatterers to illustrate collective effects such as coherent scatter from high density scatterers, scattering from moving particles and scattering from acoustic waves. Tropospheric applications are repeatedly used to illustrate the results.

10.0 Introduction

Scattering is important to remote sensing both from the point of view of a sensing interaction and because of interference with the probing electromagnetic radiation. In discussing scattering mechanisms, we will build on the previous spectroscopic chapter for quantum mechanical principles. In turn, the results of this chapter will help in understanding atmospheric constituent sensing methods and Chapter 11 on wave propagation in a scattering atmosphere.

All scattering of concern to this chapter involves the interaction of a photon with the bound electrons of a material particle. In particular, we exclude nuclear scattering, scattering of x-rays, scattering from free electrons and scattering from conductors with free electrons. The scattering of photons of moderate (ultraviolet or less) energy from dielectric systems includes all processes of significant concern to remote sensing of the troposphere. Even with these limitations, the topic is far too broad to treat in rigorous detail. We concentrate on important results, emphasizing applications and limitations, and discussing the derivative concepts important for each scattering result. Similarly, we refer the reader to particularly valuable source literature, with no attempt to do justice to the original papers in the field.

Although the interacting elements, photons and bound electrons, are identical for all types of tropospheric scattering, the subject naturally divides itself into two types of scattering. Fluorescence (10.2.2), resonance re-radiation (10.2.2) and Raman scattering (10.3.2) require a quantum mechanical treatment. Rayleigh scattering (10.3.2) is most simply, but not necessarily, treated by quantum techniques as well. Mie scattering (10.4) is best analyzed by classical electromagnetic wave formalism. Sections 10.1, 10.2, and 10.3 provide a quantum mechanical treatment of the applicable scattering types while 10.4 reviews the classical analysis of single-particle scattering. Section 10.5 analyzes scattering from collections of scattering centers.

Terminology in scattering is not uniform. One finds various definitions, exclusions and historical arguments in the literature. The labels assumed without definition in our introduction are most commonly used. They will be defined more effectively as the relevant theory unfolds. Some authors consider that fluorescence and resonance re-radiation are not true scattering since they are actually an absorption followed later by the emission of a photon, whereas Raman and Rayleigh scattering are single-step interactions involving two photons interacting simultaneously with a molecule. For atmospheric purposes, it is useful to treat these effects together. Similarly, certain authors (we are included) prefer to call classical scattering without frequency shift Tyndall scattering, although a majority of experimentalists refer to this as Mie scattering. Useful expositions of the history of scattering theory and terminology are given by Kerker (1969) and Placzek (1934). It is important to realize the existence of these different viewpoints, and to make appropriate mental bridges, when studying the source literature on scattering theory.

10.1 Review of Atomic and Molecular Theory

We specialize here the notions of 9.1 for scattering purposes. In the process we will point out significant signposts along the quantum mechanical road to the scattering results, without any attempt to be rigorous or convincing. The reader interested in a complete development will refer to the texts mentioned,

particularly Dirac (1958), Ditchburn (1963) and Placzek (1934). Bridge and Buckingham (1966) consider scattering theory as applied to the measurement of depolarization ratios in gases. This latter exposition is basically a special case in more detail of the more general theory presented by Dirac (1958) and Placzek (1934).

Quantum mechanics postulates a function $\Psi(\mathbf{r})$, called a wave function or state vector, which characterizes the state of a system of material particles. The \mathbf{r} is a generalized position vector in a space of n dimensions where n is the number of degrees of freedom of the system. The term, $\Psi(\mathbf{r})$, is normalized so that

$$\int \Psi^*(\mathbf{r})\Psi(\mathbf{r})d\mathbf{r} = 1 = \langle \Psi(\mathbf{r}) | \Psi(\mathbf{r}) \rangle \quad (10:1)$$

where we have introduced the implied integration brackets of Dirac (1958). Dirac (1958) incidentally provides an excellent, though lengthy, background for understanding quantum scattering.

Similarly we postulate matrix operators which transform one state vector into another in the form

$$A|\Psi\rangle = |\Phi\rangle \quad (10:2)$$

These operators are linear and Hermitian and correspond in the formalism to observable quantities such as energy, momentum, dipole moment and the like. The term, $|\Psi_i\rangle$ is an eigenvector of the operator A if

$$A|\Psi_i\rangle = a_i|\Psi_i\rangle \quad (10:3)$$

where a_i is a real constant called the eigenvalue. Any $|\Psi\rangle$ can be expanded in terms of the eigenvectors $|\Psi_n\rangle$ of some operator, as

$$|\Psi\rangle = \sum_{n=1}^N c_n |\Psi_n\rangle \quad (10:4)$$

where N may be finite or infinite, depending on the case, and c_n are expansion coefficients. If more than one c_n is nonzero, the arbitrary state vector $|\Psi\rangle$ is said to be a mixture of eigenstates $|\Psi_i\rangle$.

The expected physical value resulting from a measurement on a system in state $|\Psi\rangle$ is given by

$$\langle A \rangle = \langle \Psi | A | \Psi \rangle \quad (10:5)$$

where $\langle A \rangle$ is the expected result of the measurement and A is the appropriate quantum mechanical operator for the physical observable measured. If $|\Psi\rangle$ is an eigenvector $|\Psi_i\rangle$, then $\langle A \rangle \equiv a_i$, the eigenvalue. Otherwise, A will have some dispersion (i.e., distributed probability density) depending on the mixture of states $|\Psi_i\rangle$ into the state $|\Psi\rangle$ so that

$$\langle A \rangle = \sum_{n=1}^N |c_n|^2 a_n \quad (10:6)$$

Mandl (1957) discusses observables in detail.

10.1.1 Energy Levels of Atomic and Molecular Systems

In 9.1 we saw that atoms and molecules have a high probability of existing in stationary states (represented by eigenvectors) with clearly defined energy levels (eigenvalues of the energy operator in our notation). A useful diagram for understanding transitions between stationary states, depicted graphically by the associated energy levels, is shown in (F10.1). Here the initial state Ψ_i is changed to final state Ψ_f by interaction with incident photon of frequency ν_0 involving intermediate state Ψ_k of the atom or molecule.

The quantization of energy levels is a natural result of the boundary value problem implicit in (10:3) where we require real eigenvalues a_i . This sort of requirement is reasonable for stationary states; i.e., states for which the expectation value $\langle A \rangle$ is a constant in the absence of external perturbation.

Actually, the postulated wavefunction of the system (10:4) is more completely written (using normalized expansion coefficients c_n)

$$\Psi(\mathbf{r}, t) = \sum_n c_n \psi_n(\mathbf{r}) \exp(-2\pi i E_n t / h) \quad (10:7)$$

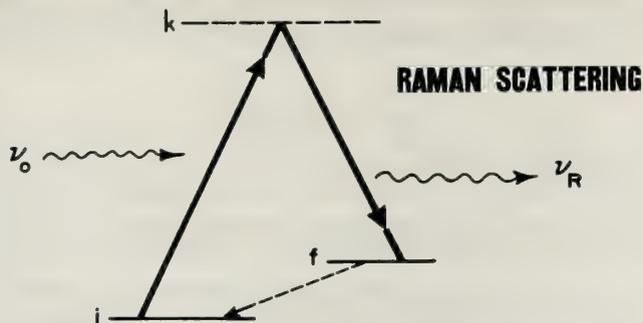


Figure 10.1 Representative energy level diagram for discussing the Raman effect.

where the ψ_n are a complete, time-independent eigenvector basis set and E_n is the corresponding eigenvalue of the total energy operator H , called the Hamiltonian. By a basis set of vectors we mean a set suitably chosen so that any arbitrary vector may be represented in terms of this set in the manner of (10:4). E_n represents one of the allowable, quantized energy levels (eigenvalues) of the atomic or molecular system. If the system is in some state i , then $c_i = 1$ and all other $c_n = 0$ in (10:7). For stationary states, only the ψ_n need be considered. Note that the eigenvector basis vectors are by definition orthonormal so that corresponding to (10:1) one has

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} \quad (10:8)$$

where δ_{ij} is the Dirac delta function.

10.1.2 Transition Probabilities

Traditionally, transition probabilities are discussed in terms of time-dependent perturbation theory where the incident photon perturbs the stationary state of the atom or molecule and induces a transition. For our consideration of scattering phenomena, a second order process, the more elegant method of Dirac (1958) displays the significant results in a better fashion than is done by time-dependent perturbation theory. In the case of Dirac's theory, the atom or molecule and the incident photons are treated as a quantum mechanical whole to which the techniques mentioned in 10.0 are applied. In the following derivation sketch, note how naturally the different types of scattering are defined as various terms in the scattering result. Ditchburn (1963) gives a useful distillation of Dirac's method.

We will represent the interaction mechanism which gives rise to the desired transition by some operator, later called H_c . H_c is, in general, time dependent. If H_c is to cause a transition, the state vector Ψ_i for the whole system cannot be an eigenvector of H_c ; for if it were, (10:3) shows that the system is unchanged (Ψ_i has not changed). In this case, no transition would have occurred. Since a transition is a time-dependent process, we retain the time dependency in (10:7), signified by the uppercase Ψ_i . All c_n but one are zero in (10:7) because the system begins and ends in a stationary eigenstate. The system eigenvector Ψ_i , not an eigenvector of H_c , will be changed by the action of H_c to some other vector Ψ' which is not necessarily an eigenvector of the system, but which can be expressed by (10:7) where the c_n are expansion coefficients in terms of the stationary states. This action results in

$$H_c |\Psi_i\rangle = h |\Psi'\rangle \quad (10:9)$$

where h is some generally complex constant. Notation can be simplified by relaxing the normalization condition on $|\Psi'\rangle$ and including h in the c_n so that

$$H_c |\Psi_i\rangle = |\Psi'\rangle \quad (10:10)$$

A fundamental postulate of quantum mechanics is that the probability of an observation (in this case a transition from state i to f) is the square of the appropriate coefficient in the state vector expansion. We write, following (10:6),

$$P(i \rightarrow f) = c_f c_f^* = |c_f|^2 \quad (10:11)$$

for the probability of finding the system in state f , or

$$P(i \rightarrow f) = |\langle \Psi_f | \Psi' \rangle|^2 = |\langle \Psi_f | H_C | \Psi_i \rangle|^2 . \quad (10:12)$$

Operator H_C changes $|\Psi_i\rangle$. The inner product picks out the coefficient c_f from the expansion of the changed vector. This postulate is really another expression of (10:6) which can be written in the form

$$\langle A \rangle = \sum_{n=1} c_n c_n^* a_n \quad (10:13)$$

where a_n is the n^{th} eigenvalue of the operator A and we have weighted the eigenvalues by the probability the system is in the n^{th} state. The probability itself is as given in (10:12).

The analysis leading to (10:12) assumes that H_C is time-independent, as for example in the case of a steady magnetic field. In the case of an incident photon, the perturbation is time dependent as photons arrive and depart. Following Dirac (1958), section 44, we separate the time-dependent operator H_C' into an operator T operating only on the time dependency of the interaction and an operator H_C (which may also be itself a function of time) operating on the other n dimensions of the space of the state vector. This separation is an alternative expression of the time-dependent Schrödinger wave equation of 9.1 (see also 10:18). That is,

$$\frac{dT}{dt} = (i\hbar)^{-1} H(t) T . \quad (10:14)$$

For the time-dependent probability,

$$P(i \rightarrow f) = |\langle \Psi_f | H_C' | \Psi_i \rangle|^2 = -\hbar^{-2} |\langle \Psi_f | H_C | \Psi_i \rangle|^2 . \quad (10:15)$$

Photon motion is removed from the transition probability in this way and only H_C for the continuous photon field need be considered. Equation (10:15) is an especially useful form for systems of interest for tropospheric scattering problems.

We now consider the form of the scattering perturbation operator H_C of (10:10). For the atom or molecule and the radiation field, we represent the Hamiltonian or total energy operator as

$$H_t = H_e + H_C + H_r \quad (10:16)$$

where H_e is the Hamiltonian for the electronic (atomic or molecular) system alone, H_r for the assembly of photons alone, and H_C is the Hamiltonian corresponding to the interaction energy. If the photons and bound electrons interact only slightly, we may to a good approximation write the wave function for the system as the product of electronic wavefunction Ψ_n (10:6) and some wavefunction Φ_q for the photon system. Here, Φ_q is formally similar to the state vectors for a system of simple harmonic oscillators. This product is of the form

$$\Psi_{nq} = \Psi_n \Phi_q . \quad (10:17)$$

Then H_e operating on Ψ_{nq} will return the eigenvalue E_n for the energy of the electronic system while H_r on Ψ_{nq} will indicate the energy of the radiation system. In other words, for the combined system without interaction, the eigenvalues are the sums of bound electronic and radiation energies, while the eigenvectors are the products of the subsystem state vectors. Ψ_{nq} is not an eigenvector of H_C , however.

The specific form of the interaction Hamiltonian is obtained using an analogy to classical mechanics, transformed by quantum mechanical correspondence postulates, and then tested on typical eigenvectors for

the correct result. For example, H_e takes the form

$$H_e = -i\hbar \frac{\partial}{\partial t}, \quad (10:18)$$

so that (10:18) applied to (10:7) by (10:3) gives E_n as the eigenvalue for the bound electronic energy of the system in the n^{th} state. The interaction (combination) Hamiltonian H_c is obtained by analogy to a system consisting of a single electron of charge e and mass m moving in a time-dependent electrostatic force field. This field is completely described by a vector potential \mathbf{A} chosen so that the scalar potential vanishes. The classical interaction between the field characterized by \mathbf{A} and the electron characterized by momentum \mathbf{p} is

$$E_{\text{class}} = \frac{1}{2m} \left[\left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - p^2 \right] = \frac{e}{mc} (\mathbf{p} \cdot \mathbf{A}) + \frac{e^2}{2mc^2} (\mathbf{A}^2). \quad (10:19)$$

This form produces the expected classical result.

Potential \mathbf{A} can be resolved into Fourier components as

$$\mathbf{A} = \int [\mathbf{A}_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{r} + 2\pi i\nu_{\mathbf{k}}t) + \mathbf{A}_{\mathbf{k}}^* \exp(i\mathbf{k} \cdot \mathbf{r} - 2\pi i\nu_{\mathbf{k}}t)] d^3k \quad (10:20)$$

or in the discrete k -value case, rather than the continuum case,

$$\mathbf{A} = \sum_{\mathbf{k}} \left[\mathbf{A}_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{r} + 2\pi i\nu_{\mathbf{k}}t) + \mathbf{A}_{\mathbf{k}}^* \exp(i\mathbf{k} \cdot \mathbf{r} - 2\pi i\nu_{\mathbf{k}}t) \right] \rho_{\mathbf{k}}^{-1} \quad (10:21)$$

where $\rho_{\mathbf{k}}$ is an arbitrary density of points in \mathbf{k} space and $2\pi\nu_{\mathbf{k}} = c|\mathbf{k}|$ so that \mathbf{k} is the photon wave vector associated with the complex amplitude vector $\mathbf{A}_{\mathbf{k}}$. $\mathbf{A}_{\mathbf{k}}$ has two orthogonal components, corresponding to two independent states of linear polarization, which components are mutually perpendicular to \mathbf{k} .

The discrete case introduces quantization of the electromagnetic field. Formally, each Fourier component can be identified with an equivalent simple harmonic oscillator as we discussed when introducing the photon field eigenvectors, $\Phi_{\mathbf{q}}$. The relation between the continuous and discrete representations of the photon field must be discussed in detail to properly evaluate matrix elements. We differentiate between photon eigenvectors $\Phi_{\mathbf{q}}$ and $\Phi_{\mathbf{q}}^{\text{D}}$ that refer to the continuous and discrete representations respectively. Equations (10:20) and (10:21) are a special case in \mathbf{k} -space of the relation, expressed here in \mathbf{p} -space (momentum),

$$\int_{\mathbf{v}} f(\mathbf{p}) d^3\mathbf{p} = \sum_{\mathbf{p}} f^{\text{D}}(\mathbf{p}) \rho_{\mathbf{p}}^{-1} \quad \text{or} \quad \int_{\mathbf{v}} f(\mathbf{k}) d^3\mathbf{k} = \sum_{\mathbf{k}} f^{\text{D}}(\mathbf{k}) \rho_{\mathbf{k}}^{-1} \quad (10:22)$$

where $\rho_{\mathbf{p}}$ is the number density of discrete points or oscillators in the appropriate momentum space and $\rho_{\mathbf{k}}$ is in \mathbf{k} or energy space.

To obtain the relation between $\Phi_{\mathbf{k}}$ and $\Phi_{\mathbf{k}}^{\text{D}}$ consider the normalization conditions (10:1). For a photon, momentum $|\mathbf{p}|$ equals $\hbar\nu/c$ so that vector momentum \mathbf{p} equals $\hbar\mathbf{k}$ and \mathbf{p} and \mathbf{k} spaces are equivalent except for a factor of \hbar . That is, in the discrete case,

$$\rho_{\mathbf{k}} = \rho_{\mathbf{p}} \hbar^3. \quad (10:23)$$

The normalization conditions are

$$\sum_{\mathbf{k}} \langle \Phi_{\mathbf{k}}^{\text{D}} | \Phi_{\mathbf{k}}^{\text{D}} \rangle = 1 = \int \langle \Phi_{\mathbf{k}} | \Phi_{\mathbf{k}} \rangle d^3\mathbf{k}. \quad (10:24)$$

From (10:22) one may write this as

$$\int \langle \Phi_{\mathbf{k}}^{\text{D}} | \Phi_{\mathbf{k}}^{\text{D}} \rangle \rho_{\mathbf{k}} d^3\mathbf{k} = 1 \quad (10:25)$$

so that by comparison, recalling the \mathbf{k} and \mathbf{p} equivalence for the normalization,

$$\Phi_{\mathbf{k}} = \Phi_{\mathbf{k}}^D \rho_{\mathbf{k}}^{1/2} \text{ or } \Phi_{\mathbf{p}} = \Phi_{\mathbf{p}}^D \rho_{\mathbf{p}}^{1/2} \quad (10:26)$$

The basic representation of eigenvectors as in (10:7) is in position or \mathbf{r} space rather than in energy-momentum space as discussed immediately above. Basic matrix elements will be evaluated between state vectors in \mathbf{r} space, but in order to use photon momentum operators, the state vectors must be transformed to \mathbf{p} space. This transformation takes the form (Dirac, 1968, sections 23 and 50)

$$\Phi(\mathbf{p}, t) = h^{-n/2} \Phi(\mathbf{r}, t) \quad (10:27)$$

where n is the number of degrees of freedom or dimensionality of the space. For photons, n is three so that

$$\Phi(\mathbf{r}, t) = h^{3/2} \Phi(\mathbf{p}, t) \quad (10:28)$$

It is interesting to note that the number of photons per unit of (momentum) phase space is h^{-3} times the number of photons per discrete state. This is closely related to the \mathbf{r} space to \mathbf{p} space transformation of (10:28). The density of oscillators in \mathbf{k} space $\rho_{\mathbf{k}}$ is an arbitrary parameter which must drop out of all calculated results for observable quantities.

As in the simple harmonic oscillator case, the excitation state of the radiation field can change by only one quantum for each \mathbf{k} in a given interaction. For example, in absorption a photon of energy $h\nu_{\mathbf{k}}$ disappears ($\mathbf{A}_{\mathbf{k}}$ is reduced by one quantum) and the bound electronic system changes to a state of higher energy, from E_i to E_f where $\Delta E = h\nu_{\mathbf{k}}$.

To pass to the quantum mechanical form of (10:19) note that the operator corresponding to momentum \mathbf{p} is simply multiplication by $m\dot{\mathbf{r}}$, and for \mathbf{A} substitute $(\mathbf{A}_{\mathbf{k}} + \mathbf{A}_{\mathbf{k}}^*)$ utilizing (10:21). The important interaction Hamiltonian is then

$$H_C = \frac{e}{mc} \sum_{\mathbf{k}} \mathbf{p} \cdot (\mathbf{A}_{\mathbf{k}} + \mathbf{A}_{\mathbf{k}}^*) \rho_{\mathbf{k}}^{-1} + \frac{e^2}{2mc^2} \sum_{\mathbf{k}\mathbf{k}'} (\mathbf{A}_{\mathbf{k}} + \mathbf{A}_{\mathbf{k}}^*) \cdot (\mathbf{A}_{\mathbf{k}'} + \mathbf{A}_{\mathbf{k}'}^*) \rho_{\mathbf{k}}^{-1} \rho_{\mathbf{k}'}^{-1}, \quad (10:29)$$

where the appropriate oscillatory exponentials of (10:20) and (10:21) are included implicitly. The $\mathbf{A}_{\mathbf{k}}$ of (10:29) represent a vector potential operator that is the quantum mechanical analog of the classical $\mathbf{A}_{\mathbf{k}}$ of (10:21). This notation emphasizes the analog between the classical and quantum quantities; it should cause no confusion since the quantum $\mathbf{A}_{\mathbf{k}}$ will be replaced by number operator $\eta_{\mathbf{k}\ell}$ immediately below. Before the operator is useful, we must convert the $\mathbf{A}_{\mathbf{k}}$ vector coefficients to simple number operators $\eta_{\mathbf{k}\ell}$. The second subscript ℓ refers to the two independent polarization states corresponding to each Fourier component of the radiation field. Operator $\eta_{\mathbf{k}\ell} \cdot \eta_{\mathbf{k}\ell}^*$ operating on $\Phi_{\mathbf{q}}^D$ as in (10:17) returns as an eigenvalue $n_{\mathbf{k}\ell}$ the incremental occupation number of the representative simple harmonic oscillator. The energy in the radiation subsystem may be expressed by classical analogy in terms of the volume integral of functions of $\mathbf{A}_{\mathbf{k}}$ and alternatively by a consideration of the energy of the ensemble of representative oscillators. The energy is expressed by the radiative Hamiltonian H_r which, using a volume integral, turns out to be

$$H_r = 4\pi^2 \sum_{\mathbf{k}} k^2 \mathbf{A}_{\mathbf{k}} \cdot \mathbf{A}_{\mathbf{k}}^* \rho_{\mathbf{k}}^{-1}, \quad (10:30)$$

or expressing each of the two non-zero components of the $\mathbf{A}_{\mathbf{k}}$ as for the $\eta_{\mathbf{k}\ell}$,

$$H_r = 4\pi^2 \sum_{\mathbf{k}\ell} k^2 A_{\mathbf{k}\ell} \cdot A_{\mathbf{k}\ell}^* \rho_{\mathbf{k}}^{-1}. \quad (10:31)$$

In both (10:29) and (10:31) the quantum mechanical $\mathbf{A}_{\mathbf{k}}$ includes implicitly the appropriate $\exp(2\pi i\nu_{\mathbf{k}} t)$ factor of (10:21) which is consistent with the notion of the representative oscillators as an equivalent formalism where one oscillator corresponds to each possible photon state. In terms of the oscillator operators,

$$H_r = \sum_{\mathbf{k}\ell} h\nu_{\mathbf{k}} \eta_{\mathbf{k}\ell} \cdot \eta_{\mathbf{k}\ell}^* \quad (10:32)$$

where we have neglected the zero-point oscillator energy which has no dynamical consequence. Equating (10:31) and (10:32), associating the square root of each coefficient with each of the operators,

$$A_{k\ell} = ch^{1/2}\nu_k^{-1/2}\rho_k^{1/2}\eta_{k\ell}(2\pi)^{-1} . \quad (10:33)$$

It is convenient to express the vector products of (10:29) in terms of the two non-zero, independent ℓ components of polarization. Substituting (10:33) into (10:29) we obtain the operator

$$\begin{aligned} H_c = & \frac{eh^{1/2}}{m(2\pi)^2} \sum_{k\ell} p\ell(\eta_{k\ell} + \eta_{k\ell}^*)\nu_k^{-1/2}\rho_k^{-1/2} \\ & + \frac{e^2h}{2m(2\pi)^4} \sum_{\substack{k\ell \\ k'\ell'}} (\eta_{k\ell} + \eta_{k\ell}^*)(\eta_{k'\ell'} + \eta_{k'\ell'}^*)(\ell \cdot \ell')\nu_k^{-1/2}\nu_{k'}^{-1/2}\rho_k^{-1/2}\rho_{k'}^{-1/2} \end{aligned} \quad (10:34)$$

The identity

$$\mathbf{p} = m\dot{\mathbf{x}} \quad (10:35)$$

is useful for clarity in (10:34). This operator replacement leads to matrix elements of the form $\langle \Psi_f | \dot{\mathbf{x}} | \Psi_i \rangle$ which can be directly differentiated by making use of the time dependence in (10:7). After differentiation, the time dependence of course remains in the Ψ_k so that we can write

$$\langle \Psi_f | \dot{\mathbf{x}} | \Psi_i \rangle = -2\pi i \nu_{fi} \langle \Psi_f | \mathbf{x} | \Psi_i \rangle \quad (10:36)$$

where ν_{fi} is $(\nu_f - \nu_i)$, the frequency (energy) difference between the *stationary states* of the system. Note that this derivation applies only to basic eigenvectors of the system. Equation (10:36) exhibits the time derivative of the displacement operator on the system. The limitation to stationary states (eigenvectors) is completely reasonable for tropospheric scattering since the bound electronic systems start and end in stationary states.

We are now prepared to discuss various types of scattering.

10.2 Two Step (Single Photon) Scattering

The first term in (10:34) can represent a single photon process where the oscillator ensemble gains or loses a single quanta (a photon appears or disappears) and the bound electron system (atom or molecule) changes from one stationary state Ψ_i to another eigenstate Ψ_g . The second term in (10:34) can only involve two photon processes. It will be treated in 10.3. If a single photon process is to be interpreted as "scattering," one must observe a series of *two* such single photon processes. First, a photon ν_o disappears and the electronic system undergoes a transition from state Ψ_i to Ψ_g . Subsequently, after a time determined by the transition probability for the system in state Ψ_g , a photon ν_s appears (is "scattered" out) and the bound electronic system changes from state Ψ_g to Ψ_f .

Some authors suggest that this process is not true scattering. Certainly it is not a two-photon process but rather two single-photon processes. Experimentally, one treats both one and two photon processes identically. Theoretically, both processes naturally follow from the same interaction Hamiltonian. For tropospheric purposes, we will discuss two-step scattering as a scattering process.

At sea level, a typical atmospheric molecule experiences a mean time between collisions in the order of 10^{-9} seconds. Lifetimes of molecules in excited states Ψ_g vary widely but 10^{-6} seconds is a representative figure. It is highly likely that a molecule in the troposphere excited in the first step of the two-step scattering process will be de-excited by collision with another molecule in a nonradiative fashion. Then the second step of the scattering never occurs and we say the scattering is *quenched*. The probability of a collision causing a nonradiative quenching is generally less than unity and can be quite small. The actual probability of a nonradiative decay depends on the excited molecule, the excited state Ψ_g and the colliding molecule. For example, a polar molecule such as HCl would likely be a more effective quenching agent than Ar, and a long lived excited state is more likely to be quenched than one which undergoes rapid spontaneous decay.

Pressure quenching is highly important to consideration of two-step scattering in the troposphere. Much experimental work remains to be done to measure quenching probabilities for natural and pollutant atmospheric molecules. Until these probabilities are known, we cannot effectively evaluate the potential of two-step scattering as a method for remote sensing of the troposphere.

10.2.1 Calculation of Single Photon Scattering

To discuss two-step scattering quantitatively, we specialize (10:15) with (10:17), (10:35) and (10:36). The $\eta_{k\ell}$ and p_{ℓ} are commuting operators which act independently on the appropriate radiation and electronic (molecular) subsystems respectively. The $\eta_{k\ell}$ destroy or create a single photon which we represent by Φ_k^D for a photon acted on which has energy $h\nu_k$. Note that the eigenvector time dependence of (10:7) is still attached to Φ_q^D and Ψ_n or to Ψ_{nq} in the double-subscript notation of (10:17).

We require for (10:15) matrix elements of the form

$$\langle \Psi_g | H_C | \Psi_i \Phi_0 \rangle \text{ and } \langle \Psi_f \Phi_s | H_C | \Psi_g \rangle \quad (10:37)$$

for absorption and emission processes respectively. It is convenient to first separate out the time dependence of the wave functions. The time variation of the initial state Ψ_{i0} depends on the total system energy so that (10:17) with (10:7) gives

$$\Psi_{i0} = \Psi_i \Phi_0 = \psi_i \exp(-2\pi i E_i t / h) \phi_0 \exp(-2\pi i \nu_0 t) = \psi_i \phi_0 \exp[-2\pi i (\nu_0 + \nu_i) t] , \quad (10:38)$$

where the lower-case ψ_g represent time-independent eigenvectors. Including the time dependence of Ψ_g as a product, the total time function for the absorption interaction is written

$$F_1(t) = \exp[-2\pi i (\nu_0 - \nu_{gi}) t] \quad (10:39)$$

with the double subscript notation of (10:36). The matrix element becomes

$$\langle \Psi_g | H_C | \Psi_i \Phi_0 \rangle = \langle \psi_g | H_C | \psi_i \phi_0 \rangle F_1(t) \quad (10:40)$$

for absorption.

For emission, a similar argument gives the time function

$$F_2(t) = \exp[2\pi i (\nu_s - \nu_{gf}) t] , \quad (10:41)$$

and the matrix element

$$\langle \Psi_f \Phi_s | H_C | \Psi_g \rangle = \langle \psi_f \phi_s | H_C | \psi_g \rangle F_2(t) . \quad (10:42)$$

If ν_0 or ν_s is close to a real energy level difference in the molecule, the first term of H_C predominates. The matrix element of the first term in H_C may be evaluated in the form $\langle \psi_g | H_C | \psi_i \phi_0^D \rangle$ or $\langle \psi_f \phi_s | H_C | \psi_g \rangle$ for absorption and emission respectively. The $\phi^D(\mathbf{r})$ photon wavefunctions are written in \mathbf{r} (coordinate) space to be consistent with the $\psi_i(\mathbf{r})$ in the form (10:7)

From this point, the development for absorption and emission differs. Consider first absorption. Fundamentally, the photon momentum operators are much more convenient than total energy operators for relativistic systems. Therefore (10:28) is used to replace the \mathbf{r} space wavefunctions with the \mathbf{p} space form; this introduces a factor $h^{3/2}$. The $\phi^D(\mathbf{r})$ are also in terms of total energy W (the energy of the representative oscillator of 10:21). Hence the oscillator density $\rho_p^{-1/2}$ associated with the photon wavevector must be corrected by a weight factor $(dW/d\mathbf{p})^{1/2} = (1/c)^{1/2}$ for photons when $\phi^D(\mathbf{p})$ is introduced into the matrix element.

Note that p is $h\nu/c$ or W/c where c is the speed of light. The discrete oscillator density in \mathbf{k} space is changed to \mathbf{p} space by (10:23); replacing $\rho_{\mathbf{k}}^{-1/2}$ by $\rho_{\mathbf{p}}^{-1/2}$ introduces a factor $h^{-3/2}$. The discrete wavefunction $\phi_{\mathbf{0}}^D(\mathbf{p})$ is replaced by continuous $\phi_{\mathbf{0}}(\mathbf{p})$ $\rho_{\mathbf{p}}^{-1/2}$ by means of (10:26). Integration over all photon variables proceeds as in (10:22). The matrix element becomes, by (10:35) with the modifications above,

$$\langle \psi_g | H_c | \psi_i \phi_{\mathbf{0}}^D(\mathbf{r}) \rangle = \frac{eh^{1/2}}{m(2\pi\nu_0 c)^{1/2}} \langle \psi_g | \mathbf{p} | \psi_i \rangle. \quad (10:43)$$

In the electronic space, (10:36) allows replacement of the electronic momentum operator by the dipole operator $e\mathbf{x}$ and a multiplicative factor of $(-2\pi i\nu_{gi})$. This yields

$$\langle \psi_g | H_c | \psi_i \phi_{\mathbf{0}}^D(\mathbf{r}) \rangle = \frac{(-2\pi i\nu_{gi})h^{1/2}}{(2\pi\nu_0 c)^{1/2}} \langle \psi_g | e\mathbf{x}_0 | \psi_i \rangle \quad (10:44)$$

for absorption. Vector \mathbf{x} is replaced by the component x_0 in the direction of polarization of the incident light. See for example the discussion preceding (10:35) for a discussion of the two independent states of polarization of the radiation field.

The transition probability is found from (10:15) with (10:40) and the preceding matrix elements (10:44) to be

$$P_a(i-g) = \frac{(2\pi)^3 \nu_{gi}^2}{h\nu_0 c} \left| \langle \psi_g | e\mathbf{x}_0 | \psi_i \rangle F_1(t) \right|^2. \quad (10:45)$$

This transition probability is still time dependent.

Consider the first step of the scattering process, which consists of the absorption of a photon of frequency ν_0 and a change in the molecular state from i to g . To evaluate the total probability of a transition in some small time interval $t = 0$ to $t = \tau$, say $P_{\tau}(i-g)$, we integrate (10:39) with respect to time giving

$$\int_0^{\tau} F_1(t) dt = \frac{\exp[-2\pi i(\nu_0 - \nu_{gi})\tau] - 1}{-2\pi i(\nu_0 - \nu_{gi})}. \quad (10:46)$$

Here $P_{\tau}(i-g)$ is on the basis of a single photon-molecule encounter in real space. To convert to measurable system variables (traditional absorption coefficient) consider an incident unidirectional photon beam with some distributed energy and a flux of one photon per unit area per unit time per unit energy ($h\nu$) range. The energy density in frequency units $\rho(\nu)$ is assumed constant in the region $\nu \simeq \nu_0$ but is otherwise arbitrary. Since the resonance denominator in (10:46) preferentially selects those $\nu_0 \simeq \nu_{gi}$, we set $\nu_0 = \nu_{gi}$ in (10:45) (at the center of the applicable ν_0 distribution) and integrate (10:46) with respect to ν_0 . Substituting $x = 2\pi(\nu_0 - \nu_{ki})\tau$, the integral becomes

$$\int_{-\infty}^{\infty} \left| \int_0^{\tau} F_1(t, \nu) dt \right|^2 d\nu = (\tau/\pi) \int_{-\infty}^{\infty} [(1 - \cos x)/x^2] dx = \tau. \quad (10:47)$$

For the transition probability per unit time we divide by τ . Performing these modifications on (10:45) we obtain

$$P_a(i-g) = \frac{8\pi^3 \nu_{gi}}{hc} \left| \langle \psi_k | e\mathbf{x}_0 | \psi_i \rangle \right|^2 \quad (10:48)$$

for the absorption coefficient per molecule for a photon beam of one photon per unit area per unit time per unit frequency range.

The emission of a photon from the excited molecule is the second step of the two step (single photon interaction) scattering process. Since the emission is dependent on the state of the molecule and not on the specific excitation process, it is obvious that on the average the emission process is isotropic.

A change of coordinate system to spherical coordinates is therefore appropriate. We are interested in the emission of a photon with momentum $|\mathbf{p}| = h\nu/c$ in any direction. The state vector ϕ_s^{Di} of interest in (10:42) is written in spherical polar coordinates, consistent with the nature of the scattering process. The superscript i for isotropic denotes the spherically symmetric state vector, ϕ_s^{Di} is in terms of frequency (total energy $W = h\nu_s$) and the polar angles θ and ϕ . This final state is one of a continuous range of possible states. Since spatial coordinates \mathbf{r} do not appear in the emission interaction, the transformation (10:28) need not be

applied. The omission of this transformation is in contrast to the absorption case. As in the case of absorption, the discrete oscillator density associated with $\phi_s^{\text{Di}}(\nu, \theta, \phi)$ requires a density weight factor $(d\nu/d\mathbf{p})^{1/2} = (h/c)^{1/2}$ when conversion from (ν, θ, ϕ) to $(\mathbf{p}, \theta, \phi)$ space is made. The discrete oscillator state vector ϕ_s^{Di} is replaced by the continuum expression $\phi_s^{\text{I}} \rho_p^{-1/2}$ as in (10:26). This substitution introduces the factor $\rho_p^{-1/2}$ into the simplification of (10:42). The density expression is still in spherical coordinates. The unit volume in continuous phase space is expressed in Cartesian and spherical coordinates as $dp_x dp_y dp_z$ and $p^2 \sin \theta dp d\theta d\phi$ respectively. A conversion from $(\mathbf{p}, \theta, \phi)$ space to (p_x, p_y, p_z) space is required to evaluate matrix elements in the conventional sense. The element in Cartesian coordinates must then be multiplied by a weight factor $(p^2 \sin \theta)^{1/2}$ to equal the matrix element in the spherical coordinates used for the emission probability. That is,

$$\langle \psi_f \phi_s^{\text{I}} | H_c | \psi_g \rangle = (p^2 \sin \theta)^{1/2} \langle \psi_f \phi_s | H_c | \psi_g \rangle . \quad (10:49)$$

Using (10:23) the density $\rho_k^{-1/2}$ in (10:34) is replaced by $\hbar^{-3/2} \rho_p^{-1/2}$ as before. The transformation factors are included with integration over all photon variables and replacement of the electronic momentum operator by the dipole operator (10:36) to yield the matrix element

$$\langle \psi_f \phi_s^{\text{I}} | H_c | \psi_g \rangle = \frac{(+2\pi i \nu_{gf})(p^2 \sin \theta)^{1/2}}{(2\pi \nu_s c)^{1/2} h^{1/2}} \langle \psi_f | e x_s | \psi_g \rangle . \quad (10:50)$$

Component x_s in the direction of polarization of the emitted light reflects the fact that the dipole matrix element must be taken in the direction of radiation polarization.

The transition probability from (10:15) with (10:42) and (10:50) is

$$P_e(g-f) = \frac{(2\pi)^3 p^2 \sin \theta \nu_{gf}^2}{h^3 \nu_s c} |\langle \psi_f | e x_s | \psi_g \rangle F_2(t)|^2 . \quad (10:51)$$

In a similar manner to the absorption case, we integrate $F_2(t)$ with respect to time from 0 to τ and integrate the square of the resultant with respect to ν_s for all ν_s since we are interested in the emission of any photon. This latter integration picks $\nu_s \simeq \nu_{gf}$ so ν_s may be set equal to ν_{gf} in (10:51). Wavefunction ϕ_s^{I} is for a particular θ and ϕ direction of \mathbf{p} . To convert to a probability per unit solid angle we integrate over all θ and ϕ and divide by 4π steradians. This averaging has the effect of eliminating the $\sin \theta$ factor. For p we substitute $h\nu_{gf}/c$. The transition probability per unit time requires division by τ . These modifications on (10:51) result in

$$P_e(g-f) = \frac{8\pi^3 \nu_{gf}^3}{hc^3} |\langle \psi_f | e x_s | \psi_g \rangle|^2 \quad (10:52)$$

for the probability of spontaneous emission. Stimulated emission is not generally applicable to remote sensing since the incident fluxes are small.

Collisions of the excited molecule with other atmospheric molecules will tend to non-radiatively de-excite the single photon (two-step) scattering molecule. The efficiency of these collisional de-excitation mechanisms is extremely difficult to calculate, even in principle, for most molecules. We represent the inverse of this pressure quenching probability as Q where $Q = 1$ for the case of the isolated molecule (no quenching) and $Q \rightarrow 0$ where the likelihood of a collisional de-excitation is extremely high in a time period characteristic of the radiative lifetime of the molecule [lifetime $\tau = 1/P(g-f)$]. Probability Q will depend on the number densities of relevant gases n_i (quenching efficiency depends on the character of the colliding molecule), the temperature T , the pressure p and the intermediate state g or $Q(n_i, T, p, g)$. This value Q is determined empirically by controlled laboratory experiments.

10.2.2 Applications

There are various ways of expressing the scattering efficiency of a molecule (or general scattering center). The preceding discussion (10.2.1) evaluated absorption and emission in traditional terms of transition probabilities. The absorption coefficient is a probability of absorption per unit photon flux (number incident per unit area per unit time) and the emission coefficient is a probability per unit time of a photon being emitted from an excited system. A useful form for scattering efficiency is the differential scattering cross section $d\sigma/d\Omega$. This is defined most simply as

$$\frac{d\sigma(\Omega)}{d\Omega} = \frac{(\text{number of photons scattered into solid angle } d\Omega \text{ per unit time per molecule})}{(\text{number of photons crossing unit area per unit time})} \quad (10:53)$$

or similarly in terms of energy, where for number of photons substitute energy. Here, $d\sigma(\Omega)/d\Omega$ has the dimensions of an area and is in general dependent on the scattering angle Ω . For the particular case of isotropic scattering considered here, $d\sigma/d\Omega$ is independent of Ω and the value of differential cross section is examined on a per molecule basis.

The transition probabilities (10:48) and (10:52) are not directly applicable to a computation of the differential scattering cross section. They are in fact initial probabilities per unit time, valid for times short with respect to the lifetime of the initial state, or equivalently, with respect to the mean time for a transition to occur. Since time does not appear explicitly, it is obvious that the probability increases linearly with time. This increase is not consistent with physical reality for times longer than the lifetime of the original state. Essentially, the model leading to probabilities (10:48) and (10:52) ignores the depletion of the initial state by the transition itself. A simple thought experiment demonstrates the reality of steady state or time integrated transition probabilities. Consider a collimated beam incident on a cell of known length and known scatterer concentration. The attenuation of the beam gives a direct measure of the steady state transition probability for absorption. Similarly, the scattered intensity into a detector subtending solid angle $d\Omega$ is a direct measure of the steady state transition probability for absorption followed by emission at the chosen frequency (including quenching effects). The product of the two time independent probabilities for absorption and emission yields the differential scattering cross section.

To obtain the time integrated (thus time independent) transition probabilities needed to evaluate the differential cross section, we introduce the notion of natural line breadth or radiation damping effects. We assume the probabilities (10:48) and (10:52) are valid for a system with probability $p(t=0) = 1$ of being in the initial state. We further assume that the probability of finding the system in an initial excited or ground state decreases with time after the perturbation begins as $p(t) = \exp(-\gamma t/2)$. Heitler (1954) in his section 18 demonstrates that for a simple two level system, γ for emission is equal to the total (initial) spontaneous transition probability per unit time for emission. This is reasonable since the higher the transition probability, the faster the probability of finding the system in the initial excited or ground state decays with time. In the sense that γ is defined to be the reciprocal of the mean lifetime of an excited or pumped state, it must just be equal to the transition probability for emission or absorption respectively.

The uncertainty relation for energy and time,

$$\Delta E \Delta t = \hbar, \quad (10:54)$$

indicates that $\Delta E = \hbar \gamma$ since the mean lifetime of the excited two level system, Δt , is just $1/\gamma$. This allows generalization to a multilevel system and assignment of a natural width γ_j to each level j . Clearly the width of any level j is the sum of initial emission probabilities from j to all lower levels as

$$\gamma_j = \sum_{h < j} P_e(j - h) \quad (10:55)$$

in the sense of (10:52). Similarly the width of an absorption or emission line is denoted by a double subscript as

$$\gamma_{jk} = \gamma_j + \gamma_k. \quad (10:56)$$

The ground state width is zero by this analysis. In particular, we consider a three level system for the two step scattering as discussed before. The γ associated with each level is found using (10:52) as

$$\gamma_i = 0$$

$$\gamma_g = P(g - f) + P(g - i) = \frac{8\pi^3}{hc^3} [\nu_{gf}^3 |\langle \psi_f | e x_s | \psi_g \rangle|^2 + \nu_{gi}^3 |\langle \psi_i | e x_s | \psi_g \rangle|^2]$$

$$\gamma_f = P(f - i) = \frac{8\pi^3}{hc^3} \nu_{fi}^3 |\langle \psi_i | e x_s | \psi_f \rangle|^2. \quad (10:57)$$

For absorption, $F_1(t)$ of (10:39) is modified by a factor $\exp(-\gamma_{ig}t/2)$ where γ_{ig} comes from the absorption probability P_a of (10:48). Integrating over t from 0 to ∞ we obtain in analogy to (10:46)

$$\int_0^{\infty} F_1'(t)dt = 1/[2\pi i(\nu_0 - \nu_{gi}) + \gamma_{ig}/2] \quad (10:58)$$

Squaring, and retaining the real part, yields a factor

$$1/[4\pi^2(\nu_0 - \nu_{gi})^2 + \gamma_{ig}^2/4] \quad (10:59)$$

In atmospheric sensing applications, ν_0 will generally be a monochromatic (laser) source with some width at half maximum w and normalized intensity density such that $I_0 w$ represents some incident power. We integrate the preceding expression over ν_0 from $\nu_0 - w/2$ to $\nu_0 + w/2$ and obtain the product

$$R(\nu_0 - \nu_{gi})/\pi\gamma_{ig} \quad (10:60)$$

where the resonance factor is given by

$$R(\nu_0 - \nu_{gi}) = \tan^{-1} 2\pi(\nu_0 - \nu_{gi} + w/2)/\gamma_{ig} - \tan^{-1} 2\pi(\nu_0 - \nu_{gi} - w/2)/\gamma_{ig} \quad (10:61)$$

This factor is just equal to $2\pi w/\gamma_{ig}$ for $\nu_0 \simeq \nu_{gi}$ and decreases monotonically as $|\nu_0 - \nu_{gi}|$ increases (note that $R(\nu_0 - \nu_{gi})$ is dimensionless). The existence of the resonance factor demonstrates that the laser source must be properly chosen to excite two step scattering. The time independent absorption probability is then

$$\overline{P_a(i-g)} = \frac{8\pi^2\nu_{gi}}{hc\gamma_{ig}} R(\nu_0 - \nu_{gi}) |\langle \psi_g | e_{x_0} | \psi_i \rangle|^2 \quad (10:62)$$

For emission, $F_2(t)$ of (10:41) is similarly modified by $\exp(-\gamma_{gf}t/2)$. We integrate over t from 0 to ∞ and integrate the square of the resultant with respect to ν_s as before. This shows that we are interested in the probability of emission of a photon of any frequency associated with the g to f transition. Of course there is a very high probability that ν_s is very near ν_{gf} . This integration of $F_2'(t)$ results in a factor $1/2\gamma_{gf}$. The result, analogous to (10:52) is

$$\overline{P_e(g-f)} = \frac{(2\pi)^3 \nu_{gf}^3}{2hc^3 \gamma_{gf}} |\langle \psi_f | e_{x_s} | \psi_g \rangle|^2 \quad (10:63)$$

which is independent of time. Linewidth γ_{gf} is obtained from (10:57). This natural linewidth is to be differentiated from the total experimental linewidth which will be dominated by pressure broadening at tropospheric pressures. The pressure broadened linewidth is typically the order of 3×10^9 Hz (0.1 cm^{-1}) at atmospheric pressure while natural γ_{gf} is typically an order of magnitude or more smaller.

Since at atmospheric pressures the molecule will in general be disturbed before both absorption and re-emission can occur, we are justified in assuming that the absorption and emission are independent. Given independence, the cross section for single photon (two step) scattering is given by the product of the two time independent probabilities $P_a \cdot P_e$. Heitler (1954) cautions that the assumption of independence is not valid if the radiation system is undisturbed. In the case of the isolated system, which case is not applicable in the troposphere, resonance fluorescence is correctly treated as a single coherent quantum process, not as an independent absorption and emission process.

The probability product, including the quenching probability Q of 10.2.1 is

$$\frac{d\sigma}{d\Omega} [i-f(g)] = \frac{(2\pi)^5 \nu_{gi} \nu_{gf}^3 R(\nu_0 - \nu_{gi})}{h^2 c^4 \gamma_{ig} \gamma_{gf}} |\langle \psi_f | e_{x_s} | \psi_g \rangle|^2 Q(\eta_i T_{pg}) |\langle \psi_g | e_{x_0} | \psi_i \rangle|^2 \quad (10:64)$$

for the scattering cross section in area/steradian with an incident photon beam of one photon per unit area per unit time of frequency ν_0 .

If state $i \equiv$ state f , then the process is called resonance re-radiation and the incident frequency ν_0 and the scattered ν_s will be the same within the uncertainty (linewidth) of the state g . If state $i \neq$ state f , then the process is called fluorescence scattering. If $\nu_{gf} < \nu_{gi}$, the fluorescence is said to be Stokes radiation; otherwise it is called anti-Stokes fluorescence.

Except for the frequency matching condition on ν_0 , the scattering cross section is dominated by the dipole matrix elements. These matrix elements are identical to those discussed in Chapter 9 for absorption, but they have been discussed here in a fashion suitable for analysis of two photon (single step) scattering as well. Section 10.2.1 serves as an introduction to 10.3 on two photon scattering.

10.3 Single Step (Two Photon) Scattering

Both terms in (10:35) are used to properly describe a two photon process. If ν_0 is chosen so that absorption or spontaneous emission is very small in (10:64), two photon processes can be important in describing the interaction between photons and bound electronic systems. Processes involving the simultaneous absorption of two photons or simultaneous emission of two photons are also described by this interaction term. These processes might be used for two-step scattering sensing as discussed in 10.2, but the resultant scattering is so weak as to be practically useless. It has never been observed in gases.

A much more useful interaction mechanism, also described by the two photon formalism, is the simultaneous disappearance of one photon and the appearance of a different photon. The bound electronic system may also undergo a (practically) instantaneous transition during the interaction. 'Simultaneous' in the experimental sense means an interaction time less than 10^{-14} seconds, or the measurement error, as mentioned by Garlick (1958). In a theoretical sense, it means that there is no necessary time order for the destruction and creation of photons in the interaction period as there was for the two-step process. All experimental evidence so far indicates that collision or 'pressure' quenching, discussed in 10.2, has no meaning for two photon scattering.

Absence of quenching is significant from the point of view of tropospheric sensing since it means that two photon processes will compete with two-step scattering more effectively at atmospheric pressures than at lower pressures. This is seen in practice. Two-step scattering of Na in the stratosphere is discussed in 24.2.3, but at ambient pressures in the troposphere two photon quantum scattering predominates.

10.3.1 Calculation of Two Photon Scattering

Transition probabilities for two photon quantum scattering are usually expressed as scattering cross sections or effective area that a photon must hit (per molecule) per unit solid angle of scattering. To evaluate the cross sections, we apply the developments of 10.1.2 to initial states given by $|\Psi_i\Phi_0\rangle$ and final states by $\langle\Psi_f\Phi_s|$. Intermediate states Ψ_g representing the bound electronic subsystem in an excited state, the appropriate photon being destroyed, are also necessary for the analysis. Although superficially similar to the excited states of 10.2, the Ψ_g for two photon scattering are not required to satisfy conservation of energy in the manner of the energy integral used in 10.2 as (10:47).

From (10:15) the photon-molecule encounter results in a transition probability

$$P(i - f) = \hbar^{-2} |\langle\Psi_f\Phi_s|H_C|\Psi_i\Phi_0\rangle|^2 \quad (10:65)$$

where H_C is given in (10:34). Since no operator on the bound electronic subsystem exists in the second term $H_C^{(2)}$, this contribution to the probability must contain a factor δ_{if} showing that the electronic system is unchanged in so called direct scattering, which is independent of any intermediate state g . Two terms in the summation contribute, i.e., when $k=0, k'=s$ and when $k=s, k'=0$, consistent with the notion that the absorption-emission and emission-absorption processes are valid and indistinguishable for two-photon scattering. This direct matrix element is

$$2\langle\Psi_f\Phi_s|H_C^{(2)}|\Psi_i\Phi_0\rangle \quad (10:66)$$

where $i\equiv f$ and therefore $s\equiv 0$. The factor of 2 relates to the fact that both the $i\rightarrow f$ and $f\rightarrow i$ orders of the interaction are valid and indistinguishable.

It is obvious that the first term in the interaction Hamiltonian (10:34), $H_C^{(1)}$, must be applied twice to go from the initial to final states since only one photon is changed each time this term is applied. The required intermediate basis states for this implicit series expansion are naturally chosen to be the eigenvectors of the electronic system, Ψ_g . Application of the operator more than once by means of intermediate eigenstates is implicit in (10:34) and is a natural extension of the operation described in (10:10) which is not to be thought

of necessarily as a single, simple mathematical operation. The multiple application of $H_C^{(1)}$ is made mathematically plausible by Dirac (1958) in terms of a matrix multiplication expansion, on the basis of the Ψ_g , of a typical matrix element.

The transition through the intermediate state can proceed in two ways. If the incident photon is destroyed before the scattered one is created, one has the elements

$$\sum_g \langle \Psi_f \Phi_s | H_C^{(1)} | \Psi_g \rangle \langle \Psi_g | H_C^{(1)} | \Psi_i \Phi_o \rangle . \quad (10:67)$$

The opposite mechanism of creation first has two photons existing in the intermediate state rather than no photons as in the first case. We denote this by a double subscript Φ_{os} and write the element as

$$\sum_g \langle \Psi_f \Phi_s | H_C^{(1)} | \Psi_g \Phi_{os} \rangle \langle \Psi_g \Phi_{os} | H_C^{(1)} | \Psi_i \Phi_o \rangle . \quad (10:68)$$

Evaluation of these three sets of matrix elements process as in 10.2.1. Equation (10:7) allows evaluation of the explicit time dependence of the product state vectors. Let $F_1(t)$ be the time factor related to the direct matrix element, $F_2(t)$ and $F_3(t)$ to the two matrix element factors of destruction—creation, and $F_4(t)$ and $F_5(t)$ to the two matrix element factors of creation—destruction. Each factor is considered separately, recalling that we postulate two different (instantaneous but formally sequential) operations for the scattering. In direct analogy to (10:39) and (10:41) we write

$$F_1(t) = \exp[-2\pi i(\nu_{os} + \nu_{if})t] = 1 \quad (10:69)$$

since $i \equiv f$ and $\nu_o \equiv \nu_s$ by the constraints on the system. For the other matrix elements, the factors are

$$\begin{aligned} F_2(t) &= \exp[-2\pi i(\nu_o - \nu_{gi})t] \\ F_3(t) &= \exp[2\pi i(\nu_s - \nu_{gf})t] \\ F_4(t) &= \exp[2\pi i(\nu_s + \nu_{gi})t] \\ F_5(t) &= \exp[-2\pi i(\nu_o + \nu_{gf})t] . \end{aligned} \quad (10:70)$$

The $F(t)$ are integrated with respect to time from $t=0$ to $t=\tau$ in the manner of (10:46) to yield appropriate resonance denominators of the form (for example) $1/[-2\pi i(\nu_o - \nu_{gi})]$.

Because the scattered photon frequency is not directly related to the $g \rightarrow f$ energy difference, we may integrate F_3 and F_5 , from the emission part of the matrix element, over variable ν_s and ν_o respectively. In the manner of (10:47), this integration returns the arbitrary time τ which is divided out to yield a transition probability per unit time or cross section. The incident frequency ν_o is not in general even approximately equal to ν_{gi} , so that integration over ν_o for the absorption part of the interaction, which involves the ν_{gi} for selection of a destroyed photon ν_o , has no meaning. Instead, the exponentials in the time integrated forms of F_2 and F_4 are averaged to one, leaving resonance denominators

$$\overline{\int_0^\tau F_2(t) dt} = 1/2\pi(\nu_o - \nu_{gi}) \quad (10:71)$$

$$\overline{\int_0^\tau F_4(t) dt} = -1/2\pi(\nu_s + \nu_{gi}) .$$

The matrix elements of (10:66) to (10:68) are expressed in time independent form with the use of (10:71). Thus (10:65) yields the probability for two photon scattering as

$$\begin{aligned} P(i - f) &= \hbar^{-2} \left| 2 \langle \Psi_f \Phi_s | H_C^{(2)} | \Psi_i \Phi_o \rangle + \sum_g \frac{\langle \Psi_f \Phi_s | H_C^{(1)} | \Psi_g \rangle \langle \Psi_g | H_C^{(1)} | \Psi_i \Phi_o \rangle}{2\pi(\nu_o - \nu_{gi})} \right. \\ &\quad \left. - \frac{\langle \Psi_f \Phi_s | H_C^{(1)} | \Psi_g \Phi_{os} \rangle \langle \Psi_g \Phi_{os} | H_C^{(1)} | \Psi_i \Phi_o \rangle}{2\pi(\nu_s + \nu_{gi})} \right|^2 . \end{aligned} \quad (10:72)$$

The appropriate matrix elements are evaluated as in 10.2. Note that the state vector for the scattered photon does not include the superscript i for isotropic since for simultaneous two photon scattering a definite relationship exists between the polarizations and directions of the incident and scattered photons. If however, \vec{k}_0 is taken as the polar axis of spherical coordinates, $P(i-f)$ will be independent of azimuth angle ϕ and $\vec{k}_0 \cdot \vec{k}_s = \cos \theta$ where θ is the angle between \vec{k}_0 and \vec{k}_s , or equivalently, the complement of the polar angle of scattered propagation.

The first matrix element term comes from the second term of (10:34). In analogy to the arguments leading to (10:44) and to (10:50), a weight factor of $(1/c)^{1/2}$ is associated with each photon state vector and of $\nu_s/c = p/h$ with the spherical to Cartesian coordinate transformation. Including these factors with (10:34), the first term is

$$2\hbar^{-1} \langle \psi_f \phi_s | H_c^{(2)} | \psi_i \phi_o \rangle = \frac{e^2}{c^2 m} \delta_{if} \cos \theta = \frac{(2\pi e)^2 \left(\frac{\nu_s}{\nu_o} \right)^{1/2}}{\hbar c^2} \frac{\hbar}{2\pi m} \delta_{if} \cos \theta \quad (10:73)$$

recalling that $\nu_s = \nu_o$ for the case $i=f$ required by the delta function. Dirac (1958) showed that

$$\frac{\hbar}{2\pi m} \delta_{if} \cos \theta = \sum_g \langle \psi_f | x_s | \psi_g \rangle \langle \psi_g | x_o | \psi_i \rangle (\nu_s + \nu_{gi}) - \langle \psi_f | x_o | \psi_g \rangle \langle \psi_g | x_s | \psi_i \rangle (\nu_s + \nu_{gf}) \quad (10:74)$$

by appropriate use of quantum identities such as $x_o s_s - x_s x_o = 0$. This allows combination of this first term with the second and third matrix element terms.

The second and third matrix element terms are evaluated in a similar manner. To the factors \hbar^{-1} from (10:65) and $eh^{1/2}/m(2\pi)^2 \nu^{1/2}$ from (10:34) are added density weighting factors, space transformation factors and time derivative factors exactly as in 10.2. The algebra, following 10.2, is left to the reader. The result of these factors is the sum

$$\frac{(2\pi e)^2 \left(\frac{\nu_s}{\nu_o} \right)^{1/2}}{\hbar c^2} \sum_g \frac{\nu_{gf} \nu_{gi}}{\nu_o - \nu_{gi}} \langle \psi_f | x_s | \psi_g \rangle \langle \psi_g | x_o | \psi_i \rangle - \frac{\nu_{gf} \nu_{gi}}{\nu_s + \nu_{gi}} \langle \psi_f | x_o | \psi_g \rangle \langle \psi_g | x_s | \psi_i \rangle. \quad (10:75)$$

In both of the two previous expressions, the angular dependence of the scattering is contained in the subscript of the dipole matrix operators in a manner identical to that illustrated by (10:34). The destruction-creation term in (10:72) has the opposite order or subscripts from the creation-destruction term since the photons are acted on in opposite order.

All three matrix element terms may be combined. The identity, obvious from F10.1, that $\nu_o - \nu_s + \nu_{gf} - \nu_{gi} = 0$, is needed for the reduction. The idealized analysis leading to (10:70) has neglected any damping factor in the oscillatory term. As was done in (10.2.1), some damping factor $\exp(-\delta_g t)$ should be associated with each real intermediate state Ψ_g . This damping differs from the factor associated with spontaneous emission (natural line breadth). It is not clear, either from theory or experiment, what size damping factor is appropriate for two photon scattering. We introduce, by postulate, at this point in the analysis a δ_g into the resonance denominator. Experimental determination either of the extent of a resonance effect in ν_o space or of the magnitude of such an effect will yield a measure of the size of δ_g . We recall that the time independent probability $P(i-f)$ is just equal to the cross section for scattering. Therefore we obtain for two photon scattering

$$\frac{d\sigma}{d\Omega} (i-f) = \frac{(2\pi)^4 \nu_o \nu_s^3}{\hbar^2 c^4} \left| \sum_g \frac{\langle \psi_f | x_s | \psi_g \rangle \langle \psi_g | x_o | \psi_i \rangle}{\nu_o - \nu_{gi} + i\delta_g} - \frac{\langle \psi_f | x_o | \psi_g \rangle \langle \psi_g | x_s | \psi_i \rangle}{\nu_s + \nu_{gi} - i\delta_g} \right|^2 \quad (10:76)$$

for the scattering cross section in area a photon must strike per steradian scattering angle with an incident photon beam of one photon per unit area. This is historically called the Kramers-Heisenberg dispersion formula.

Other forms of this relation are possible. In particular, the cross section is often given as σ for the total cross section integrated over all solid angle. For this purpose one replaces the components x_s and x_o with the displacement vectors \mathbf{r} (averaged over all angles). Since x_s and x_o are related to \mathbf{r} in a $\sin \theta$ fashion, replacement of the x components with \mathbf{r} results in a factor of $\sin^2 \theta = 2/3$ for each matrix element product. Integration over

all solid angle introduces a factor 4π outside the squared factor. This alternate form suppresses the angular dependence of (10:76) and is

$$\sigma = \frac{2^3(2\pi)^5}{3^2 h^2 c^4} \nu_o \nu_s^3 \left| \sum_g \frac{\langle \psi_f | \mathbf{er} | \psi_g \rangle \langle \psi_g | \mathbf{er} | \psi_i \rangle}{\nu_o - \nu_{gi} + \delta_g} - \frac{\langle \psi_f | \mathbf{er} | \psi_g \rangle \langle \psi_g | \mathbf{er} | \psi_i \rangle}{\nu_s + \nu_{gi} + \delta_g} \right|^2 \quad (10:77)$$

10.3.2 Applications

The two photon scattering cross section (10:76) exhibits the characteristic frequency to the fourth power scattering cross section dependence for the case when ν_o is far from ν_{gi} . The scattering for the two photon case is considerably weaker than the single photon (two step) scattering because of the resonance denominators which are generally much larger than unity. Note also the possible interference between the various terms in the summation which can further reduce the scattering intensity in certain cases.

The frequency factor $\nu_o \nu_s^3$ is given in some derivations as $(\nu_o - \nu_{fi})^4 = \nu_s^4$, depending on the approximations used. The difference in the factors for practical optical scattering is insignificant. Actual calculations using (10:76) directly are extremely difficult because the wave functions ψ are not known for the atmospheric molecules. Calculations for hydrogen have been made for the rotational (Yoshino and Bernstein, 1958) and vibrational (Victor et al., 1967) Raman transitions in molecular hydrogen. These calculations are used to provide an absolute scale to scattering cross section measurements made relative to hydrogen as an internal reference. Absolute measurements of scattering cross sections are very difficult to make. Such measurements as have been made are consistent with the prediction of (10:76) applied to molecular hydrogen.

The idealized analysis leading to (10:76) neglected a damping term δ_g in the resonance denominator. Far away from resonance this term is unimportant, but near resonance it has a significant effect. At present δ_g is difficult to estimate theoretically and has not been experimentally evaluated for gases. The damping δ_g has the effect of determining how far from ν_{gi} a resonance effect extends.

It is obvious that significant increases in cross section over the non-resonant values are possible if ν_o approaches a real energy difference ν_{gi} in the molecule. When ν_o is chosen in such a manner, one term in the sum over g predominates and (10:76) more closely approaches (10:57). The questions of whether one or two photon scattering occurs for ν_o near ν_{gi} and of what is the nature of the scattering predicted by (10:76) in this case have not been resolved. Most informed opinion suggests that one and two photon scattering coexist near resonance, with the stronger single photon interaction dominant except where pressure quenching is severe. A valid alternative view is that as ν_o approaches ν_{gi} the interaction time increases and the scattering changes smoothly from a more Raman like to a more fluorescence like scattering behavior. Some experimental evidence to clarify these views has recently been obtained (Holzer et al., 1970). The second resonance denominator will become significant only if $\nu_g < \nu_i$ or if the initial state is at a higher energy than available intermediate states. For typical laser pumps the scattered frequency ν_s will be the order of 600 THz which is far greater than the energy of any initial state with significant population for atmospheric molecules. Hence this denominator will always be large and may be practically ignored for remote sensing considerations.

If state i is identical to state f the scattering cross section (10:76) refers to Rayleigh scattering. It is also possible to analyze Rayleigh scattering by classical considerations of the average index of refraction. This classical analysis is mentioned in 10.4 and is adequately treated in standard texts (Kerker, 1969). Section 10.3.3 treats the classical result only.

If state i differs from state f the scattering cross section (10:76) refers to Raman scattering. Raman scattering is of particular importance to atmospheric composition profile studies (24.2.2). In principle, information on absorption and emission line strengths gives useful estimates of the applicable dipole matrix elements. This author is unaware of any such scattering estimates which have been made so far. The existence of absorption lines in the spectrum of a molecule of interest does provide a satisfactory reason for looking for resonance effects with laser pump frequencies in or near the absorption region. The absorption band of ozone in the blue is an example of a promising region to look for resonance Raman scattering to occur.

10.3.3 Empirical Two Photon Scattering Theory.

We have already seen that single photon scattering is necessarily isotropic and completely unpolarized. For two photon scattering a definite angular dependence and depolarization ratio exists. Figure (10:2)

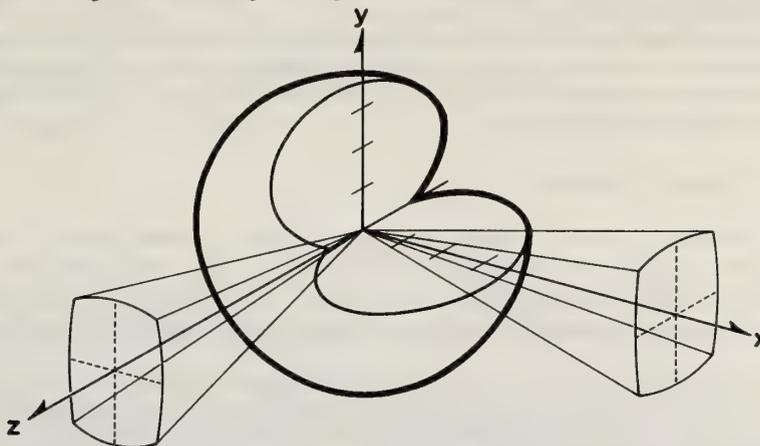


Figure 10.2 Representative three dimensional Raman scattering diagram for scattering intensity as a function of angle. Laser radiation is directed along the y axis polarized in the z direction. Instrumental collection solid angles are indicated for viewing scattering in the x and z directions.

indicates the three dimensional variation of scattered intensity with scattering angle for the specified incident laser polarization and propagation direction. The intensity follows an $I(\theta, \phi) = A + B \sin^2 \theta$ where θ is measured about the axis of incident polarization (Porto, 1966). For aid in visualization only, the pattern may be thought of as the superposition of a completely polarized dipolar pattern with the $\sin^2 \theta$ character and a completely unpolarized, spherically isotropic scattering contribution.

The two photon scattering is completely described on an empirical basis by the two intensity constants A and B. Alternatively, one can describe the scattering by the 90° differential cross section $d\sigma/d\Omega = A + B$ and the depolarization ratio $\rho = A/A+B$. Note that ρ is also the ratio of vertically to horizontally polarized intensity when viewed in the x direction in (F10.2). Some authors give the depolarization ratio as $\rho = A/B$. Care in understanding the definition used for ρ is necessary for intercomparison of reported data.

A classically derived expression useful for Rayleigh scatter (Ditchburn, 1963) is

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 \left(\frac{n-1}{N} \right)^2 \frac{\nu^4}{c^4} \sin^2 \Theta \left(\text{in } \frac{\text{area}}{\text{steradian}} \right) \quad (10:78)$$

where $d\sigma/d\Omega$ is the differential scattering cross section per molecule, n is the refractive index of the medium and N is the density of scatterers. The more familiar form of the expression results if (10:78) is integrated over Ω from 0 to 4π , multiplied by N scatterers/unit volume, and $\sin^2 \theta$ is replaced by its spherical average of 2/3. One then obtains the total scattering extinction coefficient per unit path. For example, at $\nu = 600$ THz ($\bar{\nu} = 30,000 \text{ cm}^{-1}$ in the green) and air at STP the differential cross section (10:78) is

$$\frac{d\sigma}{d\Omega} = (7.3 \times 10^{-32}) \sin^2 \Theta \frac{\text{m}^2}{\text{sr}} \quad (10:79)$$

by this formula. While the classical description is useful for numerical calculations, it does not demonstrate the important physics of the scattering, which is illustrated in (10:76). Actually (10:78) is an approximation which ignores depolarization effects (assumes $A = 0$ in F10.2). For isotropic scattering centers like the rare gases, the approximation of (10:78) is valid; experimentally one finds ρ (rare gases) < 0.005 . For more polar molecules such as CO_2 , ρ values of 0.10 and higher are found.

No such classical expression exists for Raman scattering, which is the other type of two photon scattering permitted by (10:76). Experimental measurements suggest that the magnitude of Raman scatter for nitrogen at STP and a laser pump near $\nu = 584$ THz (approximately the same conditions considered above for Rayleigh scatter) is

$$\frac{d\sigma}{d\Omega} = (1 \times 10^{-36}) + (2 \times 10^{-35}) \sin^2 \Theta \frac{\text{m}^2}{\text{sr}} \quad (10:80)$$

per molecule (Fouche and Chang, 1971). Nitrogen is free of direct absorptions in the region $\nu = 15$ THz to 1500 THz so that by (10:76) we expect no resonances in this region. With no resonances, (10:76) is used to predict the per molecule nitrogen scattering cross section for any other pump frequency ν_0 and the corresponding scattered frequency $\nu_s = \nu_0 - \nu_{fi}$.

Practical measurements of A and B for two photon scattering avoid the need of calibrating the measuring instruments for absolute photometric intensity. Instead of absolute calibration, the spectrometer is calibrated for relative spectral response and the A and B values of the molecule to be measured are ratioed to the A' and B' of a nitrogen internal standard.

10.4 Classical Analysis of Scattering

If the collection of bound electrons comprising the single particle scatterer is large enough, the quantum laws pass over to the classical limit. For such a particle one applies the laws of classical electrodynamics to a solid dielectric with a complex index of refraction. With suitable simplified particle geometry, electromagnetic scattering as a function of index, incident frequency, and particle geometry can be calculated.

Excellent, complete treatments of the theory of such scattering exist in the literature. This situation is in marked contrast to the situation in quantum scattering theory where no unified exposition of the scattering cross section exists. Since classical theory exists elsewhere, we will omit in this section much of the detail which characterizes 10.1, 10.2 and 10.3. The reader interested in details of the derivation of the following formulas should refer to van de Hulst (1957), Kerker (1969) and Born and Wolf (1965). We will mention some aspects of the derivation, following Kerker (1969) more closely than the other sources.

For tropospheric sensing with laser sources, we limit the discussion to scattering of polarized radiation from dielectric spheres. Comment on the problem of nonspherical scatterers follow from generalized aspects of the derivation.

The method for the classical analysis of scattering (Mie scattering) involves solving Maxwell's equations for the electromagnetic field when a plane wave is incident on a material interface. Maxwell's equations with the boundary conditions at the interface lead to ordinary differential equations. The infinite series solutions to these differential equations provide the desired scattering functions.

10.4.1 Review of Electromagnetic Wave Solution

A more detailed discussion of the electromagnetic wave equations is found in 8.1 to 8.3 and 8.6.1. It is instructive to compare (10:96) with (8:104).

The solution of Maxwell's equations is aided by the introduction of Debye's scalar potentials D_1 and D_2 which satisfy the scalar wave equation

$$\nabla^2 D_i - s\mu\dot{D}_i - \epsilon\mu\ddot{D}_i = 0, \quad (10:81)$$

where s is the specific conductance of the medium, ϵ and μ are the dielectric constant and magnetic permeability, and the dot superscript implies differentiation with respect to time. The electric and magnetic field vectors are expressible in spherical coordinates in terms of the Debye potentials and their spatial derivatives. Born and Wolf (1964) derive these expressions for the field vectors. Each vector is the sum of components of two linearly independent fields. For example

$$E_\theta = eE_\theta + mE_\theta = \frac{1}{r} \frac{\partial^2 (rD_1)}{\partial r \partial \theta} + k^2 \frac{1}{r \sin \theta} \frac{\partial (rD_2)}{\partial \phi} \quad (10:82)$$

where $k^2 = (\omega^2/c^2)(\epsilon + i4\pi s/\omega)$ is the square of the usual wave number (and ω is the angular frequency of the electric field $\omega = 2\pi\nu$).

Solutions for the Debye potentials are found via separation of variables in the form

$$D_i = R_i(r)\Theta_i(\theta)\Phi_i(\phi). \quad (10:83)$$

The solutions are

$$\Phi(\phi) = a_m \cos(m\phi) + b_m \sin(m\phi)$$

and

$$\Theta(\theta) = P_\ell^{(m)}(\cos \theta) , \tag{10:84}$$

the well known associated Legendre functions. For nonzero values of the function we require $|m| \leq \ell$ and $m = -\ell, \dots, \ell$. The radial equation for the D_i is the Bessel equation. The solution of this equation is the general cylindrical function which may be expressed as a linear combination of standard Bessel functions $J_{\ell+1/2}(\rho)$ and Neumann functions $N_{\ell+1/2}(\rho)$. With the parameter $\rho = kr$, the equivalence is

$$rR(r) = c_\ell(\pi\rho/2)^{1/2} J_{\ell+1/2}(\rho) - d_\ell(\pi\rho/2)^{1/2} N_{\ell+1/2}(\rho) = c_\ell\psi_\ell(\rho) + d_\ell\chi_\ell(\rho) . \tag{10:85}$$

The function

$$\zeta_\ell(\rho) = \psi_\ell(\rho) + i\chi_\ell(\rho) \tag{10:86}$$

is a useful combination for future work.

The incident, internal (within the scatterer) and scattered waves are expressed in terms of the general solutions for the D_i . Only the $\psi_\ell(\rho)$ and $P_\ell^{(m)}(\cos \theta)$ enter for the incident and internal waves, since χ_ℓ is not well behaved at the origin, while the $\zeta_\ell(\rho)$ and $P_\ell^{(m)}(\cos \theta)$ are useful for the scattered wave since $\zeta_\ell(\rho)$ vanishes at infinity. The boundary conditions at the surface of the scattering body require that the tangential components of the field vectors \mathbf{E} and \mathbf{H} be continuous across the interface. These conditions give equations for the a, b, c, and d series expansion coefficients. The complication due to nonspherical scatterers enters in the boundary condition evaluation of expansion coefficients.

For a spherical scatterer of radius a and real refractive index n (the ratio of velocity of electromagnetic wave propagation in a vacuum to that in the material), we define a dimensionless size parameter α where

$$\alpha = 2\pi a/\lambda . \tag{10:87}$$

Wavelength λ is the wavelength of the radiation in the medium surrounding the scattering particle. The refractive index of the particle relative to that of the medium is m, which is called the relative refractive index

$$m = n(\text{scatterer})/n(\text{surrounding medium}) . \tag{10:88}$$

A parameter $\beta = m\alpha$ is also useful. In terms of these dimensionless parameters, the needed series expansion coefficients are

$$a_n = \frac{\psi_n(\alpha)\psi_n'(\beta) - m\psi_n(\beta)\psi_n'(\alpha)}{\zeta_n(\alpha)\psi_n'(\beta) - m\psi_n(\beta)\zeta_n'(\alpha)} \tag{10:89}$$

$$b_n = \frac{m\psi_n(\alpha)\psi_n'(\beta) - \psi_n(\beta)\psi_n'(\alpha)}{m\zeta_n(\alpha)\psi_n'(\beta) - \psi_n(\beta)\zeta_n'(\alpha)} .$$

The prime denotes differentiations with respect to the argument.

10.4.2 Mie Scattering Results

To make these general expressions useful, consider spherical polar coordinates with the incident radiation propagating in the z direction, polarized in the x direction ($\phi=0$). The scattering plane contains the z (incident) axis and is at angle ϕ to the x axis. Angle θ measures the scattering angle to the z axis in the scattering plane. Scattered polarization is resolved parallel to (in) and perpendicular to (out of) the scattering plane.

For tropospheric scattering we are interested only in the behavior of the solutions at large distances from the scatterer. In the far field, only transverse electromagnetic waves exist. In this case, E_r and H_r may be neglected and $\zeta_\ell(\rho)$ becomes

$$\zeta_\ell(\rho) = i^{(n+1)} \exp(-i\rho) \tag{10:90}$$

$$\zeta_\ell'(\rho) = i^n \exp(-i\rho)$$

where the prime again indicates differentiation with respect to the argument.

The practical results are given in terms of two sets of functions. The angular functions for the assumed geometry are

$$\begin{aligned}\pi_n(\cos \theta) &= P_n^{(1)}(\cos \theta) / \sin \theta \\ \tau_n(\cos \theta) &= \frac{d}{d\theta} P_n^{(1)}(\cos \theta) .\end{aligned}\quad (10:91)$$

Using these and the Mie a_n and b_n coefficients (10.4.1), the intensity functions are

$$\begin{aligned}i_{\perp} &= \left| \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [a_n \pi_n(\cos \theta) + b_n \tau_n(\cos \theta)] \right|^2 \\ i_{\parallel} &= \left| \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [a_n \tau_n(\cos \theta) + b_n \pi_n(\cos \theta)] \right|^2 .\end{aligned}\quad (10:92)$$

The results for radiation polarized parallel and perpendicular to the scattering plane are

$$\begin{aligned}I_{\parallel} &= (\lambda/2\pi r)^2 i_{\parallel} \cos^2 \phi \\ I_{\perp} &= (\lambda/2\pi r)^2 i_{\perp} \sin^2 \phi .\end{aligned}\quad (10:93)$$

Both the a_n b_n depend on the size parameter α . For small α , a series approximation is useful. This case is that of a dielectric sphere of a diameter very small with respect to a wavelength ($\alpha \ll 1$). Such scattering is generally referred to as Rayleigh scattering from spheres, which is a fundamentally different analysis than molecular scattering with no change in frequency as discussed in 10.3.1. Incidentally, some aspects of both particulate and molecular scattering are the same, such as the ν^4 frequency dependence developed below. Since the a_n and b_n also depend on β , the additional restrictions must be made that $|m| \approx 1$, although it may still be complex.

For $\alpha \ll 1$ and $|m| \approx 1$ the first terms of a series expansion in terms of α are (Kerker, 1969)

$$\begin{aligned}a_1 &= \frac{2i}{3} \left(\frac{m^2 - 1}{m^2 + 2} \right) \alpha^3 & b_1 &= -\frac{i}{4\pi} (m^2 - 1) \alpha^5 \\ \pi_1(\cos \theta) &= 1 & \tau_1(\cos \theta) &= \cos \theta . \\ a_2 &= \frac{i}{15} \left(\frac{m^2 - 1}{2m^2 + 3} \right) \alpha^5\end{aligned}\quad (10:94)$$

Using these in (10:93) and keeping terms to α^3 (particle diameter a)

$$\begin{aligned}I_{\parallel} &= \left(\frac{2\pi}{\lambda} \right)^4 \frac{a^6}{r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \cos^2 \phi \\ I_{\perp} &= \left(\frac{2\pi}{\lambda} \right)^4 \frac{a^6}{r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \cos^2 \theta \sin^2 \phi\end{aligned}\quad (10:95)$$

If the incident light is unpolarized, it may be resolved into components parallel and perpendicular to the plane of scattering. Summing the scattered polarizations

$$I_{\text{unpol}} = \left(\frac{2\pi}{\lambda} \right)^4 \frac{a^6}{2r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 (1 + \cos^2 \theta) .\quad (10:96)$$

For $\alpha < 0.1$, one expects error of approximation less than 1%.

10.4.3 Application of Mie Scattering Results

The Rayleigh approximation provides a useful computational expression. Exact calculations of (10:92) are possible with modern digital computers. Much of this work has been done and is available. Kerker (1969) reviews various tabulations which can be used by interested researchers. Scattering diagrams for dielectric spheres are also available commercially (Science Spectrum, 1970), although for many purposes it is more convenient to generate the scattering functions, as needed, as part of a larger program. Computational matters are discussed by Deirmendjian (1969).

Quantitative aspects of Mie scattering are useful for considering tropospheric scattering effects. As the size parameter increases, scatter in the forward direction is much stronger than backscatter. For example, forward and backscatter are very nearly equal at $\alpha = 0.01$ but forward scatter is almost two orders of magnitude larger than backscatter at $\alpha = 1.0$. As α is increased further, the forward scattering maximum breaks into a number of smaller subsidiary maxima. Born and Wolf (1964) review some results of angular dependence of scattering intensity. The exact nature of the forward scatter enhancement of course depends on the relative refractive index, absorption (imaginary part of the complex index), and polarization of the incident and detected light as well as on α .

Similarly, the scattered intensity at a fixed angle depends strongly on size parameter α . Generally the scattering intensity increases smoothly with increasing α , with approximately an α^4 dependence, up to a maximum intensity at $\alpha \approx 1$. For larger α , the $I(\alpha)$ function exhibits a long series of irregular maxima and minima typically oscillating about a value approximately one half that of the peak intensity $I(1)$. For example, at 90° scattering for $m = 1.33$ (no absorption), the maximum and minimum are separated by typically 1.7α units and differ from each other by approximately four orders of magnitude. These large oscillations in $I(\alpha)$ are damped for absorbing particles. In practical cases, the fluctuations are further smoothed by polydisperse distributions of α (10.5.3).

Mie scattering for nonspherical particles can be calculated by a point matching method (Kerker, 1969). This calculational technique is not yet well developed, but shows promise for providing an approximate solution, to any desired accuracy, for scattering by nonspherical particles such as finite cylinders, cubes, spheroids and the like. In general, one considers a series solution in α , truncated after N terms to provide the desired accuracy appropriate to the magnitude of α . This is also done for computer calculations of spherical scattering centers. The boundary conditions give four equations involving $4N$ unknowns. Provided N discrete points on the nonspherical geometrical boundary are given, one can solve the boundary equations for the a_n , b_n , c_n , and d_n . The N geometrical points must be chosen so that they lead to independent solutions for the $4N$ unknowns. One assumes that the solution valid at N points on the particle surface is approximately valid everywhere. The point matching techniques have proven successful as an alternative solution for spheres and infinite circular cylinders. Holland and Gange (1970) discuss spherical representations for non-spherical particles.

Consideration of Mie scattering cross section provides a useful comparison with the quantum mechanical calculations of scattering cross sections. By integrating Poynting's vector for the incident plane wave and for the scattered spherical electromagnetic wave, one can obtain (Kerker, 1969)

$$\sigma_{\text{scat}} = (\lambda^2/2\pi) \sum_{n=1}^{\infty} (2n + 1)(|a_n|^2 + |b_n|^2) \tag{10:97}$$

$$\sigma_{\text{abs}} = (\lambda^2/2\pi) \sum_{n=1}^{\infty} (2n + 1)\text{Re}(a_n + b_n) ,$$

where the scattering and absorbing cross sections are expressed separately. The sum of scattering and absorbing cross sections is the extinction cross section of the particle. The extinction cross section of a large opaque obstacle is twice its geometrical cross section (Born and Wolf, 1964). For a smaller particle the same diffraction effects can cause the extinction cross section to be many times the geometrical cross section.

It is possible to obtain considerable information about scattering particles by examining the scatter function for the particle, either $I(\theta)$ as a function of angle, or $I(\alpha)$ as a function of incident wavelength or frequency. Unfortunately, the scattering functions do not uniquely characterize a particle. In general, there

will be many possible combinations of size and refractive index which will result in a particular scattering function, especially when experimental uncertainty is considered. The commercial scattering function atlas (Science Spectrum, 1970) is in fact designed for measuring particle size for scatterers of a known refractive index. Extension of this technique to polydispersions (10.5) or unknown materials is difficult.

It is important to emphasize that the scattered intensities for some particular size parameter α (10:93) are strongly dependent on relative refractive index m . For a monodisperse aerosol, the scattering will be very sensitive to slight changes in the temperature and density of the atmosphere. For example, a temperature change of 2°C can lead to a factor of 1.5 change in scattering intensity and to a change in degree of polarization of the scattered light (Cohen, 1971). These effects will be less important for polydispersions (10.5.3).

10.5 Collective Scattering Effects

We have considered so far the scattering from individual scattering centers composed of systems of bound electrons. Depending on the system and the type of interaction sought, either quantum or classical scattering analysis is appropriate. Whatever the mechanism of scattering from individual centers, when more than one scattering center is involved, the combined or collective interaction effects must be considered to accurately model the scattering.

10.5.1 Coherent Scattering

If individual scatterers are regularly spaced, as in an ideal crystal lattice, all scattering centers will scatter in a definite phase relation to each other. In this case of coherent scattering, amplitudes of scattered waves from each center will add. The result of scattered wave superposition is that only forward scattering will occur; there will be no lateral or backscatter. Lateral and back-scattered waves interfere destructively (amplitudes cancel) in coherent scattering.

In practical terms, coherent scattering is not scattering at all. The only effect on the incident beam by the phase-related scatters is to change the propagation velocity of the radiation in the medium. No energy is scattered out of the beam.

When the density of scatterers is very high on a wavelength scale, for example 10^3 or more scatters in a cube of radiation-wavelength dimension, then coherence effects begin to be important even though the individual scatters are still randomly oriented. This means that scattering from a homogeneous medium (homogeneous on the scale of a wavelength) will be coherent scatter, and scattering from an inhomogeneous medium (on the same scale) will be incoherent.

In a gas at atmospheric or lower pressures, the molecular density is low on an optical or infrared wavelength scale and the scattering will be incoherent. This means that scattered intensities (not amplitudes) will add from each random scattering center. The total scattering will be the single particle scatter intensity of (10.2) and (10.3) multiplied by the number of scattering centers. In a liquid the scattering is no longer completely incoherent. Total scattering is not proportional to molecular density, but is less than would be calculated from independent scattering assumptions. For example in ethyl ether the liquid scattering is lower by a factor of 20 (Strong, 1958) than the intensity calculated from independent scattering and verified by many experiments for gases. Rayleigh scatter in gases depends on random positions of scattering centers, but is not strictly scattering from density fluctuations as is sometimes loosely claimed. The density fluctuation model has proven useful for liquid scattering. Strong (1958) provides a brief, but clear, summary of these coherence effects.

The same random distribution of scatterers is required for effective Mie scattering. The general requirement of few scattering centers in a radiation-wavelength scale volume is also applicable. In practical tropospheric scattering, this condition for independent Mie scattering is almost always satisfied.

When large (particulate) scattering centers are involved in the system, another type of coherence interaction may occur. Even though there are very few scatters per wavelength scale volume (typically fewer than one), the near field distributions on the particulate to surrounding medium interfaces may interact to produce a coherence effect. This effect is discussed qualitatively by van de Hulst (1957). A satisfactory working rule seems to be that the scattering particles must be separated by the order of ten times their radius to eliminate near field interaction effects. Even in a very dense fog this condition is satisfied and independent scattering model calculations are valid for practical tropospheric applications.

In almost all cases of tropospheric scatter phenomena, it is possible to ignore coherent scattering effects and assume independent scattering, where the scattered intensity is simply the single center scattering function multiplied by the number of scattering centers.

10.5.2 Multiple Scattering

If each scatterer is exposed only to radiation in the incident beam, and scattered light does not strike another scatterer before detection, then the single scatter model is valid. Clearly this implies either a very low density of scatterers, a weak scattering interaction (small cross section) or both. If the scatterers are illuminated with a significant portion of previously scattered light, then some of the detected light has been scattered more than once and a multiple scattering model is appropriate.

The practical test for the absence of multiple scattering is the same as the test for independent (incoherent) scattering. If the scattered intensity and the extinction of the main beam are linearly proportional to the density of scatterers, then one is certain that both independent and single scattering conditions exist. Nonlinearity in the scattering vs. density function is evidence for either coherent or multiple scattering conditions.

In contrast to coherent interactions which depend only on scattering center density, multiple scattering effects depend principally on total concentration (in molecules per area) integrated along the radiation path. That is, both scatterer density and total extent of the scattering cloud effect multiple scattering. Even if the scatterers are completely independent, multiple scattering, as in a cloud for example, can be important.

The check of linear proportionality between scatterer density and scattered intensity is difficult to apply in the troposphere. An alternative criterion (van de Hulst, 1957) involves the extinction of the incident radiation. If the intensity at some distance x is

$$I(x) = I(0) e^{-\tau x}, \quad (10:98)$$

τx is the optical depth along the beam to point x . Here the transmission is measured from the original scatterer to the receiver. An empirical rule suggests that if $\tau x < 0.1$ (transmission $> 90\%$), single scattering is predominant. In the range $0.1 < \tau x < 0.3$ (transmission 75% to 90%) corrections for multiple scatter allow meaningful results to be obtained within the framework of independent, single scatter theory. For $\tau x > 0.3$ multiple scattering requires a complicated, approximate theory which is not yet well developed and which is beyond the scope of this short summary.

Multiple scattering is very important in the troposphere, especially as part of the overall radiative transfer problem. However such scattering is a limitation, not a tool, for remote sensing studies in the troposphere since multiple scattering tends to make interpretation of atmospheric scattering returns extremely difficult.

10.5.3 Polydisperse Size Distributions

Deirmendjian (1969) has written a definitive study of scattering on spherical polydispersions. This information is of particular application to the atmospheric questions of radiation transfer in planetary haze and clouds. In this way it is related to the multiple scattering effects mentioned above (10.5.2). From the point of view of remote sensing of the troposphere, the question of the scattering effects of polydisperse size distributions is of importance for characterizing the size parameters of a particular aerosol sample.

A qualitative effect of polydisperse distributions is the smoothing of the $I(\alpha)$ function (10.4.3) already discussed. Actual numerical $I(\alpha)$ and $I(\theta)$ functions for polydispersions are calculated precisely as outlined for Mie theory (10.4) for particles of radius r . Scattered intensities are summed for the discrete r distribution, using number density coefficients appropriate for each r . These distribution coefficients may be chosen by some theoretical model or by actual size distribution measurements. A useful size distribution is that suggested by Junge. This distribution seems to apply to some natural aerosol states and is a two-parameter function involving a multiplicative scaling coefficient c and a shaping parameter ν . If dn is the number of particles between r and r plus dr , the size distribution function is

$$f(r) = \frac{dn}{dr} = cr^{-(\nu+1)}. \quad (10:99)$$

Values of ν in range 2.2 to 4.0 are representative of actual distributions of natural particles in some instances, for example silicates in the 0.1 μm to 10.0 μm size range. A more general form discussed by Deirmendjian (1969) is

$$f(r) = ar^\alpha \exp(-br^\gamma) \quad (10:100)$$

which contains the Junge distribution as a special case.

Consideration of polydispersions adds nothing new to the previous theory (10.4.2) but rather adds computational complexity only. Calculations for Mie scattering by polydispersions are time consuming but straightforward. Deirmendjian's (1969) work tabulates many needed results.

10.5.4 Doppler Frequency Shifts

We have so far assumed that the scatterer, whether a quantum or classical bound electronic system, and the detector are at rest with respect to each other. In practice this assumption is generally not valid. Molecules are in rapid thermal motion, small particles are subject to Brownian motion and larger particulates follow microscale atmospheric flows.

Particle or molecule velocities in the troposphere are small with respect to the speed of light so that we may neglect relativistic Doppler effects and apply the classical formalism. The wave vector of the incident light has the direction of propagation and magnitude $|\mathbf{k}_0| = 2\pi\nu_0/c$ exactly as before (10:21). If n is the refractive index of the atmosphere of interest, then the speed of propagation is c/n . Considering the shift of both the incident radiation (ν_0) with respect to the scatterer and the scattered radiation (ν_s) with respect to the detector, the Doppler shift is

$$\Delta\nu_D = \nu_s - \nu_0 = (n/2\pi)v \cdot (\mathbf{k}_s - \mathbf{k}_0) . \quad (10:101)$$

In particular, for the case of backscatter $\mathbf{k}_0 = -\mathbf{k}_s$ and $\Delta\nu_D$ is negative for velocity components away from the detector, as one would expect. On an optical scale the frequency shifts are small. For example, a scatterer velocity of 30 cm/sec results in a fractional shift of 10^{-9} or a shift of approximately 1.2 MHz for green light in backscatter.

The Doppler shift of scattered light provides in principle a method for remote sensing of tropospheric wind velocities and clear air turbulence. Practical problems may severely limit such applications.

10.5.5 Brillouin Scattering

An acoustic wave in the atmosphere causes regular variations in the molecular density that are periodic in space and time. These acoustic waves propagate at some velocity s with respect to the medium, and have a characteristic frequency f and wave vector $|\mathbf{K}| = 2\pi f/s$. This periodic pressure and density wave is not consistent with the assumption of incoherent scatter made previously (10.5.1).

Since the wave is only a one dimensional ordering, in contrast to the ideal crystal lattice discussed earlier, scattering in directions other than forward occurs. Pictorially the situation for scattering from an acoustic wave is identical to scattering from a moving diffraction grating. This coherent scattering is superimposed on the generally stronger incoherent scatter. For coherent scatter, as for a diffraction grating, one adds amplitudes rather than intensities of scattered fields. This results in the condition

$$\mathbf{K} = \mathbf{k}_0 - \mathbf{k}_s \quad (10:102)$$

where the acoustic wave vector \mathbf{K} has been defined and \mathbf{k}_0 and \mathbf{k}_s are the incident and scattered optical waves respectively. This means that only acoustic waves propagating in a direction approximately bisecting the angle between \mathbf{k}_s and \mathbf{k}_0 will cause optical scatter in the prescribed direction. The density variations, which are the scattering centers, are moving with velocity s so that $|\mathbf{k}_s|$ is Doppler frequency shifted as given above (10:101). Since the acoustic waves in the allowed direction may propagate either toward or away from the source-detector direction, the spectrum of scattered light intensity as a function of frequency will exhibit a strong, central Rayleigh line (not shifted in frequency from the source) flanked by lines of the Brillouin doublet which are Doppler shifted in frequency an equal amount above and below the central frequency.

Actual experiments on Brillouin scattering in CO₂ are discussed in a useful review (Mountain, 1966). The scattering results are particularly useful for measurements of acoustic propagation and absorption characteristics in fluids. Mountain (1966) discusses theoretical scattering intensity expressions which are unfortunately not sufficiently quantitative for our purposes.

Brillouin scattering in gases under normal atmospheric conditions with random (thermal) acoustic fields is extremely weak. Typically, Brillouin studies are conducted in liquids or in gases near the critical point. In CO₂ for example, useful scattering occurs at molecular densities more than two hundred times that of the standard atmosphere. At standard atmospheric densities the scattering intensity is predicted to be at least five orders of magnitude weaker than the intensity of scattering when the gas is near the critical point. Typical Brillouin shifts are the order of 10⁹ Hz for visible light incident on CO₂ near the thermodynamical critical point.

Brillouin scattering alone has not yet been proposed as or demonstrated to be a useful remote sensor of the troposphere. Such scattering has been used with artificial acoustic sources as a velocity probe (Bogh and Cooper, 1968). However a scattering study of atmospheric gases can yield useful information on the acoustic constants of the gases such as the ratio of specific heats, sound absorption coefficients, velocity of thermal wave propagation and the like. Fabelinskii (1968) discusses Brillouin scattering particularly as applied to liquids and crystals. Some of this treatment is useful background for studies of scattering in gases. Jorna (1971) discusses the effect of stimulated Brillouin scattering on laser propagation.

10.6 Summary

The scattering intensity from an isolated scattering center can in principle be calculated from the formalism outlined in this chapter. Fundamentally different approaches are required for atmospheric scatterers that require a quantum mechanical description (free molecules) and those that can be satisfactorily treated by classical expressions (particles).

Quantum scattering results are useful for predicting changes in scattering cross section with changes in incident frequency and for understanding the relevant molecular properties. Unfortunately, incomplete knowledge of molecular wave functions does not allow quantitative calculation of the scattering intensities of many-electron molecules. Experimental measurements of scattering cross sections do give quantitative results, which are necessary for evaluating any practical tropospheric sensor. Single step (second order) Raman and Rayleigh scattering is independent of buffer gas pressure. Two step (first order) resonance and fluorescence scattering is in general much stronger than the scattering that depends on a second order process, but first order scattering is partly quenched by intermolecular collisions. Raman, fluorescence and, to some extent, resonance scattering is characteristic of a particular molecule. These phenomena may be used for species identifications and concentration measurements.

Classical Mie scattering theory is used to calculate the quantitative scattering intensity from dielectric particles (usually spheres) of a known index of refraction. Theoretical and experimental determinations of scattering from single particles agree well. Practical application to real aerosols with a distribution of sizes and with uncertain refractive indices is more difficult than pure single particle calculations. The practical calculations are straight forward but lengthy.

Collections of individual scatterers in the troposphere are of sufficiently low spatial number density that the assumption of independent scattering is generally valid on an optical wavelength scale. This means that single particle scattering intensities, rather than amplitudes, can be added to determine the overall scattering intensity. Collective scattering effects on the scattered intensity from a strongly scattering region are difficult to model accurately and are beyond the scope of the present discussion.

Applications of this scattering theory are discussed in chapters 14, 23, and 24.

10.6.1 Summary of Useful Scattering Formulas

This section collects the scattering formulas developed earlier in the chapter for various types of scattering. The original equation number is retained for reference. The variables are defined in the original derivation.

For the scattering cross section related to fluorescence or resonance re-radiation, we found the expression

$$\frac{d\sigma}{d\Omega} [i - f(g)] = \frac{(2\pi)^5 \nu_{gi} \nu_{gf}^3 R(\nu_o - \nu_{gi})}{h^2 c^4 \gamma_{ig} \gamma_{gf}} |\langle \psi_f | \text{ex}_s | \psi_g \rangle|^2 Q(n_i T p q) |\langle \psi_g | \text{ex}_o | \psi_i \rangle|^2 \quad (10:64)$$

For the scattering cross section related to Raman and Rayleigh scatterings, we developed the formula

$$\frac{d\sigma}{d\Omega} (i - f) = \frac{(2\pi)^4 \nu_o \nu_s^3}{h^2 c^4} \left| \sum_g \frac{\langle \psi_f | \text{ex}_s | \psi_g \rangle \langle \psi_g | \text{ex}_o | \psi_i \rangle}{\nu_o - \nu_{gi} + i\delta_g} - \frac{\langle \psi_f | \text{ex}_o | \psi_g \rangle \langle \psi_g | \text{ex}_s | \psi_i \rangle}{\nu_s + \nu_{gi} - i\delta_g} \right|^2 \quad (10:76)$$

For Rayleigh scattering in the classical limit, the expression

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 \left(\frac{n-1}{N} \right)^2 \frac{\nu^4}{c^4} \sin^2 \Theta \left(\text{in } \frac{\text{area}}{\text{steradian}} \right) \quad (10:78)$$

is useful.

The Mie scattered intensity for light polarized parallel and perpendicular to the scattering plane respectively is

$$I_{\parallel} = (\lambda/2\pi r)^2 i_{\parallel} \cos^2 \phi \quad (10:93)$$

$$I_{\perp} = (\lambda/2\pi r)^2 i_{\perp} \sin^2 \phi .$$

In the case of very small size parameters α (particle size to wavelength ratio) the Mie scattered intensity is approximately

$$\begin{aligned} I_{\parallel} &= \left(\frac{2\pi}{\lambda} \right)^4 \frac{a^6}{r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \\ I_{\perp} &= \left(\frac{2\pi}{\lambda} \right)^4 \frac{a^6}{r^2} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \cos^2 \theta . \end{aligned} \quad (10:95)$$

where the variables are defined in 10.4.2.

The Doppler shift in frequency resulting from scattering from a moving scattering center is

$$\Delta\nu_D = \nu_s - \nu_o = (n/2\pi)v \cdot (k_s - k_o) . \quad (10:101)$$

These formulas are useful if carefully applied with the limitations stated in the original derivation sketch. An interesting common feature is the ν^4 (or $1/\lambda^4$) dependence on incident photon frequency that exists in (10:64), (10:76), (10:78), and (10:95). The first two quantum scattering expressions involve the energy levels and wave functions of the molecular scattering centers, while the three classical scattering intensity formulas are concerned with scattering particle size and refractive index.

10.6 References

Blogh, J., and D. C. Cooper (1968), Method for measuring wind turbulence, *Nature* 220, 1118-1119.

Born, M., and E. Wolf, 1964: *Principles of Optics* (3rd ed.), Pergamon Press, Oxford.

Bridge, N. J., and A. D. Buckingham (1966), The polarization of laser light scattered by gases, *Proc. Roy. Soc. London, A*, 295, 334-349.

Cohen, A. (1972), On the influence of wavelength, air-temperature and air-density variations on Mie scattering intensities, submitted to *Appl. Opt.* 11.

- Deirmendjian, D., 1969: *Electromagnetic Scattering on Spherical Polydispersions*. American Elsevier, New York.
- Dirac, P. A. M., 1958: *Quantum Mechanics* (4th ed.), Oxford Press, Clarendon.
- Ditchburn, R. W., 1963: *Light* (2nd ed.), Interscience (Wiley), New York.
- Fabelinskii, I. L., 1968: *Molecular Scattering of Light*. Plenum Press, New York.
- Fouche, D. G., and R. K. Chang (1971), Relative Raman Cross Section for N₂, O₂, CO, CO₂, SO₂ and H₂S, *Appl. Phys., Letters* 18, 579-580.
- Garlick, G. F. J., 1958: *Luminescence*, in *Handbuch der Physik*, Bd. XXVI. Springer-Verlag, Berlin.
- Heitler, W., 1954: *The Quantum Theory of Radiation*. Oxford, London.
- Holland, A. C., and G. Gagne (1970), The scattering of polarized light by polydisperse systems of irregular particles, *Appl. Opt.* 9, 1113-1121.
- Holzer, W., W. F. Murphy and H. J. Bernstein (1970), Resonance Raman Effect and Resonance Fluorescence in Halogen Gases, *J. Chem. Phys.*, 52, 339-407.
- van de Hulst, H. C., 1957: *Light Scattering by Small Particles*. John Wiley and Sons, New York.
- Jorna, S. (1971), Atmospheric Depolarization and Stimulated Brillouin Scattering, *Appl. Opt.*, 10, 2661-2664.
- Kerker, M., 1969: *The Scattering of Light*. Academic Press, New York.
- Mandl, F., 1957: *Quantum Mechanics* (2nd ed.), Butterworths, London (and Academic Press, New York).
- Mountain, R. D. (1966), Spectral Distribution of Scattered Light in a Simple Fluid, *Rev. Mod. Phys.*, 38, 205-214.
- Placzek, G., 1934: *The Rayleigh and Raman Scattering*. in E. Marx (ed.): *Handbuch der Radiologie*. Akademische Verlagsgesellschaft VI, 2, 209-374. Translated by A. Werbin UCRL, Translation No. 526(L). Available from the Atomic Energy Commission, Division of Technical Information.
- Porto, S. P. S. (1966), Angular Dependence and Depolarization Ratio of the Raman Effect, *J. Opt. Soc. Am.*, 56, 1585-1589.
- Science Spectrum, Inc., 1970: *Atlas of Light Scattering Curves*. Science Spectrum, Santa Barbara, California.
- Strong, J., 1958: *Concepts of Classical Optics*. Freeman and Co., San Francisco.
- Victor, G. A., J. C. Brune and A. Dalgarno (1967), Optical Properties of Molecular Hydrogen, *Proc. Phys. Soc.*, 92, 42-49.
- Yoshino, T., and H. J. Bernstein (1958), Intensity in the Raman Effect, *J. Mol. Spect.*, 2, 213-240.

List of Symbols

\mathbf{r}	classical position vector and quantum position operator	\mathbf{A}	classical vector potential and quantum vector potential operator
X	a component of \mathbf{r}	\mathbf{k}	photon wave vector
$\psi(\mathbf{r})$	a wave function or state vector for a system of material particles	k	$ \mathbf{k} $, the wave vector magnitude
$\psi^*(\mathbf{r})$	complex conjugate of state vector	$\nu_{\mathbf{k}}$	photon frequency associated with wave vector \mathbf{k}
ψ_i	i th eigenvector of an operator (stationary state)	$A_{\mathbf{k}}$	k th term in discrete series decomposition of \mathbf{A} in \mathbf{k} space
\mathbf{A}	matrix operator	$\rho_{\mathbf{k}}$	arbitrary density of points in \mathbf{k} series space
a_i	i th eigenvalue of operator \mathbf{A}	$\Phi^{\mathbf{D}}$	photon eigenvector in discrete \mathbf{k} series space
c_n	n th expansion coefficient of ψ in terms of the ψ_n	$\rho_{\mathbf{p}}$	arbitrary density of points in \mathbf{p} series space (equivalent to \mathbf{k} space by transformation)
$\langle A \rangle$	expected value of physical measurement corresponding to operator \mathbf{A}	ΔE	difference in energy between two stationary states
ψ	time independent factor in ψ	$\eta_{\mathbf{k}\ell}$	number operator corresponding to the k th term in \mathbf{k} series space and to the ℓ th polarization component
E_n	energy of n th stationary state of the system (n th eigenvalue of the total energy operator)	$n_{\mathbf{k}\ell}$	eigenvalue of $\eta_{\mathbf{k}\ell}$ giving the occupation number of the ensemble
\mathbf{H}	Hamiltonian or total energy operator	$n_{\mathbf{k}\ell}$	eigenvalue of $\eta_{\mathbf{k}\ell}$ giving the occupation number of the ensemble of simple harmonic oscillators representing the k th term in the \mathbf{k} series expansion of the photon field.
t	time	ℓ_i	unit vector in direction of polarization of the i th photon state
h	Planck's constant	ν_{fi}	$\nu_f - \nu_i$, frequency difference between stationary states f and i
\hbar	Planck's constant divided by 2π	\dot{x}	time derivative of position operator \mathbf{x}
i	$\sqrt{-1}$ and, as a subscript, initial state	$F_i(t)$	various time dependent factors in matrix element
f	final state subscript	W	total energy of a representative oscillator system
g	intermediate state subscript	τ	fixed time value and lifetime of an excited state and optical depth
δ_{ij}	Dirac delta function	θ, ϕ, r	polar coordinate angles and distance
H_c	interaction Hamiltonian	Q	pressure quenching probability
$P(i-f)$	probability of a transition from a stationary state i to state f	p	total system pressure
H_e	Hamiltonian operator term acting only on the electronic subsystem	T	system temperature
H_r	Hamiltonian operator term acting only on the radiation (photon) subsystem	n_i	number density of the i th gas
Φ_q	q th eigenvector for an operator acting on a photon system	σ	scattering cross section (area)
ϕ_q	time independent factor in Φ_q	Ω	solid angle in steradians
ψ_{nq}	product of electronic and photon eigenvectors		
e	electron charge		
m	electron mass		
c	speed of light		
\mathbf{p}	classical vector momentum and quantum momentum operator		

$p(t)$	probability of a system being in a specified state at time t	a_n, b_n	Mie series expansion coefficients
γ_j	damping constant for decay of $p(t)$ for state j which is the total spontaneous transition probability per unit time of state j	π_n, τ_n	Mie angular scattering functions
		$i_{ }, i_{\perp}$	Mie intensity functions
		$I_{ }, I_{\perp}$	Mie scattered intensity
Δt	$1/\gamma$, the mean lifetime of an excited state	c, ν	Junge aerosol size distribution parameters
γ_{jk}	$\gamma_j + \gamma_k$	γ	Diermendjian aerosol size distribution parameter
w	full width at half maximum of a monochromatic source	$\Delta \nu_D$	Doppler frequency shift
I_0	Incident laser power	\mathbf{K}	acoustic wave vector
$R(\nu_0 - \nu, g_i)$	resonance factor for photon absorption		
H_c^1, H_c^2	first and second terms, respectively, of the interaction Hamiltonian		
δ_g	damping term for resonance, related to absorption by the state g		
I	scattered intensity (power)		
$A, B,$	intensity constants for two photon scattering		
ρ	depolarization ratio for quantum scattering and Mie scattering parameter kr		
n	refractive index of a medium		
N	number density of scatterers		
$\tilde{\nu}$	wavenumber of a photon (spatial frequency or reciprocal wavelength)		
D_i	Debye's scalar potential		
∇^2	Laplacian operator, $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$		
ϵ, μ, σ	dielectric constant, magnetic permeability, specific conductance of a medium		
$J(\rho), N(\rho)$	Bessel and Neumann functions, respectively		
$p_{\ell}^{(m)}$	associated Legendre functions		
$\psi_{\ell}(\rho),$			
$\chi_{\ell}(\rho)$	combinations of $J(\rho), N(\rho)$		
$\xi_{\ell}(\rho)$	combinations of $\psi_{\ell}(\rho), \chi_{\ell}(\rho)$		
\mathbf{E}, \mathbf{H}	electric and magnetic field vectors		
a	radius of spherical scattering particle		
λ	wavelength of radiation in the surrounding medium		
α	$2\pi a/\lambda$, size parameter for a scatterer and Diermendjian aerosol size distribution parameter		
m	relative refractive index, $n(\text{particle})/n(\text{medium})$		
β	$m\alpha$		

Chapter 11 PROPAGATION AND SCATTERING IN RANDOM MEDIA

Steven F. Clifford

Wave Propagation Laboratory

Environmental Research Laboratories

National Oceanic and Atmospheric Administration

The clear atmosphere is not a homogeneous medium with respect to wind, temperature, pressure, and humidity. There are both temporal and spatial variations of these quantities that will scatter and degrade electromagnetic and acoustic signals. Small-scale random fluctuations produce perturbations in the spatial refractive-index field that are extremely useful as tracers of large-scale atmospheric processes. This chapter develops a statistical characterization of the fluctuations of these quantities and a mathematical model of their interactions with EM and acoustic waves.

11.0 Introduction

The interaction of wave-like motions with the atmosphere occurs over many decades of wave-numbers. Earlier chapters have primarily concerned themselves with large-scale, internally generated motions and how they arise, transport energy and momentum, and decay; usually resulting in the turbulent mixing of background constituents, such as wind, temperature and humidity. By contrast, relatively short-wavelength, externally-generated waves are used to actually probe and continuously monitor the atmosphere's current state. Examples considered in this text are Doppler Radar, Chapter 14; Acoustic Echo Sounding, Chapters 18 and 19; FM-CW Radar, Chapter 20; and Raman Lidar, Chapter 23. These techniques are sensitive to such small-scale refractive phenomena (on the order of a wavelength) with such short interaction times that most large-scale, slow variations of atmospheric parameters have no effect. (An important exception are thin atmospheric layers having strong refractive-index gradients over distances comparable to a wavelength of the incident radiation (see Chapter 8). For the temporal and spatial scales of importance the atmosphere is essentially turbulent and may be suitably modeled as a random process. Deterministic variations such as temperature stratifications or small scale periodicities will be ignored.

The purpose of this chapter is both to illustrate the limiting effects of small-scale turbulent mixing on these techniques as well as to prepare the theoretical groundwork for the use of refractive turbulence as a tracer of atmospheric motions. The emphasis of this chapter will be on the latter, leaving the physical interpretation of the results to the chapters dealing with applications, i.e., Chapters 14, 18 and 25.

11.1 Scattering of Electromagnetic Waves

11.1.1 Maxwell's Equations

The starting point of any discussion of the propagation of electromagnetic waves in continuous media is Maxwell's equations. The form of these equations varies widely depending on the system of units and the generality of the problem (consult Ott (8.1) for a general discussion.) The atmosphere is essentially a gas mixture with zero conductivity and unit magnetic permeability, whose refractive index, n , varies as a function of position, (x, y, z) , and time, t . For EM waves n is a function of temperature, pressure, and humidity. The assumption of "frozen turbulence" discussed below, is a device that permits suppression of the time dependence of the refractivity field and results in a considerable simplification of the equations. If, in addition, we assume that the time dependence of the electric field, \underline{E} , and the magnetic field, \underline{H} , is sinusoidal, i.e.,

$$\underline{E}(\underline{r}, t) = \underline{E}(\underline{r})e^{-i\omega t} \quad \text{and} \quad \underline{H}(\underline{r}, t) = \underline{H}(\underline{r})e^{-i\omega t} \quad ,$$

Maxwell's equations take the form

$$\nabla \cdot \underline{\underline{H}} = 0 \quad (11:1)$$

$$\nabla \times \underline{\underline{H}} = -ik n^2 \underline{\underline{E}} \quad (11:2)$$

$$\nabla \cdot (n^2 \underline{\underline{E}}) = 0 \quad (11:3)$$

$$\nabla \times \underline{\underline{E}} = ik \underline{\underline{H}} \quad (11:4)$$

where k is the wavenumber of the radiation, defined by $k = \omega/c$, ω is the radian frequency of the wave and c the speed of light in free space.

Substitution of (11:2) into the curl of (11:4) results in an equation for the electric field in the form

$$\nabla^2 \underline{\underline{E}} + k^2 n^2 \underline{\underline{E}} - \nabla(\nabla \cdot \underline{\underline{E}}) = 0 \quad (11:5)$$

Substituting $\nabla \cdot \underline{\underline{E}}$ from (11:3) into (11:5), we have

$$\nabla^2 \underline{\underline{E}} + k^2 n^2 \underline{\underline{E}} + 2\nabla [\underline{\underline{E}} \cdot \nabla \log(n)] = 0 \quad (11:6)$$

The solution of (11:6) completely determines the characteristics of electromagnetic waves in a random medium since the complementary magnetic field is from (11:4), a function of $\underline{\underline{E}}$ alone. Therefore, it is sufficient to study the characteristics of the electric field vector.

At this point it becomes necessary to define the two problems being considered in this chapter. Propagation and scattering effects are distinguished primarily by the relative position of the source and receiver. Scattering usually implies that, in the absence of a mechanism that deflects some of the incident signal toward the receiver, no signal will be received. On the other hand, propagation effects, or, more explicitly line-of-sight propagation effects, involve a description of the focusing and defocusing action of the medium on the incident beam itself, producing fluctuations in the otherwise constant wave parameters such as amplitude or phase. The former problem usually considers how much power will be scattered to the receiver per unit volume of the scatterer whereas the latter considers the degradation of the received signal.

11.1.2 Statistical Description of Random Fields

A turbulent medium, such as the atmosphere for sufficiently small temporal and spatial scales, is a particular type of random field. (A random field is a random process with three independent variables.) In the same manner that stock prices fluctuate unpredictably from minute to minute on the stock exchange, the values of the temperature, humidity and wind fluctuate as their respective sensors are moved from point to point. Of course there are time changes of these quantities at a fixed position that are also random. These changes may be suppressed by considering them as different realizations of the random field. A realization is a set of measurements of the desired parameter as a function of the coordinates.

Yaglom (1962) has considered the calculus of random fields in general terms and Tatarski (1961) has an excellent discussion of how these results apply to meteorological measurements. Monographs by Batchelor (1953) and Lumley and Panofsky (1964) contain the physical and mathematical foundations of the study of homogeneous turbulence. These texts are detailed studies of material related to this section and are good reference books.

Because each realization of a random field differs substantially in detail from every other, there is very little that can be accomplished in the way of predicting the outcome of the next set of measurements from previous results. We must settle for a more meager description of its characteristics. The simplest description of a random process is its mean or expected value, i.e., for a random field, $f(\underline{r})$, the expected value, \underline{E} , is written

$$E [f(\underline{r})] = \langle f(\underline{r}) \rangle , \quad (11:7)$$

where the angle brackets indicate an ensemble average. The angle brackets operation may be written as

$$\langle f(\underline{r}) \rangle = \int_{-\infty}^{\infty} f(\underline{r}, \Omega) dP(\Omega) , \quad (11:8)$$

where $P(\Omega)$ is the probability distribution function. For continuous random variables the probability density, $p(\Omega)$, exists and (11.8) becomes

$$\langle f(\underline{r}) \rangle = \int_{-\infty}^{\infty} f(\underline{r}, \Omega) p(\Omega) d\Omega \quad (11:9)$$

where of course

$$p(\Omega) = \frac{dP(\Omega)}{d\Omega} \quad (11:10)$$

Equation (11:9) contains the ensemble parameter, Ω , which is essentially a running index labelling identical, repeated experiments. Hence the ensemble average defined in (11:9) is merely the sum of all possible outcomes of measurements of f at the position, \underline{r} , weighted by the probability of occurrence, $p(\Omega)d\Omega$. There is some difficulty in applying (11:9), as it stands, to the atmosphere. It implies that we have at our disposal an infinite number of atmospheres with identical characteristics that we may observe simultaneously to compute our desired average. Clearly, we must compromise. At this point we invoke the concept of ergodicity. For our purposes ergodicity implies that the time variations of the quantity $f(\underline{r})$, will be such that average values of functions of f over all time will result in the same number as the ensemble average. Again, we must compromise. It is impossible to make measurements for all time. How long to average and what filtering effects the averaging time, record length, and measurement technique have on the data are the subject of an entire area of research, time series analysis. Some of the early work in this field is by Wiener (1949). A concise summary is available in the work of Blackman and Tukey (1958) and Pasquill (1961) contains a well-written description of the problem as it applies to meteorological variables. It is important to be aware of these problems. They are not academic exercises but rather are the everyday concern of experimental atmospheric physicists. In discussions below we apply the ensemble average operation without any further qualification. Recall the reservations expressed above and refer to the references for amplification.

For the problem discussed in this chapter the random field of most interest is the index of refraction, $n(\underline{r})$. We restrict realizations of this function by assuming homogeneity and isotropy, which imply respectively, that the expected values of all moments of $n(\underline{r})$ do not change with translation of the origin or variations in the orientation of the system of coordinates. This assumption is quite restrictive physically since the atmosphere contains large scale events that are not isotropic, or for that matter, not even turbulent. Consider, for example, the background heating of the atmosphere during the morning hours, the sudden cooling of kilometers of air as a cloud passes overhead, etc. We would hope to remove these phenomena from our analysis of the problem of the propagation of waves. There is a technique,

discussed by Tatarski (1961, 1967), based on the use of structure functions, that effectively limits the assumption of homogeneity and isotropy to the spatial volumes of interest. This problem, although essential in modelling the atmosphere as a random field, is not critical in the development of the equations for propagation and scattering effects. As the development progresses we will note the changes introduced by the structure-function approach.

We consider the index of refraction to be composed of a mean plus a fluctuating part

$$n(\underline{r}) = \langle n(\underline{r}) \rangle + n_1(\underline{r}) \quad (11:11)$$

where $\langle n_1(\underline{r}) \rangle = 0$. A mathematical consequence of homogeneity is that $\langle n(\underline{r}) \rangle = \text{const}$. The covariance function, defined by

$$B_n(\underline{r}_1, \underline{r}_2) = \langle [n(\underline{r}_1) - \langle n(\underline{r}_1) \rangle] [n(\underline{r}_2) - \langle n(\underline{r}_2) \rangle] \rangle \quad (11:12)$$

after introducing (11:11) becomes

$$B_n(\underline{r}_1, \underline{r}_2) = \langle n_1(\underline{r}_1) n_1(\underline{r}_2) \rangle \quad (11:13)$$

The assumption of homogeneity forces the dependence, $B_n(\underline{r}_1, \underline{r}_2) = B_n(\underline{r}_1 - \underline{r}_2)$, and isotropy implies $B_n(\underline{r}_1, \underline{r}_2) = B_n(|\underline{r}_1 - \underline{r}_2|)$. There is nothing in the definition, (11:12), that restricts the size of refraction changes that could effect the determination of B_n . Thus the assumption $B_n(\underline{r}_1, \underline{r}_2) = B_n(|\underline{r}_1 - \underline{r}_2|)$, which is extremely useful to apply to the propagation equations, must be valid for all scales. This is usually violated by large scale events in the atmosphere as discussed above. To remedy this problem, we invoke the structure function, which is defined by

$$D_n(\underline{r}_1, \underline{r}_2) = \langle [n(\underline{r}_1) - n(\underline{r}_2)]^2 \rangle \quad (11:14)$$

Note that effects of refractive phenomena of size larger than the distance, $|\underline{r}_1 - \underline{r}_2|$, will be attenuated by the difference operation. The idea of locally homogeneous and isotropic turbulence arises from this definition, i.e., if we assume using (11:11) that

$$D_n(\underline{r}_1, \underline{r}_2) = \langle [n_1(\underline{r}_1) - n_1(\underline{r}_2)]^2 \rangle = D_n(|\underline{r}_1 - \underline{r}_2|) \quad (11:15)$$

we need only require that the relatively small scale eddies of size less than $|\underline{r}_1 - \underline{r}_2|$ be modeled by a homogeneous and isotropic turbulence model, certainly a more realistic assumption.

The results of the study of waves in turbulence to follow will be expressed in terms of either the covariance function, $B_n(\underline{r})$, or its Fourier transform the three-dimensional refractivity spectrum, $\Phi_n(\underline{K})$, where

$$\Phi_n(\underline{K}) = \frac{1}{(2\pi)^3} \int B_n(\underline{r}) e^{-i\underline{K} \cdot \underline{r}} d^3\underline{r} \quad (11:16)$$

and the inverse

$$B_n(\underline{r}) = \int \Phi_n(\underline{K}) e^{i\underline{K} \cdot \underline{r}} d^3K \tag{11:17}$$

(When integrals appear with no limits as in (11:16) and (11:17), integration from $-\infty$ to $+\infty$ is implied.) The parameter, $\underline{K} = (K_x, K_y, K_z)$ is the three-dimensional spatial wave number that plays the analogous role to frequency in the time-frequency transform. With the additional assumption of isotropy, i.e., $\Phi_n(\underline{K}) = \Phi_n(K)$ and $B_n(\underline{r}) = B_n(r)$, the integral over angles in (11:16) and (11:17) is easily performed to yield

$$\Phi_n(K) = \frac{1}{2\pi^2 K} \int_0^\infty B_n(r) \sin(Kr) r dr \tag{11:18}$$

and

$$B_n(r) = \frac{4\pi}{r} \int_0^\infty \Phi_n(K) \sin(Kr) K dK \tag{11:19}$$

where $K = |\underline{K}|$ and $r = |\underline{r}|$. There are other useful transforms that are derived from (11:16) and (11:17). These will be given as needed in the text.

To recapitulate, $\langle n(\underline{r}) \rangle$, $B_n(r)$ or $\Phi_n(K)$ are the parameters of our model of the statistics of the refractive index field. For first order linear theory they are sufficient to describe the statistical properties of EM and acoustic waves propagating through turbulence. There are many refinements to this description such as structure functions, many-point covariance functions, etc., that give a more complete knowledge of the medium.

11.1.3 Mean Intensity of Scattering and Scattering Cross Section

Equation (11:6) contains the description of the evolution of the field vector, \underline{E} , in a medium with arbitrary variations of refractive index. This equation is assumed to apply inside a volume, V , infinitely remote from the source of the incident radiation. This then allows the assumption of an incident plane wave when we calculate the scattered power (F11.1). If we assume that the variations are small, we may apply a perturbation procedure to (11:6) and calculate the field scattered by the irregularities in V . (The following discussions rely heavily on the developments in Chapter 4 and 7 of Tatarski, 1961.) The procedure is to assume that the refractivity field may be written as

$$n(\underline{r}) = 1 + n_1(\underline{r}) \tag{11:20}$$

where $|n_1| \ll 1$ and the mean index of the atmosphere is unity. This allows the assumption that the scattered field may be expanded as a small perturbation of the initial field, i.e.,

$$\underline{E} = \underline{E}_0 + \underline{E}_1 + \underline{E}_2 + \dots \tag{11:21}$$

where the i th term of the series is of the order $(n_1)^i$. Therefore, if the series converges rapidly, only a finite number of terms is required to describe the scattered field. (The convergence of the series is an important question. In fact, in the limit of strong refractive index perturbations for optical beams, the series does not converge, leading to some non-linear effects as discussed by Strohbehn (1968) and Tatarski (1967).) Substituting (11:20) and (11:21) into (11:6) and equating to zero all terms that are the same

order of smallness, $(n_1)^i$, we obtain a hierarchy of equations describing progressively weaker perturbations of the field, i.e.,

$$\nabla^2 [\sum_i \underline{\underline{E}}_i] + k^2(1 + 2n_1 + n_1^2) \sum_i \underline{\underline{E}}_i + 2\nabla [\sum_i \underline{\underline{E}}_i \cdot \nabla \log(1 + n_1)] = 0 \quad (11:22)$$

using the expression $\log(1 + n_1) = -\sum_{i=1}^{\infty} (-n_1)^i / i$ and equating terms of the same order in $(n_1)^i$ we obtain

$$i = 0; \quad \nabla^2 \underline{\underline{E}}_0 + k^2 \underline{\underline{E}}_0 = 0 \quad (11:23)$$

$$i = 1; \quad \nabla^2 \underline{\underline{E}}_1 + k^2 \underline{\underline{E}}_1 + 2k^2 n_1 \underline{\underline{E}}_0 + 2\nabla(\underline{\underline{E}}_0 \cdot \nabla n_1) = 0 \quad (11:24)$$

$$i = m; \quad \nabla^2 \underline{\underline{E}}_m + k^2 \underline{\underline{E}}_m + 2k^2 n_1 \underline{\underline{E}}_{m-1} + k^2 n_1^2 \underline{\underline{E}}_{m-2} + 2\nabla [\sum_{j=0}^{m-1} (-n_1)^j \underline{\underline{E}}_{m-j-1} \cdot \nabla n_1] = 0 \quad (11:25)$$

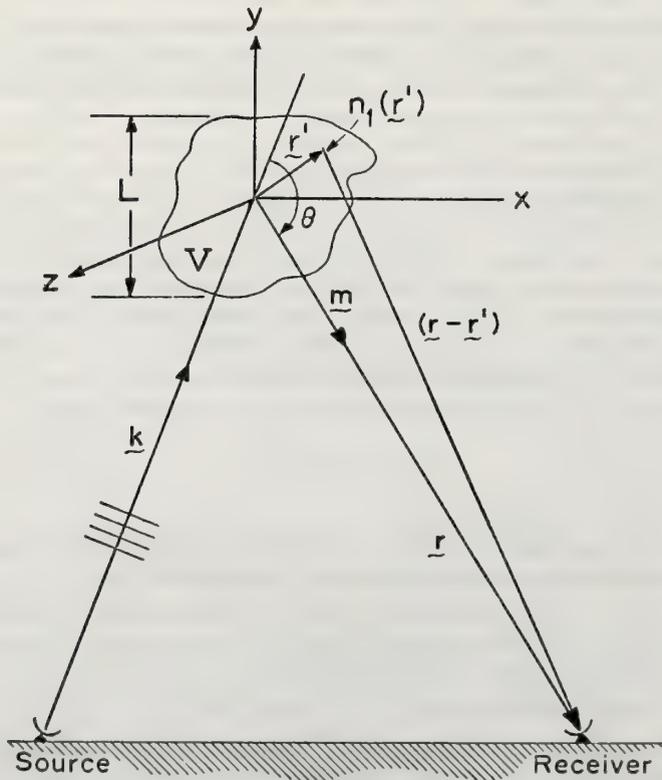


Figure 11.1 Scattering geometry for radiation of wavelength, $2\pi/k$, incident on a volume, V , containing refractive-index fluctuations, $n_1(\underline{r}')$. Energy is received at an angle, θ , with respect to the incident wave vector.

for the m th order perturbation. For most EM wavelengths n_1 is so small that the first order field scattered from these weak irregularities is sufficient to describe the received signal. Therefore, we concern ourselves with the zero and first order solutions, (11:23) and (11:24), exclusively. The higher order terms may be calculated analogously but with large increases in difficulty.

The first order field, or equivalently the scattered field, satisfies (11:24). The solution of this equation in terms of Green's functions is well known (8)

$$\tilde{E}_1(\underline{r}) = \frac{1}{4\pi} \int_V \frac{e^{ik|\underline{r}-\underline{r}'|}}{|\underline{r}-\underline{r}'|} [2k^2 n_1 \tilde{E}_0 + 2\nabla(\tilde{E}_0 \cdot \nabla n_1)] d^3 \underline{r}' \quad (11:26)$$

where the integral proceeds over the position vector \underline{r}' that describes a set of points inside the volume, V . All scattering takes place inside V and we observe the field, E_1 , at position \underline{r} external to that volume.

At this point in the development a key assumption greatly simplifies further derivation. This assumption restricts the position of the observer to be far from the limits of the scattering volume. To understand the assumption, return to (11:26). This equation expresses the fact that the observed scattered wave is a sum of spherical waves $[|\underline{r}-\underline{r}'|]^{-1} \exp[ik|\underline{r}-\underline{r}'|]$ originating at each position in the scattering volume, \underline{r}' , attenuated by the distance from point of origin to the receiver, $|\underline{r}-\underline{r}'|^{-1}$, and whose amplitude depends on the strength and gradient of the refractivity fluctuations, $n_1(\underline{r}')$, at their points of origin. (F11.1). We sum all these spherical waves (Huygen's wavelets) at the observation point, \underline{r} . We now invoke the so-called far-field assumption, which in essence says that wavelets originating from the most extreme points of the scattering volume and arriving at position \underline{r} , the origin of coordinates being chosen inside the scattering volume, must not be able to interfere. This of course implies that there be much less than a $\lambda/2$ difference in phase path to the receiver for these most extreme waves. If the scattering volume has a maximum dimension, L , this assumption may be written as $\sqrt{r^2 + L^2} - r \ll \lambda/2$, or

$$r \gg L^2/\lambda \quad (11:27)$$

The additional weak assumption $r \gg \lambda$, restricting our position to the wave zone, is also required. Tatarski (1967) shows that this rather strong restriction on r , (11:27) is not necessary to the further development of (11:26). We retain this assumption for tutorial value since it greatly simplifies the physical interpretation of the intermediate steps in the derivation. Expanding $|\underline{r}-\underline{r}'|$, we obtain

$$|\underline{r}-\underline{r}'| \cong r - \underline{m} \cdot \underline{r}' + \frac{1}{2r} [r'^2 - (\underline{m} \cdot \underline{r}')^2] \quad (11:28)$$

where \underline{m} is a unit vector parallel to \underline{r} , (F11.1). Making the assumption of (11:27), it is sufficient to retain only the first two terms in the phase of the Green's function, i.e.,

$$\exp [ik|\underline{r}-\underline{r}'|] = \exp [ik(r - \underline{m} \cdot \underline{r}')] \quad (11:29)$$

Clearly, we may also write $|\underline{r}-\underline{r}'| \sim r$ in the denominator. If we also assume a linearly-polarized, constant amplitude, infinite plane wave for the incident wave, $\tilde{E}_0 = \underline{A}_0 e^{ik \cdot \underline{r}}$, noting that it satisfies (11:23), then (11:26) assumes the following form

$$E_1(\underline{r}) = \frac{k^2}{2\pi} \frac{e^{ikr}}{r} \left[\int_V n_1(\underline{r}') \underline{A}_0 e^{i(\underline{k} - k\underline{m}) \cdot \underline{r}'} d^3 \underline{r}' + \frac{1}{k^2} \int_V \nabla [e^{i\underline{k} \cdot \underline{r}'} \underline{A}_0 \cdot \nabla n_1(\underline{r}')] e^{-ik\underline{m} \cdot \underline{r}'} d^3 \underline{r}' \right] \quad (11:30)$$

Following Tatarski (1961), (11:30) may be simplified by using Gauss' theorem on the second term in the form

$$\int_V \Psi \nabla \phi \, dV = \int_S \phi \Psi \, d\sigma - \int_V \phi \nabla \Psi \, dV . \quad (11:31)$$

Discarding the surface integral, since the value of ϕ on the surface can be reduced to zero by choosing the surface external to the turbulent scattering region, we note that the remaining volume integral is a purely longitudinal field (parallel to \underline{m}) and, of course, does not contribute to the flow of scattered energy. Therefore, when calculating the mean scattering intensity and cross section we merely ignore the last term and the resulting scattered field is

$$\underline{E}_1(\underline{r}) = \frac{k^2}{2\pi} \frac{e^{ikr}}{r} \int_V n_1(\underline{r}') \underline{A}_0 e^{i(\underline{k} - \underline{k}\underline{m}) \cdot \underline{r}'} \, d^3\underline{r}' . \quad (11:32)$$

To calculate the magnitude and direction of the flow of scattered power, recall the definition of the Poynting vector.

$$\underline{S} = (c/8\pi) \text{Re} [\underline{E}_1 \times \underline{H}_1^*] \quad (11:33)$$

where c is the wave phase velocity and \underline{H}_1 is the scattered magnetic field given by substituting (11:32) into (11:4)

$$\underline{H}_1 = \frac{-i}{k} [\nabla \times \underline{E}_1] = \underline{E}_1(\underline{r}) [\underline{m} \times \underline{A}_0] / A_0 . \quad (11:34)$$

To obtain (11:34) neglect the rapidly decreasing \underline{r}^{-2} term compared with the \underline{r}^{-1} term. Insertion of (11:34) into (11:33), with the aid of the vector identity

$$\underline{A}_0 \times (\underline{m} \times \underline{A}_0) = [\underline{m}(A_0^2) - \underline{A}_0(\underline{m} \cdot \underline{A}_0)] , \quad (11:35)$$

reduces the calculation of \underline{S} to the form

$$\underline{S} = \frac{c[E_1(\underline{r}) E_1^*(\underline{r})]}{8\pi A_0^2} [\underline{m} A_0^2 - \underline{A}_0(\underline{m} \cdot \underline{A}_0)] . \quad (11:36)$$

We are interested in the flow of energy in the direction of the observation point, $S_m = \underline{S} \cdot \underline{m}$, which from (11:36) is

$$S_m = \frac{c \sin^2 \chi}{8\pi} [E_1(\underline{r}) E_1^*(\underline{r})] \quad (11:37)$$

where χ is the angle between \underline{A}_0 and \underline{m} .

The quantity S_m is random. We are primarily interested in the average power flow, $\langle S_m \rangle$, but also recall that this function has higher statistical moments such as its variance that may be of interest in some applications, e.g., for predicting dynamic range restrictions on equipment. The averaging operation, after substituting (11:32), commutes with the integration in (11:37) and we obtain

$$\langle S_m \rangle = \frac{ck^4 A_0^2 \sin^2 \chi}{32 \pi^3 r^2} \int_V \int_V B_n(\underline{r}_1 - \underline{r}_2) e^{i(\underline{k} - \underline{km}) \cdot (\underline{r}_1 - \underline{r}_2)} d^3 \underline{r}_1 d^3 \underline{r}_2 \quad (11:38)$$

where $B_n(\underline{r}_1 - \underline{r}_2)$ is defined by (11:13) with the additional assumption of spatial homogeneity. A change of variables to $2\underline{\rho} = \underline{r}_1 + \underline{r}_2$ and $\underline{r}' = \underline{r}_1 - \underline{r}_2$, and performance of the $\underline{\rho}$ integration, resulting in the volume V appearing outside the integral, reduces (11:38) to the form

$$\langle S_m \rangle = \frac{cVk^4 A_0^2 \sin^2 \chi}{32 \pi^3 r^2} \int_V B_n(\underline{r}') e^{i(\underline{k} - \underline{km}) \cdot \underline{r}'} d^3 \underline{r}' \quad (11:39)$$

The integral featured in (11:39) in the limit as $V \rightarrow \infty$ is identical to (11:16), the definition of the refractivity spectrum. In the case of finite V , instead of having a single spatial frequency of the turbulence, $\underline{K}_0 = \underline{k} - \underline{km}$, reinforcing the ray in the direction of the receiver, spatial frequencies close to \underline{K}_0 will also contribute. (A truncated sinusoid contains a spread of frequencies about the fundamental.) Tatarski (1967) considers this effect in detail and deems it negligible. The essence of the scattering process may be understood by making no distinction between the integral in (11:39) and (11:16). Noting that $\Phi_n(-\underline{K}) = \Phi_n(\underline{K})$ a consequence of the obvious symmetry of the covariance function, i.e., $B_n(\underline{r}_1 - \underline{r}_2) = \langle n_1(\underline{r}_1) n_1(\underline{r}_2) \rangle = B_n(\underline{r}_2 - \underline{r}_1)$, and noting the definition (11:16), (11:39) finally becomes

$$\langle S_m \rangle = \frac{cVk^4 A_0^2 \sin^2 \chi}{4r^2} \Phi_n(\underline{k} - \underline{km}) \quad (11:40)$$

Equation (11:40) implies that the scattering process singles out one spatial frequency of the turbulence that satisfies the Bragg reflection condition, i.e., reinforces the scattered wave in the same manner as a spatial diffraction grating.

Following the usual practice we define the effective scattering cross section from the mean intensity of scattering given by (11:40). The differential cross section, $d\sigma$, is defined as the fraction of the incident energy flux per unit area per unit time that is scattered into the solid angle, $d\Omega$, about the direction \underline{m} . The incident energy flux is the magnitude of the time-averaged Poynting vector of the unperturbed wave

$$|\underline{S}_0| = \frac{c}{8\pi} |(\underline{E}_0 \times \underline{H}_0^*)| = \frac{c A_0^2}{8\pi} \quad (11:41)$$

Therefore, the scattering cross section is

$$d\sigma = \frac{\langle S_m \rangle r^2 d\Omega}{\frac{c A_0^2}{8\pi}} = 2\pi k^4 V \sin^2 \chi \Phi_n(\underline{k} - \underline{km}) d\Omega \quad (11:42)$$

11.1.4 Applications—The Kolmogorov Spectrum

All that remains in the determination of the scattering characteristics of homogenous turbulence is the insertion of a physically reasonable refractivity spectrum in (11:42). The most useful statistical model of atmospheric velocity turbulence is due to Kolmogorov (1941). From a dimensional analysis argument, he determined that the structure function of the velocity field should have a universal $r^{2/3}$ dependence as the spacing, r , between two sensors is varied between some lower limit, ℓ_0 , and some upper limit, L_0 . The scale

sizes of the turbulence that lie in this interval make up the inertial subrange. The sizes of the inner scale, ℓ_0 , and the outer scale, L_0 are fixed by two considerations. The inner scale determines the size of velocity fluctuations that are so small that viscous effects are important. The energy in the eddy, instead of being entirely transferred to smaller eddies, as is assumed for larger scales in the model, will be dissipated in the form of heat. The outer scale is roughly the largest length scale for which the assumption of homogeneity and isotropy hold. Both ℓ_0 and L_0 are functions of the height above the ground and the strength of turbulence. In the atmosphere's boundary layer, ℓ_0 is on the order of millimeters to centimeters and L_0 on the order of one to hundreds of meters.

The assumption that temperature fluctuations are caused by velocity fluctuations and are conservative passive additives, i.e., that in changing position they retain their temperature difference from the background, allows extending the Kolmogorov results to temperature and humidity fluctuations and ultimately to refractivity. A detailed account of this development is found in Tatarski (1961). Here we assume this result and write

$$D_n(r) = C_n^2 r^{2/3}, \quad \ell_0 \ll r \ll L_0 \quad (11:43)$$

C_n^2 is a measure of the "intensity" of the refractive-index variations. From the definitions (11:13) and (11:14), we note for homogeneous and isotropic turbulence that

$$D_n(r) = 2 [B_n(0) - B_n(r)] \quad (11:44)$$

Using this result, the definition (11:18) and simplifying, we find that $D_n(r)$ is related to $\Phi_n(K)$ by the expression

$$\Phi_n(K) = \frac{1}{4\pi^2 K^2} \int_0^\infty \frac{\sin(Kr)}{Kr} \frac{d}{dr} \left[r^2 \frac{dD_n(r)}{dr} \right] dr \quad (11:45)$$

If we substitute (11:43) into (11:45), we obtain

$$\Phi_n(K) = 0.033 C_n^2 K^{-1/3}; \quad L_0^{-1} \ll K \ll \ell_0^{-1} \quad (11:46)$$

Returning to the general bistatic geometry of (F11.1), we note that $|\underline{k} - \underline{k}'| = 2k \sin(\theta/2)$, where θ is the scattering angle measured from the direction of the incident wave to the direction of the scattered wave. The insertion of the spectrum, (11:46) evaluated at $2k \sin(\theta/2)$, into (11:42) results in the scattering cross section in its final form

$$d\sigma(\theta) = 0.016 V C_n^2 k^{1/3} \sin^2 \chi [\sin(\theta/2)]^{-1/3} d\Omega \quad (11:47)$$

for $L_0^{-1} \ll 2k \sin(\theta/2) \ll \ell_0^{-1}$.

Equation (11:47) indicates that the strength of scattering of electromagnetic waves is proportional to the volume and strength of the refractivity fluctuations in the scattering medium, is weakly dependent on wavelength, $\lambda^{-1/3}$, and strongly peaked in the forward scatter, $\theta=0$ direction. The function $\sin^2 \chi$ is the effect of polarization. Note that the restriction on the solutions, precludes using (11:47) for certain angles, e.g., forward scatter, $\theta=0$.

11.2 Scattering of Acoustic Waves

11.2.1 The Wave Equation for Sound

The scattering of sound waves from atmospheric turbulence is entirely analogous to the previous EM problem. From the equations of hydrodynamics we can generate a perturbed Helmholtz' equation analogous to (11:24) [Monin (1962), Clifford and Brown (1970)] in the form

$$\nabla^2 P_1 + k^2 P_1 + 2k^2 n_1 P_0 - 2 ik P_0 \frac{\partial n_1}{\partial x} = 0 . \quad (11:48)$$

P_1 is the acoustic field of wavelength, $\lambda=2\pi/k$, scattered by the interaction of the incident field, $P_0 = A_0 e^{ikx}$, with the refractivity fluctuations, n_1 . The fourth term is analogous to the EM polarization term. Since the acoustic wave is essentially longitudinally polarized, there will be interaction between wave motions parallel to gradients of refractivity. P_0 is assumed in (11:48) to be travelling in the x direction.

11.2.2 The Effective Index of Refraction

The development of (11:48) involves the definition of the effective refractive-index fluctuation

$$n_1 = -\left(\frac{u'_x}{c_0} + \frac{1}{2} \frac{T'}{T_0}\right) , \quad (11:49)$$

where c_0 is the mean speed of sound in air, u'_x is the perturbation velocity of the zero-mean, background flow, and T' is the temperature fluctuation from its average value, T_0 . The minus sign arises from the fact that a positive T' or a wind fluctuation parallel to the direction of the wave phase velocity, reduce the refractivity.

Although the definition (11:49) derives from the general equations of hydrodynamics, a quite simple, physical model produces the same result. The refractive index, n , of any wave is defined as the ratio of its phase velocity in some reference medium (still, dry air for acoustic waves, free space for EM waves) to that velocity currently observed. Assuming c_0 to be the speed of sound in air, we have

$$n = c_0/c . \quad (11:50)$$

If the medium is in motion, the acoustic wave index becomes

$$n = c_0 \left[c + \frac{\underline{u} \cdot \underline{k}}{k} \right]^{-1} \quad (11:51)$$

where k is the wavenumber and \underline{u} is the velocity of the medium. Equation (11:51) simply states that the motion of the fluid carries the acoustic wave along and changes its phase speed in proportion to the fluid velocity parallel to the original propagation direction. If \underline{k} is x-directed, the term is simply u_x . If we assume that the ideal gas law holds, then the speed of sound is proportional to the square root of temperature. If the temperature of the air is $T = T_0 + T'$ relative to the reference temperature, T_0 , then for $T' \ll T_0$

$$c \sim c_0 \left[1 + \frac{T'}{2T_0} \right] \quad (11:52)$$

where c_0 is the phase velocity when $T = T_0$. Inserting this result in (11:51) with the additional assumption that $\langle u_x \rangle = 0$, and therefore, $u_x = u'_x$, we obtain

$$n = \left[1 + \frac{u'_x}{c_0} + \frac{T'}{2T_0} \right]^{-1} \quad (11:53)$$

Again assume that $n = 1 + n_1$, where $|n_1| \ll 1$, and expand $\left[1 + \left(\frac{u'_x}{c_0} + \frac{T'}{2T_0} \right) \right]^{-1}$ in a binomial series, we then obtain

$$n_1 = - \left(\frac{u'_x}{c_0} + \frac{T'}{2T_0} \right) \quad (11:54)$$

exactly as in (11:49). The only additional complication in the scattering theory occurs because the acoustic wave is sensitive to motions of the propagation medium, u'_x , where EM waves are not.

11.2.3 Average Energy Flux Density and Cross Section

The solution of (11:48), found by a Green's-function approach identical to that for the solution of (11:24) is

$$P_1(\underline{r}) = \frac{1}{4\pi} \int_V \frac{e^{ik|\underline{r}-\underline{r}'|}}{|\underline{r}-\underline{r}'|} \left[2k^2 n_1 P_0 - 2ik P_0 \frac{\partial n_1}{\partial x} \right] d^3 \underline{r}' \quad (11:55)$$

a form analogous to (11:26). Again making the far field assumption in (11:55) and following the steps from (11:26) to (11:30), we obtain

$$P_1(\underline{r}) = \frac{k^2 e^{ikr} A_0}{2\pi r} \int_V e^{i(\underline{k} - k\hat{m}) \cdot \underline{r}'} \left[n_1(\underline{r}') - \frac{i}{k} \frac{\partial n_1}{\partial x'} \right] d^3 \underline{r}' \quad (11:56)$$

where \underline{k} by assumption is parallel to the x axis and \hat{m} , as before is a unit vector in the direction of the scattered wave (F11.1).

We now define the energy flux density, \underline{S} , following Tatarski (1961), as

$$\underline{S} = \frac{\rho_0 k c_0}{2} \text{Im} [P_1^* \nabla P_1] \quad (11:57)$$

where c_0 is the mean speed of sound and ρ_0 the mean density of the atmosphere. Im means the "imaginary part of". Substituting in the proper forms of P_1 with $P_1 = (2\pi r)^{-1} k^2 A_0 e^{ikr} Q$, and ignoring r^{-2} terms in (11:57), since $kr \gg 1$, we obtain

$$\underline{S} = \frac{k^6 A_0^2 c_0 \rho_0}{8\pi^2 r^2} [Q Q^*] \hat{m} \quad (11:58)$$

We again find the average energy flux in the direction \underline{m} by following the same procedure and using the same assumptions as in going from (11:37) to (11:39). With Q from (11:56), we obtain

$$\langle Q Q^* \rangle = V \int_V B_{nn}(\underline{r}') e^{i(\underline{k} - k\underline{m}) \cdot \underline{r}'} d^3 \underline{r}' \quad (11:59)$$

where

$$B_{nn}(\underline{r}_1 - \underline{r}_2) = \langle [n_1(\underline{r}_1) - \frac{i}{k} \frac{\partial n_1(\underline{r}_1)}{\partial x_1}] [n_1(\underline{r}_2) + \frac{i}{k} \frac{\partial n_1(\underline{r}_2)}{\partial x_2}] \rangle \quad (11:60)$$

To the same order of approximation as in (11:40), we may write

$$\langle S_m \rangle = \frac{\pi c_0 V k^6 A_0^2 \rho_0}{r^2} \Phi_{nn}(\underline{k} - k\underline{m}) , \quad (11:61)$$

where $\Phi_{nn}(\underline{k} - k\underline{m})$ is the three-dimensional spectrum given by (11:60) substituted into (11:16). To determine Φ_{nn} , consider (11:60) in the form

$$B_{nn}(\underline{r}_1 - \underline{r}_2) = (1 - \frac{i}{k} \frac{\partial}{\partial x_1}) (1 + \frac{i}{k} \frac{\partial}{\partial x_2}) \left[\frac{B_{xx}(\underline{r}_1 - \underline{r}_2)}{c_0^2} + \frac{B_{TT}(\underline{r}_1 - \underline{r}_2)}{4 T_0^2} \right] \quad (11:62)$$

where from (11:13) and (11:49)

$$B_{xx}(\underline{r}_1 - \underline{r}_2) = \langle u'_x(\underline{r}_1) u'_x(\underline{r}_2) \rangle \quad (11:63)$$

and

$$B_{TT}(\underline{r}_1 - \underline{r}_2) = \langle T'(\underline{r}_1) T'(\underline{r}_2) \rangle . \quad (11:64)$$

To obtain (11:62) we have assumed that the temperature and velocity fluctuations are uncorrelated, i.e., $\langle u'_x(\underline{r}_1) T'(\underline{r}_2) \rangle = \langle u'_x(\underline{r}_2) T'(\underline{r}_1) \rangle = 0$. Tatarski demonstrates that the general form of $B_{j\ell}(\underline{r})$, ($j = \ell = 1, 2, 3$) is given by the equation

$$B_{j\ell}(\underline{r}) = \int e^{i\underline{K} \cdot \underline{r}} (\delta_{j\ell} - \frac{K_j K_\ell}{K^2}) E(\underline{K}) d^3 \underline{K} \quad (11:65)$$

where $E(\underline{K})$ is the spectral density of the energy and $\delta_{j\ell}$ is defined as

$$\delta_{j\ell} = \begin{cases} 1 & j = \ell \\ 0 & j \neq \ell \end{cases} \quad (11:66)$$

with $\delta_{ij} = 3$ and $K_i K_i = K^2$, by the summation convention. The peculiar form of the spectrum inside the integral, (11:65), is a result of the assumptions of local isotropy and incompressibility of the turbulent flow. Similarly

$$B_{TT}(\underline{r}) = \int e^{i\underline{K} \cdot \underline{r}} \Phi_{TT}(\underline{K}) d^3 \underline{K} \quad (11:67)$$

Substituting (11:65) and (11:67) into (11:62), we obtain

$$B_{nn}(\underline{r}_1 - \underline{r}_2) = \int e^{i\underline{K} \cdot (\underline{r}_1 - \underline{r}_2)} \left\{ \left(1 - \frac{\underline{K} \cdot \underline{k}}{k^2}\right)^2 \left[\left(1 - \frac{(\underline{K} \cdot \underline{k})^2}{(Kk)^2}\right) \frac{E(\underline{K})}{c_0^2} + \frac{\Phi_{TT}(\underline{K})}{4T_0^2} \right] \right\} \quad (11:68)$$

It is obvious from (11:17) that what is inside the curly brackets of (11:68) is the spectrum, Φ_{nn} , needed in (11:61). Substituting the spectrum evaluated at $\underline{K} = \underline{k} - \underline{km}$ and assuming isotropy as before, i.e., $|\underline{K}| = |\underline{k} - \underline{km}| = 2k \sin(\theta/2)$, we obtain

$$\langle S_m \rangle = \frac{\pi c_0 V k^6 A_0^2 \rho_0}{r^2} \cos^2 \theta \left[\cos^2 \left(\frac{\theta}{2}\right) \frac{E(2k \sin \frac{\theta}{2})}{c_0^2} + \frac{\Phi_{TT}(2k \sin \frac{\theta}{2})}{4T_0^2} \right] \quad (11:69)$$

for the average energy flux density.

The effective scattering cross section is found analogously to the EM case as the ratio of the energy flux scattered into a solid angle $d\Omega$ to the incident energy flux density

$$d\sigma = \frac{\langle S_m \rangle}{|S_0|} r^2 d\Omega \quad (11:70)$$

S_0 from the zero-order formula similar to (11:57) is

$$|S_0| = \left| \frac{k c_0 \rho_0 A_0^2}{2} \underline{k} \right| = \frac{k^2 c_0 \rho_0 A_0^2}{2} \quad (11:71)$$

If we now specialize Φ_{TT} and E to a Kolmogorov spectrum of turbulence, we have

$$\Phi_{TT}(K) = 0.033 C_T^2 K^{-11/3} \quad (11:72)$$

and

$$E(K) = 0.061 C_V^2 K^{-11/3} \quad (11:73)$$

where C_T^2 and C_V^2 are the so-called structure constants which are measures of the strength of the temperature and velocity fluctuations respectively. Substituting (11:69), (11:71), (11:72) and (11:73) into (11:70), we finally obtain

$$d\sigma = 0.030 k^{1/3} V \cos^2 \theta \left[\frac{C_V^2}{c_0^2} \cos^2 \left(\frac{\theta}{2} \right) + 0.13 \frac{C_T^2}{T_0^2} \right] \left[\sin \left(\frac{\theta}{2} \right) \right]^{-1/3} d\Omega \quad (11:74)$$

The general comments about the EM cross section after (11:47) apply here with a few interesting differences. The $\cos^2 \theta$ term above prohibits turbulent scattering of acoustic waves at $\theta = 90^\circ$ and inside the brackets the $\cos^2(\theta/2)$ term indicates that no scattering due to wind eddies will be apparent in the backscatter, $\theta = 180^\circ$, direction. (See 18 for the practical implications of (11:74).)

11.3 Line-of-Sight Propagation Effects

Studying the influence of the turbulent atmosphere on EM and acoustic wave parameters, i.e., amplitude, phase, and angle-of-arrival, can be a valuable tool for probing atmospheric dynamics. The problem is to determine the parameter that has the desired path weighting function (see 25), or equivalently, spatial resolution. Strohbehn (1966), (1970) discusses the general problem of remote probing of turbulence using line-of-sight propagation effects and the sensitivity of each wave parameter to different assumed refractivity spectra. The treatment to follow will consider only the line-of-sight propagation of high-frequency, plane EM waves. However, Clifford and Strohbehn (1970) and Clifford and Brown (1970) have shown that the results found here for the amplitude and phase spectra apply also for low-frequency EM and a wide band of acoustic waves. We consider the relatively simple plane wave propagation problem and quote the spherical wave results, leaving the details to be found in Fried (1967) and Lee and Harp (1969). More realistic sources, and receivers i.e., finite apertures, are considered by Ishimaru (1969), Schmeltzer (1967), Lutomirski and Yura (1971).

11.3.1 Perturbation Theory

The geometry of the current problem consists of a source and receiver within line-of-sight of each other and each imbedded in the same medium containing refractivity fluctuations, n_1 . The point receiver detects the amplitude and phase fluctuations, that is, the deep fades and degraded phase coherence of the signal in the receiving plane.

The equation that applies for this problem is, again, (11:6). The last term in (11:6) relates to the change in polarization of the field as the wave propagates. The effects of this term are considered by many authors, Strohbehn and Clifford (1967), Saleh (1967) and Strohbehn (1968). The last is the most general treatment of the problem to date. All results indicate that the depolarizing effects of atmospheric turbulence are entirely negligible for optical and microwave signals. Neglecting this term, we obtain scalar wave equations for the zero and first order perturbations of the field, E_0 and E_1 , in the form

$$\nabla^2 E_0 + k^2 E_0 = 0 \quad (11:75)$$

$$\nabla^2 E_1 + k^2 E_1 + 2k^2 n_1 E_0 = 0 \quad (11:76)$$

The unperturbed field, assumed a unit amplitude, z-directed plane wave, $E_0 = e^{ikz}$, must satisfy (11:75) and, substituted into (11:76), it becomes part of the source term. The simplest technique for solving

(11:76) is to substitute the two-dimensional, Fourier-Stieltjes representation of n_1 and E_1 and solve the resulting one-dimensional differential equation, i.e., let

$$n_1(x, y, z) = \int e^{i(K_2 x + K_3 y)} d\nu(z, K_2, K_3) \quad (11:77)$$

and

$$E_1(x, y, z) = \int e^{i(K_2 x + K_3 y)} d\epsilon_1(z, K_2, K_3) , \quad (11:78)$$

and substitute into (11:76) to obtain

$$\frac{d^2}{dz^2} (d\epsilon_1) + (k^2 - K^2) d\epsilon_1 = -2k^2 e^{ikz} d\nu \quad (11:79)$$

where $K^2 = K_2^2 + K_3^2$. The solution of (11:79) is straightforward and reduces to a convolution of its impulse response with the source term

$$d\epsilon_1(z, \underline{K}) = \frac{ik^2}{\sqrt{(k^2 - K^2)}} \int_0^\infty dz' e^{ikz' + i\sqrt{(k^2 - K^2)}|z - z'|} d\nu(z', K) . \quad (11:80)$$

(Using the form of (11:77) and (11:78) makes (11:80) valid only for globally homogeneous and isotropic random fields. However, with a suitable redefinition of $d\epsilon_1$ and $d\nu$, the result also applies when n_1 and E_1 are locally homogeneous random variables.) Since we are concerned with the amplitude and phase fluctuations, we must find the relation between $d\epsilon_1$ and da , $d\sigma$, the amplitude and phase random variables, respectively. Clearly, the square root of (11:80) substituted into (11:78) times its complex conjugate would give the amplitude variable directly. However, further analysis would involve finding fourth moments of (11:80) and would be exceedingly complicated, although straightforward. If, instead, we make use of the weak scattering assumption, $\ln_1 \ll 1$, we can greatly simplify our work. We may write,

$$\frac{E}{E_0} = 1 + \frac{E_1}{E_0} = \frac{A}{A_0} \exp [i(S - S_0)] \quad (11:81)$$

i.e., the total field $E = E_0 + E_1$ in terms of its amplitude, A , and phase, S , normalized to the initial fields' amplitude and phase, A_0 and S_0 . Take the natural logarithm of (11:81)

$$\ln \left(1 + \frac{E_1}{E_0} \right) = \ln \frac{A}{A_0} + i(S - S_0) \quad (11:82)$$

and note that for $|E_1| \ll |E_0|$, with $A = A_0 + A_1$, and $S_1 = S - S_0$, we have

$$\frac{E_1}{E_0} \approx \ln \left(1 + \frac{A_1}{A_0} \right) + i S_1 . \quad (11:83)$$

Again, assuming $|A_1| \ll |A_0|$, we obtain

$$\frac{A_1}{A_0} \approx \operatorname{Re} \left(\frac{E_1}{E_0} \right) \quad (11:84)$$

$$S_1 \approx \operatorname{Im} \left(\frac{E_1}{E_0} \right) . \quad (11:85)$$

These two equations give a simple way of determining the amplitude and phase perturbations as the real and imaginary parts of the normalized perturbed field E_1/E_0 .

Returning to (11:80), there are two simplifications that can be made for our problem. First, we are interested in forward scatter. Therefore, the integral should proceed from $z' = 0$, the beginning of the random medium, up to the observation point $z' = z$. The integral for $z' = z$ to $z = \infty$ represents the contribution due to backscatter. Second, as will become clear in the development, the important values of K assuming $kL \gg 1$, lie in the region $K \ll k$; the phase results being sensitive to much smaller K 's than the amplitude (see 25). For these values of K , the following approximation holds $\sqrt{(k^2 - K^2)} \sim k \left[1 - \frac{K^2}{2k^2} \right]$. Retaining both terms in the exponential and the first term in the denominator of (11:80), we obtain

$$d\epsilon_1(z, \underline{K}) = ik \int_0^z dz' e^{ikz} - i \left(\frac{K^2(z-z')}{2k} \right) d\nu(z', \underline{K}) \quad (11:86)$$

With (11:86) inserted into (11:78) after normalizing by $E_0 = e^{ikz}$ as indicated in (11:84) and (11:85), we obtain

$$da(z, \underline{K}) = k \int_0^z dz' \sin \left[\frac{K^2(z-z')}{2k} \right] d\nu(z', \underline{K}) \quad (11:87)$$

$$d\sigma(z, \underline{K}) = k \int_0^z dz' \cos \left[\frac{K^2(z-z')}{2k} \right] d\nu(z', \underline{K}) \quad (11:88)$$

These expressions are from the Fourier expansions of the amplitude and phase fluctuations analogous to (11:78). The form of (11:87) and (11:88) requires the additional information that $d\nu^*(z, -\underline{K}) = d\nu(z, \underline{K})$, a consequence of the fact that n_1 must be real.

Relations for $d\nu$, da and $d\sigma$ that can easily be derived from the definition of a two-dimensionally homogeneous covariance function are from Tatarski (1961)

$$\langle d\nu(z_1, \underline{K}) d\nu^*(z_2, \underline{K}') \rangle = \delta(\underline{K} - \underline{K}') F_n(\underline{K}, z_1 - z_2) d^2 \underline{K}' d^2 \underline{K} \quad (11:89)$$

$$\langle da(z, \underline{K}) da^*(z, \underline{K}') \rangle = \delta(\underline{K} - \underline{K}') F_A(\underline{K}, 0) d^2 \underline{K}' d^2 \underline{K} \quad (11:90)$$

$$\langle d\sigma(z, \underline{K}) d\sigma^*(z, \underline{K}') \rangle = \delta(\underline{K} - \underline{K}') F_S(\underline{K}, 0) d^2 \underline{K}' d^2 \underline{K} \quad (11:91)$$

The functions F_n , F_A and F_S are two-dimensional spectra defined by

$$F_g(\underline{K}, \xi) = \int_{-\infty}^{\infty} \cos(K_1 \xi) \Phi_g(K_1, \underline{K}) dK_1 \quad (11:92)$$

where $\Phi_{\mathbf{g}}$ is the three-dimensional spectrum, e.g., (11:16). Multiplying (11:87) and (11:88) by their respective complex conjugates and averaging using (11:89) through (11:91) yields

$$F_A(\underline{\mathbf{K}}, 0) = k^2 \int_0^z dz' \int_0^z dz'' \sin \left[\frac{K^2(z-z')}{2k} \right] \sin \left[\frac{K^2(z-z'')}{2k} \right] F_n(\underline{\mathbf{K}}, z' - z'') \quad (11:93)$$

$$F_S(\underline{\mathbf{K}}, 0) = k^2 \int_0^z dz' \int_0^z dz'' \cos \left[\frac{K^2(z-z')}{2k} \right] \cos \left[\frac{K^2(z-z'')}{2k} \right] F_n(\underline{\mathbf{K}}, z' - z'') \quad (11:94)$$

At this point it is useful to examine the relation between the two-dimensional spectral density, $F_n(\underline{\mathbf{K}}, z' - z'')$, and the refractive index spectrum, $\Phi_n(\underline{\mathbf{K}})$, considered in 11.1.2. $F_n(\underline{\mathbf{K}}, z' - z'')$ has some important properties that simplify the integration involved in (11:93) and (11:94). First, it is apparent from (11:92) that F_n is even in $z' - z''$, that is

$$F_n(\underline{\mathbf{K}}, z' - z'') = F_n(\underline{\mathbf{K}}, z'' - z')$$

Second, since F_n corresponds to the correlation in two planes along the z axis at $z = z'$ and $z = z''$, clearly any contribution to this correlation must come from inhomogeneities that are of such a size, ℓ , that they intersect the two planes. Therefore under the assumption that $\Phi_n(\mathbf{K}_1, \underline{\mathbf{K}})$ is isotropic, i.e., $\Phi_n(\mathbf{K}_1, \underline{\mathbf{K}}) = \Phi_n(\mathbf{K}_0)$ where $\mathbf{K}_0^2 = \mathbf{K}_1^2 + \mathbf{K}^2$, the scale of the inhomogeneities for which F_n is non-zero must satisfy the inequality $\ell \cong \frac{1}{K_0} > (z' - z'')$. Therefore the function $F_n(\underline{\mathbf{K}}, z' - z'')$ falls off rapidly outside the interval $K|z' - z''| \leq 1$.

Returning to (11:93) and (11:94), we introduce the relative coordinates $\xi = z' - z''$ and $2\eta = z' + z''$. We note that the Jacobian of this transformation is unity and that the integration with respect to η may be performed directly, yielding

$$F_A(\underline{\mathbf{K}}, 0) = \int_0^L d\xi \left[k^2(L - \xi) \cos \left(\frac{K^2\xi}{2k} \right) + \frac{k^3}{K^2} \sin \left(\frac{K^2\xi}{2k} \right) - \frac{k^3}{K^2} \sin \left(\frac{K^2(2L - \xi)}{2k} \right) \right] F_n(\underline{\mathbf{K}}, \xi) \quad (11:95)$$

$$F_S(\underline{\mathbf{K}}, 0) = \int_0^L d\xi \left[k^2(L - \xi) \cos \left(\frac{K^2\xi}{2k} \right) - \frac{k^3}{K^2} \sin \left(\frac{K^2\xi}{2k} \right) + \frac{k^3}{K^2} \sin \left(\frac{K^2(2L - \xi)}{2k} \right) \right] F_n(\underline{\mathbf{K}}, \xi) \quad (11:96)$$

where the coordinate, z , has been replaced by the total path length, L .

From the argument above, we note that $F_n(\underline{\mathbf{K}}, \xi)$ falls off rapidly for $K\xi \geq 1$. Therefore, in the important region of integration $\frac{K^2\xi}{k} \leq \frac{K}{k}$. Remembering the assumption that $K \ll k$, then $\frac{K^2\xi}{2k} \ll 1$ and we may write

$$\cos \left(\frac{K^2\xi}{2k} \right) \sim 1, \sin \left(\frac{K^2\xi}{2k} \right) \sim \frac{K^2\xi}{2k}, \sin \left[\frac{K^2(2L - \xi)}{2k} \right] \sim \sin \frac{K^2L}{k} \quad (11:97)$$

One further simplification arises from the fact that we are not interested in the correlation functions over distances that are the order of the path length. Therefore we are interested in those scale sizes which satisfy the relation $K^{-1} \ll L$. But, since we have already shown that the major contribution to the integrals is for $\xi \leq K^{-1}$ then $\xi \ll L$. Using the above approximations in the integrals for $F_A(\underline{\mathbf{K}}, 0)$ and $F_S(\underline{\mathbf{K}}, 0)$ we have,

$$F_A(\underline{\mathbf{K}}, 0) \cong \left[k^2L - \frac{k^3}{K^2} \sin \left(\frac{K^2L}{k} \right) \right] \int_0^L F_n(\underline{\mathbf{K}}, \xi) d\xi \quad (11:98)$$

$$F_S(\underline{K}, 0) \cong [k^2 L + \frac{k^3}{K^2} \sin(\frac{K^2 L}{k})] \int_0^L F_n(\underline{K}, \zeta) d\zeta . \quad (11:99)$$

Since $F_n(\underline{K}, \zeta)$ falls off rapidly for large ζ , we may extend the integration to infinity with negligible error. Using the inverse of the transform relation (11:92), we finally obtain

$$F_A(\underline{K}, 0) = \pi k^2 L [1 - \frac{k}{K^2 L} \sin(\frac{K^2 L}{k})] \Phi_n(0, \underline{K}) \quad (11:100)$$

$$F_S(\underline{K}, 0) = \pi k^2 L [1 + \frac{k}{K^2 L} \sin(\frac{K^2 L}{k})] \Phi_n(0, \underline{K}) \quad (11:101)$$

Equations (11:100) and (11:101) relate the two-dimensional spectra for the amplitude and phase fluctuations in the plane $z = L$ to the refractive index spectrum. In the case where the index of refraction is an isotropic random field, we have $\Phi_n(K_1, \underline{K}) = \Phi_n(\sqrt{K_1^2 + K^2})$, $F_A(\underline{K}, 0) = F_A(K, 0)$ and $F_S(\underline{K}, 0) = F_S(K, 0)$. Then (11:100) and (11:101) become

$$F_A(K, 0) = \pi k^2 L [1 - \frac{k}{K^2 L} \sin(\frac{K^2 L}{k})] \Phi_n(K) \quad (11:102)$$

$$F_S(K, 0) = \pi k^2 L [1 + \frac{k}{K^2 L} \sin(\frac{K^2 L}{k})] \Phi_n(K) . \quad (11:103)$$

11.3.2 Spatial Covariance Functions

Equations (11:102) and (11:103) are the isotropic two-dimensional spectra of the amplitude and phase variations in the receiving plane. Their major utility arises from the fact that a two-dimensional Fourier transform reduces them to measurable quantities, $B_A(\rho)$ and $B_S(\rho)$, the spatial covariance functions. The isotropic spatial covariance functions are defined by

$$B_{A, S}(\rho) = 2\pi \int_0^\infty J_0(K\rho) F_{A, S}(K, 0) K dK . \quad (11:104)$$

For the amplitude and phase spectra in (11:102) and (11:103)

$$B_{A, S}(\rho) = 2\pi^2 k^2 L \int_0^\infty J_0(K\rho) [1 \mp (\frac{k}{K^2 L}) \sin(\frac{K^2 L}{k})] \Phi_n(K) K dK \quad (11:105)$$

where the minus applies for the amplitude and plus for phase. If we insert a Kolmogorov refractivity spectrum from (11:46) into (11:105) the amplitude covariance function is well-defined whereas the phase result diverges. The diffraction process (see Chapter 25) that causes the amplitude fluctuations, produces a $[1 - \sin(\frac{K^2 L}{2k}) / (\frac{K^2 L}{2k})]$ spectral filter function that removes the singularity in the refractivity spectrum at $K = 0$. The phase results, on the other hand, do not remove the singularity and the integral diverges unless the integration is truncated at $K \sim L_0^{-1}$. Physically, this implies that the amplitude fluctuations are produced by specific, relatively small-size eddies, $K \sim [\lambda L]^{-1/2}$, whereas the phase fluctuations are determined by the largest refracting eddies, $K \sim L_0^{-1}$, found in the medium. Phase

covariance measurements will be compromised by insufficient low-frequency cutoffs or, at the other extreme, degraded by slowly varying non-Kolmogorov processes that could produce non-stationary data and absurd results. It is best to produce a spatial filter that filters the data in a known way and reduces the influence of large eddies.

A useful spectral filter is obtained from the structure function. From the relation for phase analogous to (11:44) and the definition (11:104), we obtain

$$D_S(\rho) = 4\pi^2 k^2 L \int_0^\infty [1 - J_0(K\rho)] \left[1 + \left(\frac{k}{K^2 L}\right) \sin\left(\frac{K^2 L}{k}\right)\right] \Phi_n(K) K dK \quad (11:106)$$

The new filter function $[1 - J_0(K\rho)]$ smoothes the spectral singularity at $K=0$ and produces a stable, measurable function. The integral in (11:106) is performed by Tatarski (1961) for the Kolmogorov spectrum and yields

$$D_S(\rho) = \begin{cases} 1.46 C_n^2 k^2 L \rho^{5/3} & \ell_0 \ll \rho \ll \sqrt{(\lambda L)} \\ 2.91 C_n^2 k^2 L \rho^{5/3} & \sqrt{(\lambda L)} \ll \rho \ll L_0. \end{cases} \quad (11:107)$$

The amplitude covariance function from (11:105), generalized for a non-uniform distribution of C_n^2 is from Tatarski (1961)

$$B_A(\rho) = 0.132 \pi^2 k^2 \int_0^L dz C_n^2(z) \int_0^\infty dK K^{-8/3} J_0(K\rho) \sin^2 \left[\frac{K^2(L-z)}{2k} \right]. \quad (11:108)$$

The log-amplitude variance, $\sigma_\chi^2 = B_A(0)$, found from (11:108), with C_n^2 uniform along the z axis is

$$\sigma_\chi^2 = 0.31 C_n^2 k^{7/6} L^{11/6}, \quad \sqrt{(\lambda L)} \gg \ell_0. \quad (11:109)$$

The equivalent results for spherical waves are (Tatarski, 1967)

$$D_S(\rho) = \begin{cases} 0.54 C_n^2 k^2 L \rho^{5/3}, & \ell_0 \ll \rho \ll \sqrt{(\lambda L)} \\ 1.09 C_n^2 k^2 L \rho^{5/3}, & \sqrt{(\lambda L)} \ll \rho \ll L_0 \end{cases} \quad (11:110)$$

$$B_A(\rho) = 0.132 \pi^2 k^2 \int_0^L dz C_n^2(z) \int_0^\infty dK K^{-8/3} J_0(K\rho z/L) \sin^2 \left[\frac{K^2 z(L-z)}{2kL} \right] \quad (11:111)$$

and with C_n^2 uniform along z

$$\sigma_\chi^2 = 0.124 C_n^2 k^{7/6} L^{11/6}, \quad \sqrt{(\lambda L)} \gg \ell_0 \quad (11:112)$$

The physical interpretation and applications of (11:106) – (11:112) are extensively considered in later chapters. (see especially Chapter 25). We attempt in this section to indicate how they are obtained and provide a convenient summary of the results. Extensive details in the derivation of each equation are found in the cited references.

11.3.3 Restrictions on the Solution

The solutions for the covariance function, $B_A(\rho)$, and the log-amplitude variance are known to be incorrect for optical propagation near the ground. In fact, in the case of $\sigma_\chi^2 > 0.3$, the results appear to break down for both acoustic (Mandics, 1971) and optical beams (Ochs and Lawrence, 1969). There is evidence that the diffraction-theory results (Chapter 25) do not apply for such strong-scintillation (large σ_χ^2) conditions. This is reflected in the mathematics by the non-convergence of the perturbation series for E_1 when the strength of turbulence, C_n , or the path length L is sufficiently large. To date much work has been done to attempt an understanding of this problem with very little accomplishments.

11.4 References

- Batchelor, G. K., 1953: *The Theory of Homogeneous Turbulence*. Cambridge University Press, London.
- Blackman, R. B. and J. W. Tukey, 1958: *The Measurement of Power Spectra from the Point of View of Communications Engineering*. Dover Publ., Inc. New York.
- Clifford, S. F. and E. H. Brown (1970), Propagation of sound in a turbulent atmosphere, *J. Acoust. Soc. Am.* **48**, 1123-1127.
- Clifford, S. F. and J. W. Strohbehn (1970), The theory of microwave line-of-sight propagation through a turbulent atmosphere, *IEEE Trans. on Ant. and Prop.* *AP-18*, 264-274.
- Fried, D. L. (1967), Propagation of a spherical wave in a turbulent medium, *J. Opt. Soc. Am.* **57**, 175-180.
- Ishimaru, A. (1969), Fluctuations of a beam wave propagating through a locally homogeneous medium, *Radio Sci.* **4**, 295-305.
- Kolmogorov, A., 1941: in *Turbulence, Classic Papers on Statistical Theory*. S. K. Friedlander and L. Topper Editors, INTERSCIENCE, New York, 151, 1961.
- Lee, R. W. and J. C. Harp (1969), Weak scattering in random media, with applications to remote probing, *Proc. IEEE* **57**, 375-406.
- Lumley, J. L. and H. A. Panofsky, 1964: *The Structure of Atmospheric Turbulence*. John Wiley and Sons, New York.
- Lutomirski, R. F. and H. T. Yura (1971), Propagation of a finite optical beam in an inhomogeneous medium, *Appl. Opt.* **10**, 1652-1658.
- Mandics, P. A. (1971), Line-of-sight acoustical probing of atmospheric turbulence, Stanford Electronics Labs Tech. Rept. No. 4502-1 (SEL-71-002).
- Monin, A. S. (1962), Characteristics of the scattering of sound in a turbulent atmosphere, *Soviet Physics-Acoustics* **7**, 370-373.
- Ochs, G. R. and R. S. Lawrence (1969), Saturation of laser beam scintillation under conditions of strong atmospheric turbulence, *J. Opt. Soc. Am.* **59**, 226-227.
- Pasquill, F., 1961: *Atmospheric Diffusions*. D. Van Nostrand, New York.

- Saleh, A. A. M. (1967), An investigation of laser wave depolarization due to atmospheric transmission, IEEE J. Quant. Electron, *QE-3*, 540-543.
- Schmeltzer, R. A. (1967), Means, variances and covariances for laser beam propagation through a random medium, Quart. Appl. Math. *24*, 339-354.
- Strohbehn, J. W. (1966), The feasibility of laser experiments for measuring atmospheric turbulence parameters, J. Geophys. Res. *71*, 5793-5808.
- Strohbehn, J. W. (1968), Line-of-sight wave propagation through the turbulent atmosphere, Proc. IEEE *56*, 1301-1318.
- Strohbehn, J. W. (1970), The feasibility of laser experiments for measuring the permittivity spectrum of the turbulent atmosphere, J. Geophys. Res. *75*, 1067-1076.
- Strohbehn, J. W. and S. F. Clifford (1967), Polarization and angle-of-arrival fluctuations for a plane wave propagated through a turbulent medium, IEEE Trans. on Ant. and Prop. *AP-15*, 416-421.
- Tatarski, V. I., 1961: *Wave Propagation in a Turbulent Medium*. McGraw-Hill Book Co., New York. (Translated by R. A. Silverman)
- Tatarski, V. I., 1967: *Propagation of Waves in a Turbulent Atmosphere*. Nauka, Moscow, USSR. (in Russian)
- Wiener, N., 1949: *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. John Wiley and Sons, New York.
- Yaglom, A. M., 1962: *An Introduction to the Theory of Stationary Random Functions*. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

List of Symbols

\underline{A}_0	Amplitude vector of the unperturbed wave	F_A	Two-dimensional spectrum of the amplitude fluctuations
\underline{A}_1	Amplitude fluctuation of a plane EM wave	F_n	Two-dimensional spectrum of the refractivity fluctuations
$B_A(\underline{\rho})$	Two-dimensional spatial covariance function of the amplitude fluctuations	F_S	Two-dimensional spectrum of the phase fluctuations
$B_{j\ell}(\underline{r})$	The correlation tensor of the velocity field	$f(\underline{r})$	A general function of position
$B_n(\underline{r})$	Covariance function of the refractive index fluctuations	\underline{H}	Magnetic field
$B_S(\underline{\rho})$	Two-dimensional spatial covariance function of the phase fluctuations	\underline{H}_0	Unperturbed magnetic field
B_{TT}	Covariance function of the temperature fluctuations	\underline{H}_1	Perturbed magnetic field
B_{xx}	Covariance function of the x-directed wind velocity	i	Running index of integers
C_n^2	Refractive index structure constant	Im	Take the imaginary part of
C_T^2	Temperature structure constant	$\underline{K}=(K_x, K_y,$	Three-dimensional spatial wave number
C_v^2	Wind velocity structure constant	$\quad K_z)$	
c	Speed of phase propagation of acoustic or electromagnetic waves	\underline{K}_0	Three-dimensional wave vector describing the turbulent spatial frequencies that reinforce the incident field in the direction of \underline{m}
c_0	Mean speed of sound in air	k	Wavenumber of the radiation
$D_n(\underline{r})$	Structure function of the refractivity fluctuations	L	A dimension of the volume containing refractivity fluctuations, pathlength
$de_1(z, \underline{K})$	Fourier-Stieltjes measure of the perturbed electric field	L_0	Outer scale of turbulence
$da(z, \underline{K})$	Fourier-Stieltjes measure of the wave amplitude fluctuations	ℓ_0	Inner scale of turbulence
$d\sigma(z, \underline{K})$	Fourier-Stieltjes measure of the wave phase fluctuations	\underline{m}	Unit vector in the scattering direction
$d\nu(z, \underline{K})$	Fourier-Stieltjes measure of the refractivity fluctuations	n	Dummy position variable
$d\Omega$	Solid angle	$n(\underline{r})$	Index of refraction
$d\sigma$	Scattering cross section	$n_1(\underline{r})$	Refractive index fluctuation
\underline{E}	Total electric field	P_0	Unperturbed acoustic wave
$E[]$	Expected value operator	P_1	Perturbed acoustic wave field
\underline{E}_1	Perturbed electric field vector	$P(\Omega)$	Probability distribution function
\underline{E}_0	Unperturbed electric field vector	$p(\Omega)$	Probability density function
$E(\underline{K})$	Three-dimensional energy spectrum	$\underline{r}=(x, y, z)$	A position vector
		$\underline{r}'=(x', y', z')$	A vector describing coordinates inside the volume containing refractivity variations
		\underline{S}	Scattered wave Poynting vector
		S	Total phase of a plane wave
		S_0	Unperturbed phase of a plane wave

\underline{S}_0	Incident wave Poynting vector
S_1	Perturbed phase of a plane wave
S_m	The Component of Poynting vector in the direction of \underline{m}
T	A temperature fluctuation from background
T_0	Average background temperature
t	Time
u'_x	Fluctuations in x-directed wind velocity
V	Volume
∇	Gradient operator
$\delta_{j\ell}$	Kronecker delta
ζ	Dummy position variable
θ	Scattering angle
λ	Wavelength
$\underline{\rho}$	Position vector
ρ_0	The mean density of the atmosphere
σ^2_χ	The variance of the log-amplitude
ψ	General function
$\Phi_n(\underline{K})$	Three-dimensional spectrum of the refractive index fluctuations
Φ_{nn}	The acoustic refractive index spectrum
Φ_{TT}	Three-dimensional spectrum of the temperature fluctuations
Φ	General function
χ	Angle between \underline{m} and \underline{A}_0 , logarithm of the amplitude fluctuations
Ω	Ensemble index
ω	Radian frequency
\diamond	Averaging operator

Chapter 12 REMOTE SENSING OF SEA STATE BY RADAR

Donald E. Barrick

Electromagnetics Division
Battelle, Columbus Laboratories

Several radar techniques have evolved over recent years which permit the straightforward measurement of certain important ocean wave parameters. At MF and HF, the ocean waveheight spatial spectrum can be measured directly via the first-order Bragg-scattered signal intensity; a variety of experiments are briefly examined which involve monostatic ground-wave and ionospheric radars, bistatic HF buoy-shore systems, bistatic LORAN A signal scatter systems, and bistatic buoy-satellite systems. The second-order contributions to HF scatter produce a continuous Doppler return which varies in position and amplitude with sea state. At UHF, it is possible to measure indirectly the spatial slope spectrum of the longer ocean waves via cross-correlation of simultaneous Bragg-effect returns at two frequencies. Finally, short-pulse microwave satellite altimeters permit a direct measurement of the significant (or rms) waveheight of the sea at the suborbital point via the specular point scatter mechanism. These techniques will be important for (i) detailed oceanographic measurements of the characteristics of sea waves, (ii) routine monitoring of sea state for maritime purposes, and (iii) deduction of wind patterns above the seas for meteorological purposes.

12.1 Description of the Sea Surface

The quantitative interpretation of radar scatter from the sea requires the use and appreciation of certain properties of ocean waves. A brief review is undertaken here of the ocean-wave physics and characteristics which we will need later; also, common oceanographic nomenclature pertaining to ocean waves is defined and explained. A readable but detailed treatment of all aspects of ocean wave physics can be found in the text by Kinsman (1965); a more elementary introduction to water waves is the concise soft-cover booklet by Bascom (1964).

12.1.1 Nomenclature

Sea State. This term as used here refers to the state of the sea, or roughness, as determined by the heights of the largest waves present. Numbers have been assigned to sea states by the International Mariners' Codes, and these are related to wave heights in (T12.1).

Significant Wave Height. This term is a common maritime descriptor referring to the average of the heights—from crest to trough—of the 1/3 highest waves; it is denoted $H_{1/3}$.

RMS Wave (or Roughness) Height. This is a term describing root-mean-square height—above the mean surface level—used in rough surface scatter theories; it is denoted here by h . While there is no exact general relationship between h and $H_{1/3}$, a common approximation frequently used for wind waves is $H_{1/3} \simeq 2.83 h$.

Length. The length or spatial period of a single ocean wave is the distance from one crest to another; it is denoted L .

Period. Unless denoted otherwise, this refers to the temporal period, and is the length of time it takes two successive crests of a single wave to pass one point. It is denoted T .

Spatial Wavenumber. This is defined in terms of the length of an ocean wave as $\kappa = 2\pi/L$.

Temporal Wavenumber. This radian wavenumber is given in terms of the period by $\omega = 2\pi/T$.

Fetch. The fetch is the horizontal distance over which a nearly constant wind has been blowing.

Duration. This term refers to the length of time during which a nearly constant wind has been blowing.

Wind Waves. This term refers to a system of ocean waves which is being, or has very recently been, aroused by winds blowing locally above that area of the ocean. Wind waves result in a random appearing ocean height profile.

Fully Developed Seas. This is an equilibrium sea state condition reached after sufficient duration and fetch at a given wind speed. The estimated duration and fetch versus wind speed required to produce fully developed seas is provided in (T12.1)

Table 12.1 Deep-Water Wind Waves And Sea State

WIND VELOCITY (KNOTS)	4		5		6		7		8		9		10		20			30		40		50		60		70						
BEAUFORT WIND AND DESCRIPTION	1 LIGHT AIR		2 LIGHT BREEZE		3 GENTLE BREEZE		4 MODERATE BREEZE		5 FRESH BREEZE		6 STRONG BREEZE		7 MODE RATE GALE		8 FRESH GALE		9 STRONG GALE		10 WHOLE GALE		11 STORM											
REQUIRED FETCH (MILES)							50		100		200		300		400		500		600		700											
REQUIRED WIND DURATION (HOURS)							5		20		25				30												35					
SIGNIFICANT WAVE HEIGHT* (FEET)							2		4		6		8		10		15		20		25		30		40		50		60			
SEA STATE AND DESCRIPTION	1 SMOOTH		2 SLIGHT		3 MODE RATE		4 ROUGH		5 VERY ROUGH		6 HIGH		7 VERY HIGH		8 PRECIPITOUS																	
WAVE PERIOD (SECONDS)	1		2		3		4		6		8		10		12		14		16		18		20									
WAVE LENGTH (FEET)			20		40		60		80		100		150		200		300		400		500		600		800		1000		1400		1800	
WAVE VELOCITY (KNOTS)			5		10		15		20		25		30		35		40		45		50		55		60							
PARTICLE VELOCITY (FEET/SECOND)	1		2		3		4		5		6		8		10		12		14													
WIND VELOCITY (KNOTS)	4		5		6		7		8		9		10		20			30		40		50		60		70						
*If the fetch and duration are as great as indicated above, these waveheight and sea state conditions exist. If fetch and duration are greater, waveheight can be up to 10% greater.																																

Swell. When wind waves move out of the area in which they were originally excited by the winds, or after winds have ceased to blow, these waves change their shape and settle down to what is known as "swell". Swell appears less random and more nearly sinusoidal, of great length, and with great width along the crestlines. The usual period of swell is from six to sixteen seconds. Swell, while an occasional phenomenon, can arise from storm areas thousands of miles distant.

Deep-Water Waves. When the water is sufficiently deep that the effect of the bottom on the propagation characteristics of the waves can be neglected, they are called "deep-water" waves. Generally, if the depth is greater than 1/2 the length of a given wave, the deep-water approximation is valid. Except near beaches, ocean waves are deep-water waves, and we utilize this assumption throughout this chapter.

Gravity Waves. This term refers to waves in which the chief restoring force upon the perturbed water mass is gravity. Waves whose lengths, L, are greater than 1.73 cm (Phillips, 1966) are gravity waves. Since gravity waves are the essence of sea state, they are the only types of waves considered in this chapter.

Capillary Waves. This term refers to waves in which the chief restoring force acting on the perturbed water mass is surface tension. Less than 1.73 cm in length, they are not important for most of the topics of this chapter.

12.1.2 Wind Wave Surface Height and Slope Distributions

Patterns of wind waves having various lengths, heights, and directions of motion interact to form a random-appearing surface. Hence the quantitative characteristics of such a surface are best described statistically. One of the statistical functions frequently occurring in the analysis of radio wave interactions with the sea is the probability density function of the surface height and its spatial derivatives (or slopes). Physically, the probability density function $p(x)$ is defined such that $p(x)dx$ is the probability that the random variable lies in the interval dx between $x - dx/2$ and $x + dx/2$.

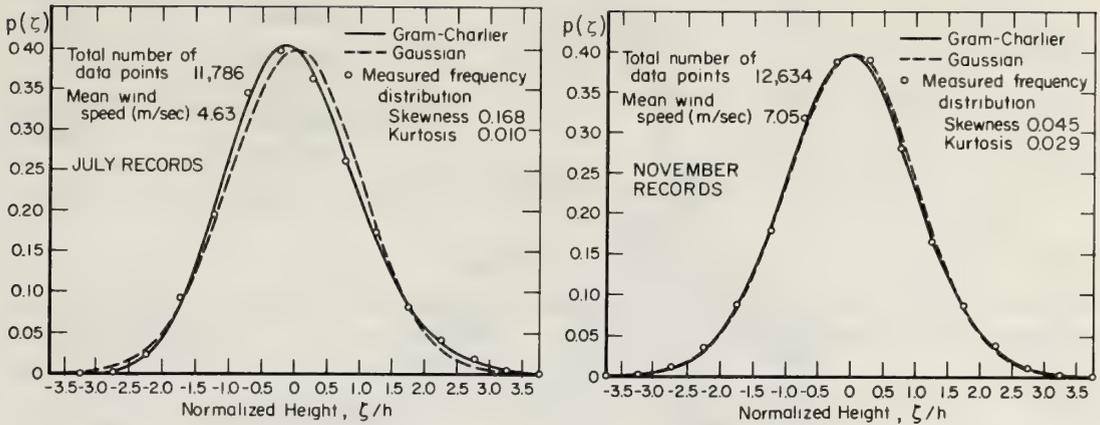


Figure 12.1 Measured versus model probability density functions for sea surface height (after MacKay, 1959).

For the sea surface height, ξ , above the mean sea level, MacKay (1959) found from detailed analyses of measured wave records that the height probability density function is nearly Gaussian (or normal). This yields

$$p(\xi) = (2\pi h^2)^{-1/2} \exp\left\{-\frac{\xi^2}{2h^2}\right\} \quad (12:1)$$

where h is the rms height of the surface, i.e., $h = (\langle \xi^2 \rangle)^{1/2}$, where $\langle \dots \rangle$ denotes average. The actual density for the sea height cannot be truly Gaussian for two reasons: (i) For Gaussian distributed waves, there is always some finite—albeit small—probability that very large waveheights can occur, whereas for the sea, wave breaking occurs when the heights and slopes exceed certain critical values. (ii) The Gaussian function is symmetric, whereas the sea height is not truly symmetric about the mean. This can be seen from looking at the sea surface profile, which tends to have sharp pointed peaks (for $\xi > 0$), but rounded shallow troughs (for $\xi < 0$). Thus the sea surface profile would not look the same upside down, whereas a true Gaussian variable would.

For the latter reason, the true height probability density function is slightly better matched by a Gram-Charlier model than by the Gaussian, as shown in (F12.1), after MacKay (1959). The difference is very slight, however, and for most analytical purposes the Gaussian model is entirely adequate†. The Gaussian height distribution will be assumed and used throughout this chapter.

If the height distribution for the sea were truly Gaussian, then the distribution of its spatial derivatives (i.e., the slopes) would also be Gaussian, because a linear operation on a Gaussian random variable (e.g., differentiation) produces another Gaussian random variable. The actual slope distributions for the sea are again almost—but not quite—Gaussian. We take the x -axis as horizontal and pointing in the dominant wind direction (i.e., along the downwind direction), and the y -axis as horizontal and pointing in the crosswind direction. Then Cox and Munk (1954), using glitter point photography to measure the directional slopes, find that $\xi_y (= \partial \xi / \partial y)$ in the crosswind direction is symmetric, but $\xi_x (= \partial \xi / \partial x)$ is skewed toward the upwind direction, probably due to wind stress. This is shown in (F12.2). Thus the departure from Gaussian is again slight, and while it could be important in applications involving radar scatterometers looking near the vertical with high angular resolution (Nathanson, 1971), the difference is ignored in this chapter. Also noteworthy from the figures is the fact that the observed rms slope in the upwind-downwind direction is not significantly different from that in the crosswind direction. It can be shown analytically that ξ_x and ξ_y are uncorrelated at any given point on the ocean. Therefore, we take the following for the *joint* probability density function for the surface slopes:

$$p(\xi_x, \xi_y) = (\pi s^2)^{-1} \exp\left\{-\frac{\xi_x^2 + \xi_y^2}{s^2}\right\} \quad (12:2)$$

† One case where one might desire a higher order correction to the Gaussian model accounting for the skewness is in a detailed analysis of the short pulse return from a radar altimeter.

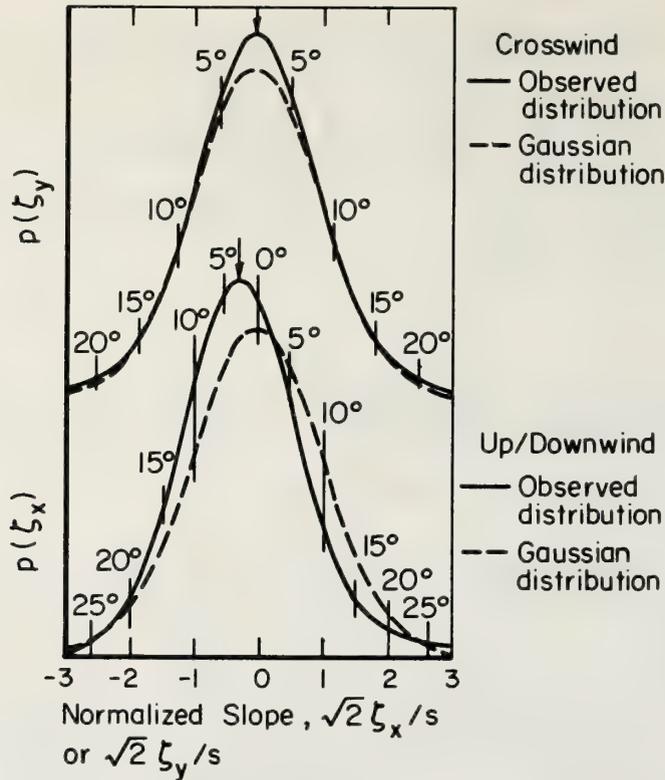


Figure 12.2 Measured versus model probability density functions for sea surface slopes (after Cox and Munk, 1954).

where, as explained above, we take $s^2 = \langle \xi_x^2 + \xi_y^2 \rangle = \langle \xi_x^2 \rangle + \langle \xi_y^2 \rangle = 2\langle \xi_x^2 \rangle = 2\langle \xi_y^2 \rangle$, s being the total rms slope of the ocean surface at a given point.

12.1.3 First-Order Gravity Wave Dispersion Relationship

The equation for the surface height, ζ , of a deep-water gravity wave is obtained from hydrodynamic theory (Kinsman, 1965). Generally, the wave surface height, ζ , is a function of the two orthogonal horizontal coordinates x, y and of time, t , i.e., $\zeta(x, y, t)$. This function—as well as the velocity potential and stream function—must satisfy Laplace's second-order differential equation; in addition, they satisfy two boundary conditions at the free surface: (i) the kinematic condition and (ii) the dynamic condition. While Laplace's differential equation is linear, the boundary conditions are not. Thus an exact solution is difficult to obtain.

The common method of solving these equations is to expand all of the functions into a perturbational series. Then the nonlinear boundary conditions are ordered into several equations, each containing terms of a higher order of magnitude. The ordering (or perturbational) parameter is the height of a wave divided by its length; this quantity is always very small for gravity waves. The lowest-order equations to emerge are linear and can be solved for the first-order height, ζ , of the surface. Second and higher-order solutions for ζ can also be obtained, and will be discussed in a later section. As the height of the water wave decreases, the first-order solution becomes increasingly valid because higher-order terms decrease in magnitude. Hence, the first-order solution is also referred to as the small-amplitude approximation for water waves.

The first-order solution, ζ , has several distinctive characteristics. It can consist of the superposition of an arbitrary number of sinusoids of different amplitudes, spatial lengths, and directions. *But*—unique to water waves—each sinusoid of a given wavelength (or wavenumber) moves at a distinct phase velocity. The relationship for the first-order phase velocity, v , is obtained also from the lowest-order surface boundary

conditions. It is (for gravity waves)

$$v = \sqrt{\frac{gL}{2\pi}} = \sqrt{\frac{g}{\kappa}}, \tag{12:3}$$

where g is the acceleration of gravity (≈ 9.81 m/s), and $\kappa = 2\pi/L$ is the total spatial wavenumber for the wave of length L .

Another way of stating (12:3) is to relate the temporal wavenumber, ω , of the wave to its spatial wavenumbers; this is commonly called a dispersion relationship in physics. It is

$$\omega^2 = g\kappa = g\sqrt{\kappa_x^2 + \kappa_y^2}, \tag{12:4}$$

where we assume for generality that the wave is moving in a direction whose angle θ with respect to the x -axis is given by $\tan^{-1}(\kappa_y/\kappa_x)$. The total wavenumber magnitude, κ , is thus the square root of the sum of the squares of the x - and y -directed spatial wavenumbers κ_x and κ_y .

From (12:4) we can obtain still another commonly seen first-order expression relating the wavelength to the period:

$$T^2 = \left(\frac{2\pi}{g}\right)L. \tag{12:5}$$

Table 12.2 Relationship Between Period, Length, And Phase Velocity Of Small Amplitude Gravity Waves

Period, T, seconds	Wave Length, L		Velocity, v	
	feet	meters	knots	meters/second
6	184	56	18.1	9.3
8	326	100	24.1	12.4
10	512	156	30.2	15.5
12	738	225	36.2	18.6
14	1000	305	42.4	21.8
16	1310	400	48.6	25.0

Thus we see for first-order gravity waves a unique square root relationship between the water wavelength and its temporal characteristics such as its velocity and period. Table (12.2) provides a ready connection between these quantities for the longer, higher waves which generally comprise sea state. It will be seen later that this square-root dispersion relationship forms the basis for several unique radar experiments involving scatter from sea waves.

12.1.4 Waveheight Spectrum of Wind Waves

The statistical quantity developed by oceanographers to relate the height of ocean waves to their length is the waveheight spectrum. It will be seen later that this spectrum also appears directly in radar scatter theories. The most general form for this spectrum contains two spatial wavenumbers (κ_x, κ_y) and one temporal wavenumber (ω) to describe the waveheight $\zeta(x,y,t)$ as a function of its three independent variables; we denote it as $S(\kappa_x, \kappa_y, \omega)$.

For a random-like system of wind waves, we assume that the dominant wind and wave direction is in the $+x$ direction. Then we can express the surface height in a Fourier series as a sum of traveling waves:

$$\zeta = \sum_{m,n,k=-\infty}^{\infty} P(m,n,k)e^{iamx+iany-i\omega t} = \sum_{m,n=-\infty}^{\infty} P(m,n)e^{ia(mx+ny)-i\omega t} \tag{12:6}$$

Here $a = 2\pi/L_f$ and $w = 2\pi/T_f$, where L_f and T_f are the wavelength and period of the fundamental components in the expansion. The wavenumbers of each sinusoid are then $\kappa_x = am$, $\kappa_y = an$, $\omega = wk$. The first summation in (12:6) is more general, assuming no particular dispersion relationship. Because first-order water waves are constrained to follow the dispersion relationship expressed by (12:4), however, it is possible to simplify this to a double summation over two independent indices (or wavenumbers); the third is given in terms of the first two as

$$\omega_+ = \text{sgn}(am)(ag)^{1/2}(m^2 + n^2)^{1/4} = \text{sgn}(\kappa_x)g^{1/2}(\kappa_x^2 + \kappa_y^2)^{1/4}, \quad (12:7)$$

where $\text{sgn}(u) = \pm 1$ depending upon whether its argument u is \pm .

The first-order spatial/temporal average waveheight spectrum can now be written in terms of the Fourier coefficients of the expansion (Barrick, 1972):

$$\langle P(m,n,k)P(m',n',k') \rangle = \begin{cases} \left(\frac{2\pi}{L_f T_f} \right)^3 S(\kappa_x, \kappa_y, \omega) & \text{for } \begin{cases} m' = -m \\ n' = -n \\ k' = -k \end{cases} \\ 0 & \text{for other } m', n', k' \end{cases} \quad (12:8)$$

and

$$\langle P(m,n)P(m',n') \rangle = \begin{cases} \left(\frac{2\pi}{L_f} \right)^2 S(\kappa_x, \kappa_y) & \text{for } \begin{cases} m' = -m \\ n' = -n \end{cases} \\ 0 & \text{for other } m', n' \end{cases} \quad (12:9)$$

Again, using the first-order dispersion relationships it is possible to express the more general $S(\kappa_x, \kappa_y, \omega)$ in terms of the directional spatial spectrum, $S(\kappa_x, \kappa_y)$:

$$S(\kappa_x, \kappa_y, \omega) = S(\kappa_x, \kappa_y) \delta(\omega - \omega_+), \quad (12:10)$$

where ω_+ is given in (12:7) and $\delta(u)$ is the Dirac impulse function of argument u . The normalization here is such that the mean-square surface height is

$$h^2 \equiv \langle \xi^2 \rangle = \int_{-\infty}^{\infty} d\kappa_x \int_{-\infty}^{\infty} d\kappa_y \int_{-\infty}^{\infty} d\omega S(\kappa_x, \kappa_y, \omega) = \int_{-\infty}^{\infty} d\kappa_x \int_{-\infty}^{\infty} d\kappa_y S(\kappa_x, \kappa_y). \quad (12:11)$$

One can define a non-directional temporal spectrum $S(\omega)$ as follows:

$$S(\omega) = \int_{-\infty}^{\infty} d\kappa_x \int_{-\infty}^{\infty} d\kappa_y S(\kappa_x, \kappa_y, \omega) = \int_{-\infty}^{\infty} d\kappa_x \int_{-\infty}^{\infty} d\kappa_y S(\kappa_x, \kappa_y) \delta(\omega - \omega_+) \quad (12:12)$$

It turns out that oceanographers can conveniently measure $S(\omega)$ directly in a number of ways. For a review of these techniques, see Kinsman (1965). Many have reported detailed observations of $S(\omega)$ for wind-driven ocean waves. Others have attempted to fit empirical laws to these observations to relate the spectrum to the wind speed. One such set of observations is shown in (F12.3a) taken from Moskowitz (1964); these carefully selected spectra for deep-water waves are fully developed only at wind speeds below 30 knots, however. Moskowitz notes that on the open seas, the fetch and duration are rarely sufficient for winds above 30 knots that the sea will reach a fully developed condition. Thus observed spectra at these higher winds will usually be lower than models developed for fully developed seas.

Several semi-empirical models for wind-wave spectra enjoy popularity; among them are the Neumann-Pierson, the Pierson-Moskowitz, and the Phillips spectra (Kinsman, 1965). These differ chiefly in the form postulated for the lower-end cutoff. Because of its mathematical simplicity and for general lack of detailed information about the cutoff (which is observed to be quite steep in the absence of swell), we shall here employ the Phillips model when this function is needed for quantitative estimates. Furthermore, since observations indicate that a specific directionality is difficult to justify (Phillips, 1966; Munk and Nierenberg,

1969), we shall assume that the model is semi-isotropic. This means that all directions in the +x half-space are equally favored in amplitude by the waves. This model then has the form

$$S(\kappa_x, \kappa_y) = \begin{cases} B/[2\pi(\kappa_x^2 + \kappa_y^2)^2] & \text{for } \kappa \equiv \sqrt{\kappa_x^2 + \kappa_y^2} > g/u^2 \\ 0 & \text{for } \kappa \equiv \sqrt{\kappa_x^2 + \kappa_y^2} < g/u^2 \end{cases} \quad (12:13)$$

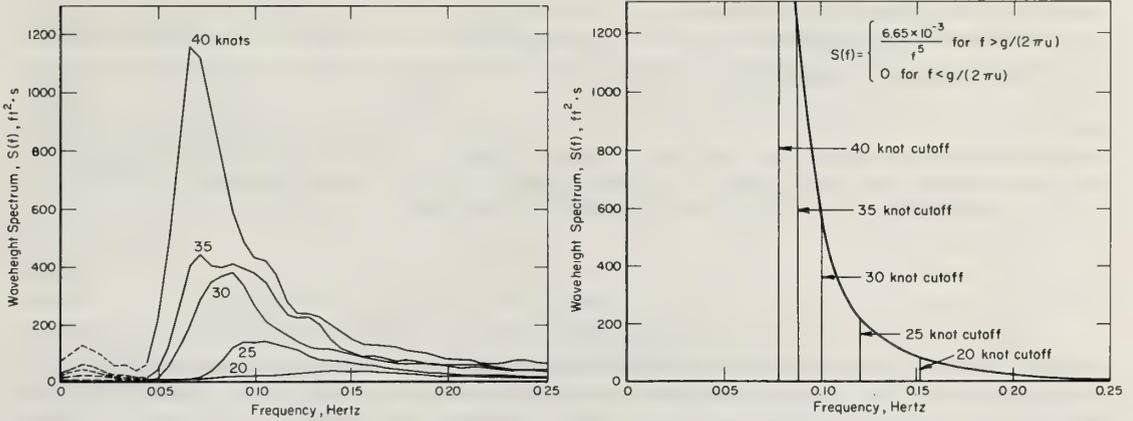


Figure 12.3 Measured and model waveheight temporal spectra. (a) Measured [after Moskowitz, 1964]; seas are not fully developed at winds above 25 knots. (b) Phillips model for fully developed seas.

where B is a dimensionless constant observed by Phillips (1966) and others to be approximately 0.005. Energy in this spectrum is spread symmetrically over both positive and negative wavenumbers. The above model shows the spectral saturation observed when the wavenumber κ exceeds the lower-end cutoff, given in terms of the wind speed u by g/u^2 . In this saturated region—often called the equilibrium region—the Phillips spectrum follows a κ^{-4} law.

From (12:13) and (12:12), the temporal non-directional version of the Phillips model is found to be:

$$S(\omega) = \begin{cases} g^2 B/\omega^5 & \text{for } \omega > g/u \\ 0 & \text{for } \omega < g/u, \end{cases} \quad (12:14)$$

where again the energy is distributed symmetrically for $\pm\omega$. Figure (12.3b) shows plots of (12:14) for comparison with spectra observed by Moskowitz.

Physically, our Phillips model implies that the wind does not affect the shape of the spectrum in the equilibrium region. As the wind increases it merely drives the cutoff lower, piling up more energy beneath the spectrum (and increasing the rms waveheight). This assumes of course that one waits until the seas are fully developed at a given wind speed. The reason for this effect on the lower-end cutoff can be explained simply. The longest (and hence fastest) waves which can be excited by the wind are those whose phase velocity, v , matches the wind speed u . The length of these waves is given by the dispersion relationship (12:3): $[(gL_{CO})/(2\pi)] = (g/\kappa_{CO}) = v^2 = u^2$. Solving this for κ_{CO} , we obtain g/u^2 , the sinusoid with the smallest (or cutoff) wavenumber which is excited by the wind with speed u .

12.1.5 RMS Height and Slope of Wind Waves

The rms (or significant) waveheight, as mentioned previously, is the essence of sea state. The rms slope of water waves, while not as directly indicative of sea state, appears frequently in scatter theories, especially specular point theories for microwave frequencies. Hence it is desirable to have quantitative estimates of the dependence of each on wind speed for fully developed seas.

The mean-square waveheight, h^2 , can be obtained directly from (12:11) by using the Phillips spectrum (12:13). This gives

$$h^2 = \frac{Bu^4}{2g^2} \text{ m}^2, \text{ or } h = .016u^2 \text{ m}, \quad (12:15)$$

where $B = 0.005$ (a dimensionless constant), u = windspeed (m/s), and g is the acceleration gravity (9.81 m/s^2).

The mean-square slope can be obtained in a similar manner. Since we have already assumed a *semi-isotropic* directional pattern for the Phillips spectrum, we have $\langle \zeta_x^2 \rangle = \langle \zeta_y^2 \rangle = \frac{1}{2} \langle \zeta_x^2 + \zeta_y^2 \rangle = \frac{1}{2} s^2$, where s^2 is the *total* slope at a point on the surface. We obtain a result for s^2 —after integration of the Phillips spectrum—which depends upon the upper (as well as the lower) bound on the spectrum. If one is interested only in the slope of the gravity waves, then it makes sense to take as the upper limit $\kappa_c \approx 0.038 \text{ m}^{-1}$, the boundary between the gravity and capillary wave regions. We then have

$$s^2 = B \ln \left(\frac{\kappa_c u^2}{g} \right), \quad (12:16)$$

Often in the specular point theories applicable at microwave and higher frequencies, the slopes of the capillary waves do in fact affect the magnitude of the scatter. In this case, the mean-square slope should include these capillaries. Phillips (1966) and Miles (1962) show that for high winds, about half of the mean-square slope comes from the capillaries, and one needs to add a term to (12:16) for $u > 5 \text{ m/s}$ to account for viscous dissipation. This correction term is

$$s_c^2 = B' \ln \left(\frac{\kappa_p}{\kappa_c} \right), \quad (12:17)$$

where $B' = 0.015$ (dimensionless), $\kappa_p \approx 3.2 u^{8/5} \text{ cm}^{-1}$, and as before $\kappa_c \approx 3.8 \text{ cm}^{-1}$.

A simpler empirical relationship derived from (F4.17) of Phillips (1966) can be obtained which includes the effect of capillary slopes as well as the slopes of the gravity waves. It is valid roughly for $1 \text{ m/s} < u < 15 \text{ m/s}$.

$$s^2 \approx 5.5 \times 10^{-3} u, \text{ or } s = 0.074\sqrt{u}. \quad (12:18)$$

12.2 MF/HF Radar Scatter from the Sea

One of the more thoroughly established radar techniques for remote sensing of sea wave characteristics uses frequencies in the MF and HF regions. Recent quantitative theories, confirmed by a variety of experimental configurations, lend considerable credence to the concept. We review first the physical mechanism and theoretically predicted echo strength, and then apply these results to several monostatic and bistatic concepts at MF and HF. Supporting experimental data for these techniques is presented where available.

12.2.1 Predicted Magnitude and Physical Nature of Sea Echo at MF/HF

Sea echo at frequencies below VHF has been observed by radars since World War II. Crombie (1955) appears to have been the first to correctly deduce the physical mechanism responsible for this sea scatter. Based upon HF experimental observations of the backscatter Doppler signal spectrum, he noted that—in contrast with a typical noiselike clutter—the sea echo always appeared at a discrete frequency shift above and below the HF carrier. These discrete Doppler shifts could not be produced by all of the ocean waves illuminated by the radar, since according to (12:3) waves of different lengths move at different velocities and hence would produce echoes at many Doppler shifts. Thus, working backwards and calculating the ocean wave

velocity from the observed discrete Doppler shift, and then the length of the ocean wave traveling at this velocity, he arrived at the following rather startling result: The only ocean wave from the entire spectrum present which produces backscatter at HF has a wavelength precisely one-half the radar wavelength and is moving directly toward and/or away from the radar. The observed Doppler shift of the sea return was seen to increase with the square root of the carrier-frequency—rather than in direct proportion, as with a discrete moving target—further confirming this explanation (following the square-root relationship between velocity and length of gravity waves, as given in (12.3)). Hence the experimentally deduced mechanism was seen to be “Bragg scatter”, the same phenomenon responsible for scatter of X-rays in crystals and light rays from diffraction gratings and holograms.

Quantitative theoretical analyses of the scatter problem lagged these experimental deductions by several years. Peake (1959) appears to have been the first to reduce the classic statistical boundary perturbation theory of Rice (1951) to σ^0 , the normalized average scattering cross section per unit area for a slightly rough surface. Barrick and Peake (1968) noted that this result, when interpreted, shows that scatter is produced by the Bragg mechanism, in agreement with Crombie’s deductions. Based upon a deterministic analysis of backscatter from sinusoidal waves, Wait (1966) independently obtained a result which was explainable via Bragg scatter.

No attempt was made until very recently to apply these scatter theories to the sea, which, as we have seen in the preceding section, has a unique waveheight spectrum and simple first-order dispersion relationship between spatial and temporal ocean wavenumbers. Barrick (1970, 1972) and Crombie (1971) both have obtained quantitative predictions for the scattered signal spectrum for sea echo, including the temporal variation and the dispersion relationship for the ocean waveheight. The results and notation of Barrick are somewhat more general and will be employed in this chapter; Crombie’s solution for backscatter agrees both quantitatively and qualitatively with Barrick’s, serving as an independent check.

The technique used by Barrick was initially applied by Rayleigh to scatter of acoustic waves from a sinusoidal surface. It was generalized by Rice to permit the analysis of the average electromagnetic signal intensity scattered from a randomly rough surface. Basically, one employs a Fourier series expansion for the surface, as given in (12:6), and then expands the three components of the electromagnetic field above the surface into the same type of series with the same wavenumbers (am, an wk), but with unknown coefficients. These coefficients are then determined by enforcing the boundary conditions at the surface. The fields at the boundary are expanded in a perturbational manner, permitting an ordering of the terms and a straight-forward solution for the unknown field coefficients. Mathematical details are found in Rice (1951), Peake (1959), and Barrick (1970, 1972).

This boundary perturbation approach requires the assumption of the following limitations in order to be *mathematically* valid: (i) the height of the surface must be small in terms of the radio wavelength, (ii) surface slopes must be small compared to unity, and (iii) the impedance of the surface medium must be small in terms of the free space wave impedance. These conditions are all satisfied by the sea below mid-VHF.

The solutions obtained from the Rice perturbation technique possess some similarity to those obtained earlier by Davies (1954) for a slightly rough surface using a physical optics technique. The perturbation results are superior, however, for two reasons: (i) they contain polarization dependence and correctly predict near-grazing scatter for vertical polarization, whereas physical optics does not, and (ii) they are mathematically valid in the low-frequency limit (as wavelength approaches infinity), whereas the physical optics approximation will eventually fail its inherent requirement that surface radii of curvature be much larger than wavelength.

Before giving the solutions for the scattering coefficients of the sea, we first review the radar range equations for average received power and its spectral density scattered from a patch of sea of area dS :

$$\left. \begin{array}{l} dP_R(\omega) \\ dP_R \end{array} \right\} = \frac{P_T G_T G_R \lambda^2}{(4\pi)^3 R_R^2 R_T^2} F_T^2 F_R^2 dS \times \left\{ \begin{array}{ll} \sigma(\omega) & W/\text{rad/s} \\ \sigma^0 & W, \end{array} \right. \quad (12:19)$$

where P_T is the transmitted power, R_R and R_T are the ranges from the scattering patch dS to the receiver and transmitter respectively, and λ is the radar wavelength. The quantities F_T and F_R are the Norton attenuation factors between the patch and the transmitter and receiver for TM propagation near the surface; they account for any propagation losses greater than the normal free-space $(1/R^2)$ spreading losses, and hence approach unity for a perfectly conducting flat earth. One must be cautious in defining the antenna gains G_T and G_R in the direction of the scattering patch. For ground-wave or line-of-sight propagation to and from the patch, G_T and G_R must be the *equivalent* free-space gains of the antenna; that gain is less than its measured gain in the presence of the conducting ground by 6 dB. For example, a vertical quarter-wave monopole fed against the ground would have an equivalent free-space gain for use in (12:19) of -0.85 dB rather than $+5.15$ dB. For over-the-horizon ionospheric propagation to the patch, however, one employs the normal gains of the antenna measured in the presence of the ground (e.g., $+5.15$ dB for the quarter-wave monopole)†.

The actual average scattering cross section for the patch of sea within the radar resolution cell of area $dS(m^2)$ is then $\sigma^0 dS(m^2)$. Hence, σ^0 is the average scattering cross section of the sea per unit area. Its counterpart in the equation for received power spectral density is $\sigma(\omega)$, the average scattering cross section per unit area per rad/s bandwidth. The normalization used here is such that $\sigma^0 = \frac{1}{2} \int_{-\infty}^{\infty} \sigma(\omega) d\omega$.

Referring to (F12.4) which defines the incidence and scattering angles at the sea surface patch dS , we can write the following expressions for $\sigma(\omega)$ and σ^0 for *vertically incident and vertically scattered polarization* (Barrick, 1972):

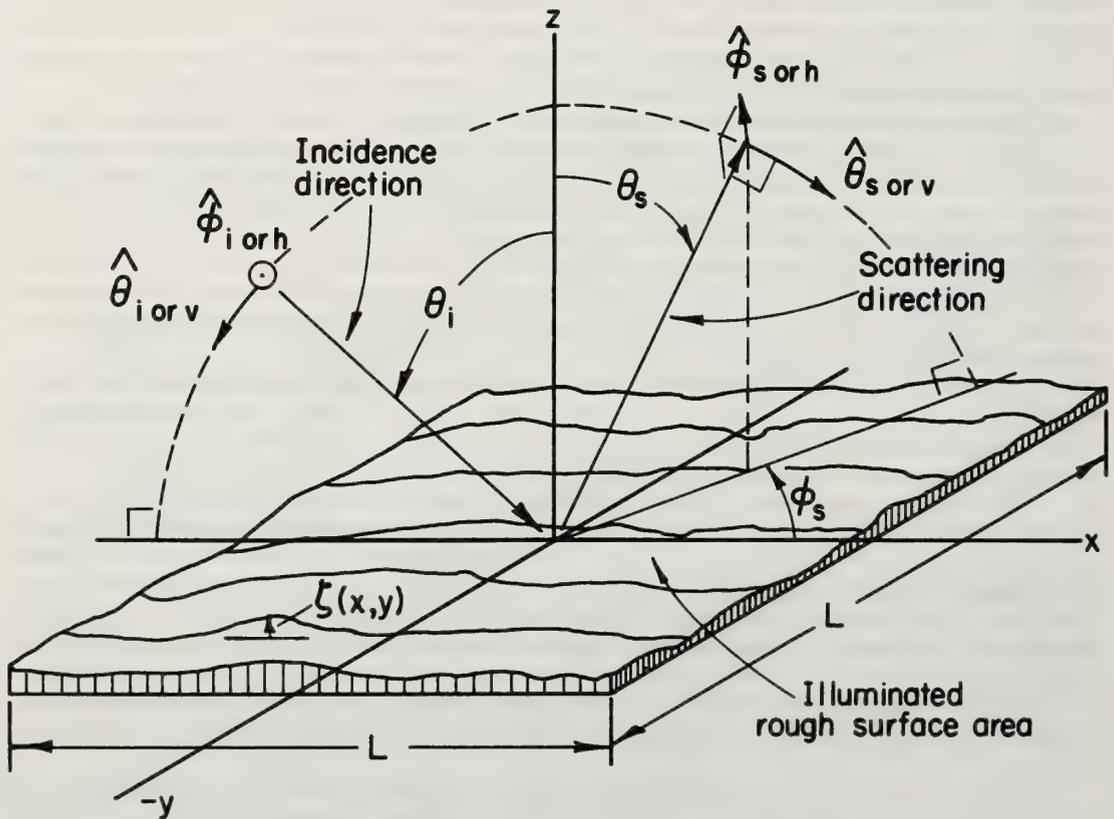


Figure 12.4 Local geometry near scattering patch.

† Any ionospheric attenuation losses could in this case be absorbed in a factor similar to $F_T^\dagger F_R^\dagger$.

$$\left. \begin{array}{l} \sigma_{VV}(\omega) \\ \sigma_{VV}^0 \end{array} \right\} = 4\pi k_0^4 (\sin \theta_i \sin \theta_s - \cos \varphi_s)^2 \times \begin{cases} W[k_0(\sin \theta_s \cos \varphi_s - \sin \theta_i), k_0 \sin \theta_s \sin \varphi_s, \omega - \omega_0] \\ W[k_0(\sin \theta_s \cos \varphi_s - \sin \theta_i), k_0 \sin \theta_s \sin \varphi_s] \end{cases}, \quad (12:20)$$

where $k_0 = 2\pi/\lambda$ is the radio wavenumber, ω_0 is the radiant carrier frequency, and $W(\kappa_x, \kappa_y, \omega)$, $W(\kappa_x, \kappa_y)$ are the first-order spatial/temporal and spatial waveheight spectra of the sea, respectively; they are related to the spectra defined in (12.1.4) by†

$$W(\kappa_x, \kappa_y, \omega) = 2^3 S(\kappa_x, \kappa_y, \omega); \quad W(\kappa_x, \kappa_y) = 2^2 S(\kappa_x, \kappa_y). \quad (12:21)$$

Equation (12:20) was derived assuming that the scattering patch is perfectly conducting. For the sea at MF/HF, this approximation is quite valid. The only seemingly confusing issue is the fact that when either the incident or scatter polarization state is vertical, and when the propagation angle for that polarization state approaches grazing, the scattered power remains finite. For a finitely conducting surface medium, the result analogous to (12:20) would always approach zero at grazing. This apparent difference is reconciled by Barrick (1972); the effect of finite conductivity for vertical polarization is separated from the scattering cross section and expressed as the Norton attenuation factors, F_T and F_R , in (12:19). Since vertical polarization is the only mode which can propagate efficiently as a ground-wave near the sea at MF/HF, most experiments would logically employ vertical (or TM) if one or the other paths to the scatter patch grazes the sea surface. Hence, one can handle the analysis of such an experiment by treating the sea surface patch as perfectly conducting, and then accounting for the finite conductivity by employing the Norton factors F_T and/or F_R , depending upon whether the incident and/or scatter states are vertically polarized.

If one or the other or both polarization states are not vertical, one can modify (12:20) in the following manner to give the other three cross sections and spectra for the sea††: $\sigma_{VH}(\omega)/\sigma_{VH}^0$, $\sigma_{HV}(\omega)/\sigma_{HV}^0$, and $\sigma_{HH}(\omega)/\sigma_{HH}^0$ are obtained by replacing the factor $(\sin \theta_i \sin \theta_s - \cos \varphi_s)^2$ in (12:20) by $(\cos \theta_i \sin \varphi_s)^2$, $(\cos \theta_s \sin \varphi_s)^2$, and $(\cos \theta_i \cos \theta_s \sin \varphi_s)^2$, respectively. Thus the dependence of scatter upon the nature of the roughness is the same for any polarization state; it is contained in the surface height spatial/temporal spectrum.

Physically (12:20) is interpreted as follows. The ocean spatial wavenumbers (κ_x, κ_y) which are producing scatter are given in terms of the radio wavenumber, k_0 , and observation angles θ_i , θ_s , φ_s by $\kappa_x = k_0(\sin \theta_s \cos \varphi_s - \sin \theta_i)$ and $\kappa_y = k_0 \sin \theta_s \sin \varphi_s$. These, however, are precisely the wavenumbers required of a diffraction grating which is to scatter a wave incident from θ_i into a direction θ_s , φ_s . Hence, the theory shows that the ocean surface produces scatter by the simple Bragg mechanism, which confirms the experimental deductions of Crombie (1955). Furthermore (12:20) implies that, in order to measure the directional spectrum of the sea, one can measure the sea echo (i.e., $\sigma(\omega)$ or σ^0) and vary k_0 , θ_i , θ_s , or φ_s in whatever manner is most convenient experimentally. Different schemes which vary one or more of these quantities are to be examined in the following subsections.

Having the radar range equation (12:19) and the expression for the sea scatter cross sections (12:20), one can now analyze any monostatic or bistatic configuration by integrating (12:19) over the area illuminated within the radar range cell and/or beam. Examples will be considered later. To obtain estimates for the received sea-scatter power magnitude for quantitative system design, one can employ the Phillips spectrum (12:13) in (12:20). Use of this spectrum for sufficiently high wind speeds (such that $u^2 > gk_0 \times \sqrt{\sin^2 \theta_i - 2 \sin \theta_i \sin \theta_s \cos \varphi_s + \sin^2 \theta_s}$) provides an upper limit on received power; for winds and seas which are lower, the received power will be always less than this amount.

† We apologize for the inconsistency in notation for the waveheight spectrum. Unfortunately, oceanographers independently established the convention using S, while scattering analysts adopted the Rice convention based upon W. Both are currently found in the literature, depending upon the discipline preferred by the user. Consequently, we employ both here and give the connection between them to facilitate reference to other works.

†† The first subscript always refers to the polarization state of the incident wave, while the second denotes the state for the scattered wave of interest.

12.2.2 Backscatter MF/HF Experiments

We consider in this section possible backscatter experiments employing MF/HF radars. These are in almost all cases either surface-based ground-wave or ionospheric sky-wave configurations. In both situations, $\varphi_s \rightarrow \pi$ and θ_i, θ_s are sufficiently close to grazing (i.e., within 20°) that the $\sin \theta$ factors appearing in (12:20) can be replaced by unity. We then have

$$\left. \begin{array}{l} \sigma_{VV}(\omega) \\ \sigma_{VV}^0 \end{array} \right\} = 16\pi k_0^4 \times \left\{ \begin{array}{l} W(-2k_0, 0, \omega - \omega_0) \\ W(-2k_0, 0) \end{array} \right. \quad (12:22)$$

It is understood of course that the waveheight spectrum wavenumbers κ_x and κ_y (in the x,y directions) are defined *at the scattering patch* dS . Hence, as one integrates (12:19) over the surface S , (12:22) implies that the x-y axis at the patch remain constant with respect to the radar line of sight, and therefore must rotate as the line of sight changes in azimuthal position on the ocean. Thus, as one swings the radar beam by 90° , he is not only looking at a different patch of ocean, but with the κ_x, κ_y wavenumber positions in (12:22) interchanged.

Let us first calculate $\sigma_{VV}(\omega)$ and σ_{VV}^0 based upon the fully developed Phillips spectrum model, and compare these predictions with experimental evidence. Using (12:13) in (12:22), we have

$$\begin{aligned} \sigma_{VV}(\omega) &= 4 \times 10^{-2} \delta(\omega - \omega_0 + \sqrt{2gk_0}) , \\ \sigma_{VV}^0 &= 0.02 = -17\text{dB} . \end{aligned} \quad (12:23)$$

It was initially assumed that waves were traveling only in the $+x$ direction (away from the radar). If waves are also moving into the $-x$ half-space (toward the radar), then we have an impulse function at $\omega = \omega_0 + \sqrt{2gk_0}$, as well as the one at $\omega = \omega_0 - \sqrt{2gk_0}$ shown above. Thus one sees that in general, according to the first-order theory, all of the energy backscattered is contained at two discrete Doppler shifts ($\pm\sqrt{2gk_0}$) from the carrier. Secondly, the magnitude of σ_{VV}^0 , the average backscattered cross section per unit area, has as its upper limit -17 dB, as defined according to (12:19).

Let us now compare both of these predictions with experimental evidence based upon ground-wave radar configurations. One set of recent ground-wave measurements of sea backscatter was made by Headrick of the Naval Research Laboratory (Barrick, 1972) at 10.087 MHz, in which he obtained measurements of σ_{VV}^0 . In the experiment, two vertical monopoles were located near Annapolis, Maryland, on the upper Chesapeake Bay. Spectral processing permitted separation of water-wave scatter from stationary ground clutter echoes. The signal format used provided a 20 nmi range resolution cell. The Norton attenuation factor $F_R (=F_T)$ was calculated for four range cells at different distances on the bay using the pertinent water conductivity (i.e., ~ 2 mho/m).

Data were recorded and processed on February 4, 1969, a day on which a moderate wind was blowing from the north. Waves receding from the radar were observed to be stronger due to the wind, and water waves of the Bragg scatter length $\lambda/2$ (15 m in this case) were estimated to be fully developed. The average received power from the water was processed at four ranges down the bay: 45, 55, 67, and 75 nmi. Propagation to all of these points was via groundwave since they were all below the radio horizon; thus one must compare measurements with σ_{VV}^0 at grazing incidence, as given in (12:22) or (12:23). With the water area within each resolution cell (i.e., dS) estimated from maps of the bay, this factor—as well as the Norton attenuation factors—were removed from the radar equation. This yielded experimental values for σ_{VV}^0 of -17 dB at all four ranges†.

The fact that the 15 m long water waves were fully developed (only a 9.4 knot wind is required to arouse waves of this length) means that the backscatter might have been expected to approach the Phillips saturation estimate in (12:23) as an upper limit. The agreement between measured and predicted values of σ_{VV}^0 not only lends credence to the theory, but confirms the oceanographic estimate of the "Phillips saturation constant", $B = 0.5 \times 10^{-2}$ used in (12:13).

† Headrick employs the actual antenna gains rather than their effective free space gains. Hence his reported values of -29 dB with (12:19) correspond to σ_{VV}^0 of -17 dB by our definition -6 dB caused by each antenna.

As further evidence of the validity of the first-order theory for ocean-wave scatter, we cite recent HF measurements by Crombie et al. (1970) from Barbados Island in the West Indies. Again the antennas were located near the water so that propagation to ranges beyond the horizon was via ground wave. In this case we examine Crombie's received signal spectrum; a very high spectral resolution of 0.002 Hz was obtained with digital signal processing. Backscatter was received with broad-band vertical monopole antennas from the half-space toward the east.

Shown in (F12.5) are the relative received power spectra measured simultaneously on August 15, 1969, at 2.9 and 8.37 MHz from the range cell at 45 km. Coherent processing at a 0.5 Hz offset (removed in the figures here) permits both negative and positive shifts above the carrier to be observed. The first-order peaks (corresponding to our impulse functions in (12:23)) occur as predicted at ± 0.174 Hz from the 2.9 MHz carrier and ± 0.296 Hz at 8.37 MHz. The relative strength of the positive spike over the negative spike at both frequencies agrees with the dominant wind direction in this area; trade winds from the east should excite west-moving water waves, producing a positive Doppler shift. Lesser spikes in the records at 0.0 Hz, +0.25 Hz for 2.9 MHz and at +0.42 Hz for 8.37 MHz are attributed by Crombie as due to higher-order hydrodynamic and electromagnetic contributions. Theory of such processes is examined in a later subsection.

We turn attention now to two ground-wave backscatter experiments which can be used to measure the waveheight spectrum of the ocean. In both cases, one must vary the frequency (and hence the Bragg wavenumber, $2k_0$) which samples the ocean waveheight spectrum in (12:22) over its significant lower end. If we assume that the spectrum cuts off at a wavenumber *somewhere* near g/u^2 ($g = 9.81$ m/s², u = wind speed, m/s,) then for higher winds and seas one must use lower frequencies. A plot of the backscatter radar frequency required versus wind speed is given in (F12.6); one of course should actually employ a frequency lower than this (by possibly 20 percent) to ascertain the spectral behavior below cutoff.

The first experiment, discussed and tested by Crombie (1971), employs azimuthally omnidirectional antennas on the coast. Using pulsed signals and a range gate set at 22.5 km from the radar, Crombie obtained the average received power at as many as eight frequencies, ranging between 1.7 MHz and 12.3 MHz. With such an experiment, one is simultaneously observing sea scatter from a semi-circular annulus, and it is assumed that the sea is relatively homogeneous over such a circle (i.e., that the directional ocean waveheight spectrum is essentially constant over 45 km). This assumption is reasonable for on-shore winds and waves; for off-shore winds, however, the limited fetch does not permit the waves nearer the shore to build up as high as those more distant. Crombie (1971) notes that even at 100 km, higher off-shore waves may still not be fully developed. Hence if one spectrally processes the signals and employs the energy only in the Doppler line *above* the carrier, the homogeneity assumption should be valid.

Since the antennas in this experiment are azimuthally omni-directional and since the received energy from everywhere in the semi-circular annulus occurs at the same \pm Doppler shift, it is not possible to obtain the *directional* waveheight spectrum. One can obtain the non-directional waveheight temporal spectrum, which was defined in (12:12) and exemplified in (F12.3). To do this and relate $S(\omega)$ to the received power, P_R , at carrier frequency f_0 , one must integrate the second version of (12:19) over the semi-circular annulus, using (12:22) and (12:12). The one-sided spectral result is

$$S(\sqrt{4\pi g f_0/c}) = \frac{4R^3}{\Delta R} \frac{P_R}{P_T} (G_T G_R F^4 \sqrt{4\pi g f_0/c})^{-1} \text{ m}^2/\text{rad/s} \quad (12:24)$$

where c is the free-space radio wave velocity and ΔR is the width of the range-resolution cell. It must be noted that several of the factors on the right side of (12:23) may vary with frequency, including antenna gains, transmitted power, and the Norton attenuation factors.

Shown in (F12.7) are several such temporal spectra reduced by Crombie from observations off Barbados Island in the West Indies (plotted versus Hertz rather than rad/s). In the upper two, he was able to compare the predicted significant waveheight (obtained from integrating the area under the spectrum) with laser profilometer measurements of significant waveheights; the agreement is good. In the bottom plot, Crombie shows also a Pierson-Moskowitz model spectrum predicted for a 20 knot wind and fully developed seas for comparison. Crombie cautions that the spectra to the left of the line marked "minimum observed frequency" are estimated; in the upper pair of records, the estimated portion is a substantial portion of the area under the curves. This points up the importance of employing a sufficiently low (MF) frequency if one wishes to obtain spectral detail in higher sea states.

A variation of the above backscatter experiment can provide the directional rather than merely the non-directional waveheight spectrum. The azimuthally omni-directional antenna is placed on a moving ship at

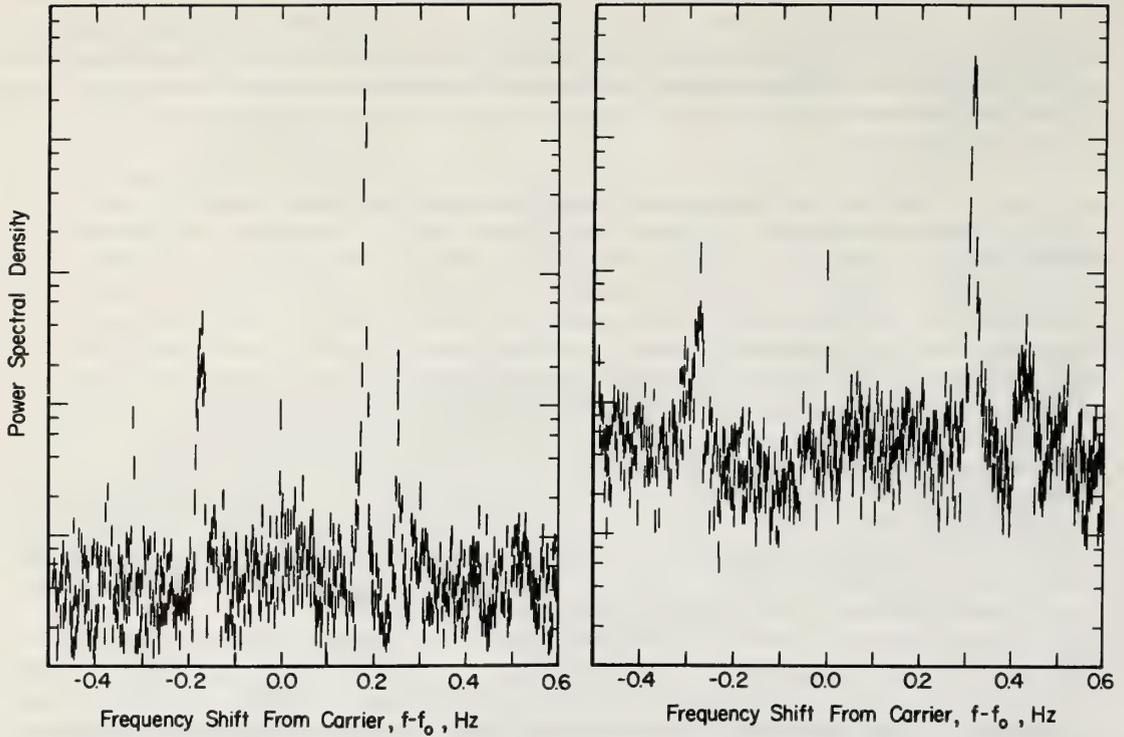


Figure 12.5 Measured backscatter signal Doppler spectra of HF sea echo (after Crombie et al., 1970) (a) 2.9 MHz carrier frequency (b) 8.37 MHz carrier frequency.

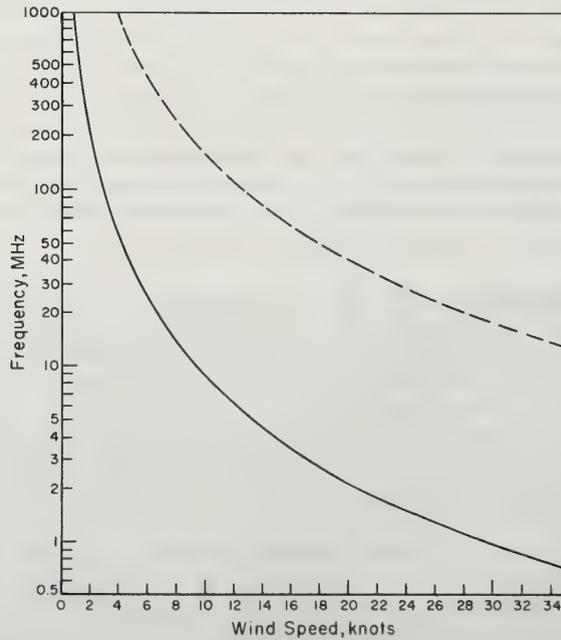


Figure 12.6 Solid curve gives frequency necessary to observe lower end (cutoff) of gravity wave spectrum for near-grazing backscatter; dashed curve gives frequency limit where slightly rough surface model fails mathematically for given wind speed.

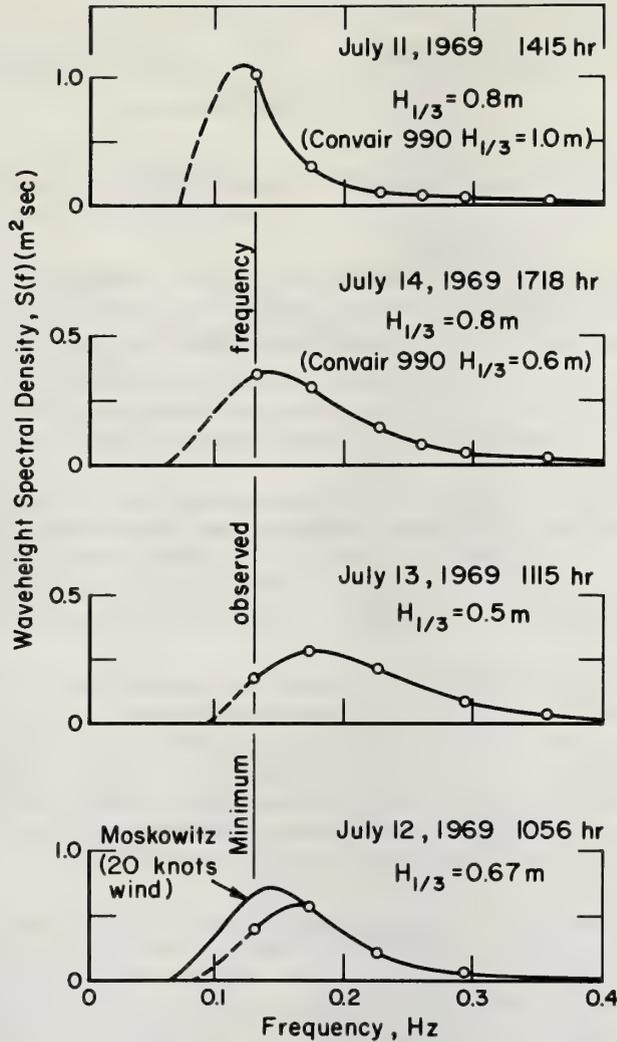


Figure 12.7 Temporal waveheight spectra deduced from MF/HF ground-wave backscatter observations (after Crombie, 1971).

sea. In this case the sea return does not appear at the previous single set of unique Bragg Doppler shifts around the carrier, for the ship's motion introduces a different Doppler bias on the return from each point around the range ring. This Doppler shift is $\omega_{sh} = 2\omega_0 v_{sh} \cos \theta / c$, where v_{sh} is the ship speed and θ is the angle from the ship's bow to a point on the range ring. We assume again that the ocean waves are moving predominantly into one half-space and that the "sea state" is homogeneous over the area of the ring. Then the one-sided power spectral density from a given element of area on the ring (i.e., $dS = R\Delta R\Delta\theta$) is found from (12:19) and (12:22) to be

$$dP_R(\omega) = \frac{2^4 P_T G_T G_R F^4 \Delta R \Delta \theta}{k_0^2 R^3} S(2k_0 \cos \theta, 2k_0 \sin \theta) \delta(\omega - \omega_0 \pm \omega_+ - \omega_m \cos \theta), \quad (12:25)$$

where $\omega_m = 2\omega_0 v_{sh} / c$ is the maximum ship-induced Doppler shift and $\omega_+ = \sqrt{2gk_0}$ is the Doppler shift from the gravity waves. To obtain the average power density spectrum, $P_R(\omega)$, from (12:25) we must integrate over θ . The resulting one-sided spectrum $P_R(\omega)$ is seen to consist of two "pedestals", each centered at $\omega_0 \pm \sqrt{2gk_0}$. Thus one can relate the spatial directional waveheight spectrum, $S(\kappa_x, \kappa_y)$ to, say, $P_R(\omega)$ within the positive

pedestal at $\omega_0 + 2gk_0$ as follows:

$$S(2k_0\eta/\omega_m, 2k_0\sqrt{1 - \eta^2/\omega_m^2}) = \frac{R^3 k_0^2}{2^4 \Delta R G_T G_R F^4} \sqrt{\omega_m^2 - \eta^2} \frac{P_R(\omega)}{P_T}, \quad (12:26)$$

where $\eta = \omega - \omega_0 - \sqrt{2gk_0}$, and where it is understood that $P_R(\omega)$ is non-zero only within the pedestal region, i.e., $|\eta| \leq \omega_m$.

Thus, by examining the received power density spectrum $P_R(\omega)$, one can obtain the directional spatial spectrum of ocean waves for spatial wavenumbers $\kappa_x = 2k_0 \cos \theta$, $\kappa_y = 2k_0 \sin \theta$ over all θ by noting that $\cos \theta = \eta/\omega_m$ and $\sin \theta = \sqrt{1 - \eta^2/\omega_m^2}$. As in the preceding experiment, one must vary the carrier frequency (or k_0) to sample the spectrum at a different *total* wavenumber, κ ($\kappa = \sqrt{\kappa_x^2 + \kappa_y^2} = 2k_0$). Also, it must be noted that there is an ambiguity with this technique, for waves propagating symmetrically with respect to the ship axis (i.e., $\pm\theta$) will be indistinguishable in their Doppler characteristics. Ship course changes can readily resolve this ambiguity, however.

Up to now, we have discussed only ground-wave backscatter radar sensors. These are limited in range to perhaps 100-200 km around the radar. Several investigators have recognized the potential of sky-wave (ionospheric) radars as remote ocean wave sensors (Tveten, 1967), (Ward, 1969), (Hasselmann, 1971); such shore-based radars could look at ocean areas as remote as 3200 km away. Since the normal range of elevation (or grazing) angles with ionospheric radars is usually 20° or less, (12:22) is adequate to describe ionospheric backscatter from the ocean.

As an example of ionospherically-propagated sea backscatter, we show in (F12.8) an averaged spectrum made by Barnum (1971) using a wide-aperture antenna at SRI in California, looking into the Pacific Ocean. Averaging was done over 16 range cells, each of which was approximately 1 km long by 10 km wide. At the particular time this record was made (~ 1500 local time, 30 March 1971), the ionosphere was quiet, and the resulting signals are relatively "clean" compared to similar signals under disturbed conditions. The frequency used was 25.75 MHz, the sea area illuminated was ~ 2700 km from the radar, and the elevation (or grazing) angle was about 6° . Use of coherent integration in the receiver/processor of 25.6 s permitted a spectral resolution of approximately 0.04 Hz.

The predicted first-order Bragg backscattered Doppler shifts at 25.75 MHz should occur at ± 0.518 Hz from the carrier. It should be noted from (F12.8) that ionospheric motion imposed an overall shift, or offset, from the carrier of about 0.07 Hz; such offsets are common with ionospheric radars. However, one should note that the distance between the two first-order Bragg lines (2×0.158 Hz in this case) is independent of such offsets, and can in fact be used to determine the offset. On the day this sea echo was recorded, Barnum ascertained from weather maps of the illuminated ocean area that the wind (~ 20 knots) was blowing in a direction 250° , and that seas with 12 ft significant waveheights were reported moving predominantly in a direction 260° . His radar beam was pointed in a direction 242° ; hence predominant wave movement was away from the radar, which is confirmed by the preponderance of the negative Doppler components over the positive components. Noted also by Barnum are the positions (relative to that of the first-order Bragg line) of other "higher-order" peaks in the scatter spectrum. Analysis of such higher-order contributions to scatter will be discussed in a subsequent section.

Ionospheric radars hold considerable promise for quick and simple routine monitoring of sea state hundreds of miles from the radar site. Detailed waveheight spectra and directional information on ocean-wave movements are probably not possible with such a radar however. The system sees only those waves in a given part of the ocean moving toward and away from the radar at about one-half the radar wavelength. At a given range and time of day, it is not possible to sweep the frequency of an ionospheric radar over a very wide band due to propagation conditions. During the daytime, one is usually restricted to about 17-30 MHz, while at nighttime one can operate between 9-17 MHz. Hence, the important lower end of the gravity wave spectrum can rarely be observed with an ionospheric radar via first-order Bragg scatter. Recent analyses and interpretations of the higher-order peaks in the received signal spectrum may provide additional useful information on the state of the sea not possible from the first-order Bragg lines. Signal fading and multi-moding in ionospheric propagation at times can produce quite confused spectral records, obscuring

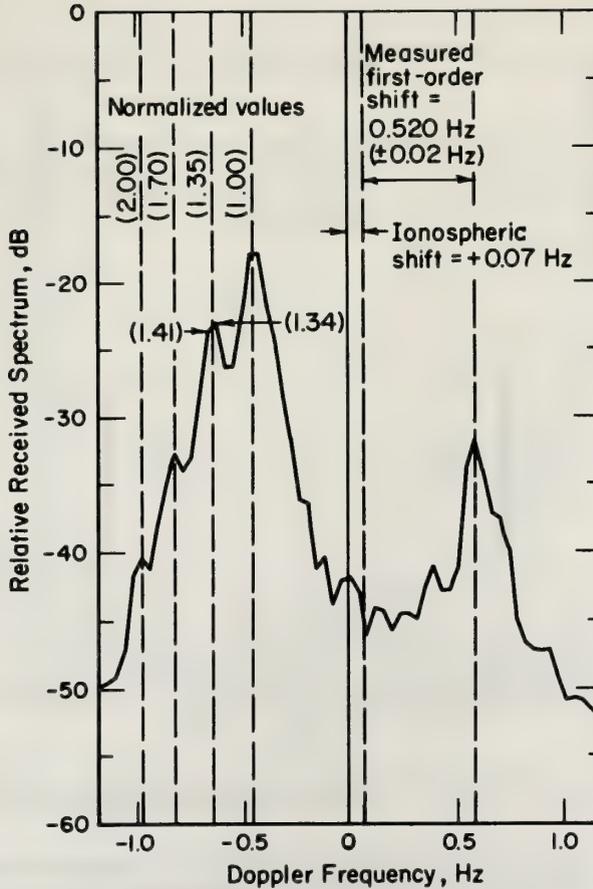


Figure 12.8 Ionospherically propagated sea backscatter spectrum at 25.75 MHz (after Barnum, 1971).

higher-order, less prominent peaks. Nonetheless, ionospheric radars—properly exploited—will eventually prove to be valuable remote sensors of sea state.

12.2.3 Bistatic Surface-Based MF/HF Experiments

Upon the discovery of the Bragg effect as the first-order mechanism responsible for sea scatter at MF/HF, several groups independently recognized the potential in bistatic concepts for ocean wave sensing. Unlike backscatter radars, a bistatic radar at a given frequency can measure the strengths of ocean waves of several wavelengths and directions of movement; furthermore, it is possible to relate the scattering ocean wavelengths and directions to the radar observables (namely time delay and Doppler shift). This can be seen by writing the first equation of (12:20) for a bistatic radar with both transmitter and receiver on the surface ($\theta_1 = \theta_s = \frac{\pi}{2}$). Using (12:10) in (12:20), we obtain:

$$\sigma_{VV}(\omega) = 2^8 \pi k_0^4 \sin^4\left(\frac{\varphi}{2}\right) S[k_0(\cos \varphi - 1), k_0 \sin \varphi] \delta\left[\omega - \omega_0 + \sqrt{2gk \sin(\varphi/2)}\right], \quad (12:27)$$

where $180^\circ - \varphi$ is the bistatic angle at the scattering patch between the lines to the transmitter and receiver. Thus one sees from (12:27) that the sea echo from different patches having different bistatic angles φ will return at different Doppler shifts. Thus it is possible to relate the echo strength at a given Doppler shift to the local sea waveheight directional spectrum evaluated at wavenumbers $\kappa_x = k_0(1 - \cos \varphi)$ and $\kappa_y = k_0 \sin \varphi$.

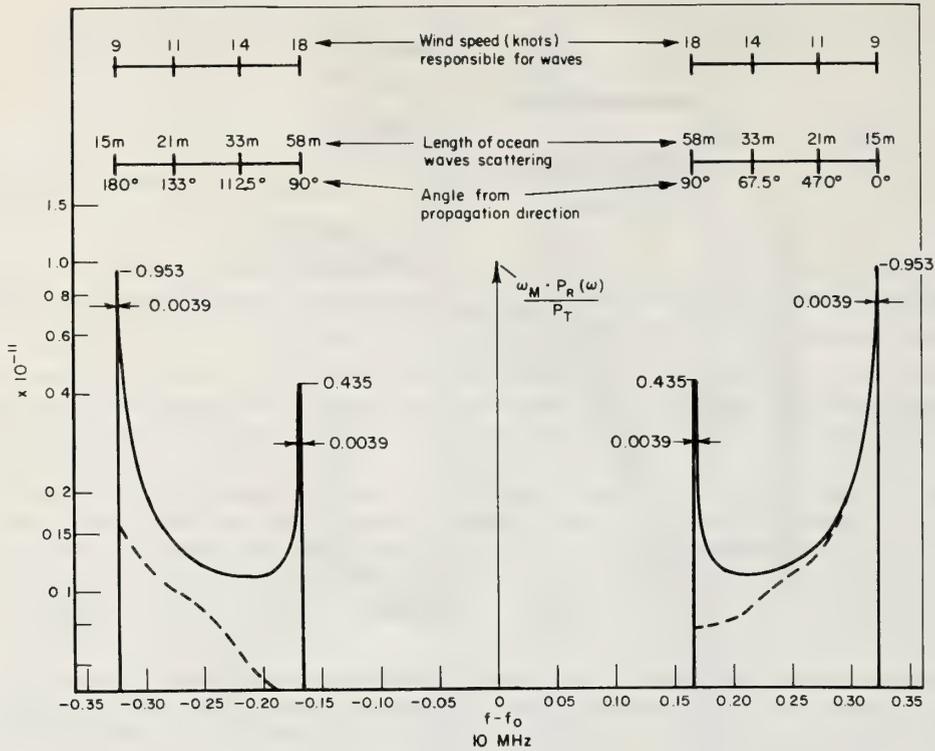


Figure 12.9 Geometry of bistatic ground-wave MF/HF radar.

Several investigators have examined such systems: Barrick (1970), Teague (1971) and Nierenberg and Munk (1969) all analyzed the relationships between time-delay and Doppler space (t, f_D) and the Bragg spatial wavenumber space (κ_x, κ_y). Signals arriving at a given time delay in a bistatic radar system are all scattered from an ellipse having the transmitter and receiver points as foci, i.e., $(R_T + R_R)/c = t$, where R_T and R_R are the transmitter-scatterer and scatterer-receiver distances. Such a configuration is shown in (F12.9). To illustrate this transformation, let us normalize the observed Doppler frequency to that for ocean-wave backscatter, i.e., $f_N \equiv f_D / (\sqrt{2gk_0}/2\pi)$. Since the observed Doppler shift, as seen from (12:27) is always greatest for backscatter, we have $0 \leq |f_N| \leq 1$. Nierenberg and Munk further define a quantity K equal to f_N^2 ; it is exactly equal to the magnitude of the spatial wavenumber of the ocean waves producing Bragg scatter at a given point divided by $2k_0$. Thus $0 \leq K \leq 1$, with $K = 1$ occurring for backscatter ($\varphi = \pi$). The other radar observable, time delay, is also normalized by dividing by t_d , the time delay experienced by the direct signal propagating between the transmitter and receiver separated by distance $2d$ ($t_d = 2d/c$): $S \equiv t/t_d$. Then $1 \leq S \leq \infty$. We also normalize the ocean spatial wavenumbers by dividing them by $2k_0$, i.e., $U = \kappa_x/2k_0, V = \kappa_y/2k_0$. Then the transformation relates K or f_N (Doppler shift) and S (time delay) for a point on the elliptical annulus to the ocean wavenumbers U, V responsible for scatter. Since $0 \leq |U|, |V| \leq 1$, we would like to be able to determine the ocean waveheight spectrum for all U, V in this range by observing K and S . Figures (12.10) and (12.11) show this transformation. The figures show that while it may be possible to find a family of time delays and Doppler shifts corresponding to a given wavenumber, U , it may not be possible to find a single time delay and Doppler shift if one specifies both U and V . To be able to do so would mean that the curve for U and the curve for V in terms of S and K intersect somewhere. It can be seen that many of the curves do not in fact cross (e.g., curves of U and V each near unity do not intersect). Furthermore, the confluence of both sets of curves to the upper right indicates that if they do cross in this region where they are nearly parallel, accuracy in the practical determination of U, V and S, K will suffer. These equations are explored in greater detail by Teague (1971).

We now proceed to find the general expression for the received signal spectrum $P_R(\omega)$ in terms of the spatial waveheight spectrum of the surface for a given time delay and Doppler shift. This is most easily done

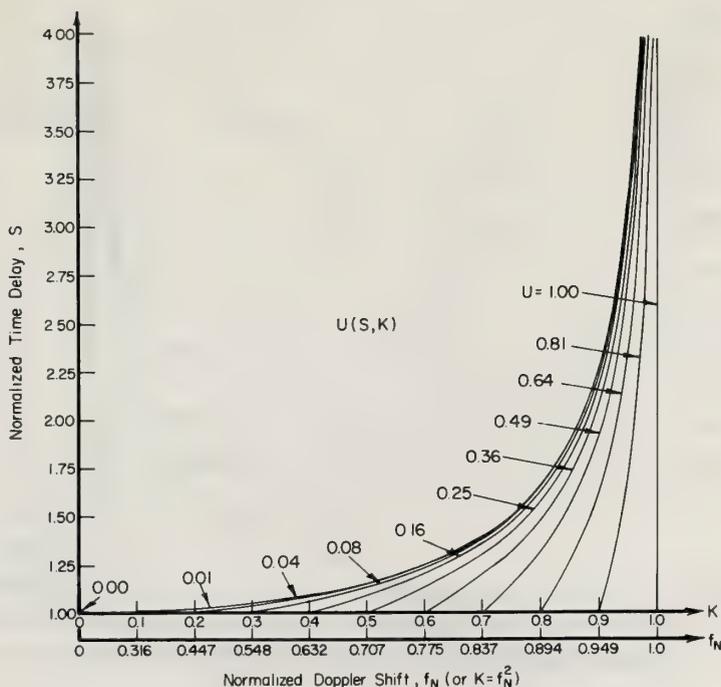


Figure 12.10 Transformation between U (normalized spatial x -wavenumber), and f_N (normalized Doppler shift) and S (normalized time delay) radar observables.

by converting to elliptic coordinates, μ , θ and then integrating over θ around a given elliptical annulus. We employ (12:27) in (12:19), but note that the wavenumbers in S of (12:27) are oriented with respect to the incidence direction at a given scattering patch; we must convert these to a space-fixed set of wavenumbers in order to determine a directional waveheight spectrum. The x -direction, and hence κ_x wavenumber direction, for this space-fixed frame is taken as the baseline connecting the receiver and transmitter (i.e., $\theta = 0$). As before, we assume that the sea is homogeneous over the area of an ellipse; this will restrict one in practice to transmitter-receiver separations no greater than about 200 km.

The elliptic coordinate μ can be defined in terms of time delay as $\cosh \mu = t/t_d = S$, where $t_d = 2d/c$ is the transmitter-receiver time of flight. The effective pulse length, τ , defines an elliptical annulus of width $\Delta\mu$ given by $\tau/t_d = \Delta\mu \sinh \mu$. The increment of area—or scattering patch—on the annulus is $dS = R_T R_R \Delta\mu d\theta$; as mentioned above, we intend to integrate over θ . The transmitter and receiver distances at μ , θ , are $R_T = d(\cosh \mu - \cos \theta)$ and $R_R = d(\cosh \mu + \cos \theta)$. We note that in general, transmitter and receiver antennas may be directional, implying that G_T and G_R are functions of μ and θ ; we assume here that both antennas are azimuthally isotropic for convenience†. Furthermore, the Norton attenuation factors F_T^2 and F_R^2 are complicated functions of range R_T and R_R . If frequency is restricted to the lower HF and MF region, and transmitter-receiver separation is kept less than 200 km, these factors are approximately unity, a fact which simplifies the integration; we make this assumption also.

We then perform the integration over θ . Two Doppler constants are defined in order to simplify the notation: $\omega_B = \omega_+ \equiv \sqrt{2gk_0}$ and $\omega_S \equiv \sqrt{2gk_0} \tanh \mu$. The former is the Doppler shift for backscatter, occurring here at the ends of the ellipses. The latter quantity is the Doppler shift which arises from ocean waves at the sides of the ellipse; hence it is the minimum Doppler shift observed for a given time delay. Defining Doppler shift as $\eta = \omega - \omega_0$, we then have a non-zero signal spectrum only for $\omega_S \leq |\eta| \leq \omega_B$. In terms of these constants and time delay, t (or $\mu = \cosh^{-1}(t/t_d)$), we obtain the following expression for the directional waveheight spectrum as a function of the observed one-sided received power density spectrum:

† Field tests of this concept to be discussed employed simple isotropic antennas.

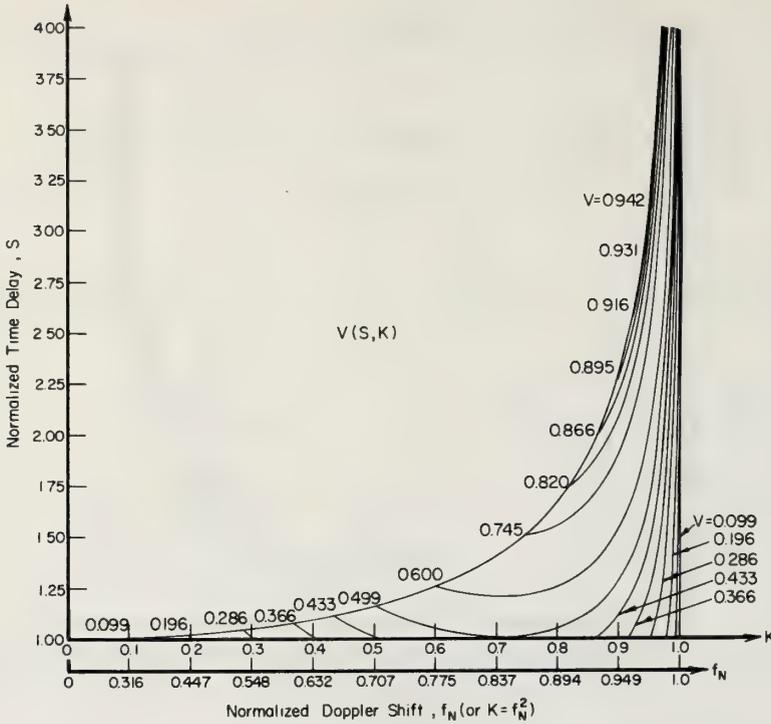


Figure 12.11 Transformation between V (normalized spatial y -wavenumber), and f_N (normalized Doppler shift) and S (normalized time delay) radar observables.

$$S \left[2k_0 \sinh \mu(\eta/\omega_B)^2 \sqrt{(\eta/\omega_S)^4 - 1}, 2k_0 \cosh \mu(\eta/\omega_B)^2 \sqrt{1 - (\eta/\omega_B)^4} \right] \\ = \frac{(k_{0d})^2 \sinh^2 \mu \sqrt{1 - (\eta/\omega_B)^4} \sqrt{(\eta/\omega_S)^4 - 1}}{2^5 G_T G_R \Delta \mu (\eta/\omega_B)^{11}} \cdot \frac{\omega_B P_R(\omega)}{P_T} \quad (12:28)$$

In order to obtain a feel for the appearance and magnitude of the received signal spectrum $P_R(\omega)$ with such a system, we employ the Phillips model for the waveheight spectrum of ocean waves, i.e., (12:13). We assume that wind speed is sufficiently great that all ocean waves which can produce scatter are present and fully developed. We employ as antennas two vertical ground-fed quarter-wave monopoles with equivalent free-space gains of 0.82; they are located $2d = 100$ km apart. Two frequencies are considered: 5 and 10 MHz. The signal pulse is rectangular in shape and taken as $12.5 \mu s$, and we consider the sea-scattered signal originating from the elliptical annulus at a delay $t = t_d + \tau$, i.e., one pulse length after the arrival of the direct pulse. The signal spectra calculated for this example are shown in (F12.12). Shown there also are the ocean wavelengths, ocean-wave directions with respect to the baseline, and the wind speeds required to excite the waves responsible for scatter at the indicated Doppler shift. The “ears” near the endpoints of the two Doppler pedestals are due to scatter from the “stationary” regions of the ellipse, viz., the sides and ends; they result from the radicals appearing in (12:28). The heights actually observed for these “ears” will depend upon the resolution of the spectral processor, since the area (or energy) contained under them is integrable and finite.

The total sea-scattered power received from this ellipse is approximately $1.5 \times 10^{-11} P_T$ at 5 MHz. This should be compared with a received signal power direct from the transmitter of $3.2 \times 10^{-9} P_T$. Hence the sea-scattered signal is down about 23 dB from the direct signal at this separation.

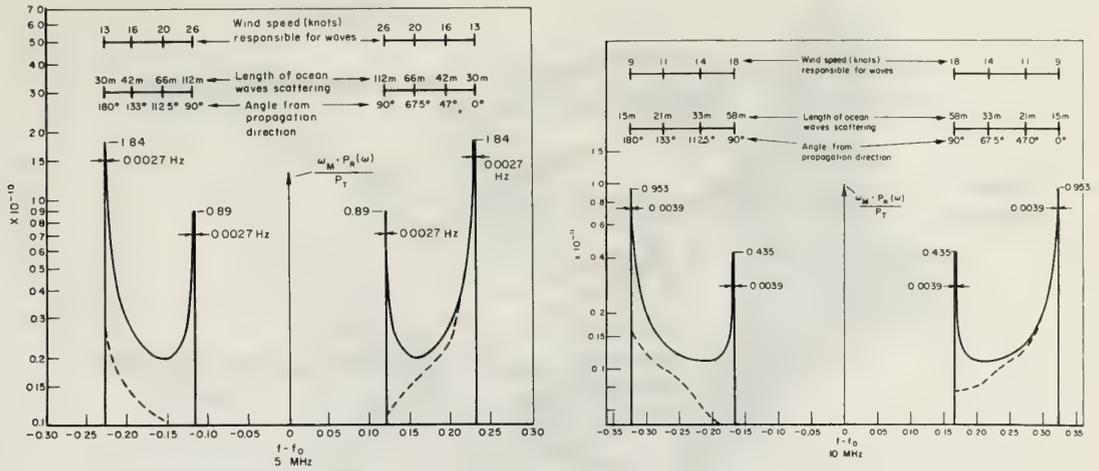


Figure 12.12 Predicted received signal spectra for bistatic sea scatter of F12.9. Solid curves were computed for fully aroused Phillips isotropic ocean wave spectrum. Dashed curves represent expected measurements for non-isotropic sea. (a) 5 MHz carrier frequency. (b) 10 MHz carrier frequency.

In practice, the received signal spectrum will be lower than that shown in (F12.12) for the semi-isotropic fully developed wind-wave model. The dashed curves show more typical shapes; the signal at the lower Doppler frequencies is likely to be weaker because the longer ocean waves responsible for their scatter are likely to be present less of the time.

Two groups have reported the results of such bistatic tests; in neither case were details of the ocean wave directional spectrum known with sufficient accuracy to allow complete quantitative comparisons. Wind and wave directions were available in an experiment reported by Barrick and Grimes (1970) at 5.8 MHz, permitting qualitative comparison and agreement; it will be described here. The other experiment was reported by Teague (1971) and Peterson, Teague, and Tyler (1970); using LORAN A signals (1.85 MHz), this experiment allowed the observation of quite long ocean waves (i.e., ~ 200 m wavelength), which are usually the essence of swell.

The experiment reported by Barrick and Grimes (1970) was conducted off Cape Kennedy, Florida, as shown in (F12.13). The transmitter was a buoy located 120 km from the receiving site on the coast. A binary phase-coded CW signal of basic interval $12.5\mu s$ was employed. The convolution of the signal with a replica of itself produces an effective pulse shape which is *not* uniform in amplitude over the $12.5\mu s$, as assumed in the example considered previously. Hence, the pattern of the illumination over the range gate one $12.5\mu s$ interval behind the reception of the direct signal tapers in amplitude from a maximum at the center of the elliptical annulus shown in (F12.13) to zero at the baseline. Therefore, one would not expect to see the innermost "ears" or the sharp inner cutoff shown in (12.12) for a rectangular pulse. Also, the receiving antenna had a beamwidth of about 12° , pointing in the direction of the buoy (the transmitter antenna on the buoy was azimuthally isotropic). Hence, the antenna pattern tended to enhance scatter from the regions close to the baseline and behind the buoy with respect to scatter from regions toward the sides of the ellipse. The average transmitted power was 10 watts, and the spectral processor resolution was better than 0.01 Hz.

Shown in (F12.14) and (F12.15) are two observed spectra (non-averaged) at 5.8 MHz. Coast Guard wind and wave data was available for this area. On March 19, 1970, wind and relatively mild waves were from

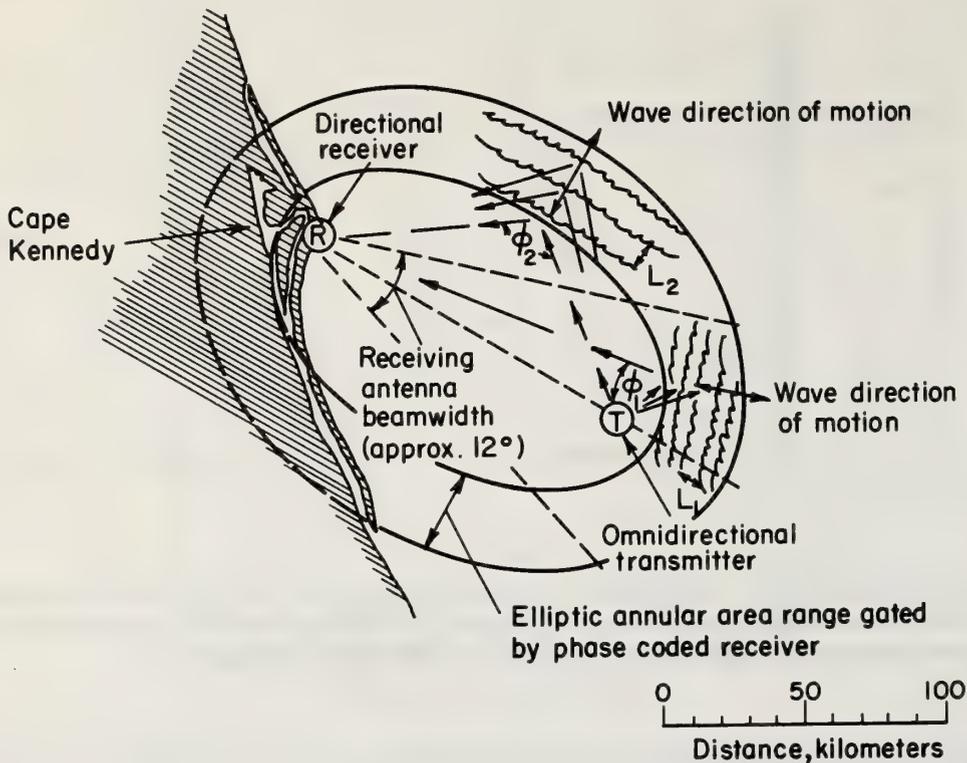


Figure 12.13 *Bistatic buoy-shore sea scatter configuration employed off Cape Kennedy.*

the east, as shown at the top of (F12.14). Thus one would expect approaching waves, predominantly from the region behind the buoy. Such waves approaching the buoy and baseline should produce positive Doppler shifts; such a spectral behavior was in fact observed, as shown at the bottom of (F12.14). On March 23, however, stronger, more variable winds blew from directions ranging between 220° and 315° . The "mean" wind and wave direction over that day was from the west, as shown at the top of (F12.15). Hence, strong negative Doppler shifts would be expected from the region behind the buoy. Again, this behavior was confirmed by the radar observations. Weaker positive Doppler shifts were produced by waves approaching the baseline, possibly near the receiving end of the path. It should be noted from these figures that the sea echoes fall within the bands predicted by the first-order Bragg-scatter theory. This represents a further confirmation of the theory and mechanism discussed here. Teague (1971) also observed that his echoes were always contained within the predicted "pedestal" region.†

12.2.4 Bistatic Surface-to-Satellite HF Experiment

As a final example of a remote sensing system based on first-order bistatic Bragg scatter, we examine an orbiting satellite receiver operating in conjunction with several ship or buoy transmitters. The geometry is shown in (F12.16). This concept was conceived and discussed by Barrick (1970), and is currently being developed by NASA into an experiment to be flown in the Skylab series of satellites (Ruck, et al., 1971). This experiment is capable of providing the directional waveheight spectrum (i.e., amplitudes and direction of ocean wave movements) near the transmitter; like preceding experiments, one must vary the transmitted frequency in some fashion over most of the lower HF band. Unlike previously

† The echo at zero Doppler shift is produced by land scatter behind the receiver, and also by some feed-through of the direct signal from the buoy.

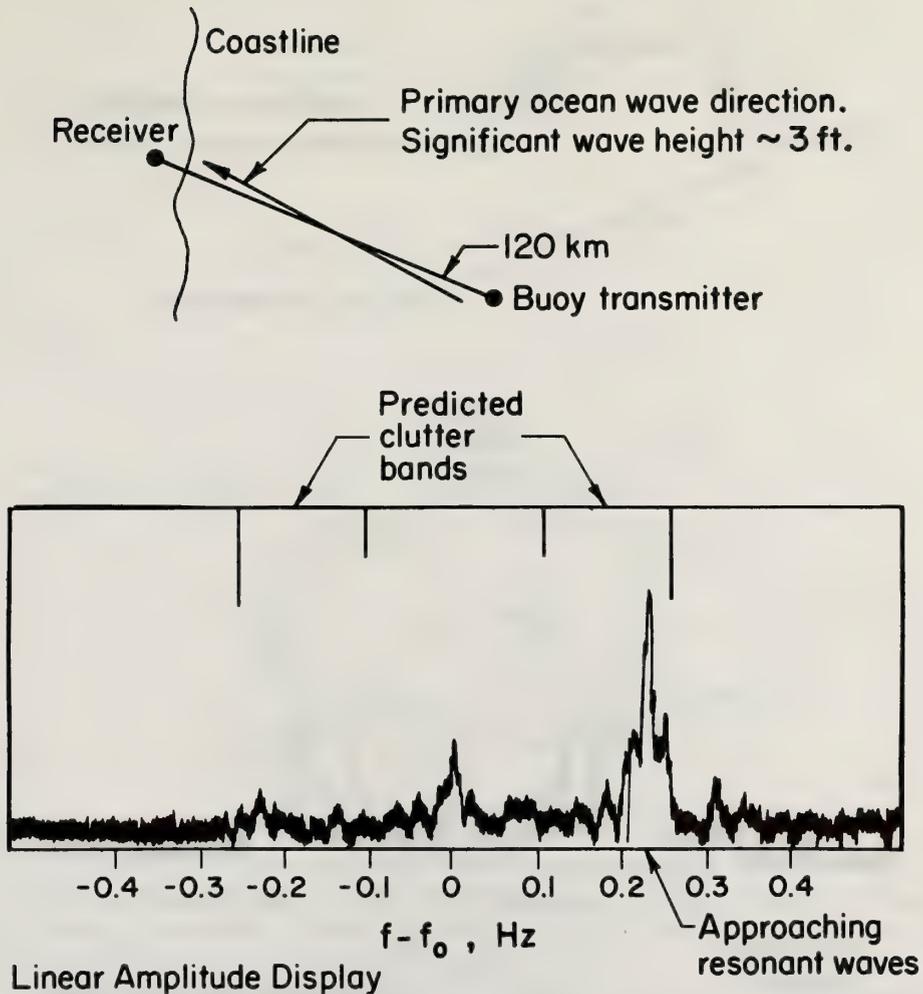


Figure 12.14 *Wind/wave directions and measured bistatic sea scatter Doppler spectrum at 5.8 MHz, 19 March 1970.*

discussed experiments, however, the receiver-processor need not be sensitive to the small Doppler shifts of the moving ocean waves, but rather obtains its directionality from the much larger Doppler shifts induced by satellite motion.

The transmitting antenna on the buoy or ship could consist of a simple vertical monopole (whip). The receiving antenna aboard the nonsynchronous satellite could be a pair of crossed loops or horizontal crossed dipoles. Short pulses (or equivalent coded signals) are employed. In general, ocean areas of constant time delay are represented by the intersection of confocal spheroids with the ocean surface; the foci of the spheroids are the transmitter and receiver points. The intersection curves on the surface are non-confocal ellipses. So as not to overcomplicate the mathematics, we consider here (for the sake of example) the case when the satellite is directly overhead, for then these elliptical annuli become circular annuli. The Doppler shift from each point on the ring due to satellite motion is different because of a varying overall range rate around the ring. Thus curves of constant Doppler shift are radial lines on the surface emanating from the transmitter. For the general case when the satellite is not directly overhead, these constant Doppler lines are hyperbolas intersecting the ellipses orthogonally; this case is analyzed in detail by Ruck et al. (1971).

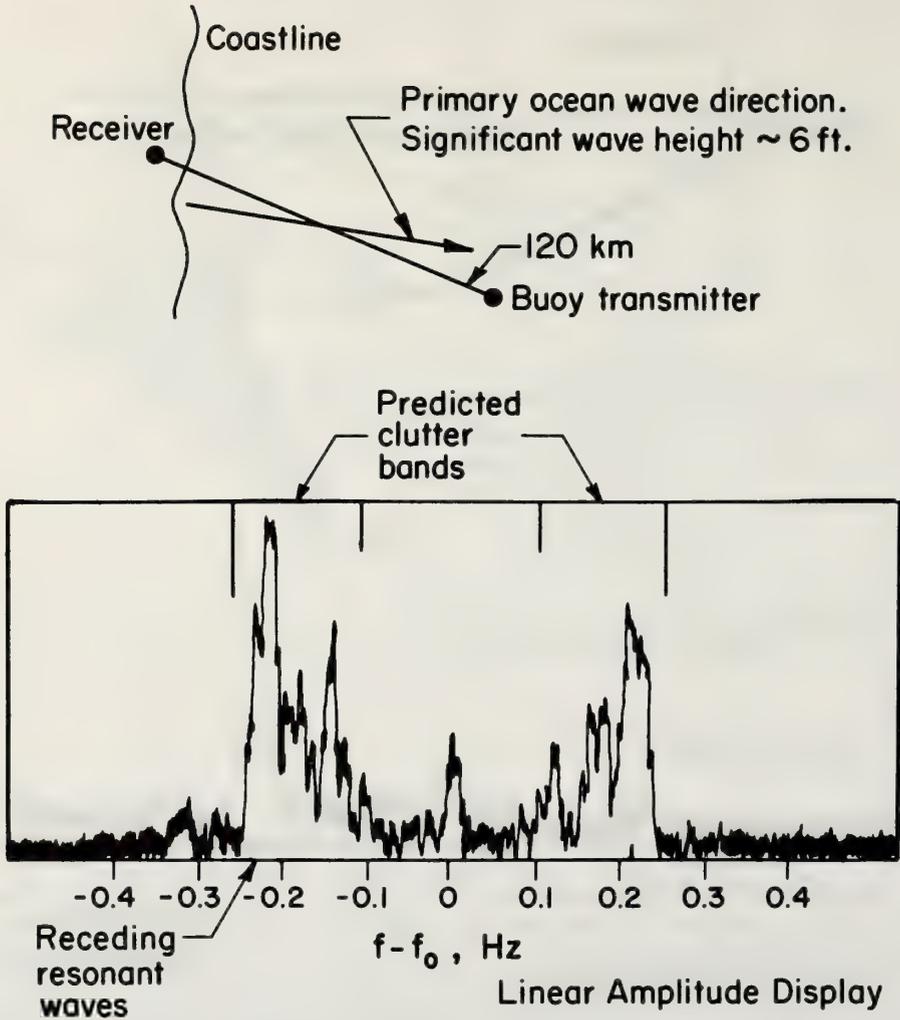


Figure 12.15 Wind/wave directions and measured bistatic sea scatter Doppler spectrum at 5.8 MHz, 23 March 1970.

In this case the bistatic angles to a point on the surface are $\theta_i = \pi/2, \theta_s = 0$. Hence the magnitude of the Bragg wavenumber of the scattering ocean waves is k_0 here, rather than $2k_0$ for near-grazing backscatter configurations. We again intend to fix the wavenumber directions in the surface height spectrum to the space around the buoy or ship, since the original scattering cross sections (12:20) were defined with respect to the local incidence directions at each point on the scattering ring. Hence we take the x -axis and κ_x direction as the horizontal line within the orbital plane. Thus the satellite-induced Doppler shift to any point on the ring of radius R_T at an angle φ with respect to the $+x$ -axis is $\omega_0 v_s (R_T/R_R) \cos \varphi/c = \omega_{dm} \cos \varphi$, where here the maximum Doppler is $\omega_{dm} = \omega_0 v_s (R_T/R_R)/c$, v_s being the velocity of the satellite, ω_0 the radian carrier frequency, and c the free space radio wave velocity. Using (12:20) in (12:19), the average received one-sided power spectrum for a patch on the ring ($dS = R_T \Delta R_T \Delta \varphi$) is found to be:

$$dP_R(\omega) = \frac{4P_T G_T G_R F_T^2 \Delta R_T \Delta \varphi}{k_0^2 R_T R_R^2} S(k_0 \cos \varphi, k_0 \sin \varphi) \delta(\omega - \omega_0 \pm \omega_+ - \omega_{dm} \cos \varphi). \quad (12:29)$$

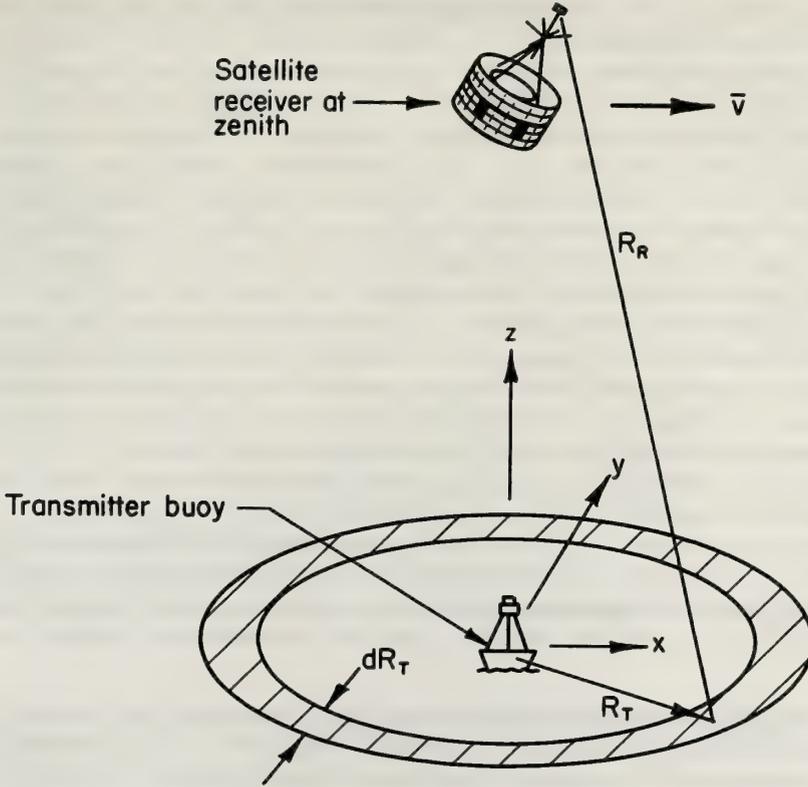


Figure 12.16 Bistatic HF surface-to-satellite radar.

Here the width of the ring, ΔR_T is related to the effective signal pulse width, τ , as $\Delta R_T = c\tau$. The direct signal from the transmitter to the satellite is observed at a delay $t_d \approx R_R/c$ (for $R_R \gg R_T$); hence the range R_T is given in terms of time delay $t = (R_R + R_T)/c$ as $R_T \approx ct - ct_d$.

In a manner similar to that following (12:25), we integrate (12:29) over φ to obtain the total signal spectrum from the range ring. The various factors preceding S in the above equation are assumed to be constant in this integration. As before, we assume that the sea height spectrum is homogeneous over the circle around the transmitter. Then we obtain an expression for the directional waveheight spectrum of the surface in terms of the received signal spectrum as follows:

$$S^{\prime}_{k_0 \eta / \omega_{dm}, k_0} \sqrt{1 - \eta^2 / \omega_{dm}^2} = \frac{R_T R_R^2 k_0^2}{4 \Delta R_T G_T G_R F_T^2} \sqrt{\omega_{dm}^2 - \eta^2} \times \frac{P_R(\omega)}{P_T}, \quad (12:30)$$

where $\eta = \omega - \omega_0 \mp \sqrt{gk_0}$. Normally, satellite-induced Doppler is so great that ocean-wave-induced Doppler can be neglected ($\omega_{dm} \gg \sqrt{gk_0}$); in this case, $\eta \approx \omega - \omega_0$. Thus by measuring the received sea echo signal spectrum at the satellite, and knowing its velocity, position, and orbital plane, one can obtain the directional waveheight spectrum for spatial wavenumbers $\kappa_x = k_0 \cos \varphi$, $\kappa_y = k_0 \sin \varphi$, for all φ by noting that $\cos \varphi = \eta / \omega_{dm}$ and $\sin \varphi = \sqrt{1 - \eta^2 / \omega_{dm}^2}$. One must vary the carrier frequency, however, to sample the spectrum at a different total wavenumber, $|\kappa|$ ($\kappa = \sqrt{\kappa_x^2 + \kappa_y^2} = k_0$). As before, there is an ambiguity, in that waves crossing the circle at $+\varphi$ cannot be distinguished from waves crossing the circle at $-\varphi$. Ruck et al. (1971) discuss several techniques for removing this ambiguity.

As an illustration of what the received signal spectrum will look like at 5 and 10 MHz for fully developed semi-isotropic waves describable by the Phillips spectrum (12:13), we consider the following example. The antenna on the satellite is assumed to be a half-wave dipole with gain 1.64, and the "free-space" gain of the quarter-wave monopole transmitting antenna is taken to be 0.82. The satellite is at an altitude of 300 km and moving with velocity 8000 m/s. We select a pulse length $\tau = 10 \mu\text{s}$, yielding a clutter ring width $\Delta R_T = 3 \text{ km}$. Let us select a time delay $t - t_d = 50 \mu\text{s}$, corresponding to $R_T = 15 \text{ km}$. Then we obtain the spectra shown in (F12.17). At the top of these plots, the angle φ of ocean wave directions producing the echo at that Doppler shift is given. The maximum Doppler shift from the satellite is $f_{dm} \approx 6.67 \text{ Hz}$ at 5 MHz, which is considerably larger than the ocean-wave Doppler $\sqrt{gk_0}$ of 0.161 Hz; hence the neglect of ocean-wave-induced Doppler shifts in this experiment seems reasonable. Also, there is no need for spectral processing resolution less than about 0.5 Hz, which alleviates the data handling requirements aboard the satellite.

The satellite should receive and process the direct signal as well as the sea-reflected signal. The direct signal will serve (a) as a time reference, (b) as a Doppler (frequency) reference, and (c) to calibrate and remove any unknown path loss through the ionosphere from the sea return. For the example considered in the preceding paragraph, the direct signal power received at the satellite is of the order of $10^{-9} P_T$ to $10^{-11} P_T$, depending upon the gain pattern of the transmitting antenna in the direction of the satellite. This compares with a total maximum received sea echo power of approximately $2.5 \times 10^{-12} P_T$ at 5 MHz.

The most serious limitations on a system such as this are imposed by the ionosphere. Orbital altitudes greater than 200 km will often not allow penetration of the lower HF frequencies through the ionosphere to the satellite. The F2 layer of the ionosphere is the densest and if the satellite is orbiting above it (i.e., above 300 km), then the following conclusions concerning ionospheric limitations were determined by Ruck et al. (1971).

(a) The operating frequency of the sensor must be confined to the range 3.5 to 30 MHz. (This permits sensing of ocean waves with lengths between 10 and 100 m.) Propagation conditions favorable to the system exist at night between 0 and 6 hr local time. At such times the minimum ionospheric penetration frequency ranges from 3.5 to 5 MHz depending upon the season and sunspot cycle. Operation throughout the rest of the day can take place at frequencies as low as 9-10 MHz.

Restriction of operation at the lowest frequencies to a six-hour period every day, however, may not limit the utility of the sensor for the following reason. At the lowest frequencies, the longest ocean waves are being observed (i.e., greater than 40 m). However, these longer ocean waves require greater times (i.e., of the order of 24 h) to build up and die down (T12.1). Thus the heights of these longer waves will not change appreciably over times less than a day, and their observation once a day should be sufficient.

(b) During normal ionospheric conditions, the (excess) absorption loss due to passage through the ionosphere will be less than approximately 15 dB providing the operating frequency exceeds the minimum penetration frequency by 0.5 MHz.

(c) The noise environment encountered by the satellite will be that due to cosmic noise, with a maximum effective noise temperature of about $4 \times 10^6 \text{ }^\circ\text{K}$ at 3 MHz.

Based on these loss and noise considerations, the study concludes that adequate signal-to-noise ratios can be obtained with average transmitter power output levels of the order of 10 W for a satellite in a 400 km orbit.

12.3 Second-Order HF Sea Echo

It was noted several places previously in this chapter (when comparing the first-order Bragg scatter theory with measured sea echo spectra) that the observed records often contain smaller—but non-negligible—peaks at Dopplers other than the first-order lines. Also, Barnum (1971) observes an overall higher "floor" under the sea spectra than would be expected from normal processor and noise clutter. He has confirmed that this "floor" is produced by sea echo by (a) looking at land scatter for comparison, and (b) shutting off the transmitter and observing the "floor" due to system and external noise. This "floor" and the higher-order peaks are not predicted from the first-order theory developed previously; one must go to a higher-order

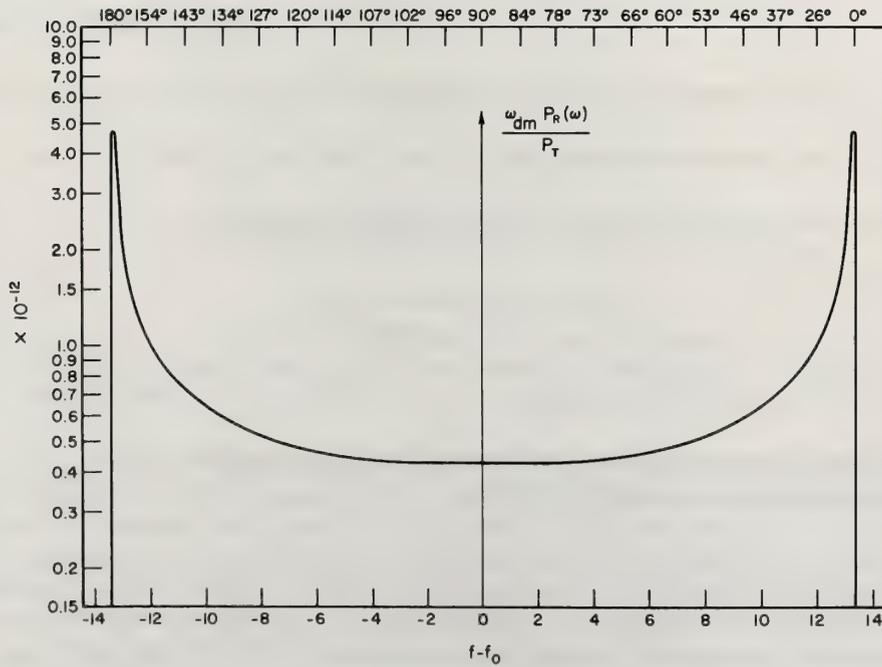
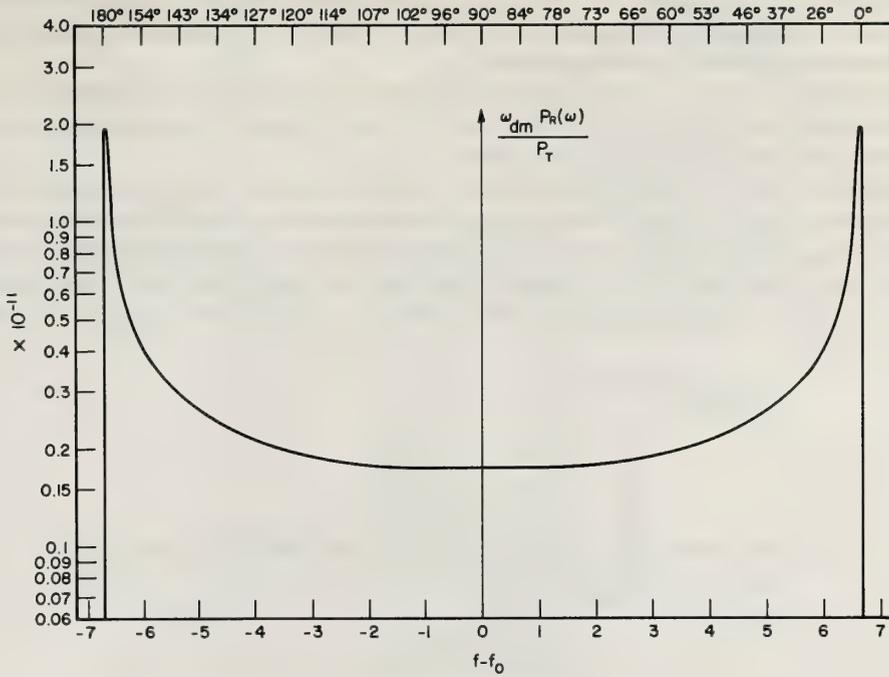


Figure 12.17 Predicted received signal spectra for bistatic sea scatter of F12.16; computed for fully aroused Phillips isotropic waveheight spectrum (a) 5 MHz carrier frequency (b) 10 MHz carrier frequency.

analysis of scattering and ocean wave interaction processes. Such spectral contributions could ultimately pose a limitation on the performance of any radar system designed for first-order Bragg scatter, and hence they should be qualitatively and quantitatively understood.

A more positive reason for examining these higher order sea echo spectral contributions is that they may in themselves provide additional information about the state of the sea. This suggestion was made by Hasselmann (1971), Crombie (1971), and Barrick (1971b).

Two obvious sources for this higher-order return suggest themselves: (a) the second-order terms for scatter from the Rice boundary perturbation theory, and (b) the second-order terms from the hydrodynamic equations describing the water surface height. To simplify matters here, we consider backscattering at grazing incidence with vertical polarization over a perfectly conducting sea. Then it can be shown that the average second-order backscatter cross section per unit surface area per rad/s bandwidth, $\sigma_{VV}(\omega)$, to be used with (12:19) is:

$$\sigma_{VV}(\omega) = 16\pi k_0^4 \iint_{-\infty}^{\infty} \left| \Gamma_T(\vec{\kappa}_1, \vec{\kappa}_2) \right|^2 \delta(\eta - \omega_1 - \omega_2) W(\vec{\kappa}_1) W(\vec{\kappa}_2) \, dpdq, \quad (12:31)$$

where $\vec{\kappa}_1 = (p - k_0)\hat{x} + q\hat{y}$; $\vec{\kappa}_2 = -(p + k_0)\hat{x} - q\hat{y}$; $\kappa_1 = |\vec{\kappa}_1|$; $\kappa_2 = |\vec{\kappa}_2|$; $\omega_1 = \text{sgn}(\kappa_{1x})\sqrt{g\kappa_1}$; $\omega_2 = \text{sgn}(\kappa_{2x})\sqrt{g\kappa_2}$; $\eta = \omega - \omega_0$ is the Doppler shift from the carrier; $\delta(x)$ is the Dirac impulse function of argument x ; and $W(\vec{\kappa}) \equiv W(\kappa_x, \kappa_y)$ is the directional spatial waveheight spectrum of the ocean, as defined in (12:21).

For the second-order electromagnetic contributions alone, Γ is found to be

$$\Gamma_{EM} = \frac{1}{2} (\kappa_{1x}\kappa_{2x} - 2\vec{\kappa}_1 \cdot \vec{\kappa}_2) / (\sqrt{\vec{\kappa}_1 \cdot \vec{\kappa}_2} + k_0\Delta), \quad (12:32)$$

where Δ is the normalized impedance of the sea surface, as discussed by Barrick (1971a).

The second-order hydrodynamic effects produce

$$\Gamma_H = -\frac{i}{2} \left[\kappa_1 + \kappa_2 + (\kappa_1\kappa_2 - \vec{\kappa}_1 \cdot \vec{\kappa}_2) (1 - 2\eta^2 / (\eta^2 - \omega_B^2)) (g/\omega_1\omega_2) \right], \quad (12:33)$$

where $i = \sqrt{-1}$ and $\omega_B = \sqrt{2gk_0}$ is the first-order Bragg Doppler shift. The total Γ_T which must be used in the integral to account for both types of second-order effects is

$$\Gamma_T = \Gamma_{EM} + \Gamma_H \quad (12:34)$$

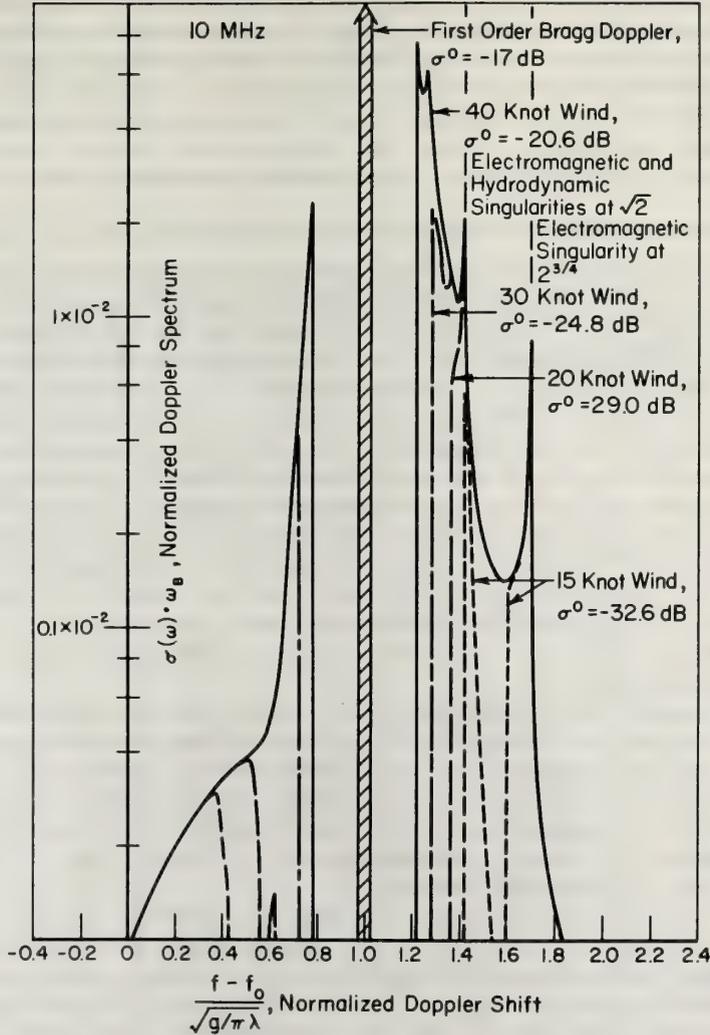


Figure 12.18 Predicted Doppler spectrum of first and second-order near-grazing sea backscatter at 10 MHz for propagation in upwind direction (Phillips semi-isotropic waveheight spectrum assumed).

The above integral clearly shows that a double-scatter Bragg process is responsible for the second-order sea return. The scattered radio wavenumber, $-k_0 \hat{x}$, is equal to $\vec{k}_1 + \vec{k}_2 + k_0 \hat{x}$, where the last term is the incident radio wavenumber. The frequency, ω , of the scattered field is identically $\omega_1 + \omega_2 + \omega_0$. In the case of the electromagnetic second-order effects, an ocean wavetrain with wavenumber \vec{k}_1 scatters the radio energy along the surface to a second wavetrain with wavenumber \vec{k}_2 , which redirects it back toward the source; the intermediate radio wave can be either propagating or evanescent. In the case of the hydrodynamic effects, two ocean wavetrains produce second-order ocean waves with wavenumbers $\vec{k}_1 \pm \vec{k}_2$; these latter ocean waves are not freely propagating because they do not satisfy the first-order gravity wave dispersion relationship, but they do produce radar scatter.[†]

[†] One of the two integration processes can be done in closed form, this resulting from the impulse-function occurring in the integrand. The remaining integration is done numerically because of the complex form of the integrand.

To illustrate the nature of both of these second-order effects at HF, we perform the integration indicated in (12:31) numerically† for the following example. The frequency chosen is 10 MHz, and we employ the Phillips semi-isotropic wind-wave model of (12:13) for $W(\vec{k})$, the first-order waveheight spectrum. Using different windspeeds as the parameter and considering propagation in both the upwind and crosswind directions, we obtain plots shown in (F12.18) and (F12.19) for the signal spectrum $\sigma_{VV}(\omega)\omega_B$. The first-order discrete Doppler lines are also shown for reference; they will be present at 10 MHz for the Phillips model at winds greater than 9-1/2 knots. Also shown is the normalized radar cross section for each spectrum, $\sigma_{VV}^0(\sigma_{VV}^0 = \frac{1}{2} \int_{-\infty}^{\infty} \sigma_{VV}(\omega) d\omega)$, where the spectra plotted here are two-sided, i.e., $\sigma_{VV}^0(-\omega) = \sigma_{VV}^0(\omega)$.

The figures illustrate that second-order continuous Doppler sidebands do occur and their magnitude will depend upon sea state. These sidebands contain continuous, integrable singularities (of the square-root-types) at positions of $\sqrt{2}$ and $2^{3/4}$ times the first-order Bragg line. The $\sqrt{2}$ singularity is due to both electromagnetic and hydrodynamic second-order effects. Electromagnetically, it is due to higher-order Bragg scatter, i.e., from ocean waves of length $L = \lambda$ (rather than $\lambda/2$). These ocean waves travel at a speed $\sqrt{2}$ greater than those at $\lambda/2$, and hence the $\sqrt{2}$ spike; this is a "grating lobe" effect occurring for larger diffraction grating spacings. Hydrodynamically, the $\sqrt{2}$ singularity is due to the second spatial harmonic of the trochoidal wave profile of fundamental length $L = \lambda$; this second harmonic is of length $\lambda/2$, producing first-order Bragg scatter, but it travels at the same phase speed as the fundamental to which it is attached. The phase speed of the fundamental is $\sqrt{2}$ greater than the normal ocean wave with length $\lambda/2$, and hence the $\sqrt{2}$ hydrodynamic contribution. Finally, the $2^{3/4}$ singularity is due to a "corner reflector" electromagnetic effect. This occurs when the two sets of (non-evanescent) scattering ocean waves pass through 45° with respect to the propagation direction. The total Doppler shift from these two sets of ocean waves, $\omega_1 + \omega_2 = \omega_B(\sqrt{\cos \alpha} + \sqrt{\sin \alpha})$, is maximum‡ at $\alpha = 45^\circ$, i.e., $2^{3/4} \omega_B$. Thus a condition of mathematical stationarity occurs for the Doppler shift when α passes through 45° .

As deduced previously from (F12.18) and (F12.19), the second-order received sea echo spectrum increases both in its amplitude and in its proximity to the first-order Bragg lines with increasing wind speed. Conversely, for a given wind speed, the same second-order spectrum increases with increasing carrier frequency. The common parameter for each curve in the figures is $g/(2k_0 u^2)$, where u is the wind speed and $k_0 (= 2\pi f_0/c)$ is the radar wavenumber. Thus for a given value of this parameter, the same spectrum curve can be obtained by doubling wind speed and reducing frequency by a factor of four, or if the frequency is increased by a factor of four, by halving the wind speed.

At present, we have no conclusive experimental validation of the theory because of a general lack of accurate HF sea echo spectrum measurements. Several available records, however, exhibit many of the principal features of our predicted second-order spectra. Crombie's ground-wave measurements at 8.37 MHz (F12.5) shows a definite second peak above the positive first-order line; its position is greater than the first-order line by a factor of 1.36. From (F12.18), this would correspond to seas aroused by a wind speed of about 21-22 knots. Also evident in this record by Crombie is a third peak which occurs at about 1.69 times the first-order line; the $2^{3/4} (= 1.682)$ singularity appears to explain this peak. Barnum's sky-wave backscatter measurements (F12.8) at 25.75 MHz also appear to contain higher-order peaks beyond the first-order line. He estimates that these occur at 1.35, 1.70, and 2.00 times the first-order line. Again, the first higher order peak at 1.35 is explainable by seas driven by about 13-knot winds, while the second peak is quite close to the predicted $2^{3/4}$ singularity. As with Crombie's record, the height of the third peak near $2^{3/4}$ is less than that of the second, while the height of the second is considerably less than that of the first-order Bragg line; these features all agree with the predicted second-order spectral behavior.

Hasselmann (1971) has suggested (based upon several approximations which were examined in detail by Stewart (1971)) that the second-order sea echo spectra above and below the first-order Bragg lines should be symmetric reproductions of the first-order temporal nondirectional waveheight spectrum of the sea, centered about the first-order lines. Our more detailed derivations show that these second-order spectra are *not* symmetric about the Bragg lines; this is especially true for the crosswind case, where no energy at all appeared above the Bragg line. However, these second-order contributions do possess *some* of the features of the first-order waveheight temporal spectrum centered around the Bragg lines. For example, they become higher in

† The angle α here is $(180^\circ - \phi)/2$, where ϕ is the bistatic angle between the incident and the first-scattered radio wave. Backscatter produced by such a double interaction process requires that the bistatic angle between the first-scattered and second-scattered (i.e., backscattered) radio wave be ϕ .

amplitude and move in closer to the Bragg lines with increasing wind speed and/or frequency. Hence we agree with Hasselmann's basic conclusions that sea state *can* be deduced at higher HF frequencies by examining the features (i.e., strength and position) of the second-order peaks in the sea echo backscatter spectrum. Upon further confirmation by measurements, this technique may prove to be quite valuable in remote sensing of sea state, especially with ionospheric radars which are restricted in their operation to the upper HF region.

12.4 UHF Indirect Bragg Scatter Using Two Frequencies

A technique currently under development for measuring the slope spectrum of the longer gravity waves will be briefly examined here. This concept employs the correlation between the sea return at two closely spaced UHF frequencies as a measure of the larger and longer ocean waves present. The interpretation of the final result of the derivation shows that the mechanism yielding the slope spectrum of the surface *resembles* Bragg scatter; the surface slope spectrum is evaluated, however, not at the wavenumber corresponding to the carrier frequency, but at the "beat" wavenumber corresponding to the difference between the two UHF frequencies.

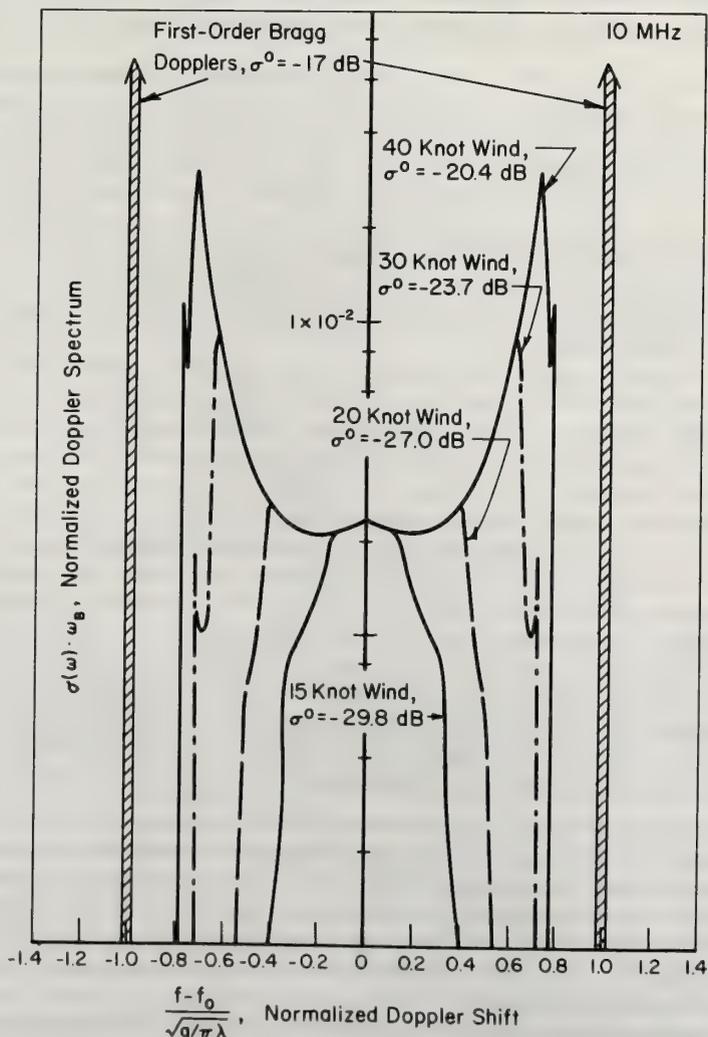


Figure 12.19 Predicted Doppler spectrum of first and second-order near-grazing sea backscatter at 10 MHz for propagation in crosswind direction (Phillips semi-isotropic waveheight spectrum assumed).

Before summarizing the analysis, let us briefly discuss the physics behind the anticipated behavior. HF scatter from the sea has been conclusively shown, both theoretically and experimentally, to be due to the Bragg (or diffraction grating) effect. The theory shows that a *mathematical* upper limit of frequency (for a given waveheight) can be expected, beyond which the perturbation approach used should not be valid; this upper frequency in terms of wind speed is given in (F12.6). Several recent experimental efforts, however, have established that the Bragg mechanism produces non-specular sea scatter at UHF, microwave, and even millimeter-wave frequencies. Wright (1968) deduced this from his signal spectra—as well as from quantitative comparisons of σ^0 for various polarizations with the previously developed theory; he observed scatter from waves generated in a controlled wind tank. Guinard and Daley (1970) established that the Bragg mechanism also explained—even quantitatively—the microwave scatter they observed on the sea. In the latter case, much larger and longer ocean waves are present; yet measurements have confirmed that the much smaller wavelets actually producing microwave backscatter are those whose lengths are $\lambda/(2 \sin \theta)$, where θ is the angle of incidence from the vertical. Again, their results measured for σ_{vv}^0 and σ_{hh}^0 (the average backscatter cross sections per unit area) agree quantitatively with theoretical predictions based on a *slightly* rough surface, i.e., (12:20), (12:22), and (12:23), in their dependence on polarization, incidence angle, and saturation effect in the wind-wave equilibrium region. Such agreement is apparent over most aspect angles, as long as one stays away from the specular direction (i.e., the vertical for backscatter) and grazing incidence (where shadowing becomes significant).

Barrick and Peake (1968) and Wright (1968) explained this behavior by considering the surface at these higher frequencies to be a “composite”, made up of two or more scales of roughness. Thus, one has the Bragg-scattering wavelets riding on top of the longer and higher gravity waves. With this model, one obtains two regions of scatter: the quasi-specular region and the diffuse region. Near the specular direction, backscatter is produced via reflections from many specular points, or facets, oriented normal to the line of sight. For the sea, this type of backscatter dominates out to 10-15° from the vertical; its magnitude and behavior is predictable from both physical and geometrical optics approaches. Farther away from the specular direction, scatter is predictable via the Bragg mechanism, as though the smaller wavelets riding on the larger waves were really the only ones present. The magnitude and polarization dependence of this “diffuse” scatter follows (12:20) for the slightly rough surface.

Valenzuela (1968) first noted that the magnitude of the return from a slightly rough surface (i.e., the Bragg scatter) does have some dependence upon the local incidence and scattering angles, as seen in (12:20)—even though the dependence may be weak for some polarization states over a large range of angles. Hence the effect of the longer gravity waves under the Bragg-scattering wavelets should be seen as a “tilting plane”, modulating the amplitude of the Bragg scatter because of the slope of the larger-scale wave underneath. Let us take as an example a uniform Bragg-scattering wavetrain on top of a single larger and longer sinusoidal wave. Now imagine a short radar pulse, less in its spatial length than one-quarter the wavelength of the longer sea wave, propagating along the surface and backscattering via the Bragg mechanism from the wavelets. Due to the slope of the longer sinusoidal wave and hence the periodic variation of the local angle of incidence to the pulse as it propagates along, the radar receiver should see a return which is amplitude modulated in a periodic manner by the longer gravity wave. If one analyzed the spectrum of this amplitude modulated signal, he would be able to relate the result to the slope of the larger wave at its own spatial frequency or wavenumber. Thus one could, with such a short-pulse experiment, measure the slope spectrum of the longer gravity waves by Fourier transforming the received signal strength and looking at its spectrum. This technique was examined recently by Soviet investigators (Zamarayev and Kalmykov, 1969).

The concept to be analyzed here is quite similar to that described above. By using two frequencies, however, and cross-correlating their received powers, one eliminates the Fourier transform process required for the short-pulse technique. Nearly CW signals can be used. The bandwidth required (i.e., frequency separation here) is much the same as for the short-pulse, however, because both techniques are essentially employing spatial range resolution to distinguish the slopes of the underlying longer gravity waves. By eliminating the spectral analysis process and the short pulse requirement, we feel that the two-frequency correlation concept offers a possibly more tractable sensing tool.

The two-frequency correlation concept was analyzed in Ruck et al., (1971) in detail. The derivation there was meant to establish quantitatively some of the features of the correlated power. To facilitate the analysis, the following assumptions were imposed: (a) Only backscatter was considered. (b) The surface was taken as perfectly conducting. (c) Horizontal polarization was examined for incidence and backscatter. (d) A

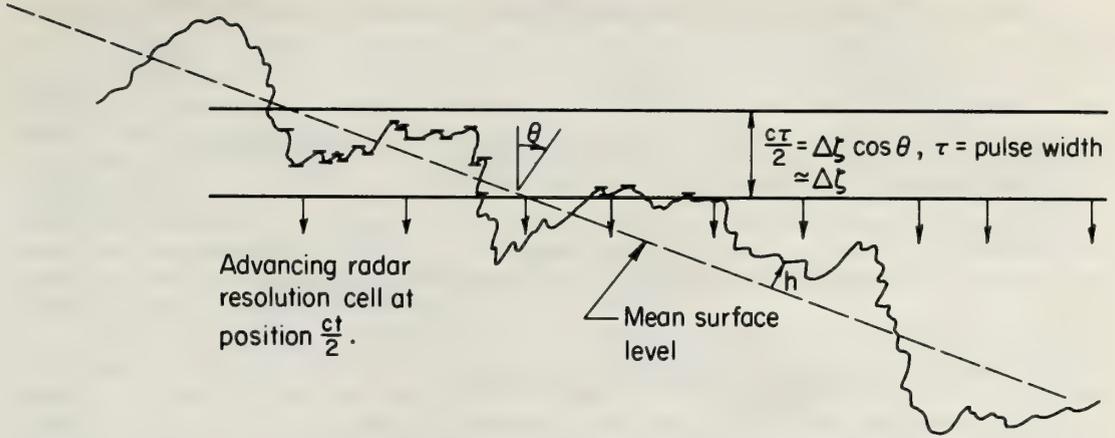


Figure 12.20 Physical picture of specular point scatter. Specular points within radar resolution cell are shown highlighted.

one-dimensional random surface was analyzed, corrugated along the plane of incidence. (e) The incidence angle region was selected to be not too close to the specular direction (i.e., the vertical), but yet not so close to grazing that shadowing is a problem. (f) The slopes of both the large-scale and small-scale sea-wave components present were assumed small.

Based upon these assumptions, the variance of the backscattered power densities at two frequencies, f_b and f_a , was obtained, i.e., $\text{Var}[P(\Delta k)] \equiv \langle P_a P_b \rangle - \langle P_a \rangle \langle P_b \rangle$, where $\Delta k = k_b - k_a = 2\pi(f_b - f_a)/c$. Thus, Δk can be considered a “beat” or difference wavenumber. The length and width of the surface patch (assumed square) subtended in the radar cell is L , and the angle of incidence from the vertical is θ . Then the result for the variance was found to be

$$\begin{aligned} \text{Var}[P(\Delta k)] = & \frac{k_a^2 k_b^2 L^6 E_0^4}{Z_0^2 q_z^2 16 \pi^2 R_0^4} \left\{ \left[\frac{\sin(\Delta k L \sin \theta)}{\Delta k L \sin \theta} \right]^2 \left[f^2(k_0 \xi) W(k_0 \xi) \right]_{s=0}^2 \right. \\ & \left. + \left[\frac{d(f^2(k_0 \xi) W(k_0 \xi))}{ds} \right]_{s=0}^2 \cdot \frac{1}{\pi L} W_{SL}(2\Delta k \sin \theta) \right\}, \end{aligned} \quad (12:35)$$

where R_0 is the distance from the scattering patch to the far-field point, E_0 is the electric field strength of the plane wave incident on the surface, and Z_0 is the free-space wave impedance. The quantity $W(\kappa_x)$ is the one-dimensional spatial waveheight spectrum of the sea surface in the wave-length range around one meter; $W_{SL}(\kappa_x)$ is the one-dimensional spatial waveslope spectrum of the larger gravity waves. The remaining quantities appearing in (12:35) are $q_z = 2 \cos \theta$, $s = d\xi_L/dx = \text{slope of the larger-scale component of the surface}$, $\xi = 2s \cos \theta - 2 \sin \theta$, and $f(k_0 \xi) = k_0 \left[\sqrt{1 - (2s \cos \theta - \sin \theta)^2} + \cos \theta + s \sin \theta \right] \times [\cos \theta + s \sin \theta]$.

In order to obtain numerical estimates of the relative magnitudes of the first and second terms in (12:35), we must evaluate the quantity in square brackets and its derivative at zero slope. To do this, a form for the waveheight spectrum of the smaller-scale (Bragg-scattering) ocean waves must be assumed. By selecting our operating frequency in the UHF band (viz., near 1 GHz), we ensure that these Bragg-scattering ocean waves are of the order of 30 cm in length. Such waves are still gravity waves (in contrast to capillary waves) and hence should follow the Phillips model in the saturation region. Furthermore, these shorter gravity waves require winds greater than only 1-1/2 knots to excite them; hence, they are nearly always present. On the other hand, their build-up time is of the order of 10 minutes, in contrast to capillary waves which build up and die down in a matter of seconds; therefore they should exist rather uniformly and stably over times and areas which are significant in making the measurements. This is the reason that 1 GHz is proposed as the operating frequency.

One can readily convert the two-dimensional Phillips model (12:13) to a one-dimensional version. The result is

$$W(\kappa_x) = \frac{B}{\kappa_x^3}, \quad (12:36)$$

where $\kappa_x > 0$ (i.e., the above spectrum is one-sided), and as before, $B = 0.005$ (dimensionless).

Upon evaluation of the indicated factors in (12:35), we obtain

$$\begin{aligned} \text{Var}[P(\Delta k)] = & \frac{k_a k_b L^6 B^2 E_0^4 \cot^6 \theta}{Z_0^2 \gamma^6 \pi^2 R_0^4} \left\{ \cos^2 \theta \left[\frac{\sin(\Delta k L \sin \theta)}{\Delta k L \sin \theta} \right]^2 \right. \\ & \left. + \frac{(5 \sin^2 \theta - 3 \cos^2 \theta)^2}{\pi L \sin^2 \theta} W_{SL}(2\Delta k \sin \theta) \right\}. \end{aligned} \quad (12:37)$$

Let us now interpret the two terms in (12:35) and (12:37). The first term is merely the Fourier transform of the range resolution cell illumination pattern on the surface (i.e., here we assumed uniform illumination over the cell of length $(L/2) \sin \theta$) evaluated at spatial frequency $2\Delta k$. If the range cell is beam-limited rather than pulse-limited, one might approximate the cell illumination by a uniform pattern between the half-power points of the antenna beam pattern on the surface. If one uses a more realistic illumination pattern along this cell, the $(\sin x)/x$ function will be replaced by the Fourier transform of the actual pattern. With a properly selected and tapered illumination function, the first term can be kept very small, so long as $\Delta k L \sin \theta$ is large compared to unity. In a pulse-limited situation, this means making the pulse length, τ , sufficiently long that $2\pi\Delta f\tau \sin \theta \gg 1$ over the range of Δf used in the experiment.

The second term in (12:35) and (12:37) contains the desired information about sea state, as represented in the waveslope spectrum of the longer gravity waves. Their waveheight spectrum is readily related to $W_{SL}(\kappa_x)$ as $W_{HL}(\kappa_x) = W_{SL}(\kappa_x)/\kappa_x^2$. Hence, measurement of the waveslope spectrum by sweeping frequency (and thus varying Δk) can be directly transformed into waveheight spectral information. It is desirable to select θ , the incidence angle, so that the magnitude of the second term is enhanced with respect to the first. For horizontal polarization and backscatter, a poor choice would be θ at or near 37.8° from the vertical, for this makes the factor in parentheses multiplying the second term identically zero. On the other hand, a value of θ near 60° will usually result in the second term being larger than the first for W_{SL} non-zero and near its equilibrium (saturation) value.

The argument of the term containing the waveslope spectrum would lead one to think that a Bragg-effect scatter were occurring at the beat wavenumber $2\Delta k \sin \theta$, rather than at the carrier wavenumber k_a or k_b . Hence we refer to this as an indirect Bragg-scatter measurement. By sweeping Δf from 2-20 MHz, one should be able to obtain sea state information by measuring the magnitude, shape, and cutoff of the larger gravity wave slope spectrum. The two frequencies can be generated quite simply by using a balanced modulator near the output of the transmitter. Since the scattered power is correlated in the receiver, it is not necessary to maintain phase coherence of the two signals through the receiver channels. Hence the equipment requirements should present no significant obstacles.

This technique, examined here for backscatter and horizontal polarization, can be used for other polarization states and in bistatic arrangements (so long as one avoids the specular reflection direction). The analysis is currently being extended to include three-dimensional scatter from two-dimensionally rough, non-perfectly conducting surfaces. The basic nature of the results are not expected to differ from those examined here, however. Up to the present, this technique has not been tested experimentally; hence we can offer no measured data for validation of the concept. Plans are underway to test the technique in the near future.

12.5 Sea State Effects on a Microwave Radar Altimeter Pulse

As a final tool for remote sensing of both geodetic and sea state information, we discuss the microwave radar altimeter. Decisions by NASA to fly short-pulse altimeters in both the GEOS-C and Skylab series of

satellites have recently accelerated theoretical and experimental efforts on radar altimetry. A sufficiently clear picture is presently available—both from analysis and experimental data—of the basic interaction process between the pulse and the sea surface. At and near the (vertical) sub-altimeter point, microwave scatter is produced by specular points which are distributed in height, thus having some obvious relationship to sea state. In most cases of interest, the effective radar spatial pulse width is smaller than the ocean waveheights it encounters on the surface; hence a stretching of the pulse will occur due to sea state. When the purpose of the experiment is to find the instantaneous mean sea level (for geodetic reasons) to an accuracy of less than one meter, one is faced with the problem of finding this position in a received echo distorted by sea state effects; hence one must remove such effects from the signal. On the other hand, one may wish to use the radar altimeter as a sea state sensor; in this case, he would like to know how to relate sea state to the received pulse distortion. Both problems are examined here.

The first subsection discusses the specular point-theory of sea scatter and obtains the distribution of these points as a function of waveheight. The second subsection then applies the specular-point scattering model to the radar altimeter configuration and determines a simple closed-form solution for the altimeter return for the case of Gaussian pulse and beam widths. The final subsection simplifies this result for certain limiting altimeter configurations commonly used in practice, and compares the model with measured data.

12.5.1 Specular Point Distribution and the Scattering Model

For the microwave frequencies at which an altimeter will operate, scatter from the sea within the near-vertical region directly beneath the altimeter is quasi-specular in nature. This means that backscatter is produced by specular or glitter points on the surface whose normals point toward the radar. Such scatter persists only to about $10\text{-}15^\circ$ away from the vertical, since gravity waves can seldom maintain slopes greater than this amount before they break and dissipate energy. A physical picture of the specular points illuminated within a short pulse radar cell advancing at an angle θ with respect to the mean surface normal is shown in (F12.20).

Specular point scatter is readily predictable from geometrical and/or physical optics principles, and has been analyzed by Kodis (1966) and Barrick (1968). Here the theory is extended to include the height of the surface, since the short radar pulse may not illuminate the entire peak-trough region at a given time. As the starting point, we note from elementary geometrical optics principles that the field scattered by N specular points (expressed in terms of the square root of the backscatter cross section) is

$$\sigma_B^{1/2} = \sum_{i=1}^N \pi^{1/2} g_i^{1/2} e^{j2k_0 \zeta_i \cos \theta} \quad (12:38)$$

where g_i is the Gaussian curvature at the i -th specular point, i.e., $g_i = |\rho_{1i}\rho_{2i}|$, with ρ_{1i} and ρ_{2i} as the principal radii of curvature at this point. Also, ζ_i is the height of the i -th specular point above the mean surface, θ is the angle of incidence from the vertical, and $k_0 = 2\pi/\lambda$ is the free-space radar wavenumber, λ being the wavelength.

We now square and average the above equation with respect to the phase, φ_{ij} , noting that $\varphi_{ij} = 2k_0(\zeta_i - \zeta_j) \cos \theta$ will be uniformly distributed between zero and 2π as long as the sea waveheight is greater than the radar wavelength. We then rewrite the remaining single summation in integral form as a distribution of specular points versus height, ζ , above the surface and versus Gaussian curvature, g . The average radar cross section per unit surface area per unit height, $\eta^0(\zeta)$, can then be written in terms of the average specular point density, $n(\zeta, g)$ as follows (details are found in Ruck et al, 1971):

$$\eta^0(\zeta) = \pi \int_0^\infty n(\zeta, g) g \, dg \quad (12:39)$$

where $n(\zeta, g)$ is the average number of specular points within the height interval $\zeta - d\zeta/2$ to $\zeta + d\zeta/2$ and with Gaussian curvatures between $g - dg/2$ and $g + dg/2$. The quantity $\eta^0(\zeta)$ is related to σ^0 , the average backscatter cross section per unit area as $\sigma^0 = \int_{-\infty}^\infty \eta^0(\zeta) \, d\zeta$. Thus a short pulse having a radar resolution cell of width $\Delta\zeta$ will produce, on the average, a radar cross section per unit area of $\eta^0(\zeta)\Delta\zeta$.

The specular point density, n , can readily be determined (almost by inspection) from the work of Barrick (1968) preceding Equation (7) of that paper; one merely includes height, ζ , in the probability densities. The following result—applicable for backscatter—is obtained for the integrand of (12:39):

$$n(\zeta, g)gdg = \pi \sec^4 \theta p(\zeta, \zeta_{xsp}, \zeta_{y sp}, \zeta_{xx}, \zeta_{xy}, \zeta_{yy}) d\zeta_{xx} d\zeta_{xy} d\zeta_{yy}, \quad (12:40)$$

where $\zeta_x, \zeta_y, \zeta_{xx}, \zeta_{xy}, \zeta_{yy}$ are the partial derivatives of the surface height up to second order, $\zeta_{xsp}, \zeta_{y sp}$ are the surface slopes required at a specular point (these latter slopes are known geometrical quantities). The quantity $p(x_1, \dots, x_n)$ is the joint probability density function for the random variables x_1, \dots, x_n .

The integration over ζ_{xx}, ζ_{xy} , and ζ_{yy} can now be performed. Furthermore, since the height ζ and the slopes $\zeta_{xsp}, \zeta_{y sp}$ at any point are uncorrelated (as discussed in 12.1.2), and since we intend to employ Gaussian distributions for the surface height and slopes (also discussed in 12.1.2), we can finally express the scatter per unit height as the product of the height and slope density functions:

$$\eta^0(\zeta) = \pi \sec^4 \theta p(\zeta) p(\zeta_{xsp}, \zeta_{y sp}), \quad (12:41)$$

where $p(\zeta)$ is as given in (12:1) and $p(\zeta_x, \zeta_y)$ is given in (12:2). Also, the required total slope at the specular point to be used in (12:2) is $\sqrt{\zeta_{xsp}^2 + \zeta_{y sp}^2} = \tan \theta$.

12.5.2 Application to Satellite Altimeter

We now apply (12:41) to the problem depicted in (F12.21). An altimeter at height H emits a spherical pulse which in turn sweeps past a spherical earth. The spatial pulse width for a backscatter radar is $\Delta\zeta = c\tau/2$, where c is the free space radio wave velocity and τ is the time width of the pulse (compressed, if applicable). As our time reference, we choose $t = 0$ to be the time that the center of the signal, reflected from the uppermost cap of a smooth spherical earth, returns to the receiver. In terms of the angles shown in (F12.21), the height, ζ , to a point at the center of the cell above the mean sea surface can then be written as

or

$$\zeta = \frac{H(1 - \cos \psi) + a(1 - \cos \varphi) - (ct/2) \cos \psi}{\cos \varphi},$$

or

$$\zeta \approx \frac{H}{2} \psi^2 \left(1 + \frac{H}{a}\right) - \frac{ct}{2} \approx \frac{H}{2} \psi^2 - \frac{ct}{2}, \quad (12:42)$$

where it is assumed that $H \ll a$, and that ψ and φ are small over the scattering region on the sea of significance in the altimeter receiver return.

Several investigators have employed the relationship in (12:42) along with a model for surface scatter, to examine the altimeter radar return. Godbey (1964) and Greenwood et al (1969) assumed quite simple models for the scattering coefficient, and hence obtained a result which was not quantitatively dependent upon ocean waveheight. Miller and Hayne (1971) made a considerable improvement over these efforts by assuming a slightly skewed Gaussian model for $p(\zeta)$ in (12:41), but did not show the remaining slope dependence contained in the specular point model developed here. Hence their radar return, while containing an unspecified multiplicative factor, nonetheless possesses the same pulse shape characteristics as our model.

For the purpose of obtaining a general closed-form containing all of the parameters of interest in the design and analysis of an altimeter experiment, we select simple Gaussian beam and pulse shapes. Thus we define the power gain of the antenna (squared to account for the use of the same antenna twice) as

$$G^2(\psi) = \exp\left\{-(8\ln 2)\psi^2/\Psi_B^2\right\}, \quad (12:43)$$

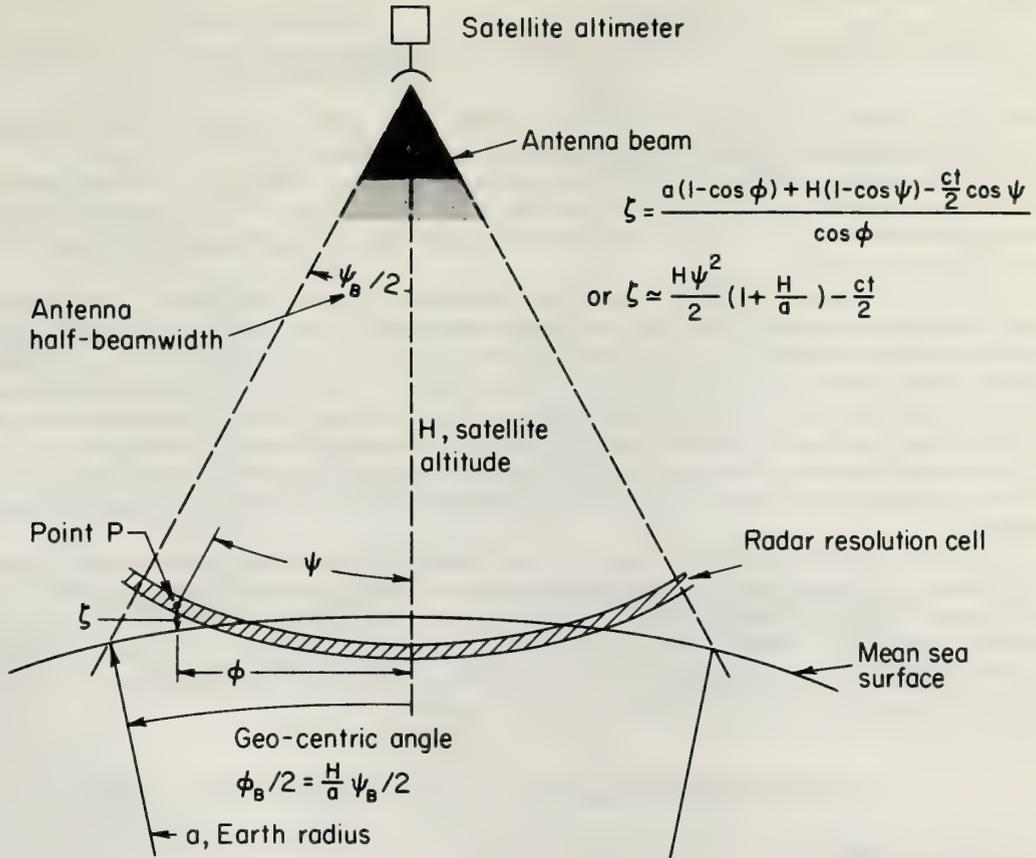


Figure 12.21 Geometry of satellite altimeter.

where ψ_B is the one-way half-power antenna beamwidth. Likewise, $P(x)$, defined as the spatial (compressed, if applicable) altimeter signal power pulse, is taken to be

$$P(x) = \exp\left\{-(4\ln 2)t_*^2/\tau^2\right\} = P(ct_*/2), \tag{12:44}$$

where t_* is time with respect to the pulse center, τ is the half-power width of the signal pulse, and the distance x is related to time in a backscatter radar as $x = ct_*/2$.

Now, the product of the average radar cross section and the antenna gain is a function of time as the spherical pulse intercepts the sea near the suborbital point. This produces an average power at the receiver output which is related to $G^2\sigma(t)$ through the radar equation†: $P_R(t) = P_T \frac{\lambda^2}{(4\pi)^3 H^4} G^2\sigma(t)$.

$$G^2\sigma(t) = 2\pi^2 a^2 \int_0^\infty G^2(\psi) p(\xi_{xsp}, \xi_{ysp}) \sec^4 \theta \sin \varphi \times \left[\int_{-\infty}^\infty P(\xi - \zeta) p(\xi) d\xi \right] d\varphi, \tag{12:45}$$

where the Gaussian height and slope distributions, $p(\zeta)$ and $p(\xi_x, \xi_y)$ are to be used in integrating (12:45). One can express θ and ψ in terms of φ as follows: $\psi \approx \frac{a}{H}\varphi$ and $\theta \approx \psi + \varphi \approx \frac{a}{H}\varphi$. Then the integrations over ξ and φ

† The height factor in the radar range equation is assumed to be constant to first order, because only the echo from the vicinity of the suborbital point is of interest.

can be carried out to obtain:

$$G^2 \sigma(t) = \left(\pi^{3/2} H x_w / s^2 \right) \exp \left[(t_p / t_s)^2 - 2t / t_s \right] \left[1 - \Phi(t_p / t_s - t / t_p) \right], \quad (12:46)$$

where $x_w = c\tau / (4\sqrt{\ln 2})$, $t_p = 2\sqrt{x_w^2 + 2h^2} / c$, $t_s = 2H\Psi_B^2 / c$, and $1/\Psi_B^2 = (8 \ln 2) / \Psi_B^2 + 1/s^2$, c being the free-space radio wave velocity. Also, $\Phi(x)$ is the error function of argument x .

The mean-square sea wave height, h^2 , and total slope, s^2 , result from the use of the Gaussian height and slope distributions given in 12.1.2. For wind-driven fully developed seas, h^2 can be expressed in terms of wind speed through (12:15). The mean-square slope is related to the wind speed through (12:16) and (12:17), or as a simpler model (12.18).

The interpretation of t_p and t_s sheds light on the interaction process. The constant t_p is the equivalent pulse width after stretching by sea waves on a planar mean surface. If the rms height h of the sea waves is greater than the spatial width x_w of the pulse, then the equivalent pulse length after scatter, t_p , is essentially the time of flight between the crests and troughs of the waves. The constant t_s is interpreted as the two-way time of flight difference between the suborbital point and the edge of the effective scattering region illuminated by the advancing spherical pulse. For a narrow antenna beamwidth, Ψ_B , this time becomes small. If the slopes of the specular-scattering waves are smaller than the antenna beamwidth, however, then the width of the scattering region is limited by the lack of specular-scattering sea waves rather than the antenna beamwidth.

A term commonly used in the literature on radar altimeters is the "impulse response" of the surface. This is merely the return from the surface illuminated by an impulse function, $P(x) = \delta(x)$; let us refer to this as $G^2 \sigma_I(t)$. To obtain the response of the surface to any other waveform, one need only convolve the surface impulse response with the desired waveform. The impulse response can be obtained very simply from (12:46) by noting that the impulse function $\delta(x)$ can be taken by writing $P(x) = (\pi x_w^2)^{-1/2} \exp -x^2/x_w^2$ in (12:44) and taking the limit of (12:46) as $x_w \rightarrow 0$. We thus obtain

$$G^2 \sigma_I(t) = (\pi H / s^2) \exp \left[(t_p / t_s)^2 - 2t / t_s \right] \left[1 - \Phi(t_p / t_s - t / t_p) \right], \quad (12:47)$$

where now $t_p = 2\sqrt{2}h/c$.

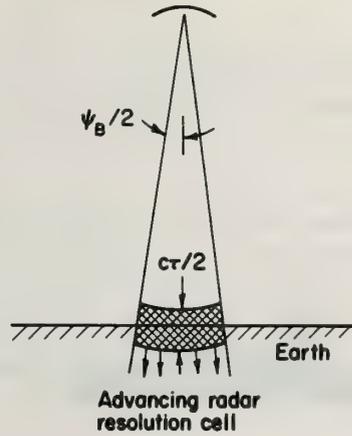
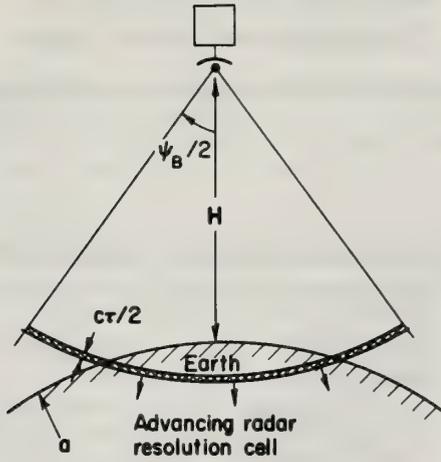
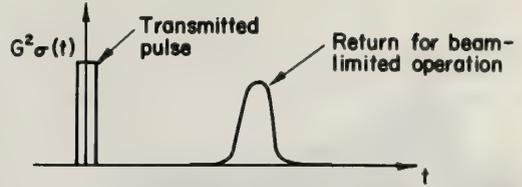
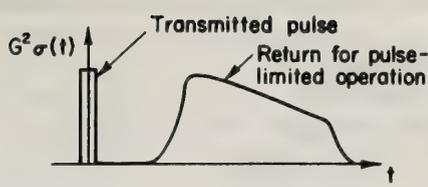
To provide further insight into the interaction process, and also to serve as a check on the model (12:46), let us consider the limiting situation where the roughness vanishes. First, we allow the rms height, h , to approach zero; we still assume, however, that the rms slope, s , is non-zero, i.e., that there are spatially distributed specular-point scatterers, but with near-zero height deviation from the mean surface. In this limit, the form of (12:46) remains the same, but $t_p = \tau / (2\sqrt{\ln 2})$. Thus the equivalent pulse width, t_p , is not stretched after the interaction because the waves have zero height. Now, as we allow the rms slope, s , to approach zero, this results in $t_s \rightarrow 0$. Then (12:46) simplifies to $G^2 \sigma(t) = (\pi H^2) \exp \{-t^2/t_p^2\}$. This merely means that the original Gaussian pulse, represented by the exponential, is reflected from a smooth surface, and hence the entire return comes from the suborbital specular reflection. The factor πH^2 is simply the radar cross section, from geometrical optics considerations, for a spherical surface with radius H . Hence, one obtains the expected result in the limit of vanishing roughness.

12.5.3 Limiting Altimeter Configurations

(a) **Pulse-Limited Operation.** When the altimeter altitude and/or beamwidth are sufficiently large that $t_s \gg t_p$, we have the situation depicted in (F12.22). The return at a given time is obtained from the area in the circular annulus subtended by the pulse. This is referred to as pulse-limited operation. The return in the limit $t_s \gg t_p$ may then be obtained from (12:46) as

$$G^2 \sigma_p(t) = \left(\pi^{3/2} H x_w / s^2 \right) \left[1 + \Phi(t/t_p) \right] \exp(-2t/t_s). \quad (12:48)$$

The above equation shows that the return consists of a rapid rise near $t=0$, as expressed by the quantity in square brackets, followed by a very gradual exponential decay to zero. In this case, the leading edge of the return contains the desired information about the mean surface position and/or sea state; the



Pulse-Limited Altimeter

Beam-Limited Altimeter

Figure 12.22 Two modes of altimeter operation and the resulting signals.

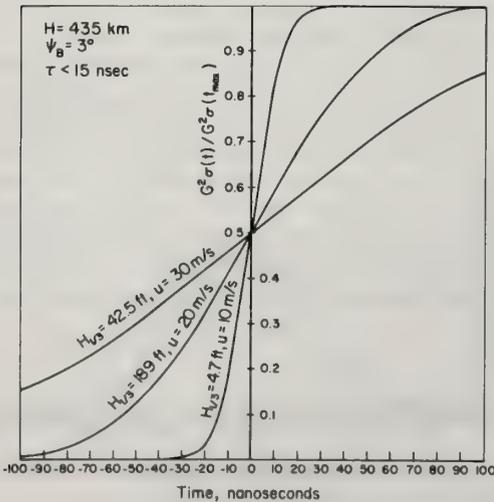


Figure 12.23 Leading edge of averaged altimeter output versus time for pulse-limited operation.

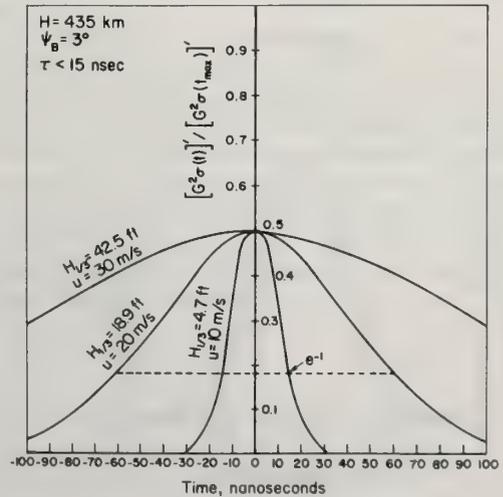
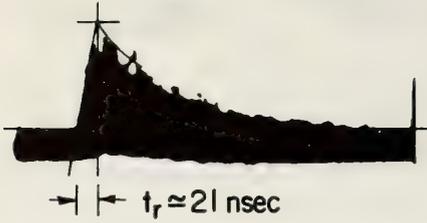
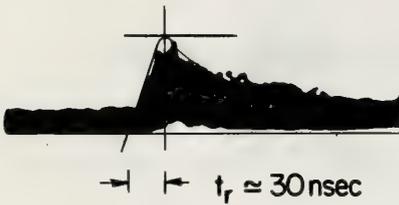


Figure 12.24 Derivative of leading edge of averaged altimeter output versus time for pulse-limited operation.



Flight #14
 Run #12
 H = 10 kft
 $\tau = 20$ nsec
 $t_r \approx 21$ nsec
 Measured wind = 12 knots
 Calculated wind = 14.1 knots

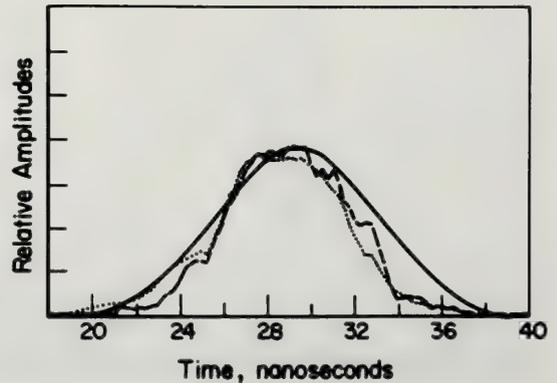
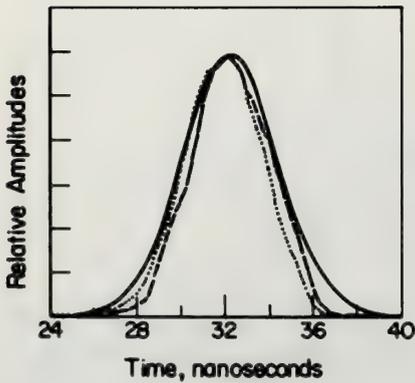


Flight #16
 Run #9
 H = 10 kft
 $\tau = 20$ nsec
 $t_r \approx 30$ nsec
 Measured wind = 22 knots
 Calculated wind = 21.2 knots

Figure 12.25 Measured aircraft altimeter responses. Wind speeds inferred from rise times are compared to observed wind speeds.

$H_{1/3} = 3.1$ ft

$H_{1/3} = 5.2$ ft



- Measured radar response
- Measured wavestaff response
- Calculated response

Figure 12.26 Measured (after Yaplee et al., 1971) altimeter (impulse) responses versus calculated using beam-limited model.

leading edge (normalized) is proportional to $\frac{1}{2} [1 + \Phi(t/t_p)]$. Shown in (F12.23) are curves of this leading edge as a function of significant waveheight (related to rms waveheight in 12.1.1). Also shown on these curves are the wind speeds required to fully arouse wind waves to these heights. The parameters chosen are typical of the NASA Skylab satellite altimeter. The figure shows that the mean surface position can be found from the mean return (in the absence of noise) by locating the half-power point on the leading edge.

For sea-state determination, one could determine the rise time of the leading edge. A possibly simpler technique would be to differentiate the mean return near the leading edge, producing a pulse-type signal as shown in (F12.24). This can be obtained from (12:48) as

$$\frac{d}{dt} [G^2 \sigma_p(t)] \approx [2\pi H x_w / (s^2 t_p)] \exp(-t^2/t_p^2). \quad (12:49)$$

Thus the width of this pulse is directly proportional to the effective pulse width, t_p . If the altimeter pulse width, τ , is kept small with respect to the expected stretching due to roughness (i.e., less than about 15 ns), then the width of this differentiated return is directly proportional to significant (or rms) waveheight. Thus an orbiting altimeter such as this could very simply monitor the significant waveheight of the oceans along its orbital path.

As limited validation of the model, we show results obtained by Raytheon (Ruck et al, 1971) from an aircraft altimeter at 10,000 ft. The pulse width of 20 ns and beamwidth of 5° resulted in a nearly pulse-limited operation. Surface wind speeds were reported as 12 and 22 knots during two of the flights; these are compared in (F12.25) with wind speeds deduced from the rise times, t_r , of the leading edge (assuming wind-driven waves in which waveheight is related to wind speed through (12:15)). The agreement seems quite reasonable.

(b) Beam-Limited Operation. If the antenna beamwidth and/or altimeter height are sufficiently small, then the illumination geometry shown in (F12.22) will result. Here, $t_s \ll t_p$, and the return at any time comes from specular points within a disc of area $\sim \pi H^2 \Psi_0^2/4$ at the suborbital points. In this limit (12:46) can be simplified to obtain:

$$G^2 \sigma_b(t) = \left[\pi H^2 \Psi_0^2 / (2 \sqrt{\epsilon n^2} s^2) \right] (\tau/t_p) \exp[-(t/t_p)^2]. \quad (12:50)$$

Normally the beamwidth and height requirements are such that a satellite altimeter could almost never be beam-limited; nor are most aircraft altimeters beam-limited. The beam-limited mode of operation does offer a very simple return to interpret, however; the width of the return pulse is directly proportional to the ocean waveheight if the signal pulsewidth, τ , is sufficiently small.

As a comparison with the beam-limited model, we show data measured by Yapple et al (1971). His measurements were made from a tower with $H = 70$ ft and $\Psi_B \approx 2^\circ$. His pulse width was 1 ns. We compare in (F12.26) the shape of the curve given by (12:50) with what he has calls his impulse response. He plots the responses measured both by radar and by a wavestaff for two days on which the significant waveheights (measured by the wavestaff) were 3.1 and 5.2 ft; these values of waveheight were used in the curves based on our model. The overall comparison is good, but the figures also illustrate where the Gaussian assumption about the waveheight is weak: both in the echo tails and in the symmetry about the center. These points of departure were mentioned previously in 12.1.2. While these differences are interesting, they should not detract from the fact that the simple Gaussian model is adequate for predicting the mean sea surface position and significant waveheight, so long as accuracies better than about 10 cm are not required.

12.6 References

Barnum, J. R. (1971), Private communication.

Barrick, D. E. (1968), Rough surface scattering based on the specular point theory, IEEE Trans. Ant. Prop. AP-16, 449-454.

Barrick, D. E. and W. H. Peake (1968), A review of scattering from surfaces with different roughness scales, Radio Science 3, 865-868.

- Barrick, D. E. (1970), The interaction of HF/VHF radio waves with the sea surface and its implications, Electromagnetics of the Sea, AGARD Conference Proceedings No. 77; also available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia, Accession No. AD-716 305.
- Barrick, D. E. and J. A. Grimes (1970), Predicted and measured ground-wave sea clutter spectrum at 6 MHz, presented at Fall USNC/URSI Meeting, September 15-17, Columbus, Ohio.
- Barrick, D. E. (1971 a), Theory of HF/VHF propagation across the rough sea, Parts I and II, Radio Science 6, 517-533.
- Barrick, D. E. (1971 b), Dependence of second-order sidebands in HF sea echo upon sea state, pp. 194-197, 1971 G-AP International Symposium Digest, September 21-24, Los Angeles, California.
- Barrick, D. E. (1972), First-order theory and analysis of MF/HF/VHF scatter from the sea, IEEE Trans. Ant. Prop. AP-20, 2-10.
- Bascom, W., 1964: *Waves and Beaches*. Anchor Books, Doubleday & Co., Inc., Garden City, New York.
- Cox, C. S. and W. H. Munk (1954), Statistics of the sea derived from sun glitter, Jour. Marine Res. 13, 198-227.
- Crombie, D. D. (1955), Doppler spectrum of sea echo at 13.56 Mc/s, Nature 175, 681-682.
- Crombie, D. D., J. M. Watts and W. M. Beery (1970), Spectral characteristics of HF ground wave signals backscattered from the sea, Electromagnetics of the Sea, AGARD Conference Proceedings No. 77; also available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia, Accession No. AD-716 305.
- Crombie, D. D., 1971: Backscatter of HF radio waves from the sea, pp. 131-162 of J. R. Wait, ed. *Electromagnetic Probing in Geophysics*, The Golem Press, Boulder, Colorado.
- Davies, H. (1954), The reflection of electromagnetic waves from a rough surface, Proc. IEE, Part III 101, 209-214; discussion on above paper Proc. IEE, Part III 102, 148.
- Godbey, T. W. (1964), Oceanographic satellite radar altimeter and wind sea sensor, pp. 21-26 of G. C. Ewing, ed. *Oceanography in Space*, Proceedings of Conference, Woods Hole Oceanographic Institution, Reference No. 65-10, August 24.
- Greenwood, J. A., A. Nathan, G. Neumann, W. J. Pierson, F. C. Jackson and T. E. Pease, 1969: Radar altimetry from a spacecraft and its potential applications to geodesy, pp. 59-80 *Remote Sensing of the Environment*, Vol. 1, American Elsevier Publishing Co., Inc., New York.
- Guinard, N. W. and J. C. Daley (1970), An experimental study of a sea clutter model, Proc. IEEE 58, 543-550.
- Hasselmann, K. (1971), Determination of ocean wave spectra from Doppler radio return from the sea surface, Nature Physical Science 229, 16-17.
- Kinsman, B., 1965: *Wind Waves*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Kodis, R. D. (1966), A note on the theory of scattering from an irregular surface, IEEE Trans. Ant. Prop. AP-14, 77-82.
- MacKay, J. H. (1959), On the Gaussian nature of ocean waves, Engineering Experiment Station, Georgia Institute of Technology, Atlanta, Georgia, Internal Technical No. 8.

- Miles, J. W. (1962), On the generation of surface waves by shear flows, Part 4, *J. Fluid Mech.* *13*, 433-448.
- Miller, L. S. and G. S. Hayne (1971), System study of the geodetic altimeter concept, Research Triangle Institute, Research Triangle Park, North Carolina, Final Report, Contract No. NAS6-1829 (March 1971), Unclassified.
- Moskowitz, L. (1964), Estimates of the power spectrums of fully developed seas for wind speeds of 20 to 40 knots, *J. Geophys. Res.* *69*, 5161-5179.
- Munk, W. H. and W. A. Nierenberg (1969), High frequency radar sea return and the Phillips saturation constant, *Nature* *224*, 1285.
- Nathanson, F. E. (1971), Radar pulse compression and high resolution sea reflectivity, Proceedings of Joint NOAA/NASA/NAVY Conference on Sea Surface Topography from Space, October 6-8, Key Biscayne, Florida.
- Nierenberg, W. A. and W. H. Munk (1969), Sea spectra and radar scattering, 1969 JASON Summer Study Working Paper, Boulder, Colorado.
- Peake, W. H. (1959), Theory of radar return from terrain, Part 1, *IRE Int. Conv. Record* *7*, 27-41.
- Peterson, A. M., C. C. Teague and G. L. Tyler (1970), Bistatic radar observation of long-period directional ocean-wave spectra with LORAN A, *Science* *170*, 158-161.
- Phillips, O. M., 1966: *Dynamics of the Upper Ocean*. Cambridge University Press, London, pp. 109-119.
- Rice, S. O., 1951: Reflection of electromagnetic waves from slightly rough surfaces, pp. 351-378 of M. Kline, ed. *Theory of Electromagnetic Waves*, Interscience Publishers, New York.
- Ruck, G. T., D. E. Barrick and T. Kaliszewski (1971), Bistatic radar sea state monitoring, Battelle, Columbus Laboratories, Columbus, Ohio, Technical Report, Contract NAS6-2006 (December, 1971), Unclassified.
- Stewart, R. H. (1971), Higher order scattering of radio waves from the sea, pp. 190-193, 1971 G-AP International Symposium Digest, September 21-24, Los Angeles, California.
- Teague, C. C. (1971), High-frequency resonant scattering techniques for observation of directional ocean-wave spectra, Stanford University, Stanford, California, Technical Report No. SU-SEL-71-039, Contract N00014-69-A-0200-6012, Unclassified; also available as Ph.D. dissertation under same title, University Microfilms, Ann Arbor, Michigan.
- Tveten, L. H. (1967), Ionospherically propagated sea scatter, *Science* *157*, 1302-1304.
- Valenzuela, G. R. (1968), Scattering of electromagnetic waves from a tilted slightly rough surface, *Radio Science* *3*, 1057-1066.
- Wait, J. R. (1966), Theory of HF ground wave backscatter from sea waves, *J. Geophys. Res.* *71*, 4839-4842.
- Ward, J. F. (1969), Power spectra from ocean movements measured remotely by ionospheric radio backscatter, *Nature* *223*, 1325-1330.
- Wright, J. W. (1968), A new model for sea clutter, *IEEE Trans. Ant. Prop.* *AP-16*, 217-223.
- Yaplee, B. S., A. Shapiro, D. L. Hammond, B. D. Au and E. A. Uliana (1971), Nanosecond radar observations of the ocean surface from a stable platform, *IEEE Trans. Geoscience Electronics*, *GE-9*, 170-174.

Zamarayev, B. D. and A. I. Kalmykov (1969), On the possibility of determining the spatial structure of an agitated ocean surface by means of radar, Izvestia USSR Academy of Sciences – Atmospheric and Oceanic Physics 5, 64-66.

List of Symbols

a	(radio) radius of earth ($\approx 4/3$ times actual earth radius)	k_o	radar wavenumber ($= 2\pi/\lambda$)
a	spatial wavenumber of fundamental in Fourier expansion of (ocean) surface waveheight ($= 2\pi/L_f$)	k, m, n	summation indices, assuming integer values
B	Phillips ocean waveheight spectrum saturation constant (≈ 0.005)	L	spatial period of ocean wave
c	free-space radio wave velocity ($\approx 3 \times 10^8$ m/s)	L_{CO}	cutoff (spatial) wavelength of Phillips ocean waveheight spectrum
d	half the distance between transmitter and receiver in bistatic radar	L_f	spatial wavelength of fundamental in Fourier expansion of (ocean) surface waveheight.
E_o	electric field intensity incident on scattering patch	$n(\zeta, g)$	density of specular points per unit area versus height and Gaussian curvature
F_T, F_R	Norton attenuation factors for near surface propagation between scattering patch and transmitter, receiver.	P_b, P_a	powers received at two frequencies, f_b and f_a
f	temporal frequency in cycles per second	$P(m, n, k)$	Fourier coefficient of expansion of ocean surface height in terms of x, y, t
f_o	radio carrier frequency	x, y, t	
f_b, f_a	carrier frequencies in two-frequency sensor	P_R	average received power at receiver
f_D	Doppler shift from carrier ($= f - f_o$)	$P_R(\omega)$	average received power spectral density at receiver
f_N	normalized Doppler shift	P_T	average transmitted power
G_T, G_R	free-space gains of transmitter and receiver antennas	$P(x)$	effective spatial (power) pulse of radar altimeter
g	acceleration of gravity (≈ 9.81 m/s ²)	$P(x_1, \dots, x_n)$	joint probability density function of n random variables x_1, \dots, x_n
g	Gaussian curvature of point on surface ($= \rho_1 \rho_2 $)	R_o	distance from antenna to scattering patch in backscatter radar
H	height of radar altimeter	R_T, R_R	ranges (distances) from scattering patch to transmitter and receiver
$H_{1/3}$	significant (ocean) waveheight, crest-to-trough	S	normalized time delay in bistatic radar ($= t/t_d$)
h	rms (ocean) waveheight with respect to mean surface	$S(\kappa_x, \kappa_y)$	spatial ocean waveheight spectrum
i	imaginary number symbol ($= \sqrt{-1}$)	$S(\kappa_x, \nu_y, \omega)$	spatial-temporal ocean waveheight spectrum

$S(\omega)$	temporal ocean waveheight spectrum	Γ_T	total coupling factor appearing in integrand for second-order sea scatter
s	rms total (ocean) waveslope ($= [\langle \xi_x^2 + \xi_y^2 \rangle]^{1/2}$)	Δ	normalized average impedance of sea surface; also, symbol denoting an increment, i.e., ΔR - increment of range
$\text{sgn}(p)$	symbol taking value ± 1 depending on whether argument p is \pm	$\delta(p)$	Dirac impulse function of argument p
T	temporal period of ocean wave	ξ	height of ocean surface above mean plane
T_f	temporal wavelength of fundamental in Fourier expansion of (ocean) surface waveheight	ξ_x, ξ_y	directional slopes of ocean surface in (orthogonal) x and y directions ($= \frac{\partial \xi}{\partial x}, \frac{\partial \xi}{\partial y}$)
t	time delay of radar echo	η	radian Doppler shift from carrier ($= \omega - \omega_0$)
t_d	time delay for direct pulse between transmitter and receiver in bistatic radar	$\eta^0(\xi)$	average backscatter cross section per unit area per height increment
t_T	rise time in altimeter radar return	θ	angle between mean normal to surface and incidence or scatter direction
t_s, t_p	time constants (parameters) appearing in altimeter return model	θ	variable in elliptical coordinate system
U, V	normalized ocean spatial wavenumbers ($= \kappa_x/2k_0, \kappa_y/2k_0$)	θ_i	angle of incident radar wave from vertical
u	wind speed	θ_s	angle of scattered radar wave from vertical
v	phase velocity of ocean wave	κ	spatial wavenumber of ocean wave ($= 2\pi/L$)
v_s	satellite velocity	κ_c	spatial wavenumber of ocean waves at threshold between gravity and capillary region ($= 3.8 \text{ cm}^{-1}$)
v_{sh}	ship speed	κ_{CO}	cutoff (spatial) wavenumber of Phillips ocean waveheight spectrum
$W(\kappa_x, \kappa_y, \omega)$	alternate (radio scatter) definition of spatial-temporal ocean waveheight spectrum ($= 2^3 S(\kappa_x, \kappa_y, \omega)$)	κ_x, κ_y	spatial (ocean) height wavenumbers in (orthogonal) x and y directions
$W_{SL}(\kappa_x)$	ocean waveslope spatial spectrum for longer gravity waves	λ	radar wavelength
w	temporal wavenumber of fundamental in Fourier expansion of (ocean) surface waveheight ($= 2\pi/T_f$)	μ	variable in elliptical coordinate systems
$\langle x \rangle$	ensemble average of random variable x	ρ_1, ρ_2	principal radii of curvature at point on surface
x_w	effective spatial altimeter pulse with after stretching by ocean waves	σ^0	average scattering cross section of sea per unit area
Z_0	wave impedance of free space ($\approx 120\pi$ ohms)	$\sigma(\omega)$	average scattering cross section of sea per unit area per rad/s bandwidth
α	grazing angle	τ	radar temporal pulse width
Γ_{EM}	coupling factor for electromagnetic second-order sea scatter terms		
Γ_H	coupling factor for hydrodynamic second-order sea scatter terms		

$\Phi(x)$	error function of argument x
φ	azimuthal angle in bistatic radar near surface between incidence and scatter plane ($\varphi = 180^\circ$ for backscatter); also, in radar altimeter, angle at earth center between altimeter and surface scattering patch
φ_s	azimuthal angle of scattered radar wave from incidence plane
Ψ_B	half-power beamwidth of radar altimeter antenna
ψ	angle of radar altimeter between vertical and scattering point on earth
ω	temporal wavenumber of ocean wave ($= 2\pi/T$)
ω_0	radian wavenumber of radio carrier frequency
ω_B	radian Doppler shift for near-grazing backscatter ($= \sqrt{2k_0 g}$)
ω_{dm}	maximum satellite-induced (radian) Doppler shift
ω_m	maximum ship-induced (radian) Doppler shift
ω_s	radian Doppler shift from sides of ellipse in bistatic radar

Chapter 13 ATMOSPHERIC MOTION BY DOPPLER RADAR

D. A. Wilson and L. J. Miller
Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

This chapter reviews the principles and applications of Doppler radar to the study of atmospheric motion. The first part of the chapter describes the amplitude and phase information contained in the backscattered signal and how this information can be extracted and processed to obtain backscattering cross sections and radial velocities of targets in the atmosphere. The second part of the chapter then describes how this reflectivity and velocity information can be used in the investigation of the structure and motion fields associated with atmospheric phenomena. It is seen that Doppler radar can be usefully employed as a research tool in several areas of meteorology including small scale turbulence, cloud physics, and the dynamics of convective storms and large scale storms.

13.0 Introduction

Radars transmit radio frequency (r-f) energy which is intercepted and reradiated by land, sea, and atmospheric targets. Changes in the received signal from that transmitted give clues to important target parameters such as scattering cross section as well as indirect measurements of atmospheric absorption and attenuation. Radar, as used in the field of meteorology, allows one to detect the presence of objects in the atmosphere, to recognize their character, and to determine their position in three-dimensional space. Doppler radar differs from conventional radar in that it is a very accurate phase measuring device, allowing one to determine the position of the targets as a function of time; thus, their velocities can be determined.

13.1 Basic Radar Principles

Most signals can be represented by the real or imaginary part of

$$s(t) = A(t) \exp \{ j[\omega(t)t + \Phi(t)] \}, \quad (13:1)$$

where the amplitude A , frequency ω , and phase Φ can be functions of time and j is the $\sqrt{-1}$. Some coding schemes for information transmission and reception are amplitude, frequency, or phase modulation. These are generated by varying A (AM), ω (FM), or Φ (PM) with time. Radar return signals are quite complex since generally they consist of combined modulation. To extract information, therefore, it is necessary to demodulate the return signal by appropriate means.

Radiated signals can be of a continuous wave (CW) nature or pulsed. An additional transmission form is the so-called pulsed CW whereby CW source characteristics are maintained pulse-to-pulse. This chapter is restricted to a pulsed CW coherent Doppler radar. This system maintains constant amplitude, frequency, and phase of the transmitted signal. Interpretation of amplitude and phase fluctuations of the r-f return leads to the Doppler or power density spectrum as a measure of target radial speed.

13.1.1 The Backscattered Signal from a Point Target

A point target (small cross-sectional area) within the scattering volume defined by the antenna beamwidth and pulse duration τ , see (F13.1), returns a signal whose instantaneous voltage is

$$E(t) = A(t) \cos [\omega_c t + \Phi(t)] \quad , \quad (13:2)$$

where $\omega_c = 2\pi f_c$ is the constant carrier frequency and Φ is the phase relative to the carrier phase. If the target is fixed, the phase is constant and a function of the distance r from the radar. A moving target having radial velocity V_R returns a signal whose phase varies with time and is given by

$$\Phi(t) = \frac{4\pi}{\lambda} (R_0 + V_R t) \quad , \quad (13:3)$$

where λ is the incident radiation wavelength and R_0 is the initial distance. For narrow beam radiation patterns, the radial velocity is approximately the projection of the vector velocity on the radar beam axis.

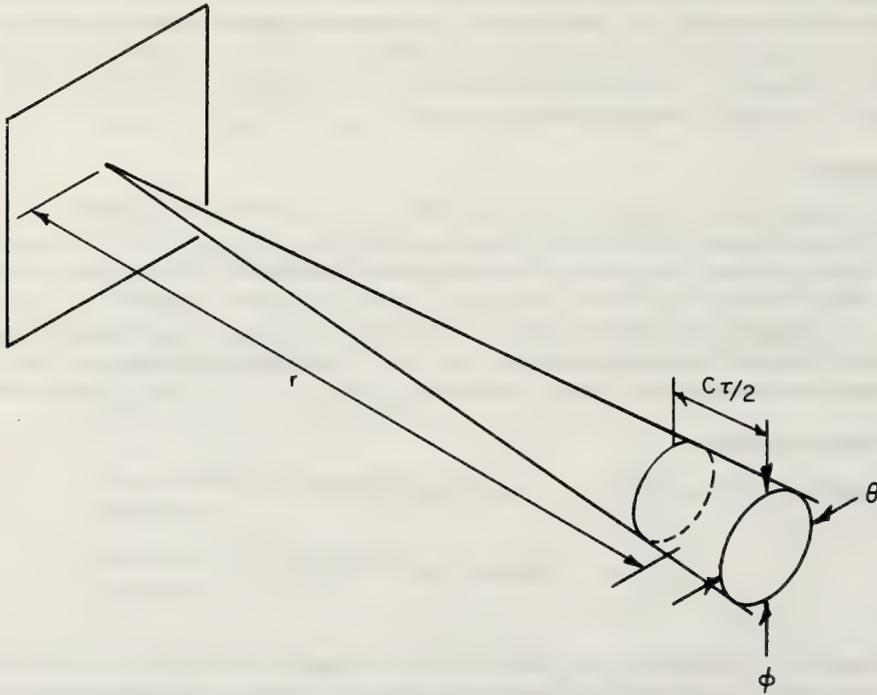


Figure 13.1 Scattering volume at distance r from a pulsed radar having a pulse duration τ and beamwidths θ and ϕ in the horizontal and vertical, respectively.

13.1.2 The Backscattered Signal from Distributed Targets

When the scattering volume contains N point targets, the return signal is the superposition of individual returns. The instantaneous return voltage is then

$$E(t) = \sum_{n=1}^N A_n(t) \cos [\omega_c t + \Phi_n(t)] \quad , \quad (13:4)$$

where A_n is the amplitude and Φ_n is the phase of the return signal from the n^{th} scatterer. The above expression assumes secondary scattering effects are negligible compared to the first order scattering. In addition, target motion must be statistically independent and targets must move freely for several radar wavelengths (Lhermitte, 1968c) for a clear definition of individual phases. A more complicated expression is needed when these conditions are not satisfied. With the possible exception of heavy rain, snow, or hail, the above expression is valid for atmospheric scattering.

13.1.3 The Radar Equation

A target having a cross-sectional area A_c located at a distance r from the radar will intercept an amount of power, $P_t G A_c / 4\pi r^2$, where P_t is the transmitted power and G is the transmitting antenna gain factor. If the target reradiates isotropically, the return power at the receiver is (Battan, 1959)

$$P_r = \frac{P_t G A_e A_c}{16\pi^2 r^4} \tag{13:5}$$

for a receiving antenna having effective area A_e . The relationship between effective area and gain is (Kraus, 1950)

$$A_e = \frac{G\lambda^2}{4\pi} \tag{13:6}$$

A semiempirical expression for antenna gain of a paraboloidal reflector with aperture A_p is (Battan, 1959)

$$G = \frac{8\pi A_p}{3\lambda^2} \tag{13:7}$$

This expression is sufficiently accurate for most weather radars.

Since most targets do not scatter isotropically, it is convenient to introduce the backscattering cross section σ , defined as "the area intercepting that amount of power, which, if scattered isotropically, would return an amount of power equal to that actually received" (Battan, 1959), that is,

$$\sigma \equiv \frac{\left(\text{Power reflected toward the receiving aperture per unit solid angle} \right)}{\left(\text{Incident power density per } 4\pi \text{ steradians} \right)}$$

Substituting backscatter cross section for geometric cross section and replacing the effective area with (13:6), the return power, (13:5), becomes

$$P_r = K_r \frac{\sigma}{r^4} \tag{13:8}$$

where the constant $K_r = P_t G^2 \lambda^2 / 64\pi^3$ depends only on the particular radar system used and not the scatterer. For N targets where σ_n is the cross section of the n^{th} scatterer, on the average, the return power is

$$\bar{P}_r = \frac{K_r}{r^4} \sum_{n=1}^N \sigma_n \tag{13:9}$$

where r is the range to the center of the scattering volume. The above expression assumes random phase of the individual return voltages. A slightly more useful meteorological form is obtained by using the average radar cross section per unit volume and multiplying by the volume, V , effectively illuminated. This leads to

$$\bar{P}_r = \frac{K_r V \eta}{r^4} \tag{13:10}$$

The quantity $\eta = \sum \sigma_n / V$ is called the radar reflectivity. The effective volume for distances much greater than a pulse length is approximately

$$V = \frac{\pi r^2 \theta \phi c \tau}{8}, \quad (13:11)$$

where θ, ϕ are, respectively, the off-axis horizontal and vertical beam angles (assumed to be at most a few degrees) and c ($\sim 3 \times 10^8$ m sec⁻¹) is the propagation speed. The above expressions assume constant gain across the antenna beam. Approximating the antenna pattern by a Gaussian beam (Lhermitte, 1963; Nathanson and Reilly, 1968), the gain is

$$G(\theta, \phi) = G_0 \exp \left[- \left(\frac{\theta^2}{2\sigma_\theta^2} + \frac{\phi^2}{2\sigma_\phi^2} \right) \right] \quad (13:12)$$

where $\sigma_\theta, \sigma_\phi$ are the standard deviations of the two-way pattern (assumed to be at most a few degrees) and G_0 is the on-axis gain factor. Accounting for gain variations across the beam, the exact form of the radar equation, (13:9), becomes

$$\bar{P}_r = \frac{P_t \lambda^2}{64\pi^3} \sum_{\text{vol}} \frac{G^2(\theta, \phi) \sigma_n}{r_n^4} \quad (13:13)$$

Introducing the radar reflectivity $\eta = \eta(r, \theta, \phi)$, the summation can be expressed as a volume integral over the pulse or contributing region so that

$$\bar{P}_r = \frac{P_t \lambda^2}{64\pi^3} \int_{\text{vol}} \frac{G^2(\theta, \phi) \eta(r, \theta, \phi) dV}{r^4} \quad (13:14)$$

Using the Gaussian beam approximation over a volume having uniform reflectivity, integration leads to (Probert-Jones, 1962)

$$\bar{P}_r = \frac{c}{1024\pi^2 \ln 2} (P_t \tau \lambda^2 G_0^2 \theta \phi) \left(\frac{\eta}{r^2} \right) \quad (13:15)$$

where $\ln 2$ is the natural logarithm of 2. Equation (13:15) has been grouped according to the constant ($c/1024\pi^2 \ln 2$), the measurable radar parameters ($P_t \tau \lambda^2 G_0^2 \theta \phi$), and target parameters (η/r^2).

13.1.3.1 Radar Backscattering Cross Section and the Weather Radar Equation

A general treatment of plane wave scattering by a sphere was made by Mie (1908) and later restated by Stratton (1941), Goldstein (1946), and Kerr (1951). Ryde (1946) presented a theoretical paper on Mie scattering applied to radar echoes from water and ice particles. From these theoretical considerations, it has been shown (Gunn and East, 1954) that the backscattering cross section for particle diameters $D \ll \lambda$ is

$$\sigma = \frac{\pi^5 D^6}{\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2, \quad (13:16)$$

where m is the complex index of refraction. The above expression is the Rayleigh approximation to Mie

scattering. The average radar cross section $\bar{\sigma}$ in a pulse volume containing N scatterers is the sum of individual cross sections so that

$$\bar{\sigma} = \frac{\pi^5}{\lambda^4} |K|^2 \sum_{\text{vol}} D_n^6 \quad (13:17)$$

assuming all scatterers consist of the same material such as only water droplets. When there is a mixture (e.g., water and ice), the dielectric factor $|K|^2 = |(m^2 - 1)/(m^2 + 2)|^2$ should be left inside the summation. The dielectric factor varies from about 0.9 for water to 0.2 for ice. Division of (13:17) by the scattering volume gives the average radar reflectivity

$$\eta_{\text{av}} = \frac{\bar{\sigma}}{V} = \frac{\pi^5 |K|^2}{\lambda^4} Z \quad (13:18)$$

where $Z = \sum D_n^6 / V$, the radar reflectivity factor, is the average sum of the sixth powers of particle diameter per unit volume. Generally, Z is given in units of mm^6/m^3 since drop diameters are rarely larger than 5 mm. The above expression can now be substituted for η in (13:15). Therefore, in terms of the radar reflectivity factor Z, the weather radar equation is

$$\bar{P}_r = \frac{c\pi^3}{1024 \ell n 2} \left(\frac{P_t \tau G_o^2 \theta \phi}{\lambda^2} \right) \left(\frac{|K|^2 Z}{r^2} \right) \quad (13:19)$$

Transposing, the radar reflectivity factor is

$$Z = \frac{1024 \ell n 2}{c\pi^3} \left(\frac{\lambda^2}{P_t \tau G_o^2 \theta \phi} \right) \left(\frac{\bar{P}_r r^2}{|K|^2} \right) \quad (13:20)$$

which is also related to meteorological parameters such as rainfall rate and drop-size distributions.

13.1.4 The Backscattered Signal from Coherent Radars for Point and Distributed Targets

Coherence in a radar system does not alter the form of the backscattered signal. It simply increases the amount of information which can be extracted from (13:2) and (13:4). Incoherent radars can measure average power or backscattered signal amplitude. (Also, the variance associated with amplitude fluctuations can be related to the velocity spectrum. However, this is difficult to implement and is rarely used.) In addition to this capability, coherent radars provide information on the change of phase with time. The frequency and phase of the transmitter and local oscillator must be stable for a time at least equal to the pulse repetition period. This is achieved with a stable local oscillator (STALO) and phase-locking loop (see F13.2).

The time rate of phase change, time derivative of (13:3), is an angular frequency $\omega_d = 4\pi V_R/\lambda$. It is therefore equivalent to a Doppler frequency shift

$$f_d = \frac{2V_R}{\lambda} \quad (13:21)$$

Approaching targets, negative velocity in standard kinematic coordinates, have increasing phase with time which corresponds to a positive Doppler frequency shift. Equation (13:21) is an approximation valid for the motion of atmospheric (non-relativistic) particles.

In a pulsed Doppler radar system, the time functions (13:2) for point or (13:4) for distributed targets are available only at discrete time intervals corresponding to the radar pulse repetition period.

Atmospheric Motion by Doppler Radar

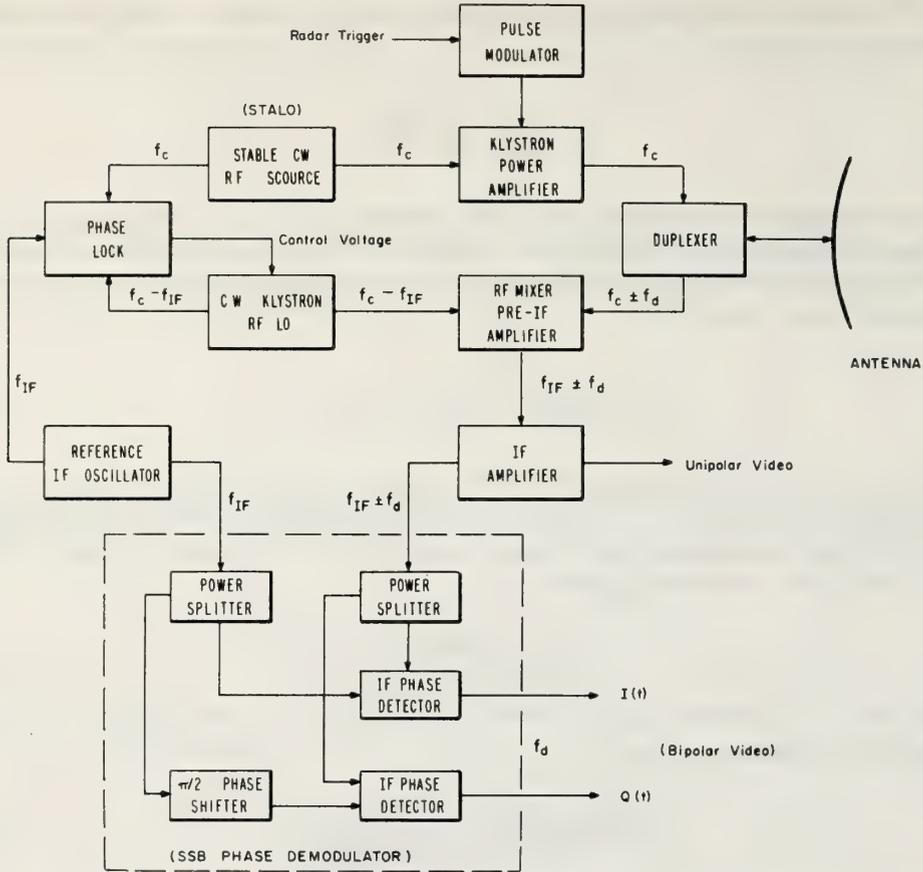


Figure 13.2 Block diagram of a typical coherent Doppler radar system having bipolar video outputs.

Therefore, if the radial velocity of the scatterers is such that the phase changes by more than π (Doppler frequency shift greater than one half the pulse repetition rate), an ambiguity in velocity exists. This is equivalent to aliasing at the folding or Nyquist frequency (Blackman and Tukey, 1958) given by

$$f_N = \frac{1}{2T} \quad , \quad (13:22)$$

where T is the pulse repetition period. If positive and negative Doppler frequency shifts can be resolved, the unambiguous frequency range is doubled. This is accomplished by a scheme similar to the phase-shift method of single-sideband signal reception (Norgaard, 1956). Two balanced mixers are used, each having the Doppler return signal as one of the inputs. The other input for one mixer is a reference signal at the intermediate frequency (IF); whereas, the second mixer has the reference signal phase shifted by $\pi/2$ radians as an input. This is generally a four-port device (F13.2) having two inputs, IF plus Doppler frequency and IF, and two outputs, in-phase and quadrature signals at the Doppler frequency. The phase demodulator outputs are

$$I(t) = E'(t) \cos 2\pi f_d t$$

$$Q(t) = E'(t) \sin 2\pi f_d t \quad , \quad (13:23)$$

where I and Q are respectively the in-phase and quadrature signal amplitudes. The quantity $E'(t)$ is the return signal $E(t)$ from (13:2) or (13:4) after r-f mixing which removes the carrier frequency f_c (see F13.2). This method of phase demodulation gives an unambiguous Doppler frequency range of $-\frac{1}{2T} \leq f_d \leq \frac{1}{2T}$.

13.1.5 The Doppler Spectrum

Equation (13:23) is equivalent to the signal representation

$$a(t) = E'(t) e^{j2\pi f_d t} \quad (13:24)$$

The above expression, for constant amplitude E' , has a Fourier transform consisting of a single impulse of strength $2\pi E'$ at $f = f_d$. This is the case for a single point target having constant backscattering cross section and moving toward the radar at velocity $V_R = f_d \lambda / 2$. The signals from most distributed targets, especially meteorological scatterers, exhibit amplitude variations and contain multiple velocity components rather than a single velocity. The power density spectrum of $a(t)$, where now many frequencies are present, is called the Doppler spectrum.

13.1.5.1 Meaning

The Doppler or power density spectrum is the return power as a function of Doppler shift. Since Doppler frequency and radial velocity are linearly related (13:21), the velocity spectrum can be written as

$$S(V_R) dV_R = P(f) df \quad (13:25)$$

where the power density spectrum $P(f)$ is

$$P(f) = |A(f)|^2 \quad (13:26)$$

The spectral density function $A(f)$ is the Fourier transform of the return signal $a(t)$ and is given by

$$A(f) = \int_{-\infty}^{\infty} a(t) e^{-j2\pi f t} dt \quad (13:27)$$

for continuous signals. The velocity spectrum $S(V_R)$ is the Doppler velocity distribution weighted by the backscattered power. Explicitly, the power spectrum of $a(t)$ is related to the velocities of scatterers by (Smith, 1970)

$$S(V_R) = k \int_{-\infty}^{\infty} \rho^2 W(\rho, V_R) d\rho \quad (13:28)$$

where $W(\rho, V_R)$ is the joint probability of amplitude ρ and radial velocity V_R of the target. This is to say that $W(\rho, V_R) d\rho dV_R$ is the probability that a scatterer picked at random has a velocity between V_R and $V_R + dV_R$ while returning a signal amplitude between ρ and $\rho + d\rho$. The factor k in (13:28) is a normalization factor such that the average return power \bar{P}_r given by

$$\bar{P}_r = \int_{-\infty}^{\infty} S(V_R) dV_R \quad (13:29)$$

is the same as that obtained by using the weather radar equation (13:19). Meteorological interpretations of Doppler velocity spectra are based on (13:28) and (13:29).

13.1.5.2 Measurement

As previously stated (13.1.4), the voltages (13:23) at the output of the phase demodulator are available only at discrete times corresponding to the pulse repetition period. These outputs are sampled at some fixed range and stored as a discrete time series. For N time series samples, the discrete Fourier transform is

$$A(k) = \sum_{n=0}^{N-1} a(n) e^{-j \frac{2\pi kn}{N}}, \quad -\left(\frac{N-1}{2}\right) \leq k \leq \frac{N}{2}, \quad (13:30)$$

where the voltage $a(n) = I(n) + jQ(n)$ is the complex sum of the phase demodulator output voltages for the n th sample taken at time t_n . The power associated with the k th frequency interval is

$$P(k) = |A(k)|^2 \quad (13:31)$$

and the frequency intervals are centered at discrete frequencies

$$f_k = \frac{k}{NT} \quad (13:32)$$

where T , the radar pulse repetition period, is the time between time series samples. Equivalent discrete velocity points can be obtained by using the frequency-velocity relationship (13:21). The minimum velocity resolution obtainable is

$$\Delta V_R = \frac{\lambda}{2NT}, \quad (13:33)$$

and the maximum unambiguous velocity is

$$V_{\max} = \pm \frac{\lambda}{4T}. \quad (13:34)$$

The above restrictions need to be considered in the design of any pulsed Doppler radar system. There is also a limitation on maximum unambiguous range given by

$$r_{\max} = \frac{cT}{2}, \quad (13:35)$$

and minimum range resolution

$$\Delta r_{\min} = \frac{cT}{2} \quad (13:36)$$

These restrictions and limitations necessitate compromise since (13:34) dictates decreasing the pulse repetition period T to increase the maximum unambiguous velocity while (13:35) dictates increasing T to achieve greater unambiguous ranges. Once a pulse repetition period is established, velocity resolution (13:33) is a somewhat less stringent restriction. Required resolution is easily achieved by varying the number of time series samples.

13.2 Doppler Radar Applications

Doppler radar is a powerful research tool for remotely probing atmospheric motion on scales from the microscale (cm wavelength) to the mesoscale (100 km wavelength). Thus, it has been used for studies in 1) Micrometeorology – to learn about the turbulence structure of the atmospheric boundary layer, 2) Cloud Physics – to study the trajectories and growth of precipitation particles in a storm, 3) Cloud Dynamics – to map the circulation inside a convective storm or a hurricane, 4) Synoptic Scale Meteorology – to determine patterns of the wind field, deformation, and divergence in a large scale storm, 5) Air Pollution Meteorology – to derive relationships between the detailed structure of the particle motion and the spread of contaminants. The following subsections present briefly the techniques used and a few results of investigations in some of these areas.

13.2.1 The Radial Velocity Components in a Meteorological Coordinate System

Meteorologists commonly use a right-handed coordinate system with the origin located on the earth's surface and the x axis positive toward the east, the y axis positive toward the north, and the z axis positive upward. For a radar located at the origin, the coordinate transformation equations are

$$\begin{aligned} x &= R \cos \theta \sin \beta \\ y &= R \cos \theta \cos \beta \\ z &= R \sin \theta \end{aligned} \quad (13:37)$$

where R is the range, and θ, β are respectively the elevation and azimuth angles. The geometry of these systems is shown in (F13.3).

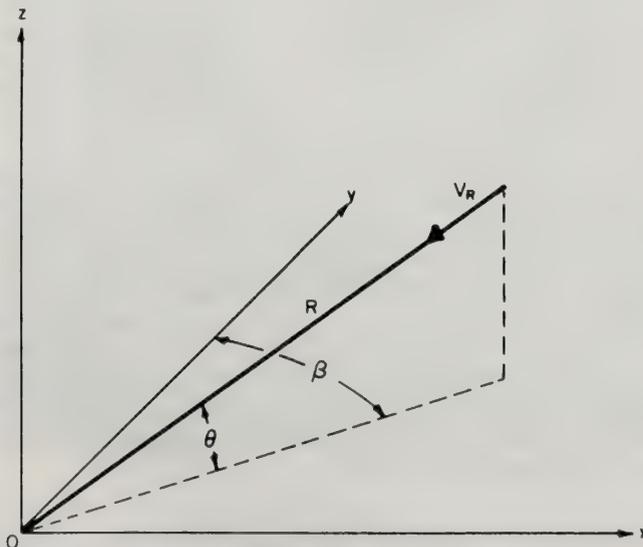


Figure 13.3 Radar coordinate system.

The Doppler radial velocity, V_R , is the projection on the beam axis of the target vector velocity and is given by

$$V_R = u \cos \theta \sin \beta + v \cos \theta \cos \beta + V_f \sin \theta . \quad (13:38)$$

The variables u , v , and V_f are the orthogonal components of target motion along the x , y , and z axes, respectively. For some investigations it is more convenient to express (13:38) in an alternate form

$$V_R R = ux + vy + V_f z , \quad (13:39)$$

where the range $R = \sqrt{x^2 + y^2 + z^2}$.

The vertical component of target motion, V_f , is the sum of the vertical air motion, w , and the target terminal fall velocity, V_t , and is given by

$$V_f = w + V_t . \quad (13:40)$$

The inseparability of the w and V_t contributions, which are often comparable, has been one of the most difficult obstacles in interpretation of Doppler data. Some of the various methods that can be used to surmount this difficulty will be discussed later.

It is assumed, for most studies, that the targets are carried horizontally with the wind. Thus, u and v are equal to the horizontal components of the wind speed. However, it has been shown by Stackpole (1961) and Wilson (1970b) that some targets such as larger raindrops or hailstones respond sluggishly to wind accelerations and caution must be used in interpreting Doppler spectra in turbulent wind fields or in regions of sharp velocity gradients.

13.2.2 Cloud Physics Applications

One simple experiment that can be performed with Doppler radar utilizes a fixed, vertically pointing beam. Here, the horizontal component contribution of (13:38) is zero and the radial velocity is given by (13:40). The radial velocity (13:38) has been written as a function of β , the azimuth angle of the beam axis. Since the beam has finite width, radial velocities for those particles at off-axis locations should strictly be written in terms of $\beta \pm \Delta\beta$. This effect can be interpreted as an additional increase in the spectral width. However, for narrow beams ($\sim 1^\circ$), this effect can be neglected provided the horizontal wind is less than 25 msec^{-1} .

Thus, with suitable range gating, the vertical beam method allows the study of spectra of precipitation fall velocities as a function of height and time. The character of these spectra and their spatial and temporal behavior can provide insight into the precipitation growth and evaporation mechanisms in some situations.

For example, if the contribution due to vertical air motion in (13:40) can be neglected, and the precipitation consists entirely of water drops, the size distribution of the raindrops can be evaluated. The technique involves rewriting the joint probability $W(\rho, V_R)$ in (13:28) as an amplitude distribution times a conditional distribution of radial velocity, given ρ . The amplitude is related to the backscattering cross section and hence particle diameter D , (13:16). Introducing the drop-size distribution $N(D)$, where $N(D)dD$ is the number of particles per unit volume of air having diameters from D to $D + dD$, and performing the integration indicated by (13:28) leads to (Smith, 1970)

$$S(-V_t) = \frac{\bar{P}_r}{Z} D^6 N(D) \frac{dD}{dV_t} . \quad (13:41)$$

This is an expression of the Doppler spectrum as a function of drop diameter. The ratio \bar{P}_r/Z is obtained from (13:19). Generally, the relationship between terminal velocity and drop diameter must be derived from other experimental data (e.g., Best, 1950 or Gunn and Kinzer, 1949). When vertical air motion is present, the left-hand side of (13:41) must be generalized to $S(w-V_t)$.

It must be emphasized that the technique is only valid when the vertical air motion contribution either is small and can be neglected or is known. For stratiform rain where the vertical air motion is almost zero, several excellent examples of drop-size distributions have been obtained (Probert-Jones, 1960; Rogers and Pilié, 1962; Caton, 1963; Rogers, 1966a). In convective precipitation the technique is more questionable since the vertical air motion contribution is unknown. However, through careful analysis of the Doppler information, meaningful results can sometimes be obtained (Sekhon and Srivastava, 1970).

Some examples which show the typical character of fall velocity spectra and drop-size distributions obtained by the vertical beam method are shown in (F13.4, 13.5, 13.6). Figure (13.4) shows a Doppler spectrum of falling snow. The spectrum is quite narrow and appears to be nearly Gaussian in shape. The mean (0.6 msec^{-1}) of the spectrum is in agreement with the expected terminal velocity of snowflakes. The spectral width is the sum of the intrinsic terminal fall velocity variance and the vertical air motion variance. The spectra in (F13.5) were obtained just below the 0°C level in the so-called radar bright band. Since the freezing level is a transitional region, both ice and water particles may be present in the scattering volume. Because these particles do not have the same fall velocities, the spectra become wider. In fact, the spectra often exhibit a distinct bimodal character.

Finally, (F13.6) shows drop-size distributions calculated from Doppler spectra taken in Hawaiian rain (Rogers, 1966a,b). Note that these distributions appear to depart significantly from the often assumed exponential distribution. The concentration of drops smaller than 300 microns in diameter may be questionable since the power returned from the smaller drops may be equal to or less than the noise power or minimum detectable signal. Note also, that the distributions broaden with decreasing altitude. This broadening suggests that a coalescence mechanism may be depleting the concentration of small drops and increasing the number of large drops.

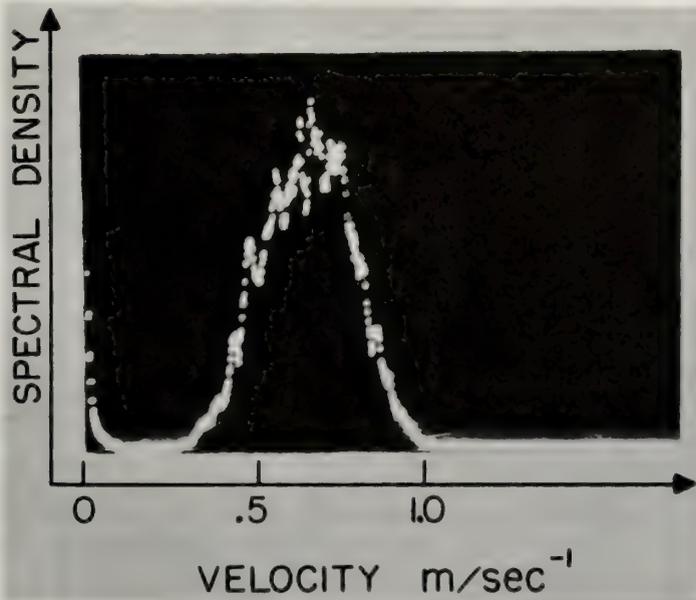


Figure 13.4 Doppler spectrum obtained in snow with a vertically pointing Doppler radar beam. Note the excellent velocity resolution of the system (after Lhermitte, 1968b).

Atmospheric Motion by Doppler Radar

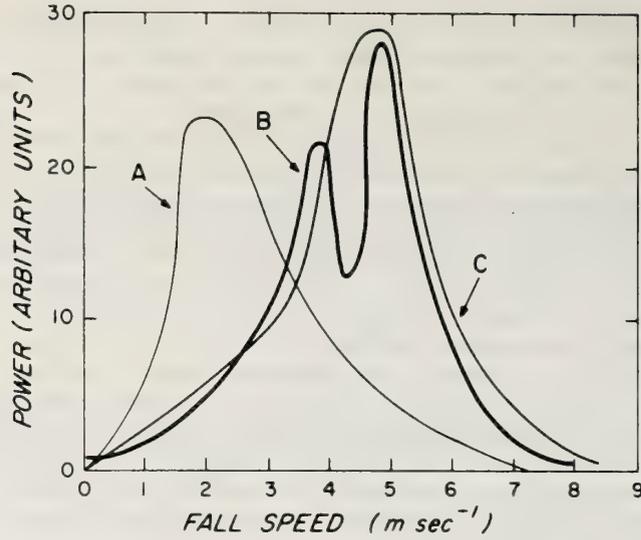


Figure 13.5 Velocity spectra at three altitudes within the melting layer, February 3, 1970. Curve A is 100, Curve B, 250, and Curve C, 400 meters below the height of the 0°C isotherm, as measured by radiosonde. Spectra were averaged over a 5-minute interval (after Dyer, 1970).

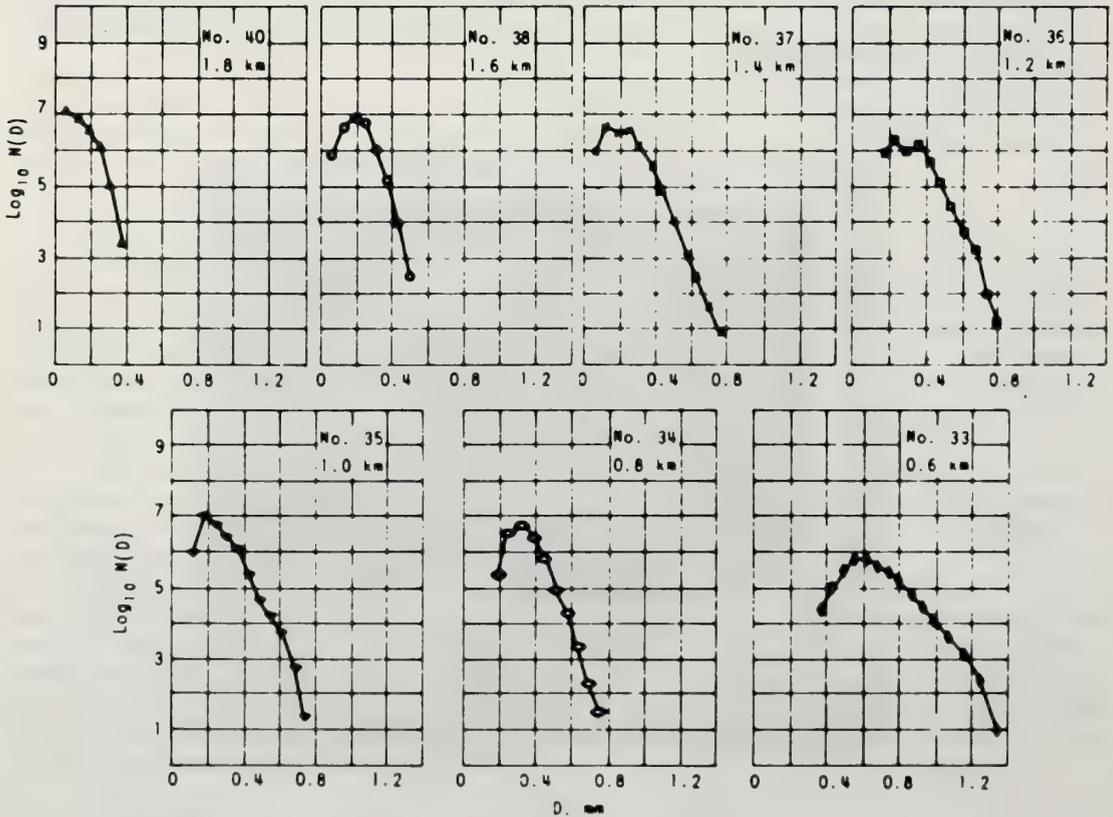
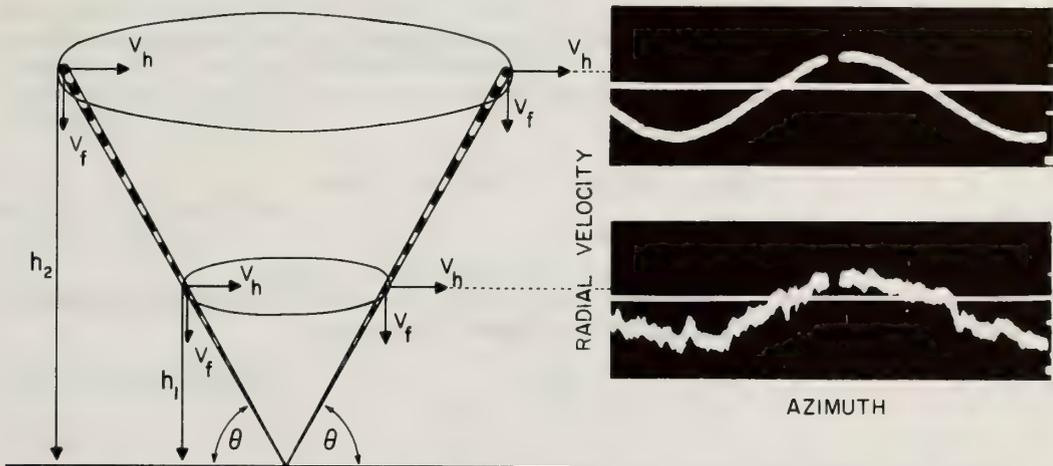


Figure 13.6 Drop-size distribution as a function of altitude (after Rogers, 1966a).

13.2.3 The Kinematic Structure of Large Scale Storm Systems

In this section, a large scale storm system is defined as a storm in which precipitation is falling continuously over a large area. The storm may be producing rain or snow or a combination of the two with associated freezing level and radar bright band. The most common example is the cyclonic disturbances in the westerlies that move through mid-latitudes during the winter months.

Lhermitte and Atlas (1961) recognized that, if there were detectable targets at all azimuths being carried horizontally in a uniform wind field and moving with a constant vertical velocity, then (13:38) would be a sine wave. A fixed elevation angle and range scan then shows the radial velocity as a function of azimuth. This type of scan is the velocity-azimuth display (VAD). The geometry of the VAD technique and some examples of displays taken at two different heights, h_1 and h_2 , are shown in (F13.7).



$$V_R = V_h \cos \theta \cos (\beta - \beta_0) + V_f \sin \theta$$

Figure 13.7 Velocity azimuth display with examples of results (after Lhermitte, 1968c).

The mean wind direction and speed are inferred from the VAD more easily by rewriting (13:38) as

$$V_R = V_H \cos (\beta - \beta_0) + V_f \sin \theta \tag{13:42}$$

For a constant elevation scan, the horizontal wind, V_H , determines the amplitude of the sine wave. The wind direction, β_0 , is a phase shift of the wave and $V_f \sin \theta$ is a DC component due to the vertical motion.

Returning to (F13.7), the horizontal line in each of the VAD displays corresponds to zero radial velocity. A uniform wind is shown in the h_2 trace; whereas, an irregular wind is shown in the h_1 trace. Such irregularities are caused by random as well as organized eddies or wind field perturbations on a scale larger than the pulse volume. These irregularities can also be a result of horizontal gradients of the mean precipitation fall velocity. The width of the VAD trace is caused by the distribution of target motion within a pulse volume. The contribution of the vertical component in (13:42) contains information on the divergence of the mean flow as we shall see in the next section.

13.2.3.1 The Mean Wind Field

Profiles of the mean wind speed and direction as a function of height through the storm can be determined by least-squares analysis of the VAD at each range gate. There is, however, an ambiguity in the prediction of vertical motion when there are convergent or divergent wind fields over the scanning circle.

This ambiguity was first recognized by Caton (1963) who pointed out that a convergent wind field produces a net radial velocity toward the radar that is additive to the contribution from the mean target fall speed. This is easily seen by integrating (13:38) over a complete azimuth revolution such that

$$\int_0^{2\pi} V_R(\beta) d\beta = \cos \theta \int_0^{2\pi} (u \sin \beta + v \cos \beta) d\beta + \sin \theta \int_0^{2\pi} V_f d\beta \quad (13:43)$$

The first term on the right-hand side of (13:43) will be zero only if there is no horizontal wind convergence. Otherwise, this term will be a measure of the divergence which is then additive to the mean fall velocity. As can be seen, scanning at zero elevation angle results in a horizontal divergence measurement since the last term on the right-hand side of (13:43) would then be zero.

A straightforward and thorough method of analysis of the VAD record is outlined by Browning and Wexler (1966). This method is summarized in the following paragraphs. The technique assumes that the radial velocity $V_R(\beta)$ can be expressed as a mean, $\bar{V}_R(\beta)$ plus a fluctuating component, $V'_R(\beta)$. Similarly, $\bar{u} = u + u'$, $\bar{v} = v + v'$, and $V_f = \bar{V}_f + V'_f$. Substituting these expressions into (13:38) and averaging results in

$$\bar{V}_R(\beta) = \bar{u} \cos \theta \sin \beta + \bar{v} \cos \theta \cos \beta + \bar{V}_f \sin \theta \quad (13:44)$$

where the averages of the fluctuating components have been assumed to be zero.

Browning and Wexler then recognized that there may be gradients of the mean wind across the circle and expanded the horizontal components in a truncated Taylor series

$$\begin{aligned} \bar{u} &= \bar{u}_0 + \frac{\partial \bar{u}}{\partial x} x + \frac{\partial \bar{u}}{\partial y} y \\ \bar{v} &= \bar{v}_0 + \frac{\partial \bar{v}}{\partial x} x + \frac{\partial \bar{v}}{\partial y} y \end{aligned} \quad (13:45)$$

where the zero subscript refers to velocities at the center of the scanned circle. Expression (13:45) assumes that horizontal gradients of \bar{V}_f and that higher order terms in the Taylor series expansion can be neglected. Obviously, if the precipitation is showery in nature or if there are considerable irregularities in the wind field due to medium scale turbulence, a noise level is introduced that will lead to uncertainties in the results.

Substituting (13:45) and (13:37) into (13:44) results in

$$\begin{aligned} \bar{V}_R(\beta) &= \frac{R \cos^2 \theta}{2} \left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} \right) + \bar{V}_f \sin \theta + \bar{u}_0 \cos \theta \sin \beta \\ &+ \bar{v}_0 \cos \theta \cos \beta + \frac{R \cos^2 \theta}{2} \left(\frac{\partial \bar{u}}{\partial y} + \frac{\partial \bar{v}}{\partial x} \right) \sin 2\beta \\ &+ \frac{R \cos^2 \theta}{2} \left(\frac{\partial \bar{v}}{\partial y} - \frac{\partial \bar{u}}{\partial x} \right) \cos 2\beta \end{aligned} \quad (13:46)$$

The terms in (13:46) can be evaluated through Fourier series least-squares analysis of the $\bar{V}_R(\beta)$ measurements. This analysis leads to

$$\bar{V}_R(\beta) = A_0 + \sum_{k=1}^2 (A_k \cos k\beta + B_k \sin k\beta) \quad (13:47)$$

where

$$A_0 = \frac{1}{2\pi} \int_0^{2\pi} V_R(\beta) d\beta$$

$$A_k = \frac{1}{\pi} \int_0^{2\pi} V_R(\beta) \cos k\beta d\beta \tag{13:48}$$

$$B_k = \frac{1}{\pi} \int_0^{2\pi} V_R(\beta) \sin k\beta d\beta$$

The Fourier coefficients are then related to the following kinematic properties of the wind field:

HORIZONTAL DIVERGENCE:
$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} = \frac{2}{R \cos^2 \theta} (A_0 - \bar{V}_f \sin \theta) \tag{13:49}$$

MEAN WIND COMPONENTS:
$$\begin{aligned} \bar{v}_0 &= A_1 / \cos \theta \\ \bar{u}_0 &= B_1 / \cos \theta \end{aligned} \tag{13:50}$$

STRETCHING DEFORMATION:
$$\left(\frac{\partial \bar{u}}{\partial x} - \frac{\partial \bar{v}}{\partial y} \right) = -\frac{2A_2}{R \cos^2 \theta} \tag{13:51}$$

SHEARING DEFORMATION:
$$\left(\frac{\partial \bar{u}}{\partial y} + \frac{\partial \bar{v}}{\partial x} \right) = \frac{2B_2}{R \cos^2 \theta} \tag{13:52}$$

Browning and Wexler also considered error sources that might contaminate their results. They concluded, for widespread precipitation, that reasonable estimates of divergence and deformation can be made within 20 km of the radar at low elevation angles. The upper limit on the elevation angles where the target fall velocity can still be neglected was determined to be 9° in snow and 7° in rain.

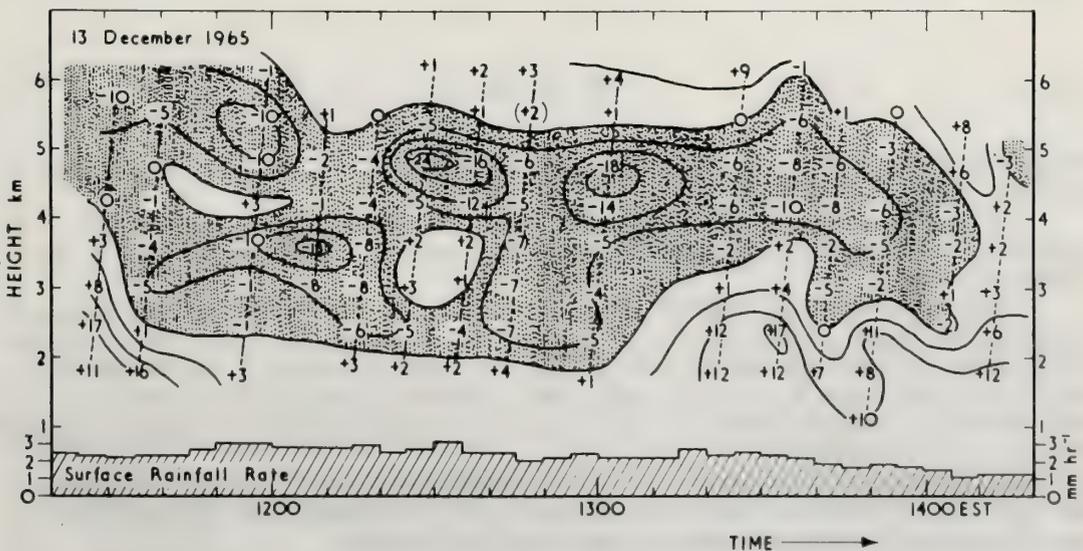


Figure 13.8 Time-height pattern of divergence in units of 10^{-5} sec^{-1} . Surface rainfall rate is shown along the abscissa (after Browning and Wexler, 1966).

An example of the divergence pattern as a function of time and height is shown in (F13.8) which is taken from Browning and Wexler's paper. The melting layer was located at 1800 meters; thus, most of the data was taken in snow conditions and at the restricted elevation angles and ranges stated above. The pattern reveals that even in steady precipitation conditions the divergence changes quite rapidly. Notice near the end of the record (after 1400 EST) that the surface rainfall rate decreases as the convergence aloft changes to divergence.

Vertical air motion can be obtained by applying the continuity equation. For an incompressible, steady state fluid the equation of continuity is

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial z} = 0 \quad (13:53)$$

Integration of (13:53) through a depth of the atmosphere from z_1 to z_2 gives

$$\bar{w}(z_2) = \bar{w}(z_1) - \int_{z_1}^{z_2} \left(\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} \right) dz \quad (13:54)$$

as an estimate of the mean vertical motion through the layer. Since the atmosphere does not favor compression and is generally locally steady, this method gives estimates of mean vertical motion that are usually within 10% of the true values. If z_1 is at the ground, $\bar{w}(z_1) = 0$ and the vertical velocity profile can be obtained by integration of (13:53) through successive layers of the storm. Measurements at low elevation angles are sometimes contaminated by ground clutter which contributes to a spectrum at zero velocity. This spectrum biases the $\bar{w}(0)$ estimates toward zero velocity, thereby introducing uncertainties in absolute values of \bar{w} at higher altitudes.

13.2.3.2 The Turbulence Field

Numerous attempts have been made to infer turbulent wind field properties from the Doppler spectral variance. For example, the variance can be related to small scale turbulent processes that occur at wavelengths smaller than the characteristic dimensions of the pulse volume. The variance, σ_V^2 , of the Doppler spectrum is

$$\sigma_V^2 = \sigma_S^2 + \sigma_F^2 + \sigma_B^2 + \sigma_T^2 \quad (13:55)$$

where the individual terms are the wind shear, σ_S^2 , the fall speed, σ_F^2 , the finite beamwidth, σ_B^2 , and the turbulent kinetic energy, σ_T^2 , variances.

The quantity of interest is the variance, σ_T^2 , due to the turbulent kinetic energy of the air. To measure σ_T^2 , it is necessary to separate the various contributions to (13:55). More often than not, this is a formidable problem.

Despite this difficulty, turbulent kinetic energy estimates have been made under restricted conditions. For example, Rogers and Tripp (1964) were able to measure the contribution to turbulent kinetic energy from vertical motion. A vertically pointing radar beam was used in snow conditions. This effectively eliminates wind shear variance unless strong convection is present. The fall speed spectrum of snow is narrow so that the variance, σ_F^2 can be estimated. The σ_B^2 contribution is minimal for narrow beams or light wind conditions.

Mel'Nichuk et al. (1968) have shown that if the turbulence is obeying the inertial subrange laws, σ_F^2 and σ_B^2 could be effectively eliminated by observing σ_V^2 at different pulse lengths. From their measurements they were able to estimate the dissipation rate of turbulence.

Lhermitte (1968b) recognized that for scales larger than the pulse volume, the turbulent energy could be estimated from the variability of the mean radial velocity. By considering only the variability of

the mean radial velocity, the difficulties of separating variances in (13:55) are largely eliminated. Lhermitte derived from (13:38) an expression for the variance of the radial velocity as a function of the beam angular position

$$\begin{aligned} \sigma_R^2(\beta) = & \sigma_u^2 \cos^2 \theta \sin^2 \beta + \sigma_v^2 \cos^2 \theta \cos^2 \beta + \sigma_w^2 \sin^2 \theta \\ & + \text{cov}(uv) \cos^2 \theta \sin 2\beta + \text{cov}(uw) \sin 2\theta \sin \beta \\ & + \text{cov}(vw) \sin 2\theta \cos \beta \end{aligned} \quad (13:56)$$

where the variance of the terminal fall velocity has been neglected. The interesting feature of (13:56) is that the right-hand side contains both variances and covariances of the velocity components. The covariances are related to the Reynolds stresses or momentum fluxes. By observing σ_R^2 in the upwind and downwind directions, Lhermitte was able to derive the energy spectra of the longitudinal wind and the covariance between the longitudinal wind and the vertical wind.

Wilson (1970a) extended these ideas to derive information on the turbulent energy and momentum fluxes. The scheme involves a subtraction of (13:44) from (13:38) to get an expression for the fluctuating velocities. Assuming horizontal homogeneity and time stationarity of the turbulence field, the expression for the fluctuating component can be squared and averaged to give

$$\begin{aligned} \overline{V_R'^2}(\beta) = & \overline{[V_R(\beta) - \bar{V}_R(\beta)]^2} = (\overline{u'^2} + \overline{v'^2}) \frac{\cos^2 \theta}{2} + \overline{w'^2} \sin^2 \theta \\ & + (\overline{v'^2} - \overline{u'^2}) \frac{\cos^2 \theta}{2} \cos 2\beta + \overline{u'v'} \cos^2 \theta \sin 2\beta \\ & + \overline{u'w'} \sin 2\theta \sin \beta + \overline{v'w'} \sin 2\theta \cos \beta \end{aligned} \quad (13:57)$$

Expression (13:57) is analogous to (13:56) except the variances and covariances are replaced by estimates. Expression (13:57) is then evaluated by Fourier analysis of $\overline{V_R'^2}(\beta)$ in a manner similar to the technique used for the mean wind field. Here, the Fourier coefficients of (13:48) are related to properties of the mean turbulence field and are given by

$$A_0 = \left(\frac{\overline{u'^2} + \overline{v'^2}}{2} \right) \cos^2 \theta + \overline{w'^2} \sin^2 \theta \quad (13:58)$$

$$A_2 = \left(\frac{\overline{v'^2} - \overline{u'^2}}{2} \right) \cos^2 \theta \quad (13:59)$$

$$B_2 = \overline{u'v'} \cos^2 \theta \quad (13:60)$$

$$B_1 = \overline{u'w'} \sin 2\theta \quad (13:61)$$

$$A_1 = \overline{v'w'} \sin 2\theta \quad (13:62)$$

Expression (13:58) can be identified as an estimate of the horizontal turbulent kinetic energy if the elevation angle is low. At higher angles, the vertical turbulent kinetic energy makes an increasingly important contribution to the estimate. Expression (13:59) is an estimate of the relative importance of the horizontal turbulent energy or is a measure of the horizontal isotropy of the turbulence field.

An important result of the analysis is that the individual momentum fluxes $\overline{u'v'}$, $\overline{u'w'}$, and $\overline{v'w'}$ can be estimated. These fluxes are related to the mechanical production rate of turbulent kinetic energy, P, by

$$P = -\overline{u'w'} \frac{\partial \bar{u}}{\partial z} - \overline{v'w'} \frac{\partial \bar{v}}{\partial z} \quad (13:63)$$

where the terms involving the horizontal shears have been neglected by homogeneity assumptions (Lumley and Panofsky, 1964). The vertical gradients of the mean wind components are obtained from the wind profiles calculated using the scheme described in (13.2.3.1).

Results of an analysis of both the mean wind field and the turbulence field (Wilson, 1970a) are

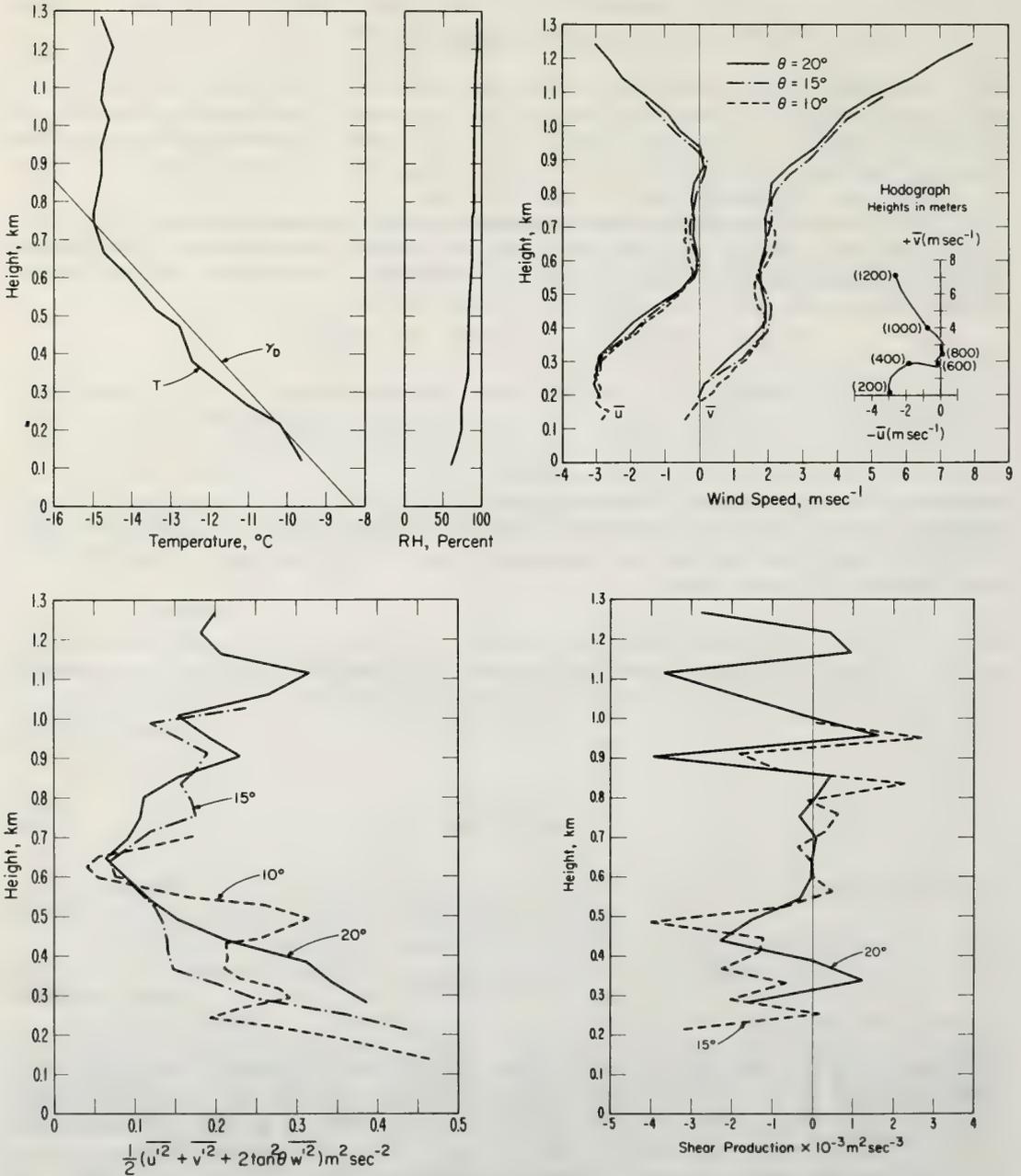


Figure 13.9 Mean wind field and turbulence field as determined from VAD Doppler information in snow. a) Temperature sounding. b) Mean wind profiles. c) Turbulent kinetic energy profiles. d) Mechanical production rate (after Wilson, 1970a).

shown in (F13.9). The data were taken in the vicinity of Boulder, Colorado, during steady light snow. Details of the interpretation of the data in terms of the boundary layer structure are given in the original paper. The important points to be made here are concerned with the technique.

1) A single VAD scan allows calculation of the mean wind profiles with excellent velocity resolution – better than 0.25 msec^{-1} . This resolution is evidenced by the differences in the profiles obtained at successive VAD scans taken two minutes apart at three different elevation angles (F13.9b). For better resolution, data from a series of scans should be analyzed.

2) The profiles of turbulent energy and the mechanical production rate show some variability, but general features of the structure can still be inferred (F13.9 c, d).

3) Simultaneous measurements of the detailed temperature profiles (F13.9a) are highly desirable for comparison with the wind structure and for calculation of Richardson numbers.

13.2.4 The Circulation Inside a Convective Storm

One research area which uses meteorological Doppler radar is the study of three-dimensional velocity fields inside convective storms – the most common example of which is the thunderstorm. Despite other observational techniques such as aircraft and conventional weather radar as well as theoretical and numerical modeling, only general features of storm circulation are known. A review of these features will help determine the applicability of Doppler radar probing of thunderstorms.

Listed below are some of these general features (taken, in part, from Newton, 1967).

1. Several types of thunderstorms exist which range from short lived (20-30 minutes), single convective cells (air mass thunderstorms) to multiple cell complexes. Additionally, it is now recognized that the squall line associated severe storm is a much longer lived phenomena – five hours or more. These squall line storms can be likened to a giant heat engine moving great horizontal distances while being continually fed by moist unstable air.

2. The lifetime of individual cells is characterized by three stages – a cumulus cloud stage primarily composed of updrafts, a mature stage with both updraft and downdraft regions, and a dissipating stage with generally weak downdrafts throughout the cloud.

3. Circulation patterns of well developed thunderstorms are usually more complicated than a simple up-downdraft model. Regions of sloping motion rather than strictly vertical motion are likely to exist in some storms. Additionally, high values of vorticity and strong velocity gradients exist as evidenced by in situ aircraft observations. Vertical velocities can also be as high as 30 msec^{-1} .

4. There is a strong interaction between the circulation of the storm and that of the storm environment at all altitudes. Each is continually modifying the other through inflow, outflow, and entrainment.

Doppler radar is suitable for remote sensing of those areas of a storm having naturally occurring targets such as raindrops. However, probing of the storm environment is not possible since naturally occurring targets are not present. This restriction can be partially overcome by dispensing artificial targets such as chaff (one-half wavelength dipoles) in the storm environment. Certainly the motion inside the storm can be measured with several radars probing the same area at the same time.

Clearly, an instantaneous “picture” of the entire storm and its environment is necessary to completely resolve the entire motion field. However, a great deal of knowledge can still be gained under some restricted scanning techniques such as those described in the next two subsections.

13.2.4.1 The Vertical Beam Method

Early attempts to measure the velocity field inside a thunderstorm used the vertically pointing radar beam (e.g., see Battan, 1964). The technique assumes that a thunderstorm in a significant stage of development passes directly over the radar. Through use of a suitable range gating system, measurements of the vertical component of target motion throughout the depth of the cloud are made. Repeated measurements of V_f as a function of time as the storm passes over the radar allows one to construct a vertical slice through the cloud. This analysis is obscured by the unknown evolution of the storm as it passes overhead and the unknown location of the slice with respect to the configuration of the storm. These unknowns can be partially determined by a surveillance radar probing the storm from some distance

away. Caution is advised, however, in the interpretation of time-height data from the Doppler data when the wind is changing direction with height.

To determine the vertical air motion by the vertical beam method, the contributions from particle fall velocity must be subtracted from the measured radial velocity. Two methods have been proposed to accomplish this. Battan (1964) suggested that the vertical motion of the air could be inferred from the upper velocity bound of the Doppler spectrum. He reasoned that the smallest size particles that could be detected by X-band Doppler radar would be about 300 microns. The terminal velocity of such targets is about 1 msec^{-1} and thus the vertical velocity of the air could be inferred by adding 1 msec^{-1} to the observed upper bound of the spectrum. The method assumes that detectable targets of the 300 micron size are always present in the radar beam. Joss and Waldvogel (1970) pointed out that the upper bound of the spectrum is also controlled by the dynamic range of the recording equipment and by the presence of turbulent motion within the pulse volume. Battan's technique, therefore, might introduce significant uncertainties in the vertical velocity in regions of high reflectivity or strong turbulence.

Rogers (1964) used a slightly different approach for determining the vertical air motion. He reasoned that if one knew the mean terminal velocity of the targets then \bar{w} could be inferred by

$$\bar{w} = \bar{V}_f - \bar{V}_t \quad (13:64)$$

where \bar{V}_f is the mean of the Doppler spectrum and \bar{V}_t is the mean terminal fall velocity. Expressions for the mean terminal velocity as a function of the measured total radar reflectivity factor, Z , can be derived if a drop-size distribution for rain is assumed. For an exponential model of drop-size distribution he derived an expression for \bar{V}_t given by

$$\bar{V}_t = -3.8 Z^{1/14} \quad (13:65)$$

In a later study, Rogers (1967) found that the relationship

$$\bar{V}_t = -0.46 Z^{1/3} \quad (13:66)$$

was more applicable in Hawaiian rainfall.

Battan and Theiss (1968) point out that the Rogers method will lead to uncertainties in the estimate of \bar{w} when ice or melting ice particles are present in the radar beam. In a comparison of the two methods they found that either technique will give estimates of vertical air motion that are accurate within 3 msec^{-1} when the spectral width is between 5 and 15 msec^{-1} . The largest discrepancies occur for very narrow or very wide spectra.

An example of a height-time section obtained by the vertical beam method is shown in (F13.10) (Donaldson, 1967). The vertical velocity was obtained from an assumed relationship between the reflectivity and mean particle fall speed, similar to the Rogers technique. Note that the highest vertical velocities occur near the forward edge of the storm and that updrafts are prevalent in the upper portions of the cloud. Also, there are strong horizontal gradients of \bar{w} at 1636 EST and large values of $\partial\bar{w}/\partial z$ near 1642 EST.

13.2.4.2 Multiple Doppler Radar Methods

To overcome restrictive time stationarity and inhomogeneity assumptions inside a thunderstorm, it has been proposed (Lhermitte, 1968a) that a network of two or three Doppler radars be used to triangulate on the storm and to resolve the components of (13:38). The geometry of such a scheme (F13.11) shows one radar located at $(0,0,0)$ in the meteorological coordinate system, while the other radars are located at $(x_2, 0, 0)$ and $(x_3, y_3, 0)$, respectively. Solutions of (13:38) are not obtained when the three radars are located along the same baseline or when the storm is over any one of the baselines.

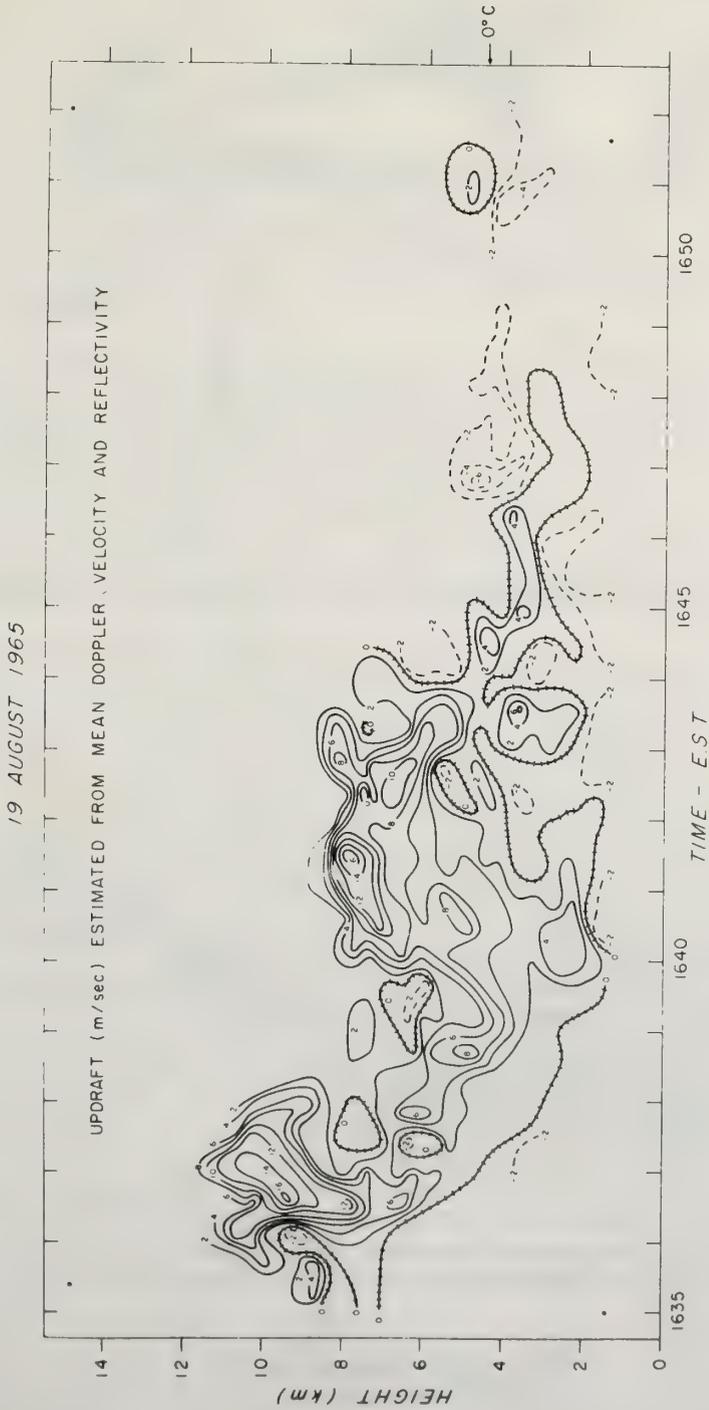


Figure 13.10 Vertical air motion field estimated from Doppler velocity and an assumed relationship between reflectivity and mean particle fall speed. Updrafts are solid contours, and downdrafts are dashed contours, with railroad tracks showing location of zero vertical air motion (after Donaldson, 1967).

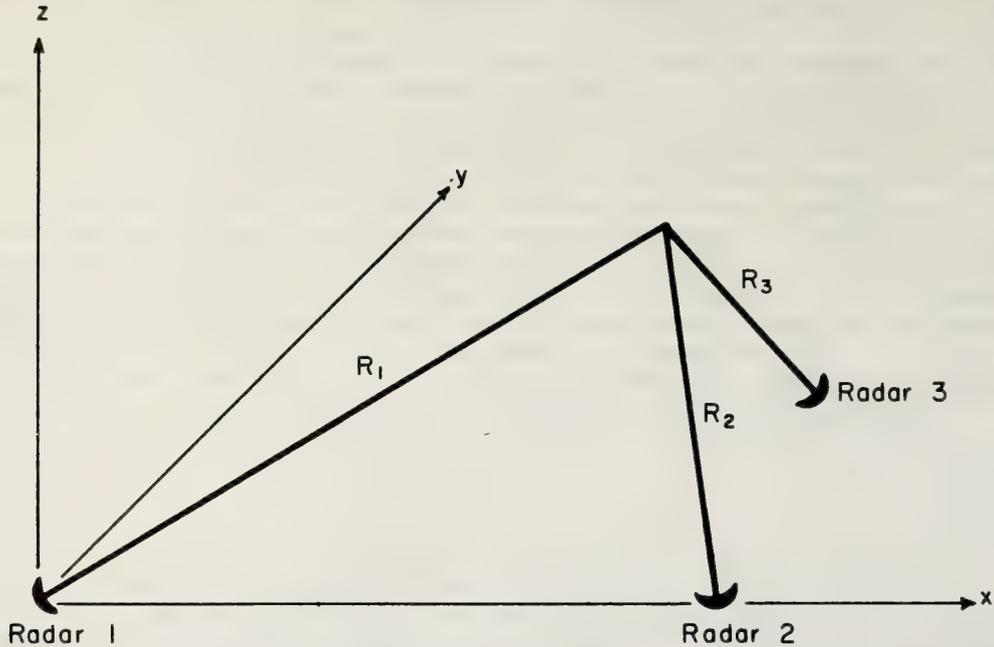


Figure 13.11 Geometry of the 3-radar coordinate system. Radar 1 is located at $(0, 0, 0)$; Radar 2 at $(x_2, 0, 0)$; Radar 3 at $(x_3, y_3, 0)$.

The equations to be solved for u , v , and $V_f = w + V_t$ are

$$\begin{aligned} V_1 R_1 &= ux + vy + (w + V_t) z \\ V_2 R_2 &= u(x-x_2) + vy + (w + V_t) z \\ V_3 R_3 &= u(x-x_3) + v(y-y_3) + (w + V_t) z \end{aligned} \quad (13:67)$$

where the subscripts refer respectively to radars 1, 2, and 3. Expression (13:67) contains four unknowns and a unique solution is obtained only with inclusion of the equation of continuity (13:53). Methods of solution for both two and three Doppler radar networks are outlined by Armijo (1969). The solution of (13:67) for two radars assumes knowledge of $V_t(x, y, z)$ or alternately restricts the observations to the lower regions of the convective storm where the vertical component can be neglected.

The proposed multi-Doppler methods are difficult to implement because of logistics problems. For example, it is not practical to require that all the radars observe the same point in the storm at the same time. The time required to mechanically steer the radar beam so that observations are obtained coincidentally at all points in the storm would be large compared to the evolution time of the storm. It is more practical to independently scan the radar beams and simultaneously observe the radial velocity at many range gates. This technique in turn assumes sufficient time stationarity so that data from separate radars at different times can be assembled in space and time and solutions to (13:67) can be obtained. A partial solution to this problem for two radars has been proposed by Lhermitte and Miller (1970). Their technique, called COPLAN scanning, requires that two radars simultaneously scan the radial velocity field in a plane which is tilted with respect to the ground (F13.12). The elevation angle of each radar is controlled electronically so that as the radars scan in azimuth the following condition is satisfied

$$\frac{\tan \theta_1}{\cos \beta_1} = \frac{\tan \theta_2}{\cos \beta_2} = \tan \alpha = \text{constant} \quad (13:68)$$

where α is the angle that the plane makes with the horizontal and the subscripts 1 and 2 refer to the two radars. The angle β is measured clockwise from the direction perpendicular to the baseline.

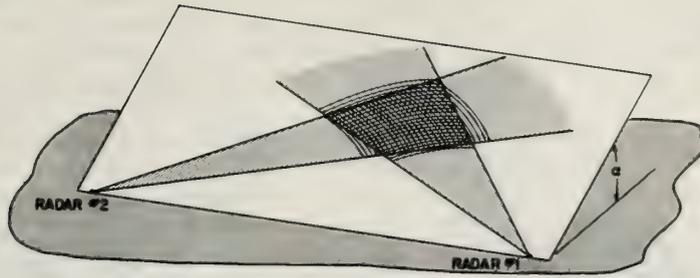


Figure 13.12 *COPLAN scanning. The azimuth and elevation angles of the two radar beams are controlled so that coplanar scanning is realized in a plane with tilt α (after Lhermitte and Miller, 1970).*

Other difficulties of the multiple radar system involve the high rate of data acquisition. Lhermitte (1968c) suggested that a practical means of handling this enormous amount of data would be to record the digitized return signal for several range gates along with appropriate "housekeeping" data on magnetic tape for later processing by a large general purpose computer. Such a scheme has been implemented at NOAA-WPL and operates quite satisfactorily. The difficulty of such a system is that real-time knowledge of the velocity field is not obtained. Such knowledge is desirable to make adjustments in the radar and in the scanning mode to optimize the experiment. Real-time knowledge of the radial velocity spectra can be realized by use of a fast Fourier transform apparatus and a spectral moment computer installed in each radar. Such devices are in various stages of exploration at NOAA-WPL, NOAA-NSSL, and the University of Miami.

Finally, let us look at an example of the wind field inside a thunderstorm as determined from analysis of the radial velocity fields obtained from two X-band Doppler radars (Lhermitte, 1970). The data were acquired just east of Boulder, Colorado, on August 29, 1969 in a fairly "well behaved" thunderstorm which moved toward the northeast at 9 msec^{-1} during the several minutes of simultaneous data collection. Observations were restricted to the lower (subcloud) region of the thunderstorm so that the vertical components in (13:67) could be neglected. Figure (13.13) shows the horizontal streamline and reflectivity patterns at four different mean altitudes. Note that the horizontal circulation is fairly complicated, but apparently well organized, with regions of confluence and vorticity that are displaced toward the northwest at increasing altitudes. Figure (13.13a) includes computation of the convergence and (F13.13b) includes an estimate of the vertical motion.

Although Lhermitte's data were acquired by independent scanning with radial velocities thus obtained as much as several minutes apart, the results appear to be internally consistent. The concept of multiple Doppler observations of a thunderstorm has at least been proven to be feasible. Many refinements of the technique, along with implementation of coordinated beam scanning (such as COPLAN) and real-time data processing need to be implemented. Many thunderstorms of various types and intensities and in various stages of development need to be probed with multiple Doppler networks before the best scanning technique is realized and before any generalizations can be made about the structure of the velocity fields.

13.2.5 Special Observational Techniques

Several specialized observational schemes that involve the Doppler radar concept have been applied or proposed for observations of limited types of atmospheric phenomena. A few of these will be briefly mentioned here.

1. Airborne Doppler Radar (Lhermitte, 1971). This technique proposes that a meteorological Doppler radar be installed in an aircraft for studies of thunderstorms and hurricanes. The technique takes advantage of the excellent mobility of the aircraft to rapidly scan the storm and minimize the time stationarity/space homogeneity problems encountered with ground based scanning techniques. The airborne Doppler can also go "where the action is," for example it can be used to map the velocity field in a hurricane that is far from any ground based radar. There are many expected difficulties in implementation

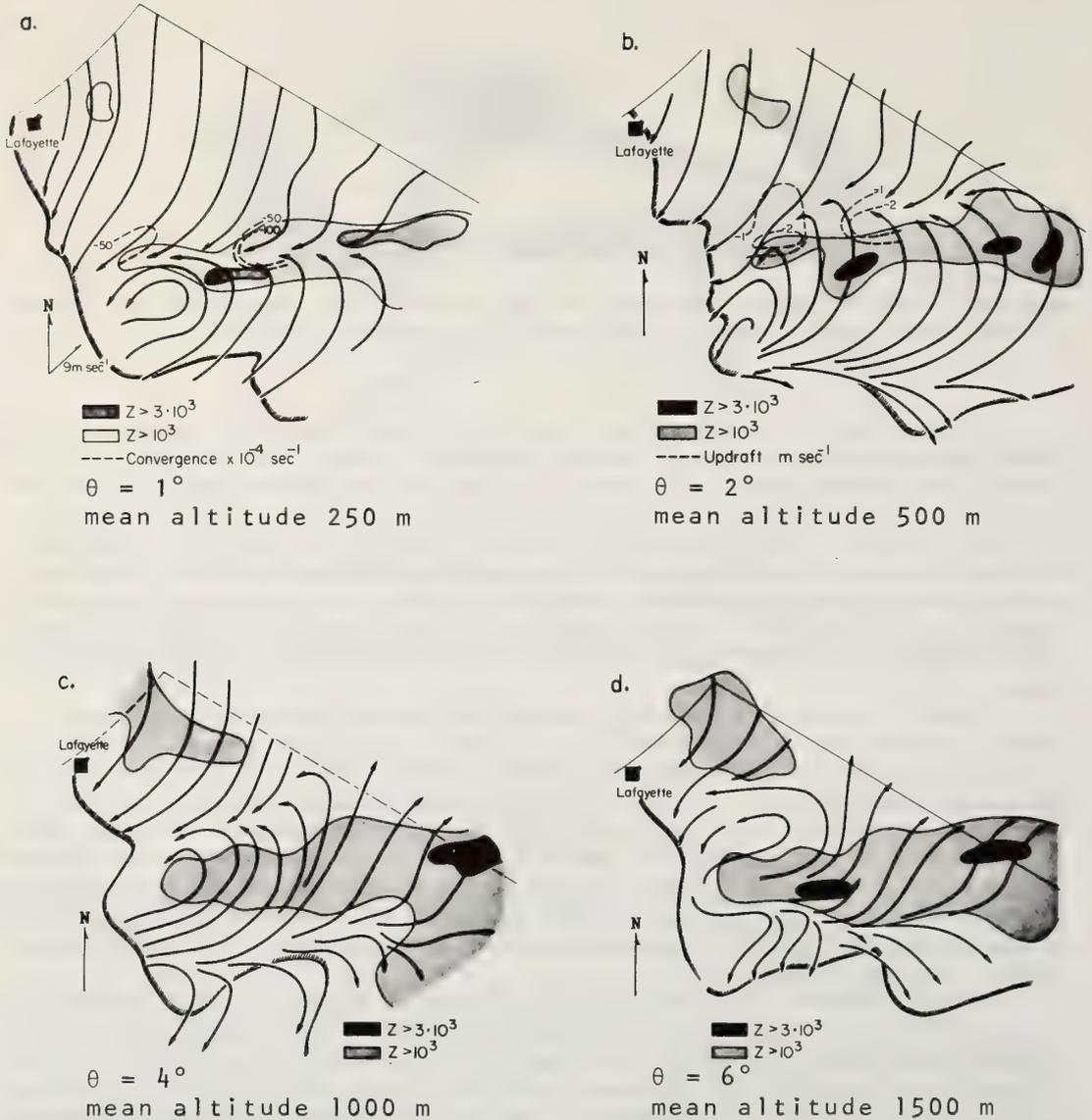


Figure 13.13 Streamline analysis from dual-Doppler radar data obtained on August 29, 1969, at 1414 MST. The data are shown for several mean altitudes. Figure 13.13a includes computation of convergence from the wind field, and Figure 13.13b includes an estimate of vertical air motion. Note the vorticity and confluence regions and their displacement towards the northwest with increasing altitudes. Echo advection (230° , 9 m sec^{-1}) is also indicated (after Lhermitte, 1970).

of the technique (cost, antenna size limitations, attitude and velocity of the aircraft) but none of these are insurmountable and the scheme has great promise for acquiring meteorologically significant results. Lhermitte proposes that initially a vertically pointing beam be used so that the ground speed of the aircraft can be measured. This permits determination of target velocities with respect to the ground rather than the aircraft.

2. Use of Artificial Targets – Chaff. A primary shortcoming of radar used in probing velocity fields in the atmosphere is that detectable targets must be present. (Some of the more advanced ultrasensitive radars can sometimes see enough “angel” echoes to derive information on the velocity spectrum in the clear air (Dobson, 1970)). The absence of detectable targets can be partially overcome by dispensing artificial targets, chaff dipoles, in the atmosphere. These dipoles are typically thin strands (~ 1 mil) of aluminum coated fiberglass that are cut to one half the radar wavelength. A few chaff targets present in the radar beam at several kilometers will backscatter sufficient electromagnetic energy to be detectable. The chaff needles are excellent tracers of the air motion and have an advantage over precipitation targets in that their terminal velocity is small and known (~ 25 cm sec⁻¹). The main disadvantage is that it is often difficult to dispense the chaff over a sufficiently large area to determine the velocity field.

3. Plan Shear Indicator (Armstrong and Donaldson, 1969, Donaldson, 1970a, 1970b). This technique provides a real-time indication of radial and tangential gradients of velocity by displaying the velocity information on an intensity modulated plan position indicator (PPI) scope called a plan velocity indicator. The PVI allows qualitative determination of the location and intensity of regions of convergence and vorticity as well as an indication of abnormal shear. The real-time advantage is achieved at the expense of mediocre resolution in range and velocity. Details of the scheme and examples of its application are described in the above referenced papers.

4. Random Signal Radar. The principle limitations of the coherent pulse Doppler radar method of measuring the motion of atmospheric tracers are that the range resolution is limited to $c\tau/2$ (τ is the pulse width of the coherent transmitter), and that the product of maximum unambiguous range (13:35) and maximum unambiguous velocity (13:34) is $c\lambda/8$. Note that both r_{\max} and V_{\max} depend on the pulse repetition frequency so that a higher pulse repetition frequency leads to larger V_{\max} but smaller r_{\max} . The range resolution limitation arises because of the difficulty of generating high peak power pulses with a width of less than 0.1 microseconds – range resolution better than about 15 meters.

A method of overcoming these limitations is potentially available in the random signal radar (McGillem, et al., 1969). This radar transmits pulses of wideband noise at a high repetition rate which can easily be varied to eliminate velocity ambiguity for meteorological targets. The transmitted pulses are independent because the transmitted signal is random noise. The signal return from a given transmitted pulse can then occur after the transmission of many other pulses without leading to range ambiguity. The range resolution, determined by the bandwidth (B) of the transmitted signal can be made to match most meteorological applications. The range resolution is of the order of $c/2B$. Large bandwidths can be obtained by amplifying the noise in a traveling wave tube, so that 100 MHz bandwidths (1.5 meters resolution) are easily generated.

A further advantage of the random signal radar is that Doppler information can be extracted with a simple bandpass filter, with a velocity resolution given by $\lambda/2W$ where W is the filter bandwidth. A 3-cm radar with a 10-hz filter would give a velocity resolution of .15 m/sec.

The principle disadvantage of the random signal radar for meteorological targets is that the backscatter from ranges other than the range being examined will appear as additional (uncorrelated) noise. Random signal radar has been demonstrated at short ranges with point targets. The potential of random signal radar in meteorology has yet to be explored.

13.3 Conclusions

This chapter contains a brief summary of the principles, applications, and potential applications of pulsed Doppler radar for remote sensing of the atmosphere. No attempt has been made to describe any of the techniques or results in detail. Instead, the authors have attempted to provide a reasonably complete bibliography, with the text serving to acquaint the reader with the field and to point him toward the appropriate literature.

It has been shown that a pulsed Doppler radar is a very accurate phase measuring device. Proper mathematical treatment (Fourier transformation) of the time dependent backscattered signal allows

estimation of the radial component of the target velocity spectrum within a scattering volume. In addition, measurements of the backscattered signal amplitude are related to cross sections, reflectivity, and rainfall rate.

Utilization of the radial velocity spectra to derive information on the three-dimensional air circulation depends on the investigator's ability to relate target velocity to the air velocity. For example, it was seen that knowledge of the character of the targets (i.e., rain, snow, chaff, etc.) often allows one to estimate the relationship between target motion and air motion. In studies of many types of large scale phenomena such as cyclonic storms, time stationarity and horizontal homogeneity assumptions are often valid allowing one to properly assemble velocity information at many points in space and time to derive estimates of the mean wind and turbulence fields. Conversely, it was shown that for certain types of atmospheric phenomena such as convective storm systems, such assumptions can only be made for short time periods and over small regions of space. In this case, it is necessary to employ a network of two or more Doppler radars to resolve the three-dimensional wind field from independently determined radial velocity components.

13.4 Acknowledgments

The authors wish to thank Mr. Richard Strauch for contributing the section on random signal radar and for his general review of the entire paper.

13.5 References

- Armijo, L. (1969), A theory for the determination of wind and precipitation velocities with Doppler radars, *J. Atmos. Sci.* 26, 570-573.
- Armstrong, G. M. and R.,J. Donaldson, Jr. (1969), Plan shear indicator for real-time Doppler radar identification of hazardous storm winds, *J. Appl. Meteorol.* 8, No. 3, 376-383.
- Battan, L. J. 1959: *Radar Meteorology*, Univ. of Chicago Press, Chicago.
- Battan, L. J. (1964), Some observations of vertical velocities and precipitation sizes in a thunderstorm, *J. Appl. Meteorol.* 3, 415-420.
- Battan, L. J. and J. B. Theiss (1968), Measurement of draft speeds in convective clouds by means of pulsed-Doppler radar, *Proc. 13 Radar Meteorol. Conf.*, Boston, Am. Meteorol. Soc. 26-29.
- Blackman, R. B. and J. W. Tukey, 1958: *The Measurement of Power Spectra*, Dover Publications, New York.
- Best, A. C. (1950), Empirical formulae for terminal velocity of water drops falling through the atmosphere, *Quart. J. Roy. Meteorol. Soc.* 76, 302-311.
- Browning, K. A. and R. Wexler (1966), The determination of kinematic properties of a wind field using a single Doppler radar, *Proc. 12th Weather Radar Conf.*, Boston, Am. Meteorol. Soc., 125-127.
- Caton, P. G. F. (1963), The measurement of wind and convergence by Doppler radar, *Proc. 10th Weather Radar Conf.*, Boston, Am. Meteorol. Soc., 190-196.
- Dobson, E. B. (1970), Doppler radar measurements of mean wind variations in the clear atmosphere, *Preprints 14th Radar Meteorol. Conf.* Boston, Am. Meteorol. Soc., 69-72.
- Donaldson, R. J., Jr. (1967), Measurement of air motion in a thunderstorm anvil by Doppler radar, Paper presented at the 48th meeting Am. Meteorol. Soc., San Francisco, *American Meteorol. Soc. Bulletin*, 48, 838.

- Donaldson, R. J. Jr. (1970a), Vortex signature recognition by Doppler radar, *J. Appl. Meteorol.* 9, No. 4, 661-670.
- Donaldson, R. J. Jr. (1970b), Severe weather warning by plan shear indicator, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 123-126.
- Dyer, R. M. (1970), Particle fall speeds within the melting layer, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 157-160.
- Goldstein, L. (1946), Radio wave propagation experiments, Summary Tech. Rept. Commit. Propagation National Defense Research Council, 2, Washington.
- Gunn, K. L. S. and T. W. R. East (1954), The microwave properties of precipitation particles, *Quart. J. Roy. Meteorol. Soc.*, 80, 522-545.
- Gunn, R. and G. D. Kinzer (1949), The terminal velocity of fall of water droplets in stagnant air, *J. Meteorol.* 6, 243-248.
- Joss, J. and A. Waldvogel (1970), Raindrop size distributions and Doppler velocities, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 153-156.
- Kerr, D. E., ed., 1951: *Propagation of Short Radio Waves*, MIT Radiation Laboratory Series, 13, McGraw-Hill, New York.
- Kraus, J. D., 1950: *Antennas*, McGraw-Hill, New York.
- Lhermitte, R. M. (1963), Motions of scatterers and the variance of the mean intensity of Weather Radar Signals, Sperry Rand Research Program 38310 SPRC-RR-65-57, Atmospheric Physics Dept., Sudbury, Mass.
- Lhermitte, R. M. (1968a), New developments in Doppler radar methods Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 14-17.
- Lhermitte, R. M. (1968b), Turbulent air motion as observed by Doppler radar, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 489-503.
- Lhermitte, R. M. (1968c), Atmospheric probing by Doppler radar, Proc. Tech. Sessions Remote Atmospheric Probing Panel of the Natl. Acad. Sci. Committee on Atmospheric Science (NAS-CAS), Chicago, 253-285.
- Lhermitte, R. M. (1970), Dual-Doppler radar observation of convective storm circulation, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 139-144.
- Lhermitte, R. M. (1971), Probing of atmospheric motion by airborne pulse-Doppler radar techniques, *J. Appl. Meteorol.* 10, No. 2, 234-246.
- Lhermitte, R. M. and D. Atlas (1961), Precipitation motion by pulse Doppler radar, Proc. 9th Weather Radar Conf., Boston, Am. Meteorol. Soc., 218-223.
- Lhermitte, R. M. and L. J. Miller (1970), Doppler radar methodology for the observation of convective storms, Preprints 14th Radar Meteorol., Conf., Boston, Am. Meteorol. Soc., 133-138.
- Lumley, J. L. and H. A. Panofsky, 1964: *The Structure of Atmospheric Turbulence, Monographs and Texts in Physics and Astronomy XII*, Interscience Publishers, a Division of John Wiley & Sons, New York.

- McGillen, C. D., Cooper, G. R., and W. B. Waltman (1969), Use of wideband stochastic signals for measuring range and velocity, EASCON Record, IEEE Electronics and Aerospace Convention, 305-311.
- Mel'Nichuk Yu, V., G. A. Smirnova, and A. A. Chernikov (1968), Radar measurements of the rate of turbulent energy dissipation in clouds and precipitation, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 486-489.
- Mie, G. (1908), Beiträge zur Optik trüber Medien, speziell kolloidales Metallösungen, Ann. Physik, 25, 377.
- Nathanson, F. E. and J. P. Reilly (1968a), Radar precipitation echo experiments on temporal, spatial, and frequency correlation, IEEE Trans., v. AES-4, No. 4, 505-514.
- Newton, C. W. (1967), Severe convective storms, Adv. in Geophys. 12, 257-308.
- Norgaard, D. W. (1956), The phase-shift method of single-sideband signal reception, Proc. IRE, No. 44, 1735-1743.
- Probert-Jones, J. R. (1960), The analysis of Doppler radar echoes from precipitation, Proc. 8th Weather Radar Conf., Boston, Am. Meteorol. Soc., 377-385.
- Rogers, R. R. (1964), An extension of the Z-R relationship for Doppler radar, World Conf. Radar Meteorol. and 11th Weather Radar Conf., 158-160.
- Rogers, R. R. (1966a), Doppler radar investigation of Hawaiian rain, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc. 128-134.
- Rogers, R. R. (1966b), Project Hawaii—an investigation of rain on the island of Hawaii, Technical Report, Cal No. VC-2049-P-1, Cornell Univ.
- Rogers, R. R. (1967), Doppler radar investigation of Hawaiian rain, Tellus 19, 432-455.
- Rogers, R. R. and R. J. Pilié (1962), Radar measurements of drop size distribution, J. Atmos. Sci. 19, 503-506.
- Rogers, R. R. and B. R. Tripp (1964), Some radar measurements of turbulence in snow, J. Appl. Meteorol. 3, 603-610.
- Ryde, J. W. (1946), The attenuation and radar echoes produced at centimetre wavelengths by various meteorological phenomena, *Meteorological Factors in Radio Wave Propagation*, London, Phys. Soc., 169-188.
- Sekhon, R. S. and R. C. Srivastava (1970), Doppler radar observation of droplet size distribution in a thunderstorm, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 171-176.
- Smith, P. L., Jr., (1970), Weather Radar (unpublished manuscript), South Dakota School of Mines and Technology, Rapid City.
- Stackpole, J. (1961), The effectiveness of raindrops as turbulence sensors, Proc. 9th Weather Radar Conf., Boston, Am. Meteorol. Soc., 212-217.
- Stratton, J. A., 1941: *Electromagnetic Theory*, McGraw-Hill, New York.
- Wilson, D. A. (1970a), Doppler radar studies of boundary layer wind profile and turbulence in snow conditions, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 191-196.

Wilson, D. A. (1970b), The kinematic behavior of spherical particles in an accelerating environment, U.S. Dept. of Commerce, NOAA Technical Report ERL 187-WPL 13.

13.6 Bibliography

Aoyagi, J., M. Fujiwara, Z. Yangisawa, and N. Kodaira (1966), Doppler radar observations of snow showers, Proc. 12th Weather Radar Conf., Boston, Am. Met. Soc. 112-116.

Aoyagi, J. (1968a), Mean Doppler velocities of precipitation near the ground, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 22-25.

Aoyagi, J. (1968b), A multi-channel Doppler frequency analyzer, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 324-327.

Armijo, L. (1970), Computer simulation of severe storm observations with Doppler radars, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 145-146.

Armstrong, G. M. and R. J. Donaldson, Jr. (1968), A convenient indicator of tangential shear in radial velocity, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 50-52.

Atlas, D. (1963), Radar analysis of severe storms, Meteorol. Mono. 5, 27, 177-220.

Atlas D. (1964), Advances in radar meteorology, Adv. Geophys. 10, New York, Academic Press, 317-478.

Atlas, D., J. Aoyagi, and R. J. Donaldson, Jr. (1965), Doppler analysis of the physical dynamics of a convective storm, Proc. Conf. Cloud Physics, Tokyo, 314-318.

Atlas, D. (1966), The balance level in convective storms, J. Atmos. Sci. 23, 635-651.

Atlas, D. and R. Tatehira (1968), Precipitation-induced mesoscale wind perturbations, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 166-175.

Atlas, D., R. Tatehira, R. C. Srivastava, W. Marker, and R. F. Carbone (1969), Precipitation-induced mesoscale wind perturbation in the melting layer, Quart. J. Roy. Meteorol. Soc. 95, 544-560.

Battán, L. J. (1963a), Some observations of vertical velocities and precipitation sizes in a thunderstorm, Proc. 10th Weather Radar Conf., Boston, Am. Meteorol. Soc., 303-308.

Battán, L. J. (1963b), The vertical velocities of angel echoes, Proc. 10th Weather Radar Conf., Boston, Am. Meteorol. Soc., 309-315.

Battán, L. J. and J. B. Theiss (1967), Measurement of draft speeds in convective clouds by means of pulsed-Doppler radar, Univ. of Arizona, Inst. Atmos. Phys., Sci. Rept. No. 22.

Borresen, J. A. (1970), Use of Doppler VAD-pattern to detect shear zones and turbulence in a snowstorm, J. Appl. Meteorol. 10, No. 3, 433-442.

Boucher, R., R. Wexler, D. Atlas, and R. Lhermitte (1965), Meso-scale wind structure revealed by Doppler radar, J. Appl. Meteorol. 4, 590-597.

Boucher, R. J. (1968), Characteristics of turbulent structures observed by Doppler radar in snow, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 480-485.

Boucher, R. J. and H. Ottersten (1971), Doppler radar observation of wind structure in snow, J. Appl. Meteorol. 10, No. 2, 228-233.

- Boyenvall, E. H. (1960), Echoes from precipitation using pulse Doppler radar, Proc. 8th Weather Radar Conf., Boston, Am. Meteorol. Soc., 57-64.
- Brantley, J. Q., Jr. (1957), Some weather observations with a continuous wave Doppler radar, Proc. 6th Weather Radar Conf., Boston, Am. Meteorol. Soc., 297-306.
- Brook, M., and D. Latham (1968), The fluctuating radar echo. I. Modulation by vibrating drops. Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 2-7.
- Brown, R. A. and R. L. Peace, Jr. (1968), Mesoanalysis of convective storms utilizing observations from two Doppler radars, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 188-191.
- Browning, K. A. and D. Atlas (1966), Velocity characteristics of some clear air dot angels, J. Atmos. Sci. 23, 592-604.
- Browning, K. A. and R. Wexler (1968), A determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteorol. 7, 105-113.
- Browning, K. A., T. W. Harrold, A. M. Whyman, and J. G. D. Beimers, (1968a), Horizontal and vertical air motion and precipitation growth within a shower, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 122-127.
- Browning, K. A., T. W. Harrold, A. M. Whyman, and J. G. D. Beimers, (1968b), Horizontal and vertical air motion and precipitation growth within a shower, Quart. J. Roy. Meteorol. Soc. 94, 498-509.
- Browning, K. A. and T. W. Harrold (1969), Air motion and precipitation growth in a wave depression, Quart. J. Roy. Meteorol. Soc. 95 288-309.
- Browning, K. A., T. W. Harrold, and J. R. Starr (1970), Richardson number limited shear zones in the free atmosphere, Quart. J. Roy. Meteorol. Soc. 96, 40-49.
- Browning, K. A. and T. W. Harrold, (1970), Air motion and precipitation growth at a cold front, Quart. J. Roy. Meteorol. Soc. 96, 369-389.
- Chmela, A. C. (1970), Propagation of the severe convective storm from Doppler observations, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 121-122.
- Cooley, J. W. and J. Tukey (1965), An algorithm for the machine calculation of complex Fourier series, Mathematics of Computation 19, 297-301.
- Cox, E. G. and H. L. Groginsky (1966), An analysis of the estimates of wind parameters using the VAD technique, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 44-51.
- Donaldson, R. J., Jr., G. M. Armstrong, and D. Atlas (1966), Doppler measurements of horizontal and vertical motions in a paired instability line, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 392-397.
- Donaldson, R. J., Jr. (1967a), A preliminary report on Doppler radar observation of turbulence in a thunderstorm, Air Force Cambridge Research Laboratories Environmental Research Paper No. 255, ARCRL 67-0015.
- Donaldson, R. J., Jr. (1967b), Horizontal wind measurement by Doppler radar in a severe squall line, Air Force Cambridge Research Laboratories, Proc. Conf. on Severe Local Storms, St. Louis, 89-98.
- Donaldson, R. J., Jr. and R. Wexler (1968). Notes on thunderstorm observation by fixed-beam Doppler radar, J. Atmos. Sci. 25, No. 1, 139-144.

- Donaldson, R. J., Jr., and A. C. Chmela (1968), Distribution of vertical velocity mean and variance in a thunderstorm, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 492-297.
- Donaldson, R. J., Jr., (1971), Mapping a thunderstorm anvil flow by Doppler radar, J. Appl. Meteorol. 9, No. 6, 911-915.
- Dyer, R. M. (1968), Doppler measurements in stratiform rain, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 144-147.
- Easterbrook, C. C. (1967), Some Doppler radar measurement of circulation patterns in convective storms, J. Appl. Meteorol. 6, 882-888.
- Ekpenyong, B. E. and R. C. Srivastava (1970), Radar characteristics of the melting layer—A theoretical study, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 161-166.
- Foote, G. B. (1968), Variance spectrum analysis of Doppler radar observations in continuous precipitation, J. Appl. Meteorol. 7, No. 3, 459-464.
- Foote, G. B. and P. S. duToit (1969), Terminal velocity of raindrops aloft, J. Appl. Meteorol. 8, No. 2, 249-253.
- Glover, K. M. (1966), The feasibility of detecting shock waves by pulse Doppler radar, AFCRL-66-378, Environmental Research Paper No. 200.
- Glover, K. M., A. W. Bishop, and W. Lob (1968), Wind measurement by dual beam radar, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 456-463.
- Gorelik, A. G. (1965), Simultaneous measurements of Lagrangian and Eulerian turbulence in snow precipitation, Atmospheric Oceanic Physics Series 1, No. 9, 989-991 (translated by Kenneth Syers).
- Gorelik, A. G. (1968), Wind structure investigations of boundary layer by radar "clear air" returns, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 248-251.
- Gorelik, A. G., V. V. Kostarev and A. A. Chernikov (1962), New possibilities of radar measurement of the wind, Meteorologiya i Gidrologiya 7, Moscow, 34-39.
- Gorelik, A. G., V. Mel'Nichuk, and A. A. Chernikov (1965), The statistical characteristics of the radar echo as a function of the dynamic processes and microstructure of the meteorological entity, AFCRL, T-R-479.
- Gorelik, A. G. and V. Mel'Nichuk (1966), Radar measurements of turbulent parameters in clouds and precipitations (abstract only), Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 104.
- Gorelik, A. G. and V. E. Logunov (1968), Determination of vertical air motion velocity in rainfall by Doppler radar, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 18-21.
- Groginsky, H. L. (1966), Digital processing of the spectra of pulse Doppler radar precipitation echoes, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 30-34.
- Groginsky, H. L. (1968a), Unambiguous measurement of updraft velocity and drop size distribution, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 30-34.
- Groginsky, H. L. (1968b), Scanning requirements of the (VAD) operation of a pulse Doppler weather radar (abstract of late paper), Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 35.

- Harris, F. I., B. E. Ekpenyong, and D. Atlas (1970), Doppler winds, stalactites, and secondary generators at the evaporation base of an ice crystal overcast, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 181-186.
- Harrold, T. W. (1966), Measurement of horizontal convergence in precipitation using a Doppler radar—a case study, *Quart. J. Roy. Meteorol. Soc.* 92, 31-40.
- Harrold, T. W. and K. A. Browning (1967), Mesoscale wind fluctuations below 1500 meters, *Meteorological Magazine* 96, 367-376.
- Harrold, T. W. and K. A. Browning (1968), Low-level airflow at a cold front, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 222-225.
- Holmes, D. W. and R. L. Smith (1958), Doppler radar for weather investigations, Proc. 7th Weather Radar Conf., Boston, Am. Meteorol. Soc., F-29 to 4-36.
- Hunt, F. R. (1970), Precipitation measurements with a variable velocity notch receiver, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 177-180.
- Kodaira, N. (1964), A pulsed-Doppler radar for weather observations, Proc. World Conf. on Radar Meteorol., Boston, Am. Meteorol. Soc., 300-303.
- Kodaira, N. and Z. Yanagisawa (1965), The MRI pulse Doppler radar, Proc. Internatl. Conf. on Cloud Physics, Tokyo, 309-313.
- Kraus, M. J. (1970), Doppler radar investigations of flow patterns within severe thunderstorms, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 127-132.
- Lhermitte, R. M. (1960a), Variations de la vitesse de chute des particules d'une précipitation étendue, à différents niveaux, *Comptes Rendus des Séances de l'Académie des Sciences* 250, 899-901.
- Lhermitte, R. M. (1960b), The use of a special pulse Doppler radar in measurement of particles fall velocities, Proc. 8th Weather Radar Conf., Boston, Am. Meteorol. Soc., 269-275.
- Lhermitte, R. M. (1962), Note on wind variability with Doppler radar, *J. Atmos. Sci.* 19, 343-346.
- Lhermitte, R. M. (1963), Weather echoes in Doppler and conventional radars, Proc. 10th Weather Radar Conf., Boston, Am. Meteorol. Soc., 323-329.
- Lhermitte, R. M. (1964), Doppler radars as severe storm sensors, *Bul. Am. Meteorol. Soc.* 45, 587-596.
- Lhermitte, R. M. (1966a), Probing air motion by Doppler analysis of radar clear air returns. *J. Atmos. Sci.* 23, 575-591.
- Lhermitte, R. M. (1966b), Application of pulse Doppler radar technique to meteorology, *Bul. Am. Meteorol. Soc.* 47, No. 9, 703-711.
- Lhermitte, R. M. (1966c), Doppler observation of particle velocities in a snowstorm, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 117-124.
- Lhermitte, R. M. and D. Atlas (1963), Doppler fall speed and particle growth in stratiform precipitation, Proc. 10th Weather Radar Conf., Boston, Am. Meteorol. Soc., 297-302.
- Lhermitte, R. M. and J. Dolley (1966), Study of the motion of clear air targets, Proc. 12th Weather Radar Conf., Boston, Am. Meteorol. Soc., 293-299.

- Lhermitte, R. M. and E. Kessler (1964), An experimental pulse Doppler radar for severe storm investigations, Proc. World Conf. on Radio Meteorol. Boston, Am. Meteorol. Soc., 304-309.
- Lofgren, G. R. and L. J. Battan (1969), Polarization and vertical velocities of dot angel echoes, J. Appl. Meteorol. 8, No. 6, 948-851.
- Peace, R. L., Jr. and R. A. Brown (1968), Single and double radar velocity measurements in convective storms, Proc. 13th Radar Meteorol. Conf. Boston, Am. Meteorol. Soc., 464-470.
- Peace, R. L., Jr., R. A. Brown, and H. G. Camnitz (1969), Horizontal motion field observations with a single pulse Doppler radar, J. Atmos. Sci. 26, No. 5, 1096-1103.
- Pilie, R. J., J. E. Jiusto, and R. R. Rogers (1963), Wind velocity measurement with Doppler radar, Proc. 10th Wea. Radar Conf., Boston, Am. Meteorol. Soc., 329a-329L.
- Probert-Jones, J. R. and W. G. Harper (1961), Vertical air motion in showers as revealed by Doppler radar, Proc. 9th Weather Radar Conf., Boston, Am. Meteorol. Soc., 225-232.
- Rogers, R. R. (1963), Investigation of precipitation processes, Cornell Aero. Lab. Technical Report, VC-1660-P-2.
- Rogers, R. R. and A. J. Chimera (1960), Doppler spectra from meteorological radar targets, Proc. 8th Weather Radar Conf., Boston, Am. Meteorol. Soc., 377-385.
- Serafin, R. J. and L. C. Peach (1970), Calibration of vertically pointing Doppler radar for meteorological echoes, Proc. 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 147-152.
- Sloss, P. W. and D. Atlas (1968), Wind shear and reflectivity gradient effects on Doppler radar spectra, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 44-49.
- Spilhaus, A. F. (1948), The distribution of raindrops with size, J. Meteorol. 5, 161-164.
- Sweeney, H. J. (1970), Turbulence in the lower atmosphere, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 197-202.
- Tatehira, R. and R. C. Srivastava (1968), Note on updraft estimation with Doppler radar, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 36-43.
- Theiss, J. B. (1963), More target data with sideband coherent data, Electronics 36, 40-43.
- Tripp, R. (1964), The CAL pulse Doppler radar, World Conf. Radar Meteorol. and 11th Weather Radar Conf., Boston, Am. Meteorol. Soc., 330-337.
- Wexler, R. (1968), Doppler radar measurements in a rainstorm with a spiral band structure, Proc. 13th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 192-193.
- Wexler, R., A. C. Chmela, and G. M. Armstrong (1967), Wind field observations by Doppler radar in a New England snowstorm, Monthly Weather Review, 95, 12, 929-935.
- Wiley, R. L., K. A. Browning, J. Joss, and A. Waldvogel (1970), Measurement of drop size distribution and vertical air motion in widespread rain using pulsed Doppler radar and distrometer, Preprints 14th Radar Meteorol. Conf., Boston, Am. Meteorol. Soc., 167-170.
- Wilson, D. A. (1963), Drop size distribution as recorded by pulsed Doppler radar, Univ. of Arizona Masters' thesis.

List of Symbols

ω	Angular frequency	Σ	Sum of $n = 1, 2, \dots, N$ particles
ϕ	Phase and off-axis vertical radar beam angle	vol	Dielectric factor
λ	Wavelength	$ K ^2$	Radar reflectivity factor
R_r	Radar range	Z	Doppler shifted frequency
P_r	Radar return power	f_d	Nyquist frequency
P_t	Average peak transmitted power	f_N	Pulse repetition period
G	Antenna gain function	T	In-phase signal from phase demodulator
A_e	Antenna effective area	I	Quadrature signal from phase demodulator
A_c	Geometric cross-sectional area	Q	Doppler velocity spectrum
V_R	Doppler radial velocity	S	Doppler frequency spectrum
t	Time	P	Doppler signal amplitude
A_p	Antenna aperture area	ρ	Azimuth angle
σ	Back scattering cross section	β	Verticle velocity of particles
η	Radar reflectivity	V_f	Terminal fall velocity of particles
V	Radar pulse volume	V_t	Horizontal components of air motion
θ	Off-axis horizontal beam angle, elevation angle	u, v	Vertical component of air motion
c	Propagation speed for electromagnetic radiation	w	Horizontal air motion
τ	Pulse width (time)	V_H	Temporal or spatial average
$\sigma()$	Standard deviation of ()	\bar{O}	Covariance
\int_{vol}	Volume or triple integral	cov	Turbulent kinetic energy production rate
\ln	Natural logarithm	P	Fluctuating component
m	Complex index of refraction	$()'$	COPLAN tilt angle
		α	

Chapter 14 STUDIES OF THE CLEAR ATMOSPHERE USING HIGH POWER RADAR

Kenneth R. Hardy
Meteorology Laboratory
Air Force Cambridge Research Laboratories

This chapter describes some of the key results on the structure of the clear atmosphere as derived from observations with high power radars having steerable antennas. Most of the dot echoes in the air are caused by insects or birds. Backscattering from refractive index variations is the other major source of clear air radar echoes. Theoretical relationships between the radar reflectivity, the refractive index microstructure, and the intensity of clear air turbulence are presented. Experimental results of radar investigations within the clear atmosphere of convection, turbulence, and a variety of wave structures are summarized.

14.0 Introduction

14.0.1 Clear Air Radar Echoes

Radar echoes from an apparently clear atmosphere have been observed since the early days of radar. Because the origin of the echoes was usually a mystery, the clear air echoes were often called "radar angels." There is a variety of angel phenomena reported in the literature, and there have been numerous attempts to arrive at a plausible and satisfactory explanation of their origin. Knowledge covering the period up to 1964 on the characteristics of angel echoes has been ably presented by Plank (1956) and Atlas (1959 and 1964). These authors were principally concerned with the cause of the echoes, and in some cases they were not able to provide satisfactory explanations of the observations. One of the main reasons for the lack of understanding was the limited quality and quantity of the radar observations. With the application of high power and high resolution radars to the investigation of clear air echoes, Atlas and Hardy (1966) presented what now appears to be a correct explanation of two types of echoes. Although high power radars with steerable antennas were necessary for the verification of the causes of the echoes, in recent years such radars have been applied to investigations of clear air structures and processes, and it is these investigations which will be emphasized in this chapter.

Perhaps another difficulty in the correct identification of angel echoes arose because of the numerous possible causes of the echoes. These include anomalous propagation, ground targets seen by the sidelobes of the antenna beam, second-sweep echoes beyond the unambiguous range of the radar, birds and insects, and variations in the atmospheric index of refraction. It is the scattering from fluctuations in the refractive index which provides the most useful type of radar echoes from the clear atmosphere, although on occasion there are sufficient insects in the air to provide suitable tracers of atmospheric structure and motion. Since the scattering from refractive index variations is generally very weak, it is essential to use sensitive radars for investigations of the clear atmosphere. It is primarily for this reason that the explanation of the echoes was not widely accepted until sensitive radars were applied to meteorological studies of the clear air.

There are only three known installations of high powered radars with steerable antennas which have been devoted to studies of the clear atmosphere. One installation includes three sensitive radars at Wallops Island, Virginia. These radars, operating at wavelengths of 3.2, 10.7, and 71.5 cm, were the first to be used intensively for investigations of the clear air (Hardy et al., 1966; Katz, 1966). A few years later Crane (1970) presented measurements with a 23.2-cm wavelength radar located in Westford, Massachusetts. The third installation includes a powerful pulsed Doppler 10.7-cm radar located in Defford, England (Browning and Watkins, 1970; Browning, 1972). The approximate values for some of the important parameters of these five radars are given in (T14.1). Also included in the table are the values for the WSR-57, a 10.4-cm radar used by the National Weather Service of the United States for detection of precipitation. Note that the minimum detectable reflectivity (assuming the contributing region is filled) is at least a factor of 30 greater for the WSR-57 radar than for the other two 10-cm radars. The Westford 23.2-cm radar is the most sensitive of all of the radars. However, the ground clutter is fairly severe for this radar, and atmospheric

measurements at ranges less than 50 km are difficult to obtain. This chapter will be concerned primarily with the numerous investigations which have been carried out with the five high powered radars listed in (T14.1).

A description of the key properties of clear air echoes from insects and birds and from variations in refractive index will also be included. Because the atmospheric concentration of birds or the larger insects is small, usually only one such target appears in the radar contributing region at a given time, and the echoes appear as dots or as point targets when viewed on a radar scope. In contrast, the echoes from variations in refractive index have considerable horizontal extent and may occur in rather narrow layers or they may outline the boundaries of convective cells.

Table 14.1. Characteristics of the Five Radars with Steerable Antennas Which Have Been Used for Studies of the Clear Atmosphere. For comparison, the characteristics of the WSR-57 radar are included.

	Wallops Island, Virginia			Defford, England	Westford, Mass.	(WSR-57)
Wavelength (cm)	3.2	10.7	71.5	10.7	23.2	10.4
Antenna diameter (m)	10.4	18.4	18.4	25	25.6	3.6
Beamwidth (deg)	0.2	0.5	2.9	0.33	0.6	2.0
Pulse length (μ sec)	2	1.3	1	1.25	10	4
Peak transmitted power (10^6 watts)	0.9	3.0	6.0	1	4	0.4
Minimum detectable reflectivity (cm^{-1})	6×10^{-17}	4×10^{-18}	4.5×10^{-19}	10^{-17}	6×10^{-19}	3×10^{-16}
At a range of (km)	10	10	10	10	100	10

14.0.2 Atmospheric Structure and Clear Air Echoes

Progress on the understanding of radar angels and on the description of the radar structure of the clear atmosphere has been rapid over the past six years. Yaglom and Tatarski (1967) were the editors of a book which included papers on radar angels presented at a colloquium in Moscow in 1965. Also Lane (1967a) organized a study institute which included several presentations on the radar structure of the clear atmosphere. More recently Hardy and Katz (1969) and Ottersten (1969a) have reviewed the progress in the probing of the clear atmosphere using high powered radars. Much of the work to be described in this chapter is taken from the above papers as well as from the papers by Browning (1971) and Harold and Browning (1971).

There are two ways in which the radar backscattering provides information on atmospheric structure. First, the backscattered power is related to the intensity of fluctuations in refractive index within a very narrow range of eddy sizes, centered at one half the radar wavelength. The maximum wavelength which has been used extensively for clear air radar investigations is 71.5 cm, and the eddy sizes responsible for the backscatter will be near 36 cm for this radar, whereas 5 cm will be the atmospheric scale of importance for backscattering at radar wavelengths of 10 cm. Despite the small scale of these eddy sizes, valuable information on the variability of temperature and water vapor and the small scale velocity field can be extracted. There is also some correlation between this small scale structure of refractive index and the mean vertical gradients of refractive index and velocity, and this will be illustrated in 14.1.2. Perhaps of greater significance is the second way that information is obtained. This is through the direct analysis of the echo regions and patterns which are outlined by the radar. Radar patterns provide information about the structures and motions in clear air and will be described in detail in 14.4-14.6.

Investigators have recognized two sets of atmospheric conditions which favor radar scattering from variations in refractive index. One is associated with the variations in refractive index at the boundaries of convective cells. In this case radar provides information on the three-dimensional convective structure associated with rising moist air. In this convective domain, water vapor is the dominant contributor to refractive index variations and positive buoyancy-forces contribute significantly to the generation of small scale turbulence. The other set of conditions occurs above the convective mixing zone where layers of enhanced static stability often form and negative buoyancy-forces tend to suppress turbulent motions. In

these stable regions, turbulence is generated by mechanical energy which is extracted from the mean flow either by local breakdown of wind shear or by overturning of unstable waves. High power radars detect the hydrostatically stable layers once the vertical wind shear is sufficiently large to overcome the stabilizing negative buoyancy-force. Breakdown may occur locally over very small depths (a few meters) or, in cases of rapid shear development, breakdown may take place throughout layers having a depth of more than 1 km. These deep layers are often associated with clear air turbulence (CAT) which aircraft experience. The radar characteristics of these shear zones and their association with CAT form the major portion of the experimental results described in 14.5 and 14.6.

14.1 Theory

14.1.1 Main Features of Scattering from a Turbulent Medium

It is now firmly established that sensitive radars with wavelengths of 10 cm or longer often detect echoes which are caused by the backscattering from inhomogeneities of refractive index in the clear atmosphere. The general theory of the scattering of electromagnetic waves by refractive index fluctuations has been treated by Tatarski (1961). Smith and Rogers (1963) and independently Ottersten (1964) derived an explicit expression relating the radar reflectivity to the intensity of the refractive index variations. However, Ottersten (1968a, 1969a, b, and c) has carried through with a clear exposition of the numerous interrelationships between radar measurements and direct observations, and the theoretical description which follows is largely taken from his excellent papers.

Before the theory is developed, it might help to review qualitatively the process whereby turbulence acts on the mean gradients to produce the variations in refractive index which are responsible for the radar backscattering. When air parcels are displaced in the atmosphere by turbulent mixing, inhomogeneities are created because the characteristics of the displaced air parcels will differ from those of the environment. The pressure of the displaced parcels undergoes a continuous equalization with the environmental pressure. This process will change the temperature and water vapor pressure of the air parcels. Their potential temperature and specific humidity, however, will be preserved. The displaced air parcels will, to the first approximation, keep their identity. Consequently, the resulting inhomogeneities in refractive index are best characterized by the differences in potential temperature and specific humidity or of potential refractive index between the regions exchanging air parcels. The sharper the original mean gradient and the more violent the turbulent mixing the stronger will be the inhomogeneities which are created, and consequently the radar scattering will be increased. Similarly, for a given mean gradient of potential refractive index, the stronger the turbulence the stronger will be the inhomogeneities. As we shall see more clearly in the following paragraphs, for a given wavelength, the power backscattered originates from inhomogeneities of the potential refractive index at essentially one scale.

The basic expression which relates the radar reflectivity, η , or radar cross section per unit volume, to the random refractive index fluctuations is

$$\eta(\vec{k}) = \frac{\pi^2}{2} k^4 \Phi_n(\vec{k}) \quad (14:1)$$

where the direction of the vector wave number k is the radar's radial direction and it has magnitude $k = |\vec{k}| = 4\pi/\lambda$, λ is the radar wavelength, and $\Phi_n(\vec{k})$ is the spatial power-spectral density which is a three-dimensional representation of the refractive index field. $\Phi_n(\vec{k})$ is the Fourier transform of the three-dimensional refractive index covariance function and is related to the variance of the refractive index $\overline{n'^2}$ by

$$\int_{-\infty}^{\infty} \Phi_n(\vec{k}) dk = \overline{n'^2} \quad (14:2)$$

where the integration is carried out over the entire wave number space. Although $\Phi_n(\vec{k})$ is defined for all wave number space, only one particular value of the wave number contributes to the radar backscattering. The radar essentially samples the refractive index spectrum at the particular wave number vector which is

directed along the radar radial direction and has a magnitude of $4\pi/\lambda$. This wave number corresponds to the spatial scale of $L = 2\pi/k = \lambda/2$. The eddy sizes near $\lambda/2$ contribute to the backscatter because it is only these sizes that produce additive phases and that consequently produce a signal which is detectable at the receiver. The only requirement is that some spectral energy is present at a scale of $\lambda/2$. The process is identical to constructive interference of the waves diffracted by the appropriate spacing of the grating for Bragg scattering. This selective isolation of only one eddy size by the radar is equivalent to the application of a narrow-band filter to the spectrum of refractive index irregularities.

$\Phi_n(\vec{k})$ is a three-dimensional space spectrum, but often a one-dimensional spectrum appears in expressions for the radar reflectivity because this is the spectrum which can be obtained from measurements with a single refractometer. Ottersten (1969a) uses $F_n^X(k_X)$ to denote the one-dimensional spectrum which describes the refractive index field along the x-axis in an x, y, z coordinate system; it is obtained through the integration

$$F_n^X(k_X) = \iint_{-\infty}^{\infty} \Phi_n(\vec{k}) dk_y dk_z \quad (14:3)$$

The normalization convention is

$$\overline{n'^2} = \int_{-\infty}^{\infty} F_n^X(k_X) dk_X = \int_0^{\infty} S_n^X(k_X) dk_X \quad (14:4)$$

where $S_n^X(k_X) = 2F_n^X(k_X)$ is the one-dimensional spectrum that would be measured by a discrete refractive index sensor carried through the air along the x-axis. $F_n^\alpha(k_\alpha)$ describes the refractive index field in the arbitrary direction α , and in general will vary with the direction of α . For an isotropic field, however, $\Phi_n(\vec{k}) = \Phi_n(k)$, and $F_n^\alpha(k_\alpha)$ is independent of the direction of α ; for this situation (14:3) reduces to

$$F_n^\alpha(k_\alpha) = \int_{|k_\alpha|}^{\infty} 2\pi k \Phi_n(k) dk \quad (14:5)$$

or to

$$\frac{dF_n^\alpha(k_\alpha)}{dk_\alpha} = -2\pi k_\alpha \Phi_n(|k_\alpha|) \quad (14:6)$$

when expressed in the differential form. Since $S_n^\alpha(k) = 2F_n^\alpha(k)$ is the measurable one-dimensional refractive index spectrum, substitution of (14:6) in (14:1) gives

$$\eta(k) = -\frac{\pi}{8} k^3 \frac{dS_n^\alpha(k)}{dk} \quad (14:7)$$

Equation (14:7) is the expression to be used for the comparison of the radar reflectivity measurement, $\eta(k)$, to measurements of the one-dimensional refractive index spectra, provided that the refractive index field is isotropic for $k = 4\pi/\lambda$. With an isotropic field it is no longer necessary to carry the superscript α because all directions will have the same spectral properties.

The one-dimensional spectrum of velocity in a turbulent medium, $S(k)$, has a well-known shape in the inertial sub-range; this range is postulated as a result of the Kolmogorov similarity theory of locally isotropic turbulence in an incompressible fluid when the Reynolds number is large. The average properties of the turbulent motion in the inertial sub-range are determined uniquely by ϵ , the average rate of dissipation of turbulent kinetic energy per unit mass. For the field of velocity fluctuations along any single direction, the one-dimensional spectrum $S(k)$ in the inertial sub-range is given by

$$S(k) = A_1 \epsilon^{2/3} k^{-5/3} \quad (14:8)$$

where k is the radian wave number corresponding to the scale $2\pi/k$. The value of A_1 is about 0.47 for longitudinal velocity spectra and is $(4/3)0.47$ for the transverse spectra. Implicit in (14:8) is the fact that the energy dissipation rate per unit mass, ϵ may serve as a measure of turbulence intensity in the inertial sub-range; it has units of $\text{cm}^2\text{sec}^{-3}$.

Tatarski (1961) summarizes the consequences of the similarity theory for the properties of the spectrum of passive additives in the turbulent flow. If the potential refractive index is considered as a conservative passive additive, its one-dimensional spectrum of the refractive index variability in the inertial subrange is

$$S_n(k) = A_{1n} \epsilon_n \bar{\epsilon}^{-1/3} k^{-5/3} \quad (14:9)$$

where the variance of the refractive index is given by $\bar{n}^2 = \int_0^\infty S_n(k) dk$. In (14:9), the quantity ϵ_n is the average rate of dissipation of the inhomogeneities in the refractive index field by molecular diffusion; it is analogous to the energy dissipation rate ϵ but has units of sec^{-1} because n is non-dimensional. Gurvich et al. (1967) gives a value of the numerical constant A_{1n} of about 0.7. By taking the derivative of (14:9) and substituting into (14:7) the expression

$$\eta(k) = \left(\frac{5}{3}\right) \frac{\pi}{8} k^2 S_n(k) \quad (14:10)$$

is obtained. Equation 14:10 is a special case of (14:7); it applies when the atmospheric layer of interest is known to be isotropic at the $\lambda/2$ scale and, in addition, the $\lambda/2$ scale is known to fall within the inertial subrange. When these conditions are valid, (14:10) can be used to compare radar reflectivity and refractometer measurements.

Often the structure constant, C_n^2 , for the field of refractive index is used as a measure of the variability of the refractive index within the inertial sub-range. It is given by

$$C_n^2 = a^2 \epsilon_n \bar{\epsilon}^{-1/3} \quad (14:11)$$

where a^2 is a universal non-dimensional constant and is approximately equal to $4A_{1n}$. Usually C_n^2 is given in units of $\text{cm}^{-2/3}$. Using (14:11) in (14:9),

$$S_n(k) \approx \frac{1}{4} C_n^2 k^{-5/3} \quad (14:12)$$

and by substituting (14:12) in (14:10) for $k = 4\pi/\lambda$, the equation

$$\eta \approx 0.38 C_n^2 \lambda^{-1/3} \quad (14:13)$$

is obtained. Provided conditions are isotropic and the wave number $4\pi/\lambda$ falls within the inertial sub-range, radar measurements of η provide direct estimates of C_n^2 . In the cgs system, η has units of cm^{-1} .

In contrast to the wavelength dependence expected for the scattering from refractive index inhomogeneities, the scattering by particles which are small relative to the radar wavelength has a strong dependence. The reflectivity from these small particles is given by

$$\eta = \pi^5 |K|^2 \lambda^{-4} \frac{\sum D^6}{\text{Unit Vol.}} \quad (14:14)$$

where $K = m^2 - 1)/(m^2 + 2)$ and m is the complex index of refraction of the particle, D is the particle diameter, and the sum is taken over a unit volume. Equation (14:14) is the Rayleigh approximation; it can be applied with very small errors for cloud droplets and raindrops when radar wavelengths greater than 3 cm are used. Comparing (14:13) and (14:14), one sees that particle scatter can be distinguished from scatter due to variations in refractive index by measuring the reflectivity at more than one radar wavelength.

14.1.2 Relationship of Radar Reflectivity to Turbulence Intensity

Qualitative reasoning in the first few paragraphs of 14.1.1 led to the suggestion that under certain conditions turbulence intensity should be related to radar reflectivity. This relationship will now be developed in more quantitative terms, and again the presentation depends largely on papers published by Ottersten (1968, 1969a).

Much of the significant clear air turbulence (CAT) occurs at high tropospheric levels where the contribution of moisture to the refractive index is generally small. At temperatures below about -45°C Ottersten (unpub. manuscript) has estimated that neglecting the moisture influence on the strength of the radar reflectivity in regions of turbulence will introduce an error of less than 3 dB. It is also known that CAT is generally associated with a statically stable layer. Consequently, in what follows, it will be assumed that the atmosphere is statically stable (buoyancy force is negative) and that the moisture contribution to the refractive index field is small. With these assumptions, Ottersten (unpub. manuscript) has shown that C_n^2 can be expressed in terms of C_T^2 , the structure constant for temperature fluctuations, and the relation is:

$$C_n^2 = [77.6 \times 10^{-6} P_0 T^{-2}]^2 C_T^2 \quad (14:15)$$

where P_0 is the mean pressure of the layer in millibars and T is the temperature in degrees Kelvin. In the cgs system C_T^2 has units of $(^\circ\text{K})^2 \text{cm}^{-2/3}$ and the term within the bracket has units of $(^\circ\text{K})^{-1}$.

In analogy with (14:11), C_T^2 can now be expressed by:

$$C_T^2 = a^2 \epsilon^{2/3} \epsilon_T \bar{\epsilon}^{-1} \quad (14:16)$$

where ϵ_T is the average rate of dissipation of the inhomogeneities in the temperature (or refractive index) field by molecular diffusion. It is now assumed that the ratio $\epsilon_T \bar{\epsilon}^{-1}$ is determined by the ratio between the production of mean square temperature fluctuations and the net production of turbulent kinetic energy:

$$\frac{\epsilon_T}{\epsilon} = \frac{\overline{\theta'w'} \, d\theta/dz}{\overline{u'w'} \, du/dz - \overline{\theta'w'} \, g/\theta} \quad (14:17)$$

where u and w are the horizontal and vertical components of motion, z is the vertical coordinate, and θ is the mean value of the generalized potential temperature (see Ottersten, 1969c) for the level under consideration. The primed quantities indicate deviations from the mean, and the bars indicate averages in the horizontal. Ottersten (unpub. manuscript) explains that expressing $\epsilon_T \bar{\epsilon}^{-1}$ by (14:17) does not neglect divergences of the fluxes of temperature variability and velocity perturbations but only requires that the two divergence terms depend to the same extent on their respective net production terms. Similarly, (14:17) allows a certain degree of non-stationarity because it only requires that the rate of change in the mean square temperature fluctuations and the rate of change in turbulent kinetic energy depend in analogous ways on their respective net production terms. These assumptions appear reasonable where there is a large transfer of kinetic energy from wind shear into turbulent motion as would be expected in regions of CAT.

The flux Richardson number, R_f , is defined as the ratio of the kinetic energy removed by buoyancy forces to the kinetic energy introduced by the wind shear and is given by:

$$R_f = \frac{g}{\theta} \frac{\overline{\theta'w'}}{\overline{u'w'} du/dz} \quad (14:18)$$

where g is the acceleration of gravity. The flux Richardson number is positive in statically stable conditions and less than unity where turbulence is produced. Substitution of (14:18) in (14:17) gives:

$$\frac{e_T}{\epsilon} = \frac{\theta}{g} \frac{d\theta}{dz} \frac{R_f}{1-R_f} \quad (14:19)$$

By combining (14:19) and (14:16), C_T^2 may be expressed by

$$C_T^2 = a^2 \epsilon^{2/3} \frac{\theta}{g} \frac{d\theta}{dz} \frac{R_f}{1-R_f} \quad (14:20)$$

Also, in regions where moisture gradients are negligible, (14:15) is applicable and with the use of (14:20), the radar reflectivity in (14:13) can be expressed by

$$\eta = 0.38\lambda^{-1/3} (77.6 \times 10^{-6} P_O T^2)^2 a^2 \epsilon^{2/3} \frac{\theta}{g} \frac{d\theta}{dz} \frac{R_f}{1-R_f} \quad (14:21)$$

Provided the moisture gradients are small, this later equation is valid under some rather general assumptions for regions where there is a large kinetic energy transfer from shear into turbulence. It can be seen that the radar reflectivity is directly related to the mean vertical gradient of generalized potential temperature ($d\theta/dz$) and to $\epsilon^{2/3}$ which is a measure of the intensity of the velocity fluctuations in the inertial sub-range.

The relationship (14:21) is complicated, however, by the factor $R_f/1-R_f$. As a first approximation, R_f can be assumed to be 1/4, the critical value of the gradient Richardson number (Ri) for the initiation of dynamic shear instability as indicated by theoretical studies (Miles and Howard, 1964) and experiments (Browning, 1971). Thus, the factor $R_f/1-R_f$ is about 1/3. This value may vary with the type and size of the generating instability and also with the stage of the turbulence development.

Direct measurements with an instrumented aircraft in conjunction with radar observations are required for an improved understanding of the interrelationships indicated by (14:21). Since strong CAT appears to be associated with high values of $d\theta/dz$ (Browning, 1971), it is apparent that sensitive radars should be able to detect regions of significant turbulence within the upper troposphere and stratosphere.

Although the estimate will be crude, it might be instructive to substitute some representative values in (14:21) in an attempt to arrive at the radar reflectivity which might be expected in regions of CAT near the tropopause. The value of a^2 is about 2.8. Near the tropopause, typical values might be $P_O = 200$ mb, $T = 220^\circ K$, and $\theta = 350^\circ K$. Substitution of these values in (14:21) and using $R_f/1-R_f = 1/3$ gives

$$\eta = 1.3 \times 10^{-14} \lambda^{-1/3} \epsilon^{2/3} \frac{d\theta}{dz} \quad (14:22)$$

In regions of moderate CAT, Atlas et al. (1966a) estimate ϵ to be about $100 \text{ cm}^2 \text{ sec}^{-3}$, and Browning (1971) observed $d\theta/dz$ to be generally greater than $5 \times 10^{-5} \text{ }^\circ\text{C cm}^{-1}$. With these values and for a radar wavelength of 10 cm, the reflectivity of moderate CAT is about $3 \times 10^{17} \text{ cm}^{-1}$. From (T14.1), it is seen that the four sensitive radars with a wavelength of 10 cm or longer would be able to detect moderate CAT out to ranges of at least 10 km. This is supported by the experimental results on the radar detection of CAT which are summarized in 14.6.

14.2 Experimental Verification of the Theory

14.2.1 An Experiment to Test Theory

As suggested earlier (14:10) is the equation which relates the radar reflectivity to the spectral energy measured with a refractometer, provided that the scale length of $\lambda/2$ is known to fall within the inertial sub-range. Experiments directed toward a quantitative check of (14:10) were conducted at Wallops Island, Virginia. A sensitive 10.7-cm wavelength radar was used to measure the radar reflectivity and a fast response refractometer mounted on an aircraft obtained the required measurements of refractive index (Kropfli et al., 1968).

The technique which was used during the experiment was devised by Konrad (see Konrad and Randall, 1966). It involved the tracking of a meteorologically-instrumented aircraft with the radar, making radar measurements and meteorological measurements simultaneously. One or several radar "data" gates were located just ahead of the aircraft tracking gate; the radar echo strength was measured for the contributing region defined by the position of the data gate and the antenna pattern. The airplane flew into this "contributing region" about 10 seconds after the radar had a measure of its reflectivity. By taking this lag into account, one could compare the radar reflectivity with the corresponding refractive index characteristics.

Since the radar used in the experiments had a wavelength of 10.7-cm, the spectral power of refractive index at a scale of 5.35 cm was required. The refractometer did not respond to this small scale, and an extrapolation procedure was introduced (Lane, 1967b). For this purpose, a curve with a $-5/3$ slope (inertial sub-range) was drawn through the computed spectral powers. Extrapolation occurred from a scale of about 1 m down to 5.35 cm. The spectral power at the 5.35-cm scale was denoted by Kropfli et al., (1968) as $\langle(\Delta n)^2\rangle F_n(k_r)$ where $\langle(\Delta n)^2\rangle$ is the variance and $F_n(k_r)$ is the normalized one-dimensional wave number spectrum of refractive index along the direction of the aircraft flight path. In the notation used by Ottersten (1969a) and followed in 14.1, $S_n(k)$ of (14:10) is identical to $\langle(\Delta n)^2\rangle F_n(k_r)$, and these values, estimated from various refractive index spectra, are plotted along the abscissa of (F14.1). The left ordinate in (F14.1) is the 10.7-cm radar reflectivity which was estimated to be most representative of the space where the refractometer measurements were obtained; the right ordinate is the corresponding C_n^2 values which were computed using (14:13). The solid line marked $(5/3)\pi/8k_r^2$ was drawn for comparison to the theoretical relationship between the radar reflectivity η and the spectral density of refractive index $\langle(\Delta n)^2\rangle F_n(k_r)$ or $S_n(k)$ as given by the refractometer measurements. Most of the points fall within the ± 3 dB bounds as illustrated in the figure. This agreement was taken by Kropfli et al. (1968) as evidence that the scattering processes and assumptions which led to the development of (14:10) were essentially correct.

The measurements described by Kropfli et al. (1968) were obtained in the lower 3 km of the atmosphere in June and the moisture contribution to the refractive index variability was predominate. This accounts for the large radar reflectivity values (10^{16} - 10^{14} cm^{-1}) which were observed. From (T14.1) it can be seen that these reflectivities would produce signals which would be 10-30 dB above the minimum detectable for the two sensitive 10-cm radars. The radar reflectivities of (F14.1) were determined on the assumption that the contributing region of the radar was filled with a uniform scattering field. If the size of the scattering field were less than the contributing region, then the radar estimates of this field would be too low. Because Kropfli et al. (1968) obtained their measurements in portions of the troposphere which were generally well mixed, it is likely that the radar contributing region was usually filled and their reflectivity values would be representative.

Atlas et al. (1970) have described their radar measurements within extremely narrow layers (about 2 m thickness), and they found reflectivity values at a wavelength of 10 cm which were greater than 10^{-12} cm^{-1} (see also 20). Consequently, in the lower troposphere where moisture is the major contributor to the refractive index variability, very large reflectivities can occur over narrow layers. As demonstrated in 14.1.2, however, the reflectivity from CAT near the tropopause could be as much as 40 dB less than the reflectivity values which occur in the lower atmosphere.

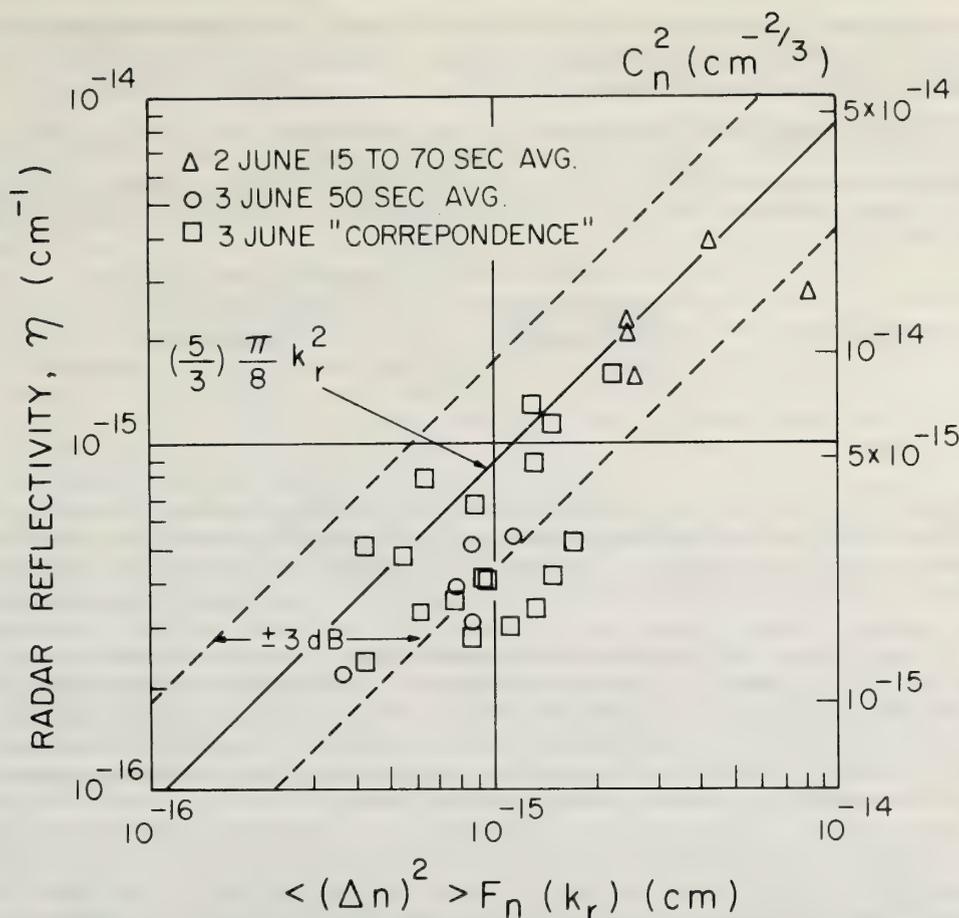


Figure 14.1 Observed radar reflectivity as a function of spectral power of refractive index at a scale of 5.35 cm. Spectral power was obtained by extrapolating computed spectra (based on airborne microwave refractometer measurements) down to wave number k_r corresponding to one-half the 10.7-cm radar wavelength. For almost all cases the radar reflectivity corresponds to within 3 dB of that expected from direct refractometer measurements (from Kropfli et al., 1968).

14.2.2 Identification of Scattering Sources by Multiwavelength Radar Measurements

It was suggested earlier that multiwavelength radar measurements could be used to identify the probably source of the scattering by observing the wavelength dependence of the reflectivity. The wavelength dependence for scattering from refractive index variation is $\lambda^{-1/3}$ as given by (14:13) and for Rayleigh scatterers is λ^{-4} as given by (14:14). The multiwavelength radar facility at Wallops Island has the capability to carry out such an identification. Some of the characteristics of these radars are given in (T14.1), and the way in which it is possible to distinguish the scattering from particles which are small relative to the wavelength (Rayleigh scatterers) from that expected from refractive index variations is illustrated in (T14.2). If the pulse volumes are filled with Rayleigh scatterers, then the signal above the minimum detectable (MDS) at 3.2-cm wavelength will be about 5 dB stronger than the signal above the MDS of the 10.7-cm radar and 30 dB stronger than that of the 71.5-cm radar. On the other hand, if the reflectivity of the scattering medium has a wavelength dependence of approximately $\lambda^{-1/3}$, then the 3.2-cm signal above its MDS will be about 15 dB lower than the signal above the 10.7-cm MDS and 20 dB lower than the signal above the 71.5-cm MDS. Thus, if the signal above the MDS is stronger at the longer wavelength, it is certain that the mechanism giving rise to the radar echoes is not scattering by Rayleigh particles. Since the ratios in (T14.2) depend on the individual radar parameters, the tabulated values can only be applicable for the Wallops Island radars. The ratios will also be inaccurate if the scattering from the

clear air approaches that from particles for one or more of the radars. This situation is quite rare, however, and as will be shown in the data which follow, qualitative evaluations of the scattering source can be obtained with the help of information contained in (T14.2).

Table 14.2. Ratios of Signal Strengths Above Minimum Detectable Expected from Two Types of Scatterers (Applicable Only for the Wallops Island Radars)

Ratio of 3.2-cm received power above minimum detectable to that at 10.7 and 71.5 cm (dB)

Wavelength(cm)	Rayleigh scatterers (λ^{-4} dependence)	Scattering from refractive index fluctuations ($\lambda^{-1/3}$ dependence)
10.7	5	-15
71.5	30	-20

Figure (14.2) shows simultaneous photographs of the range height indicators (RHI) of the three Wallops Island radars, taken while the beams were scanning synchronously in elevation angle. Except for an overcast cirrus layer, the sky was clear at the time of this observation. The cirrus cloud appears between a height of 7 and 10 km; it is easily identified as it is strongest at 3.2-cm wavelength (left photo), weaker at 10.7-cm wavelength (middle photo), and not visible at 71.5-cm wavelength (right photo). Such a drop-off in signal strength with wavelength is expected for Rayleigh scatterers as indicated in (T14.2). The fact that the numerous dot echoes, appearing between 1 and 3 km in height, are seen most prominently with the 3.2-cm radar and not at all with the 71.5-cm radar indicates that these targets are small relative to the radar wavelength. This observation on the dot echoes, as well as many other types of measurement (for example, the tracking by radar of single known insect species), has led to the conclusion that most of the dot targets are insects (Glover and Hardy, 1966; Glover et al., 1966).

In (F14.2), the thin layers seen near the surface and at 1 km with the two longer wavelength radars are not attributed to particle scatter. If this were so, the layers would appear much stronger at the shorter wavelength (T14.2). Instead, the layers are caused by backscatter from variations in refractive index (Hardy et al., 1966). These authors show that the wavelength dependence of the echo layers is consistent with the theory of scatter by refractive-index variations (14:13). In addition, Kropfli et al. (1968) have confirmed that these clear-air radar layers are associated with increased refractivity fluctuations as measured directly with refractometers mounted on aircraft or suspended below a helicopter.

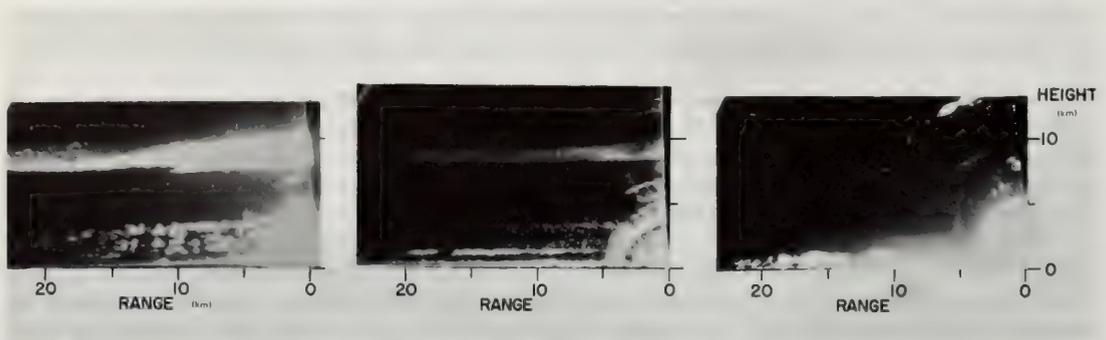


Figure 14.2 Photographs of range-height indicators at 3.2-, 10.7-, and 71.5-cm wavelengths (left to right) along an azimuth of 260 degrees for 1740 EST, September 3, 1966, at Wallops Island, Virginia. A cirrus cloud layer between heights of 8 and 10 km appears at the shorter wavelengths, whereas the longer wavelength detects only the clear-air variations in refractive index. The numerous dot echoes which appear uniformly distributed between 1 and 3 km at the two shorter wavelengths are due to single insects (from Hardy and Katz, 1969).

14.3 Clear Air Dot Echoes

14.3.1 The Nature of Dot Echoes

Dot or point echoes, as their name implies, appear as point targets on a PPI or RHI scope photograph (F14.2). They are also very commonly observed with a fixed vertically-pointing radar beam. Plank (1956) and Ottersten (1970) have given excellent descriptions of the major characteristics of dot echoes including the diurnal and seasonal variation of dot echo activity. An individual dot echo usually has a uniform cross section, but a very wide range of backscattering cross sections are possible. With vertically pointing antennas the scatterers generally appear for a fraction of a second up to a few seconds. Their appearance suggests targets of limited extension which drift through the radar beam at a speed which is close to the wind speed. Examples of dot echoes observed by Ottersten (1970) with a vertically pointing 10.0-cm wavelength radar are shown in (F14.3). This time-height record was obtained by intensity modulating a scope which was photographed while the film moved continuously at a slow speed. Sometimes on hot, clear summer days the dot echoes appear very close, almost forming thick layers as seen in the upper portion of (F14.3). Often there is a tendency for the dot echoes to concentrate at some preferred heights as was noted by Plank (1954) and Hardy et al. (1966).

Prior to 1962, there was little known about the wavelength dependence of the radar backscattering cross section of the dot echoes. Certainly, dot echoes were seen consistently with radars having wavelengths of less than 3 cm (Crawford, 1949; Plank, 1954; Vrana, 1961). However, at wavelengths of about 10 cm, only the more sensitive radars were able to detect the relatively small cross section of the vast majority of dot echoes.

Let us look at some of the observations of dot echo cross sections as reported by various investigators. These are indicated in (F14.4). The data by Borchardt (1962), Roelofs (1963), and Fehlhaber and Grosskopf (1964) show cross sections which range from about 5×10^{-4} to 10 cm^2 . The cross sections reported by Vrana (1961) and Plank (1956) vary from 2×10^{-5} to $8 \times 10^{-2} \text{ cm}^2$. Note that the minimum detectable cross section of the TPQ-11 radar (a vertically pointing 0.86-cm wavelength radar) at a range of 1 km is $5 \times 10^{-6} \text{ cm}^2$ which is almost 20 dB less than the cross section of a 1-mm diameter water drop.

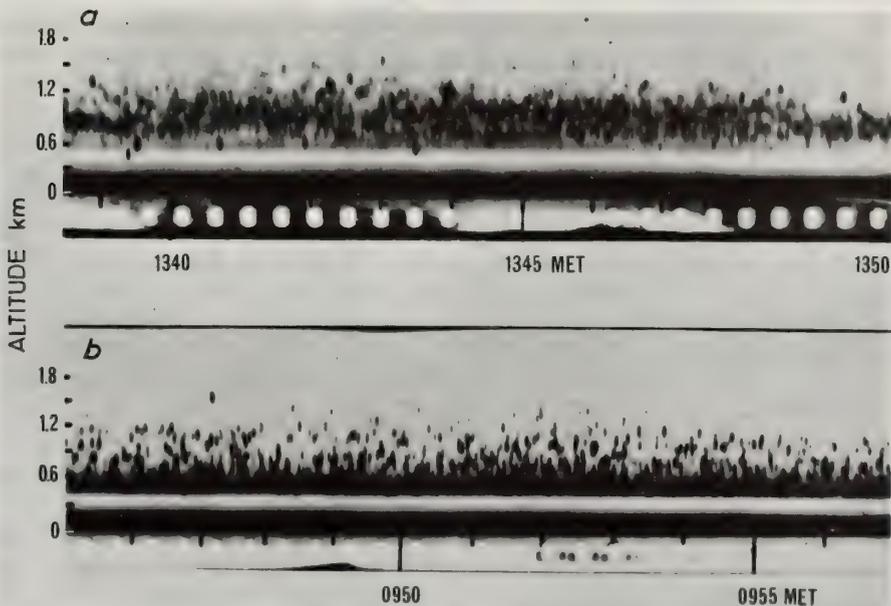


Figure 14.3 Two 12-minute samples of dot echo tropograms (time-height records) obtained with a 10.0-cm wavelength radar near Stockholm, Sweden for (a) 1345 LST, 3 July 1963, and (b) 0950 LST, 1 August 1963. The appearance of the dot echoes suggests small targets drifting with the wind through the radar beam (from Ottersten, 1970).

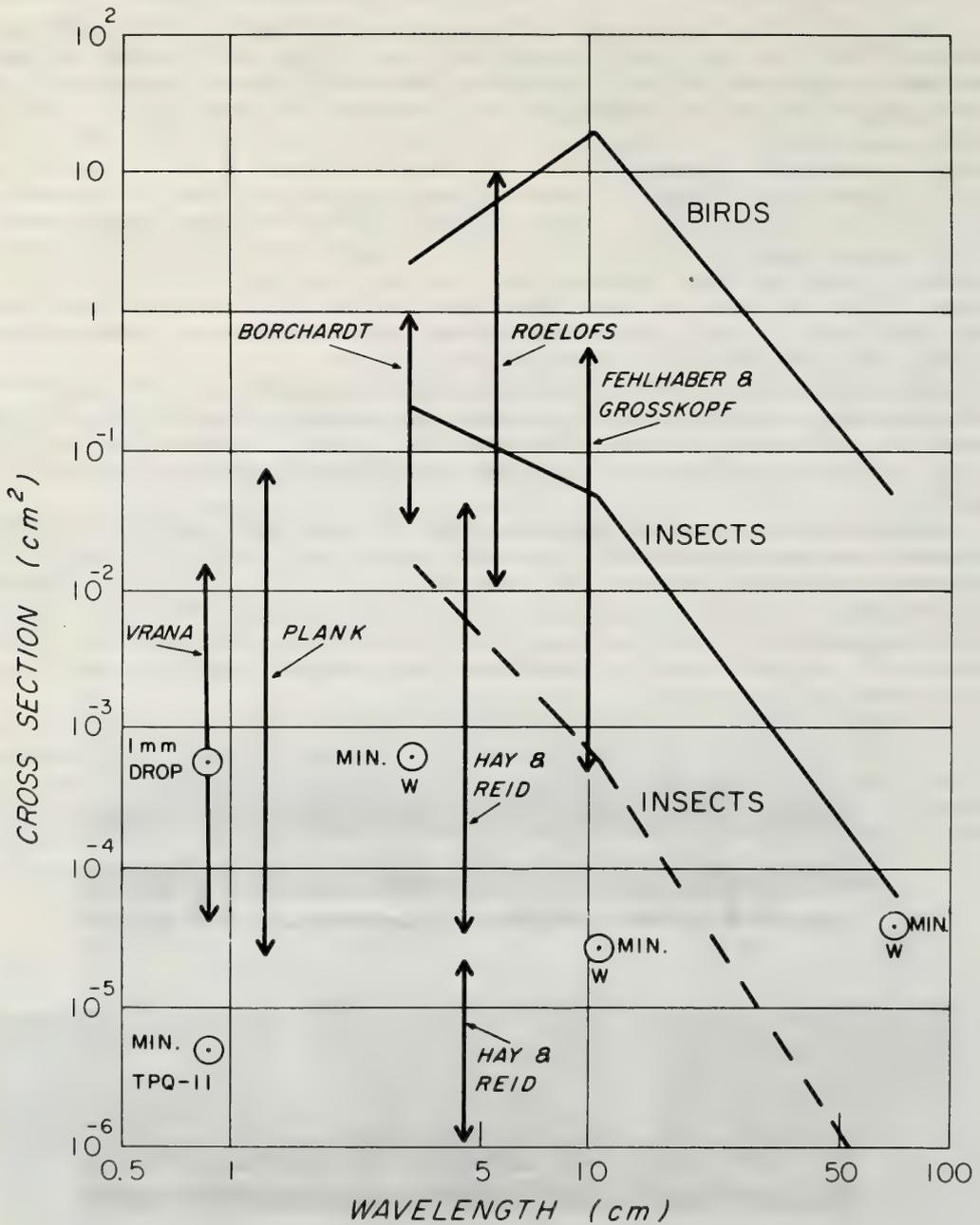


Figure 14.4 Cross sections of dot angels, insects, and birds as a function of wavelength. The range of dot angel cross sections reported by various investigators and the insect and bird cross sections are illustrated schematically. The minimum detectable cross sections for the TPQ-11 radar and the Wallops Island radars are indicated by "min TPQ-11" and "min W," respectively (from Hardy and Katz, 1969).

The results of cross section measurements of insects and birds at various wavelengths are indicated schematically in (F14.4). The curve for birds was chosen from a consideration of observations at three wavelengths on known birds (Konrad and Hicks, 1966; Konrad et al., 1968) and of additional information reported by Glover and Hardy (1966). There is naturally a large variation in bird cross sections. Even for the same bird, the cross section can change by as much as ± 10 dB as the bird's configuration or the viewing aspect changes.

The curves for known insects were similarly obtained from multiwavelength measurements using the Wallops Island radars (Glover et al., 1966). The two curves indicate a probable range of values for fairly large insects (maximum size equivalent to a worker honey bee). The cross sections of insects having dimensions of less than 3 mm are generally undetectable with the Wallops Island radars. The insect cross sections vary by about ± 5 dB at wavelengths greater than 3 cm. There are no extensive cross section measurements of known insects at wavelengths less than 3 cm. Nevertheless, the cross sections of water spheres, which have volumes corresponding to those of a large class of insects, fall within the range of 10^{-4} to 10^{-1} cm² at wavelengths from 1-3 cm.

The data in (F14.4) indicate that, taking into account the method of observation, all the cross section measurements of dot angels fall within the range of values expected for insects or birds. Moreover, based on simultaneous multiwavelength measurements (Glover and Hardy, 1966) and on bistatic and depolarization studies by Chernikov (1966), it was concluded that all of the dot angels observed in detail have characteristics which identify them as either insects or birds.

14.3.2 Dot Echoes for Scientific Studies

Birds have been identified or detected with radar since the mid-1940's (Lack and Varley, 1945), and radar has been used extensively in the study of bird flight, roosting, feeding, and migration patterns. These investigations have recently been described admirably in a text on radar ornithology by Eastwood (1967). Because of their smaller cross sections, insects have not been studied intensively by radar. Nevertheless, radar cross sections of a few insects have been measured (Glover et al., 1966; Hajovsky et al., 1966), and the value of radar to the entomologist was pointed out by Glover et al. (1966). Although this may be a meager beginning, there is little reason to doubt that, with suitably designed radars, radar entomology could become as important a subject for study as radar ornithology has been in the past.

Because insects are relatively weak flyers, they are reasonable tracers of the mean wind. For example, insects were probably the source of the clear air echoes on several occasions when Doppler radar was used to investigate a nocturnal jet (Lhermitte, 1966; Browning and Atlas, 1966), the passage of a cold front (Lhermitte and Dooley, 1966), and Bénard-like convection (Section 14.4.4). Figure (14.5), taken from Browning and Atlas (1966), illustrates the time-height pattern of mean horizontal velocity derived from a Doppler radar when insects were sufficiently plentiful to permit wind observations to be made up to a height of about 1.3 km. The velocity field varies in an orderly fashion and there is evidence of the formation of a strong low level nocturnal jet. Browning and Atlas (1966), however, advise caution in the detailed interpretation of the wind field because they believe that some of the insects may have been flying with a significant speed in a preferred direction. Although they warn about the selectivity which must be exercised in using dot echoes as reliable wind tracers, there appear to be some occasions when insects can be relied upon to give useful information of motions in the clear atmosphere.

14.4 Clear Air Convection

14.4.1 General Radar Patterns of Clear Air Convection

Using appropriate radars, it is possible to study convective processes in the clear atmosphere. Powerful radars at wavelengths of 10 cm or longer consistently detect variations in refractive index associated with free convection in the clear air (Atlas and Hardy, 1966). During conditions of strong surface heating and light winds, sensitive 10-cm radars reveal thermal-like, clear-air structures aloft which are about 1-3 km in diameter and several hundred meters in height (Hardy and Ottersten, 1969; Konrad, 1970; Harrold and Browning, 1971). These clear air convective cells may persist for about 30 minutes and are characterized by updrafts in their centers.

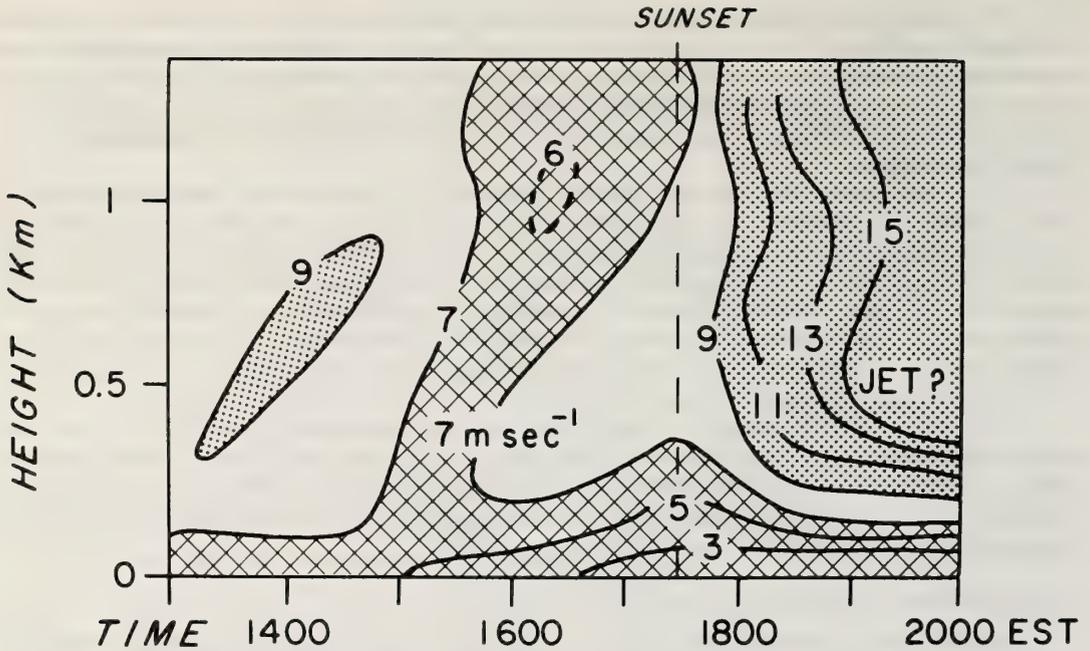


Figure 14.5 Time-height pattern of the mean horizontal velocity of the population of dot echoes obtained with a 5.4-cm wavelength radar at Sudbury, Massachusetts on 27 Sept 1965 (from Browning and Atlas, 1966).

Figure (14.6) is a range height indicator (RHI) scope photograph for the sensitive 10.7-cm wavelength radar at Wallops Island. The photograph was taken while the sky was clear, and it shows the typical convective cells over the land. As described by Hardy et al. (1966) the echoes in (F14.6) arise by virtue of the scattering from irregular refractive-index variations in the clear air. These fluctuations are particularly strong at the cell boundaries and probably most intense at the apex of the cell.

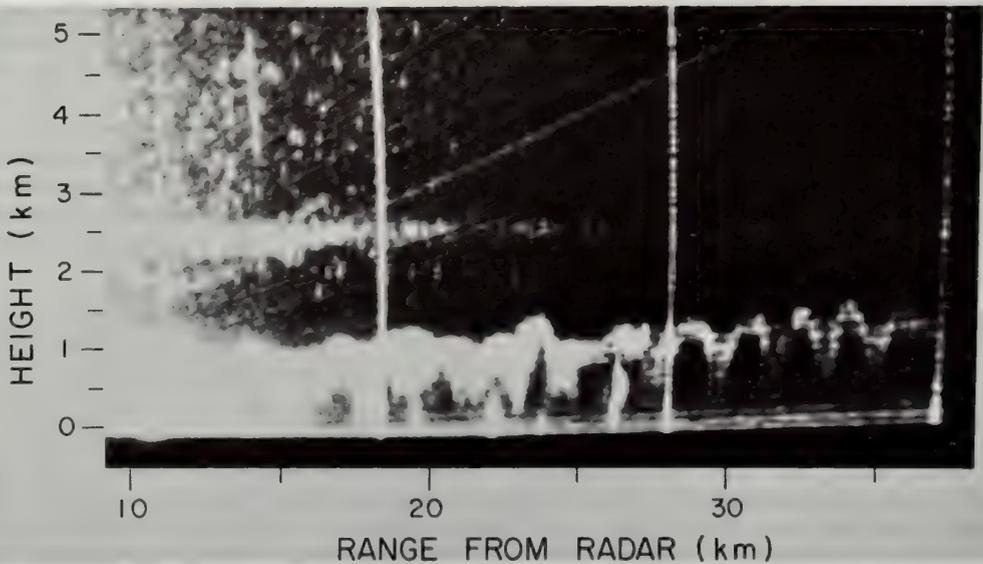


Figure 14.6 Photograph of RHI at 10.7-cm wavelength along an azimuth over land at 1330 EST on 30 July 1968 at Wallops Island, Virginia. The echo structure outlines the boundaries of convective cells while the sky was clear. A clear air layer occurs between 2 and 2.5 km (from Konrad, 1970).

Echoes from these small convective cells in horizontal section are shown by the sector plan position indicator (PPI) photo in (F14.7a). The echo structure is doughnut-shaped; that is, the cell echoes typically appear circular or elliptical and have echo-free centers. At this time the convection had developed to a point where it also could be seen visually in the sky as scattered (1/10-2/10) fair weather cumulus clouds which probably were associated with the tops of some of the mature cells. The droplet sizes in these clouds, however, were too small to be detected, and the 3.2-cm radar, which is more sensitive to backscattering from small particles than the other radars, did not show any cloud echoes. Thus, the echoes beyond 25 km in (F14.7a) are not due to any type of particle scatter. In fact, doughnut echoes are often observed when the sky is completely clear. Figure (14.7b) is a sketch of an atmospheric structure which would lead to the observed echo pattern; i.e., the radar detects only the boundary of the cell where the refractive-index fluctuations are relatively large, and when the radar is scanned in azimuth, a doughnut-shaped echo results. The sketch is meant to illustrate how a pattern with an echo free center could arise. As indicated in the sketch, the flow within the cell is upward in the center, and the relative flow around the periphery is outward and possibly downward.

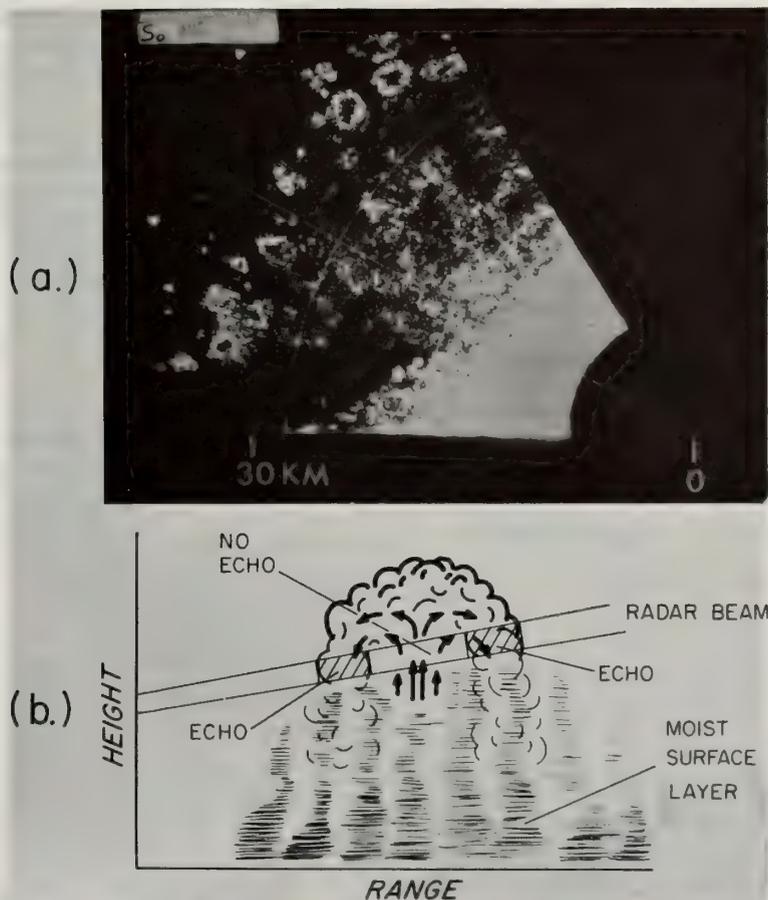


Figure 14.7a Sector PPI photo at 10.7-cm wavelength for 3-deg. elevation angle taken at 1052 EST 15 August 1967, at Wallops Island, Virginia. The strobe line indicates the 300-deg. azimuth. Echoes of cells in horizontal section at the appropriate altitude and range display the characteristic doughnut-shape (from Hardy and Ottersten, 1969).

Figure 14.7b Sketch in vertical section of the cell structure. The radar outlines the boundary of the cell, where the refractive index fluctuations are largest; when the radar is scanned in azimuth, a doughnut-shaped echo results. The air flow within the cell, indicated by the arrows, has been deduced from detailed studies of the cell evolution by examining three-dimensional radar patterns of individual cells (from Hardy and Ottersten, 1969).

Using the radars at Wallops Island, Konrad (1968) studied clear air convective patterns and reported that these convective cells or thermals are seen virtually throughout the year over both land and sea. At times the cells are aligned in parallel rows, and at other times they are randomly distributed as in (F14.7a). In the spring and summer months the cells typically appear over land. During the late fall and early winter months, on those occasions when the ocean water temperature is much higher than the air temperature, these cells are common over the ocean. The aligned clear air convective cells may be termed "thermal streets" in analogy with the type of organized convection, termed "cloud streets," that sometimes is observed. Konrad and Kropfli (1968), confining their observations to clear air convection over the sea, report that the entire radar echo pattern moves along with the mean wind within the convective layer. They further report that the convective cells may be rather uniformly distributed at altitudes less than 300 m but show a definite organization at higher altitudes. They note that random distribution of the cells is expected for light surface winds (small shear) while the convective cells may get aligned in bands parallel to the wind if the surface wind is stronger (presumably larger shear).

14.4.2 Detailed Features of Individual Convective Cells

Using stepped elevation sector PPI's Konrad (1970) carried out a detailed time-space investigation of convective cells. Consistent with reports by Hardy and Ottersten (1969), he found that the cells growing in the lower levels rise rapidly. The tops may be dome-shaped or they may consist of two smaller cells. The upward velocity of the cell tops often increase with time from about 0.5 m sec^{-1} to 1.5 m sec^{-1} . The cell accelerates until it rises above its equi-density or neutral level at which time it slows down and cools off by adiabatic expansion. It becomes denser than its environment and sinks back unless condensation occurs and the latent heat released provides additional lift. Thus, in the mature or decaying stage, the tops of the clear air cells overshoot their equilibrium level and the cell boundaries then start to dissipate. Konrad (1970) has concluded that the strong patterned echo cells from near the top of the convective field are in fact cold and moist relative to a slightly stable environment, and therefore the actual buoyancy force would be downward in many of the observed cells.

14.4.3 Features of the Convective Field

The large scale convective field has quite different properties than the individual convective cells. Figure (14.8), taken from Konrad (1970), illustrates how the convective field or the depth of convection increases during the heating of the ground surface in the morning. In nine of the ten cases studied, the height of the convective domain rose almost linearly at a rate of about 0.08 m sec^{-1} which is considerably smaller than the updrafts in the individual cells.

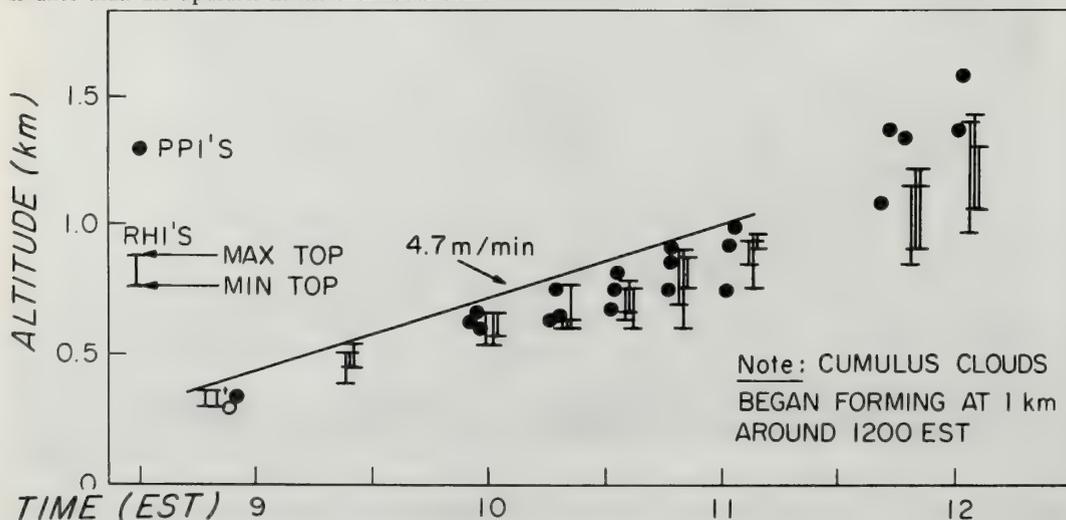


Figure 14.8 The height of the convective field on 26 August 1969 measured from time sequence RHI and PPI photographs with the 10.7-cm radar at Wallops Island, Virginia. The line indicates a linear growth rate of about 0.08 m sec^{-1} (from Konrad, 1970).

Using the powerful 10.7-cm wavelength radar in Defford, England Harrold and Browning (1971) found that the upper limit to the convection varies in height by several hundred meters over horizontal distances of tens of kilometers. This is illustrated in (F14.9). They also identify discrete areas of deeper convection in the clear atmosphere, and, more importantly, they show that when the thermal structure of the atmosphere is suitable for showers or thunderstorms to develop, the showers form only within those discrete areas of deeper convection which have persisted for some hours. The relationship between these persistent areas of deeper convection and the occurrence of future convective growth may be of great value for improving the accuracy of short-range shower forecasts. For the first time, it is also possible to study the convective process throughout all stages of storm development.

14.4.4 Meso-scale Organization of Convective Elements

As mentioned earlier, the small convective cells may be randomly oriented or they may be aligned in "streets." On occasion, another type of organization may be present. This is illustrated in (F14.10) which is a photograph of a PPI scope of a 3.2-cm wavelength CPS-9 radar taken at 0 deg elevation. Except for a few fair weather cumulus clouds (<1/10 coverage) the sky was clear at the time of the observation. Clear-air echoes extend out to 50 n miles, and a Bénard-like cellular pattern can be seen especially in the southwest quadrant. The diameters of the cells vary from 5-10 km. There is also a tendency for the stronger echoes to be arranged in lines oriented from the northwest to southeast. At the time of the observation the winds near the surface were light and from the northwest. Hardy and Ottersten (1969) have investigated this pattern and concluded the following: first, the echoes were due to the scattering from insects which were plentiful at the time of observations; second, each Bénard-like cell was composed of several smaller cells organized around its periphery; and third, the Bénard-like cells were characterized by downward motion in their centers and general updraft motion around their perimeters.

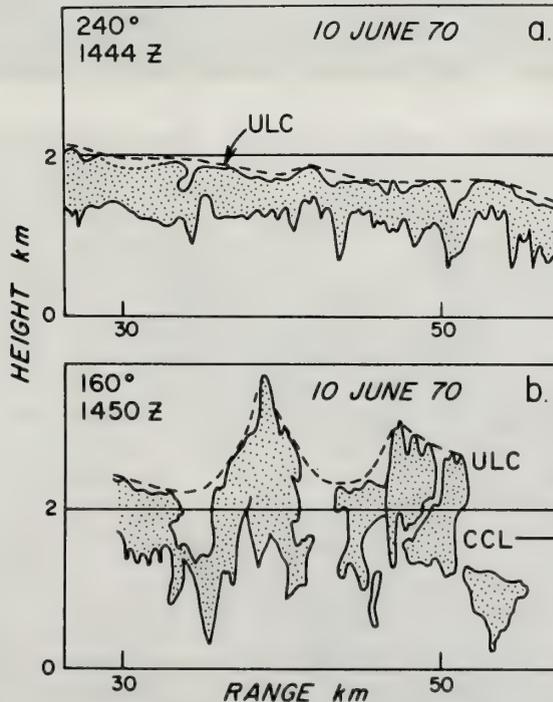


Figure 14.9 Sketches from RHI photographs obtained with the 10.7-cm radar at Defford, England showing the nature of the echo over a limited range on 10 June 1970. The dashed lines (ULC) are the upper limits of convection. The CCL mark is the convective condensation level. The two sketches are for different azimuths at about the same time and they emphasize that the depth of the convective layer can have considerable spatial variation over short distances (from Harrold and Browning, 1971).

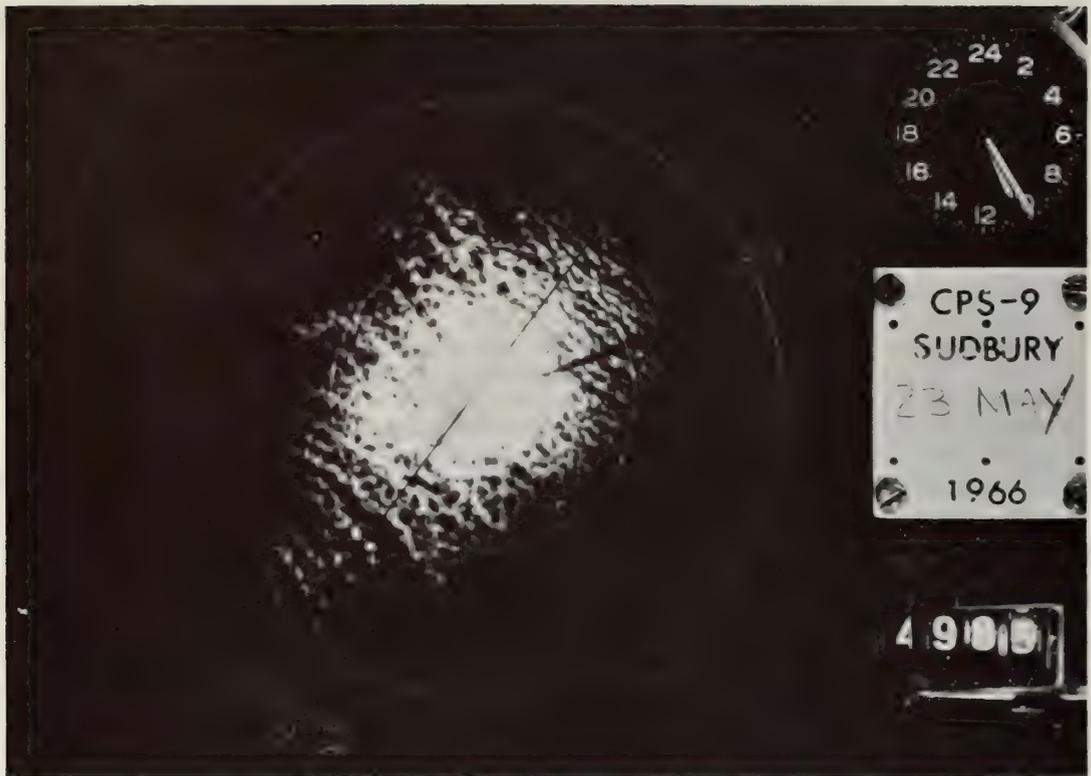


Figure 14.10 PPI photo at 0 degrees elevation angle taken at 1120 EST on May 23, 1966, while the sky was essentially clear. The major range marks are at 25-n mi intervals. The radar is a 3.2-cm CPS-9 located at Sudbury, Massachusetts. A mesoscale cellular pattern, similar to the Bénard circulation patterns studied extensively in the laboratory can be seen. The echoes are due to scattering from insects which occurred at abnormally high concentrations during the observations. The insects apparently provide excellent tracers of the mesoscale atmospheric circulation (from Hardy and Ottersten, 1969).

The flow pattern of the Bénard-like cells is opposite to that of the smaller cells discussed previously. However, evidence for the correctness of the flow direction for the Bénard cells is provided by the numerous studies of similar patterns seen in satellite photographs of clouds. Although the CPS-9 is not a particularly powerful radar, it is able to provide valuable information on convective processes when insects are plentiful. Bénard-like patterns, however, are also observed in clear air on rare occasions with sensitive 10-cm radars by virtue of the scattering from refractive index fluctuations.

14.5 Clear Air Radar Patterns in Shear Zones

14.5.1 Characteristics of Stratified Radar Echoes

Atlas (1964) has reviewed some of the early studies of radar echoes from stratified layers. These layers usually corresponded in height to regions having sharp vertical gradients in refractive index. Additional evidence for the existence of clear air radar layers associated with large variations of refractive index is given in Saxton et al. (1964), Hardy et al. (1966), and Kropfli et al. (1968).

It will be recalled from 14.1 that the radar reflectivity should be dependent on both the mean vertical gradient of potential refractive index and the intensity of the turbulence. In a well mixed layer where the potential refractive index is constant, no degree of turbulence can lead to detectable radar echoes. However, except in the convective mixing zone within the lower 2-3 km of the atmosphere, it is

believed that significant clear air turbulence will always be associated with a vertical gradient of potential refractive index. For a layer with a given mean gradient, increased turbulence will produce increased refractive index inhomogeneities at a scale of $\lambda/2$ (because of the known behavior of turbulent energy cascading to smaller scales), and consequently the radar reflectivity will also increase. Provided that the atmosphere is known to be isotropic and that the scale $\lambda/2$ falls within the inertial subrange, (14:13) predicts that the scattering from refractive index irregularities will be weakly wavelength dependent. Experimental evidence, although limited, also suggests that the clear air radar reflectivity in stratified layers is weakly dependent on the radar wavelength (Hardy et al., 1966; Atlas et al., 1966b). In the viscous dissipation range (i.e., the small scales beyond the inertial subrange) the refractive index spectral power falls off very rapidly. It is believed that this "fall-off" often occurs at a scale of a few centimeters. Consequently, the $\lambda/2$ scale for the 3.2-cm wavelength radar at Wallops Island is usually within the viscous dissipation region. For this reason and also because the minimum detectable reflectivity of the 3.2-cm radar is relatively high, clear air scattering from refractive index inhomogeneities is rarely detected with 3.2-cm radars.

Ottersten (1970) describes the clear air scattering layers as occurring within stable regions at zones of enhanced static stability across which vertical shear of the horizontal wind is accentuated. As the shears are strengthened by larger scale processes, small-scale overturning and turbulence may break out within the thin zones or layers. If the turbulence within the thin layers is not able to erode the shear fast enough, breakdown at a progressively larger scale will take place. Tilting of very stable layers by dynamical processes such as internal fronts, gravity waves, and mountain waves accentuates the vertical wind shear and on occasion may lead to final breakdown over large vertical depths with escalation of the turbulence to scales and intensities of concern for aviation.

An example of the multiple stratifications which sometimes occur is illustrated in (F14.11). This is an RHI photograph of the 10.7-cm radar at Wallops Island. More than ten separate, horizontally stratified, clear air layers are visible below 4 km. A thin cloud layer appeared near 6 km. However, the major scattering even from this layer was due to refractive index inhomogeneities at the cloud boundaries because the layer was not detected with the more sensitive particle detecting 3.2-cm radar [see (T14.2)]. The echo layer near 5 km has a "braided" appearance and similar structures will be discussed in later sections.

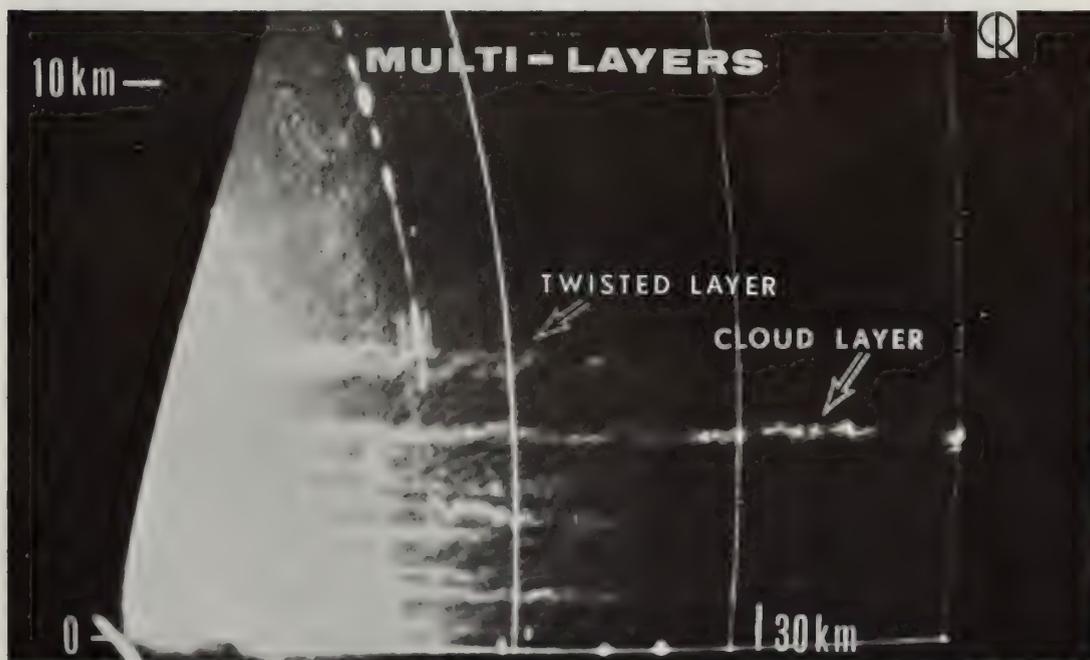


Figure 14.11 RHI photo at 120 deg. azimuth taken at 1515 EST on 9 January 1969 with the 10.7-cm radar at Wallops Island, Virginia. More than ten separate, horizontally stratified, clear air layers are visible below 6 km (from Ottersten, 1969).

Multiple stratifications have also been reported recently by Katz (1969), Crane (1970), Kropfli (1971), and Starr and Browning (1972). The scattering layers may be quite shallow, perhaps less than 10 m thick as reported by Gossard et al. (1970), or they may be in the order of 1 km in depth.

14.5.2 Wave-like Patterns in the Clear Atmosphere

Perhaps high power radars have made their greatest contribution to the atmospheric sciences by their capability to view or map a variety of wave structures in the clear air. Hicks and Angell (1968) carried out the first investigation of the wave-like structures observed by high powered radars in the clear atmosphere, and this was rapidly followed by other studies of atmospheric waves (Hicks, 1969; Gossard et al., 1970; Atlas et al., 1970; Boucher, 1970; Browning and Watkins, 1970).

Figure (14.12) adapted from Scorer (1969a) and Browning and Watkins (1970) illustrates the stages in the development of "billows," a wave pattern which often leads to cloud formation at the wave crests (Ludlam, 1967). The effect of the wave development on the mean profiles of temperature and wind velocity is also indicated. Billow development as sketched in (F14.12) is also referred to as Kelvin-Helmholtz instability. It is a form of dynamic instability which occurs within a hydrostatically stable flow wherever there is both an inflection in the velocity profile and a sufficiently strong vertical shear (see profiles in the upper left of the figure). The stage with the closed rotor corresponds to the overturning of the wave into a structure which is often observed with radar. In fact, the radar structures which are observed throughout the history of the wave development are indicated by the darker line. The light line represents the stream surface in the portion of the wave which breaks; usually this detailed structure is too small to be resolved with radar. Note that in the overturned or "braided" stage, the radar outlines mainly the boundaries of the rotor circulation. This is so because the turbulence inside the rotor tends to destroy the refractive index gradients within the center of the layer but will accentuate the gradients at the boundaries where the air parcels have their maximum displacement. As the final stage, the echoes are in the form of "stretched filaments" and a layer forms both at the bottom and top boundary of the original overturned layer. As seen in the lower right, the mean profiles have been markedly changed by the development and overturning of the wave and by turbulence. A layer with a dry adiabatic lapse rate and uniform wind speed is bounded by two sharp inversions which are characterized by very strong wind gradients. It is the heights of the sharp inversions which define the radar "double-layer."

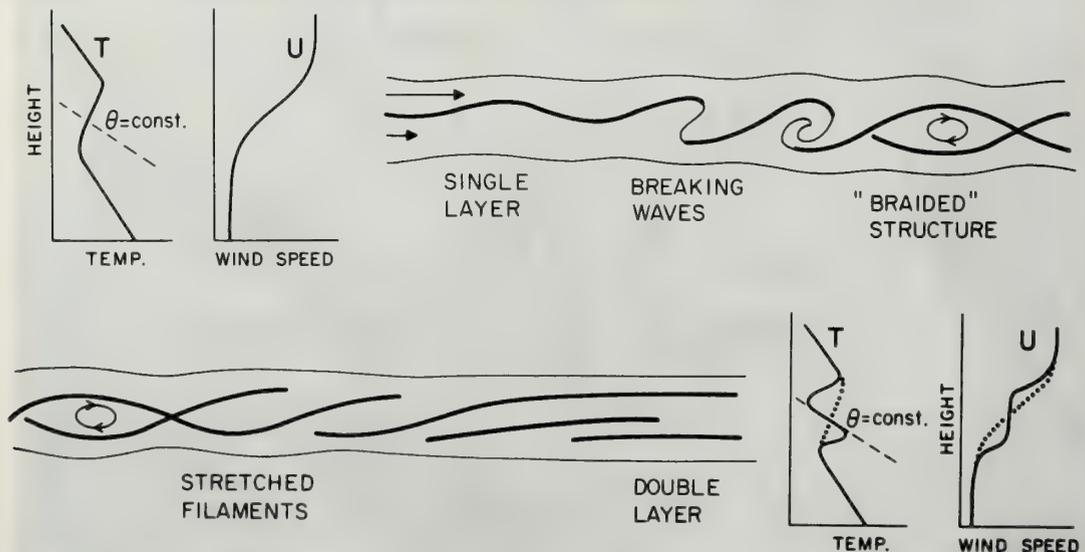


Figure 14.12 Illustration of the stages in the development of an individual Kelvin-Helmholtz billow, a wave pattern sometimes leading to cloud formation at the wave crests. Thick lines correspond to the detectable clear air radar echo, which starts as a single layer (upper half of figure) and finishes as a double layer (lower half of figure). Schematic vertical profiles of temperature and wind speed are indicated before and after the occurrence of Kelvin-Helmholtz instability (adapted from Browning and Watkins, 1970).

Figure (14.13), taken from Boucher (1970), is one of the better examples of a clear air wave structure seen by radar. Note the similarity of the pattern to the "braided" structure sketched in (F14.12). In the analysis of the meteorological data associated with the wave pattern, Boucher (1970) found that the radar echoes occurred within a statically stable (baroclinic) zone characterized by strong wind shear and low Richardson number. In this case clear air radar echoes were observed over the 3 hours of observations and an aircraft involved in the program reported moderate turbulence throughout most of the time spent within the baroclinic zone.

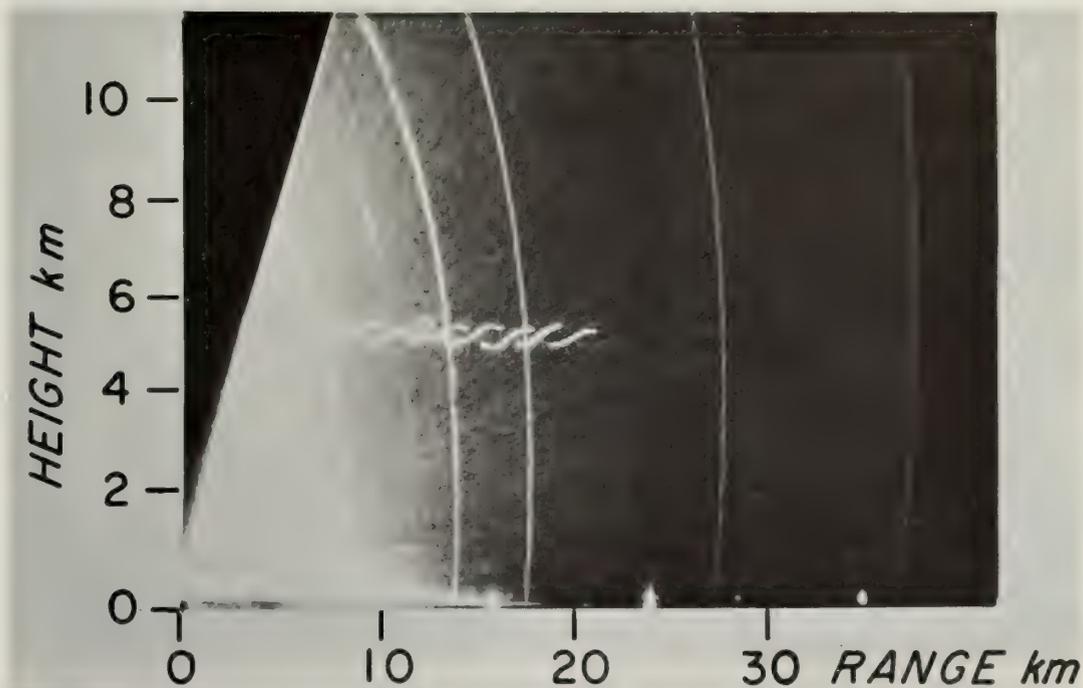


Figure 14.13 *Retouched RHI photo taken 15 January 1969 at 1235 EST with the 10.7-cm radar at Wallops Island, Virginia, azimuth 145 deg. The clear air wave structure was associated with moderate CAT and occurred within a statically stable zone of strong shear (from Boucher, 1970).*

Another outstanding example of a "braided" wave pattern and the associated meteorological sounding are shown in (F14.14) and (F14.15) respectively. The wave pattern occurs near a height of 11 km and is above an extensive region of cloud and precipitation. In (F14.15), it is seen that the wave is within a region of strong stability and marked wind shear. The gradient Richardson number, which is the ratio of the negative buoyancy force to the inertial forces, is plotted to the right of (F14.15). It has its lowest value within the zone of the radar echoes, and this feature is consistent with the theory for the development of Kelvin-Helmholtz instability.

An entirely different type of wave has been studied by Starr and Browning (1972). They used the powerful 10.7-cm Defford radar which is situated 60 km east of the lee slope of the South Wales mountains. This is a suitable location for observing organized lee waves, and an example of these waves is shown in (F14.16). The RHI photograph was taken while the radar scanned toward 270° which was in the upwind direction. Echoes below 4.5 km were mainly from the clear air, those above 6 km were from cirrus, and those between 4.5 and 6 km probably had contributions from both sources. On this occasion two families of lee waves existed. One family, with a wavelength of 8-10 km, had a maximum crest to trough amplitude of only about 200 m at a height of 4 km; this is difficult to see in (F14.16). The other family of lee waves, with a wavelength of 19 km, attained a maximum amplitude of as much as 700 m at a height of 6-7 km. Both the short and long waves maintained fixed positions with respect to the ground in the lee of the Welsh mountains.



Figure 14.14 Photograph of RHI at wavelength of 10.7-cm, 1509 EST, 7 Feb 1968 at Wallops Island, Virginia, azimuth 270 deg. Cloud and precipitation extend from the surface to about 10 km. The unusual clear-air echo structure of apparently crossing waves out of phase occurs at 11.3 km, the height of the tropopause. This is a good example of the "braided" structure illustrated in (F14.12) (from Hardy et al., 1969).

Reed and Hardy (1972) described some long waves which, in contrast to the stationary lee waves observed by Starr and Browning (1972), were embedded within the mean flow. A portion of one of the waves is shown in (F14.17). It has a wavelength of about 30 km and a crest to trough amplitude of 1.8 km; other RHI photographs showed more than one full wavelength of the long wave. Large secondary billows (Scorer, 1969b) formed in the up-slope of the long wave as shown in (F14.17). These billows developed where the environmental wind (increasing with altitude) and the tilting by the wave together accentuated the vertical wind shear and induced dynamic instability. The billows have a wavelength of about 1.6 km and a crest to trough amplitude of about 200 m. Moderate to severe turbulence was encountered by a NASA T-33 aircraft as it probed the clear air wave patterns observed by the radar.

14.5.3 Atmospheric Structure in the Vicinity of Kelvin-Helmholtz Billows

Browning (1971) has obtained radiosonde measurements within one hour of the occurrence of large amplitude billows as observed with the 10.7-cm radar in Defford, England. He restricted his study to billows of large amplitude because these were easy to resolve with high power radar and because it was only in these cases that it was possible to obtain profiles of stability and shear with the necessary resolution for a meaningful comparison with the radar data. Most of the billows had a crest to trough amplitude between 300 and 400 m and occurred at heights between 5.6 and 10.7 km.

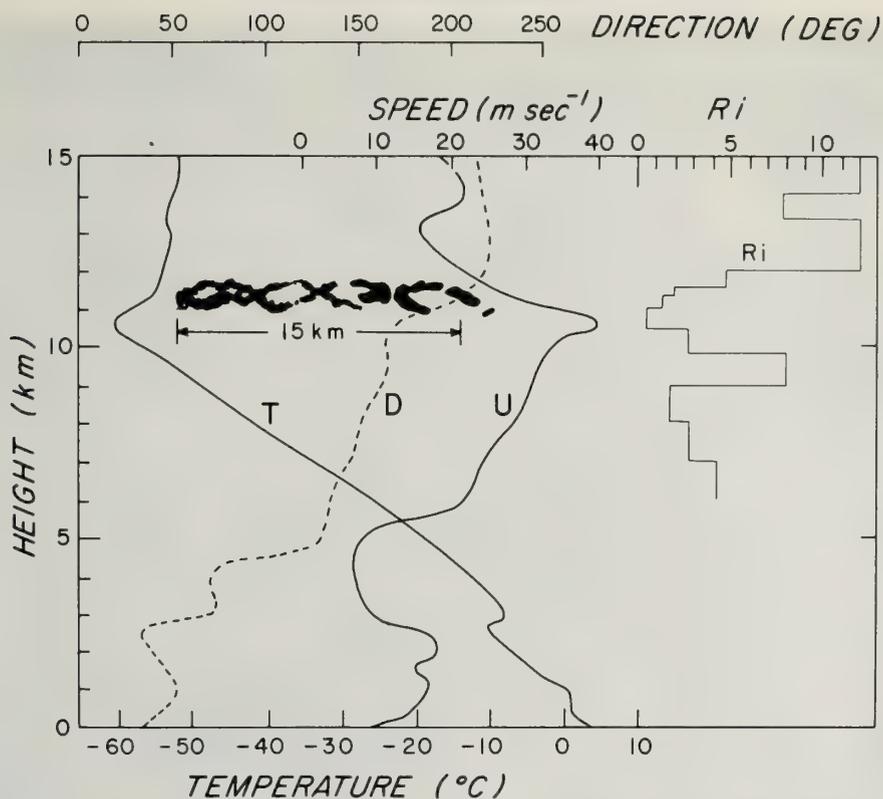


Figure 14.15 Meteorological situation associated with the clear air radar echoes of (F14.14). The curve marked *T* is the temperature; *U*, the wind speed; *D*, the wind direction, and *Ri*, the Richardson number. Note that the wave structure occurs in a region of pronounced static stability and strong shear with a low value of *Ri*.

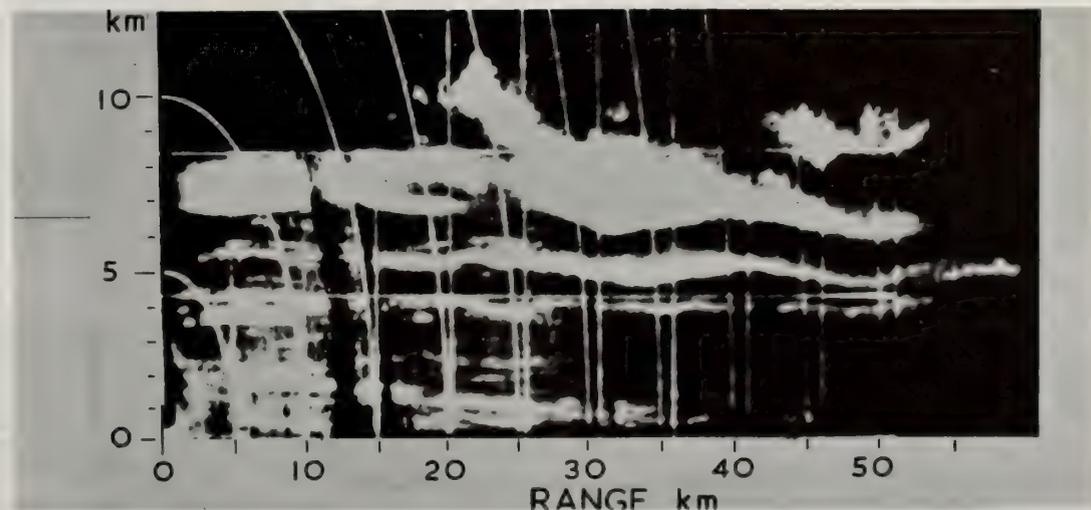


Figure 14.16 Photograph of the RHI display along azimuth 270 deg (upwind direction) at 1124 GMT on 15 April 1970 taken with the 10.7-cm radar at Defford, England. Height markers are at 4.3 and 8.3 km. Vertical markers represent horizontal range at 5.1 km intervals and the curved markers represent slant ranges at 5.0 km intervals. Notice the waves of rather large amplitude between 5 and 7 km height. Echoes above 6 km are from cirrus particles; those below 4.5 km are from the clear air (Starr and Browning, 1972).



Figure 14.17 Photograph of the RHI at radar wavelength of 10.7-cm, 0332 GMT, 18 March 1969, azimuth 270° , Wallops Island, Virginia. The wind and shear are directed toward the left in this photo. Note the portion of the long arch and the superposition of the billow structure. The vertical scale is expanded by a factor of about three and this accounts for the seemingly large slope of the long wave. The inset is a tracing from the original record (from Reed and Hardy, 1972).

Theoretical studies indicate that Ri , the Richardson number, should be less than or equal to 0.25 before Kelvin-Helmholtz instability can be initiated (Miles and Howard, 1964). In the 17 cases investigated, Browning (1971) found that the minimum value of Ri when evaluated over layers 200 m deep was usually in the range 0.15-0.3. The results for the 17 cases are shown in (F14.18). Considering the difficulty of investigating this phenomenon, the agreement between experimental and theoretical results is remarkable.

14.5.4 Waves and Turbulence in a Stratified Atmosphere

As discussed by Stewart (1969) and Bretherton (1969), it is often difficult to distinguish clearly between wave motion and turbulence in a stratified fluid. Turbulence, as used in this chapter, is described by an input of energy into the energy-containing eddies at large scales (in the order of 100 m) and by a downward cascading within the inertial subrange into smaller and smaller scales until the turbulent energy is dissipated into heat in the viscous dissipation range (scales of less than about 1 cm). The eddies are random and the turbulent field is isotropic.

Billows or Kelvin-Helmholtz instabilities form within layers of strong static stability in the presence of sufficiently strong wind shear. The "wave-instability" drifts with the mean wind of the shear layer. While the billow is in the process of breaking (F14.12), the eddies are essentially two-dimensional, but they probably break down into three-dimensional turbulence at the smaller scales. A pilot flying within a layer of billows would describe the region as being turbulent, but he may also experience periodic buffeting as he traverses successive billows.

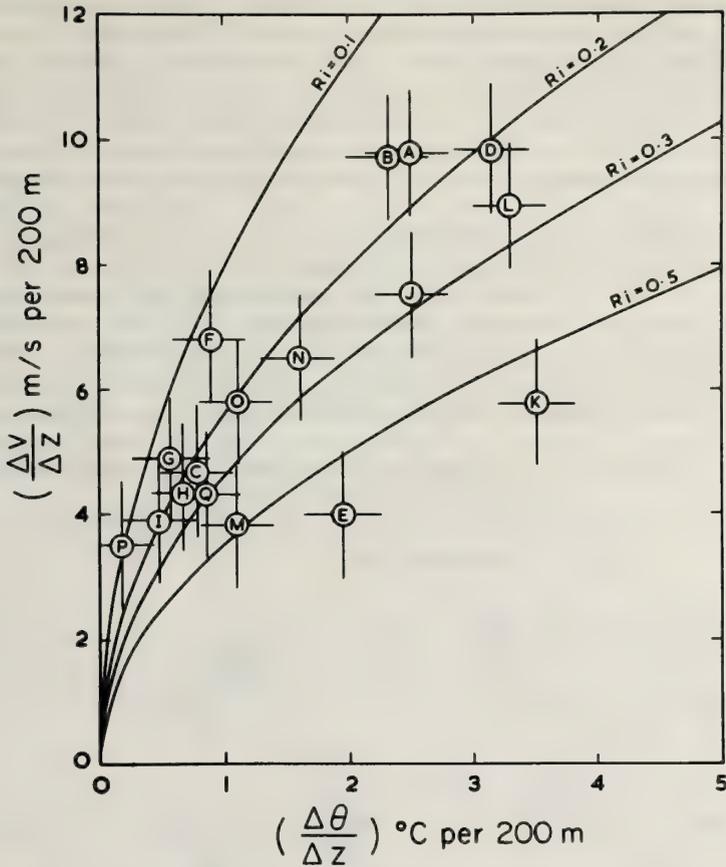


Figure 14.18 The maximum value of the wind shear, $\Delta V/\Delta Z$, over layers 200 m deep within the height interval occupied by the Kelvin-Helmholtz billows, plotted against the corresponding value of the vertical gradient of potential temperature, $\Delta\theta/\Delta Z$, for 17 cases. Solid lines are isopleths of Ri over the corresponding layers (from Browning, 1971).

True waves propagate relative to the flow; some propagate to great heights, carrying energy and momentum with them. Although these internal gravity waves may have a large range of wavelengths in the atmosphere, generally their wavelength is sufficiently long that they are undetected by aircraft. These long waves may be important for the generation of some turbulence, however, because with a stratified shear flow the tilting produced by the wave creates local regions of decreased Richardson number where dynamic instability may be induced (Scorer, 1969a; Reed and Hardy, 1972).

Dutton and Panofsky (1970), on the basis of available empirical knowledge, have concluded that at least some of the clear air turbulence results from the hydrodynamic instability of internal fronts in accord with the Kelvin-Helmholtz model. Probably this is the dominant mechanism for the breakdown of shear layers. However, Scorer (1969a) states that there are several possible sources of CAT and encourages investigators not to overlook these possibilities.

14.6 Radar Investigations of Clear Air Turbulence

14.6.1 Clear Air Radar Echoes and Aircraft Turbulence

As indicated in (14.1.2), it is expected that there will be some correlation between the strength of the clear air radar reflectivity and severity of clear air turbulence (CAT) especially in the stratosphere and upper troposphere where moisture contributions to the refractive index are small. It is important to realize, however, the large gap between the turbulent eddy sizes which affect aircraft and those which create the

refractive index inhomogeneities detectable by radar. The radars are only sensitive to refractive index variations over scales of one half of the radar wavelength. Thus, for the 10.7-cm and 71.5-cm wavelength radars at Wallops Island, the clear air returns depend on the turbulent energy confined to eddy sizes of about 5 and 35 cm respectively. On the other hand, aircraft are mainly responsive to turbulent eddies of scales in the order of 10-1000 m, the high values applicable for heavy, high speed aircraft. However, because the energy at large eddy sizes gradually breaks down into smaller and smaller turbulent eddies, it is expected that the radar measurements give some information on the turbulent kinetic energy at the larger scales.

Various investigators have described joint radar and aircraft studies of CAT (Hicks et al., 1967; Glover et al., 1968; Glover et al., 1969; Hardy et al., 1969; Crane, 1970; Browning et al., 1970; Glover and Duquette, 1970). Using the radars at Wallops Island, Glover and Duquette (1970) found that all altitude intervals corresponding to clear air radar layers between 0.5 and 15 km, when probed with fighter jet-aircraft, were turbulent. The results of the 53 flights during the winters of 1969 and 1970 are given in (T14.3). About 12% of the light or greater turbulence above 6 km was not detected by the radar. However, the stronger turbulence is generally detected with greater probability than the lighter turbulence.

Table 14.3. Percentage of observed CAT detected by Wallops Island radars during 1969 and 1970 (from Glover and Duquette, 1970).

CAT INTENSITY	ALTITUDE (km)				
	0.5-3	3-6	6-9	9-12	12-15
LESS THAN LIGHT	100	88	44	67	100
LIGHT OR GREATER	100	100	81	93	92

Crane (1970) has verified that more sensitive radars are able to detect stratified layers which occur in the lower 10 km of the stratosphere and which may extend out to ranges of 200 km. He used the 23-cm wavelength radar (T14.1) which, for scattering from refractive index variations, is at least 10 dB more sensitive than the radars at Wallops Island. A series of five tests was conducted during May and June, 1968; each test consisted of nearly simultaneous observations of thin layers in the stratosphere using the radar and an Air Force high altitude aircraft. The pilot encountered 28 layers of turbulence throughout the tests: he characterized the turbulence as very light in 23 layers, as light in 4 layers, and as moderate in 1 layer. For these tests, the radar detected thin scattering layers at the height of the aircraft encounters and near the horizontal position of the aircraft for 5 of the very light encounters, 3 of the light encounters, and the single moderate encounter.

The results of Crane's tests showed that thin turbulent layers which affected both the aircraft and the radar were present in each of the five test series. Layers reported as having light or moderate turbulence by the aircraft were also indicated by temperature inversions in the radiosonde data and were generally detected by the radar.

14.6.2 Radar and Instrumented Aircraft Measurements of Turbulence

Although there have been many occasions of turbulence observations with radar and aircraft, only a few of the flights have been with aircraft equipped with meteorological sensors. Browning et al. (1970) described the observations of moderate CAT in which measurements of a fully instrumented aircraft were related to simultaneous radar observations of a specific patch of Kelvin-Helmholtz billows. The aircraft and radar observations on that occasion are shown in the distance-time diagram in (F14.19) where distance is the range upwind from the radar. The shaded areas in this figure represent patches of billows detected by

the radar at heights from 10.5 to 11 km. The two lines AB and CD represent two aircraft flights made at a height of 10.9 km. Although the aircraft appears to have missed the billows of largest amplitude, the maximum rms vertical acceleration of 0.65 g (corresponding to moderate, almost severe turbulence) occurred close to the patch of 500-m amplitude billows observed by radar. It is also noted that there is a tendency for the maximum billow amplitude to occur in the vicinity of the most pronounced lee wave crest.

Mather and Hardy (1970) also present instrumented aircraft measurements in the vicinity of clear air radar structures. Figure (14.20) shows the aircraft data observed at the same time that the radar was detecting a clear air wave structure. A portion of the radar-observed wave which occurred at the same time and in the same vicinity as the aircraft is shown as an inset in the figure. The line on the photograph represents the altitude of the flight path with respect to the wave.

The top trace in (F14.20) is the component of the wind parallel to the aircraft track (073 deg), averaged over 2.5 seconds. The second trace is the differential static temperature averaged in the same way. Trace 3 (ΔT_T) is the differential total temperature measured by a fast-response temperature probe with a frequency response flat to 10 Hz. It is shown here to illustrate how sharp the density gradient becomes across the interface between the smooth and turbulent flow. The bottom two traces are the lateral (v) and vertical (w) components of turbulence measured in the layer.

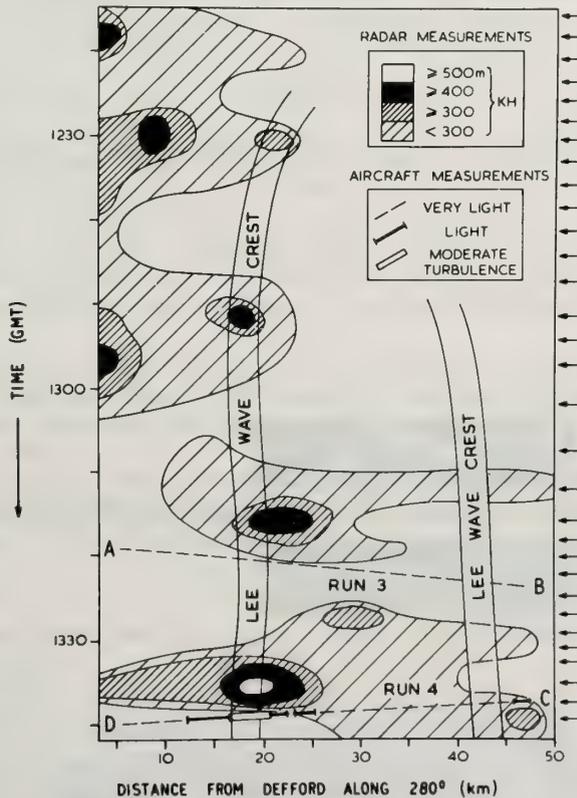


Figure 14.19 Locations of detectable 10.7-cm wavelength radar echoes associated with Kelvin-Helmholtz billows at tropopause level on 3 February 1970 near Defford, England with the radar pointing towards 280 deg. The crest-to-trough amplitude of the billows is indicated by shading (see key in figure). Also shown are the positions of crests in the lee wave pattern as observed by radar. Arrows on the right side of the diagram denote times of individual RHI scans. Aircraft turbulence reports along the path AB and CD are plotted according to the code indicated in the key. Time and space registration between the radar and aircraft data is thought to have been closer than 1 minute and 3 km, respectively (from Browning et al., 1970).

Since the displacement of the air by the wave is approximately adiabatic, temperatures will be lower than the mean when the aircraft is sampling air brought up from below and vice versa for air brought down. Consequently, the stream surface will be exactly out of phase with the temperature structure. Similarly, since the wind increased with height throughout the layer containing the wave, the U-component of wind is also out of phase with the stream surface.

Figure (14.20) can be thought of as a slice through the wave at an instant in time, since the aircraft speed is large compared to the wave speed. Six cycles of the wave have been measured. As the amplitude of the wave grows, bursts of turbulence are observed in the wave troughs (1727:24 and 1728:40 as indicated by the arrows) until, at 1729:45 (horizontal bracket), a 20-second outbreak of moderate turbulence is experienced in the trough of the large amplitude wave. As seen from the sharp changes in ΔT_T in the last two cycles of the wave, the turbulent mixing process appears to be intensifying the density gradient across the bottom interface between the smooth and turbulent layers. This could also enhance the radar echo from the structure.

The brief description of two cases of radar and instrumented aircraft observations serves to demonstrate the value of these combined observational techniques. Glover and Duquette (1970) conclude that sensitive ground based radars are reliable sensors of regions of CAT. Of equal importance, however, is that these sensitive radar systems provide an unequalled opportunity for detecting and investigating atmospheric structures associated with the production and life cycle of turbulence in the free atmosphere. In experimental programs on CAT, the radars can also provide the vital information for guiding instrumented aircraft into the regions of most interest.

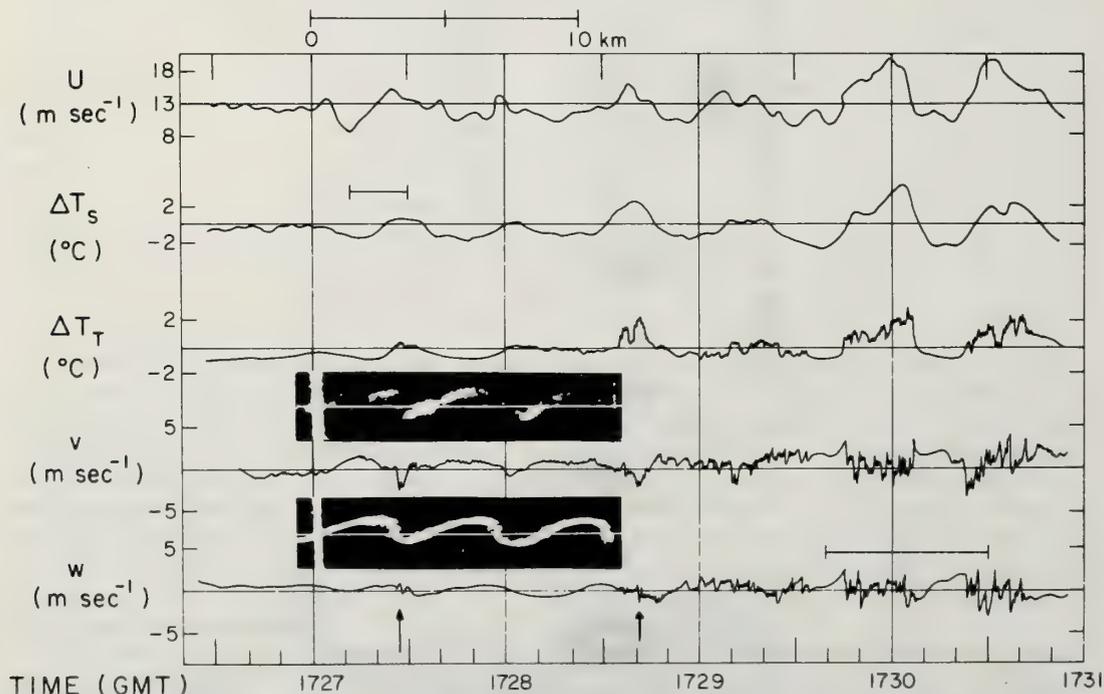


Figure 14.20 Data obtained by a T-33 aircraft on 19 February 1970 at a height of 2600 m near Wallops Island, Virginia, along an azimuth of 073 deg. Time histories show longitudinal wind speed U , differential static temperature ΔT_S , differential total temperature ΔT_T , and lateral component v and vertical component w of turbulence measured in a layer giving clear air radar echoes. The upper inset is a photograph of the 10.7-cm wavelength RHI scope and the lower inset is a retouched photo of the same data. The RHI photo was taken about 1727:15 GMT along an azimuth parallel to the aircraft track but the two sets of data were displaced by about 5 km. The radar pattern was enlarged to match the equivalent range sampled by the aircraft and the pattern was shifted slightly in the figure in order that the wave structure and aircraft data were mutually consistent. The white horizontal line is the altitude of the flight path with respect to the waves. Note that the turbulence is associated mainly with the troughs of the waves (adapted from Mather and Hardy, 1970).

14.7 Summary

Some of the key results on the structure of the clear atmosphere as derived from observations using high power radars with steerable antennas have been described. During the initial application of sensitive radars to atmospheric probing, research was directed toward the verification of the causes of the echoes. By 1966 it had been confirmed that although some echoes were due to birds and insects, most clear air echoes seen at wavelengths of 10 cm or longer were due to the scattering from refractive index inhomogeneities.

These inhomogeneities are associated with sharp gradients of potential refractive index (potential temperature and/or specific humidity) which occur in the atmosphere near the boundaries of convective cells and in stratified layers; in the low troposphere the contribution due to humidity is nearly always dominant. Turbulence within the regions of sharp gradients leads to inhomogeneities of refractive index which are responsible for the backscatter detectable with sensitive radars. The inhomogeneities occur over a range of eddy sizes, but the intensity of the energy scattered depends on the intensity of turbulence on a scale of half the radar wavelength and on the mean gradient of refractive index. Investigations with radar and direct sensors corroborate the theory as outlined, and the radar reflectivity of the clear air echoes is in quantitative agreement with the reflectivity expected from observed refractive index structures.

Although insects occasionally occur in sufficient numbers that they can be used as effective tracers of atmospheric motions, the scattering from the refractive index inhomogeneities is detected consistently with sensitive radars. It is a rare occasion when sensitive radars of 10-cm or longer wavelength fail to detect some clear air scattering within the troposphere. During the summer months, convective patterns in the lower 3 km are detected. The details of individual convective elements 1-3 km in diameter have been described with the help of high power radar. In addition the height and character of the convective field have been monitored throughout the day. The convective process can now be investigated throughout all stages of storm development. Moreover, it has been suggested that a careful evaluation of clear air convective patterns could lead to the determination of the most likely locations for future convective growth and thus provide a greatly enhanced short-range forecasting capability.

Investigations utilizing sensitive radars have led to an improved understanding of wave structures in the clear atmosphere. It appears that Kelvin-Helmholtz instability is the dominant mechanism for exchanging momentum in stratified shear flows. Radar outlines the stages in the development of Kelvin-Helmholtz instability and makes it possible to investigate the details of the structures which lead to turbulence.

It has been established that sensitive radars will detect most of the significant CAT out to ranges of about 30 km. This result is consistent with theoretical investigations which indicated that the radar reflectivity in the upper troposphere and stratosphere (where moisture influences are small) will be a good indicator of the CAT severity.

The vigorous use of sensitive radars over the past few years has demonstrated the value of these remote probes for deducing atmospheric structures and processes in the clear atmosphere. However, the full potential of using sensitive radars for atmospheric research has not been realized. As suggested by Ottersten (1969a), we must now proceed from an essentially descriptive science to one which is quantitative and thereby establish much closer ties with meteorological theory.

14.7.1 Acknowledgements

Many of the results presented in this chapter were obtained by my colleagues, and I would like to thank them for the many discussions which helped in my understanding of radar meteorology. In particular, I have drawn heavily on the theoretical work carried out by Mr. Hans Ottersten while he was with Air Force Cambridge Research Laboratories and on the experimental work completed by Messrs. Kenneth M. Glover and Roland J. Boucher of AFCRL and by Mr. Isadore Katz and his colleagues at Applied Physics Laboratory, Johns Hopkins University. The chapter has been significantly improved as a result of the painstaking reviews of Mr. Hans Ottersten and two anonymous reviewers and I would like to thank them for their important contributions. The research conducted at Wallops Island, Virginia has been supported in part by the National Aeronautics and Space Administration, Wallops Station.

14.8 References

- Atlas, D. (1959), Radar studies of meteorological "angel" echoes, *J. Atmos. Terr. Phys.*, *15*, 262-287.
- Atlas, D. (1964), Advances in radar meteorology, *Advan. Geophys.*, *10*, 317-478.
- Atlas, D., and K. R. Hardy (1966), Radar analysis of the clear atmosphere: angels, *Proc. XV General Assembly of URSI, Munich, Germany, 5-15 Sept.*, 401-469.
- Atlas, D., K. R. Hardy, and K. Naito (1966a), Optimizing the radar detection of clear-air turbulence, *J. of Appl. Meteor.*, *5*, 450-460.
- Atlas, D., K. R. Hardy, and T. G. Konrad (1966b), Radar detection of the tropopause and clear air turbulence, *Proc. 12th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass.*, 279-284.
- Atlas, D., J. I. Metcalf, J. H. Richter, and E. E. Gossard (1970), The birth of "CAT" and microscale turbulence, *J. Atmos. Sci.*, *27*, 903-913.
- Borchardt, H. (1962), Wolkenbeobachtungen mit einem doppelwelligen Radargerät, *Beit. Physik Atmosphäre*, *35*, 43-68. See also Atlas (1964).
- Boucher, R. J. (1970), CAT at a subsidence inversion: a case study, *J. Appl. Meteor.*, *9*, 534-537.
- Browning, K. A. (1971), Structure of the atmosphere in the vicinity of large-amplitude Kelvin-Helmholtz billows, *Quart. J. Roy. Meteor. Soc.*, *97*, 283-299.
- Browning, K. A. (1972), Atmospheric research using the Defford radar facility, *Weather*, *27*, 2-13.
- Browning, K. A., and D. Atlas (1966), Velocity characteristics of some clear-air dot angels, *J. Atmos. Sci.*, *23*, 592-604.
- Browning, K. A., and C. D. Watkins (1970), Observations of clear air turbulence by high power radar, *Nature*, *227*, 260-263.
- Browning, K. A., C. D. Watkins, J. R. Starr, and A. McPherson (1970), Simultaneous measurements of clear air turbulence at the tropopause by high-power radar and instrumented aircraft, *Nature*, *228*, 1065-1067.
- Chernikov, A. A. (1966), Some new Soviet investigations of angel echoes, *Proc. 12th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass.*, 291-292.
- Crane, R. K. (1970), Measurement of clear air turbulence in the lower stratosphere using the Millstone Hill L-Band radar, *Preprints 14th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass.*, 101-106.
- Crawford, A. B. (1949), Radar reflections in the lower atmosphere, *Proc. IRE*, *37*, 404-405.
- Dutton, J. A., and H. A. Panofsky (1970), Clear air turbulence: a mystery may be unfolding, *Science*, *167*, 937-944.
- Eastwood, E., 1967: *Radar Ornithology*, Methuen, London.
- Fehlhaber, L., and J. Grosskopf (1964), Untersuchung der structure der troposphäre mit einem vertikalradar, *Nachrichtentechnische Zeitschrift*, *17*, 503-507. English Translation - AFCRL, Bedford, Mass., TG-242, March 1965.

- Glover, K. M., and E. F. Duquette (1970), A study of clear air turbulence using sensitive radars, Preprints 14th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 89-94.
- Glover, K. M., and K. R. Hardy (1966), Dot angels: insects and birds. Proc. 12th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 264-268.
- Glover, K. M., R. J. Boucher, H. Ottersten, and K. R. Hardy (1968), Radar, aircraft, and meteorological investigation of clear air turbulence, Proc. 13th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 242-247.
- Glover, K. M., R. J. Boucher, H. Ottersten, and K. R. Hardy (1969), Simultaneous radar, aircraft, and meteorological investigations of clear air turbulence, J. Appl. Meteor., 8, 634-640.
- Glover, K. M., K. R. Hardy, T. G. Konrad, W. N. Sullivan, and A. S. Michaels (1966), Radar observations of insects in free flight, Science, 154, 967-972.
- Gossard, E. E., J. H. Richter, and D. Atlas (1970), Internal waves in the atmosphere from high-resolution radar measurements, J. Geophys. Res., 75, 3523-3536.
- Gurvich, A. S., M. Koprov, L. R. Tsvang, and A. M. Yaglom (1967), Empirical data on the small-scale structure of atmospheric turbulence, in atmospheric turbulence and radio wave propagation, Proc. Int. Colloq., Moscow, June 15-22, 1965, Nauka, Moscow, 30-52.
- Hajovsky, R. G., A. P. Deam, and A. H. LaGrone (1966), Radar reflections from insects in the lower atmosphere, IEEE Trans. on Antennas and Propagation, 14, 224-227.
- Hardy, K. R., and I. Katz (1969), Probing the clear atmosphere with high power, high resolution radars, Proc. IEEE, 57, 468-480.
- Hardy, K. R., D. Atlas, and K. M. Glover (1966), Multiwavelength backscatter from the clear atmosphere, J. Geophys. Res., 71, 1537-1552.
- Hardy, K. R., and H. Ottersten (1969), Radar investigations of convective patterns in the clear atmosphere, J. Atmos. Sci., 26, 666-672.
- Hardy, K. R., K. M. Glover, and H. Ottersten 1969: Radar investigations of atmospheric structure and CAT in the 3 to 20-km region, in *Clear Air Turbulence and Its Detection*, edited by Y. H. Pao and A. Goldberg, 402-416, Plenum Press, New York.
- Harrold, T. W., and K. A. Browning (1971), Identification of preferred areas of shower development by means of high power radar, Quart. J. Roy. Meteor. Soc., 97, 330-339.
- Hay, D. R., and W. M. Reid (1962), Radar angels in the lower troposphere, Can. J. Phys., 40, 128-138.
- Hicks, J. J. (1969), Radar observations of a gravitational wave in clear air near the tropopause associated with CAT, J. Appl. Meteor., 8, 627-633.
- Hicks, J. J., and J. K. Angell (1968), Radar observations of breaking gravitational waves in the visually clear atmosphere, J. Appl. Meteor., 7, 114-121.
- Hicks, J. J., I. Katz, C. R. Landry, and K. R. Hardy (1967), Simultaneous radar and aircraft observations of clear-air turbulence, Science, 157, 808-809.
- Katz, I. (1966), Probing the clear atmosphere with radar. APL Technical Digest, Applied Physics Laboratory, Johns Hopkins University, Sept.-Oct., 2-8.

- Katz, I. 1969: Probing the optically clear atmosphere with radar, in *Clear Air Turbulence and Its Detection*, 417-424, Plenum Press, New York
- Konrad, T. G. (1968), The alignment of clear-air convective cells, Proc. Int. Conf. on Cloud Physics, Amer. Meteor. Soc., Boston, Mass., 539-543.
- Konrad, T. G. (1970), The dynamics of the convective process in clear air as seen by radar, J. Atmos. Sci., 27, 1138-1147.
- Konrad, T. G. and J. Hicks (1966), Tracking of known bird species by radar, Proc. 12th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 259-263.
- Konrad, T. G. and D. Randall (1966), Simultaneous probing of the atmosphere by radar and meteorological sensors, Proc. 12th Conf. on Radar Meteorology, Amer. Meteor. Soc., Boston, Mass., 300-305.
- Konrad, T. G. and R. A. Kropfli (1968), Radar observations of clear-air convection over the sea, Proc. 13th Radar Meteor. Conf., Amer. Meteor. Soc., Boston, Mass. 262-269.
- Konrad, T. G., J. J. Hicks, and E. B. Dobson (1968), Radar characteristics of birds in flight, Science, 159, 274-280.
- Kropfli, R. A. (1971), Simultaneous radar and instrumented aircraft observations in a clear air turbulent layer, J. Appl. Meteor., 10, 796-802.
- Kropfli, R. A., I. Katz, T. G. Konrad, and E. B. Dobson (1968), Simultaneous radar reflectivity measurements and refractive index spectra in the clear atmosphere, Radio Sci., 3, 991-994.
- Lack, D., and G. C. Varley (1945), Detection of birds by radar, Nature, 156, 446.
- Lane, J. A., 1967a: *Structure of the Lower Atmosphere and Electromagnetic Wave Propagation*, NATO Advanced Study Institute, directed by J. A. Lane, 2-15 Sept. 1967, Aberystwyth, Wales.
- Lane, J. A. (1967b), Radar echoes from tropospheric layers by incoherent backscatter, Electronic Lett., 3, 173-174.
- Lhermitte, R. M. (1966), Probing air motion by Doppler analysis of radar clear air returns, J. Atmos. Sci., 23, 575-591.
- Lhermitte, R. M., and J. T. Dooley (1966), Doppler study of the motion of clear-air targets, Proc. 12th Conf. on Radar Meteorology, Amer. Meteor. Soc., Boston, Mass., 293-299.
- Ludlam, F. H. (1967), Characteristics of billow clouds and their relation to clear air turbulence, Quart. J. Roy. Meteor. Soc., 93, 419-435.
- Mather, G. K., and K. R. Hardy (1970), Instrumented aircraft measurements in the vicinity of clear air radar structures. Preprints 14th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 49-55.
- Miles, J. W., and L. N. Howard (1964), Note on a heterogeneous shear flow, J. Fluid Mech., 20, 331-336.
- Ottersten, H. (1964), A theoretical treatment of microwave backscattering from tropospheric turbulence, Forsvarets Forskningsanstalt, Stockholm, Report A623, 16 pp.
- Ottersten, H. (1968), Theoretical aspects on CAT detection by radar. Proc. 13th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 252-257.

- Ottersten, H. (1969a), Atmospheric structure and radar backscattering in clear air, *Radio Sci.*, **4**, 1179-1193.
- Ottersten, H. (1969b), Radar backscattering from the turbulent clear atmosphere, *Radio Sci.*, **4**, 1251-1255.
- Ottersten, H. (1969c), Mean vertical gradient of potential refractive index in turbulent mixing and radar detection of CAT, *Radio Sci.*, **4**, 1247-1249.
- Ottersten, H. (1970), Radar angels and their relationship to meteorological factors, Forsvarets Forskningsanstalt, Stockholm, FOA Reports, **4**(2), 1-33.
- Ottersten, H. (1970), Radar observations of the turbulent structure in shear zones in the clear atmosphere, Preprints 14th Radar Meteorology Conf., Amer. Meteor. Soc., Boston, Mass., 111-116.
- Plank, V. G. (1956), A meteorological study of radar angels, *Geophys Res. Paper* 52, 117 pp., Air Force Cambridge Research Laboratories, Bedford, Mass.
- Reed, R. J., and K. R. Hardy (1972), A case study of persistent, intense clear air turbulence in an upper level frontal zone, *J. Appl. Meteor.*, **11**, Vol. 3, 541-549.
- Roelofs, T. H. (1963), Characteristics of trackable radar angels, Center for Radiophysics and Space Research, Res. Rep. 137, 52 pp., Cornell Univ., Ithaca, N. Y.
- Saxton, J. A., J. A. Lane, R. W. Meadows, and P. A. Mathews (1964), Layer structure of the troposphere – Simultaneous radar and microwave refractometer investigations, *Proc. IEE*, **3**, 275-283.
- Scorer, R. S. (1969a), Mechanisms of clear air turbulence. Symposium on Clear Air Turbulence and its Detection, Seattle, Wash., 14-16 Aug. 1968, Plenum Press, New York, 34-50.
- Scorer, R. S. (1969b), Billow mechanics, *Radio Sci.*, **4**, 1299-1307.
- Smith, P. L., and R. R. Rogers (1963), On the possibility of radar detection of clear-air turbulence, *Proc. 10th Weather Radar Conf.*, Amer. Meteor. Soc., Boston, Mass., 316-322.
- Starr, J. R., and K. A. Browning (1972), Observations of lee waves by highpower radar, *Quart. J. Roy. Meteor. Soc.*, **98**, 73-85.
- Tatarski, V. I., 1961: *Wave Propagation in a Turbulent Medium*. McGraw-Hill Book Co., Inc., New York, N. Y., 285 pp.
- Vrana, N. (1961), Some characteristics of radar angel echoes, Center for Radiophysics and Space Research, Res. Rept. No. 32, 29 pp., Cornell University, Ithaca, New York.
- Yaglom, A. M., and V. I. Tatarski, 1967: *Atmospheric Turbulence and Radio Wave Propagation*, *Proc. Int. Colloq.*, Moscow, June 15-22, 1965, edited by A. M. Yaglom and V. I. Tatarski, 376 pp., Nauka, Moscow.

List of Symbols

D	drop diameter	ϵ	the average rate of dissipation of turbulent kinetic energy per unit mass
g	acceleration of gravity		
k	radian wave number	ϵ_n	the average rate of dissipation of the inhomogeneities in the refractive index field by molecular diffusion
L	scale length = $2\pi/k$		
m	complex index of refraction	ϵ_T	the average rate of dissipation of the inhomogeneities in the temperature field by molecular diffusion
n	refractive index of air		
P_0	atmospheric pressure at reference level o	η	radar reflectivity
T	atmospheric temperature	θ	potential temperature
u	horizontal component of motion	λ	radar wavelength
w	vertical component of motion	$\phi_n(k)$	spatial power-spectral density of refractive index at vector wavenumber k where k is the radar's radial direction
C_n^2	structure constant for the field of refractive index		
C_T^2	structure constant for the field of temperature		
$F_n^x(k_x)$	one-dimensional spectral density of refractive index along the x-axis at wavenumber k.		
R_f	flux Richardson number		
$S_n^x(k_x)$	one-dimensional spectral density of refractive index as obtained from measurements of a discrete sensor carried through the air along the x-axis		

Chapter 15 RADIOMETRY

Jack B. Snider and Ed R. Westwater
Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Radiometric measurements of thermal energy emitted or scattered by the atmosphere or by natural surfaces can be used to infer physical properties of the medium under observation. These passive methods of remote sensing require knowledge of the generation and propagation of the radiation as well as a complete understanding of its measurement by the radiometer. This chapter discusses remote sensing by several ground- and satellite-based passive techniques after first presenting basic concepts in radiative transfer and radiometry.

15.0 Introduction

Remote sensing by passive radiometry requires sensitive measurements of non-coherent electromagnetic energy that is emitted or scattered by the medium under observation. Applications of passive techniques range from satellite determination of surface parameters to ground-based observations of temperatures in the stratosphere. In this chapter, we will summarize basic concepts in the transfer of microwave and infrared radiation. This presentation is followed by a general discussion of radiometers. Finally, a few examples of remote sensing by radiometers are given. Attention is confined to microwave and infrared portions of the spectrum.

15.1 Radiative Transfer

In describing an unpolarized radiation field within a medium, it is convenient to consider the energy dE_ν in the frequency interval $(\nu, \nu+d\nu)$ passing through an area dA within a solid angle $d\Omega$ about a direction specified by a unit vector \vec{s} in a time interval dt . This energy is related to the specific intensity I_ν by

$$dE_\nu = I_\nu \vec{s} \cdot \vec{n} dA d\Omega d\nu dt \quad , \quad (15:1)$$

where \vec{n} is the unit normal to dA (see F15.1). In general, I_ν is a global function of both position P and viewing direction \vec{s} in a manner determined by the emitting, absorbing, and scattering characteristics of the entire medium: $I_\nu = I_\nu(P, \vec{s})$. We will briefly describe these characteristics below, following Kondratyev (1969).

The emission from a volume can be defined by an emission coefficient, j_ν , such that the emission into $d\Omega$ by a mass dm , again within $(\nu, \nu+d\nu)$ and $(t, t+dt)$ is

$$dE_\nu = j_\nu dm d\Omega d\nu dt \quad , \quad (15:2)$$

In general, j_ν contains contributions from both thermal emission and scattering. For most atmospheric applications, the thermal contribution to j_ν is independent of direction (an exception is high altitude microwave emission from O_2 in which the earth's magnetic field affects average molecular orientation). When j_ν describes scattering, a directional dependence on scattering angle is usually present.

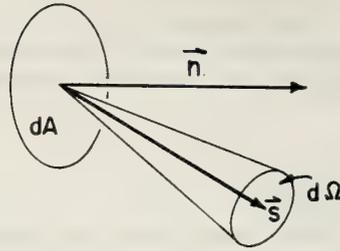


Figure 15.1 Geometry defining specific intensity.

The volume extinction coefficient, e_ν , is defined such that the energy removed per unit time per unit solid angle per unit frequency by a volume of cross section dA and length dr from a beam of intensity I_ν in a direction normal to dA is

$$\frac{d E_\nu}{d\nu d\Omega dt} = -e_\nu I_\nu dA dr \quad , \quad (15:3)$$

or

$$d I_\nu = -e_\nu I_\nu dr \quad , \quad (15:3a)$$

This extinction of energy can occur either by absorption, in which energy from the radiation field is converted into heat, or by scattering, in which the electromagnetic energy is simply redirected. For simplicity, we will neglect higher-order processes in which a photon is absorbed at one frequency and emitted at others. The extinction can thus be expressed as the sum of an absorption coefficient α_ν and a scattering coefficient σ_ν

$$e_\nu = \alpha_\nu + \sigma_\nu \quad , \quad (15:4)$$

We will now consider in detail the processes which enhance the energy in the beam. To describe scattering, the angular distribution of scattered energy must be given. This is done by introducing the scattering function $p_\nu(\vec{s}', \vec{s})$ which describes the fraction of the energy incident from the direction \vec{s}' which is scattered into the direction \vec{s} . The scattering function is usually normalized such that

$$\int p_\nu(\vec{s}', \vec{s}) \frac{d\Omega}{4\pi} = 1 \quad , \quad (15:5)$$

Thus the spectral power density incident from the solid angle $d\Omega'$ which is scattered into $d\Omega$ by an element of volume dV (see F15.2) is (per unit solid angle)

$$\frac{d^2 E_\nu}{d\nu dt d\Omega} = \frac{\sigma_\nu}{4\pi} p_\nu(\vec{s}', \vec{s}) I_\nu(\vec{s}') d\Omega' dV \quad , \quad (15:6)$$

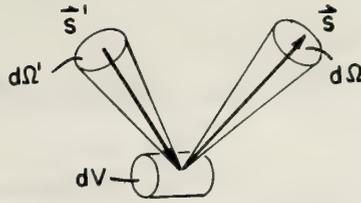


Figure 15.2 Geometry defining scattering volume.

and the total amount of energy scattered into $d\Omega$ per unit time per unit bandwidth per unit solid angle is

$$\frac{d E_{\nu}}{d\nu dt d\Omega} = \frac{\sigma_{\nu}}{4\pi} dV \int p_{\nu}(\vec{s}', \vec{s}) I_{\nu}(\vec{s}') d\Omega' \quad , \quad (15:7)$$

or

$$d I_{\nu} = \frac{\sigma_{\nu}}{4\pi} dr \int p_{\nu}(\vec{s}', \vec{s}) I_{\nu}(\vec{s}') d\Omega \quad , \quad (15:8)$$

Combining (15:2) and (15:8), we get, for the scattering contribution to the emission coefficient $j_{\nu}^{(s)}$

$$j_{\nu}^{(s)} = \frac{\sigma_{\nu}}{\rho} \int p_{\nu}(\vec{s}', \vec{s}) I_{\nu}(\vec{s}') \frac{d\Omega'}{4\pi} \quad . \quad (15:9)$$

where ρ is the mass density of the emitting volume.

An atmosphere for which $j_{\nu} = j_{\nu}^{(s)}$ is called a scattering atmosphere. The converse to this is the situation in which the emission occurs by energy conversion from the matter field to the radiation field. To study this case, the assumption of local thermodynamic equilibrium (LTE) allows familiar ideas of black body radiation to be applied (Goody, 1964). Here Kirchhoff's law is valid and the ratio of coefficients of absorption and emission is a universal function of temperature T and frequency, $B_{\nu}(T)$, where

$$\frac{j_{\nu}}{(\alpha_{\nu}/\rho)} = B_{\nu}(T) \quad , \quad (15:10)$$

The Planck function, $B_{\nu}(T)$, is given by

$$B_{\nu}(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT) - 1} \quad , \quad (15:11)$$

where h is Planck's constant, k is the Boltzmann constant, and c is the speed of light. As discussed by Goody (1964), conditions of LTE are well satisfied below 50 km for both microwave and infrared radiation.

We can now derive the time-independent equation of transfer (for an unpolarized medium) by requiring that the increase in energy in a given direction in a small volume equal the difference between

total emission and total extinction

$$\begin{aligned} \frac{dE_\nu}{d\nu} = & -(\alpha_\nu + \sigma_\nu) I_\nu dA dr d\Omega dt + \alpha_\nu B_\nu dV d\Omega dt \\ & + \frac{\sigma_\nu}{4\pi} \int p_\nu(\vec{s}', \vec{s}) I_\nu(\vec{s}') d\Omega' dV d\Omega dt \end{aligned} \quad (15:12)$$

or

$$\begin{aligned} \frac{dI_\nu}{dr} = & -(\alpha_\nu + \sigma_\nu) I_\nu + \alpha_\nu B_\nu \\ & + \frac{\sigma_\nu}{4\pi} \int p_\nu(\vec{s}', \vec{s}) I_\nu(\vec{s}') d\Omega' \end{aligned} \quad (15:13)$$

This equation is frequently expressed in a slightly different form by introducing the optical distance between points P_1 and P_2 , $\tau_\nu(P_1, P_2)$, where

$$\tau_\nu(P_1, P_2) = \int_{P_2}^{P_1} (\alpha_\nu + \sigma_\nu) dr \quad (15:14)$$

As defined above, the non-negative quantity, τ_ν , is symmetric in its arguments. In terms of this variable, (15:13) becomes

$$\frac{dI_\nu}{d\tau_\nu} = -I_\nu + (1-f_\nu) B_\nu + \frac{f_\nu}{4\pi} \int p_\nu(\vec{s}', \vec{s}) I_\nu(\vec{s}') d\Omega' \quad (15:15)$$

$$= -I_\nu + J_\nu \quad (15:16)$$

where

$$f_\nu \equiv \sigma_\nu / (\alpha_\nu + \sigma_\nu) \quad (15:17)$$

and

$$J_\nu(\vec{s}) = (1-f_\nu) B_\nu + \frac{f_\nu}{4\pi} \int p_\nu(\vec{s}', \vec{s}) I_\nu(\vec{s}') d\Omega' \quad (15:18)$$

For an atmosphere in LTE, $f_\nu = 0$, while for a scattering atmosphere, the source function, J_ν , is determined by $f_\nu = 1$. The *formal* solution to (15:16) may be shown to be (see (F15.3))

$$\begin{aligned} I_\nu(P, \vec{s}) = & I_\nu(P'', \vec{s}) \exp[-\tau_\nu(P'', P)] \\ & + \int_0^{\tau_\nu(P'', P)} J_\nu(P', \vec{s}) \exp[-\tau_\nu(P'', P)] d\tau_\nu \end{aligned} \quad (15:19)$$

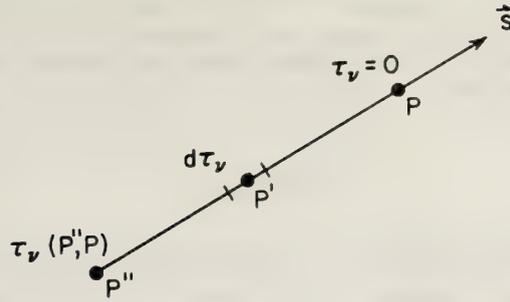


Figure 15.3 Integration variables in solution to radiative transfer equation.

The physical meaning of (15:19) is clear: the intensity of emission at any point P in direction \vec{s} is the resultant of emission at all anterior points, P', reduced by the factor $e^{-\tau_\nu(P',P)}$ to account for extinction by the intervening medium. Equation (15:19) is only a formal solution because in general, J_ν depends on the intensity in all directions and this is unknown. However, when $J_\nu = B_\nu$, the formal solution is the exact solution and is the basic equation in a multitude of atmospheric problems. Solutions of (15:19) for a scattering atmosphere, for various types of phase functions, $p_\nu(\vec{s}, \vec{s})$ are discussed by Chandrasekhar (1960). Analytic solutions to this equation when both scattering and thermal emission are present, i.e., when $0 < f_\nu < 1$, are difficult if not impossible to obtain.

For altitudes below about 50 km, molecular scattering is negligible and the transfer equation with $J_\nu = B_\nu$ is valid for a clear atmosphere. However, when clouds are present, their scattering properties must be taken into account for infrared radiation. For many, but not all, clouds, scattering of microwave energy is much less than absorption.

When scattering and emission are both important, numerical techniques, such as finite-difference methods (Volchok and Chernyak, 1968) have given useful results for a cloud layer above a partially reflecting earth surface.

To correctly formulate scattering problems, polarization of the radiation field must be taken into account. As discussed by Chandrasekhar (1960), this can be done by introducing the four Stokes parameters to completely describe the radiation and then by deriving, in a similar manner to the scalar unpolarized problem, a matrix equation of transfer. A treatment of this equation is beyond the scope of this chapter.

Above, we have considered the transfer of radiation through a continuous medium by defining volume absorption, scattering and emission coefficients. To describe the radiation properties of natural surfaces, it is convenient to relate these properties to those of a black body. Thus, for example, the spectral power density, emitted into a solid angle $d\Omega$ in a direction \vec{s} by an area dA with normal \vec{n} is

$$\frac{dE_\nu}{d\nu dt} = \epsilon_\nu B_\nu dA d\Omega \vec{n} \cdot \vec{s}, \tag{15:20}$$

where, in general, the emissivity ϵ_ν is a function of direction. The corresponding ability of a body to reflect or absorb energy is expressed in terms of its reflectivity, r_ν , or absorptivity, a_ν , and

$$a_\nu = \epsilon_\nu = 1 - r_\nu \tag{15:21}$$

To completely specify the reflection properties of a surface, the angular distribution of reflected energy in the upper hemisphere must be given. A black surface has $a_\nu = \epsilon_\nu = 1$.

As discussed above, for an atmosphere in LTE, the emission from a volume or a surface can be related to the universal function of frequency and temperature $B_\nu(T)$. It is common, especially in microwave radiometry, to describe a radiation field in terms of an equivalent black body temperature, or brightness temperature, $T_{b\nu}$, such that

$$I_\nu = B_\nu(T_{b\nu}) \quad (15:22)$$

Since the intensity is a function of frequency and viewing direction, the brightness temperature will vary accordingly. Furthermore, for the microwave region, in which $h\nu \ll kT$, the Rayleigh-Jeans approximation to (15:22) yields

$$I_\nu = 2 \frac{\nu^2}{c^2} k T_{b\nu} \quad (15:23)$$

Combining (15:22) and (15:19) yields

$$T_{b\nu}(P, \vec{s}) = T_{b\nu}(P'', \vec{s}) \exp(-\tau_\nu(P'', P)) + \int_0^{\tau_\nu(P'', P)} T_{b\nu}(P', \vec{s}) \exp[-\tau_\nu(P', P)] d\tau_\nu \quad (15:24)$$

Calculations of clear air microwave brightness looking downward from a satellite over earth and looking upward from earth are shown in (F15.4). Absorption and emission profiles were determined from atmospheric temperature and density profiles given in the 1962 US Standard atmosphere. Note the dependence of the total upward emission on surface emissivity for the fixed surface temperatures assumed in these calculations. The absorption coefficients for the 22.235 GHz water vapor line and the 60 GHz oxygen band were calculated from the Van Vleck-Weisskopf line shape (Van Vleck and Weisskopf, 1945). Above 50 GHz, only calculations at 53.5, 54.5 and 55.5 GHz are shown. Examination of plots such as (F15.4) indicate promising spectral regions for remote sensing. For example, microwave measurements at frequencies below about 20 GHz are insensitive to the clear atmosphere but are strongly affected by surface conditions. In contrast, measurements around 22.235 GHz or in the 60 GHz region are potential sources of information about water vapor and temperature structure. Similar considerations in the infrared have shown:

- a. measurements in the 4.3 and 15 μm bands of CO_2 can yield information on vertical temperature structure (Wark and Fleming, 1966);
- b. measurements in the 6.3 μm band of H_2O and in the 9.6 μm band of O_3 provide information on the vertical distribution of H_2O and O_3 (Conrath, 1969; Shafrin, 1970); and
- c. measurements in spectral "window" regions, such as 10.5 to 12.5 μm can yield information on surface temperature (Rao, et al., 1971).

Above, we have discussed the generation and transfer of microwave and infrared electromagnetic noise energy in the atmosphere. To apply measurements of this energy to remote sensing requires a thorough understanding of instrumental effects on spectral resolution, spatial resolution, and angular resolution, and of the contamination of the atmospheric signal with instrumental noise. In the next section, we discuss commonly used radiometric measurement techniques.

15.2 Measurement Techniques

In this section we consider methods for the measurement of radiation in two frequency regions: the microwave, defined here as extending from about 1 to 300 GHz; and the infrared, extending, in terms of

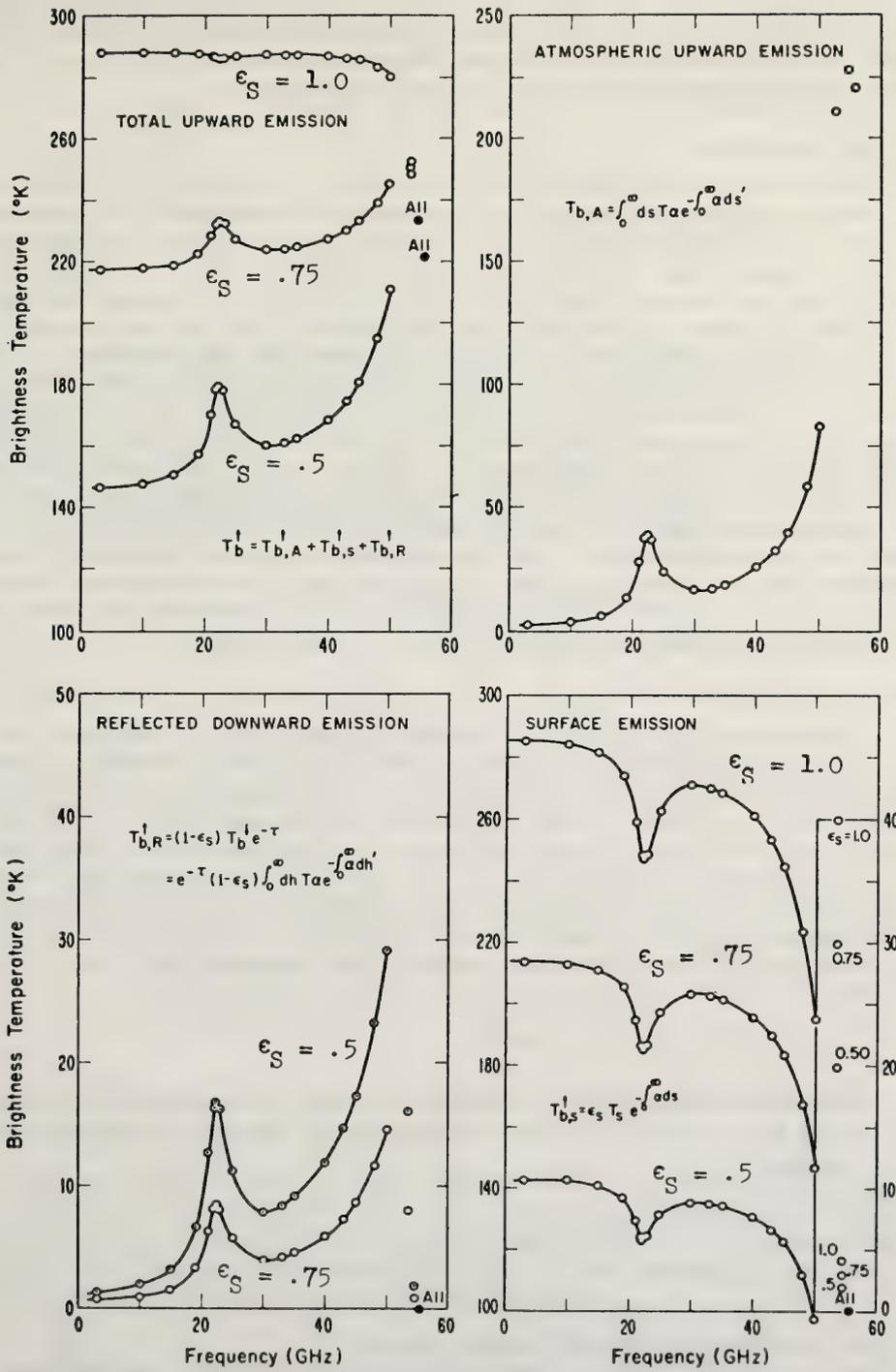


Figure 15.4 Nadir brightness temperatures for clear atmosphere.

wavelength from about 0.7 to 1000 μm . The radiometer is an instrument which measures radiant power incident on an aperture. Since all radiometers measure the same quantity (perhaps expressed in different units), it follows that all radiometers must be at least somewhat similar. We shall begin, therefore, by examining some similar features of microwave and infrared radiometers, and then proceed to examine specific differences in instruments and techniques in the two frequency regions.

15.2.1 The Basic Radiometer

The components of the basic radiometer are shown in (F15.5); they consist of an energy collecting aperture, a detector, amplifying electronic circuitry, and a device for recording the radiometer output. In the microwave case the energy collecting aperture is an antenna while in the infrared radiometer, the aperture might be a lens or mirror. The detector in the infrared radiometer directly converts the impinging radiation into some easily measured quantity, e.g., a voltage or a change in resistance. The microwave detector may also be operated as a direct detector provided sufficient amplification can be achieved at the signal frequency to yield a usable signal to noise ratio from the detector. Direct detection at microwave frequencies is generally limited to applications where bandwidths of the order of several GHz may be employed. In most atmospheric remote sensing applications, the bandwidths of interest are relatively narrow, say 100 MHz. As a result, direct detection is not feasible and the detector is in reality a superheterodyne receiving system in which the signal frequency band is translated to an intermediate frequency band by mixing with a strong local oscillator source. The intermediate frequency is located in a frequency region where high gain and the desired bandwidth may be easily achieved.

Since the output recording scheme often depends upon the particular application, it will not be discussed in detail here. In microwave radiometry the output is usually an electrical analog of the incident radiation and thus may be recorded using a variety of techniques. Infrared radiometry outputs may also be electronic in nature, but photographic techniques are employed in some instruments.

15.2.2 Microwave Radiometers

In our discussion of the techniques of microwave radiometry, we will concentrate upon those systems which employ the superheterodyne receiver since this type of system is most likely to be used in atmospheric remote sensing. We shall also assume that the electromagnetic radiation of interest is incoherent and has the properties of noise. As shown in the preceding section, the Rayleigh-Jeans approximation to the Planck function can usually be made in the microwave region so that the brightness of an emitter is linearly related to temperature. For these reasons, plus the fact (as we discuss below) the noise power collected by an antenna is equivalent to the noise temperature generated in a resistance, it is customary to calibrate a microwave radiometer in units of temperature.

The noise power, w , per unit bandwidth available at the terminals of any resistance is given by Nyquist (1928) as

$$w = kT, \quad (15:25)$$

where k is Boltzmann's constant and T is the absolute temperature of the resistance. In the following argument, we shall show that the noise power per unit bandwidth collected by a microwave antenna is given in a similar manner by

$$w = kT_a, \quad (15:25a)$$

where T_a is the temperature of the antenna "radiation resistance."

Consider an antenna connected to a resistance by a matched transmission line. Both antenna and resistance are enclosed in perfectly absorbing surroundings at absolute temperature T_a as shown (F15.6). We assume that the transmission line and antenna are lossless and that all energy collected by the antenna is absorbed by the resistance. Since the system is in thermal equilibrium, it follows that all energy generated in the resistance must be radiated by the antenna. Therefore, the noise power collected by an

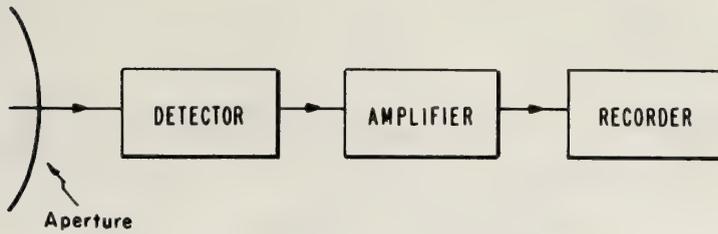


Figure 15.5 *Basic radiometer circuit.*

antenna is equivalent to the noise power generated in the antenna radiation resistance. The radiation resistance is, in fact, a fictitious resistance that represents the antenna. This argument leads to the concept of “antenna temperature” which is defined as the temperature of the antenna radiation resistance required to produce the observed noise power available at the antenna terminals.

For the real situation where the antenna is not immersed in a blackbody enclosure, the antenna noise temperature, T_a , is given by the weighted average of the noise temperature distribution surrounding the antenna; this will be discussed in greater detail in 15.2.2.1, Antenna Effects.

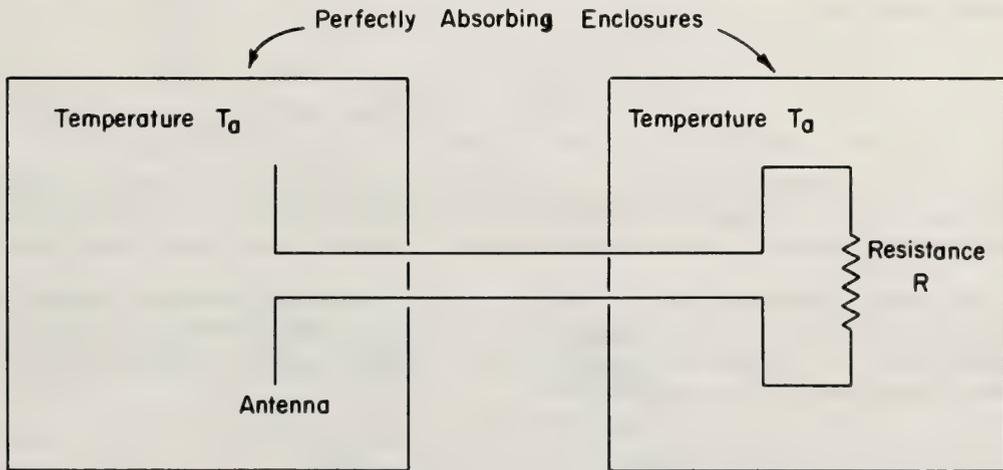


Figure 15.6 *System used to explain concept of antenna noise temperature.*

A typical microwave radiometer is shown in block form in (F15.7); the configuration shown is the “total-power radiometer” since the antenna remains connected to the radiometer during operation. The sensitivity of the total-power radiometer is defined as the minimum detectable temperature change referred to some convenient reference plane, normally the radiometer input terminal. The sensitivity ΔT_{\min} is determined by the system characteristics according to Kraus (1966)

$$\Delta T_{\min} = C \frac{(T_r + T_a)}{\sqrt{B\tau}} \quad (15:26)$$

where T_r is the effective noise temperature of the radiometer, T_a is the antenna temperature, B is the pre-detection bandwidth, τ is the integration time, and C is a detectability constant.

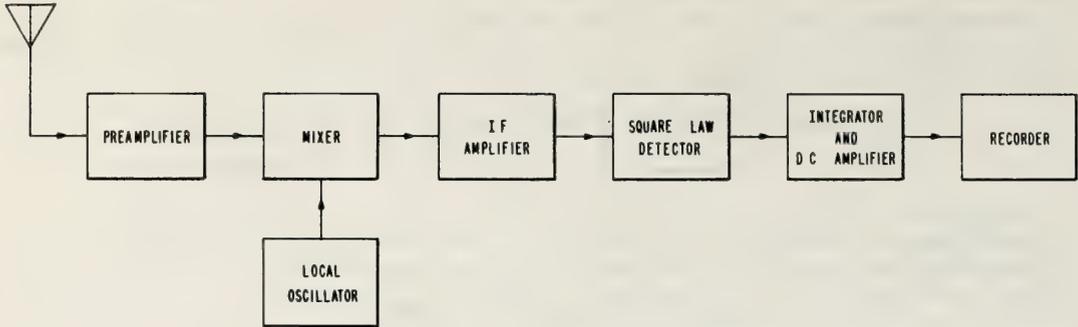


Figure 15.7 Total power microwave radiometer.

The effective radiometer noise temperature, T_R , is a measure of the internally generated receiver noise caused by resistive losses in transmission lines, and noise generated in the preamplifier, mixer, and other components. Since the antenna temperature, T_a , and T_R have similar characteristics, they add to yield a system noise temperature that is a fundamental limitation to the achievable radiometer sensitivity. The voltage output, V_o , of the total-power radiometer is proportional to the noise power incident upon the square-law detector, i.e.,

$$V_o = C_1 k(T_a + T_R) B G_o \quad (15:27)$$

where C_1 is a proportionality constant, G_o is the predetection gain, and the other terms are defined in (15:26). Since the detector is unable to distinguish between a change in T_a or T_R , the total-power system must employ some method of monitoring the magnitude of T_R in order to make accurate measurements of T_a .

The classical radiometer equation (15:26) is strictly valid only if the predetection radiometer gain remains constant. In practice (15:26) must be modified to account for the inevitable gain variations. Since gain variations are not correlated with radiometer output fluctuations caused by system noise, the actual sensitivity equation may be written as (Kraus, 1966)

$$\Delta T_{\min} = C(T_R + T_a) \sqrt{\frac{1}{B\tau} + \left(\frac{\Delta G}{G_o}\right)^2} \quad (15:28)$$

where ΔG represents the effective value of pre-detection gain fluctuations, G_o is the total pre-detection gain, and the other terms are as previously defined. The "switching radiometer" first described by Dicke (1946) was developed to reduce the effect of low frequency gain variations (see F15.8). The radiometer input is switched between the antenna and the reference termination at a rate which is sufficiently removed from the spectrum of gain variations that their effects may be largely ignored. The switching frequency is usually in the range from 30 Hz to 2 KHz. The reference termination may be either an actual waveguide or transmission line load or, in some cases, the thermal emission from the modulator in its highly attenuating position. An example of the latter type is the rotary waveguide attenuator. The physical temperature of the reference termination should be controlled to maintain a constant emission temperature. The radiometer output in the switched mode is proportional to the difference in the noise temperature of the antenna and the reference termination, i.e.,

$$V_o = C_2 k(T_a - T_r) B G_o \quad (15:29)$$

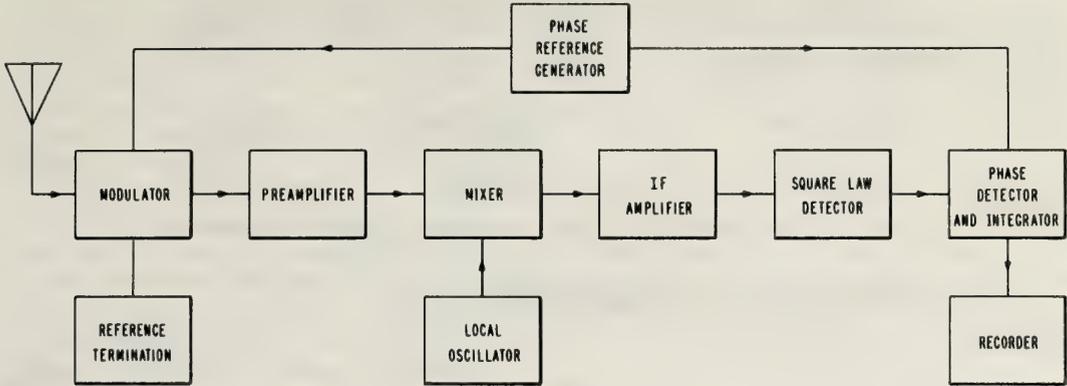


Figure 15.8. Switched microwave radiometer.

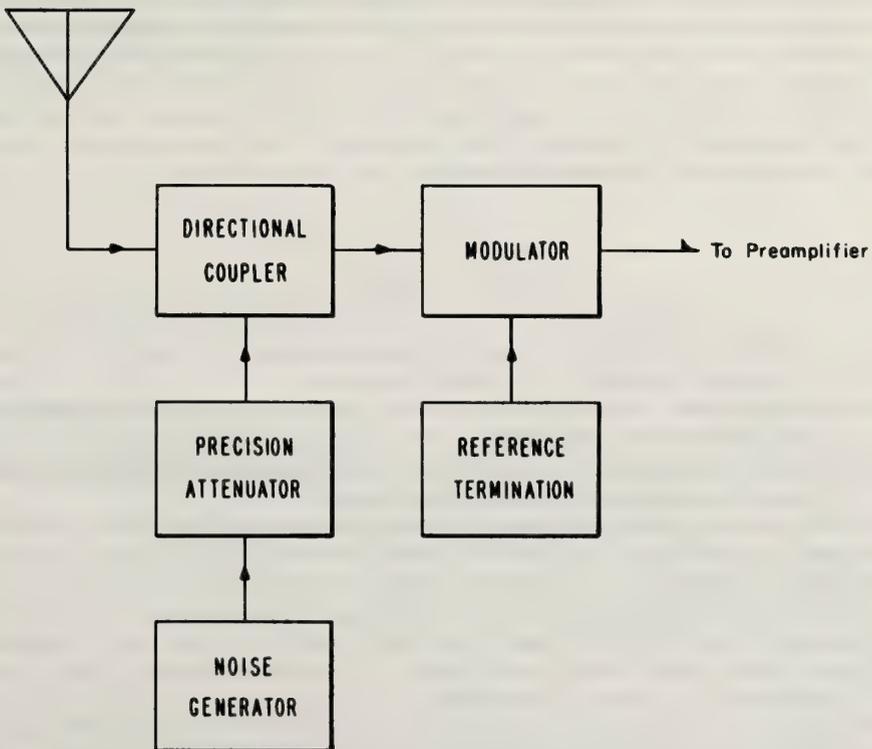


Figure 15.9 Microwave radiometer with noise adding circuitry.

where C_2 is a proportionality constant, k is Boltzmann's constant, T_a and T_t are the antenna and reference termination noise temperatures, respectively, B is the predetection bandwidth, and G_O is the predetection gain. Since only differences in antenna and reference temperatures are detected in the switched radiometer, continuous monitoring of the effective radiometer noise temperature is not required and the equipment is considerably less complicated than for the total-power system.

Table 15.1. Modulation constant M for various switching waveforms.

Modulation Waveform	Detection Waveform	M
Square Wave	Square Wave	2.0
Square Wave	Sine Wave	2.22
Sine Wave	Sine Wave	2.83

Gain variations ΔG now cause an output fluctuation given by

$$\Delta V_o = C_3(T_a - T_t) \frac{\Delta G}{G_o} \quad (15:30)$$

where C_3 is a proportionality constant and the other quantities are defined above. Thus, if T_a and T_t are made equal, gain fluctuations will not affect the radiometer output. Since the antenna temperature is colder than the reference source in most applications, the usual method of equalizing T_a and T_t is to add noise to the antenna signal; the method is shown in (F15:9). Inasmuch as the noise adding circuitry is in effect a calibration device, and indeed in some systems is used only in radiometer calibration, we shall henceforth call it a calibration circuit. Noise sources in the microwave region are either gas discharge tubes or diode noise generators. The magnitude of the added noise is varied by means of the precision variable attenuator. The excess noise temperature, T_e , added through the noise circuitry is

$$T_e = \frac{(T_g - T_o)}{C_d L_{at}} \quad (15:31)$$

where T_g = the effective noise temperature of the noise generator,
 T_o = the ambient temperature of the calibration circuitry,
 C_d = coupling factor of the directional coupler, defined as the ratio of the input power to the directional coupler to the output power from the auxiliary arm, and
 L_{at} = resistive loss factor (ratio of input power to output power) of the precision attenuator and connecting wave-guide in the calibration circuitry.

The noise adding scheme offers a convenient and accurate means of relating antenna temperature to the amount of noise required to zero the radiometer output. Assuming that the terms in (15:30) and the loss factor and physical temperature of the waveguide before the modulator are known, T_a can be related to the setting of the variable attenuator in the calibration circuit. This is the basis for the so-called "null balance" operating mode in which the value of attenuation required to obtain zero radiometer output is recorded rather than the phase detector output voltage. The radiometer output may be set to zero either manually or automatically by means of a servo system.

The sensitivity of the switched radiometer is poorer than that of the total-power radiometer by a factor of at least 2; the theoretical sensitivity depends upon the switching waveform as well as the reference waveform applied to the phase detector. Assuming a balanced radiometer (i.e., $T_a = T_t$), the theoretical sensitivity is given by

$$\Delta T_{\min} = MC \frac{(T_a + T_r)}{\sqrt{B\tau}} \quad (15:32)$$

where M is a constant that depends upon the modulation used and other quantities are as defined for (15:25). The following table after Kraus (1966) lists M for several modulation and phase detection waveforms.

One method of radiometer calibration has been mentioned above in our discussion of null-balanced operation. In many applications it may neither be desirable nor feasible to operate in the null balanced mode, e.g., if a servo system is not available, or for wideband multi-spectral measurements where T_a varies with frequency. In this event, calibration may be accomplished by replacing the antenna with a source of known noise temperature, e.g., a liquid nitrogen cooled termination, or a heated termination. Varying amounts of noise are then added via the calibration circuit to provide incremental calibration of the radiometer output.

In order to calculate the noise added to the system, we must know not only the resistive loss factors of the precision attenuator and connecting waveguide sections, but the physical temperature of these losses as well. Most radiometric systems control the temperature of critical components and in addition monitor and sense their physical temperature by means of thermistor probes. The waveguide loss factors can be measured in the laboratory. By careful measurement and analysis, it is possible to measure antenna temperature with an accuracy of the order of 1 or 2 K. We now turn to the problem of converting antenna temperature to brightness temperature.

15.2.2.1 Antenna Effects

The quantity required as input data for the remote sensing inversion schemes discussed in 15.3 is the brightness temperature in a given direction. However, the quantity measured by the radiometer is the net antenna temperature determined by the weighted average of the brightness temperature distribution in the sphere surrounding the antenna, i.e.,

$$T_a = \frac{1}{4\pi} \int_{4\pi} G(\Omega) T_b(\Omega) d\Omega \quad (15:33)$$

where $G(\Omega)$ = the antenna gain in the direction of the solid angle, Ω , and $T_b(\Omega)$ = the brightness temperature from this direction. Antenna power patterns are determined by the diffraction pattern of the antenna aperture. Directional antennas have patterns that peak in a primary direction (i.e., the main beam) with many secondary, less intense, maxima.

For convenience of analysis, (15:33) may be rewritten as the sum of two integrals: the first shall be taken over the main beam region of the antenna pattern while the second is taken over the remainder of the pattern. Solving for the brightness temperature integrated over the main beam we obtain

$$T_{bm} \equiv \int_{\text{main}} G(\Omega) T_b(\Omega) \frac{d\Omega}{4\pi} = T_a - \int_{\text{remainder}} G(\Omega) T_b(\Omega) \frac{d\Omega}{4\pi} \quad (15:34)$$

For a radiometric system pointing toward the earth from space, the T_a to T_b conversion problem is fairly simple since the brightness temperature distribution in the back and side lobe regions of the antenna is the radiation from cold space. In the case of a ground-based radiometer where the back and side lobes are illuminated by the atmosphere and ground and where the antenna pointing angle may be variable, the conversion process is considerably more involved. The difficulty of the conversion problem depends in large part upon the performance of the antenna; the important figure of merit for our purpose is the antenna beam efficiency: the ratio of power contained in the main beam to the total power collected by the antenna. The difficulty of converting T_a to T_b is clearly inversely proportional to the beam efficiency. Many remote sensing applications require high angular resolution so that the large aperture-wavelength ratios which are necessary inherently result in fairly high beam efficiencies. Another method that may be used to maximize the beam efficiency is to employ reflector antennas which minimize direct radiation into

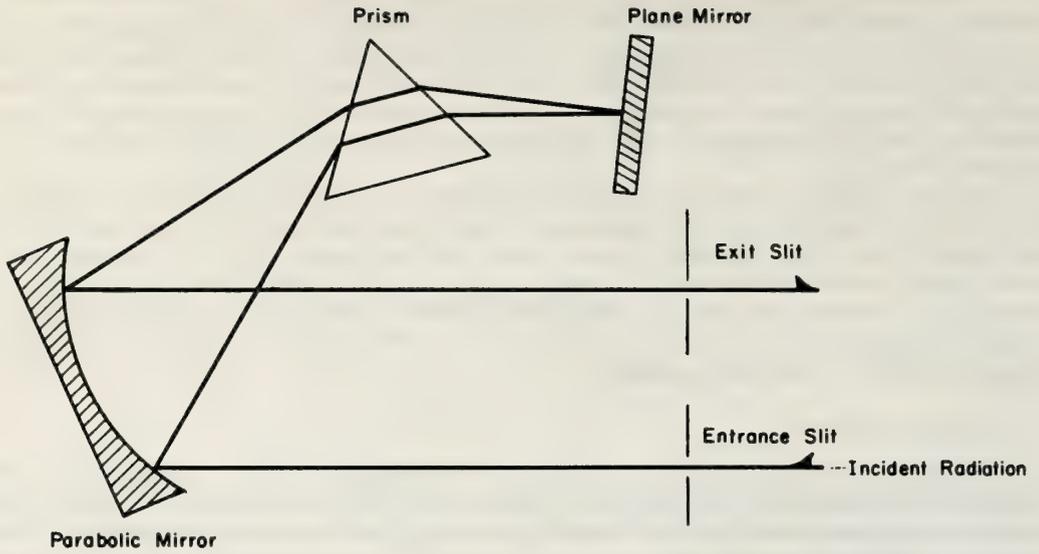


Figure 15.10. *Single pass monochromator.*

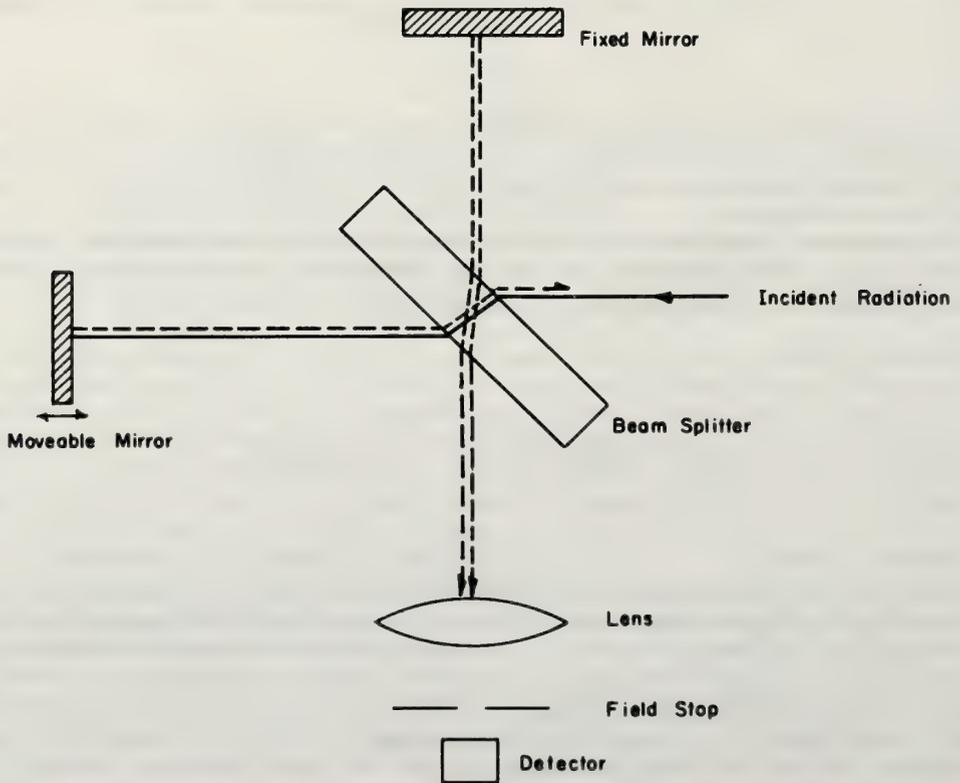


Figure 15.11 *Michelson interferometer.*

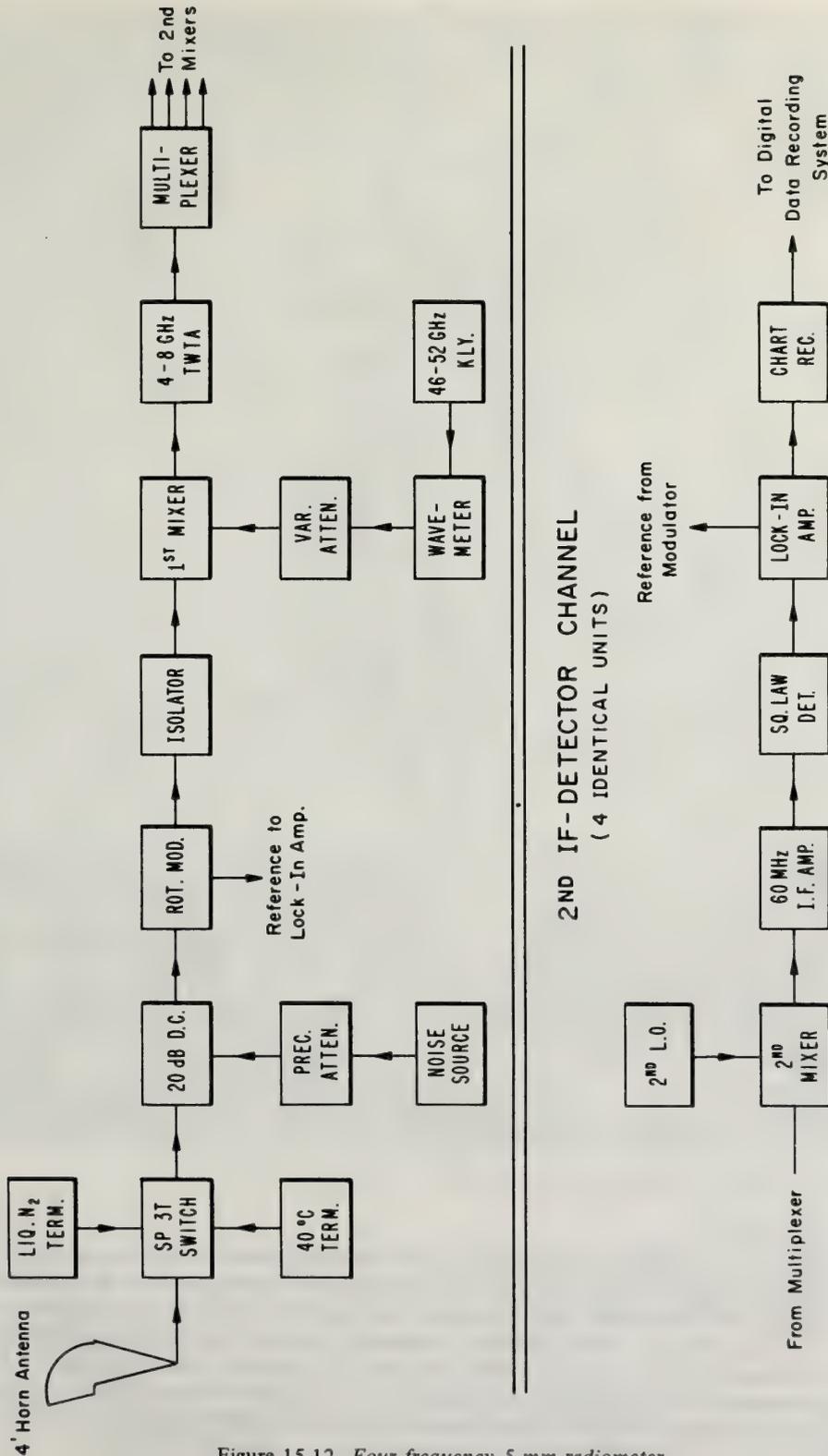


Figure 15.12 Four frequency 5 mm radiometer.

the antenna feed; an example of this type of arrangement is the hog-horn antenna (see F15.13). Of course, the optimum antenna configuration must be based upon the particular application.

In practice, the antenna temperature brightness-temperature characteristic for a particular antenna system is estimated by numerically integrating (15:33). Since the antenna pattern is usually measured only in the two principal planes, it is often necessary to make some assumption about the antenna response in intermediate planes. The distribution of brightness temperature can usually be very closely approximated from knowledge of the atmospheric absorption (and hence emission) properties at the frequency of operation and the variation of atmospheric temperature with height. Suitable assumptions regarding the emissivity of the ground must also be made. If the radiometric system is to be pointed at different angles, (15:33) should be evaluated at all angles likely to be used. The results of this study will determine the actual technique required to convert from antenna to brightness temperature. Provided the beam efficiency is fairly high, it may be feasible to make the conversion at each angle using a simple correction factor based upon a single evaluation of (15:33).

Variations in the basic microwave radiometer are generally made to achieve either spatial resolution of radiation sources or to measure the variation of radiation as a function of signal frequency. Some radiometers may obtain resolution in both quantities simultaneously. Spectral resolution can be considered from large and small-scale viewpoints. For example, large-scale spectral resolution would be the variation with signal frequency of the total power contained in the intermediate frequency (IF) pass-band. Spectral resolution on the small scale would be the variation of the power in relatively narrow bandwidths within the IF pass-band at a fixed signal frequency.

Large scale resolution in spectral power can be obtained in a variety of ways. In the simplest case, the local oscillator frequency may be changed in a systematic manner while the radiometer output is recorded. In more sophisticated systems, several radiometers may be operated simultaneously from a common antenna through suitable coupling networks. Where measurements are to be made at widely spaced frequencies, two or more basic radiometers may be operated from a single antenna using suitable channel separation filters working at the signal frequency. An alternate method is to perform the separation into individual frequency bands at the intermediate frequency; the division may be accomplished through either a second conversion to a new intermediate frequency of narrow bandwidth or by active or passive filtering by means of tuned circuits. The latter method is most suitable when measurements are desired in relatively narrow frequency intervals within a fairly small IF bandwidth, e.g., 100MHz or so.

Provided the separation between individual frequency bands is not too great, calibration of multi-frequency radiometers can be performed in a manner similar to that described earlier for the single frequency radiometer. The major requirement is that the calibration noise source have a known output over all frequencies of operation; in general, this requirement limits use of a single calibration source to a given standard waveguide band when gas discharge tube calibration sources are used. Of course, waveguide resistive losses and transmission-reflection characteristics must be accurately known for all frequencies. A block diagram of a multi-frequency radiometer is shown in 15.3.

Angular resolution in the microwave region is determined by the size of the antenna aperture. Where resolution of radiation sources over an area is desired, the antenna main beam is scanned over the area either mechanically or electrically; the most common scanning method is the mechanically driven raster scan. The antenna temperature at each pointing angle is given by (15:33). When the antenna main beam efficiency is not high, determination of brightness temperature may be subject to serious error because of the variation of radiation into back and side lobes with movement of the antenna. It is essential, therefore, that high quality antennas be used in ground-based scanning applications.

15.2.3 Infrared Radiometers

In schematic form, the infrared (IR) radiometer is considerably less complicated than a microwave radiometer; thus the diagram shown in (F15.5) is a fairly adequate representation of a so-called "DC" radiometer. The IR "DC" radiometer does not employ switching between signal and a reference and is analogous to the total power microwave radiometer. By inserting a chopper between the input aperture and the detector we obtain the basic schematic of the more typical IR radiometer. Use of chopping, of course, is the exact counterpart of the Dicke microwave radiometer and is employed, as we shall point out later, for some of the same reasons. With these similar features in mind, we now examine some of the details of IR radiometers.



Figure 15.13 *Four frequency radiometer and conical horn antenna.*

Following the pattern used in our discussion of the microwave radiometer, we begin with detectors of IR radiation. However, in order to clarify our discussion, we shall first point out the customary division into sub-bands within the IR spectrum extending from 0.72 to 1000 μm . Within this spectrum three subdivisions are usually made:

Near IR	0.72 to 1.5 μm
Middle IR	1.5 to 5.6 μm
Far IR	5.6 to 1000 μm

These spectral divisions have been made somewhat arbitrarily as the result of detector and optical development (Hackforth, 1960). A spectral region is sometimes referred to in terms of the detector most widely used in the region. For example, the term "lead sulfide region" is fairly common and implies that part of the IR spectrum from about 1 to 3 μm . While we do not advocate this rather vague system of classification, the reader should be aware of the practice.

Infrared detectors are divided into two major types: thermal detectors and quantum detectors. Examples of thermal detectors are thermocouples, thermistors, bolometers, the Golay (pneumatic) cell and the relatively new pyroelectric detector. As implied by the name, thermal detectors experience some change in their characteristics, i.e., voltage output, resistance, gas pressure, electrostatic charge, as a result of heating caused by incident radiation. Thermal detectors are characterized by their uniform response over wide spectral intervals and their usefulness in the far IR region. Although operable over the entire IR spectrum, thermal detectors are less sensitive in the near-IR and middle-IR bands. Since they operate on the basis of self-heating, the response time of thermal detectors is longer than for quantum detectors and varies from 3 or 4 milliseconds to several seconds.

Quantum detectors operate by directly changing their characteristics through absorption of photons when irradiated. For example, absorption of photons may result in a change in conductivity of a semiconductor through electron excitation; this type of detector is the well known photoconductor. Other quantum detectors include the photoemissive type which emit electrons when irradiated, and photovoltaic detectors which generate a voltage between two conductors when irradiated. Because of the frequency dependent nature of the photon absorption process and resultant affinity for operation in relatively narrow spectral bands, quantum detectors are often called selective detectors. Photoemissive devices are used mainly in the visible and near IR bands. In general, photoconductive and photovoltaic IR detectors find greatest use in the middle IR region although by cooling some units to nearly 4K, operation may be extended to 30 or 40 μm (Wolfe, 1965). Since this class of detectors does not operate by self-heating, response times are typically 2 or 3 orders of magnitude faster than for thermal detectors.

A number of methods of quantitatively measuring the performance of IR detectors have evolved over the years. Since the manner in which the detector is used and the characteristics of the radiation of interest have such an important influence upon how the detector is calibrated, the use of a uniform figure of merit for intercomparison of detector performance has been slow to develop. The most common measures of performance in present use are listed below; a more comprehensive list of performance descriptors is given by (Wolfe, 1965).

Noise equivalent power – The noise equivalent power, NEP, is defined as that quantity of radiation that produces a rms detector output equal to the rms output caused by detector noise.

Detectivity – The detectivity, D , is defined as the signal to noise ratio per unit of input radiation. D is the reciprocal of the NEP.

Responsivity – The responsivity, R , is defined as the detector output per unit quantity of incident radiation. For electrical detectors, the units of responsivity might be volts/watt.

In these definitions we have assumed unit detector area and unit frequency interval. In practice these quantities must be stated in more specific terms. For example, calibration conditions, i.e., the measured detector spectral response, detector area, the modulation frequency, and the measurement bandwidth must be specified. If the detector performance is measured using a blackbody source, the relative spectral response of the detector should be stated. When this information is combined with knowledge of the blackbody source temperature and theoretical spectrum, the detector's absolute spectral response may be determined in a straightforward manner.

Jones (1959) advocates use of a specific detectivity, D^* , which removes the areal and amplifier bandwidth dependencies from the detectivity, D . The specific detectivity is defined as

$$D^*(\lambda, f_0) = DA_d^{1/2} (\Delta f)^{1/2} \quad (15:35)$$

where D = detectivity in watts^{-1} ,
 A_d = detector area in cm^{-1} ,
 Δf = measurement bandwidth in Hz,
 λ = wavelength of incident radiation, and
 f_0 = chopper frequency.

The specific detectivity simplifies comparison of detectors having different areas but should be used with caution since the expression assumes a uniform response over the surface area of the detector. Since the response of many detectors does, in fact, vary across their surface, it is possible that comparisons based upon D^* may be subject to error. It should also be borne in mind that D^* may not be proportional to $A_d^{1/2}$ in some detectors.

Table (15.2) lists the performance of several detectors used in the IR spectrum. The table is not intended to be a comprehensive list but rather is intended to illustrate performance differences between the two major types of IR detectors. The table is based upon information contained in Jamieson, et al. (1963).

Table 15.2. Performance of typical infrared detectors.

Detector Type	Operating Temperature (K)	Wavelength of Maximum Response (μm)	D^* $\lambda_p, 900$ ($\text{cm Hz}^{1/2}\text{w}^{-1}$)	Response Time (s)
Lead Sulfide	295	2.5	10^{11}	2×10^{-4}
Lead Sulfide	195	2.7	3×10^{11}	2×10^{-3}
Lead Sulfide	77	3.2	1.2×10^{11}	3×10^{-3}
Thermistor	295	Flat	8×10^9	2×10^{-4}
Thermocouple	295	Flat	2×10^8	1.5×10^{-2}
Golay Cell	295	Flat	2×10^9	1.5×10^{-2}
Pyroelectric	295	Flat	5×10^8	10^{-3}

The ultimate limit to IR detector performance is, of course, noise. The source of the limiting noise may be external to the detector, i.e., background noise, or internal to the detector. Some internal noise sources include the usual Johnson noise, shot noise, current noise, etc. An additional type of noise occurring in quantum detectors is photon noise, i.e., noise generated by fluctuations in the rate of photon absorption by the detector. Many quantum detectors are capable of performance limited only by radiation noise from the background. The basic limit to the performance of thermal detectors, however, is Johnson noise.

In a manner analogous to the Dicke radiometer used in the microwave region, infrared radiometers employ chopping (modulation) of the IR signal to reduce gain instabilities caused by drift in d.c. amplifiers following the detector as well as to provide a controlled reference radiation level. Chopping may also be used for spatial filtering and/or background discrimination. In atmospheric remote sensing, or when measuring radiation from extended rather than point sources, the use of chopping for spatial filtering is not common. We shall therefore limit our present discussion to the use of chopping for improved radiometer stability and to generate a reference level. Spatial discrimination by means of chopping will be discussed later.

The reference radiation level is the incident radiation that produces a zero level in the radiometer output. The chopper used in the IR radiometer may generate the reference level if the emissivity of the chopper blades is quite high; the usual procedure is to blacken the chopper blades to obtain a high emission. The self-emission method finds greatest application in the near IR spectral region where thermal radiation from instrument components is not a serious problem. At longer wavelengths, the method may not be satisfactory because non-uniformity in chopper blade temperature and emissivity may make determination of the exact reference radiation level extremely difficult.

A second technique is to make the chopper blades highly reflecting and to utilize an optical arrangement that causes the chopper blades to reflect the radiation from a blackbody source onto the IR detector. A useful feature of this method is that the temperature of the blackbody source may be easily controlled to vary the reference level; thus the radiometer can be operated in a null-balance mode. Extreme caution must be taken to insure that only radiation from the blackbody is reflected into the detector to insure that the reference level is indeed determined by the temperature of the blackbody. It may also be necessary to compensate for differences in solid angles through which the signal source and the reference source illuminate the detector.

In most IR radiometers the chopper is located between the collecting aperture and the detector. If a field stop is used in front of the detector, the chopper is located before the stop so that radiation from chopper blades or the collecting aperture arrives at the detector only through the field stop. Ideally, filters should be located between the chopper and detector so that any radiation from the filter will not be chopped and thus will not contribute to the signal.

The problem of stray radiation from filters and other radiometer components is somewhat similar to that of converting antenna temperature to brightness temperature in microwave radiometry. While the field of view is much more limited in the IR region than at microwaves so that radiation into side lobe regions can usually be ignored, thermal radiation from components can cause errors in the radiant quantity being measured. Since corrections for stray radiation by calculation are nearly impossible, extremely good optical design must be employed to prevent undesired thermal radiation from contaminating the signal of interest.

In our earlier discussion we considered IR detector performance chiefly from the point of view of comparing the merits of different types of detectors for use in a given application. We now turn to the ultimate system problem: the calibration of the total instrument. Realizing that instrument calibration depends not only upon how the device is constructed and is operated but also upon the radiant quantity to be measured, we shall be forced to limit our discussion to general principles and considerations.

There are three basic characteristics of any radiometer that must be known to interpret the results of radiation measurements: the responsivity, the reference level to which the incident radiation is compared, and the detectivity. In some applications where the signal-to-noise ratio is considerably greater than unity, it may not be necessary to know the detectivity extremely accurately; in the majority of applications, however, the responsivity and the reference radiation level must be accurately known. We have previously discussed techniques for generating the reference radiation level and shall assume that the reference is known unambiguously. In the following sections we emphasize measurement of the responsivity and briefly consider the instrument detectivity.

The responsivity of a radiometer is the output per unit quantity of incident radiation as a function of the following characteristics: power level, wavelength, spatial variations, temporal fluctuations, and polarization. The unit in which the radiometer is calibrated depends upon the application. For example, if the radiometer is to measure radiation from a near point source, the appropriate unit is power per unit solid angle. If isotropic extended sources are being measured, the appropriate unit is radiance, or power per unit solid angle per unit area. Other radiometric quantities can also be used; some of the more common radiometric units are listed in (T15.3) after Bell (1959). The radiance is sometimes expressed in terms of temperature, although to be correct, the source temperature defines the radiance for blackbodies only; for this reason, many workers discourage the practice of calibrating IR radiometers in temperature units. However, in applications such as atmospheric remote sensing in which we are interested in non-coherent thermal radiation, calibration in temperature units is proper. It should be emphasized that careful attention must be given to any spectral characteristic that might affect the temperature calibration. When possible, the calibration source should have the same spectral distribution as the signal source.

With this brief introduction, we shall examine calibration of the instrument in terms of the specific parameters listed above. Where appropriate, parallels between IR and microwave radiometric methods will be indicated.

Wavelength

The spectral responsivity of IR radiometers should be determined by direct measurement as opposed to calculation from piecewise knowledge of the spectral characteristics of the detector and the optical system including any filters. Spectral responsivity should be measured using a variable wavelength source of high spectral purity; the output of the calibration source as a function of wavelength, of course, must be

known. The output of the source, probably a monochromator, is usually measured with a reference detector whose responsivity is relatively independent of wavelength. Therefore, the radiometer is actually calibrated relative to the reference detector. In general it is not practical to obtain absolute spectral calibrations when using a monochromator since the beam geometry will not be of the required area and solid angle. Absolute calibrations are customarily made at a few wavelengths employing a blackbody source and filters to limit the bandwidth of the radiation incident upon the radiometer. The possibility that instrument spectral responses exist outside the region of maximum responsivity should be investigated.

At microwave frequencies, the spectral response is determined by the local oscillator frequency and the IF amplifier characteristics. Because of relation 15:25, calibration of microwave spectral responsivity is considerably less complicated than in the IR case; in fact, at a given local oscillator frequency, the effect of the IF passband is automatically taken into account when using the noise source calibration technique. Spurious frequency responses must be considered for microwave as well as for IR radiometers.

Power Level

In applications where the radiant power level is variable, the detector output versus incident power level must be determined. Ideally, the calibration should be made over the full range from threshold of detectability to detector saturation. It is extremely important when making linearity tests of IR radiometers that the level of the incident radiation is varied in such a manner that the instrument output changes unambiguously, e.g., not as the result of a change in the size of the image in the detector. Perhaps the most satisfactory way of varying the power of the radiation incident upon the radiometer is to employ a collimator and precision absorbing filters which remove known fractions from the incident beam. Any variation in spatial responsivity must be taken into account when using this technique.

Variation of the power level when calibrating the microwave radiometer has already been discussed.

Spatial Responsivity

The spatial responsivity does not require measurement when the incident radiation completely fills the receiving aperture during both calibration and measurement; however, the possibility that radiation may be scattered into the radiometer detector from sources outside the field of view of the radiometer must always be investigated. As mentioned earlier, such external source contributions can usually be held to small values by good optical design. When necessary, spatial responsivity can be measured by employing a small movable source located at the correct distance from the radiometer so that the source image is sharply focussed on the field stop. The radiometer output is then measured for different positions of the source relative to some convenient set of reference axes.

The spatial responsivity of the microwave radiometer is determined by the antenna power pattern. Response of the instrument to radiation sources located anywhere in the sphere surrounding the antenna can be calculated using (15:33).

Table 15.3. Radiometric Quantities

Quantity	Definitions	Units
Radiant power, P	Rate of transfer of radiant energy	Watt
Radiant intensity, J	Radiant power per unit solid angle from a source	Watt steradian ⁻¹
Radiance, N	Radiant power per unit solid angle per unit area from a source	Watt steradian ⁻¹ cm ⁻²
Irradiance, H	Radiant power per unit area incident upon a surface	Watt cm ⁻²

Temporal Responsivity

Instrument response to rapid changes in the power level of the incident radiation can be easily determined for "DC" radiometers by modulating the source radiation and noting the time required for the output to reach its steady-state value. For switching radiometers, the calibration source should be modulated at a variable rate. The temporal responsivity is then obtained by measuring the radiometer output versus source modulation frequency. It is extremely important that the temporal responsivity of scanning radiometers (to be discussed below) be precisely known because of interdependence with the scanning rate. In some cases, the temporal responsivity of an instrument can be calculated using the integration time, chopping frequency, etc.

Polarization Effects

Unless it has been established that the target radiation is not polarized, the IR radiometer output should be measured employing different polarizations for the calibration source. For microwave radiometers, the polarization to which the instrument responds is determined by the antenna system.

Typical IR Instruments

Now that some of the fundamental considerations of infrared radiometry have been discussed, we shall describe a few instruments commonly used in IR measurements.

A frequent requirement in IR work is for a radiation source of high spectral purity, e.g., in spectral responsivity measurements. The monochromator is a device that generates such radiation by means of prisms or gratings which disperse radiation from a broadband IR source into relatively narrow spectral intervals. Several dispersions may be required to achieve high spectral purity and eliminate contamination of the output by scattering; in addition, filters and choppers are sometimes used to eliminate or discriminate against radiation in an undesired spectral region. A diagram of a simple monochromator is given in (F15.10).

In some IR studies, it is required to investigate some characteristic, e.g., thermal radiation from some material, a solid, a gas, or perhaps the atmosphere, over a large wavelength interval. A class of instruments that have been developed for use in such applications is the spectrometer. This device may be considered as a special case of a tunable IR radiometer. Spectrometers consist of two major optical components in addition to the usual detector and amplifying electronics. These optical components are (1) the entrance optics which focus the incident radiation onto the instrument's entrance slit and onto the detector after spectral dispersion, and (2) the optical system that disperses the incident radiation into a spectrum. The focussing system generally employs lenses or mirrors.

Several techniques may be used for the second optical system. The most common devices and those that must be employed to obtain high spectral resolution are prisms and gratings. Another method makes use of narrow passband optical filters, often arranged in a wheel, which are inserted systematically in the beam of incident radiation to vary the spectral response of the instrument. In prism spectrometers, the wavelength is varied by changing the angle of incidence of the incoming radiation upon the prism. For diffraction gratings, the wavelength is varied by alternating the position of the grating relative to the beam of incident radiation. The grating is superior to the prism in resolving power.

Spectrometers ordinarily use thermal detectors because of their relatively uniform broad spectral responsivity. However, some high speed instruments may contain a selective detector in order to take advantage of its relatively fast response time.

Another type of spectrometer has come into use because of its rapidity in making measurements over a wide spectrum and its improvement in the signal-to-noise ratio by virtue of the effective increase in the number of spectral elements observed in a short period of time. This instrument is called the interference spectrometer, or sometimes, the multiplex spectrometer to distinguish this class of instrument from the conventional or monochromator spectrometer. The basic principle of operation is to generate an interferogram of incoming radiation in an interferometer; the spectrum is then retrieved by Fourier transformation of the interferogram. The operation may be better understood by considering the schematic of the Michelson interferometer shown in (F15.11). Incident radiation is divided into two rays by the beam splitter. After reflection from the plane mirrors and a second division by the beam splitter, two recombined beams are formed of which one is focussed on the detector. The detector output is recorded versus displacement of the movable mirror producing an interferogram. The spectrum is usually recovered

by digitizing the interferogram and performing the Fourier transformation in a digital computer. The spectrum may also be recovered by means of a wave analyzer but with much poorer spectral resolution than can be obtained with a digital computer. A more detailed introductory article on Fourier transform spectroscopy is given by Lowenstein (1966).

In conventional spectrometers, choppers are employed to define a reference radiation level and to improve stability in the same manner as in ordinary IR radiometers. In the interferometer spectrometer, the movement of the mirror can serve as the chopper and the detector temperature provides the reference level.

Calibration of spectrometers is performed in much the same way as for an ordinary IR radiometer and should follow the basic principles outlined earlier. Some spectrometers perform wavelength calibration along with measurements of the source by observing the radiation from a line source of known wavelength.

Another special class of instrument is the scanning radiometer; in general, scanning devices are used when spatial resolution of a source of radiation is desired. One form of device used to discriminate between target and background radiation makes use of reticle scanning. In this system, a reticle with clear and opaque areas interrupts the incident beam in a manner similar to chopping described earlier. The reticle pattern may be arranged to provide background discrimination or to indicate angular position and motion of targets in the radiometer's field of view.

A second important scanning system is employed when it is necessary to observe the variation in a radiant quantity over a large area: the optical-mechanical scanner is an IR radiometer which systematically scans an area of interest. The detector output may be electronically recorded as a function of the radiometer pointing angle for later analysis or may modulate the intensity of a light source that is coupled to the scanning motion to provide an image that may be recorded photographically. Various scanning patterns may be used, e.g., circular, spiral, raster. For general imaging work, the raster scan is probably the most commonly used pattern.

In this section we have described some basic radiometric techniques and methods and have discussed a few specialized instruments employed at microwave and infrared frequencies. We now present some specific applications of radiometry in remote sensing of various properties of the atmosphere; a few examples of passive geophysical remote sensing will also be given.

15.3 Applications

15.3.1 Ground-based microwave determination of low altitude temperature profiles.

Ground-based measurements of thermal emission in the 60 GHz band of O_2 are a source of information about vertical temperature profiles. The emission at any frequency and any angle is the weighted spatial average of the temperature profile over a distance whose optical depth is roughly equal to unity. By varying the optical depth (by angular or frequency diversity) a set of brightness observations are generated that are sensitive to different layers of the temperature profile. The temperature profile is then recovered from the brightness measurements by inversion of an integral equation (see chapter 16 for a discussion of inversion techniques).

We will now describe a multi-channel mm wave radiometer which has been used to implement the temperature profiling technique. Major objectives in the design of the radiometer were capability of operation on 3 or 4 frequencies from a common antenna, high angular resolution in the antenna system, good spectral resolution, and the greatest possible sensitivity consistent with the other design goals.

A block diagram of the final radiometer design is given in (F15.12). The basic system is the well known Dicke radiometer in which the receiver is switched between the antenna and a temperature-controlled reference termination. In the present system, the Dicke switch (modulator) is a rotating vane attenuator that alternately transmits and absorbs the energy from the antenna. The components located between the antenna and rotary modulator serve the following functions: during calibration the radiometer is connected to one of the temperature controlled terminations through the waveguide switch to establish a known temperature reference. (At 0° elevation angle and in the 60 GHz spectral region the atmosphere is essentially a blackbody having an effective temperature very nearly equal to the surface temperature; hence, atmospheric emission may also be used as a calibration reference point). Incremental calibration of the radiometer output is accomplished by injecting noise from the gas discharge noise source into the signal channel via the precision attenuator and auxiliary arm of the directional coupler. During measurement, the antenna is connected to the radiometer and noise is injected into the signal channel to allow operation in a

“semi-balanced” condition. Since a true balance cannot be obtained at more than a single frequency, an appropriate amount of noise is injected to obtain an approximate balance at all frequencies.

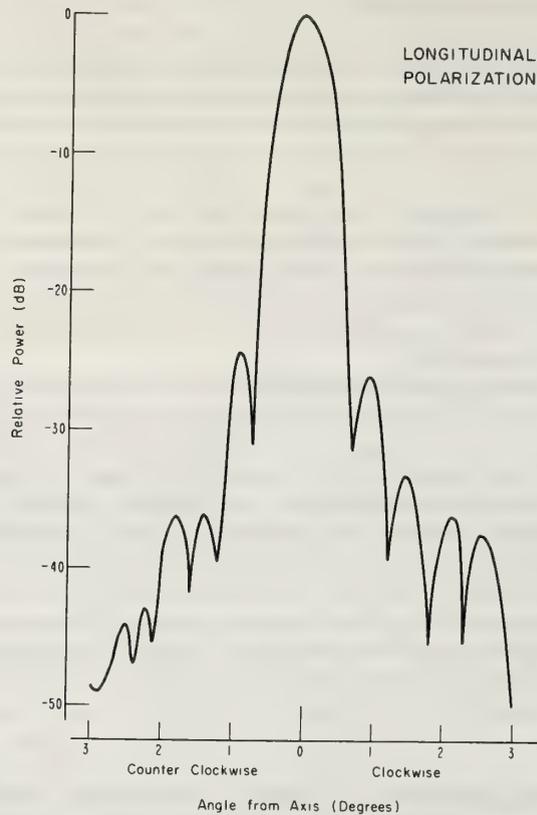


Figure 15.14 Radiation pattern of conical horn antenna at 54 GHz—longitudinal polarization.

The mixer translates a 4 GHz band of frequencies in the 50 GHz region to a 4-8 GHz first intermediate frequency. After a stage of amplification, separation into four individual channels takes place in a strip-line multiplexer; center frequencies are 4.5, 5.5, 6.5, and 7.5 GHz. A second conversion to a 60 MHz center frequency follows the multiplexer; the noise bandwidth of the 60 MHz IF amplifiers is approximately 30 MHz to achieve a fair degree of spectral resolution. Following the second IF amplifiers are four identical stages of video detection, synchronous detection using a lock-in amplifier, and analog output recording.

All components before the lock-in amplifiers are located in an equipment enclosure mounted at the antenna. The waveguide compartment is heated to maintain the ambient temperature reasonably constant. Since ohmic losses in the waveguides generate noise, waveguide temperatures are measured by thermistor probes located at several points and remotely recorded for subsequent use in computing the antenna temperature.

The radiometer enclosure and its 1.2 m diameter conical horn antenna are installed on an elevation over azimuth antenna mount which provides the capability of angular scan operation. (See F15.13). The antenna position is read out remotely at the data recording position. A relatively large antenna aperture was employed to obtain a high angular resolution mentioned earlier, while the conical horn configuration was selected to minimize stray radiation into the back and side lobes and thereby simplify the conversion of antenna temperature to brightness temperature. The first few degrees of the longitudinally polarized power pattern for this antenna are shown in (F15.14).

A summary of the essential radiometer characteristics is presented in (T15.4).

Table 15.4. Multi-frequency radiometer characteristics.

Operating frequency	— 52.5, 53.5, 54.5, 55.5 GHz
Antenna characteristics	— 1.2 meter diameter conical horn, 3 dB beamwidth 0.3°
Receiver type	— Dicke switching radiometer, dual-conversion superheterodyne 1st IF: 4 to 8 GHz 2nd IF: 60 MHz center frequency, noise bandwidth approx. 30 MHz
Sensitivity	— Approx. 1 K for 60s integration time

A few results from a recent temperature sensing experiment (Snider, 1971) are shown in (F15.15). These profile recoveries were determined by applying statistical techniques (Westwater, 1972) to multi-spectral and multiangle data gathered by the radiometer described above. As is evident from inspection of (F15.15), the correspondence between the temperature profiles measured by a radiosonde and by the radiometer is good.

15.3.2 Microwave atmospheric range correction

Sophisticated missile tracking systems require accurate methods of microwave range correction along slant paths through the atmosphere. Similar problems arise in atmospheric corrections for radio astronomy measurements. Currently used correction techniques are based on surface or radiosonde observations or both. These techniques are limited primarily by the high variability of water vapor in both space and time, and by departures of water vapor from horizontal stratification. In this section we describe a technique of measuring thermal emission near the 22.235 GHz water vapor absorption line to correct for the wet component of integrated refractivity (Westwater, 1967).

The correction technique depends on first determining the total atmospheric attenuation from an emission measurement and then relating the attenuation to integrated wet refractivity. In the absence of external sources, the microwave brightness, $T_{b\nu}$, is given by

$$T_{b\nu} = \int_0^{\infty} T \alpha_{\nu} e^{-\int_0^r \alpha_{\nu} dr'} dr, \quad (15:36)$$

$$= T_m (1 - e^{-\int_0^{\infty} \alpha_{\nu} dr'}) \quad (15:37)$$

$$= T_m (1 - e^{-\tau_{\nu}}) \quad (15:38)$$

where the mean radiating temperature, T_m , is a functional of temperature and composition profiles along the viewing direction. In general, T_m is not known, but it can be estimated from surface observations, from local climatology, or from a recent radiosonde release. In the weak attenuation limit (i.e., $\tau_{\nu} \ll 1$) to (15:38), the brightness is much more sensitive to τ_{ν} than T_m , as we will show qualitatively with a simple linear analysis. First, we expand (15:37) in a Taylor expansion (assuming $\tau_{\nu} \ll 1$)

$$T_b \doteq (1 - e^{-\tau_{\nu}}) \delta T_m + e^{-\tau_{\nu}} T_m \delta \tau_{\nu}, \quad (15:39)$$

$$\doteq \tau_{\nu} \delta T_m + T_m \delta \tau_{\nu}. \quad (15:40)$$

The relative sensitivity of brightness fluctuations to fluctuations in τ_ν or T_m is determined by the ratio

$$\frac{\delta\tau_\nu}{\tau_\nu} \bigg/ \frac{\delta T_m}{T_m}$$

Reasonable numerical values of $\frac{\delta\tau_\nu}{\tau_\nu} = .5, \delta T_m = 3 \text{ K}$ and $T_m = 300 \text{ K}$

result in the ratio of about 50 to 1, showing that in this case the brightness depends much more on total attenuation than average temperature.

Solving (15:37) for the total attenuation yields

$$\int_0^\infty \alpha_\nu dr' = -\ln \left[1 - \frac{T_{b\nu}}{T_m} \right] \quad (15:41)$$

It remains to relate the attenuation to the integrated wet refractivity, N_w , where N_w is related to absolute humidity, ρ_w by

$$\int_0^\infty N_w ds, \quad N_w = 1721.4 \rho_w / T \quad (15:42)$$

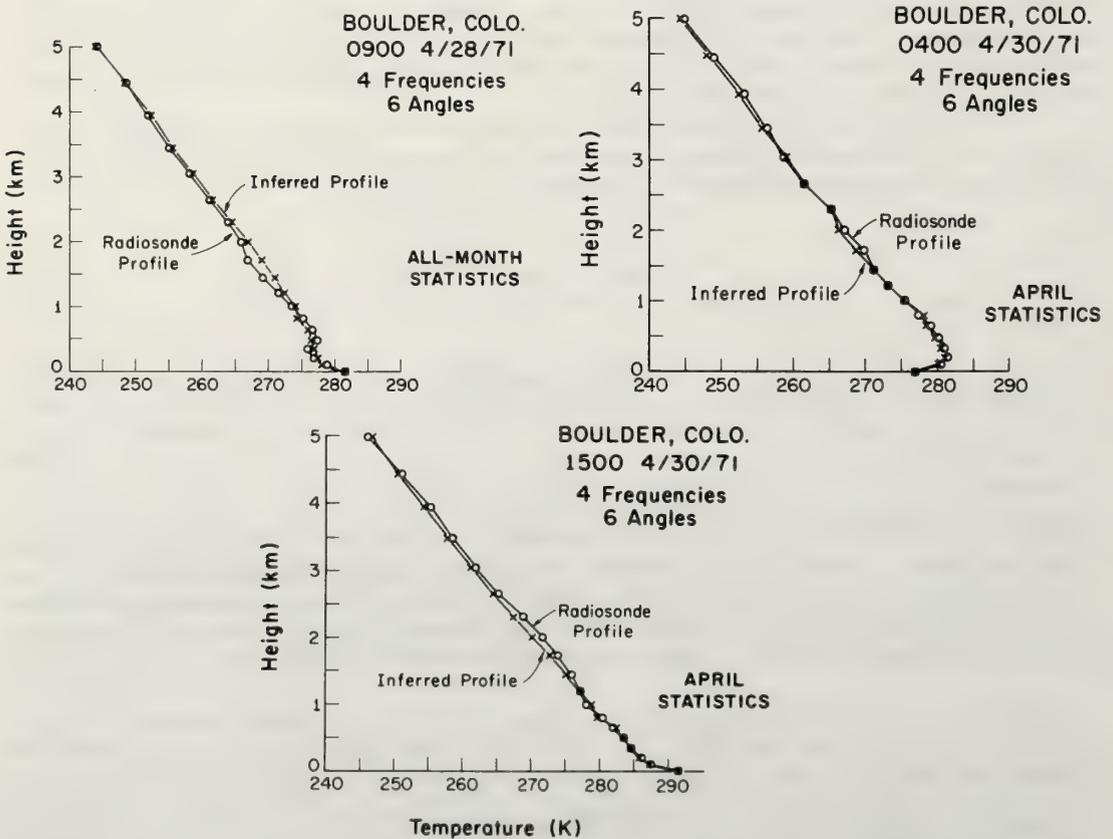


Figure 15.15 Inferred and measured temperature profiles using combination of angle-scan and multi-spectral input data.

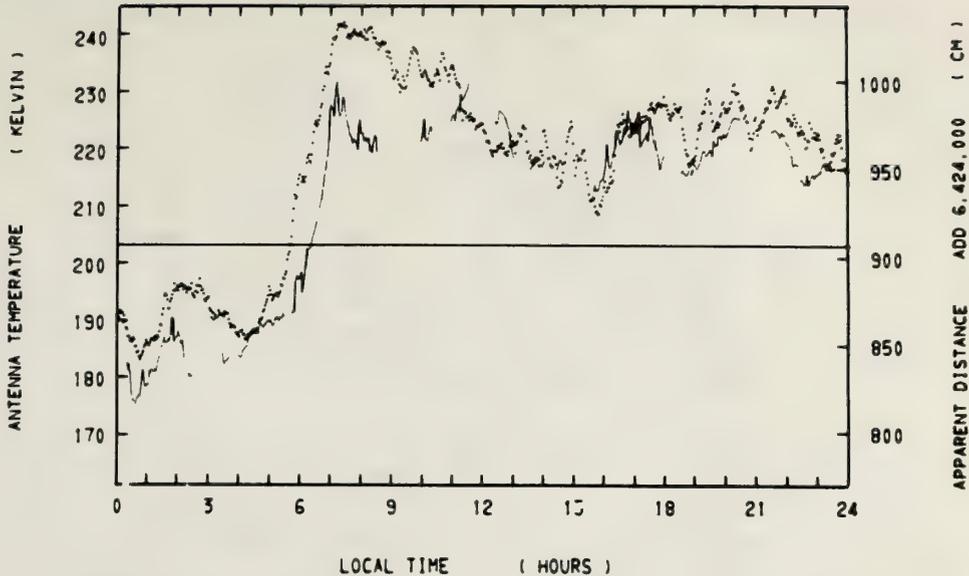


Figure 15.16 Comparison of radiometer antenna temperature at 7.48 degrees elevation solid curve, and apparent distance, dotted curve, for data collected 7/30/68. The correlation coefficient is 0.926.

The method by which this is achieved is rather involved and will not be presented here. However, since the dominant variable in both N_w and $\alpha_{\nu w}$ is the absolute humidity, it should be plausible that the integrals of the two should be closely related. Further details of this part of the reductive process are given by Westwater (1967).

To test the above technique, a recent experiment was conducted on a calibrated range path in the Hawaiian islands. Microwave electrical path measurements, taken over a 64 km ground to mountain top propagation path, were compared with 20.6 GHz thermal emission measurements made along an adjacent path. This experiment clearly showed that, during conditions of clear and light cloud cover, fluctuations in range and antenna temperature were well correlated. Some of the results of this experiment are shown in (F15.16). A complete discussion of the experiment is given by Guiraud, et al. (1972).

The technique discussed in this section, has further potential as a method of determining water vapor content along any slant path through the atmosphere. A single frequency technique can determine this quantity during clear and partially overcast conditions, while the addition of a properly chosen radiometric cloud measurement could correct for all but rain clouds.

15.3.3 Vertical distribution of water vapor from satellite Infrared Spectrometer measurements.

The estimation of vertical humidity profiles from satellite radiance observations has been discussed by Smith and Howell (1971), Conrath (1969), and Gorchakova, et al., (1970). This technique requires mathematical inversion of high resolution ($\sim 5 \text{ cm}^{-1}$) radiance observations in well chosen spectral intervals in H_2O absorption bands (such as $6.3 \mu\text{m}$) to yield vertical humidity profiles.

In deriving the humidity profile from the radiance observations, a linear perturbation analysis of the transfer equation, leads to an equation of the form

$$\delta I_{\nu} = \int_0^{P_s} \delta \rho(P) W(\nu, P) dP \quad (15:43)$$

where δI_{ν} is the difference between the measured radiance and that calculated from a first estimate to the profile, $\delta \rho(P)$ is the difference between true and estimated humidity at pressure P , and the weighting function $W(\nu, P)$ is determined from vertical temperature and transmission profiles. The integration in (15:43) ranges from the top of the atmosphere ($P = 0$) to the surface ($P = P_s$).

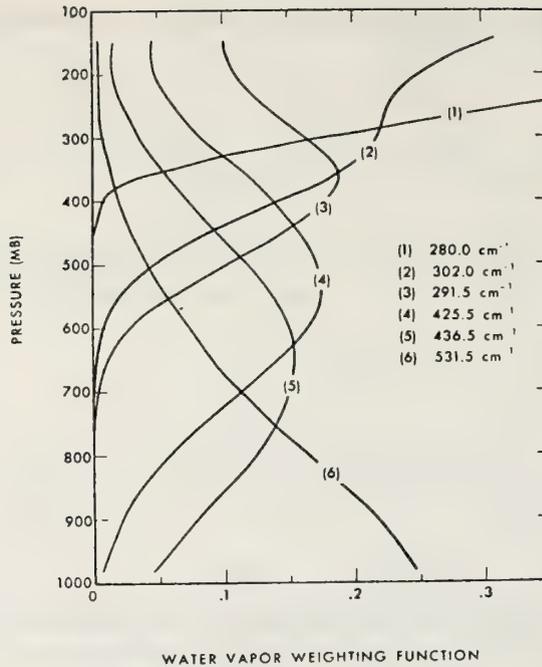


Figure 15.17 Atmospheric weighting functions (see below) for the water vapor channels of SIRS-B (from Smith and Howell, 1971).

The vertical height resolution and number of independent pieces of information about the humidity structure are related to both the widths and to the extent of overlap of the weighting functions. A set of weighting functions appropriate to six water vapor channels of the Nimbus IV SIRS-B experiment are shown in (F15.17). (Smith and Howell, 1971). Note that these functions are relatively broad and that they overlap to a certain extent, but that they peak at different levels in the atmosphere.

In addition to weighting functions, several other factors affect the achievable accuracy in humidity profile estimation. Among these are uncertainties in the temperature profile, errors in transmission functions, and measurement noise levels. The most serious limitation to this and other IR profile retrieval techniques is the presence of clouds which often partially fill the radiometer beam. Clouds are generally opaque and tend to reduce upwelling radiance; i.e., in a given direction, they extinguish radiance from below, while radiating as black bodies at the colder temperature of the cloud top. A cloudy atmosphere profile retrieval given by Smith and Howell (1971) is shown in (F15.18). Results both with and without a cloud correction are compared to a nearby radiosonde sounding. Agreement above the cloud level (~ 800 mb) is excellent. Although it is clear that problems still remain with this technique, the potential for global retrieval of humidity profiles is evident.

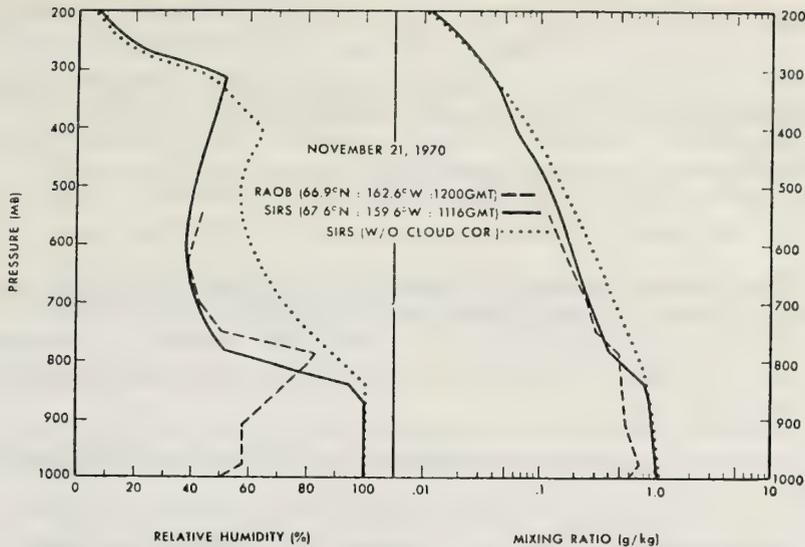


Figure 15.18 *SIRS-B* retrieval of a water vapor profile in a cloudy atmosphere (from Smith and Howell, 1971).

15.3.4 Further Applications

In 15.3.1, 15.3.2, and 15.3.3 we have shown in detail a few applications of radiometers to remote atmospheric sensing. Among other of the many promising geophysical applications are:

- a. Remote sensing of sea surface temperatures from satellite and aircraft measurements of surface radiance. Infrared emissivities of sea water are essentially unity, and measurements of upward radiance in window channels permits determination of sea surface temperatures to about 1°C (Smith, et al., 1970) during clear conditions. Scanning radiometer measurements of the NASA ITOS-1 satellite have been used to construct maps of sea surface temperatures [Rao, et al., 1971].
- b. Mapping of ice cover in oceans by passive microwave techniques from satellites and aircraft. The large difference in emissivity between ice and water allows the percentage of ice cover over smooth oceans to be estimated (Basharinov, et al., 1970). Measurements of emission as a function of frequency, polarization, and viewing angle also permit inferences about sea state, but many competing effects are present and unique information is difficult to obtain. However, microwave measurements below 15 GHz are only weakly affected by clouds and offer promise of near all weather capability in surface sensing.
- c. Measurements of vertical temperature profiles from satellites. See chapter 16, where this technique is discussed in detail.
- d. Ground-based microwave radiometric observations of liquid water in thunderstorms. Observations of thermal emission at 10.7 GHz have been used to derive contours of integrated liquid water content in thunderstorm cells (Decker and Dutton, 1970). This technique derives attenuation from emission measurements, which, with simplifying assumptions, can be directly related to the desired quantity. With additional frequencies, the technique has potential to infer liquid-ice ratios in thunderstorms.
- e. Ground-based microwave sensing of temperatures in the stratosphere and lower mesosphere. On the edges of the 60 GHz absorption band, thermal emission from narrow O_2 lines in the 30 to 60 km region can penetrate the troposphere and provide a means of monitoring temperatures at these altitudes (Waters, 1971). Height resolution of about 16 km and accuracy of $\sim 2.5^{\circ}\text{K}$ are possible with present state of the art instruments.

15.4 References

- Basharinov, A. E., A. S. Gurvich, L. T. Tuchkov and K. S. Shifrin (1970), The terrestrial thermal radio-emission field, *Izv. Atmos. and Ocean Phys.* 6, 366-380.
- Bell, Ely E. (1959), Radiometric quantities, symbols, and units, *Proc. IRE* 47, 1432-1434.
- Chandrasekhar, S., 1960: *Radiative Transfer*. Dover Publications, Inc., New York.
- Conrath, B. J. (1969), On the estimation of relative humidity profiles from medium-resolution infrared spectra obtained from a satellite. *J. Geophys. Res.* 74, 3347.
- Decker, M. T. and E. J. Dutton (1970), Radiometric observations of liquid water in thunderstorm cells, *J. Atmos. Sci.* 27, 785-790.
- Dicke, R. H. (1946), The measurement of thermal radiation at microwave frequencies, *Rev. Sci. Instr.* 17, 268-275.
- Goody, R. M., (1964): *Atmospheric Radiation, I. Theoretical Basis*. Oxford University Press, London.
- Gorchakova, I. A., M. S. Malkevich, and V. F. Turchin (1970), Determination of the vertical humidity profile of the atmosphere from measurements of the earth's natural radiation, *Izv. Atmos. and Ocean. Phys.* 6, 565-576.
- Guiraud, F. O., M. T. Decker and E. R. Westwater (1972), Experimental investigation of the correction of electrical range errors by passive microwave radiometry. NOAA Technical Report, ERL 221-WPL 19.
- Hackforth, Henry L., 1960: *Infrared Radiation*, McGraw-Hill Book Company, New York.
- Jamieson, John A. et al., 1963: *Infrared Physics and Engineering*, McGraw-Hill Book Company, New York.
- Jones, R. Clark (1959), Phenomenological description of the response and detecting ability of radiation detectors, *Proc. IRE* 47, 1495.
- Kondratyev, K. Ya., 1969: *Radiation in the Atmosphere*. Academic Press, New York and London.
- Kraus, John D., 1966: *Radio Astronomy*, McGraw-Hill Book Company, New York.
- Lowenstein, Ernest V. (1966), The history and current status of Fourier transform spectroscopy, *Applied Optics* 5, 845-854.
- Nyquist, H. (1928), Thermal agitation of electrical charge in conductors, *Phys. Rev.* 32, 110-113.
- Rao, P. K., A. E. Strong and R. Koffler (1971), Sea surface temperature mapping off the eastern United States using NASA's ITOS satellite. *Proceeding of the Seventh International Symposium on Remote Sensing of Environment*, Vol. I, 17-21 May 1971, 683-703.
- Shafrin, Yu A. (1970), On the possibilities of the infrared method for the measurement of the vertical distribution of ozone with the aid of satellites, *Izv. Atmos. and Ocean. Phys.* 6, 696-703.

- Smith, W. L. and H. B. Howell (1971), Vertical distributions of atmospheric water vapor from satellite infrared spectrometer measurements, *J. App. Met.* *10*, 1026-1034.
- Snider, J. B. (1971), Recent results in ground-based sensing of atmospheric temperature profiles, Proceedings of Symposium on Air Pollution, Turbulence and Diffusion, 7-10 December 1971, 56-65.
- Van Vleck, J. H. and V. F. Weisskopf (1945), On the shape of collision-broadened lines, *Rev. Mod. Phys.* *17*, 227-.
- Volchok, B. A. and M. M. Chernyak (1968), Microwave radiation transfer in clouds and precipitation, chapter in *Transfer of Microwave Radiation in the Atmosphere*, K. S. Shifrin, Ed. Available from U. S. Department of Commerce Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 22151.
- Wark, D. Q. and H. E. Fleming (1966), Indirect measurements of atmospheric temperature profiles from satellites; I. Introduction, *Mon. Wea. Rev.* *94*, 351-362.
- Waters, J. W. (1971), Ground-based passive microwave sensing of temperatures in the stratosphere and lower mesosphere, Proceeding of the Seventh International Symposium on Remote Sensing of Environment, Vol. III, 17-21 May 1971, 1765-1776.
- Westwater, E. R. (1967), An analysis of the correction of range errors due to atmospheric refraction by microwave radiometric techniques, U.S. Department of Commerce ESSA Technical Report IER 37-ITSA 37.
- Westwater, E. R. (1972), Ground-based determination of low altitude temperature profiles by microwaves, *Monthly Weather Rev.*, *100*, 1, 15-28.
- Wolfe, William L., 1965: *Handbook of Military Infrared Technology*, U. S. Government Printing Office, Washington, D. C.

List of Symbols

a_ν	Absorptivity of a surface at frequency ν	$p(\vec{s}', \vec{s})$	Function describing the fraction of energy incident from direction \vec{s}' that is scattered into direction \vec{s}
A	Area		
A_d	Detector area		
B	Pre-detection bandwidth	P	Position or pressure
$B_\nu(T)$	Planck function	r	Length
c	Velocity of light	r_ν	Reflectivity of a surface at frequency ν
C	Detectability constant	R	Responsivity
C_1, C_2, C_3	Proportionality constants	\vec{s}	Unit Vector
C_d	Coupling factor of directional coupler	t	Time
dB_ν/dp	Rate of change of Planck function with pressure	T	Temperature
$d\tau/dU$	Rate of change of transmittance with path length	T_a	Antenna temperature
D	Detectivity	$T_b(\Omega)$	Brightness temperature in direction of solid angle Ω
D^*	Specific detectivity	T_{bm}	Brightness temperature integrated over antenna main beam
e_ν	Volume extinction coefficient at frequency ν	$T_{b\nu}$	Equivalent black body temperature, or brightness temperature at frequency ν
E_ν	Energy at frequency ν	T_e	Excess noise temperature
f_ν	Ratio of scattering coefficient to total extinction coefficient	T_g	Effective noise temperature of microwave noise source
f_o	Chopper frequency	T_m	Mean radiating temperature
$G(\Omega)$	Antenna gain in direction of solid angle Ω	T_o	Ambient temperature of calibration circuit
G_o	Pre-detection gain	T_I	Effective noise temperature of a microwave radiometer
h	Planck's constant = 6.63×10^{-34} joule sec	T_t	Effective noise temperature of reference termination
H	Irradiance	V	Volume
I_ν	Specific intensity at frequency ν	V_o	Radiometer output voltage
j_ν	Emission coefficient at frequency ν	w	Noise power per unit bandwidth from a resistance
$j_\nu^{(s)}$	Contribution to emission coefficient by scattering	$W(\nu, P)$	Weighting function
J	Radiant intensity	α_ν	Absorption coefficient at frequency ν
J_ν	Source function at frequency ν	$\alpha_{\nu w}$	Contribution to absorption coefficient by water vapor
k	Boltzmann's constant = 1.38×10^{-23} joule K^{-1}	$\delta\rho(P)$	Difference between true and estimated water vapor density at pressure ρ
L_{at}	Attenuator resistive loss factor	Δf	Measurement bandwidth of amplifier following an infrared detector
m	Mass	ΔG	Pre-detection gain fluctuations
M	Modulation constant	ΔT_{min}	Minimum detectable temperature
\vec{n}	Unit normal to a surface	ΔV_o	Radiometer output voltage fluctuations
N	Radiance		
N_w	Integrated wet refractivity		
NEP	Noise equivalent power		

ϵ_ν	Emissivity of a surface at frequency ν
λ	Wavelength of incident radiation
λ_p	Wavelength at which detectivity is a maximum
ν	Frequency
Ω	Solid angle
ρ	Mass density
ρ_w	Water vapor density (absolute humidity)
σ_ν	Scattering coefficient at frequency ν
τ	Integration time
$\tau_\nu(P_1, P_2)$	Optical distance between points P_1 and P_2

Chapter 16 INVERSION TECHNIQUES

Ed R. Westwater and Otto Neall Strand

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

The determination of profiles from indirect observations is a commonly occurring problem in atmospheric physics. In many cases, this inverse problem can be well approximated by a Fredholm integral equation of the first kind. Physical examples of radiometric determination of temperature and density profiles are related to such equations. Mathematical properties of this equation are discussed and smoothing, statistical and iterative inversion techniques are presented. The chapter concludes with examples of temperature profile retrievals from ground-based microwave and satellite infrared radiance observations.

16.0 Introduction

The application of remote sensing techniques to geophysics has generated a variety of inversion problems. Well-known examples of inverse problems include determination of vertical temperature and density profiles from remote radiation measurements (Wark and Hilleary, 1969; Staelin, 1969), estimation of particle drop-size distributions from scattering measurements (Dave, 1971), and prediction of density distribution within the earth from measurements of seismic wave travel times (Backus and Gilbert, 1968). Although the underlying physics of many inverse problems is widely different, their mathematical structure is similar.

In this chapter, we discuss specifics of a few examples of radiometric inverse problems that are encountered in tropospheric remote sensing, abstract from these examples features common to many inversion problems, and finally we discuss several of the commonly used inversion algorithms.

16.1 Physical Examples of Inverse Problems

We consider methods of determining profiles from remote radiation measurements. The radiative intensity, I_ν at frequency ν emitted by a non-scattering atmosphere in local thermodynamic equilibrium is given by the integral form of the radiative transfer equation (Goody, 1964)

$$I_\nu = \int_0^{\infty} B_\nu(T) \alpha_\nu \exp\left[-\int_0^h \alpha_\nu(h') dh'\right] dh + I_\nu^o \exp\left[-\int_0^{\infty} \alpha_\nu(h') dh'\right] \quad (16:1)$$

where $B(T)$ is the Planck function, T is the absolute temperature, h is the height, α_ν is the absorption coefficient, and I_ν^o is the unattenuated intensity at the opposite boundary of the emitting medium. In general, the absorption coefficient is a function of composition, density, and temperature. Physically, the amount of energy emitted at h in the direction of I_ν is $\alpha_\nu B_\nu(T) dh$; the fractional amount of this energy surviving to the measuring point is $\exp\left[-\int_0^h \alpha_\nu(h') dh'\right]$. The total intensity is then the sum of incremental emission plus the attenuated external contribution. In the earth's atmosphere, the principal gaseous absorption and emission is from the IR bands of CO_2 , H_2O , and O_3 and the microwave lines of O_2 and H_2O . We will not consider cloud emission here.

The determination of vertical temperature profiles from satellite measurements of upward radiance is a well-studied example of an inverse problem. Information about vertical temperature structure is obtained in the following way: Consider a well-mixed absorbing gas such as CO_2 or O_2 . At the top of the atmosphere, the density, and hence the emissivity, is zero. As one penetrates the atmosphere, the pressure broadening of absorbing spectral lines increases the upward emission. The increase persists until the attenuation above the emitting layer becomes dominant. By varying the wavelength from strongly absorbing to weakly absorbing spectral regions, the contribution from different height layers can be varied, and a degree of height resolution can be achieved. (The surface contribution, I_ν^o , can be determined from a "window channel" measurement.)

Thus, we may view a radiance observation as a weighted spatial average of the vertical Planck function profile with a weighting function

$$K(\nu, h) = \alpha_\nu(h) \exp\left[-\int_0^h \alpha_\nu(h') dh'\right] \quad (16:2)$$

For a well mixed gas, this function is known, except perhaps for a small temperature dependence. Typical weighting functions for downward microwave observations are shown in (F16.1). Note the overlapping of the functions; this illustrates a not always realized fact that n measurements do not necessarily yield n independent pieces of information. Except for a small temperature dependence, the weighting functions can be calculated as a function of altitude (or pressure) using empirical or theoretical pressure broadening equations with standard atmospheres: By considering departures from a standard atmosphere, both the infrared problem (Rodgers, 1966) and the microwave problem (Westwater, 1970) can be linearized and reduced to an integral equation of the form

$$\int_a^b K(x, y) f(y) dy = g(x) \quad (16:3)$$

where $g(x)$ is a measured quantity, $K(x, y)$ is the known kernel or weighting function, and $f(y)$ is the desired solution.

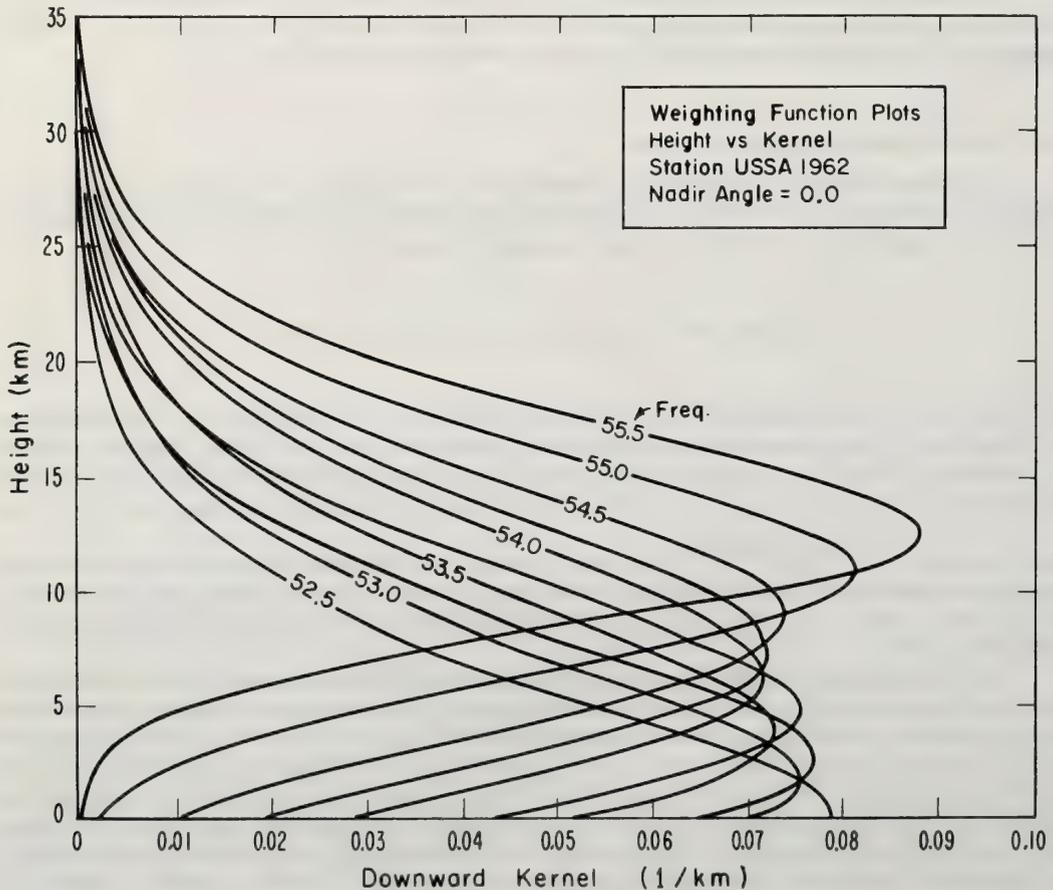


Figure 16.1 Weighting functions for downward looking multispectral temperature probing.

WEIGHTING FUNCTIONS FOR ANGULAR PROBING

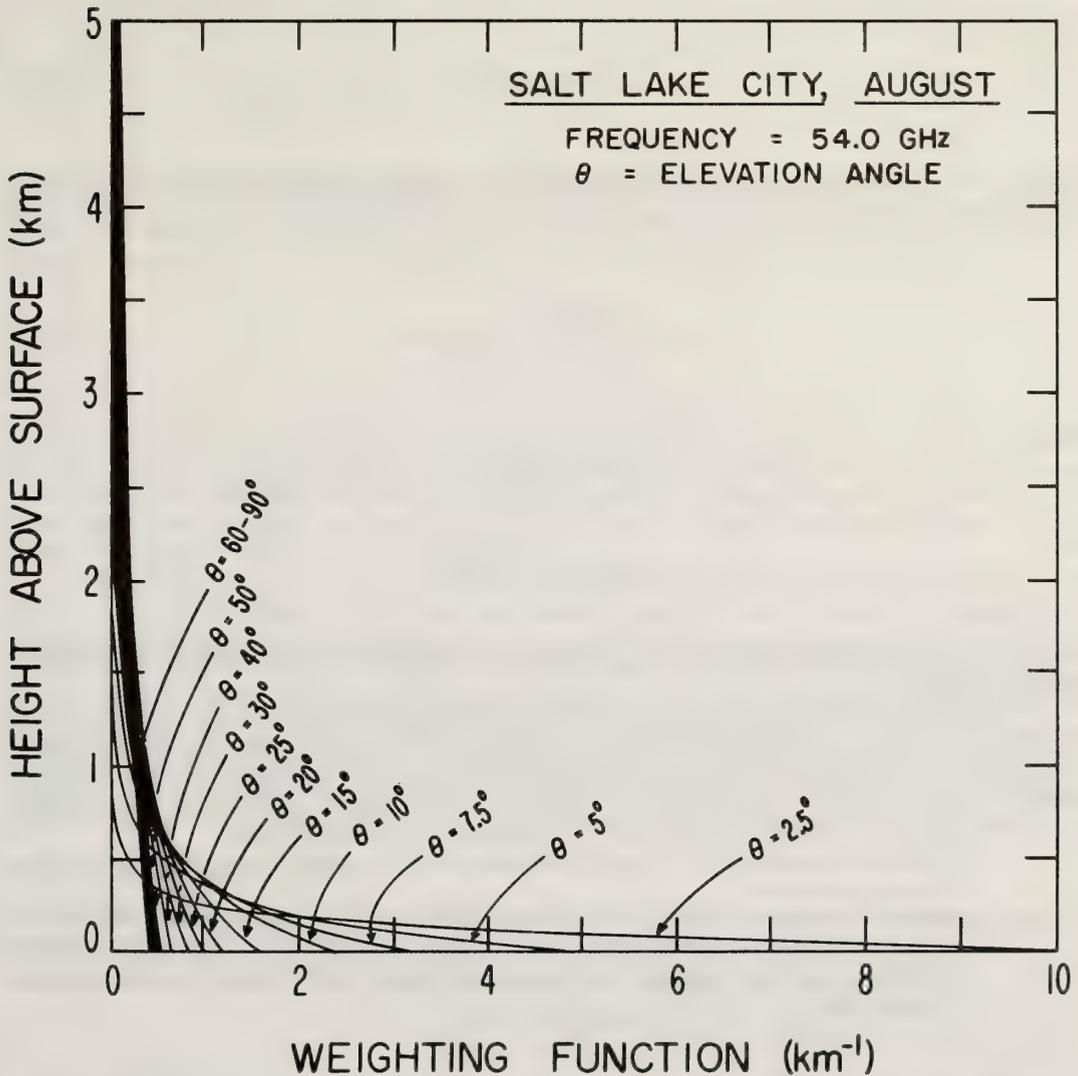


Figure 16.2 Weighting functions for upward looking angular temperature probing.

A similar inverse problem is encountered when determining low altitude temperature profiles (Westwater, 1972) by ground-based measurements of downward emission by oxygen. Data can be generated by multi-spectral or variable angle techniques. An example of ground-based weighting functions for an angular technique is shown in (F16.2). The form of the integral equation to solve is again (16:3), but direct surface measurements of temperature can be added as an exact constraint on the unknown profile.

Radiometric determination of water vapor and ozone profiles supply the final example of an inverse problem that we discuss here. Consider (16:1) when $B_\nu(T)$ is known (or equivalently, when errors in estimating $B_\nu(T)$ are much less than measurement errors in I_ν) and when the surface term is either known or negligible. Proper wavelength selection to minimize radiation from contaminating variable absorption bands leads to

$$\alpha_\nu = \rho\kappa_\nu + \beta_\nu \quad (16:4)$$

where ρ is the desired density and κ_ν and β_ν are approximately known. Measurements of radiance are now non-linear functionals of the ρ profile

$$I_\nu = \int f_\nu(\rho(h)) dh \quad (16:5)$$

Equation (16:5) can be solved by using the Newton-Raphson method (Conrath, 1969). A Taylor expansion is made about an initial guess $\rho^0(h)$ and truncated after the linear term. This yields

$$\delta I_\nu = \int \frac{\partial f_\nu[\rho^0(h)]}{\partial \rho} \delta \rho(h) dh, \quad (16:6)$$

where

$$\delta I_\nu = I_\nu - I_\nu(\rho^0)$$

and

$$\delta \rho(h) = \rho(h) - \rho^0(h).$$

Finally, to solve the non-linear equation (16:5), we solve (16:6) and iterate the process until some convergence criterion is satisfied.

From the above physical examples we summarize the common features. First, we start with a set of observations, $g(x)$, that are related to the desired profile $f(y)$. We assume that we have solved the direct problem, in the sense that given f we can calculate g to well within the measurement error. We assume the problem is linear, i.e., that (16:3) is satisfied, or that iterative techniques based on solving (16:3) will converge. Since g is the result of a measurement, two features are always present in a real problem:

- a. g is always measured with some error (however small). Frequently, the statistics or a bound on the error are known;
- b. Only a finite (however large) number of measurements is available. Many measurement techniques sample only at discrete intervals. However, even continuous measurements are band limited and can be completely specified by their samples at the correct Nyquist frequency (Parker, 1971).

16.2 The Fredholm Integral Equation of the First Kind: An Ill-Posed Problem

According to the discussion in previous sections, many remote-sensing problems may be reduced to, or reasonably approximated by, a Fredholm integral equation of the first kind:

$$\int_a^b K(x, y) f(y) dy = g(x) \quad (16:3)$$

where $K(x, y)$ is a smooth function, $-\infty < a < b < \infty$, and an attempt is made to infer the function f from experimental measurements of $g(x)$. The left-hand side of (16:3) will often be referred to as the *transform* of f .

Although some interesting classical theory relates to the case in which $g(x)$ is known for a continuous interval of x -values (Smithies, 1958; Landweber, 1951), the usual practical problem involves only a finite number of measurements $g(x_1), \dots, g(x_n)$ in which errors are usually present. For the most part we confine ourselves to this finite case in the following discussion.

16.2.1 Inherent Instability

Consider the integral

$$I_N(f, x_i) = \int_0^\pi K(x_i, y) [f(y) + C \sin Ny] dy, \quad i = 1, \dots, n, \quad N = 1, 2, \dots, \tag{16:7}$$

where C is an arbitrary constant. It follows from the Riemann-Lebesgue lemma (Whittaker and Watson, 1963) and (16:3) with $a = 0$ and $b = \pi$ that $I_N(f, x_i) \rightarrow g(x_i)$, $i = 1, 2, \dots, n$, as $N \rightarrow \infty$. Thus, the set of transformed values can be brought as close to $[g(x_i)]$ as desired by taking the frequency of the sinusoidal term $\sin Ny$ large enough. However, $|C \sin Ny|$ can be arbitrarily large and is not small even in the mean-square sense, as

$$\int_0^\pi \sin^2 Ny dy = \pi/2$$

for any N . The function $f_1(y) = f(y) + C \sin Ny$ can therefore be transformed into a set of values only slightly different from the corresponding values $g(x_i)$. One may look at this result backwards to see that *mere closeness of the transforms to the given numerical values does not guarantee closeness of the unknown function f to its correct value*. The reader is hereby suitably warned that even now occasional attempts are made to justify methods only by the closeness of the transforms to $g(x_i)$.

The consequences of the instability shown above determine the nature of inversion methods which must be used. The successful methods have one feature in common: either explicitly or implicitly they incorporate desired properties of f other than its ability to satisfy (16:3). For example, the method may force f to be smooth, or it may impose conditions such as boundedness, having a known statistical distribution, or being a linear combination of certain well-chosen functions. In any case, the conditions imposed remove the instability and render the problem capable of unique or near-unique solution. Of course, the conditions imposed must correspond with whatever a priori knowledge we may have about f . There is an infinite set of possible functions f whose transforms are within reasonable measurement error of $g(x_i)$, and the successful methods pick out from this set a function f that is to be regarded as superior in some sense to the others.

16.2.2 Quadrature Approximation

In the usual case where the $g(x_i)$ have measurement errors it is often useful to replace the integral in (16:3) by a quadrature, i.e., we approximate all integrals by a formula of the form

$$\int_a^b h(y) dy \cong \sum_{j=1}^m w_j h(y_j).$$

Then (16:3) takes the form

$$\sum_{j=1}^m w_j K(x_i, y_j) f(y_j) = g(x_i) \quad , \quad i = 1, \dots, n. \tag{16:8}$$

or in abbreviated matrix form

$$Af = g \quad , \tag{16:9}$$

where

$$A = (w_j K(x_i, y_j)), \quad f = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}, \quad g = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{bmatrix} .$$

The problem is then reduced to solving for the vector f of unknowns. Several cautions are in order: First, the use of the quadrature is justified if it does not itself produce serious errors in f . This will be true if the measurements of $g(x_i)$ are known to have errors *considerably larger* than those introduced by the quadrature. If the errors in $g(x_i)$ are negligible, then either a formula of known high accuracy must be used or the quadrature must be justified by other means. Second, if $m \leq n$ and the number n_0 of independent equations is at least equal to m , it is possible to solve (16:9) either exactly ($n_0 = m$) or in the least-squares sense ($n_0 \geq m$). In these cases one could obtain estimates \hat{f} given by (A^T is the transpose of A)

$$\begin{cases} \hat{f} = A^{-1}g, & m=n_0=n \\ \hat{f} = (A^T A)^{-1} A^T g & m \leq n_0 \leq n \end{cases} \quad (16:10)$$

Unfortunately, this procedure is often (but not always) unsuccessful for the following reasons: for small m , an interpolation of the values f does not resemble the unknown function f because the set $\{y_i\}$ is too sparse or the quadrature approximation is too crude or both. Furthermore, for larger m the components of \hat{f} become highly oscillatory and the matrix A becomes ill-conditioned. The ill-conditioning of A is related to maximum and minimum eigenvalues of $A^T A$ (Wilkinson, 1965). Certain alternative procedures exist which remedy these difficulties; they are discussed in the next section (Phillips, 1962; Tihonov, 1963; Twomey, 1963). Finally, we note that if $n_0 < m$ one is forced to incorporate more information about f than is contained in (16:9) before any estimate can be obtained at all.

In 16.3 we discuss several well-known inversion methods, all of which are successful because they impose a priori knowledge either implicitly or explicitly on the estimate of f .

16.3 Inversion Techniques

16.3.1 Smoothing Methods

For the present we assume that the matrix equation (16:9) is a suitable approximation in the presence of measurement error, but that, as usual, (16:10) provides an inadequate solution for f . The methods discussed in this section provide the necessary smoothing in an explicit way. The approach here includes the methods of Phillips (1962), Twomey (1963), and Tihonov (1963). Consider minimizing the positive expression

$$U(f) = (g - Af)^T(g - Af) + \gamma_1 (f - f_0)^T(f - f_0) + \gamma_2 (Bf)^T(Bf), \quad (16:11)$$

where $\gamma_1, \gamma_2 > 0$, f_0 is a vector describing a desired solution for f , i.e., a *preconceived-notion* function, and B is an $m \times m$ matrix describing some desirable smoothing of f . The three main terms of $U(f)$ have the following significance: The value $(g - Af)^T(g - Af)$ is a measure of the accuracy with which f satisfies (16:9), and $(f - f_0)^T(f - f_0)$ is determined by the departure of f from the preconceived-notion vector, f_0 . Finally, $(Bf)^T(Bf)$ has a value determined by the departure of f from ideal smoothness, represented by $Bf = 0$. The most common assignments of B describe smallness of zeroth, first, or second derivatives. The following examples will clarify the meaning of (16:11).

Case I (Phillips, 1962; Twomey, 1963). $\gamma_1 = 0$,

$$B = \begin{matrix} & \underbrace{\hspace{10em}}_m & \\ \left. \begin{matrix} -2 & 1 & 0 & \cdot & \cdot & 0 \\ 1 & -2 & 1 & 0 & \cdot & \cdot \\ 0 & 1 & -2 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & 1 & -2 & 1 & 0 \\ \cdot & \cdot & 0 & 1 & -2 & 1 \\ 0 & \cdot & \cdot & 0 & 1 & -2 \end{matrix} \right\} & m. \end{matrix}$$

Here we assume (for simplicity) an evenly-spaced quadrature (such as Simpson's rule) in which the y-values are

$$a, a + \frac{b-a}{m-1}, a + \frac{2(b-a)}{m-1}, \dots, a + \frac{(m-1)(b-a)}{m-1} = b.$$

Then Bf is the discrete analog of the second derivative of the unknown function f . It is often desirable to demand that the unknown function be as smooth as possible consistent with a reasonable approximation of the transforms to $g(x_i)$. One, therefore, chooses γ_2 to balance these two opposing tendencies in a reasonable manner. Thus we require

$$(g-Af)^T (g-Af) + \gamma_2 (Bf)^T (Bf) = \text{minimum.}$$

By matrix differentiation we obtain an equation for the estimate, \hat{f} :

$$2A^T (A\hat{f}-g) + 2\gamma_2 B^T B\hat{f} = 0$$

or

$$\hat{f} = (A^T A + \gamma_2 B^T B)^{-1} A^T g. \tag{16:12}$$

Case II $\gamma_2 = 0$ (Twomey, 1963)

We require

$$(g-Af)^T (g-Af) + \gamma_1 (f-f_0)^T (f-f_0) = \text{minimum.}$$

In a manner similar to that of Case I we obtain

$$\hat{f} = (A^T A + \gamma_1 I)^{-1} (A^T g + \gamma_1 f_0). \tag{16:13}$$

Again γ_1 must be chosen to give a reasonable compromise between (16:9) and the condition $f = f_0$.

Tihonov (1963) has solved what is essentially the non-discretized analog to Case I above, and Wahba (1969) has shown that Tihonov's solution gives the "smoothest" of all possible exact solutions of (16:3) as a limit when $\gamma_2 \rightarrow 0^+$.

16.3.2 Statistical Techniques (Strand and Westwater, 1968b; Rodgers, 1966).

If we set $\gamma_2 = 0$ in the form (16:11) and generalize the result slightly to include weighting by positive definite matrices Γ_1 and Γ_2 we obtain the form

$$V(f) = (g-Af)^T \Gamma_1 (g-Af) + (f-f_0)^T \Gamma_2 (f-f_0) \tag{16:14}$$

where Γ_1 and Γ_2 are, respectively, $n \times n$ and $m \times m$ positive definite matrices. We regard f_0 as representing the mean vector (obtained from previous statistical data on f), let ϵ be the measurement error vector for g (where ϵ

has zero mean) and assume that we have the covariance matrices S_f and S_ϵ defined by

$$S_f = E [(f - Ef)(f - Ef)^T]$$

and

$$S_\epsilon = E [\epsilon\epsilon^T],$$

where E is the expected-value operator. Then, if ϵ and f are normally distributed, the Gauss-Markov theorem states that a statistically optimum estimate \hat{f} is obtained by minimizing $V(f)$ when $\Gamma_1 = S_\epsilon^{-1}$ and $\Gamma_2 = S_f^{-1}$. Inserting these matrices into (16:14) and solving $V(f) = \text{minimum}$ for f gives

$$\hat{f} = X^{-1} (A^T S_\epsilon^{-1} g + S_f^{-1} f_0) = X^{-1} A^T S_\epsilon^{-1} [g - A f_0] + f_0 \quad (16:15)$$

where

$$X = S_f^{-1} + A^T S_\epsilon^{-1} A.$$

It can be shown that the resulting covariance matrix of $f - \hat{f}$ is X^{-1} . The *expected-mean-square* error $E [(f - \hat{f})^T (f - \hat{f})]$, is then given by

$$E [(f - \hat{f})^T (f - \hat{f})] = \text{trace } X^{-1}. \quad (16:16)$$

Thus the trace in (16:16) may be obtained without solving (16:9) as an estimate of the expected accuracy of the estimate \hat{f} under the prevailing statistical conditions. The details of the statistical method may be found in the references cited above. Further generalizations have been made by Franklin (1969), Wahba (1969), Westwater (1970), and Rodgers (1971).

16.3.3 Iterative Techniques (Landweber, 1951; Fleming and Smith, 1971; Conrath and Revah, 1971)

In the methods discussed in this section an initial guess $f^{(0)}$ for f in (16:3) (or (16:9) in the discrete case) is successively improved to generate a sequence $f^{(1)}, f^{(2)}, \dots$ of functions (or vectors) such that the transforms of $f^{(n)}$ approach g as $n \rightarrow \infty$. The success of these methods depends on the fact that highly oscillatory functions (or vectors) $f^{(n)}$ do not usually begin to appear for any reasonably small value of n ; it is usually possible to approximate g (within reasonable measurement error) before such functions appear. Thus some smoothing is automatically provided if the process is terminated when the residuals are comparable in size to the assumed measurement errors in g . Iterative techniques have been successfully used to obtain atmospheric temperature profiles from satellite radiometer measurements (Fleming and Smith, 1971; Conrath and Revah, 1971). An advantage of iterative methods is that they may be applied with only slight modification to non-linear equations.

The method of Landweber (1951) was originally presented to solve the complete non-discreted equation (16:3) in the case where the right-hand side is an exact function for which the solution exists. We summarize it here. Let the interval of observation of $g(x)$ be $a \leq x \leq b$ (this can be accomplished by a linear transformation of x). Write the integral in (16:3) in the abbreviated operator form

$$Kf = \int_a^b K(x, y) f(y) dy. \quad (16:17)$$

Then we may define the *adjoint operator* K^* by means of

$$K^*g = \int_a^b K(x, y) g(x) dx. \tag{16:18}$$

(Note that the integration is with respect to x , not y .) Then the Landweber iteration is defined by

$$f^{(n)} = f^{(n-1)} + K^*(g - K f^{(n-1)}), n=1,2, \dots,$$

or

$$f^{(n)} = f^{(n-1)} + K^*(\text{residual using } f^{(n-1)}), \tag{16:19}$$

where $f^{(0)}$ is a suitable starting function. ($f^{(0)} = 0$ can be used.) The following facts can be proven. One may refer to Smithies (1958) for a discussion of the concepts involved.

- a. If (16:3) has a solution f for the given right-hand side g , if $f^{(0)}$ has the form K^*h for some h , and if all eigenvalues λ of the operator K^*K (i.e., numbers λ such that $K^*Kv = \lambda v$ has a nontrivial solution v) are such that $0 < \lambda < 2$, then $f^{(n)} \rightarrow f$ as $n \rightarrow \infty$.
- b. For increasing n the eigenfunctions v for which λ is closest to 1 appear first in the solution, and the v 's corresponding to λ 's near 0 or 2 appear only after many iterations. This effect provides a natural smoothing.
- c. If g has errors, only those parts of the error function projecting onto v 's such that $\lambda \cong 1$ will appear for small n . Thus the procedure is not too vulnerable to errors in g .

The corresponding version of the Landweber iteration for the discretized equation (16:9) is immediate: just let $f, f^{(0)}, f^{(1)}, \dots$, be vectors and replace K by the matrix A and K^* by A^T . Then the iteration becomes

$$f^{(n)} = f^{(n-1)} + A^T(g - Af^{(n-1)}), \tag{16:20}$$

where $f^{(0)}$ is a suitable initial vector.

A modification of (16:20) which seems to improve the convergence rate, at least when all $a_{ij} \geq 0$, as in many applications, has been proposed and successfully used in satellite temperature probing by Smith (1970). This modification consists of replacing A^T in (16:20) by DA^T , where $D = \text{diag}(d_{11}, d_{22}, \dots, d_{mm})$,

$$d_{jj} = 1 / \sum_{i=1}^n a_{ij}, j = 1, 2, \dots, m,$$

and $A \equiv (a_{ij})$. Smith's modification reads

$$f^{(n)} = f^{(n-1)} + DA^T(g - Af^{(n-1)}). \tag{16:21}$$

Other iterative methods exist, notably that of Chahine (1970), but they are not discussed in this short summary.

16.4 Examples of Profile Retrieval by Indirect Sensing

In (16.1), we discussed radiometric determination of temperature profiles. To illustrate some of the inversion methods described above, we will present examples of indirect retrieval of temperature profiles.

Ground-based determination of low altitude temperature profiles by angular scan microwave measurements has been investigated experimentally by Mount et al. (1970). Statistical inversion techniques have been applied to these data by Westwater (1971). An example of a retrieval is shown in (F16.3). The angular brightness temperature data at 54.5 GHz were taken at Cincinnati, Ohio, while a priori statistics were gathered from Dayton, Ohio, roughly 50 miles away. The ensemble consisted of five years of twice-a-day radiosonde soundings. The direct measurement of the surface temperature was used as an exact constraint on the solution profile, and, in addition, the original statistics were modified to those conditional on known surface conditions. Note the closeness of the solution, \hat{T} , to the directly measured radiosonde profile, T , in the region of good coverage by the weighting functions (F16.2). The elevated inversion at 2.5 km is above this region and is smoothed out by the solution.

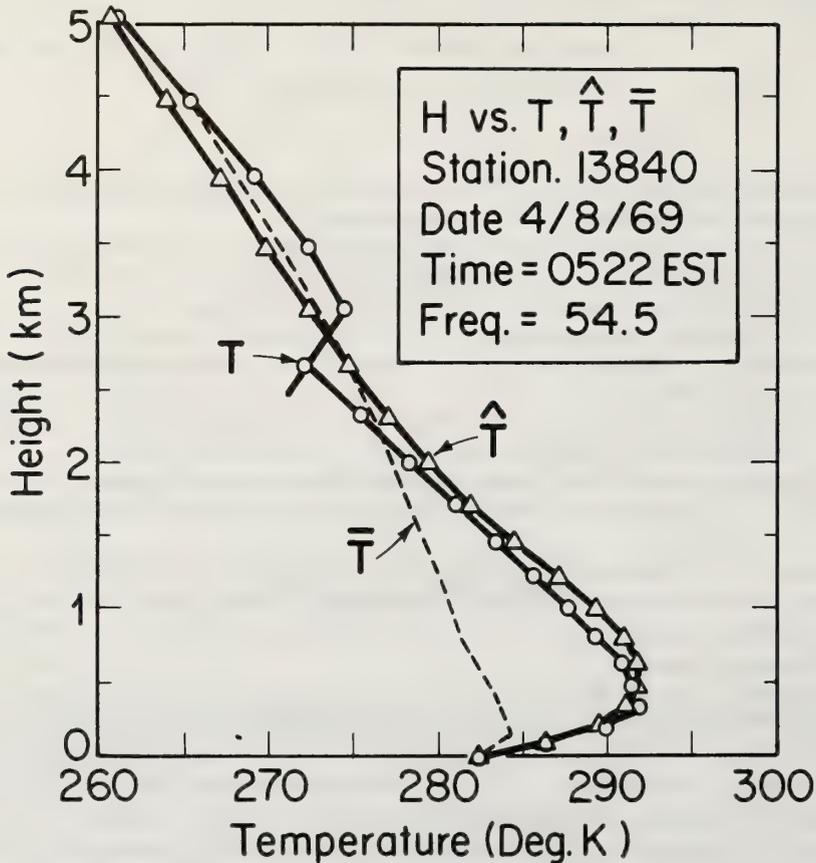


Figure 16.3 *Temperature profile derived from radiometer measurements at Cincinnati, Ohio.*

At the present time, global determination of vertical temperature profiles from satellite observations is a reality (Wark and Hilleary, 1969). Infrared observations have been successfully inverted by a priori statistical techniques (Wark and Hilleary, 1969), regression techniques (Smith, Woolf, and Jacob, 1970), and by several of the non-statistical techniques (Conrath and Revah, 1971; Chahine, 1970; Shaw et al., 1970). An example of profile inversion using the non-statistical Landweber iteration algorithm is shown in (F16.4) (Conrath and Revah, 1971). The radiance data were obtained with the Nimbus 4 IRIS experiment (Hanel and Conrath, 1970) at selected spectral intervals in the $15 \mu\text{m}$ CO_2 band. This particular example was chosen to illustrate the dependence of an iterative solution on the initial guess. The solution obtained with an initial isothermal atmosphere does not recover the structure at the tropopause (100 mb level). This structure is already present in the climatological first guess and is retained through subsequent iterations. In general, structure at spatial

frequencies greater than those inherent in the weighting functions cannot be recovered from the radiance observations alone.

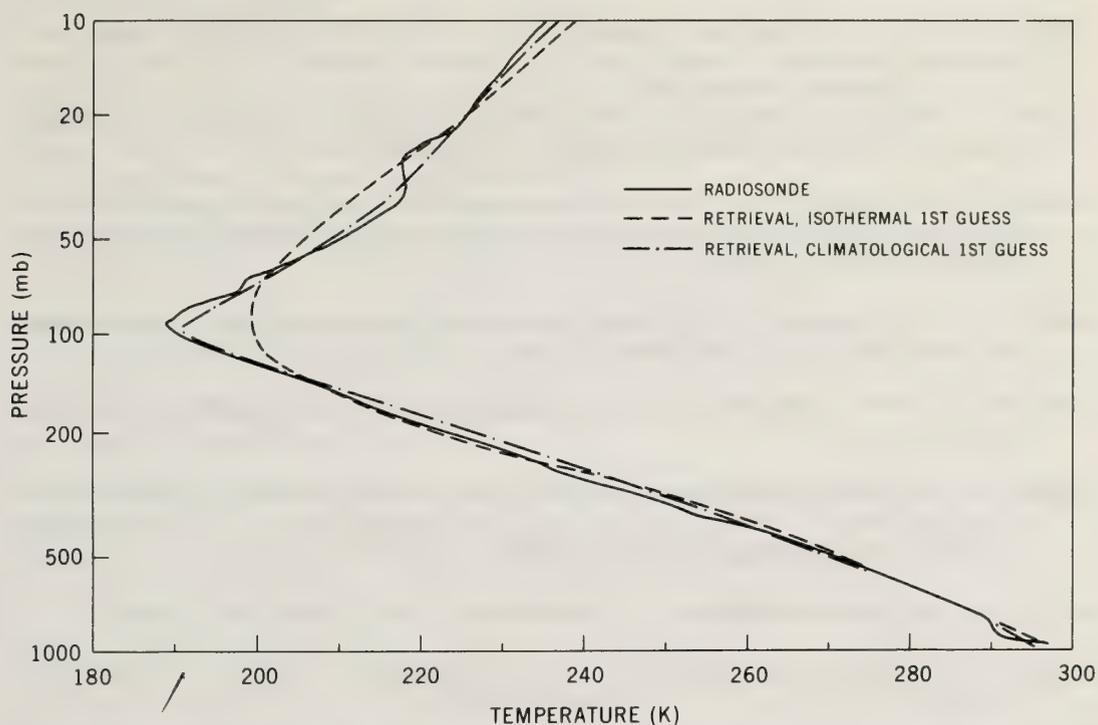


Figure 16.4 *Temperature profile derived from multispectral infrared satellite observations using the Landweber iteration technique. (Courtesy of Barney J. Conrath, NASA Goddard Space Flight Center.)*

16.5 Conclusions and Final Remarks

The evolution of the radiometric inversion problem has been a turbulent one. Initial attempts at retrieving profiles were frequently unsuccessful, because of instabilities occurring in the presence of instrumental noise. Introduction of smoothing methods achieved stability, but a quantitative estimate of the correct amount of smoothing was not always available. Statistical methods provided an optimized smoothing criteria, but required a not always available history of direct data. More recently, iterative methods have provided useful results, without a priori data. Problems still exist for iterative techniques in the dependence of the solution on the number of iterations and on the initial guess. A major problem in indirect radiometric sensing is still the inversion of data contaminated by clouds, although Rodgers (1971) has approached this problem statistically. With all the progress achieved on the clear air inversion problem in the past decade, it does not seem overly optimistic that this problem can also be overcome.

16.6 Acknowledgements

The authors thank Vincent Falcone of Air Force Cambridge Research Laboratories, for his suggestions on the manuscript. Thanks are also due to Dr. Barney Conrath of NASA Goddard Space Flight Center for providing (F16.4) and to Dr. Paul Wacker of the National Bureau of Standards for critically reviewing the manuscript.

16.6 References

- Backus, G. E., and J. F. Gilbert (1967), Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astr. Soc.* *13*, 247-276.
- Chahine, M. T. (1970), Inverse problems in radiative transfer: Determination of atmospheric parameters, *J. Atmos. Sci.* *27*, 960-967.
- Conrath, B. J. (1969), On the estimation of relative humidity profiles from medium-resolution infrared spectra obtained from a satellite, *J. Geophys. Research* *74*, No. 13, 3347-3361.
- Conrath, B. J., and I. Revah (1971), A review of non-statistical techniques for the estimation of vertical atmospheric structure from remote infrared measurements, Presented at the "Workshop on the Mathematics of Profile Inversion" at Moffett Field, California, July 1971.
- Dave, J. V. (1971), Determination of size distribution of spherical polydispersions using scattered radiation data, *Appl. Optics* *10*, No. 9, 2035-2044.
- Fleming, H. E., and W. L. Smith (1971), Inversion techniques for remote sensing of atmospheric temperature profiles, to appear.
- Franklin, J. N. (1969), Well-posed stochastic extensions of ill-posed linear problems, California Institute of Technology, Willis H. Booth Computing Center, Programming Report No. 135 (Invited address presented to the Soc. of Industrial and Applied Mathematics, April 19, 1969).
- Goody, R. M., 1964: *Atmospheric Radiation, I. Theoretical Basis*. Oxford University Press, London.
- Landweber, L. (1951), An iteration formula for Fredholm integral equations of the first kind, *Amer. J. Math.* *73*, 615-624.
- Parker, R. L. (1971), Part 1: The Backus-Gilbert method for inverse problems, presented at the "Workshop on the Mathematics of Profile Inversion" at Moffett Field, Calif., July 1971.
- Phillips, D. L. (1962), A technique for the numerical solution of certain integral equations of the first kind, *J. Assn. Computing Machinery* *9*, 84-97.
- Rodgers, C. D. (1966), Satellite infrared radiometer – A discussion of inversion results, Memorandum No. 66.13, Clarendon Lab., Oxford University.
- Rodgers, C. D. (1971), Statistical retrieval techniques for sounding the meteorological structure of the atmosphere, presented at the "Workshop on the Mathematics of Profile Inversion" at Moffett Field, Calif., July 1971.
- Smith, W. L. (1970), Iterative solution of the radiative transfer equation for the temperature and absorbing gas profile of an atmosphere, *Appl. Optics* *9*, No. 9, 1993-1999.
- Smithies, F., 1958: *Integral Equations*. Cambridge University Press, London.
- Staelin, D. H. (1969), Passive remote sensing at microwave wavelengths, *Proceedings of IEEE* *57*, 427-439.
- Strand, O. N., and E. R. Westwater (1968a), Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind, *J. Assn. Computing Machinery* *15*, 100-114.

- Strand, O. N., and E. R. Westwater (1968b), Minimum rms estimation of the numerical solution of a Fredholm equation of the first kind, S.I.A.M. J. of Numerical Analysis 5, 287-295.
- Tihonov, A. N. (1963), Solution of incorrectly formulated problems and the regularization method, Soviet Math. Dokl 4, 1035-1038.
- Twomey, S. (1963), On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, J. Assn. Computing Machinery 10, 79-101.
- Wahba, G. (1969), On the numerical solution of Fredholm integral equations of the first kind, Univ. of Wisconsin, Dept. of Statistics, Report UWIS-DS-69-217.
- Wark, D. Q., and D. T. Hilleary (1969), Atmospheric temperature: successful test of remote probing, Science 165, 1256-1258.
- Westwater, E. R. (1970), Ground-based determination of temperature profiles by microwaves, Ph.D. Thesis, Dept. of Physics, University of Colorado, Boulder, Colorado 121 pp.
- Westwater, E. R. (1972), Ground-based determination of low altitude temperature profiles by microwaves, Monthly Weather Review 100, 1, 15-28.
- Whittaker, E. T. and G. N. Watson, 1963: *A Course of Modern Analysis*. Cambridge University Press, London.
- Wilkinson, J. H., 1965: *The Algebraic Eigenvalue Problem*. Oxford University Press, Fair Lawn, N. J.

List of Symbols

A	Matrix representation of kernel	ν	Frequency
n	Number of measurements of g	I_ν	Radiative intensity or radiance
m	Number of quadrature points of f	B_ν	Planck Function
U(f)	Generalized quadratic forms used in	T	Absolute Temperature
V(f)	profile inversion	α_ν	Absorption coefficient
γ_1, γ_2	Smoothing parameters	h, h'	Height
B	Matrix representation of derivative operator	$f_\nu(\rho[h])$	Non-linear integrand of integral equation
Γ_1, Γ_2	Positive definite matrices	$K(\nu, h)$	Weighting function for vertical temperature profiles
S_f	Covariance matrix of unknown function f	δI_ν	Difference of measured and calculated radiance
S_ϵ	Covariance matrix of experimental errors	$K(x, y)$	Kernel of integral equation
E	Expected value operator	f(y)	Solution of integral equation
\bar{x}^{-1}	Covariance matrix of statistical solution of integral equation	g(x)	Measured quantity in integral equation
K^*	Adjoint operator of K	κ_ν	Mass absorption coefficient
λ	Eigenvalues of K^*K	β_ν	Residual absorption coefficient
f ⁽ⁿ⁾	Iterative solution to f at n th iteration	ρ	Mass density
D	Diagonal matrix used to improve convergence in Landweber's iteration scheme	$\delta\rho$	Difference of true and estimated density
		\hat{f}	Estimate of f

Chapter 17 ATMOSPHERICS AND SEVERE STORMS

William L. Taylor

Wave Propagation Laboratory

Environmental Research Laboratory

National Oceanic and Atmospheric Administration

The identification of severe storms and the selection of electrical parameters indicative of tornadic activity can be made only after developing an understanding of the many facets of thunderstorm electrification and the characteristics of atmospherics. A brief review relative to thunderstorm electricity, the electromagnetic radiation from lightning discharges and the effects of radio propagation reveal many areas where our knowledge of some aspects of these subjects is surprisingly limited. Some of the problems involved with the tracking of thunderstorm regions and with the selection of practical parameters to monitor severe storm activity in general and tornado activity in particular are extremely difficult to solve. The results of all recent work indicate that most tornadoes have major radio-frequency electrical activity associated with them. It is suggested that the parameter most indicative of tornadic activity is the number of bursts of high impulse counting rates at frequencies above about 1 MHz.

17.0 Introduction

Considerable effort has been devoted in the last few decades to the identification and clarification of mechanisms responsible for the production of electric charge centers within clouds, and those responsible for the reduction of these charges through the subsequent current surges of lightning flashes. Our knowledge of thunderstorm electricity certainly has been greatly expanded since the days of Benjamin Franklin. Many facets of electrification and lightning are well known. And yet, we remain shockingly ignorant of many processes involved in the generation, nurture and dissipation of thunderstorms.

No attempt is made to give a history of research or to review our current knowledge in the various fields of research touched upon in this paper. Only a general overview of thunderstorm electricity and the characteristics of atmospherics is presented in (17.1) and (17.2). A more detailed account of recent work concerning the identification and tracking of severe storms in general and of tornadoes in particular by remote sensing techniques is presented in (17.3).

17.1 Thunderstorm Electricity

A thunderstorm cloud is characterized by lightning and the resulting thunder from which it derives its names. Although other clouds can produce appreciable electric fields and sometimes lightning, it is the cumulonimbus formation that we look upon for lightning activity. Such clouds are characterized by nearly vertical chimneys of convective updrafts extending to heights of 7 km or so, but often reaching to 16 or 18 km for a severe storm. A thunderstorm frequently exhibits a cellular nature which may have as many as 10 or more individual cells that are in various stages of activity. A cell is a single organized updraft within a thunderstorm cloud. These cells appear to be the units of convection, precipitation and electrification. Large thunderstorm complexes along frontal zones may contain 50 or more cells, each of the order of 1 km in diameter with many cells or composites of cells active simultaneously.

17.1.1 Electric Charge Concentration

Electric charges are somehow separated by the convection and precipitation processes and appear to be strongly associated with the freezing level of the cloud. The freezing level is the height where the temperature is 0°C. The work of Imyanitov, et al., (1969), Kasimir (1965), Malan (1965) and Workman (1965) are a few examples of attempts to integrate the tremendous amount of detail into a complete picture of thunderstorm electrification. Yet, we can do little more than guess at the actual amount of positive and negative charges and their distributions with respect to height and time inside a thundercloud.

It is generally believed that the upper part of the cloud contains a net positive charge, the lower part contains a net negative charge, while often a small secondary positive charge volume is located near the very bottom part of the cloud base. The total quantity of separated charge is believed to range from 20 coulombs or so to a 100 coulombs or more at the time the cloud is sufficiently charged to produce lightning. Impressive as the lightning discharge of tens of thousands of amperes may be, it only requires a continuous charging current of about 1 ampere to provide for the lightning discharge current in an average thunderstorm.

17.1.2 The Lightning Discharge

The electrical activity of the thunderstorm is very complex. From the recent work of Berger (1967), Kitagawa (1965), Mackerras (1968), and Ogawa and Brook (1964), for example, we are able to construct a reasonably clear picture of the lightning discharge process. Prior to the first lightning stroke, many smaller sparks and minor streamer processes begin to take place. A streamer is a low current discharge slowly progressing between successive charged volumes. Eventually, 15 or 20 minutes after the beginning of charge separation, sufficient charge becomes concentrated to ionize the air through leader processes. A leader is a low current discharge ionizing the air either by small intermittent steps or by a long continuous dart. The first leader is initiated in a region of high electric field and moves in a direction of large potential gradient through a series of rapid, successive steps. This stepped leader advances a distance of perhaps 50 m every 50 μ s or so and occasionally branches until a number of ionized channels are formed. If a concentration of charge is encountered by the stepped leader a surge of current is generated. Such encounters as this along the stepped leader path within the cloud is called an intracloud discharge. The stepped leader sometimes progresses outside the cloud and terminates considerable distance away in clear air. This is called an air discharge.

The cloud usually contains a net negative charge while the earth acquires a positive charge in the proximity of the cloud. A large potential gradient between the cloud and the earth will often produce a stepped leader to earth. This is immediately followed by a very large surge of positive current flow from the earth to the cloud charge center. This return-stroke causes the intense luminosity of the lightning flash. Very soon after the first-return stroke, the dart leader will advance out of the cloud along the previous discharge channel at a constant speed of about 1/3 that of light. This will connect other principally negatively charged volumes within the cloud to the earth and initiate another return-stroke. This sequence of events may recur every 20 ms or so for half a second or longer until there is insufficient charge available for a dart leader to re-ionize the channel. A complete lightning discharge may contain from 2 to 20 or more return-strokes with many small current surges between strokes and after the last stroke. There may be also what is called a continuing current, i.e., a slowly decaying discharge current, amounting to a few hundred amperes associated with one or more return-strokes in a cloud-ground discharge.

17.2 Characteristics of Atmospherics

Current surges during lightning discharge processes produce a redistribution in the charge concentrations in the cloud and thus cause a change in the electrostatic field. Of course, any variations in current produce magnetic induction fields and electromagnetic fields. An often used representation of the total instantaneous electric field at distance d in meters from the discharge is

$$E_T = E_S + E_I + E_R = \frac{1}{2\pi\epsilon_0} \left[\int \frac{Mdt}{d^3} + \frac{M}{cd^2} + \frac{dM/dt}{c^2 d} \right] \quad (17:1)$$

where the subscripts S, I, and R and the three terms in the bracket represent respectively the static, induction and radiation components, ϵ_0 is the permittivity of free space ($1/36\pi \times 10^9$ farad per meter), M is the current moment in ampere-meters (instantaneous product of current and discharge length), and c is the velocity of light (3×10^8 m/s). This formulation applies to a dipole source in which d is very large compared to the dimensions of the dipole, and for which the three terms would be equal in magnitude at a distance $d = c/2\pi f$, where f is the frequency in Hertz. An atmospheric is any observed change in the electric field given by (17:1) produced by lightning discharge processes.

It is readily apparent that (17:1) has limited applicability since M must be the instantaneous sum of all currents in the discharge evaluated at retarded values of time ($t-d/c$). Some practical use of (17:1) has been realized in the VLF band (very low frequency = 3-30 kHz) at distances from 10 km or so to several 100 km, and in the ELF band (extremely low frequency = 30-3000 Hz) at distances exceeding 1000 km. Only the radiation term is important at these lower frequencies and at distances greater than about $1/6$ of the wavelength and thus (17:1) can be simplified to

$$E = \frac{2(dM/dt)}{10^7 \times d} \quad (17:2)$$

Both (17:1) and (17:2) are inapplicable for frequencies above the VLF band.

The electrostatic field amplitude decreases very rapidly with distance, but still has measurable values to about 100 km. A considerable amount of information can be inferred about discharge mechanisms from the measure of this field (Mackerras, 1968). At very close distances, charge quantity and location of charge within the cloud can be estimated (Ogawa and Brook, 1964).

The induction term also decreases rapidly in amplitude with distance. This magnetic component is directly proportional to the current moment. Very close to the lightning channel, where the discharge length is greater than d , the induction field is proportional to the reciprocal of the distance from the current. Discharge channel currents can be readily estimated from such measurements (Williams and Brook, 1963).

Many other measurements may be made on lightning discharges in addition to the electric and magnetic field measurements. Acoustic measurements, direct measurement of current to structures, photographic measurements and spectroscopic measurements are among the most fundamental. It is not within the scope of this presentation to consider such phenomena.

The remaining discussion on characteristics of atmospherics will be limited to the radiation component of the field. In addition, only the vertical component of the electric field will be considered, since the vertical electric component of the electromagnetic field is more stable and predictable than other field components when monitored from the earth's surface.

17.2.1 The Source Function

A surge of current along a conductor will radiate an electromagnetic signal. The characteristics of this signal are determined by the variations of the current moment with time. If the signal is analyzed into Fourier components, the resulting amplitude spectra will extend over a very wide band of frequencies. The spectral amplitude at frequencies less than the frequency of the maximum amplitude of the spectrum is determined largely by the decay of the current moment, while the amplitude at frequencies above the peak is a function of the risetime of the current moment. In general, the longer the conductor the lower the frequency at which the maximum amplitude component occurs.

A typical lightning return-stroke may have a current maximum of 20,000 A at ground level within a few μs after initiation. The current surge will move up along a channel determined by a prior stepped leader or dart leader at a velocity of about $1/3$ that of light or less, and reach a total length of 4 or 5 km. Such a stroke will produce a current moment maximum at about 50 μs after initiation. The radiation field waveform that would be observed with wideband equipment will attain a peak positive amplitude within about 1 μs of the beginning, return to a zero field value at 50 μs or so, reach a negative amplitude sometime later and then slowly return to a final zero field value. Spectrum analysis of such a pulse shows a maximum amplitude around 5 or 6 kHz. The amplitude of the spectrum at lower frequencies is approximately proportional to the square root of frequency which results from the approximate exponential decay of current. Spectral amplitude at higher frequencies is approximately proportional to the reciprocal of the frequency through a few MHz. At still higher frequencies, the finite risetime of the waveform will cause the spectral amplitude to decrease approximately as the reciprocal of the frequency squared. Additional details on the VLF radiation of atmospherics can be found in the theoretical work of Dennis and Pierce (1964) and in the experimental work of Taylor (1963).

ELF radiation components produced by the large, short duration current surges are at least an order of magnitude smaller than the corresponding VLF components. Since ELF components are often observed

with amplitudes comparable with the VLF components, it is believed that the continuing currents are the primary sources of larger ELF signals. (See for example, Kitagawa, et al., 1962; Pierce, 1963; and Sao, et al., 1970). The peak magnitude of these currents are only of the order of 100 A but 10 to 100 ms or more may be required for the current to decay to zero.

It is very impractical to attempt to observe the characteristics of atmospherics with a single channel of wideband equipment because of the large variations in amplitude and the wide range of time elements characteristic of different frequency components radiated from lightning discharges. The ELF, VLF and higher frequency components are more easily observed using different equipment channels that are optimized in response time and sensitivity to separate identifiable and analyzable components within each frequency band from unwanted background of interference. The bandwidths must be narrow enough to minimize the effects of interfering signals and yet must be wide enough to preserve the amplitude and time distributions of the wanted signals.

The characteristics of the electromagnetic radiation from a nearby lightning discharge varies drastically with frequency as has been reported by Oetzel and Pierce (1969). The responses of restricted bandwidth receivers tuned to frequencies less than about 30 kHz are discrete and obviously associated with the individual return-strokes. As frequency increases, the number of responses increase until at about 1 MHz there are many responses of comparable amplitude to the few recognizable responses associated with the return-strokes. Above 1 MHz, the number of responses further increase until they are almost continuous at 50 MHz or so at rates that may exceed 10^5 s^{-1} . The number of pulses tend to decrease at higher frequencies.

Responses of restricted bandwidth receivers, with bandwidths proportional to observing frequency, will be about the same value through a frequency range from about 10 KHz to 1 MHz or so from a return-stroke. The responses at higher frequencies will decrease approximately as the reciprocal of the frequency. The frequency at which the responses change to a reciprocal frequency relationship depends on the current moment variations. The shorter length current surges of the various intracloud processes tend to produce equal value receiver responses to much higher frequencies than were attained by the return-stroke prior to producing a reciprocal frequency relationship. This assumes, of course, that all radiation impulses from the discharge mechanism are separated in time relative to the response time of the receiver. The fact is, that regardless of the receiver bandwidth, there will be some impulses interfering with each other when they reach the receiver input.

The responses of narrow band receivers with bandwidths less than about one percent of the observing frequency will be primarily that of the complex sum of the spectra of many impulses taken over the response time of the receiver. These receivers will produce responses to a lightning discharge that will in general decrease in amplitude proportional to an inverse frequency relationship over a band of frequencies extending from about 10 kHz to 10 GHz. Considerable deviations from this simple relationship are experienced in practice as would be expected in the response to the radiation from a source as complex as a lightning discharge.

Very generalized average spectra of lightning discharge processes are shown in (F17.1) to aid in visualizing the distribution of the frequency components from the various processes. The spectral amplitude may vary by an order of magnitude from what is indicated here during a single discharge. The two spectra represented as "other intracloud processes" should not be considered as resulting from discrete processes or as limits in the spectrum of intracloud processes, but rather considered as representing two of a large family of spectra produced during a discharge.

A single lightning discharge may include several return-strokes that produce vertical electric fields of about 30 v/m at a distance of 10 km and corresponding peak amplitude spectra of about $10^3 \mu\text{Vs/m}$ (microvolt-seconds per meter) with over 90 percent of all flashes producing less than a 14 dB variation from these values. As the frequency of observation is increased, it becomes obvious that the radiation source changes from a few coherent impulses from return-strokes, to a great quantity of incoherent impulses or noise from other processes. The correlation between receiver responses at different frequencies, for both amplitude and time, becomes very low as the wavelength of the observations approaches the dimensions of the discharge.

17.2.2 Effects of Propagation

The radiation component of the vertical electric field can reach a receiver located on the earth's surface by several means (Jordan, 1950). Very close to the source, the radiation propagates in a straight

line until a conductor such as the earth's surface or the ionosphere is encountered. This line-of-sight propagation suffers little loss within the band of frequencies involved here. The maximum range for this type of propagation is limited by the earth's curvature and the height of the source. The resultant electric field at the receiving antenna can be expressed in the form

$$E(\text{space}) \propto \frac{e^{-j\beta d_1}}{d_1} + R \frac{e^{-j\beta d_2}}{d_2} \tag{17:3}$$

where the exponential terms are the phase factors, $\beta = 2\pi/\lambda$, R is the reflection coefficient of the earth's surface, d_1 and d_2 are the distances from the source and its image, respectively, to the receiver. The first term is the direct component or ray from the source to the antenna and the second term is the ray from the source reflected from the ground and thence to the antenna. The amplitudes of these two rays are approximately equal and are in phase so that the received field is about twice the incident field when the rays arrive at the antenna at a large angle of elevation above the horizon. At very low elevation angles, the order of a degree or so, the phase of the reflected component becomes almost 180° out of phase with the direct component and the resultant field is very small. The approximate losses for line-of-sight propagation of a ray that is 2 degrees above the horizon is shown as curve E in (F17.2).

Propagation past the line-of-sight range is primarily along the surface of the earth to some greater distance. The field of this groundwave component is inversely proportional to distance d , times an attenuation factor F , which is a function of surface conductivity and frequency. The field is of the form

$$E(\text{surface}) \propto F(1 - R) \frac{e^{-j\beta d}}{d} \tag{17:4}$$

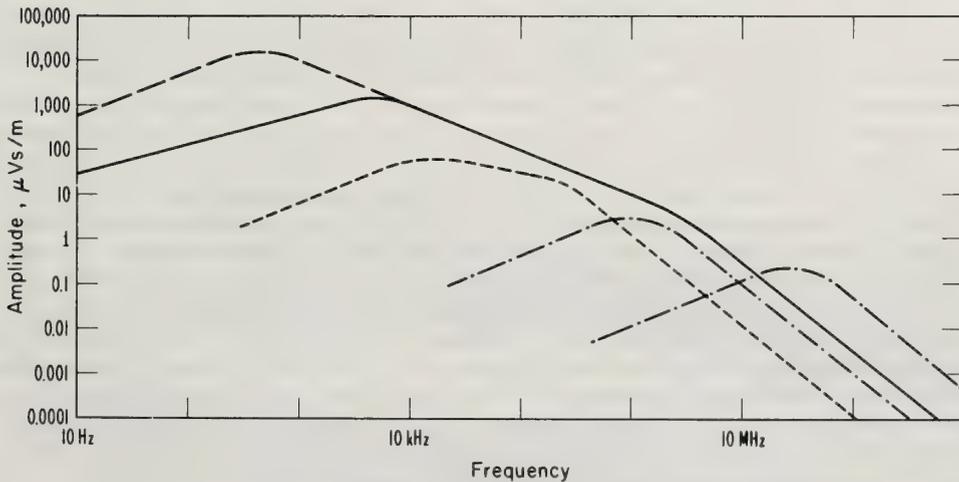


Figure 17.1 Amplitude spectra of the radiation component for lightning discharge processes relative to 1 Hz bandwidth and normalized to 10 km.

- return stroke
- - - - - continuing current supplement
- - - - - leader strokes
- · · · · other intracloud processes

Losses due to attenuation along a path of even good conductivity are appreciable at the higher frequencies. For example, the groundwave field will be attenuated about an additional 1 dB, 20 dB, 70 dB and 120 dB, respectively, at 100 kHz, 1 MHz, 10 MHz and 100 MHz below the inverse distance term at 100 km from the source for an average surface conductivity of $\sigma = 10^{-2}$ mhos/m. The lower frequency components of a curve representing this attenuation is shown as D in (F17.2). Poor ground conductivity will produce even greater attenuation. It is clear that only the frequencies below a few hundred kilohertz will propagate very far via the groundwave. The total loss for the groundwave in excess of the inverse distance term for 1000 km is shown in (F17.2) as curve C.

Atmospheric waveforms in the VLF region are usually considered in terms of individual pulses arriving via the groundwave and ionospherically reflected rays to ranges of about 2000 km. Waveform characteristics are primarily determined by the height and reflection coefficient of the D-region of the ionosphere. The VLF reflection coefficients are of the order of 0.5 during the day and somewhat larger at night. This means that the "one hop skywave" (once reflected ray) will be 15-20 dB smaller in amplitude than the groundwave at a distance of 100 km. At about 500 km the amplitudes will be about equal, and at 1000 km the skywave amplitude will exceed that of the groundwave by 6-10 dB.

VLF propagation to distances greater than 2000 km is more easily explained in terms of waveguide modes in which the earth and the ionospheric D-region are considered as walls of a spherical waveguide (Wait, 1962). To very great ranges where the first-order mode is dominant, the vertical electric field is

$$E = A_{(f)} L_{(f)} (a \sin d/a)^{-1/2} \exp(-\alpha_{(f)} d) \quad (17:5)$$

where $A_{(f)}$ is an amplitude coefficient dependent on the spectrum of the source, $L_{(f)}$ is a mode launching term or excitation factor, a is the radius of the earth, d is the great circle distance and $\alpha_{(f)}$ is the attenuation in nepers. Minimum daytime attenuation of atmospherics (Taylor, 1960) is usually only 1 or 2 dB/1000 km (where dB/1000 km = 8.68×10^3 nepers) in the center of the VLF band which means that the larger atmospherics can be observed after propagating around the world. Curve B in (F17.2) shows the general characteristics of the first-order wave guide mode expressed in dB/1000 km. There is little change in total transmission loss between day and night in the center of the VLF band (Taylor, 1967).

Propagation in the ELF band is via the zero-order mode in the earth-ionosphere waveguide which can be represented also by (17:5). It is this type propagation that produce the "slow-tail" portion of atmospherics (Taylor and Sao, 1970). The phase and group velocities for ELF propagation are less than the velocity of light and this causes the ELF "tail" to be delayed and thus separated from the VLF "head" of an atmospheric. The attenuation in dB/1000 km for the zero-order mode is shown as curve A in (F17.2).

Lightning discharge signals in the upper LF, the MF and HF bands reach ground receivers at distances greater than the limits of the groundwave component almost entirely by ionospherically reflected rays (Davies, 1965). At frequencies below a few MHz, the signals are greatly attenuated in passing through the D-region along a ray path to and from the E- or F-region of the ionosphere. Higher frequencies, above 20 MHz or so, pass through the ionosphere and are not returned to the earth. Temporal variations are large and during periods of ionospheric disturbances no useful signals above a few hundred kilohertz may be returned to the earth because of increased penetration of the ionosphere by the higher frequencies and increased absorption in the D-region by the lower frequencies. A representation of equivalent attenuation in dB/1000 km for good daytime conditions is shown in (F17.2) as curve F.

17.3 Sensing Severe Storm Activity

A thunderstorm is classified as severe by the National Severe Storms Forecast Center in Kansas City whenever it produces one or more severe storm criteria of heavy rain, large hail, high winds, extreme turbulence, intense lightning or tornadoes. When a tornado or a funnel cloud is present, usually one and sometimes all of the other criteria are also present within the area covered by the severe storm. Sanders (1971) reported that during the years 1963-1970 there were 2394 deaths caused by cumulus convective storms out of a total 3783 weather related deaths in the United States. Lightning caused 659 deaths and tornadoes were responsible for 868 deaths. It is interesting to note that hurricanes and other tropical storms caused only 467 deaths. Because the tornado is a big killer, a frightening spectacle, and a definite

manifestation of a severe storm (as opposed to criteria determined by degree as for rain, hail and wind), the following remarks are generally limited to the severe storms associated with tornadoes.

The concentration of energy within the relatively small volume of a tornado has been the subject of much investigation by prominent scientists for more than a century, but this concentration has not been explained in terms of the energy budget. The power required to drive a tornado vortex has been estimated to exceed 10^8 kW (Vonnegut, 1960), yet this is small compared to the energy available in a single thunderstorm cell which may exceed 10^9 kW (Braham, 1952). Various theories have been proposed over the years suggesting that a tornado may be produced by the concentration of kinetic energy resulting from the conservation of angular momentum through some not yet understood mechanism associated with the intense electrical activity associated with severe storms.

Jones reported that the occurrence rate of atmospherics in the 10 kHz region of the spectrum increased as the intensity of a thunderstorm increased, but decreased to a relatively small value prior to a tornado formation. However, the occurrence rate at 150 kHz greatly increased during the formation of a tornado. More recently, the effects observed on television sets tuned to channel 2 (54 MHz) have been reported by Waite and Weller (1969) and by Biggs and Waite (1970).

Many eyewitness accounts of unusual electrical activity in and around tornadoes have been reported during the last 20 or 30 years. Jones (1950) gives accounts of lightning and thunder being decidedly different during a tornado than during ordinary thunderstorms, for example, the presence of St. Elmo's fire in the vicinity of a tornado funnel, and the rapid rate of "one stroke right after another" at the base of a cloud just ahead of a funnel. Jones (1965) also reported nighttime observations of approximately circular patches of flashing pale blue illumination originating from within severe storms. Vonnegut (1960) references some observations which are indicative of intense electrical discharges near and within the funnel.

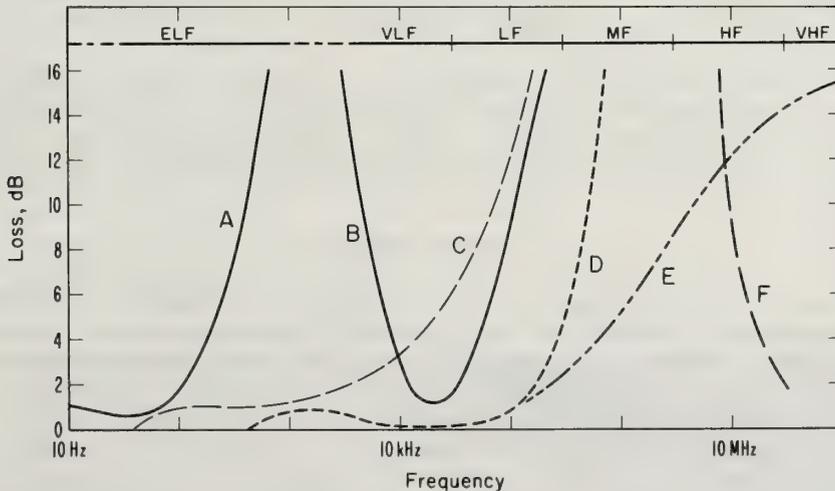


Figure 17.2 *Approximate propagation losses.*

- A — zero order mode attenuation per 1000 km
- B — first order mode attenuation per 1000 km
- C — groundwave, total loss for 1000 km path with $\sigma = 10^{-2}$ mhos/m.
- D - - - - groundwave, total loss for 100 km path with $\sigma = 10^{-2}$ mhos/m.
- E - - - line-of-sight losses for source 2 degrees above horizon, $\sigma = 10^{-2}$ mhos/m.
- F — ionosphere F-region reflection losses

17.3.1 Practical Parameters to Monitor

If appreciable currents do flow either within or closely associated with the tornado funnel, as indicated by Brook (1967), and dispersed into the severe storm cloud, the process might well tend to be almost continuous in the form of corona sparks or a gaseous glow. This would tend to reduce the energy radiated at lower frequencies from return-strokes and increase the energy partitioned into higher frequencies from short-distance, rapid-occurrence dispersive processes within the cloud. A necessary preliminary step is to determine what parameters are practical to monitor and have a high probability of being sensitive indicators of tornadic conditions.

Referring to (F17.1) and (F17.2) and the related discussions in (17.2), a few decisions can be made concerning frequencies to observe and maximum distances over which to monitor. Provided the intracloud processes are enhanced during severe storm and tornado producing conditions, the most affected region of the radio spectrum will be above a few hundred kilohertz. This does not mean that observations at lower frequencies are useless, but rather that they are not likely to contain parameters that are most sensitive to tornadic activity. There should be at least one frequency in the VLF region used primarily to monitor the occurrence of return-stroke discharges. ELF signals from the long, slowly varying discharge currents are unlikely indicators of severe storm activity and are extremely difficult to monitor at close ranges because of the large induction and electrostatic field components. The potential gradient, defined as the electric field at the earth's surface, may be worth observing at very close ranges to monitor the gross movement of electric charge within the storm.

Observations at frequencies from approximately one hundred kilohertz to several hundred megahertz can be reliably made only if the source is in line-of-sight of the receiver. If losses of the order of 20 dB in addition to the inverse distance reduction in field can be tolerated, then signals at frequencies of a few hundred megahertz can be observed from sources 2 degrees above the horizon as indicated in (F17.2). Losses at lower ray path elevations will be greater and, conversely, losses at higher ray path elevations will be less, than shown by curve E in (F17.2).

It is generally believed that the thundercloud electrification processes or charge separation mechanisms are associated with the freezing level of the cloud (see 17.1.1). This is usually estimated to be in the neighborhood of 5 km above the earth's surface in the central and eastern parts of the United States. The height h of the ray path in meters above the earth is given approximately by the relation

$$h = 0.059 d^2 + 17.4 d \theta \quad (17:6)$$

where d is distance in km, θ is the angle of elevation in degrees (for small angles) and the constants are determined for an effective $4/3$ radius of the earth to account for atmospheric refraction. Therefore, for an elevation of 2 degrees, a height of 5 km is attained at about 119 km range. A maximum observing range of 100 km is suggested. This would allow observation of the electromagnetic radiation from electrical processes occurring above the freezing level at elevation angles greater than 2 degrees.

The total dynamic range of amplitude necessary to observe 90 percent of the impulses radiated during a lightning discharge over the frequency band from 10 kHz to 100 MHz and for distances extending to 100 km at greater than 2 degree elevation above the horizon is about 100 dB. This can be reduced in practice to about 50 dB by using separate equipment for different frequency bands and discarding the very small impulses at the greater ranges.

To record and analyze the amplitude responses of each resonant circuit for individual impulses would require elaborate and expensive equipment. A practical solution to this problem is to design observational equipment that does much of the initial data reduction prior to the recording process. The proper selection of parameters would greatly reduce the required recording facilities and minimize the amount of final analysis. It should be realized, however, that to accomplish this requires some prejudgment of the phenomena which may result in discarding significant data which can never be reclaimed.

17.3.2 Tracking Thunderstorm Regions

Many methods have been developed to locate individual lightning strokes and thunderstorm areas. Only the crossed loop direction finder method will be considered here. This method is relatively simple in construction and operation, at some sacrifice of accuracy (Horner, 1954).

The crossed loop direction finder consists of two orthogonally positioned loop antennas in the vertical plane. A cathode ray tube of an oscilloscope is used to conveniently indicate direction of arrival of electromagnetic signals based on the relative amplitude responses of the two loops. The basic 180° directional ambiguity of the system is eliminated through the use of a vertical electric antenna response to unblank or turn on the electron beam of the tube. Errors produced by horizontally polarized components in the received signals and by interference from other atmospheric are minimized by using wide bandwidth or restricted bandwidth channels (not narrow band) and gating the signal so that only the initial part is observed.

Lightning stroke location can be determined through the use of two or more receiving stations in a network. Each station measures the direction of arrival and the origin of the atmospheric is fixed through triangulation. The VLF region of the radio spectrum can be employed to very great ranges with errors of only about ± 2 degrees.

The use of VLF favors observation of the return strokes, while higher frequency direction finder networks can fix other discharge processes. Distant observations at frequencies above the upper LF region would require ionospheric reflected components with resultant large directional errors. Within line-of-sight the errors are small and the location determined quite accurately except at very close ranges. Large errors from high frequency crossed loop direction finders close to the source are produced by the horizontally polarized components which becomes more serious as frequency increases.

Another way of locating the general areas of thunderstorm activity is the use of a single direction finder in conjunction with weather radar. Figure (17.3) shows the precipitation returns on one weather radar frame (a) and a time exposure of the direction of arrival display on an oscilloscope (b) taken from Taylor (1971). The straight arrows pointing up in the upper center of both portions of the figure indicate north and east is toward the right. Range from the center of the radar frame is given in nautical miles. The length of the trace from the center of the direction of arrival display is proportional to the atmospheric pulse amplitude. The more active thunderstorms are determined by the density of traces in the direction of arrival display. The active thunderstorm directions and the corresponding darker portions of the radar frame from which the atmospheric most likely originated are indicated by numbers.

There is not a one-to-one correspondence between thunderstorms detected by precipitation echos on radar screens and the occurrence of detectable atmospheric. For example, thunderstorm areas 2, 3, 6, and 7 are distinctive on the direction of arrival display of (F17.3) and readily located on the radar frame. Some of the other indicated thunderstorms are not as easily related. The large precipitation area between thunderstorms 2 and 3 produced few large atmospheric as was likewise the situation between thunderstorms 3 and 4. Precipitation area 8 produced no large atmospheric. The very small area 1, however, produced a recognizable number of atmospheric.

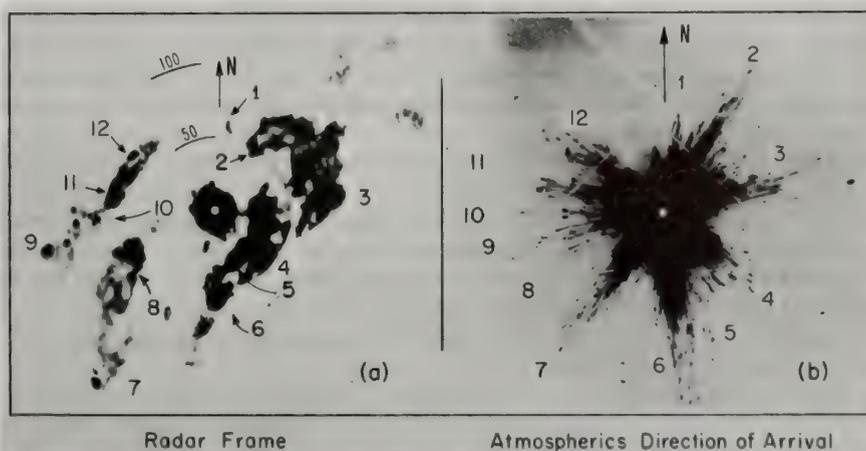


Figure 17.3 Example of radar and direction of arrival: April 29, 1970, 2230 GMT. (a) weather radar frame, (b) atmospheric direction of arrival.

17.3.3 Problems in Identifying Severe Storm Characteristics

The identification of electrical activity associated with tornadoes and severe storms is very difficult because of the many variables in both the observational data and the severe weather reports. The observation and recording operations in the field under actual severe storm conditions are perhaps the most demanding tasks of all. Both the equipment and the operator are required to function under extreme environmental influences. During storm situations, the equipment should continue to function with little or no variations in response sensitivity, observing frequency, bandpass, time constant, threshold level or any other parameter that influences the final result. In addition, the operator should properly adjust the equipment, tabulate operational logs and perform without human error. There are actually few periods when the equipment and the operator perform properly and the data can be accepted without some reservations.

The identification of severe weather has not become an objective science. Weather radar is extensively used for tracking thunderstorm areas, but it cannot independently recognize tornadoes or measure the severity of a storm. It is, however, the best tool presently available. A particular weather condition is classified as severe if it causes damage or is potentially damaging through the presence of one or more of the selected manifestations of destructive weather. These manifestations include tornado, water spout, funnel cloud, hail ($\geq 3/4$ inch diam.), wind (≥ 50 knots) and extreme turbulence (from aircraft reports). Some of these can be measured while others must be viewed or experienced. The funnel cloud is a poorly defined category because many tornado funnels are undoubtedly not observed while often unusual cloud appendages are reported as vortex formed funnels. A tornado is often not seen because of low clouds, dust or darkness, although the path of destruction that is left gives evidence that a tornado was present at a particular location. Sometimes a tornado is initially reported to have occurred, but subsequent investigation indicates that the damage was probably caused by strong winds.

Severe storms in the United States are verified by the National Severe Storms Forecast Center in Kansas City, Missouri and disseminated through the Severe Local Storms (SELS) log. These listings give the type of storm activity, time, latitude and longitude, reporting station code, distance and direction from the reporting station and the wind speed or hail size if that particular category is reported. As carefully as these reports are screened, there are many errors in identification of severe weather, location of the storms, time of the phenomena, etc. It seems probable that no one-to-one correspondence will be found between observed electrical characteristics of thunderstorms and reported weather conditions.

Attempts to identify the electrical characteristics of severe storms by comparing and summarizing the results of investigations by various workers in this field is often a very disappointing and frustrating task. Many reports present results that appear to be ambiguous, conflicting and even erroneous. This should not cast doubt upon the integrity of the investigators nor should it indicate that even honest errors were made. It does indicate that either the observational equipment was inadequate to properly respond in a predictable manner to the phenomena, that the information concerning the severe storm selected for study was in error, or that the results were biased because observations were made of phenomena resulting from electrical conditions far removed from the "characteristic" condition and perhaps near an extreme of some probability distribution.

It would be difficult to resolve some of the differences in reported results obtained from observations of the same storm by independent workers using their own instrumentation and employing their own methods of analysis. Data can often be re-examined and many differences in results can be minimized by using other methods of analysis. Variations in equipment responses and effects of many data pre-selection techniques cannot always be resolved, however, by varying analysis or by applying correctional terms obtained from the transfer functions of the equipment. Each unit of equipment distorts the signal in various ways. It is virtually impossible to reconstruct even the gross characteristics of the radiated signal from much of the data presently available.

It should not be surprising, therefore, in this field of relating atmospheric characteristics to severe storm conditions, that no warning system has been developed for one or more severe storm categories. No electromagnetic parameters have been found that are consistently indicative of tornadic activity or any other manifestation of severe weather. Until such parameters are found, it is advantageous to have several groups active in this field using widely differing techniques in order that no probable scheme will be overlooked.

17.3.4 Recent Advances in Selecting Parameters to Monitor

Expanding on the work of Jones (1959), Hughes and Pybus (1970) have studied severe storms using an instrumented airplane to gather data close to lightning discharges. They used a vertical antenna installed atop the fuselage and a horizontal antenna towed from the tail of the aircraft. The signals from each antenna were presented to a set of filters with center frequencies of 10 kHz, 50 kHz, 100 kHz, 150 kHz, 200 kHz and 250 kHz and a bandwidth of 1 kHz. The envelope detected outputs with a 1 millisecond time constant was recorded on a multi-channel magnetic tape recorder. Atmospheric pulse rates per minute was computed for each channel by counting the number of pulses whose amplitude exceeded a certain threshold. They show that the pulse rates are associated with a storm's vertical updraft or convective processes and that the higher frequencies seem to be better indicators of tornadic activity.

More recently, Shanmugam and Pybus (1971) used the same instrumentation employed by Hughes and Pybus (1970) and computed the average power each minute for 10 kHz and 50 kHz channels from the vertical and horizontal antennas as well as the average pulse rate per minute. Their data from four storms show an increase in the horizontal polarization relative to the vertical polarization as the severity of a storm increases. This implies an increase in horizontally oriented intracloud discharge processes while the proportion of vertically oriented cloud-to-ground discharges decreases as a storm becomes more severe. They further show that periodic changes occur in the pulse rate at 50 kHz that may be associated with a severe storm's fundamental periods of growth.

Stanford, Lind and Takle (1971) recorded the noise from six severe storms during the 1970 tornado season in Iowa. Their antennas were mounted on the roof of a building, 85 feet above ground level, and consisted of a five-foot whip for observations at 670 kHz, a ten-foot vertical antenna tuned to 11 MHz, and multi-element yagi antennas for observations at 53 MHz and 144 MHz. An MF broadcast-band receiver was used for 11 MHz, and combination frequency converters and communications receivers were used for 53 MHz and 144 MHz. The receiver bandwidths were approximately 5 kHz. Only the 670 kHz data and either the 53 MHz or 144 MHz data were recorded simultaneously, while the 11 MHz data was recorded during only one storm.

The envelopes of the radio-frequency (RF) signals were recorded on magnetic tape. Atmospheric pulse rates were later computed by playing back the tapes into an electronic frequency counter for a selected period of integration time. The data was then presented in histogram fashion versus time. Pulse rates were generally less than 500 per second when integrated for periods of 20 seconds, with occasional bursts of 1500 per second for a 0.2 second time scale.

In eleven out of twelve tornadoes occurring during the six storm periods reported by Stanford, Lind and Takle (1971), some type of enhanced electromagnetic pulse rates were observed. Activity from "ordinary" lightning discharges of non-severe storms were characterized by strong amplitude, isolated burst of pulses. Occasional periodic maxima in pulse rates were observed and sometimes well defined enhanced pulse rates occurred during non-severe thunderstorms. Their work also suggests that frequencies above about 1 MHz may be better indicators of tornadic activity than lower frequencies.

The most extensive recent studies to examine the possibility that distinguishable electrical radiation from severe storm areas could be found indicative of tornadic activity was started by Taylor (1971) during the 1970 tornado season at a stationary site in Oklahoma. Frequencies of observation were 10 kHz, 31.6 kHz, 100 kHz, 316 kHz, 1 MHz, 3.16 MHz, 10 MHz, 31.6 MHz and 137 MHz using single tuned filters adjusted to produce a bandwidth that was 10 percent of the center frequency through 10 MHz and a fixed 1 MHz bandwidth at higher frequencies. The rate of occurrence of the electromagnetic impulses were recorded at each frequency simultaneously at five amplitude levels separated by 10 dB. A mobile unit was used in addition to the stationary site during the 1971 observations of Taylor (1972). For this period, both observing units simultaneously recorded the rate of occurrence of atmospheric at three amplitude threshold levels (separated by 14 dB) at each frequency. Ten percent bandwidth filters were used at 31.6 kHz, 316 kHz, 3.16 MHz and 10 MHz, and 1 MHz bandwidth selected for 54 MHz and 137 MHz.

Two vertical monopole antennas with associated wideband gain controls and amplifiers were used to cover the frequency band from 10 kHz to 31.6 MHz. Each antenna was a 0.63 cm diameter aluminum rod, one meter in length, with a 15 cm diameter hemispherical corona cap at the top and an antenna-to-coax cable coupler at the bottom. The coupler unit was secured to a 1.25 m square aluminum base and the whole unit enclosed in plexiglass. One antenna system was used for frequencies of 316 kHz and lower, while the other antenna system supplied signals to the 1 MHz and higher frequency channels.

A 25 cm vertical monopole antenna and coupler mounted flush with the equipment van roof and a modified communications receiver was used for the 54 MHz channels. The 137 MHz responses were obtained from a half wave vertical dipole antenna, approximately 1.05 m overall length, mounted with the center of the antenna 1.5 m above the van roof and used a 137 MHz to 10 MHz frequency converter.

The transient responses of the vertical monopole antenna filters, the 54 MHz receiver and the 137 MHz converter were presented to logarithmic amplifiers and were then rectified and smoothed. Up to five threshold trigger circuits for each frequency were used by Taylor (1971) to activate one shot multivibrators each time the response exceeded a predetermined level. The output of each one shot was integrated and presented to multichannel magnetic tape recorders through logarithmic output drivers.

An example of atmospheric rate data from Taylor (1971) for an approximate 2 minute period on April 18, 1970 is shown in (F17.4). Each channel is aligned in time and arranged with the 5 channels of 10 kHz rates at the top, channels 6-10 for 31.6 kHz next, etc. Only channels 24 and 25 for the 1 MHz frequency and channels 28 and 29 for the 3.16 MHz frequency are shown. The frequency for each group of channels is indicated in a column on the far left of the figure. The threshold in volts per meter, which is the amplitude level the impulses must exceed before activating the counting circuits in each channel, is presented in the second column on the left for the odd numbered channels and in the far right column for the even numbered channels. Thus the lower frequencies are nearer the top with the threshold level increasing from top to bottom within each frequency group. The data format is rate versus time with the impulse rates on an approximate logarithmic scale presented just to the left of the data for odd numbered

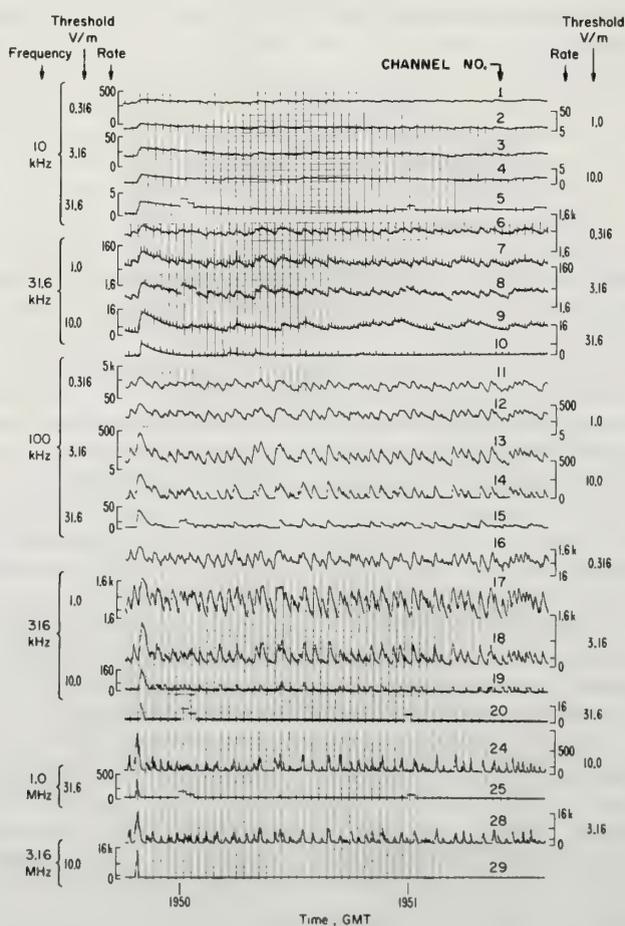


Figure 17.4 Atmospheric rate data – Severe storm activity, April 18, 1970.

channels and just to the right of the data for even numbered channels. Rates were calibrated each decade beginning at either 1.6 or 5.0 impulses per second. Channel 17, for example, is the second channel of the 316 kHz group. This is an odd numbered channel for which the threshold level of 1.0 V/m is found on the left together with a rate scale that extends from 1.6 to 1.6 k per second. The two other marks on the rate scale are for 16 and 160 impulses per second. A zero rate is indicated on some scales; in such cases the rate varies approximately linearly between zero and the first rate mark, and logarithmically thereafter. Greenwich mean time is shown along the bottom as indicated. The vertical grid lines are separated by 2 seconds. One minute time marks are shown on some channels as square pulses 2 seconds in length. Data for 1971 (Taylor, 1972) are similar to this with only three amplitude threshold levels recorded at 0.6, 3.0 and 15.0 V/m.

The data for (F17.4) were recorded during a funnel cloud reported 95 km NNE of Norman associated with the SW edge of a precipitation area as determined by weather radar and direction of arrival of atmospherics. This data was typical of atmospheric rates observed during other periods of tornadic activity. Taylor (1972) found that the characteristics of the rate data changed very little at the lower frequencies between non-severe storm activity and tornadic conditions, but there were usually an increase in the atmospheric rates at frequencies above 316 KHz as severe weather activity increased. He selected data channels 18, 24, and 28 as representative of electrical activity and counted the number of bursts of enhanced impulse rates per minute observed on each channel. Figure (17.4) shows an average of about 25 bursts per minute on each of these channels. Non-severe thunderstorms usually produced less than 5 bursts per minute.

The average burst rate—a measure of the frequency of occurrence of bursts of high impulse rates determined by averaging the burst counts per minute at the three selected channels—was used to estimate the electrical activity. All 1970-71 observations were separated into three categories determined by the measured burst rate. High rates, medium rates and low rates were determined, respectively, by burst counts exceeding 15 bursts per minute, between 7½ and 15 bursts per minute and less than 7½ bursts per minute. A summary of Taylor's (1972) preliminary analysis is presented in Table (17.1).

Table 17.1 1970-71 Data Summary

Weather Conditions	Range	No. of Events	Burst Rate Category		
			High	Medium	Low
Tornado/funnel	0-50 km	6	5	1	0
Tornado/funnel	50-100 km	12	2	5	5
Other severe storms*	0-50 km	8	1	2	5
Other severe storms*	50-100 km	13	0	2	11
All severe storms**	100-200 km	70	0	0	70
Non-severe storms	<30 km	49	4	11	34

* May have had tornado or funnel activity, though none was reported

** Included tornadic storms as well as other severe storms.

It is interesting to note from Table (17.1) that of the 18 tornadoes and/or funnels reported between 0-100 km from observation sites, 13 of them or 72% were accompanied by a high or medium burst count rate. Only 24% of all other severe storms within 100 km produced high or medium burst count rates, while no severe storm including tornadic storms at distances greater than 100 km produced a burst count rate above the low category.

If the monitoring range were limited to less than 50 km and all high and medium burst count rates were assumed to be indicative of tornadic activity, then no tornadic condition would have been missed, almost half of the other severe storms would have been classified tornadic, and slightly less than a third of the local thunderstorms would have caused false alarms. But if only the high burst count rate were used for the same range limitation, then one tornado would have been missed, only one severe storm would have attained tornadic classification, and only 4 out of 49 local thunderstorms would have produced false alarms.

There were a number of other parameters examined by Taylor (1972) that did not prove to be sensitive indicators of tornadic activity. Variations of the earth's magnetic field observed during storm conditions were traced in every case to currents produced by either the voltage fluctuations on local A.C. power lines or the disruption of power in nearby areas by the storm. Integrated noise measurements at 10 kHz, 100 kHz, 1 MHz and 10 MHz were sensitive to variations in atmospheric activity dependent on distance to the storm and impulse amplitude and appear very inferior to the rate data at the same frequencies. No prolonged brightening of the TV picture tubes of two black and white receivers were observed associated with tornadic conditions, although the nearest tornadic activity occurring when the TV receivers were operative was in excess of 20 km.

The possibility of using a television set to warn of an approaching tornado received considerable recent publicity (Waite and Weller, 1969 and Biggs and Waite, 1970). The technique involves darkening the picture tube to almost complete blackness while the set is on channel 13, and then changing to channel 2 and observing the picture tube brightness. If a tornado is closer than 25 to 30 km from an "average" TV set operated by an "experienced" observer, the picture tube should become a glowing white. This "Weller Method" caused considerable controversy between its proponents and those in the National Weather Service who preferred having people continue to follow TV broadcast of tornado alerts and warnings rather than to attempt doing their own warning using untested methods. Although the results of Stanford, Lind and Takle (1971) and Taylor (1972) indicate considerable electromagnetic energy often accompany tornadic activity, no work has been published concerning the response characteristics of television receivers to transient signals or concerning the reliability of the "Weller Method" as a tornado warning device in an average home.

The results of recent work covered in this section are consistent with the whole history of radio and visual observations of tornadoes, namely that most tornadoes, but perhaps not all, have major radio-frequency electrical activity associated with them. The radio-signal characteristic most indicative of tornadic activity is a high rate of atmospheric impulses as measured at frequencies in the range of 1 MHz to 100 MHz or so. This characteristic is readily observed by either counting the number of impulses per unit of time using high gain narrow band receivers as did Stanford, Lind and Takle (1971) or counting the number of burst of impulses per unit of time using low gain restricted band filters as did Taylor (1971, 1972). And yet, many questions remain unanswered concerning the electrical identification of tornadoes and other severe storm characteristics and the selection of critical parameters to monitor. More research is needed before a monitoring system can be devised with high tornado warning reliability and yet with few false alarms from local thunderstorm conditions.

17.4 Electricity in the Non-Violent Atmosphere

Nothing has been mentioned so far about the natural electrical phenomena occurring outside of the violent structures of thunderstorms, severe storms and tornadoes. The precipitation from clouds that do not produce lightning is usually charged, and usually the small non-precipitating clouds show some electrification. But even at great distances from violent and non-violent weather conditions, the potential gradient or vertical electric field above the earth's surface is of the order of 100 V/m. Few details of this important field will be covered here, but further information can be found in survey papers by Israël (1965), Pierce (1959) and Webb (1969).

This fair weather electricity results from the conducting envelope of the ionosphere that is maintained at a potential of about 3.5×10^5 volts more positive than the earth's surface. The problems of generation, dispersion and consumption, i.e., the separation of charges, movement of charges and recombination of charges, respectively, are very complex because the entire global electrical structure is involved. It is generally believed that thunderstorm activity around the world produces the approximate 2000 amperes charging current needed to maintain the fair weather field. The circuitry, however, includes not only the troposphere but also the earth, the ionosphere and the magnetosphere, and involves telluric currents, airglow, solar radiation, tidal forces, magnetic storms, auroral activity and other facets of an obviously complex system.

There are many variables of a local nature that affect the observed field near the surface of the earth. Since the conductivity of the earth ranges from 4 mhos/meter for sea water to about 10^{-3} mhos/meter for poor conducting ground and the conductivity of the lower ionosphere attains values approximately the same as the lower values of ground conductivity, the region of lowest conductivity and

consequently the highest potential gradient is in the atmosphere near the earth's surface where the conductivity is about 2×10^{-14} mhos/meter. Variations in wind, vegetation on the ground, radioactivity of the ground, cosmic rays, land elevation, concentration of aerosols present, nuclear test fall out, etc., affect the long and short-term values of the fair weather field.

17.5 Summary

A brief review of thunderstorm electricity and characteristics of atmospheric was presented at the beginning of this chapter to give us a foundation upon which to examine electromagnetic techniques for the remote sensing of severe storms. The problem is made more difficult because of limited knowledge of thunderstorms in general, and of large destructive severe storms in particular. It is still not possible to devise an accurate electrical model of a thunderstorm that includes the processes of charge concentration and the resultant lightning discharge.

The study of electrical activity associated with tornadoes and severe storms is made more difficult by the many variables in observational data and the complexities of weather conditions. A number of problems were examined that relate to tracking electrically active regions and to identifying severe storm characteristics. The recent results of several workers in this field were reviewed. It is interesting to realize that no actually conflicting results are found in viewing these works, although the observing frequency, the monitoring techniques and data analyzing methods were in some cases drastically different. The overall consensus indicates that most tornadoes and some severe storms as well as a few non-severe thunderstorms have an unusually high level of electromagnetic activity associated with them. This activity is best recognized in the form of very high rate of occurrence of impulses measured at frequencies above about 1 MHz.

17.6 References

- Berger, K. (1967), Novel observations on lightning discharges: results of research on Mount San Salvatore, *J. Franklin Inst.* 283, No. 6, 478-525.
- Biggs, W. Gale, and Paul J. Waite (1970), Can TV really detect tornadoes?, *Weatherwise* 23, No. 3, 120-124.
- Braham, R. R. Jr. (1952), The water and energy budgets of the thunderstorm and their relation to thunderstorm development, *J. Meteorol.* 9, 227-242.
- Brook, M. (1967), Electric currents accompanying tornado activity, *Science* 157, No. 3795, 1434-1436.
- Davies, K. 1965: *Ionospheric Radio Propagation*, NBS Monograph 80, U. S. Government Printing Office, Washington D. C.
- Dennis, A. S. and E. T. Pierce (1964), The return stroke of the lightning flash to earth as a source of VLF atmospheric. *Radio Sci.* 68D, No. 7, 777-794.
- Horner, F. (1954), The accuracy of the location of sources of atmospheric by radio direction-finding, *Proc. IEE* 101, Part III.
- Hughes, W. L. and E. J. Pybus (1970), Severe storm sferics-stroke rate, *Proc. 14th Radar Meteorology Conference*, Tucson, Arizona, 315-318.
- Imyanitov, I. M., B. F. Evteev, and I. I. Kamaldina, 1969: A thunderstorm cloud, *Planetary Electrodynamics I*, Gordon and Breach Science Publishers, Inc., New York, 401-425.
- Israël, H., 1965: Problems of fair-weather electricity, *Problems of Atmospheric and Space Electricity*, Elsevier Publishing Co., New York, 59-67.
- Jones, H. L. (1950), The identification and tracking of tornadoes, *URSI Comm. 4 Meeting*, April 18.

- Jones, H. L., 1959: The identification of lightning discharges by spheric characteristics, *Recent Advances in Atmospheric Electricity*, Pergamon Press, New York, 543-556.
- Jones, H. L. (1965), The tornado pulse generator, *Weatherwise* 18, No. 2, 78-80.
- Jordan, E. C., 1950: *Electromagnetic Waves and Radiating Systems*. Prentice-Hall, Inc., New York.
- Kasemir, H. W., 1965: The thundercloud, *Problems of Atmospheric and Space Electricity*, Elsevier Publishing Co., New York, 215-231.
- Kitagawa, N., 1965: Types of lightning, *Problems of Atmospheric and Space Electricity*, Elsevier Publishing Co., New York, 337-348.
- Kitagawa, N., M. Brook and E. J. Workman (1962), Continuing currents in cloud-to-ground lightning discharges, *J. Geophys. Res.* 67, No. 2, 637-647.
- Mackerras, D. (1968), A comparison of discharge processes in cloud and ground lightning flashes, *J. Geophys. Res.* 73, No. 4, 1175-1183.
- Malan, D. J., 1965: The theory of lightning, *Problems of Atmospheric and Space Electricity*, Elsevier Publishing Co., New York, 323-331.
- Oetzel, G. N. and E. T. Pierce, 1969: Radio emissions from close lightning, *Planetary Electrodynamics I*, Gordon and Breach Science Publisher, New York 543-569.
- Ogawa, T. and M. Brook (1964), The mechanism of the intracloud lightning discharge, *J. Geophys. Res.* 69, No. 24, 5141-5150.
- Pierce, E. T., 1959: Some topics in atmospheric electricity, *Recent Advances in Atmospheric Electricity*, Pergamon Press, New York, 5-15.
- Pierce, E. T. (1963), Excitation of earth-ionosphere cavity resonances by lightning flashes, *J. Geophys. Res.* 68, No. 13, 4125-4127.
- Sanders, F. (1971), Toward defining human needs: How does the atmosphere hurt us? *Bulletin Am. Meteorological Society* 52, 446-449.
- Sao, K., M. Yamashita, S. Tamahashi and W. L. Taylor (1970), Genesis of slow tail atmospheric deduced from frequency analysis and association with VLF components, *J. Atmospheric and Terrest. Phys.* 32, 1147-1151.
- Shanmugam, K. and E. J. Pybus (1971), A note on the electrical characteristics of locally severe storms, *Proc. 7th Conf. on Severe Local Storms, Kansas City, Missouri*, 86-90.
- Stanford, J. L., M. A. Lind, and G. S. Takle (1971), Electromagnetic noise studies of severe convective storms, *J. of Atmospheric Sciences* 28, 436-448.
- Taylor, W. L. (1960), Daytime attenuation rates in the very low frequency band using atmospheric, *J. Res. NBS 64D (Radio Prop.)* No. 4, 349-355.
- Taylor, W. L. (1963), Radiation field characteristics of lightning discharges in the band 1 kc/s to 100 kc/s, *J. Res. NBS 67D (Radio Prop.)* No. 5, 539-550.
- Taylor, W. L. (1967), VLF transmission loss calculated from spectral analyses of atmospheric, *Radio Science* 2, No. 2, 139-145.

- Taylor, W. L. and K. Sao (1970), ELF attenuation rates and phase velocities observed from slow-tail components of atmospherics, *Radio Science* 5, No. 12, 1453-1459.
- Taylor, W. L. (1971), Review of electromagnetic radiation data from severe storms in Oklahoma during April, 1970, NOAA Technical Memorandum, ERLTM-WPL 6.
- Taylor, W. L. (1972) The search for electromagnetic radiation indicative of tornadic activity, To be published.
- Wait, J. R., 1962: *Electromagnetic waves in Stratified Media.*, Pergamon Press, New York.
- Waite, P. J., and N. Weller (1969), The Weller method: tornado detection by television, Proc. of Sixth Conf. on Severe Local Storms, Chicago, 169-171.
- Webb, W. L., 1969: Global electric structure, *Planetary Electrodynamics*, 2, Gordon and Breach Science Publisher, New York, 245-271.
- Williams, D. P. and M. Brook (1963), Magnetic measurements of thunderstorm currents, *J. Geophys. Res.* 68, No. 10, 3243-3247.
- Workman, E. J., 1965: Thunderstorm electricity, *Problems of Atmospheric and Space Electricity*, Elsevier Publishing Co., New York, 296-303.
- Vonnegut, B. (1960), Electrical theory of tornadoes, *J. Geophys. Res.* 65, No. 1, 203-212.

List of Symbols

E	Vertical electric field
ϵ_0	Permittivity of free space ($1/36\pi \times 10^9$ farad per meter)
M	Current Moment, ampere-meters
t	Time in seconds
d	Distance in meters
c	Velocity of light (3×10^8 meters per second)
β	$2\pi/\lambda$
λ	Wavelength in Meters c/f
f	Frequency in Hertz
R	Reflection coefficient of the earth's surface
F	Groundwave attenuation factor
A	Spectral amplitude coefficient of source
L	Waveguide mode launching term or excitation factor
a	Radius of the earth, 6370 meters
α	Waveguide mode attenuation in nepers
h	Height of ray path above earth
θ	Ray elevation angle above horizon

Chapter 18 TEMPERATURE AND WIND STRUCTURE STUDIES BY ACOUSTIC ECHO-SOUNDING

Freeman F. Hall, Jr.

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Acoustic waves are scattered by temperature and velocity inhomogeneities in the atmosphere. In a typical acoustic echo sounder, short pulses of audible sound are collimated by an acoustic antenna and directed into the atmosphere. The returning scattered signals are collected by the same antenna for monostatic measurements of thermal structure, or at a different antenna some distance removed for bistatic measurements for both wind and thermal structure. Wind produces a Doppler frequency shift in the scattered sound, allowing horizontal and vertical wind components to be measured by a frequency analysis of the signals collected by a three antenna array.

18.0 Introduction

Acoustic echo sounding of atmospheric structure is one of the newest and yet one of the oldest remote sensing techniques. It was in 1873 that John Tyndall detected acoustic backscatter from temperature and wind structure in the atmosphere. He attributed the scattering to "acoustic clouds." The effect was discovered while investigating the propagation of acoustic energy from a large fog horn (Tyndall, 1875). He was even able to demonstrate that air, heated by a flame, could attenuate the direct propagation of sound and lead to backscattering. After Tyndall's initial investigation, the technique lay dormant for over 70 years before studied again as an atmospheric probe.

Shortly before the end of World War II, Gilman, Coxhead, and Willis (1946) used acoustic backscatter to study the structure of low level temperature inversions as they affected propagation in microwave communication links. A loudspeaker source and microphone in a 60 cm parabolic dish were used as source and receiver, and the backscatter echoes displayed on an oscilloscope, which was photographed with a time-lapse camera. A surprising amount of atmospheric structure under temperature inversions was displayed. Strong correlation was found between the acoustic inversion echo and microwave fading over a low level, 64 km path. Breakup of nocturnal thermal inversions was monitored, and the change to typical daytime convective echoes could be observed. The authors were not able to explain completely the reason why such echoes were obtained. The publication of these results in the Journal of the Acoustical Society of America may be the reason why meteorologists were not attracted to the advantages of acoustic remote sensing at an earlier date.

During the late 1950's acoustic scattering from the atmosphere was investigated experimentally in the Soviet Union by Kallistratova (1959a, 1959b, and with Tatarski, 1960), but it was the investigation by McAllister (1968) which showed that echoes could be reliably obtained to heights of several hundred meters. McAllister also showed the value of the facsimile recording media for illustrating the temporal structure of atmospheric scatter for both stable and unstable lapse rates. More recently the program at the Wave Propagation Laboratories of NOAA has demonstrated the ability to measure winds in the atmosphere by means of the Doppler shift in the scattered sound (Beran et al., 1971, Beran and Willmarth, 1971) and has studied the feasibility of utilizing sounders in urban environments as a tool for monitoring temperature inversion structure (Simmons et al., 1971).

Before considering in detail how sounders may be utilized for atmospheric studies, it will be necessary to discuss some basic acoustical physics concerning the propagation, attenuation, and scattering of sound. Following these fundamental discussions, the details of a typical sounder design are presented, and atmospheric structure which may be detected with acoustic echo sounders will be discussed.

18.1 Propagation and Attenuation of Sound in the Atmosphere

Acoustic energy propagates through the atmosphere as a longitudinal wave of pressure variations. The *intensity* of a sound wave is defined as the time rate of energy passage through a unit area perpendicular to the direction of wave propagation. In the MKS system of units, sound intensity is

measured as watts per square meter ($w m^{-2}$). The attenuation of acoustic waves as they propagate through the atmosphere varies with the frequency of the waves but in a continuous, smooth fashion. Therefore, when small frequency ranges are considered, the intensity of a plane wave train I at some distance ℓ from a reference plane where the intensity is I_0 is given by (18:1)

$$I = I_0 e^{-k\ell} \quad (18:1)$$

Here k is defined as the attenuation coefficient. It is made up of three independent components,

$$k = k_c + k_m + k_s . \quad (18:2)$$

The subscripts stand respectively for classical, molecular, and scattering (or excess) attenuation.

Classical attenuation of sound is caused by the finite viscosity of air which leads to heating of the atmosphere as the sound wave passes, and by radiation and heat conduction from the high pressure regions of the sound wave. Classical absorption was first studied by Stokes (1849), and elaborated upon by Rayleigh (1896). Classical absorption varies as the frequency ν raised to the second power, or

$$k_c = 4.24 \nu^2 \times 10^{-11} m^{-1} \quad (18:3)$$

as given by Beranek (1954). Compared to the other two sources of attenuation, classical absorption is so small as to be essentially negligible in the audible range of frequencies under normal, temperate atmospheric conditions.

The first careful investigation to determine the cause of atmospheric absorption of sound above that predicted by the classical theory was performed by Knudsen (1931). He discovered the marked dependence of absorption on the humidity of the air. In a later investigation Knudsen (1933) showed that the abnormally high absorption in air is determined by an interaction between oxygen and water molecules since the test chamber filled only with nitrogen and water vapor showed a much lower attenuation coefficient than when oxygen was present in the proportions found in the atmosphere. The mechanism by which molecular absorption occurs was explained by Kneser (1933) as the transfer of collision excited vibrational energy from oxygen molecules, produced by the acoustic field, to water vapor molecules. The energy of the excited water molecules is then radiated away in the infrared. This transfer of energy from oxygen to water vapor is facilitated by the nearly exact overlap of energy levels in the two molecules, but whereas the oxygen molecule is forbidden from radiative decay, this method of energy removal is highly favored by water vapor. The molecular interaction and transition rates have been studied more fully by Henderson and Herzfeld (1965).

The most comprehensive measurements of molecular absorption, studied by means of the decay of acoustic energy in closed chambers, are those of Harris (1966). Examples of data from this study are given in (F18.1). Molecular absorption decreases with temperature, and helps explain why in cold climates cases of unusually long range sound propagation are sometimes reported, although ducting of sound between severe arctic temperature inversions and the ground may also be important in such cases.

As an example in comparing the importance of classical and molecular attenuation, consider a frequency of 2 kHz in air at 20°C and 15% relative humidity. From (18:3) the classical absorption coefficient is $k_c = 1.7 \times 10^{-4} m^{-1}$. From (F18.1), $k_m = 6 \times 10^{-3} m^{-1}$ or a factor 35 times the classical attenuation. From this comparison it is easy to appreciate why classical attenuation is usually disregarded in considering sound propagation in standard, temperate atmospheres.

Acoustic echo sounding is feasible because sound is scattered from a probing acoustic beam by temperature structure and wind turbulence in the atmosphere. Obviously, if energy is scattered from the beam the intensity of the beam will decrease. This excess attenuation, over and above classical and molecular contributions, is expressed by the third coefficient on the right hand side of (18:2). The

magnitude of k_s will depend upon the structure of the atmosphere. Field tests by Beran, et al., (1970) utilizing a sound source on a sailplane, indicate that excess attenuation may be negligible under some atmospheric conditions, but may be of the same order of magnitude as molecular absorption for other conditions. A similar range of variation for excess attenuation was found by Dneprovskaya et al., (1963) who measured excess values of from 5 to 23 dB km⁻¹ near 2 kHz. Until we know more about sound propagation and scattering in the free atmosphere, little more can be said about excess attenuation. It should be possible eventually to predict its magnitude by observing the acoustic energy scattered from the atmospheric with an echo sounder. Accurate, quantitative interpretation of echo sounding returns depends critically upon understanding all aspects of attenuation in the real atmosphere, and more research needs to be accomplished, both theoretical and experimental, on the propagation of sound in realistic atmospheres.

18.2 Scattering of Sound Waves

Waves traveling in a perfectly homogeneous and continuous medium are not scattered. There must be variations in the propagation velocity of the wave, that is to say, inhomogeneities in the refractive index of the medium, in order to produce scattering. The velocity of sound waves in the atmosphere is a function of the temperature and of any velocity component of the air. Stated differently, temperature or wind structure effectively changes the index of refraction of the atmosphere carrying the sound waves, and one would expect energy to be scattered from the waves when such inhomogeneities are encountered.

The scattering problem may be studied analytically by a direct application of Fermat's principle of least time of propagation. This principle is directly deducible from Huygen's principle of wave front construction by means of contributory wavelets, and it states that wave motion always chooses that path of propagation requiring an extremum (a maximum or minimum) in time to cover (Richtmyer and Kennard, 1947). Once an expression for the propagation velocity has been derived, it is only required to minimize transmission time by means of the calculus of variations and the Euler equation (Margenau and Murphy, 1956). With both the Euler equation statement of Fermat's principle and the conventional three dimensional wave equation, one has two equations, solvable in the two unknowns of acoustic pressure and particle velocity. This is exactly the technique used by Monin (1962) in deriving the intensity and angular characteristics of sound scattered in a turbulent atmosphere. Following Monin, a useful form of the wave equation is

$$\frac{\partial p}{\partial t} = -\gamma p \nabla \cdot \vec{u} \quad , \quad (18:4)$$

where p is the atmospheric pressure, γ the ratio of specific heats of the atmospheric gas at constant pressure and at constant volume, and \vec{u} is the instantaneous velocity of particles in the wave. The derivation of the wave equation in this form is given by Beranek (1954). It is convenient to include the pressure dependence by means of a new variable.

$$\Pi = \frac{1}{\gamma} \ln p \quad , \quad (18:5)$$

so that the wave equation becomes (Monin, 1962)

$$\frac{\partial \Pi}{\partial t} = - \left[\nabla \cdot \vec{u} + (\vec{u} \cdot \nabla) \Pi \right] \quad . \quad (18:6)$$

Since the velocity of sound propagation is proportional to the square root of the atmospheric temperature, or

$$c^2 = c_0^2 \left(1 + \frac{T'}{T} \right) \quad , \quad (18:7)$$

where c_0 is the acoustic velocity in the undisturbed atmosphere and T' is the temperature of the disturbed atmosphere while T is the mean atmospheric absolute temperature, the Euler equation may be written in the form

$$\frac{\partial \vec{u}}{\partial t} + c_0^2 \nabla \Pi + (\vec{u} \cdot \nabla) \vec{u} + c_0^2 \frac{T'}{T} \nabla \Pi = 0 \quad (18:8)$$

Monin then introduces the atmospheric characteristics by means of

$$\vec{u} = \vec{u}' + \vec{v} \quad (18:9)$$

and

$$\Pi = \Pi' + \tau \quad (18:10)$$

where \vec{u}' and Π' represent the atmospheric turbulence contributions, and \vec{v} and τ are determined by the acoustic wave only. Assigning known statistical properties to \vec{u}' and Π' (18:6) and (18:8) are solved for the two unknowns \vec{v} and τ which describe the characteristics of the acoustic field. Monin shows that the solution for $\sigma(\theta)$, the power scattered from a unit volume per unit incident flux into a unit solid angle is given by (Clifford, 11.2)

$$\sigma(\theta) = \frac{32\pi^5 \cos^2 \theta}{\lambda^4} \left[\frac{\Phi(V) \left(\frac{4\pi}{\lambda} \sin \frac{\theta}{2} \right) \cos^2 \frac{\theta}{2}}{c_0^2} + \frac{\Phi(T) \left(\frac{4\pi}{\lambda} \sin \frac{\theta}{2} \right)}{4T^2} \right] \quad (18:11)$$

Here λ is the wavelength of the sound for the average atmospheric temperature T , θ is the scattering angle measured from the direction of the propagation of the incident wave, and $\Phi(V)$ and $\Phi(T)$ are respectively the three dimensional spectral densities of fluctuations in wind velocity and temperature evaluated at the spatial scale λ' where

$$\lambda' = \frac{\lambda}{2 \sin \frac{\theta}{2}} \quad (18:12)$$

It is pointed out by Kallistratova (1961) that the acoustic energy scattered at angle θ is determined only by spectral components of turbulence on the spatial scale λ' . When a Kolmogorov spectrum of turbulence (Meecham, 4.2.2) is assumed, (18:11) becomes

$$\sigma(\theta) = 0.055 \lambda^{-1/3} \cos^2 \theta \left[\frac{C_V^2}{c^2} \cos^2 \frac{\theta}{2} + 0.13 \frac{C_T^2}{T^2} \right] \left(\sin \frac{\theta}{2} \right)^{-11/3} \quad (18:13)$$

Here C_V^2 is the velocity structure parameter, C_T^2 is the thermal structure parameter defined by

$$C_V^2 = \overline{\left[\frac{u(x) - u(x+r)}{r^{1/3}} \right]^2} \quad , \quad C_T^2 = \overline{\left[\frac{T(x) - T(x+r)}{r^{1/3}} \right]^2} \quad (18:14)$$

Wind speed is u at position x in the positive x direction, and r is measured along the x axis. The x axis may be chosen in any arbitrary direction to suit the experiment.

The analytical study of the capabilities of acoustic sounding by Little (1969) was a necessary first step in predicting how such techniques might be useful to meteorologists. Substituting typical values for the two atmospheric structure parameters in (18:13), the possibility of obtaining readily detectable acoustic echoes with reasonably sized equipment was shown to be feasible. It should be noted from the scattering equation that the scattered acoustic power is only a weak function of acoustic wavelength, varying as $\lambda^{-1/3}$, that direct backscatter for $\theta = \pi$ is a function of the thermal structure only, that there is no scattering at $\pi/2$, and that there is strong scattering in the forward direction because of the final factor in the equation. The scattering equation obviously cannot be applied for extremely small values of θ , but its general correctness for scattering angles as small as 20° has been confirmed in experiments by Kallistratova (1961) in the real atmosphere and more recently by Baerg and Schwarz (1965) in a wind tunnel. Further details and discussions of the scattering of acoustic waves are given by Clifford (11.2).

In addition to acoustic backscatter from random temperature structure in the atmosphere, uniform temperature gradients may also contribute to the scattered power. To have a measurable acoustic reflection from a temperature gradient requires a significant change in index of refraction in a small percentage of the acoustic wavelength. Present knowledge of temperature stratification under stable conditions does not yield much insight on gradients at such a fine spatial scale. Acoustic echo sounder records indicate, however, an excess return which is frequently observed under stably stratified conditions. Fast response, differential temperature readings, which provide a measure of the random temperature structure, lead to a prediction of smaller acoustic scattering than is actually observed under such conditions. Better backscatter agreement is obtained with predictions from differential temperature readings under convectively unstable conditions, when a more uniformly mixed, isotropic turbulence condition is probably more likely to occur. Further research must be done on the relative importance of so-called specular reflections from uniform temperature gradients as compared to acoustic scatter from random temperature fluctuations under stable atmospheric conditions.

It is straightforward to calculate the acoustic reflectivity, r , of temperature steps or gradients in the atmosphere. For layers with an abrupt change of absolute temperature from T_1 to T_2 in a distance much less than an acoustic wavelength, the reflectivity is, $r = (T_1 + T_2 - 2\sqrt{T_1 T_2})/4T_0$, where T_0 is the reference temperature for unity index of refraction. A change of temperature of but 1°K , in such an abrupt step increment, would lead to larger acoustic reflections from thin layers than have been observed to date. Taking more realistic, smoothly varying, sinusoidal temperature fluctuations in one acoustic wavelength does give a reflectivity on the same order as that observed from stably stratified structure, so perhaps the interpretation of enhanced returns as due to specular reflections will prove to be correct when all the results are in.

With these qualifications, the functional equation relating transmitted and received power, and the scattering function $\sigma(\theta)$ may now be considered. This expression is entirely analogous to the meteorological radar equation (Wilson and Miller, 13.1). For distributed scattering centers in the atmosphere, the acoustic echo sounder equation takes the form

$$P_r = P_t \sigma \frac{ct}{2} \frac{A}{\varrho^2} \tau^2 G \quad (18:15)$$

where P_r is the received power, P_t is the transmitted power, A is the collecting area of the receiving antenna, t is the time duration of the acoustic pulse, τ is the transmittance of the atmosphere for acoustic energy from the antenna to the scattering volume, and G is a gain factor which accounts for the antenna beam pattern characteristics. For a perfect piston source this gain factor G is ~ 0.24 .

Consider now the received power levels to which an efficient acoustic echo sounder system must respond. The fundamental theoretical limit of the acoustic noise power is that generated by the random thermal motion of the atmospheric molecules. For the standard atmosphere and a 100 Hz receiver bandwidth, this is 4×10^{-19} w (Little, 1969). This noise power is of the same order of magnitude as the Johnson noise which is generated in the voice coil of an acoustic transducer. Using a well shielded acoustic antenna with 90° sidelobe rejection 60 dB below the vertically pointed main lobe, this low theoretical limit

of background noise has only been observed once under quiet, early morning, rural conditions on the Colorado plains. Under most conditions, other acoustic background noise will exceed the theoretical limit. Sources of such noise include wind across the transducer, insects and birds, or human activity, especially transportation or industrial devices. More details on background noise characteristics are given by Simmons, et al. (1971).

The acoustic power return, P_r , for a nominal transmitted power of 20 acoustic watts, may range from values near the theoretical noise limit, at ranges of from 100 m to 1 km for non-turbulent atmospheric regions, to values 60 or 70 dB above the noise limit for an atmosphere with a strong thermal structure function at a range of 50 m. It is therefore necessary that the acoustic echo sounding system be capable of responding over an extremely wide dynamic range. Of course, P_r is also a function of the distance to the echoing parcel. An echo sounder system with a total dynamic range of 95 dB does not seem to exhibit saturated echoes from strong C_T structure, even at short ranges of but 30 m. Methods of achieving this large dynamic range will be described in the following section.

18.3 Acoustic Echo Sounder Design

The following discussion is on a "second generation" acoustic echo sounding system, embodying many worthwhile engineering improvements while carrying on the general design philosophy of an earlier sounder (Wescott, et al. 1970). As much of the system as possible was built with the laboratory quality, commercially available electronic instruments. The purpose of this section is to acquaint the reader with sounder design considerations, and to describe the operation of a typical system.

The operation of the equipment can be followed from the block diagram, (F18.2). All timing functions are regulated by a crystal stabilized clock in the time code generator. Timing logic consists of a configuration of diode gates which pass a pulse only when a particular combination of zero and plus binary coded decimal levels occur that correspond to a desired pulse repetition frequency (PRF). Timing logic gates are currently provided for PRF's of 0.5, 0.2, 0.1 Hz corresponding to maximum echo ranges of approximately 330, 825, and 1650 m.

The PRF signal is used to trigger a digital timing generator which in turn passes a tone burst of carrier frequency from the receiver beat frequency oscillator (BFO), generates a gate pulse coincident with the tone burst, and generates two additional pulses having start and stop times that may be delayed independently be desired amounts from the start of the tone burst. Tone burst durations ranging from 10 milliseconds to 1 second are normally used with the sounder depending upon whether maximum resolution, or maximum range is desired.

As (F18.2) shows, the tone burst from the timing generator is gated a second time as it passes through a transmit gate on the way to the transmitting power amplifier. The transmit gate is a field effect transistor (FET) solid-state switch with ON and OFF modes controlled by the gate pulse from the timing generator. The need for redundant gating of the tone burst is necessary because the digital timing generator provides only 60 dB of carrier frequency attenuation between gate pulses. The resulting carrier leakthrough was enough to create a very low level "false echo" signal during the echo sounder receive mode.

The tone burst from the transmit gate is applied to the input of the power amplifier at a level of 1 volt rms. The gain of the power amplifier is adjusted to deliver 100 watts rms of electrical power through a pair of back-to-back diodes to an 8-ohm acoustic transducer for the duration of the tone burst. The transducer is coupled acoustically to a reflector-horn antenna. The overall transmitting efficiency of the transducer-antenna configuration is about 20% so that the antenna radiates about 20 watts rms of acoustic power into the atmosphere during the tone burst. t.

Between tone bursts the power amplifier produces less than a millivolt of hum and instrument noise. This noise is nonetheless more than an order of magnitude larger than some of the voltages produced by weak atmospheric echoes. Back-to-back silicon diodes were therefore placed between the power amplifier and the acoustic transducer. These diodes readily conduct large-amplitude transmit tone bursts, but are virtual open circuits to signals of less than 300 millivolts rms. In order for the diode isolation circuit to work reliably, however, it is essential that the power amplifier have a well regulated power supply. Otherwise power line surges will produce voltage fluctuations at the amplifier output large enough to pass through the diodes, and receiver overload will occur.

During transmission it is necessary to isolate and protect the extremely sensitive receiver circuits of the echo sounder from overload and possible damage. This is done in three stages, and the circuit elements involved, mostly silicon diodes as shown in (F18.2), perform a task equivalent to that of a radar transmit-receive (T-R) device. The first stage is a set of back-to-back diodes in series with resistor R1. The second stage is a diode bridge shorting gate. The action of this gate is controlled by the polarities of a pair of equal but opposite receiver gate pulses. When the polarities of these pulses are as shown in (F18.2), all diodes of the bridge conduct, and the receiver preamplifier input terminal is thereby clamped to ground. When the polarities of the pulses are reversed, as they are throughout the receiver ON time, all diodes of the bridge are biased beyond cutoff, and the gate has no effect on the receiver circuit. The third stage of isolation is a receiver gate which blocks any leakthrough during transmit.

The pulses that control the action of the diode bridge shorting gate are obtained from the gate pulse amplifier. The start time of these pulses is chosen to delay the turn on of receiver circuits until the ringing or reverberation in the acoustic antenna caused by the powerful transmit tone burst has decayed to a negligible level. This is about 100 msec from the end of the tone burst for the fiberglass horn-reflector antenna now in use, and is equivalent to about the first 16 m of echo range. It should be possible to reduce the reverberation time and subsequent loss of close-range echoes by future improvements in acoustic antenna design. The stop time of the pulses from the gate pulse amplifier is set to turn off the receiver circuits a few msec before the start of the next transmit tone burst. This is done as a precaution to insure against possible damage to the receiver circuits by the initial transient of the transmit signal.

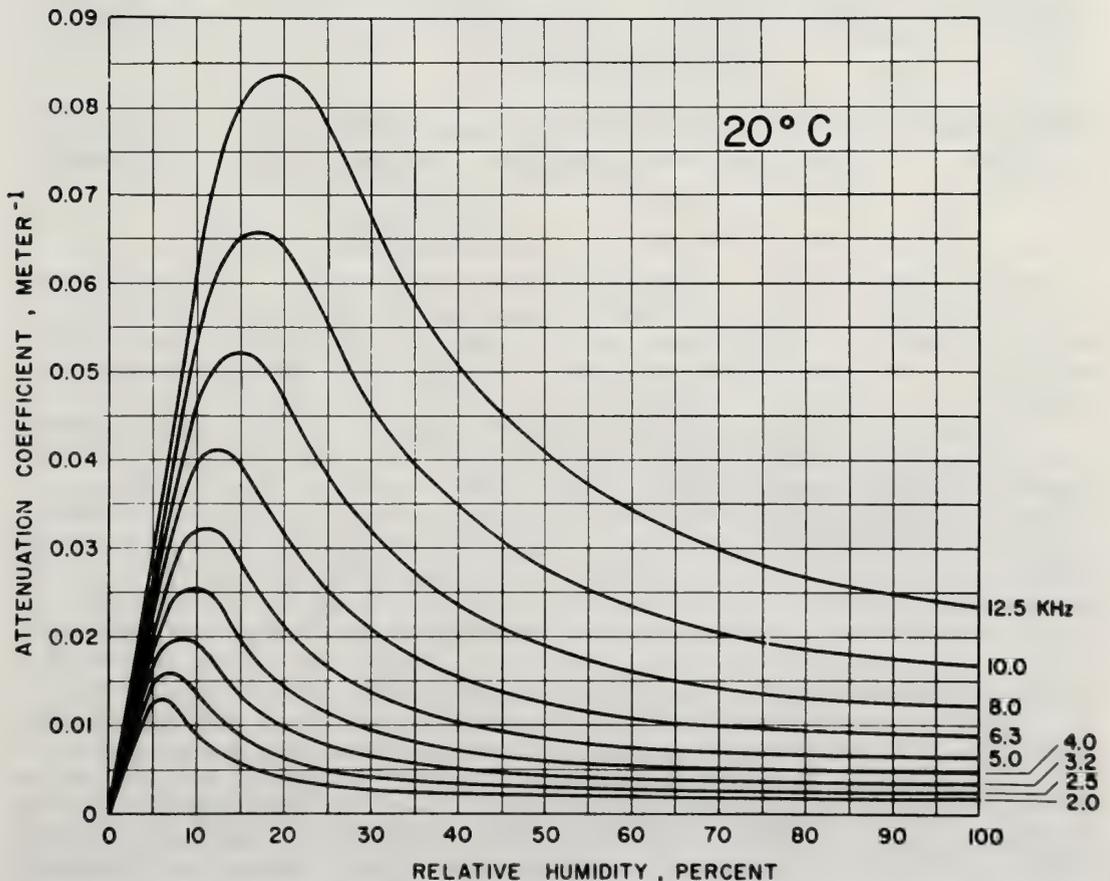


Figure 18.1 Molecular attenuation coefficients for acoustic waves as a function of humidity, for various frequencies (after Little, 1969).

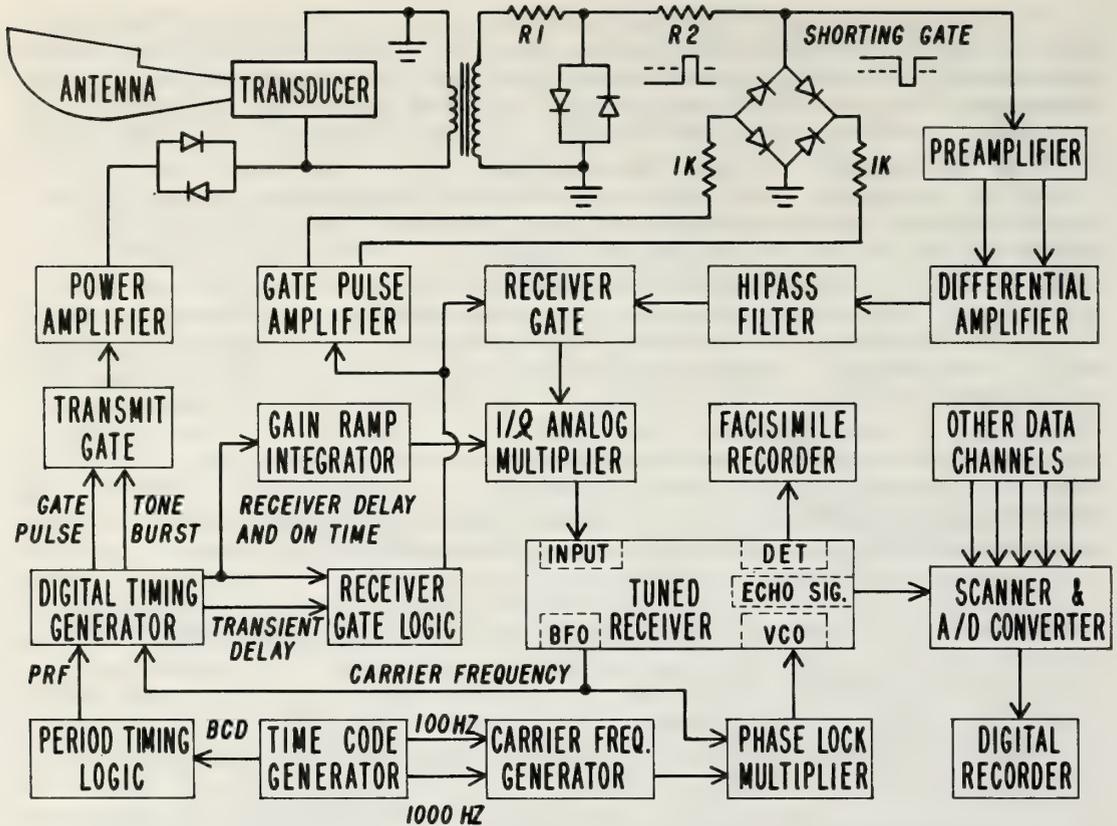


Figure 18.2 Block diagram, Mark II echo sounder (after Simmons, et al., 1971).

During reception of echoes the gating actions described earlier enable the acoustic transducer to act as a microphone which is transformer coupled to a variable-gain preamplifier as shown in (F18.2). The coupling transformer has a 30 to 1 step-up turns ratio which serves to increase the irreducible Johnson noise voltage of the microphone to a level that substantially exceeds the input self-noise of the preamplifier.

The preamplifier is located remotely at the antenna-transducer site in order to increase echo signal levels to about 1 volt rms before transmission through over 100 m of cable to the rest of the receiver system. For this reason the preamplifier includes a calibrated, remote gain control subsystem and a balanced, low-impedance, differential output stage. Preamplifier gain may be varied from 60 db to 0 db by varying a dc gain control signal from 0 to +5 volts. The differential output signal of the preamplifier is delivered through the long, multiconductor cable to the differential amplifier of the receiver system. This amplifier converts the balanced and therefore relatively hum-free signal from the preamplifier to a single-ended signal for subsequent passage through an 800 Hz highpass filter. The purpose of the highpass filter is to eliminate strong, low-frequency wind noise, man-made noise and other low-frequency components of background noise that might otherwise overload subsequent stages of the receiver system.

The output of the highpass filter is applied to the receiver gate. This is a redundant gate that operates in unison with the previously described diode bridge shorting gate. Both gates are controlled by the output pulse from the receiver gate logic section. The purpose of the receiver gate is to block a small leakthrough signal that develops during the transmit tone burst. The leakthrough path is in the multi-conductor cable that interconnects the remotely located antenna, transducer and preamplifier sections of (F18.2) to the rest of the sounder system. The cable consists of separately shielded pairs of conductors to minimize cross-coupling effects among transmit, receiver and gate control signals. Since up to 300 m of cable may be used, however, there is still enough cross-coupling to warrant the use of the redundant receiver gate.

The echo signal from the receiver gate is applied to one of the two inputs of the $1/\ell$ analog multiplier. The other input to the multiplier is a linear ramp signal generated by the gain ramp integrator. The output of the $1/\ell$ analog multiplier at any time is simply one tenth the product of the applied echo signal and the instantaneous amplitude of the linear ramp. The linear ramp from the integrator always starts at zero volts dc and grows with time at the rate of 1 volt/sec. This causes the multiplier to act as an echo signal amplifier the gain of which periodically starts at zero and increases linearly with time at the rate of 0.1/sec. By the end of a 10 second echo receiving period, for example, the gain of the $1/\ell$ analog multiplier has been swept from zero to unity. This type of gain sweep compensates for the spherical divergence or spreading loss of acoustic echoes as they are received from progressively longer ranges. The gain is varied inversely with range ($1/\ell$) rather than as the inverse square of range ($1/\ell^2$) because the output signal of the receiving transducer is proportional to acoustic pressure rather than acoustic power.

The range-compensated echo signal from the $1/\ell$ analog multiplier is applied to the input of the tuned receiver. The receiver is constructed so that the center of its pass band automatically coincides with the frequency of an integral BFO that is the source of carrier frequency for the sounder. The receiver tuning is thus locked to the mean frequency of the echo signal. Receiver bandwidth is adjustable, and normally is set to accommodate just the major sidebands of the transmitted tone burst in order to achieve an optimum S/N with minimal loss of resolution. For example, a bandwidth of 100 Hz generally has been used for tone bursts lasting 20 msec, whereas a bandwidth of 10 Hz has been used for tone bursts of 200 msec and longer.

The tuned receiver has two outputs. One of these is an amplified replica of the echo signal as bandwidth-limited by the receiver tuned circuits. This is the type of signal that is used for determining Doppler shifts of frequency between a transmitted tone burst and its received echo. It is planned to digitize this signal (and others) before recording in order (1) to eliminate frequency errors associated with analog tape recorder wow and flutter, and (2) to shorten the number of steps required for subsequent digital data analysis. The digital recording scheme, though not yet completed, is shown connected to the echo signal output of the tuned receiver on (F18.2). The other output of the tuned receiver is the detected level of the echo signal. This is a fluctuating dc signal the amplitude of which is proportional to the acoustic pressure of echoes received after correction for spherical divergence. This signal is displayed as a function of range and time of day by means of the facsimile recorder.

The acoustic antenna is an important component of the system since it serves to collimate the transmitted power into a narrow beam in the transmit mode, and serve as a highly directional microphone in the receive mode. One highly efficient type is the cone-paraboloid antenna originally designed for microwave use. The electro-dynamic transducer is bolted to the cone section in place of the microwave conical wave guide. The only other modification of the antenna found desirable is to coat it with an acoustical dampening material to limit reverberation time after the transmitting pulse is turned off. Such an antenna is shown in (F18.3). Other types of antennas which have been used in acoustic sounders include arrays of electro-dynamic speakers, and transducers mounted at the focus of parabolic dishes.

A desirable feature of any antenna is the limitation of side lobes so that unwanted acoustic power is not directed elsewhere and so that noise in the environment surrounding the sounder does not affect sounder performance. The measured sidelobe characteristics of a cone paraboloid antenna are shown in (F18.4) for different acoustic frequencies. It will be seen that at a frequency of 1 kHz, the half power beam width is about 16° , whereas at 5 kHz the main beam is measured to be only 3 or 4 degrees wide to the half power point. Note also that the 90° sidelobe is 38 dB below the main beam at 1 kHz and nearly 50 dB below the main beam at 5 kHz.

One easy way to improve sidelobe rejection is to surround the antenna with an absorbing cuff. For a vertically pointed antenna it has been found that an open cylinder of haybales is an effective absorber. If the cylinder aperture is several times that of the antenna and the height several times the antenna aperture, 20-30 dB additional 90° sidelobe attenuation is obtained. An analysis of this absorbing cuff has been performed by Strand (1971) and the theoretical improvement closely checks with that experimentally measured in the field. Further improvements should be possible with internal attenuators in the horn to achieve the effect of a tapered feed in the microwave case. The present antenna acts nearly like a perfect piston source, but by decreasing the amplitude of the pressure fluctuations near the edge of the aperture, less energy will be diffracted at the edge, further improving the sidelobe rejection.

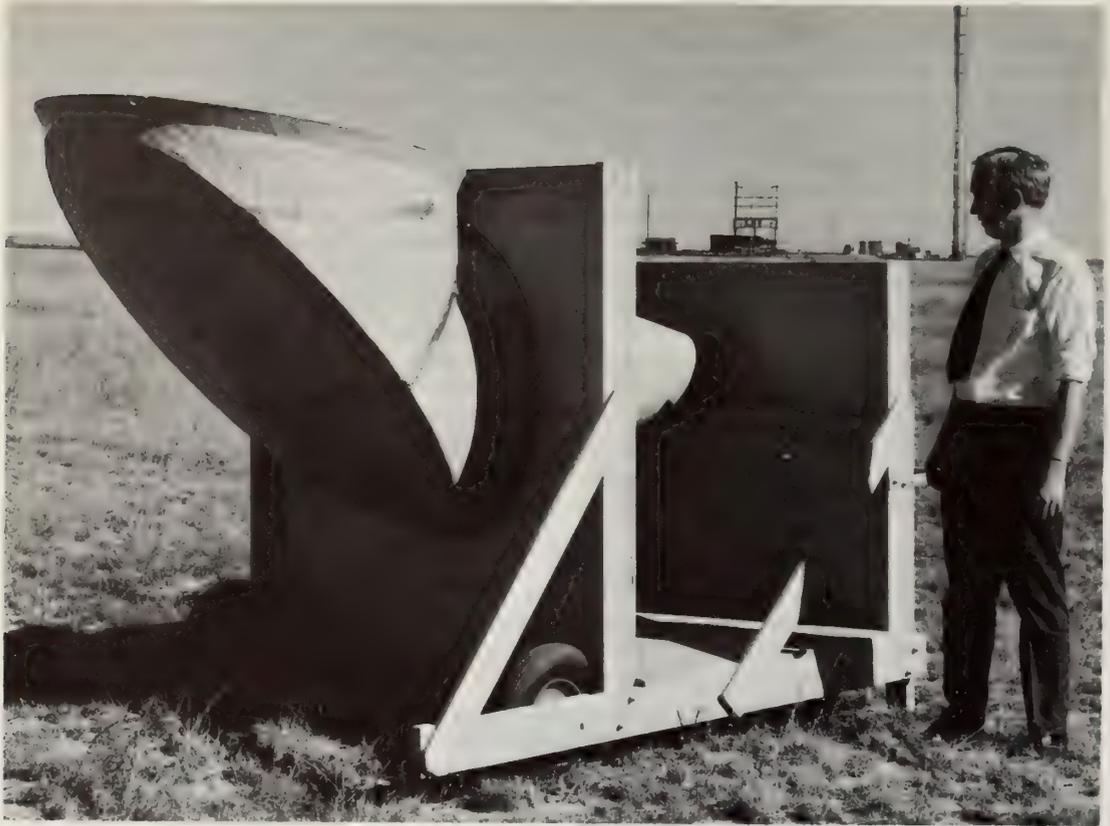


Figure 18.3 *Fiberglass conical horn, parabolic reflector antenna, highly suitable for acoustic echo sounding. (After Wescott, et al., 1970).*

18.4 Structure Observed in Returned Echo

With a co-located acoustic transmitter and receiver, the backscatter from the atmosphere is caused only by thermal structure which may be either random temperature inhomogeneities or uniform temperature gradients. This effect has been discussed earlier in connection with (18:13). The temperature structure parameter is not a usually measured atmospheric variable, and we still have much to learn about its magnitude and variation in the atmosphere. It is not surprising that considerable temperature structure is observed near midday when the atmosphere is convectively unstable, for under such conditions heat is added rapidly to the low lying levels by conduction from the solar heated ground. Typical acoustic returns obtained under convective conditions will be discussed next, followed by an investigation of the echo returns from the stably stratified atmosphere.

18.4.1 Convective Plumes

The typical pattern observed on the facsimile recording showing back-scattered intensity with height as a function of time, for the unstable atmosphere is shown in (F18.5). This is the thermal plume structure, indicating the rise of heated and thermally nonuniform air parcels from the ground due to surface heating from solar radiation. That the upwelling air is characterized by variation in temperature is well known from investigations with fast response temperature sensors on aircraft, such as the study by Myrup (1967). Myrup also shows how the variation of C_T with height may be used to differentiate between different convective models. With a well calibrated acoustic sounder, and if the atmospheric acoustic attenuation is known, it is possible to measure C_T as a function of height and compare results with the Townsend convective theory, where a variation as the -0.75 power with height is predicted, or with similarity theory where a -0.67 power law is found (Myrup, 1967). Acoustic sounding is just to the stage where it will be possible in the near future to complete such studies. By comparing the acoustic

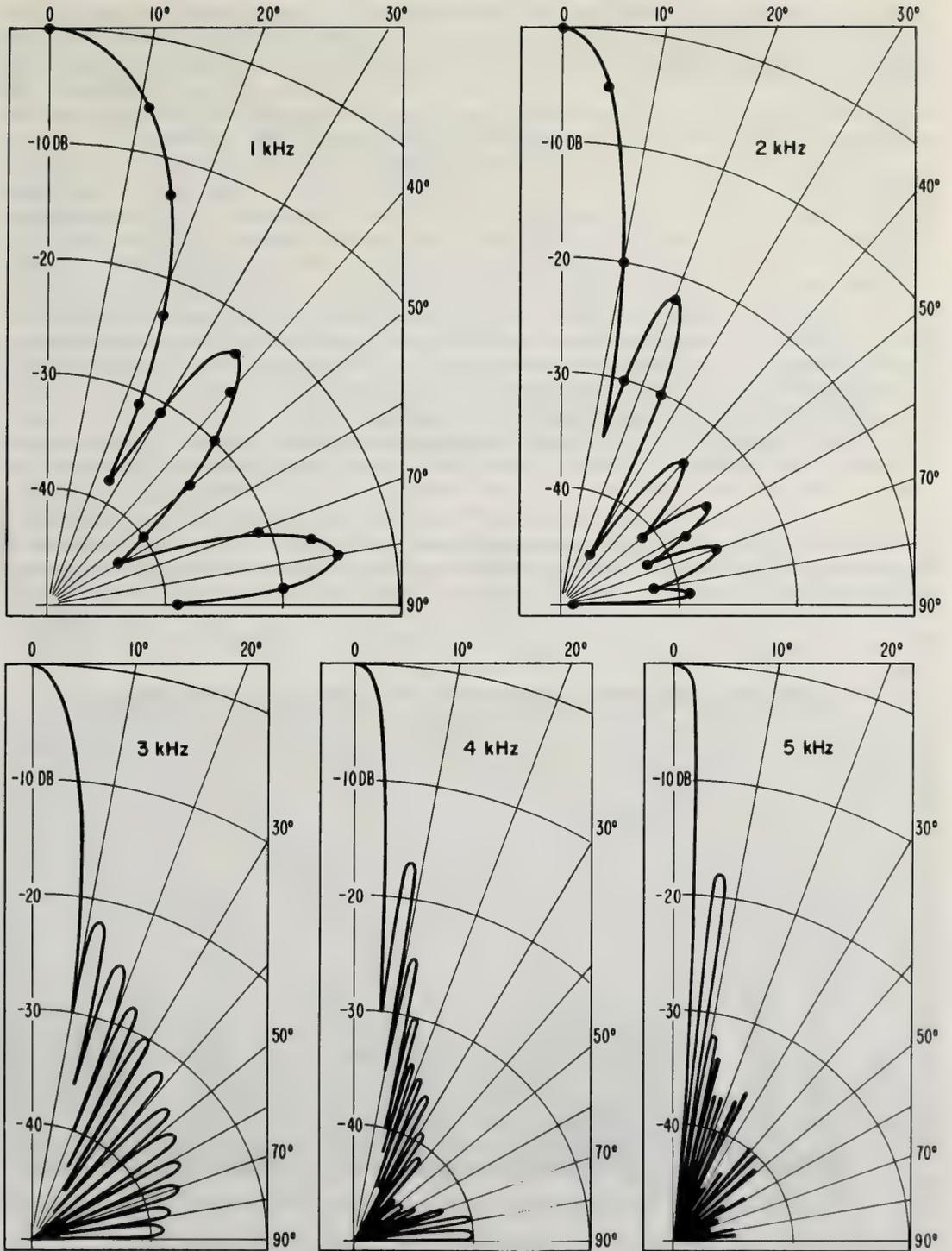


Figure 18.4 Polar diagrams of the relative gain pattern of the horn- for acoustic antenna (after Simmons et al., 1971).

sounder measurements of C_T with temperature differences recorded by fast response thermometers, spaced one-half of the acoustic wavelength apart, on a nearby micrometeorological tower, we are now studying the accuracy of the acoustic measurements. Preliminary results indicate rather good agreement, within 50%, between the sounder and tower measurements of mean C_T , and even in the variance of this parameter. The great advantage in using the sounder over tower or aircraft measurements is that the entire spatial structure of the plume is visible, instead of one or several finite horizontal slices through the structure.

The sounder facsimile record clearly shows the vertical organization of the convective plumes, seemingly putting to rest the controversy over the existence of plumes as opposed to isolated heated bubbles of rising air. However, regions of higher C_T do occur within plumes, presenting a combined bubble-plume picture. One improvement which would be advantageous is more rapid transmit damping to enable the detection of echoes from lower levels in the atmosphere than the present minimum of 30 m. It is in these lower levels below 30 m that wind shear frequently produces a marked slope in the plume structure, as observed by Kaimal and Businger (1970). Plumes detected with the sounder in the levels above 30 m are observed to slope randomly into or with the wind at a prairie location. Little wind shear is observed between 30 m and 300 m height when convection is vigorous on sunny days. Wind data for such field experiments have been measured by means of an anemometer on a tethered balloon. It therefore appears that thermal plumes significantly modify wind fields in the boundary layer.

With the calibration of the sounder well understood to an estimated accuracy of 3 dB, the facsimile recordings of plume structure can be used to study the diurnal and seasonal variation of plume height. We define a plume as being that column of the atmosphere where the thermal structure parameter $C_T > 4 \times 10^{-2} \text{ } ^\circ\text{K cm}^{-1/3}$. In (F18.5), the darkest areas in the plumes correspond to C_T values as large as $14 \times 10^{-2} \text{ } ^\circ\text{K cm}^{-1/3}$. By overlaying the records of plumes with a transparent grid, it is possible to count off the percentage of time when the plume is discernible as a shade of light grey or darker, corresponding to the discriminating C_T value defined above. The results of such an analysis are shown in (F18.6) where the increase in plume height from early morning hours to a maximum shortly after noon is documented by the three curves drawn. This data was taken at Haswell, Colorado, a flat prairie site 200 km southeast of Boulder, on a sunny day with light and variable winds below 2 ms^{-1} .

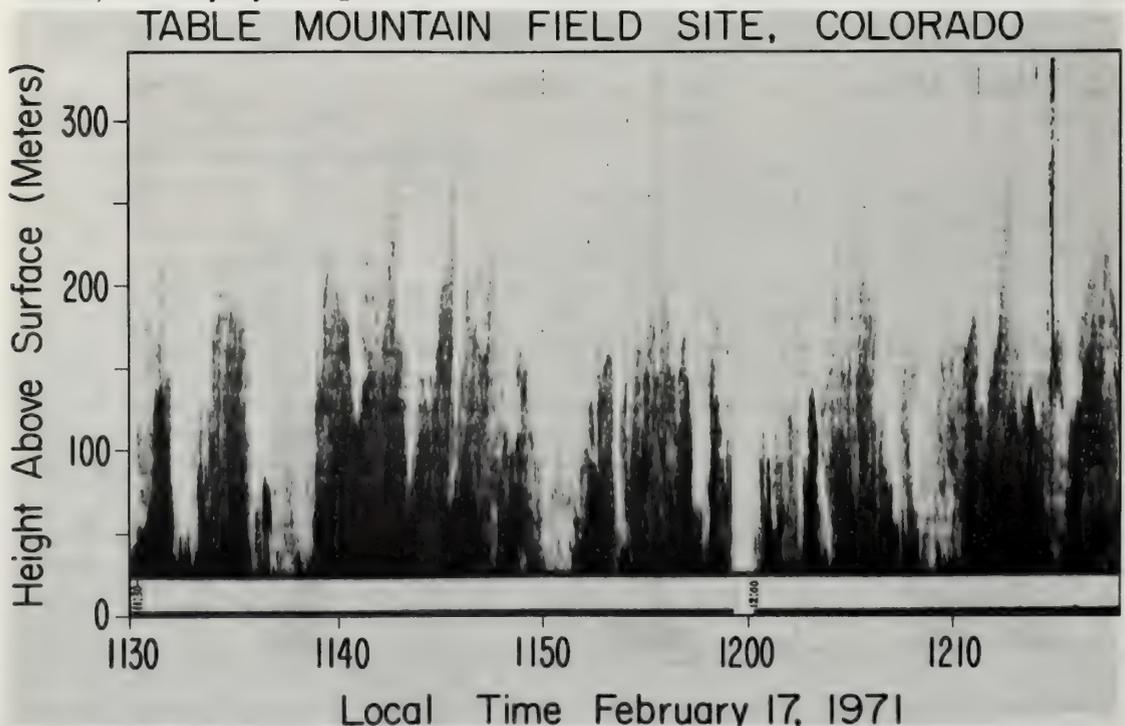


Figure 18.5 Facsimile recording of acoustic echo intensity from the unstable atmosphere, showing convective plume structure. (After Hall, et al., 1971).

After the plumes reach a maximum height shortly after noon, the height begins to decrease but on a slower time scale than the rise in height during the morning hours. The 1400-1500 plume height characteristic is nearly the same as the 1000-1100, and the 1600-1700 late afternoon plumes have nearly the same characteristic as the 0900-1000.

Turbulent mixing at the edges of the plume, and turbulence within the plume, reduces the C_T with height as shown in (F18.5). The lower regions of the plume are clearly a darker grey on the facsimile record than the upper regions. It must be emphasized, however, that C_T is a class of atmospheric variable about which we have much to learn. The thermal structure parameter does not exactly correspond with plume vertical velocity, as verified by acoustic Doppler measurements of vertical winds. Therefore low values of C_T do not necessarily delimit the convective zone in the boundary layer, but there is a clear correlation between convection and the C_T . Further theoretical and experimental work will be required to determine how the unstable convection region interacts with thermal inversions.

A second way to study the characteristics of thermal plumes is to investigate seasonal effects on plume height. In (F18.7), the 1200-1300 plume height characteristics are examined for three dates in the summer, fall, and winter months. All data were taken over Colorado prairie land on sunny days with winds 0.2 ms^{-1} . They show surprising little variation in plume characteristics with season. This somewhat surprising result needs further study, for the angles of solar radiation incidence differ by more than 50° between the extreme seasonal cases. Surface temperature gradients, or in other words the lapse rate in the lowest few tens of centimeters does not change greatly with season, (Whitman and Wolters, 1967), and this may be the explanation.

The facsimile recordings of C_T serve as a useful pictorial representation of the temporal structure in the atmosphere, and can be used for semi-quantitative work, as in finding gradients in thermal structure. To assign absolute values to C_T , it is necessary to have a more accurate means of recording the back-scattered sound intensity. We have used analog magnetic tape recording for this purpose to date. These analog records

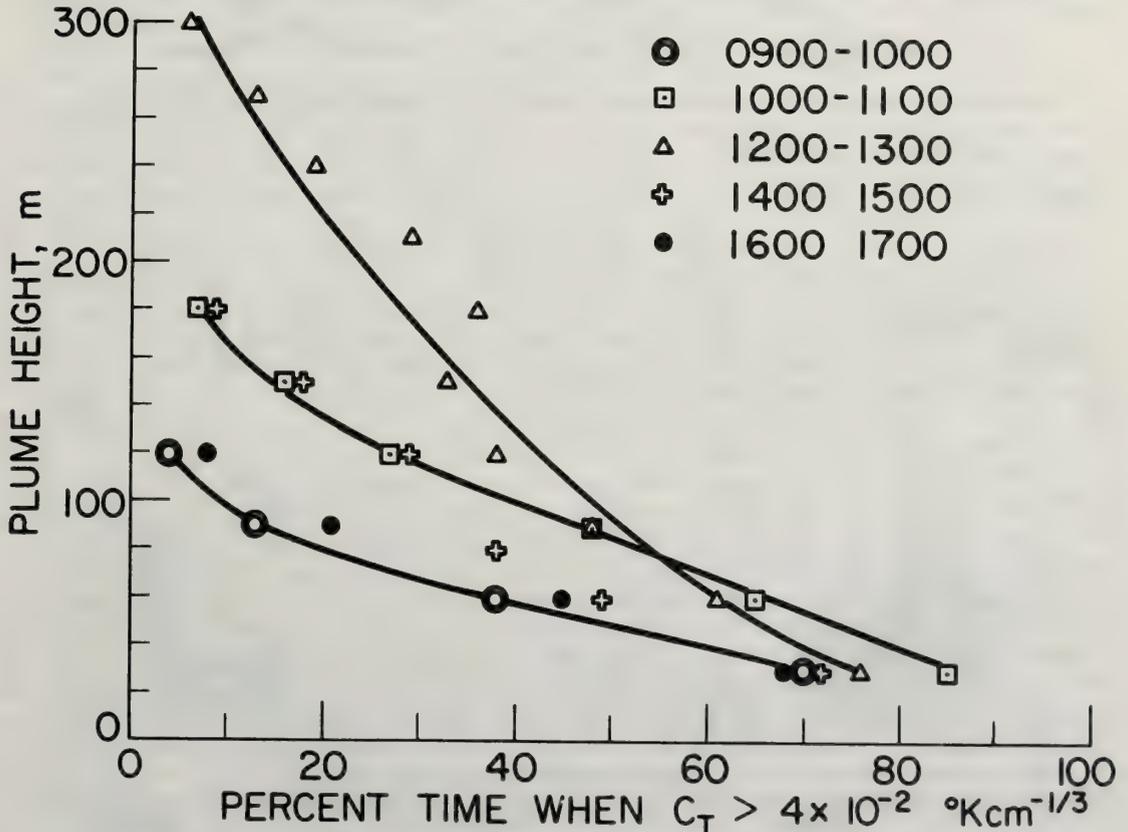


Figure 18.6 Diurnal variation in height of convective plumes (after Hall, et al., 1971).

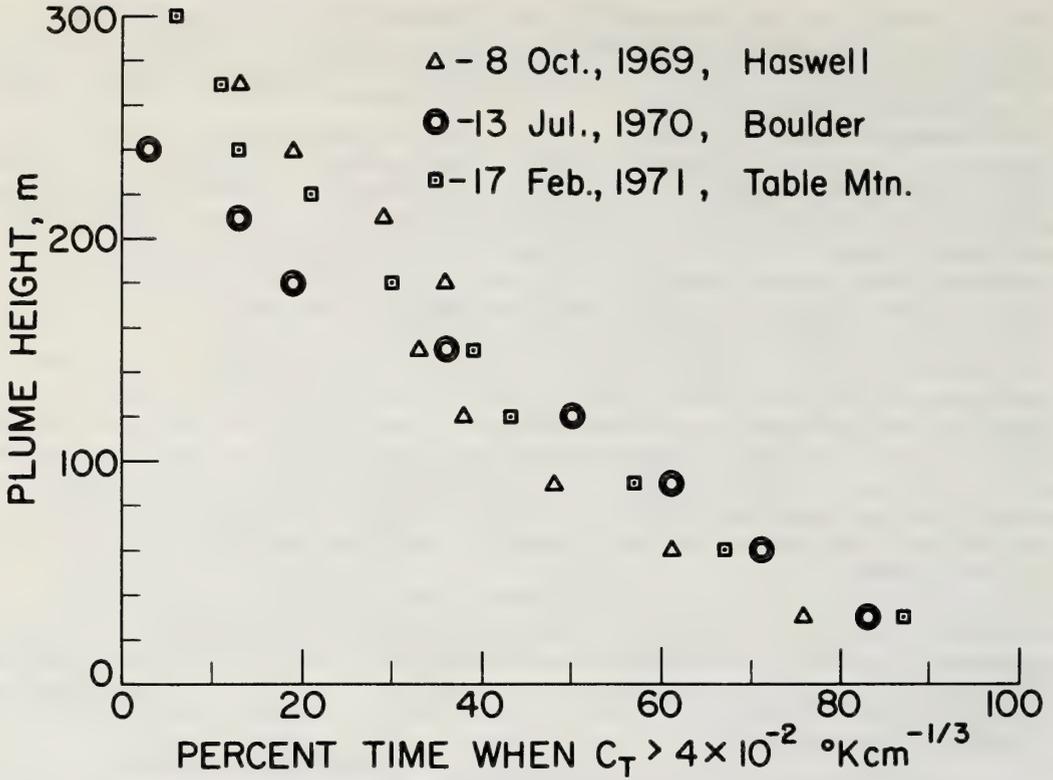


Figure 18.7 Seasonal variation in height of convective plumes at midday (after Hall, et al., 1971).

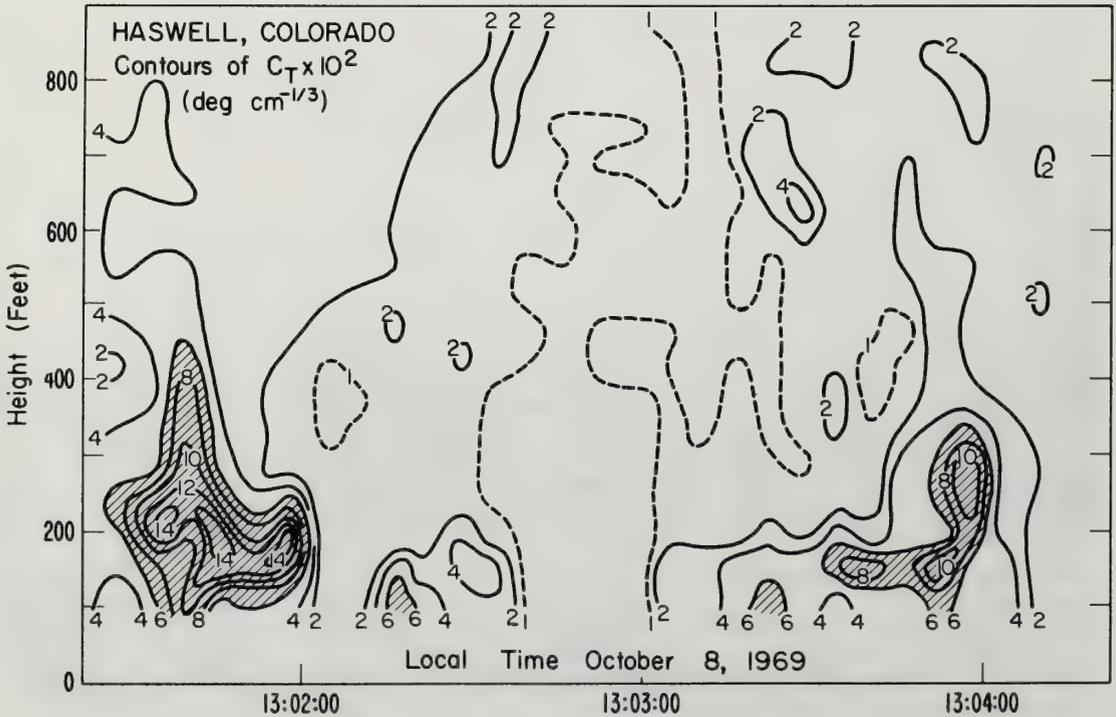


Figure 18.8 Contours of equal values for the thermal structure parameter in convective plumes.

are returned to the laboratory for digitization and computer processing, introducing in the computational scheme atmospheric humidity and temperature effects, to account for molecular absorption, and the echo sounder system parameters so that quantitative values of C_T may be derived. A plot of temperature structure detected acoustically in convective plumes in shown in (F18.8). When more extensive data is available, it will be possible to obtain a good statistical average of C_T structure with height for comparison with differing convective theories. The scattering equation (18:13) and acoustic "radar equation" (18:15) are used to relate backscattered power to C_T .

18.4.2 Temperature Inversions

A quite different type of acoustic sounder return facsimile display is obtained under stable conditions in the boundary layer. On a typical clear night, thermal radiation from the earth's surface results in the ground having a lower temperature than air immediately adjacent to it. This is because the infrared emissivity of the solid earth is much greater than the atmosphere. The dryer and more clear the atmosphere, the poorer it is as an infrared radiator. Turbulence or molecular conduction within the low lying levels of the atmosphere causes a heat loss to the earth, and a cold layer of air forms just above the solid surface. As heat conduction losses continue throughout the night, the depth of this cold layer of air increases, usually as the square root of the time interval after sunset (Sutton, 1953). If the conduction of heat occurred principally on the scale of molecular diffusion, one would not expect much random thermal structure, and we might predict that the acoustic sounder would detect only the temperature gradient separating the cold, surface air from the warmer air above. Under most atmospheric conditions, there is more heat conducted through turbulent eddy diffusion, so that throughout the deepening cooler air a large value of C_T is observed.

A typical acoustic facsimile display illustrating the formation of a nocturnal radiation inversion is given in (F18.9). The last vestige of convective plume structure dies away at about 1700 hours and the increase in the radiation inversion depth is indicated by the thickening lower dark trace beginning at 1900 hours. The blanked-out echo in the lowest 30 meters prevents detection before this time. As time progresses, the depth increases through 0200 hours. Minor variations in inversion height are probably caused by changes in wind conditions, and mixing at the top of the temperature inversion. After 0200 it appears there was sufficient mixing, possibly from wind, to dissipate the initially formed inversion structure, and to lead to the formation of yet another nocturnal inversion growing from the surface after 0230.

A second type of temperature inversion is caused by large scale downward motion, or subsidence, within the troposphere. The relatively rapid descent of air parcels in subsidence results in adiabatic heating of the air. By the time a subsiding air mass has approached the surface, its temperature significantly exceeds that of the lower lying levels. This is because typical lapse rates in the atmosphere are less than dry adiabatic. The subsiding air is prevented from reaching the earth's surface by the trapped, stable, lower levels so that horizontal divergence of the subsiding air mass frequently occurs above the low layers, leading to wind shear and turbulent mixing of the warmer upper air with the underlying levels.

The acoustic detection of the formation of a subsidence inversion is also shown in (F18.9), entering the picture at a height of 700 meters at 0240 hours. As the subsidence continues through 0800, the increasingly dark acoustic return indicates an increase in the thermal structure. The wavy nature of the return is probably caused by wind shear induced instabilities or Kelvin-Helmholtz billows (Sekera, 1948). The merging of the subsidence inversion and the radiation inversion after 0800 resulted in a temperature differential across the inversion of 10°C , as measured by radiosonde ascent in Denver, 40 km from where the record was obtained at Table Mountain, Colorado, just north of Boulder. The eventual weakening and breakup of the inversion by convective thermal plumes may be noted after 1000 with all traces of the inversion gone by 1200. This timing also matched the observed inversion breakup in Denver, and clearly indicates how the sounder may be used to monitor mesoscale phenomena which may affect points separated by many kilometers. (The dark vertical traces in the facsimile (F18.9) at the upper levels after 1000, and 1400 were caused by insect noise, while the darkening of the trace after 1600 was caused by increasing wind noise. These two sources of background noise sometimes place limitations on acoustically detected atmospheric variables.)

A second example of temperature inversion structure is shown in (F18.10). The dramatic wave structures and fine lamina of C_T indicate that boundary layer temperature inversions can be highly complex. If the depth of an inversion is to be accurately monitored, a single radiosonde ascent with temperature sensor could easily be in error should the balloon borne instrument pass through the crest or trough of one of the gravity waves frequently observed. The study of this same inversion, correlating microbarograph detected pressure variations with the acoustic record, is discussed by Hooke (7.2) in this volume.

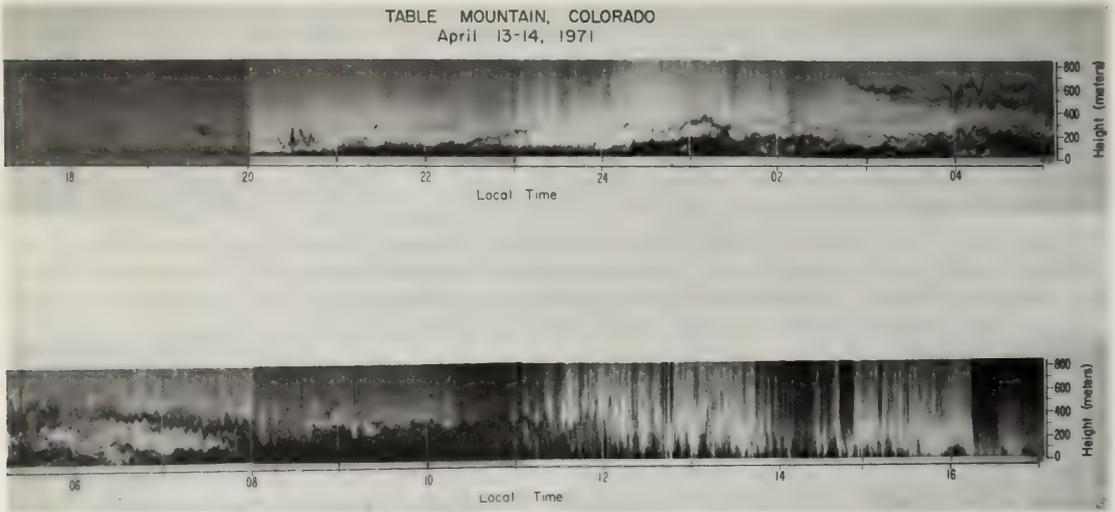


Figure 18.9 Facsimile record of 24 hours of acoustic echo sounding, illustrating returns from radiation and subsidence inversions, and the dissipation of an inversion by convection.

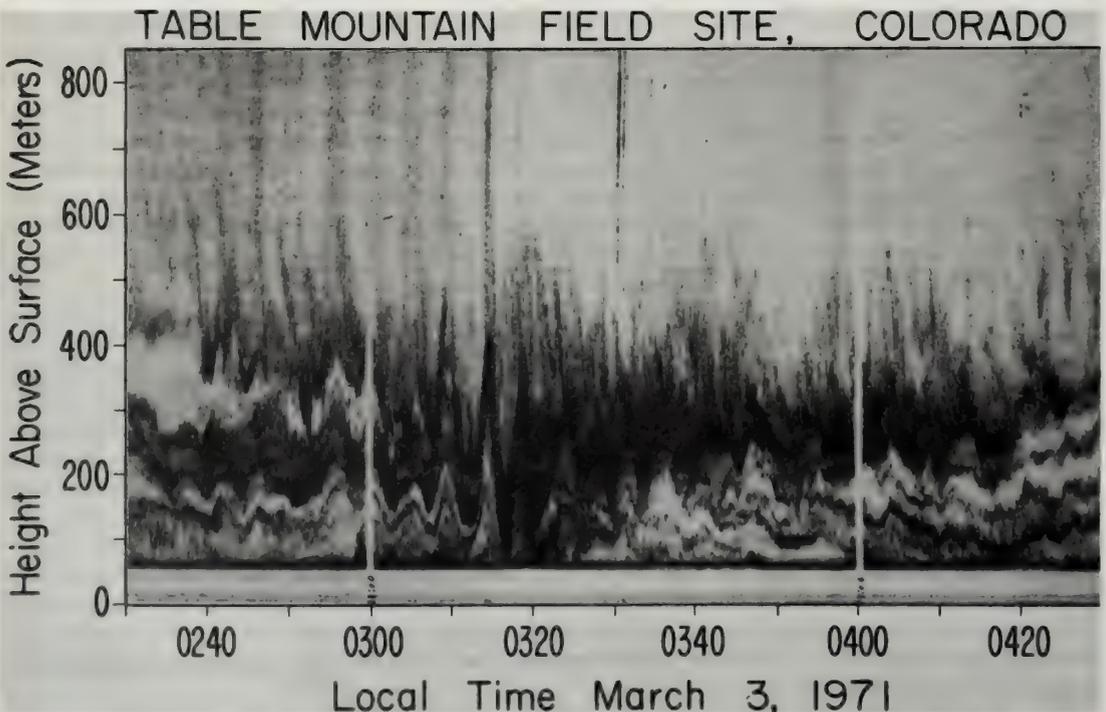


Figure 18.10 Facsimile record of acoustic returns from a stable atmosphere, showing gravity waves (after Hall, 1971).

Comparisons of acoustically measured C_T with tower measurements, under temperature inversion conditions, sometimes do not agree as well as when the atmosphere is convectively unstable. When thin, horizontal laminae are present, the acoustic technique seems to estimate consistently too high a value of C_T , sometimes 100 percent higher than the value measured with differential thermometers on a tower at the same height. It is believed that this lack of agreement is caused by the inadmissibility in the scattering theory of assuming isotropic turbulence under the stably stratified inversion conditions, so that a Kolmogorov spectrum of turbulence may not exist. Also, since the layers are nearly horizontal, and gradients of temperature present nearly specular surfaces to the advancing wave fronts in the far field of a vertically pointed acoustic antenna, some partially coherent scattering may occur to enforce the returned echo strength. When sharply defined lamina are not present below an inversion, tower and sounder measured values for C_T agree nearly as well as for the convectively unstable atmosphere or within 50 percent.

18.5 Wind Studies by Acoustic Doppler Methods

So far we have considered only the atmospheric structure revealed by the *intensity* of the backscattered signal. By analyzing the frequency content of the returned signal, and comparing this with the transmitted frequency, one can derive the velocity of the air parcels along the acoustic beam, from the Doppler frequency shift in the scattered sound. Consider a vertically pointed sounder. The vertical velocity, w , of the scattering volume along the direction of the beam is given by

$$w = \frac{c}{2} \left(\frac{\nu_O}{\nu_s} - 1 \right) \quad (18:16)$$

where c is the velocity of sound, ν_O is the transmitted carrier frequency, and ν_s is the frequency of the scattered signal. This equation is correct to first order accuracy, where wind velocity is small compared to the velocity of sound. This is nearly always true for vertical velocities in the atmosphere, and certainly for those velocities measured under nominal convective plume conditions. As pointed out by Beran, et al. (1971), the much slower speed of acoustic waves compared with electromagnetic waves can introduce some errors into the acoustic Doppler measurements. If the total velocity of the medium along the path to the scattering volume is varying with time, which is usually the case in the atmosphere, additional frequency shifts caused by beam bending may be added to the true Doppler shift produced by the radial motion of the target. Fluctuations in the vertical velocity of the air along the vertical path of the sound do not, however, produce spurious frequency shifts, provided that the vertical velocity field does not change between the upward and the downward traverse of the sound pulse through the medium. On the other hand, temperature fluctuations produce phase changes of equal magnitude and sign in both propagation directions, and do not cancel out. Overall, spurious propagation-induced frequency shifts should not produce errors in derived vertical velocities of more than about 0.1 ms^{-1} for the heights which concern us.

The frequency spectrum of the returned signal is compared with that of the original transmitted pulse to obtain the Doppler frequency shift. The method of obtaining the spectra differs slightly in the acoustic and electromagnetic cases, primarily because of the characteristically much higher pulse repetition rate used with electromagnetic radar. The final spectrum generated from a radar return is extracted from information contained in each of many separate pulses and it is this spectrum which is used to reproduce the Doppler frequency; in the acoustic case, a spectrum can be generated for each pulse. The higher pulse repetition rate of the radar does not increase the accuracy if the signal-to-noise ratios and overall duty cycles (fraction of time that the transmitter is on) are equal in the two cases. A further improvement can be made by averaging the acoustic spectra from succeeding pulses, although this reduces the horizontal resolution by increasing the size of the total volume being considered because of the ambient air motion during the time required to collect the sample.

A typical Doppler derived vertical velocity profile for a convective plume is illustrated in (F18.11). On the left is the backscatter intensity depiction of the thermal plume structure and on the right are the profiles of vertical velocity in ms^{-1} .

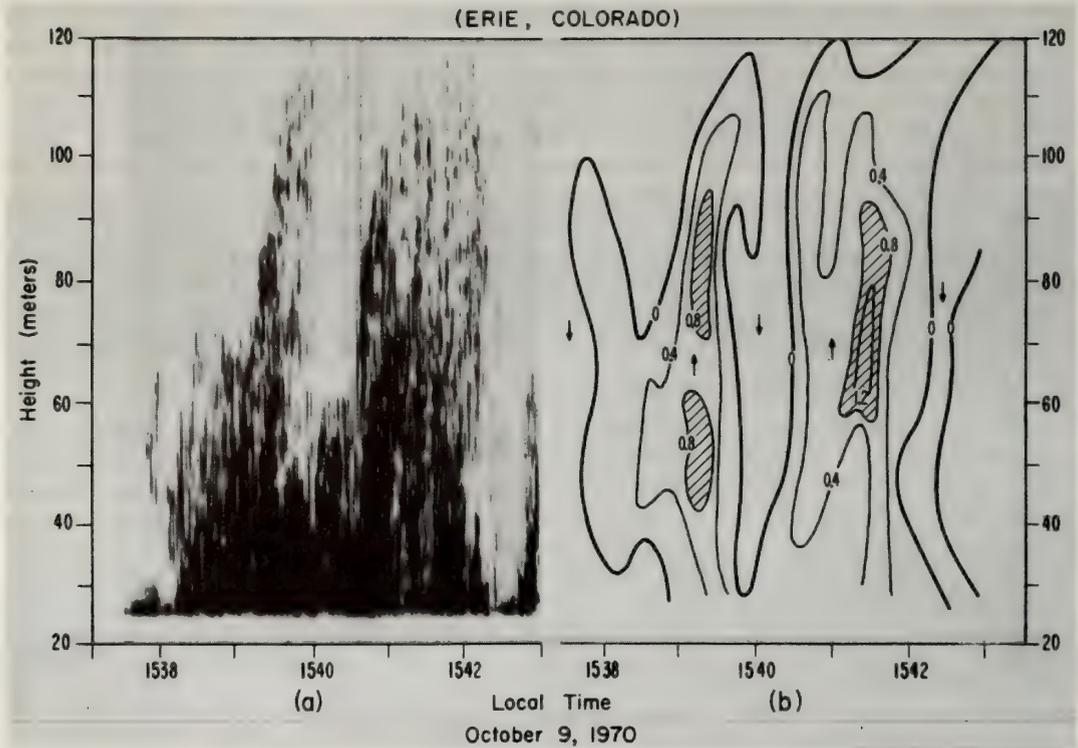


Figure 18.11 (a) Facsimile record of convective plumes, and (b) Doppler detected wind in ms^{-1} , for the same plumes as in (a) (after Beran, et al., 1971).

Studies of the vertical velocity observed under breaking gravity wave conditions when strong temperature inversions are present have revealed vertical velocities of the same order of magnitude as those observed in convective plumes, or several meters per second in both the positive and negative z direction. Further studies of this type, especially when coupled with microbarograph arrays for detection of gravity wave propagation direction, will add much more to our understanding of the dynamics of such wave motion in the boundary layer. The importance of wave energy in boundary layer dynamics has not been extensively investigated but it may turn out that wave energy budgets are of comparable importance to the energy transported through convection or dissipated through turbulence.

By employing three acoustic echo sounders in the geometry illustrated in (F18.12) it is possible to measure the Doppler component along each of the three beams. In this manner, the total wind vector may be measured in the volume intersected by all three beams. By subtracting the vertical component from horizontal components measured elsewhere on the sloping beams, and averaging over suitable periods so that the assumption of vertical velocity stationarity is valid, the vertical profile of the total wind vector may be measured throughout the boundary layer with such an arrangement. A wind sensor flown on a captive balloon borne Boundary Layer Profiler (BLP) can provide for in situ verification of the wind measurement accuracies.

The derivation of the total wind vector using three Doppler acoustic sounders has been given by Beran and Clifford (1972). The derivation of total wind speed from a Doppler shift can best be explained by referring to (F18.13). Sound transmitted from an antenna at T and having a frequency ν_0 impinges upon some scattering volume at O. If this scattering volume is moving with the velocity, \vec{V} , the frequency of the sound wave received at R will be some new value, ν_s , and we can derive a difference frequency, $\Delta\nu = \nu_s - \nu_0$, proportional to \vec{V} . With the transmitted frequency expressed as a wave vector \vec{K}_0 and the scattered wave as \vec{K}_s , the difference frequency, or Doppler shift, can be written in vector notation as

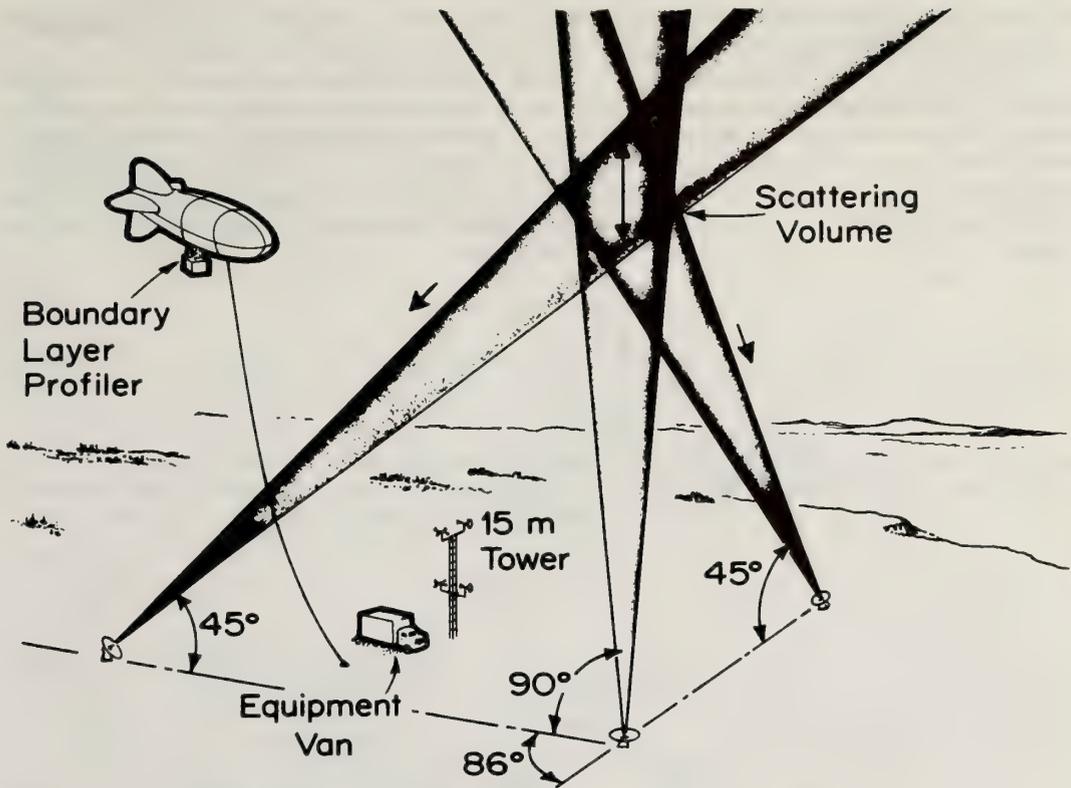


Figure 18.12 Arrangement of three acoustic echo sounders, with supporting equipment, used to measure the total wind vector by analyzing the Doppler shift in back scattered sound (after Beran and Clifford, 1972).

$$\Delta v = \frac{1}{2\pi} (\vec{K}_s - \vec{K}_o) \cdot \vec{V} \quad (18:17)$$

where \vec{V} is the wind component in the plane formed by the transmitted and received beams. Writing (18:17) in terms of the wavelength λ_o of the carrier wave, the angle β between the total wind vector \vec{V} in the plane and the component being measured, and the scattering angle θ we have

$$\Delta v \approx \frac{2V}{\lambda_o} \sin\left(\frac{\theta}{2}\right) \cos \beta \quad (18:18)$$

where $V \cos \beta$ is now the magnitude of the wind resolved along the vector $\vec{K}_s - \vec{K}_o$. Note that the direction of this component is the bisector of the angle formed by the intersection of the transmitter and receiver beams and will only be truly radial for a monostatic system.

If λ_o is replaced by c/v_o where c is the speed of sound, and the magnitude of the wind component being measured is solved for, we have

$$V \cos \beta = v' = \frac{c}{2 \sin \frac{\theta}{2}} \left[\frac{\Delta v}{v_o} \right] \quad (18:19)$$

This development assumes that refraction or bending of the beams is negligible. The effects of refraction and a more detailed development of the Doppler equation are given by Beran and Clifford (1972).

Using (18:19) and the knowledge that the direction of this wind component will be along the bisector of the angle formed by the antenna axes, several equipment configurations can be envisioned for measuring the wind vector. For example, using a single, vertically pointing, pulsed, monostatic system, a profile of the vertical wind vector can be measured by selecting gates at successive distances along the antenna beam. The vertical wind can also be measured at a single height by using a bistatic system with the elevation angles (θ_t and θ_r in F18.13) set equal. The first method was demonstrated by Beran, Little, and Willmarth (1971).

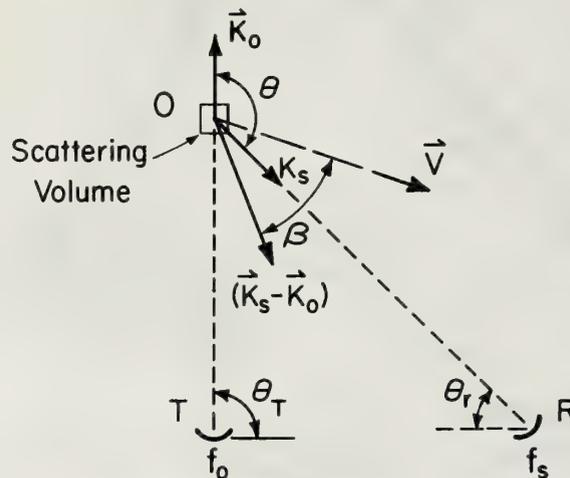


Figure 18.13 Vector diagram showing resolution of the wind in the plane defined by two sounder antenna beams (after Beran and Clifford, 1972).

Clearly, in order to sense a component of the horizontal wind, the axis of Doppler resolution must be off vertical. For the monostatic system, this simply requires that the beam be tilted. For the bistatic case, the two antennas must be arranged with asymmetric elevation angles. A second experiment by Beran and Willmarth, (1971) tested both tilted monostatic and bistatic systems. It was concluded from this work that while the monostatic system gives better accuracy for a given elevation angle (because the component being measured is closer to the horizontal), the bistatic configuration is superior for applications requiring continual operation (because of the much stronger returned signal typical of angular scatter).

The bistatic work used only two horns, hence, measured only the component of the wind in the plane formed by the intersection of their beams. This concept can be extended to include three antennas, one pointing vertically and the other two positioned along orthogonal axes centered on the first and tilted so as to intersect a common volume along the beam of the vertically pointing antenna. By pulsing only the vertical system and receiving the return from the common volume at all three antennas, three components of the wind can be measured and converted into the total wind vector at that height.

A field experiment was conducted to test this procedure in mid 1971. The tethered kytoon BLP was provided and operated by the National Center for Atmospheric Research. Allowance has been made for the time lag in wind detection, since the BLP and Doppler volume were not collocated, in plotting the comparison in Doppler and BLP observed wind shown in (F18.14). The rather good agreement between the in situ and remotely sensed wind indicates that the Doppler technique, which is low pass filtered in (F18.14) with a time constant of one minute, is measuring the same mean values and fluctuations detected with the cup anemometer on the balloon. Complications are introduced by the balloon swaying back and forth with the wind, so that a more meaningful comparison of the two types of data, with a Doppler wind filter time constant of two minutes and representing a 30 minute data sample, is shown in (F18.15). The regression line slope of 0.95 indicates very good agreement between the two wind measurements. The finite y-intercept for the BLP measured wind speed is an effect of the cup anemometer instrumentation system,

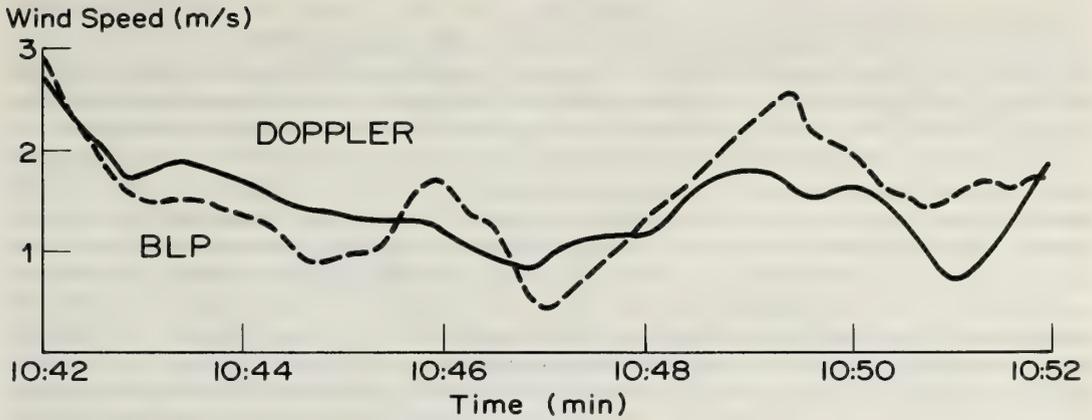


Figure 18.14 Comparison of wind measurements by acoustic Doppler and an anemometer on the boundary layer profiler (BLP) balloon, time lagged to compensate for differing positions, and low pass filtered using a one minute time constant (after Beran and Clifford, 1972).

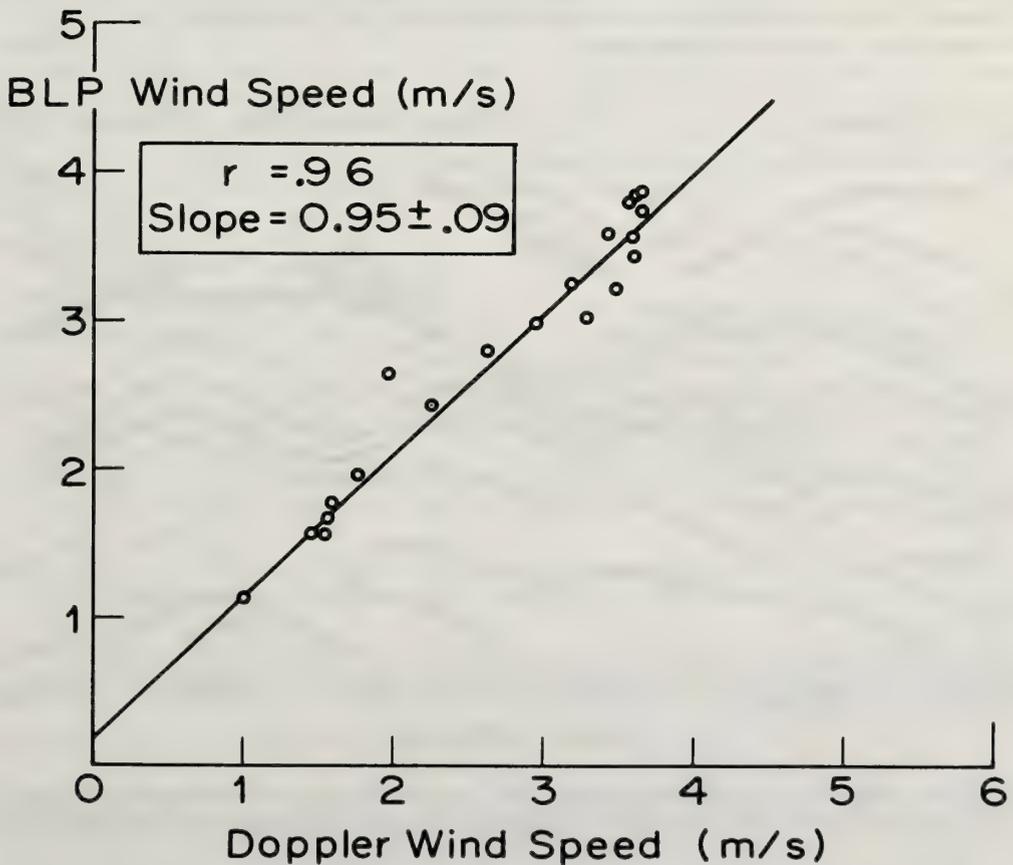


Figure 18.15 Scatter diagram comparing Doppler and BLP wind measurements, time lagged to compensate for differing positions, using a two minute constant and 30 minutes of data (after Beran and Clifford, 1972).

where no values smaller than 0.5 ms^{-1} are measured because of starting friction. It is estimated that, within the lower 500 m in the boundary layer, present acoustic Doppler techniques should provide an accuracy in total wind vector measurement of $\pm 0.5 \text{ ms}^{-1}$ for carrier frequencies near 2 kHz, pulse durations of 50 ms and taking the Doppler shift for each pulse transmitted. For 10 pulse averages, the resolution should be improved to $\pm 0.15 \text{ ms}^{-1}$, but at the sacrifice of some spatial resolution because of the averaging period.

So far the data presented for Doppler detected winds have been considered only for that volume where the three beams intercept. By assuming stationarity for the vertical velocity fields over 1 minute periods, an assumption shown to be quite valid by earlier vertical-only Doppler wind measurements, it is possible to derive the total wind vector at all heights where reasonably strong echoes are obtained. Under most nocturnal inversion conditions, and with the present acoustic equipment radiating 20 watts of acoustic power from each antenna, this height reaches at least 500 m. A time section of isotachs for horizontal winds as shown in (F18:16). A wind velocity maximum is shown in heights between 300 and 400 m. In the facsimile record obtained, this height corresponded to the regions slightly above the top of the inversion structure. It is known from previous studies that such nocturnal low level jets frequently develop above inversions, and it is believed this is the first successful remote detection of this phenomenon.

Time Section of Isotachs for Horizontal Wind (m/s) Under Oscillating Inversion. Winds were Derived Solely from Acoustic Doppler Measurements Using 3 separate Monostatic Systems, One Aimed Vertically The other two at 45 degrees.

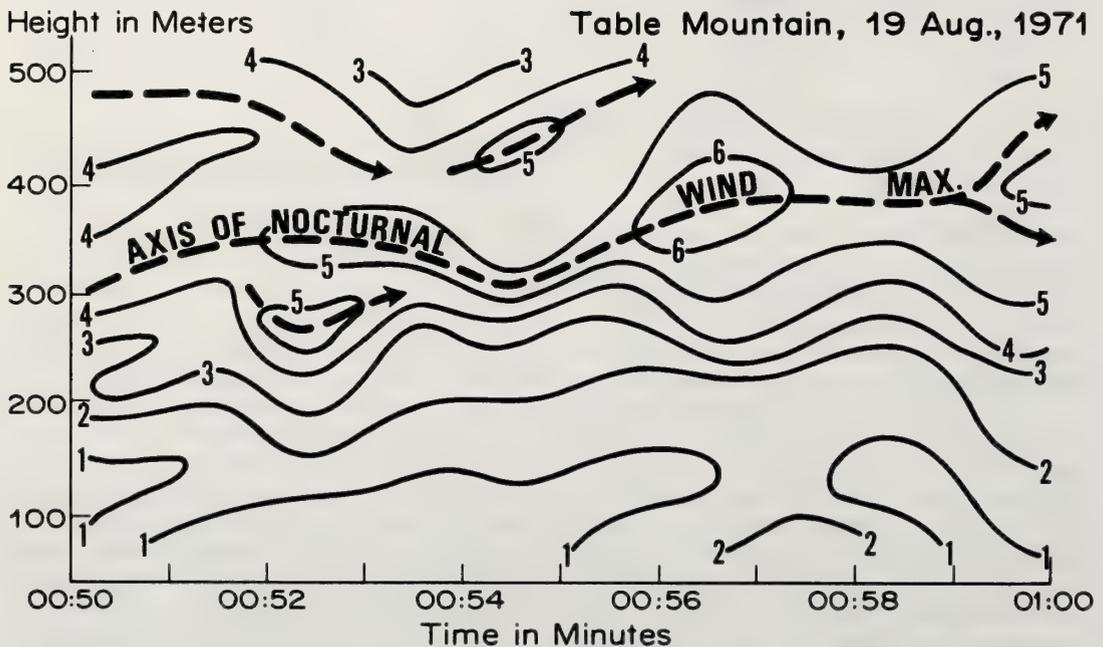


Figure 18.16 *Isotachs of the horizontal wind in ms^{-1} , measured by the acoustic Doppler technique; an oscillating temperature inversion extended to a height of about 250 m (after Beran and Clifford, 1972).*

18.6 Conclusions

Acoustic echo sounding has been found to provide a new and valuable tool for sensing temperature and velocity structure in the earth's boundary layer. So far the principal sensing technique has utilized the thermal structure signature of the atmosphere only, and even more exciting results may be expected in the future for velocity structure patterns which must be detected using a bistatic scattering geometry. Even the semiquantitative facsimile intensity displays provide new insight into the organization of convective plumes and the wave oscillations found in temperature inversions. The developing capability to measure quantitatively plume structure and vertical velocity is providing new insight into plume dynamics, especially relevant in understanding the fluxes of heat, momentum, and water vapor in the boundary layer, and the interaction of convective plumes with stable temperature inversions. The ability to sense the total wind vector in the boundary layer is of great importance in air pollution studies. Acoustic Doppler shows potential for measuring wind shear conditions, which may cause aircraft landing safety hazards if undetected. The simplicity and relatively low expense of fabricating acoustic echo sounder systems, as compared with radar or lidar devices, makes it probable that acoustic techniques will find even wider application in boundary layer studies in the future.

18.6.1 Acknowledgments

The program in Atmospheric Acoustic Echo Sounding reported here is a group effort at the NOAA Wave Propagation Laboratories. The program was initiated by WPL Director, C. G. Little, and his continued counseling and guidance are gratefully acknowledged. The equipment design and construction has been the responsibility of J. W. Wescott, B. C. Willmarth, and E. J. Owens and much of the data evaluation is due to W. D. Neff. The development of Doppler wind sensing techniques has been largely through the initiative of D. W. Beran who has also supervised much of the data evaluation. All of those listed have participated in vigorous brain storming sessions, experiment planning and data evaluation tasks, as well as contributing many long and odd hours in the performance of the field experiments. This work was supported both by NOAA research funds and by the Air Pollution Control Office, EPA.

18.7 References

- Baerg, W. and W. H. Schwarz (1966), Measurements of the scattering of sound from turbulence, *J. Acoust. Soc. Amer.* *39*, 1125-1132.
- Beran, D. W., R. M. Reynolds, and J. T. Gething (1970), Sound attenuation in the free atmosphere, Project EAR Report V, Meteorol. Dept., Univ. Melbourne, Australia.
- Beran, D. W., C. G. Little, and B. C. Willmarth (1971), Acoustic Doppler measurements of vertical velocities in the atmosphere, *Nature* *230*, 160-162.
- Beran, D. W. and B. C. Willmarth (1971), Doppler winds from a bistatic acoustic sounder, *Proc. 7th Int. Symp. Remote Sensing of Environment*, U. Mich. pp. 1699-1714.
- Beran, D. W. and S. F. Clifford (1972), Acoustic Doppler measurements of the total wind vector, to be publ., *Proc. 2nd Symp. Meteorol. Obs. and Insts.*, Amer. Meteorol. Soc.
- Beranek, L. L., 1954: *Acoustics*. McGraw-Hill Book Co., New York.
- Dneprovskaya, I. A., V. K. Iofe, and F. I. Levitas (1963), On the attenuation of sound as it propagates through the atmosphere, *Soviet Physics—Acoustics* *8*, 235-239.
- Gilman, G. W., H. B. Coxhead, and F. H. Willis (1946), Reflection of sound signals in the troposphere, *J. Acoust. Soc. Amer.* *18*, 274-283.

- Hall, F. F., Jr. J. W. Wescott and W. R. Simmons (1971), Acoustic Echo Sounding of Atmospheric Thermal and Wind Structure, Proc. 7th Int. Symp. Remote Sensing of Environment, U. Mich. 1715-1732.
- Hall, F. F., Jr., 1971: "Acoustic Remote Sensing Temperature and Velocity Structure in the Atmosphere," in *Statistical Methods and Instrumentation in Geophysics*, A. G. Kjelaas, ed. Teknologisk Forlag Oslo, 167-180.
- Harris, C. M. (1966), Absorption of sound in air versus humidity and temperature, *J. Acoust. Soc. Amer.* 40, 148-159.
- Henderson, M. C. and K. F. Herzfeld (1965), Effect of water vapor on the Napier frequency of oxygen and air, *J. Acoust. Soc. Amer.* 37, 986-988.
- Kaimal, J. C. and J. A. Businger (1970), Case studies of a convective plume and a dust devil, *J. Appl. Meteorol.* 9, 612-620.
- Kallistratova, M. A. (1959a), An experimental investigation into the scattering of sound in a turbulent atmosphere, *Dokl. Acad. Nauk SSSR* 125, 69-72.
- Kallistratova, M. A. (1959b), Procedure for investigating sound scattering in the atmosphere, *Soviet Physics—Acoustics* 5, 512-514.
- Kallistratova, M. A. and V. I. Tatarski (1960), Accounting for wind turbulence in the calculation of sound scattering in the atmosphere, *Soviet Physics—Acoustics* 6, 503-505.
- Kallistratova, M. A. (1961), Experimental investigation of sound wave scattering in the atmosphere, *Trudy Inst. Fiz. Atmos., Atmos. Turbulentnost*, No. 4, 203-256, USAF FTD translation TT-63-441.
- Kneser, H. O. (1933), The interpretation of the anomalous sound-absorption in air and oxygen in terms of molecular collisions, *J. Acoust. Soc. Amer.* 5, 122-126.
- Knudsen, V. O. (1931), The effect of humidity upon the absorption of sound in a room, and a determination of the coefficients of absorption of sound in air, *J. Acoust. Soc. Amer.* 3, 126-138.
- Knudsen, V. O. (1933), The absorption of sound in air, in oxygen, and in nitrogen-effects of humidity and temperature, *J. Acoust. Soc. Amer.* 5, 112-121.
- Little, C. G. (1969), Acoustic methods for the remote probing of the lower atmosphere, *Proc. IEEE* 57, 571-578.
- Margenau, H. and G. M. Murphy, 1956: *The Mathematics of Physics and Chemistry*, D. Van Nostrand Co., Princeton, N. J.
- McAllister, L. G. (1968), Acoustic sounding of the lower troposphere, *J. Atmos. Terr. Phys.*, 30, 1439-1440.
- Monin, A. S. (1962), Characteristics of the scattering of sound in a turbulent atmosphere, *Soviet Physics—Acoustics* 7, 370-373.
- Myrup, L. O. (1967), Temperature and vertical velocity fluctuations in strong convection, *Quart. J. Roy. Meteorol. Soc.* 93, 350-360.
- Rayleigh, J. W. S., 1896: *The Theory of Sound*, 2nd ed the Macmillan Co., London republished by Dover Publ. Inc., New York, 1945.

- Richtmyer, F. K. and E. H. Kennard, 1947: *Introduction to Modern Physics*, McGraw-Hill Book Co., New York.
- Sekera, Z. (1948), Helmholtz waves in a linear temperature field with vertical wind shear, *J. Meteorol.* 5, 93-102.
- Simmons, W. R., J. W. Wescott, and F. F. Hall, Jr. (1971), Acoustic echo sounding as related to air pollution in urban environments, Natl. Oceanic Atmos. Admin. Tech. Rept. ERL 216 WPL 17, Boulder, Colorado.
- Stokes, G. G. (1849), On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids, *Trans. Cambridge Phil. Soc.* 8, 287-319.
- Strand, O. N. (1971), Numerical study of the gain pattern of a shielded acoustic antenna, *J. Acoust. Soc. Amer.* 49, 1698-1703.
- Sutton, O. G., 1953: *Micrometeorology*, McGraw-Hill Book Co., New York.
- Tyndall, J. 1875: *Sound*, Appleton Co., London, 3rd ed, Chap. 7.
- Wescott, J. W., W. R. Simmons, and C. G. Little (1970), Acoustic echo sounding measurements of temperature and wind fluctuations, ESSA Tech. Memo. ERLTM-WPL 5, Boulder, Colo.
- Whitman, W. C. and G. Wolters 1967: Microclimatic gradients in mixed grass prairie, in *Ground Level Climatology*, ed by R. H. Shaw, Amer. Assn. Adv. Sci. Wash., D. C. 165-185.

List of Symbols

A	Area of antenna collecting aperture	\vec{V}	Wind component
c	Velocity of sound	w	Vertical wind speed
c_0	Velocity of sound at a reference condition	x	Position coordinate on x axis
C_v	Velocity structure parameter	β	Angle between total wind vector and Doppler measured component
C_T	Temperature or thermal structure parameter	γ	Ratio of specific heats of a gas at constant pressure to constant volume
G	Antenna gain factor	$\Delta\nu$	Difference frequency
I, I_0	Intensity of sound	Θ	Angle of scattering, measured from direction of incident sound propagation
k	Attenuation coefficient of sound	Θ_r	Elevation angle of received sound
k_c	Classical absorption attenuation coefficient	Θ_t	Elevation angle of transmitted sound
k_m	Molecular absorption attenuation coefficient	λ	Wavelength of sound
k_s	Scattering or excess absorption attenuation coefficient	λ_0	Transmitted sound wavelength
\vec{K}_0	Wave vector of transmitted wave	λ'	Spatial scale for evaluation of $\Phi(V)$, $\Phi(T)$
\vec{K}_s	Wave vector of scattered wave	ν	Frequency of sound
ℓ	Distance, sound path length	ν_0	Transmitted sound frequency
p	Pressure, atmospheric pressure	ν_s	Scattered sound frequency
P_r	Acoustic power received	Π	Pressure dependent variable = $\frac{1}{\gamma} \ell n p$
P_t	Acoustic power transmitted	Π'	Pressure dependent variable in a turbulent atmosphere
r	Distance increment; reflectivity	$\sigma, \sigma(\Theta)$	Scattering function, acoustic power scattered at an angle Θ , per unit incident flux in a unit solid angle
t	Time, acoustic pulse duration	τ	Sound wave pressure contribution to Π ; atmospheric transmittance
T	Absolute temperature	$\Phi(T)$	Spectral density of fluctuations in temperature
T'	Absolute temperature of disturbed atmosphere with mean temperature T	$\Phi(V)$	Spectral density of fluctuations in wind velocity
u	Wind speed	∇	Vector operator del
\vec{u}	Velocity of gas particles		
\vec{u}'	Velocity of gas particles in a turbulent atmosphere		
\vec{v}	Velocity of gas particles in response to sound wave pressure gradients		
v'	Doppler measured wind component speed		

Chapter 19 PROSPECTS FOR ACOUSTIC ECHO SOUNDING

C. Gordon Little
Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

This chapter looks to the future of acoustic echo sounding. The physics of the interaction of acoustic waves with a turbulent, inhomogeneous atmosphere is summarized and the atmospheric information conceptually available is identified. The present capabilities of acoustic echo sounders are indicated, and techniques for the expansion of these capabilities are outlined. The chapter closes with a brief discussion of some of the major atmospheric science problems likely to be studied by acoustic echo sounding methods.

19.1 The Physics of Acoustic Echo Sounding of the Atmosphere

The physics of the scatter of acoustic waves by the atmosphere is best described in terms of the fluctuation in the phase velocity of the sound wave due to localized fluctuations in temperature, humidity and velocity. Quantitatively, the phase velocity of stationary, dry air at normal temperatures is given by

$$C_d = 20.05 T^{1/2} \text{ ms}^{-1}, \quad (19:1)$$

where C_d is the phase velocity of sound in dry air and T is the absolute temperature of the air in degrees Kelvin.

If the air is moist, a small increase in the phase velocity occurs, and we have

$$C_m = C_d \left(1 + 0.14 \frac{e}{p}\right) \text{ ms}^{-1}, \quad (19:2)$$

where C_m is the phase velocity of sound in moist air for which the total pressure is p and e is the partial pressure of water vapor, in millibars.

If the air is moving with a velocity W at an angle θ relative to the direction of propagation of a plane acoustic wave, a stationary observer will record the phase velocity as the sum of the velocity of the wave through the air, plus the component of the air velocity in the direction of propagation. Thus

$$V_p = C_{d,m} + W \cos \theta \text{ ms}^{-1}. \quad (19:3)$$

At low intensities, the phase velocity is not a function of frequency in the audio range, and so the group and phase velocities are equal for sound waves travelling in the atmosphere.

Typically, the small scale fluctuations in temperature, humidity and velocity in the boundary layer are within about one order of magnitude of 0.1° Kelvin, 0.1 mb and 0.1 ms^{-1} respectively. This means that the acoustic refractive index is fluctuating as a function of time and space by (very roughly) 300 parts per million (ppm) due to turbulence, 170 ppm due to temperature eddies and 14 ppm due to humidity fluctuations. (The corresponding fluctuations in radio frequency refractive index of the boundary layer are very much less – typically about 0.1 ppm.) This fluctuating refractive index ensures that a plane wave is distorted by more than about 1 radian after propagating roughly 100 to 1000λ ; the acoustic energy is no longer confined to a single direction but, in fact, is scattered in all directions (other than through 90° – see 19:4).

The absorption of sound has been discussed in the preceding chapter (see 18.1 and F18.1). The theory of the scatter of sound leads to the equation

$$\sigma(\theta) = 0.055 \lambda^{-1/3} \cos^2 \theta \left[\frac{C_v^2}{C^2} \cos^2 \frac{\theta}{2} + 0.13 \frac{C_T^2}{T^2} \right] \left(\sin \frac{\theta}{2} \right)^{-11/3}, \quad (19:4)$$

(see 18:13) where $\sigma(\theta)$ is the scattered power per unit solid angle at a scatter angle θ , per unit volume, per unit incident flux from a region containing a Kolmogorov spectrum of homogeneous, isotropic turbulence and temperature fluctuations. Here C_v and C_T are the structure constants for velocity and temperature respectively, λ is the acoustic wavelength, C is the mean velocity of sound and T the mean absolute temperature of the scattering region.

19.2 Atmospheric Information Potentially Available by Acoustic Echo Sounding

The first attempt to identify in some detail the information conceptually inherent in acoustic echoes from the atmosphere was due to Little (1969). Pointing out that acoustic waves interact much more strongly with atmospheric irregularities than electromagnetic waves, he showed that good signal-to-noise ratios could be expected utilizing a small acoustic sounder. He concluded that the implications of (19:4) were such that acoustic echo sounders could be developed to measure, to heights of at least 1500 meters:

- 1) the vertical profile of wind speed and direction (by utilizing the Doppler technique),
- 2) the vertical profile of humidity (by means of a multi-wavelength system),
- 3) the location and intensity of temperature inversions (by using a monostatic system to study the echo power and frequency spectrum, as a function of elevation angle, range, and wavelength),
- 4) the three-dimensional spectrum of temperature inhomogeneity (by using a monostatic system to study the echo power as a function of wavelength, azimuth, and elevation),
- 5) the three-dimensional spectrum of mechanical turbulence (by using both monostatic and bistatic systems to study the echo power as a function of scatter angle, wavelength, and direction).

Typical height resolutions attainable would be of the order one half of the pulse length, i. e., about 10 meters if a 60-ms pulse is used. Spatial wave numbers could be explored over the range from $4\pi/\lambda_{\min}$ (about 400 m^{-1}) to about 10^{-2} m^{-1} in the case of mechanical turbulence, and 400 m^{-1} to 40 m^{-1} for the temperature inhomogeneities, the difference being due to the ability of the Doppler sounder to measure the radial velocities at different points along the beam simultaneously. Time resolutions for successive independent measurements to heights of 1500 meters would be about 10 seconds.

Since (conceptually at least) the u , v , and w components of the wind, and their fluctuations, can be measured as a function of height and time, it should also be possible to calculate the vertical flux of momentum, by computing the correlation between the fluctuations in the vertical and horizontal components of the wind. In addition, Gething and Jenssen (1971) have shown that it should be possible to measure both the temperature and the humidity profiles if accurate measurements of the scattering cross section can be made as a function of height and probing frequency.

Given the above amount of information on the state of the atmosphere, it is important to recognize that theoretical and/or empirical relations should permit derivation of other meteorologically significant parameters. Thus, the use of turbulence theory and measurements of $C_v(z)$, $C_T(z)$ and the horizontal wind (and hence its gradient) as a function of height should eventually permit derivation of the temperature profile, $T(z)$ and the Richardson number as a function of height. As shown by Wyngaard et al., (1971), the heat flux may be calculated from measurements of C_T ; in addition, the energy dissipation rate ϵ , (due to viscosity) may be calculated as a function of height from the relation

$$\epsilon = C_v^3 / a^{3/2} . \quad (19:5)$$

A recent analysis (Little, 1972) of the interaction of acoustic waves with hydrometeors indicates that acoustic echoes should be readily obtainable from rain, snow, fog, and cloud particles, though (because of increased absorption at the higher frequencies required because of the smallness of the targets) probably only out to ranges of several hundred meters.

The above discussion relates to the *conceptual* ability of acoustic echo sounding to measure parameters of the atmosphere such as $C_v(z)$; $C_T(z)$; $u(z)$, $v(z)$, $w(z)$; $\epsilon(z)$; $T(z)$; the spectrum of turbulence $\phi_v(\mathbf{k})$; the spectrum of temperature fluctuations $\phi_T(\mathbf{k})$; the momentum flux; the heat flux; and hydrometeor reflectivity, all as a function of time. In addition, the technique offers extremely important opportunities to monitor the structure and processes of the atmosphere, especially such dynamical processes as thermal plumes, gravity waves, and turbulence. Implicit in the interaction of acoustic waves with the atmosphere is the ability to measure, continuously and quantitatively, the internal structure and dynamical processes of the boundary layer of the atmosphere, on scales ranging from the maximum interrogation range (typically a kilometer or two) down to the acoustic half wavelength (typically about 15 cm).

19.3 Present Measurement Capabilities of Acoustic Echo Sounding

The present measurement capabilities of acoustic echo sounding have been discussed in the previous chapter. Acoustic echoes have typically been limited to the height range from ~ 30 meters to about 1000

meters, though some echoes have been obtained to perhaps 6 km height (Beran, et al., 1971c). Careful measurements of C_T have indicated that this parameter can be measured to within about 6 dB during unstable convective situations but that echoes during stable conditions often are stronger than indicated by in situ measurements of C_T . This latter result is to be expected since a partial reflection mechanism due to quasi-horizontal stratification of the temperature profiles could enhance the acoustic echoes above those predicted on the basis of measurements of the *fluctuating* component of temperature difference between two vertically-spaced fast-response thermometers.

In a series of sequential tests of the Doppler method of deriving atmospheric velocities using single monostatic, bistatic, tristatic and eventually 3 monostatic systems, Beran et al., (1971a) Beran and Willmarth, (1971b) and Beran and Clifford, (1972) have investigated the accuracy of the method. The current analyses indicate that, under fairly low wind speed conditions, velocities can be measured to better than 0.5 ms^{-1} up to heights of about 500 meters. It seems likely that this capability can be extended to about 1 km without major difficulty. As yet, no investigation has been made of the ability of the sounder to measure the spectrum of the turbulence either by studying the broadening of the echo spectrum due to relative motions of different eddies within the pulse volume, or by comparing the radial velocities of different pulse volumes as a function of spatial separation.

Alternative ways to study the velocity field include the refraction and spaced receiver methods (Derr and Little, 1970). Preliminary tests of the two techniques (McAllister, 1971; Wescott et al., 1970) indicate that both hold promise, but suggest to the author that the Doppler method is likely to be the most useful one.

Multi-frequency attempts to derive temperature and/or humidity profiles are under way in the U.S.A. and in Australia (Parry and Sanders, 1972; Gething and Jenssen, 1971), but as yet no information is available as to their success.

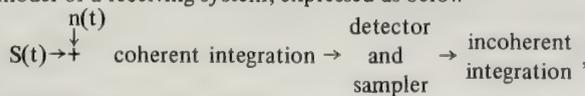
Although bistatic observations have been made by McAllister (private communication) and by Beran and Willmarth (1971b), no quantitative measurements of C_v profiles have yet been made. Similarly, there are no published accounts of quantitative studies of echoes from hydrometeors.

The most striking success of acoustic echo sounding to date has been the ability to monitor the internal structure of the atmospheric boundary layer. Thermal plumes, internal gravity waves and breaking gravity waves have all been identified and studied (e.g., McAllister, 1969; Beran et al., 1971d; Hall et al., 1971; Hooke, et al. 1972); the reader is referred to the figures of the previous chapter for examples of acoustic soundings taken under these diverse meteorological conditions.

19.4 Methods for Expansion of Current Acoustic Echo Sounding Capabilities

19.4.1 The Signal-to-Noise Equation

In any echo-sounding system, the expected ratio of output signal power to output noise power, usually termed the signal-to-noise ratio, is an important parameter identifying the inherent measurement sensitivity and capabilities of the system. Chadwick and Little (in publication) analyze the performance of echo-sounding systems in terms of a general model of a receiving system, expressed as below



where the received signal is $S(t)$ and the noise is $n(t)$. In acoustic echo sounding, the noise is usually due to random acoustical noise received by the antenna, and since the noise is relatively wideband Gaussian noise it can be assumed to have a constant spectral density over the passband of the receiver.

Their analysis for the output signal-to-noise ratio of a monostatic system yields

$$\text{SNR} = \frac{(M)^{1/2} T P A \eta \Delta L f}{4\pi R^2 N_0} \tag{19:6}$$

where M is the number of samples integrated incoherently, T is the coherent integration time, P is the transmitted peak power, A is the effective collective area of the antenna, η is the scattering cross section per unit volume, Δ is the range resolution cell, R is the range to the target, L is the round-trip attenuation factor due to absorption or scatter, f is the performance of the receiver relative to an idealized system with coherent integrator and matched filter, and N_0 is the (two-sided) noise spectral density. Prospects for improving the SNR and hence the measurement capabilities of standard pulsed acoustic echo sounders are indicated below in

terms of changes to the nine main system parameters listed in (19:6). (The range R of a given target is not a variable under the control of the researcher, and so is not discussed.)

19.4.2 Possible Steps for Increasing Acoustic Echo-Sounder Sensitivity

a. Radiated Power. Typical radiated powers are currently of the order 10 acoustic watts. By the use of arrays of transmitting transducers, increases of at least 20 dB in this parameter seem practicable, e.g., the array of 100 transducers used by Parry and Sanders (1972).

b. Coherent Integration Time. The coherent integration time T is typically of the order 0.03 sec (i.e., equivalent to the reciprocal of a bandwidth of about 30 Hz). Because of the random fluctuations of acoustic propagation, receiver bandwidths probably cannot be reduced to less than about 3 Hz, and T is therefore limited to about 0.3 sec, for a potential improvement in SNR of 10 dB. This would, of course, result in a 10-fold reduction in range resolution.

c. Incoherent Integration Time. Most acoustic echo sounders to date have displayed each pulse separately, i.e., no incoherent integration has been used. (The facsimile recorders used have, of course, tended to supply some visual integration.) By accepting the resulting degradation in time resolution, integration over (say) 100 pulses could be used; this would provide a 10 dB improvement in output SNR, though at one pulse per 10 seconds this would imply integrating the echoes for approximately 15 minutes.

d. Effective Antenna Area. The effective antenna area has typically been of the order of 3 m^2 , before introducing transducer losses. For special purposes it should be able to increase this by at least 10 dB; in the unique case of the University of Illinois cylindrical paraboloid antenna, a collecting area of the order 10^4 m^2 would be attained.

It is important to recognize that the full benefits of the increased antenna aperture may not be realized if too large a ratio of aperture diameter to wavelength is used. This results from three factors: (a) the near field distance of the antenna, $R_{\min} = 2D^2/\lambda$ may become comparable to the desired range of the system, (b) the half power beamwidth (of order λ/D) may become comparable to or smaller than the refraction of the acoustic beam produced by wind shear, with the result that the returned echo arrives from directions outside the main beam, and therefore is registered at greatly reduced sensitivity, (c) even in the absence of atmospheric refraction, the presence of atmospheric turbulence and temperature inhomogeneities will tend to cause the angular spectrum of the radiated signal to broaden as it propagates; the received echo therefore comes from a broader cone than the main beam of the antenna, and again will be recorded at reduced sensitivity. In general, these effects are not important for $D < 5\lambda$ but are likely to be of dominant importance for $D > 20\lambda$.

e. Scattering Cross Section. Most acoustic echo sounding has been done monostatically, with the echoes being produced solely by the small scale ($\lambda/2$) fluctuations in temperature. When operating bistatically, energy scattered by velocity fluctuations is also received; in most meteorological circumstances, it is expected that this echo will significantly exceed that from the temperature fluctuations (Beran and Willmarth, 1971b). The actual magnitude of this potential increase is, however, not yet known.

f. Range Resolution Cell. The depth of the atmosphere contributing to the echo at any instant is determined by the pulse length, which may be lengthened up to one half the pulse repetition period. Typical pulse lengths used to date are of the order 0.2 seconds; in a long range system, where low range resolution is presumably acceptable, pulse lengths of the order 2 seconds (equivalent to a 330 meter range integration) could be used, for a 10 dB increase in output SNR.

g. Atmospheric Attenuation. The atmospheric attenuation experienced along the two-way path to and from the region being interrogated increases with frequency and range. On the other hand, the antenna beamwidth, and the ambient noise level both tend to decrease with increasing frequency. For this reason, a compromise must be set; too low a frequency will give too high a noise level; too high a frequency will give too much absorption. Experience indicates that a 10 dB round-trip absorption is about optimum.

h. Matched Filter. Most systems operate at close to matched filter performance, and relatively little gain in performance can be obtained by such modifications.

i. Ambient Noise. The ambient noise level on most acoustic sounders is 10-40 dB above the theoretical limit of kT_aB (k = Boltzmann's constant, T_a = ambient temperature, B = receiver bandwidth). It is the author's opinion that it is important to improve the SNR first by reducing the noise level, before using "brute force" methods such as increased radiated power or antenna size, or before accepting the poorer range or time resolutions implicit in increased coherent or incoherent integration. By careful design of the antenna side lobes; by locating the antenna below the surface of the ground and therefore out of the wind; and by

using quiet field sites, it should be possible to make important gains in SNR, often amounting to 20 dB or more. In addition, it would probably be desirable to use noise-subtraction methods to reduce the effects of any background acoustic noise picked up on the antenna. This noise is usually broadband compared with the receiver bandwidth, and two adjacent passbands of equal width therefore receive essentially equal power. If one passband is now tuned to the wanted signal, it receives noise plus signal; the neighboring channel receives noise only, and by subtracting the outputs of the two channels it should be possible to greatly reduce the effects of variations in external noise level.

Table(19.1)lists the potential gains discussed above for each parameter in (19:6).

Table 19.1 Anticipated Gains in Signal-To-Noise Ratio to be Expected in the Next Several Years

Parameter	Potential Gain (dB)
$M^{1/2}$	10
T	10
P	20
A	10
η	+ ?
Δ	10
L	—
f	—
N_0	<u>20</u>
Total	<u>80+</u>

This suggests that very large ($> 10^8$) increases in overall system sensitivity are conceivably available by feasible improvements in each parameter, though some of these changes (specifically in M, T and Δ) degrade the range and/or time resolution. These improved sensitivities will probably be used in two ways – to improve the range resolution at fixed distance (by using short pulses and therefore wider receiver bandwidths), and to greatly extend the distance to which observations can routinely be made. As an example of the first direction, present sounders usually do not give quantitative data at heights below about 30 meters. A high resolution, low range system ($\Delta = 1$ meter, $R = 10 - 100$ m) could readily be built by going up in frequency to perhaps 10 kHz with 2 ms pulses radiated every 0.6 second.

Table (19.1) indicates that a more than 80 dB increase in SNR is potentially available, if all parameters could be optimized simultaneously. Since, for a given η and L, the SNR varies inversely as R^2 , this seems to imply that the range could be increased by a factor of 10^4 ! In fact, it will be difficult to increase the range by more than about a factor of ten, because the number of dB of attenuation *at a given frequency* is roughly proportional to range. (In a homogeneous isothermal atmosphere the attenuation would be accurately proportional to range, but changes in humidity, temperature and density with height tend to cause the absorption to increase somewhat more rapidly than linearly with height.) Thus, a 10 dB round-trip attenuation to 1 km would imply at least 100 dB for a range of 10 km. In order to avoid this totally excessive attenuation, a reduction in frequency by perhaps two octaves would be required, but (as indicated above) this increases the noise level. Analysis by Little (1969) indicates that the expected increase in noise level per Hz of receiver bandwidth is about 14 dB per octave decrease in frequency; since the scattering cross section is expected to decrease by 1 dB per octave even if C_T stays constant (see 19:4), the decrease in SNR implicit in a two-octave decrease in frequency is 30 dB. The total decrease in SNR due to changes in R, η and N_0 would therefore amount to 50 dB. This would have to be offset by improvements in radiated power, pulse length and collecting area, and by using increased coherent and incoherent integration, as well as antennas having improved noise rejection qualities. Using the above figures, application of 19:6 indicates that an echo sounder radiating 1000 acoustic watts at 250 Hz in 2-second long pulses from a 30 m² antenna should give sensitivities at 10 km range comparable to the performance at 1 km range of current boundary layer systems radiating 10 watt, 1000 Hz, 0.2 second pulses from a 3 m² collecting area. Improvements in acoustic antenna design to reduce the ambient noise level would be desirable to cover the fact that in general C_T decreases with height.

Several additional possibilities warrant consideration. One important limitation of current acoustic

echo sounders is the low information rate (typically, 1 echo per 10 seconds for a 1.5 km unambiguous range). For many research purposes this represents a serious limitation, and we may expect that various techniques such as FM-CW modulation, continuous wave bistatic systems, multi-beam arrays or even acoustic holography will be used to overcome the inherently limited, 1-dimensional, slow probing of the present sounders.

In considering these systems, it is important to recognize that the intervening atmosphere will introduce phase and amplitude distortions of both transmitted and scattered signals. These distortions will modify or distort the echo from the target region and will limit the potential SNR advantages of the FM-CW and pseudo-random modulation systems by limiting the maximum coherent integration time. In addition, the receiver bandwidth of current acoustic echo sounder is typically five orders of magnitude less, and the duty cycle (i.e., the fraction of time the transmitter is on) one order of magnitude greater, than for microwave radars, and so the very great increase in sensitivity obtained by using FM-CW modulation of microwave radars will not be obtained with acoustic signals; instead, gains of only 10 to 20 dB can be anticipated.

19.5 Future Atmospheric Studies

In this section, we will try to identify some of the directions of growth of acoustic echo sounding. We will first identify the atmospheric parameters which will be studied, and then comment on the atmospheric problems to which acoustic echo sounding may be expected to contribute.

19.5.1 The Velocity Field

Paramount in the use of echo sounding will be its application to measurements pertaining to the velocity field. As shown by Beran and Clifford (1972), a 3-station Doppler system can be used to derive the total wind vector, as a function of height. The next step would appear to be to measure the momentum flux, through studies of the correlation between the horizontal and the vertical components of velocity. The spectrum of turbulence in the boundary layer will probably be the next problem tackled, through measurement of the structure function of velocity determined by measuring the mean square velocity difference at two positions along the beam as a function of their spacing. Such measurements should give the spectrum of turbulence for spatial scales ranging from the pulse resolution length Δ to the maximum range of the Doppler system. In addition, two methods can be used to derive spectral information at smaller scales. The echo will show a Doppler broadening due to the random motions within the pulse volume; the magnitude of this broadening is a measure of the power at scales smaller than the pulse resolution length Δ . As indicated in 19.2, the intensity of turbulence at a spatial scale of the order $\lambda/2$ can be obtained by measuring the scattered power both monostatically and bistatically. The combination of these different methods provides an opportunity to measure the spectrum of turbulence over spatial scales from $\lambda/2$ to the maximum operating distance of the system.

This combination of measurements of profiles of wind speed, momentum flux, spectrum of turbulence and structure constant represents a very full expression of the velocity field. Nevertheless, the full 4-dimensional description of the velocity field does not seem likely to be available through acoustic echo sounding, primarily because of the relatively low information rate of acoustic echo sounding. However, by using multiple beam antenna arrays, it is likely that the velocity field will be explored in two dimensions within the next several years.

19.5.2 The Temperature Field

Present efforts to measure the temperature field by acoustic echo sounding have been largely limited to the measurement of C_T profiles, and to theoretical investigation of the possibility of measuring temperature profiles by multi-frequency measurements of the absorption of the waves as a function of height. Presumably this multi-frequency approach will be tested in the next few years, though the author is sceptical that measurements of adequate accuracy (say $\pm 1^\circ \text{K}$) can be made by this technique. An alternative hope is that the single-frequency monostatic acoustic echo sounding records, which are, of course, a product of the atmospheric temperature and velocity profiles, will eventually be fully interpretable. Since the velocity profile is measurable by Doppler techniques, the possibility exists that it may be practicable to derive the temperature profile from the known C_T and \mathbf{V} profiles (and, if necessary, the C_V profiles).

At the moment, it does not appear that the heat flux can be measured directly by an acoustical

technique. Wyngaard et al. (1971) have shown that C_T and heat flux are related, and that measurements of C_T can, therefore, be used as measurements of heat flux.

The spatial spectrum of temperature inhomogeneity can be explored in three dimensions over a rather limited range of wavenumbers by the use of a steerable, multi-wavelength echo sounder. This technique will probably be useful in identifying the sharpness of the quasi-horizontal temperature stratifications produced under many meteorological conditions.

19.5.3 The Humidity Field

Although water vapor does slightly affect the phase velocity of acoustic waves (see 19:2), the contribution of water vapor fluctuations to atmospheric acoustic echoes is believed small. The main effect of the water vapor is upon the *absorption* of the acoustic energy, i.e., upon the imaginary term in the refractive index, rather than upon the real term (see 18.1). Little (1969) showed that, since the absorption is a known function of frequency and humidity, it should be possible to derive information on the humidity profiles from accurate measurements of the echo strengths on two or three frequencies, as a function of height. Later, Gething and Jessen (1971) extended this work to show that the use of three or four frequencies should permit derivation of both the temperature and the humidity profiles. This multi-frequency approach requires accurate measurements of the changing ratio of echo strengths on the different frequencies as a function of height, and makes a number of critical assumptions as to the propagation conditions and the nature and constancy of the scattering process. The author, therefore, remains doubtful that these techniques will in fact prove useful in practice.

19.5.4 Hydrometeors

The scatter of acoustic waves by hydrometeors has been analyzed by Little (1972) who shows that hydrometeor echoes would be expected out to 300 meters range (using quite modest echo-sounder parameters) from snow, rain, and drizzle, and also from most fogs and clouds. Of particular importance is the expected angle-dependence of the scatter. For a scatter angle of 90° , (19:4) predicts no scattered power from the temperature or velocity field; as shown in (F19.1) the corresponding null for hydrometeors is at a scatter angle of 48.2° . Operation with a narrow beam bistatic system at a scatter angle of 90° should, therefore, give hydrometeor echoes free from contamination by echoes from temperature or velocity fluctuations.

So far as the author is aware, no measurements have yet been published on acoustic echo-sounding of hydrometeors. Such measurements presumably will be made during the next few years, though, as pointed out in the referenced paper, acoustic echo-sounding of hydrometeors appears to offer no important advantages over its microwave counterpart, (except, perhaps, in cost).

19.5.5 Some Meteorological Problems Amenable to Study by Acoustic Echo Sounding

As emphasized earlier, acoustic echo sounding offers a unique ability to continuously monitor the internal structure and processes of the atmosphere, especially the boundary layer. Much has already been deduced from qualitative scanning of the facsimile recordings (e.g., the identification of thermal plumes, internal gravity waves, breaking waves, etc.) but the real payoff to atmospheric science will come as the quantitative capabilities of echo-sounding are applied broadly to meteorological problems of the boundary layer and troposphere. Some areas of possible application are listed below:

a. **Thermal Plumes.** Here the ability of a small array of portable sounders to monitor profiles of C_T and vertical and horizontal wind should offer totally new insights into the details of 3-dimensional convection as a function of meteorological conditions and the underlying terrain.

b. **Temperature Inversions.** Already the sensitivity of acoustic echo-sounders is such that the internal structure of inversions can be continuously monitored. One of the most striking features of acoustic echo sounding under stable conditions is its ability to detect internal gravity waves (e.g., Hooke et al., 1972). These are made manifest by oscillations in height of quasi-horizontal echoing layers, or as oscillations in velocity. We may expect that spaced echo-sounders, and especially spaced Doppler echo-sounders will be used in studies of the role of gravity waves (of wavelength say 1 km to 10 km) in momentum and energy exchange in the lower atmosphere. Studies at high spatial resolution of the "herringbone" echo patterns seen in ground-based radiation inversions, are likely to provide new insights into the internal structure and processes of radiation inversions.

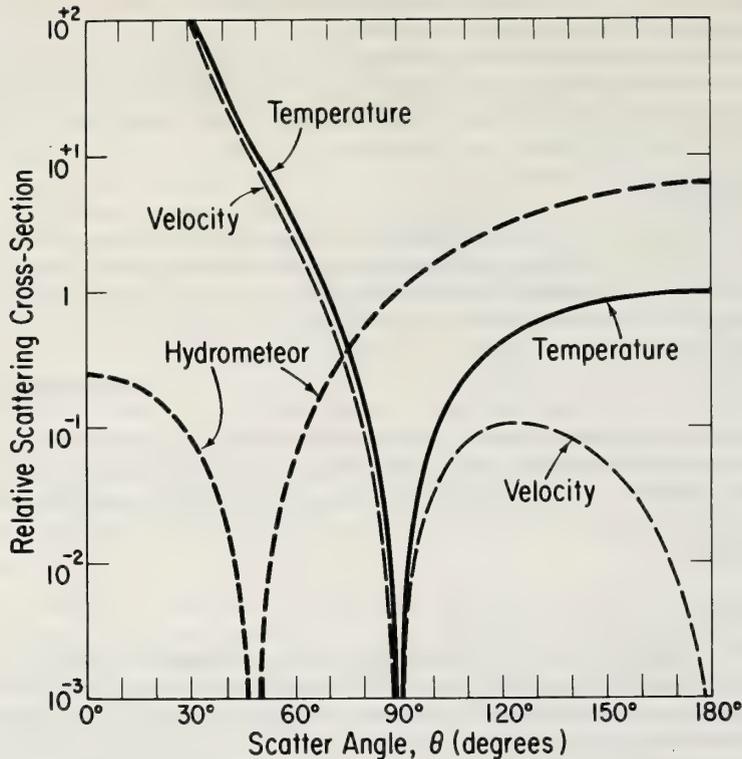


Figure 19.1 *Angle dependence of acoustic scatter from hydrometeors and a Kolmogorov spectrum of temperature and velocity fluctuations.*

c. **The Relationship Between Gravity Waves and Turbulence.** One of the unexplained features of past boundary layer studies has been the “intermittency” (in space and time) of fluctuations in parameters such as temperature. Acoustic echo-sounders are already enabling the researcher to identify coherent internal wave-like motions in the atmosphere, and on occasion to watch their breakdown into localized patches of small-scale turbulence. Further studies of “intermittency” and the relationships between gravity waves and turbulence are likely to prove rewarding.

d. **The Marine Atmosphere.** Although C_T may be expected to be much smaller in the atmospheric boundary layer over the ocean than over land, we may expect that acoustic echo sounding will provide the ability to monitor the changing structure of the lower atmosphere over the ocean, and to throw major light on any differences between the properties and processes of the boundary layer of the atmosphere over land and water.

e. **Urban Meteorology.** Several acoustic echo-sounding groups, including those at the University of Melbourne, the University of Toronto and the Wave Propagation Laboratory have shown that acoustic echo-sounding records can readily be obtained in an urban environment. This suggests that the various capabilities of the technique should be immediately applied to the problems of urban meteorology. The field is sufficiently far advanced that already one can foresee a large role for acoustic echo sounding in monitoring the structure of radiation inversions, and the associated wind profiles, for use in air pollution studies and operations. For example, widely spaced Doppler wind sensing systems could be used to monitor convergence of air and transport of pollutants in an urban area.

f. **Airport Operations.** Acoustic Doppler techniques would seem to be immediately applicable to two important airport problems – low level wind shear, and wake turbulence. In addition, it is possible that acoustic echo sounders may find a role in monitoring fog conditions at airports, including fog dispersal operations.

The above represent some of the main applications of acoustic echo-sounding in the immediate future. Many other potential applications no doubt exist, including land/sea breeze studies; studies of wind profiles, convergence and divergence in the clear air around convective storms; valley and mountain winds; subsidence inversions; and clear air turbulence investigations, to name only a few.

In summary, we conclude that the rapid progress in acoustic echo-sounding of the past few years will continue for the next several years, as echo-sounding techniques (both quantitative and qualitative) are applied to an expanding range of meteorological problems in the boundary layer and troposphere.

19.6 References

- Beran, D. W., C. G. Little and B. C. Willmarth (1971a), Acoustic Doppler measurements of vertical velocities in the atmosphere, *Nature*, 230, 160-162.
- Beran, D. W. and B. C. Willmarth (1971b) Doppler winds from a bistatic acoustic sounder, Proc. 7th Int'l. Symp. on Remote Sensing of Environment, May 17-21, University of Michigan.
- Beran, D. W., L. G. McAllister and A. R. Mahoney (1971c), Extension of acoustic sounding experiments to the upper atmosphere, WRE Tech Note A208 (AP) Weapons Research Establishment, Department of Supply, Salisbury, South Australia.
- Beran, D. W., F. F. Hall, J. W. Wescott and W. E. Neff (1971d), Application of an acoustic echo sounder to air pollution monitoring, Air Pollution Turbulence and Diffusion Symposium, December 7-10, New Mexico State University, Las Cruces, New Mexico.
- Beran, D. W. and S. F. Clifford (1972), Acoustic Doppler measurements of the total wind vector, A.M.S. Second Symp. on Meteorological Observations and Instrumentation, March 27-30, San Diego, California.
- Chadwick, R. B. and C. G. Little (1972), The comparison of sensitivities of atmospheric echo sounders, (in publication).
- Derr, V. E. and C. G. Little (1970), A comparison of remote sensing of the clear atmosphere by optical, radio and acoustic radar techniques, *Appl. Opt.* 9, 1976-1992.
- Gething, J. T. and D. Jenssen (1971), Measurements of temperature and humidity by acoustic echo sounding, *Nature* 231, 198-200.
- Hall, F. F., Jr., J. W. Wescott and W. R. Simmons (1971), Acoustic echo soundings of atmospheric thermal and wind structure, Proc. 7th Int'l. Symp. on Remote Sensing of Environment, May 17-21, University of Michigan.
- Hooke, W. H., J. M. Young and D. W. Beran (1972), Atmospheric waves observed in the planetary boundary layer using an acoustic sounder and a micro-barograph array, *Boundary Layer Meteorology* (in press).
- Little, C. G. (1969), Acoustic methods for the remote probing of the lower atmosphere, *Proc. IEEE*, 57, No. 4, 571-578.
- Little, C. G. (1972), On the detectability of fog, cloud, rain and snow by acoustic echo-sounding methods, *J. Atmos. Sci.* (in press).
- McAllister, L. G., J. R. Pollard, A. R. Mahoney and P. J. R. Shaw (1969), Acoustic sounding – a new approach to the study of atmospheric structure, *Proc. IEEE*, 57, 579-587.

- McAllister, L. G. (1971), Wind velocity measurements in the lower atmosphere using acoustic sounding techniques, WRE Tech Note A204 (AP) Weapons Research Establishment, Department of Supply, Salisbury, South Australia.
- Parry, H. D. and M. J. Sanders (1972), The design and operation of an acoustic radar, IEEE Geoscience Trans. (in press).
- Wescott, J. W., W. R. Simmons and C. G. Little (1970), Acoustic echo-sounding measurements of temperature and wind fluctuations, ESSA Tech. Memo. ERLTM-WPL5.
- Wyngaard, J. C., Y. Izumi and S. A. Collins, Jr. (1971), Behaviour of the refractive-index-structure parameter near the ground, J. Opt. Soc. Am., 61, 1646-1650.

List of Symbols

A	effective collective area of the antenna	p	total pressure, in millibars
B	receiver bandwidth	P	transmitted peak power
C	mean velocity of sound	R_{\min}	near field distance of the antenna
C_d	phase velocity of sound in dry air	R	range to target
C_m	phase velocity of sound in moist air	S(t)	received signal
C_T	structure constant for temperature	T	absolute temperature in degrees Kelvin
C_v	structure constant for velocity	T	coherent integration time
e	partial pressure of water vapor, in millibars	T_o	ambient temperature
f	performance of receiver relative to matched filter system	V_p	phase velocity
k	Boltzmann's constant	W	velocity of air
L	round-trip attenuation factor due to absorption or scatter	$\sigma(\theta)$	scattered power per unit solid angle
M	number of samples integrated incoherently	θ	scatter angle
N_o	(two-sided) noise spectral density	λ	acoustic wavelength
n(t)	effective noise level of system	ϵ	energy dissipation rate (due to viscosity)
		η	scattering cross section per unit volume
		Δ	range resolution cell (pulse resolution length)

Chapter 20 APPLICATION OF FM-CW RADAR AND ACOUSTIC ECHOSOUNDER TECHNIQUES TO BOUNDARY LAYER AND CAT STUDIES

B. R. Bean

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Recent results are given comparing acoustic and FM-CW radar sounders with in-situ measurements of atmospheric structure. The agreement of temperature and wind structure with the sounder records suggests that the two types of measurements are highly complementary, the tower data aiding in understanding the atmospheric processes traced by the sounder and the sounder aiding in selection of periods of stationarity suitable for detailed analysis of the tower data. The same conclusion can be reached for studies of exchange of energy between the mean and eddy flow. Examination of structures that appear similar to those of instability waves produced by wind shear at the stable interface of a temperature inversion indicate that such waves "break" when the Richardson number attains the value of 0.25. The wavelength associated with such "breaking" waves conforms to the theoretically predicted range of values.

20.0 Introduction

Classical meteorological research on fine-scale atmospheric structure has concentrated on the profiles of temperature and wind near the ground, say, in the first 10-20 meters. This work has also included detailed studies of the heat carried by atmospheric eddies and the partition of energy between the mean and turbulent flow. By and large, such studies have relied upon point or in-situ measurements. Such measurements are, in fact, the basis for our knowledge of the dynamics of the atmosphere's boundary layer. We are now in a time when new methods of viewing and recording properties of the atmosphere from a distance are being intensely studied. Examples are laser and radio determination of mean temperature and humidity over distances of a few tens of miles (Bean and Warner, 1967), high-powered radar interrogation of hundreds of square miles for convective laminar structure (Ottersten, 1969), and numerous other techniques (see, for example, the special issue of the Proceedings of the IEEE, April 1969).

This chapter attempts to complement the above studies by discussing experience gained in comparing classical in-situ tower measurements with those of simultaneous acoustic and EM clear-air radars. To do so, the discussion will be primarily restricted to the results-to-date of two experiments carried out on the high plains of southeastern Colorado, one involving the FM-CW (EM) radar, the other an acoustic echo-sounder.

The reason for this restriction is quite obvious when surveying the literature of radar observation of atmospheric instability events. One notes, for example, in chapter 14 the great effort that has gone into obtaining simultaneous spatial and temporal radar and meteorological observations, generally to the frustration of the best efforts of the observer. The discussion of this chapter is mostly concerned with detailed meteorological and radar measurements taken over a 500-foot height interval with the radar and meteorological tower separated by 250 m. The meteorological measurements are such as to allow study of the time history of mechanical (wind) and buoyant (heat) energy. The cases reported here are those for which the wind is *normal* to a line connecting the meteorological tower and the radar.

Thus, the reason for this seemingly arbitrary restriction of attention to the author's own work is for the sake of quantitative illustration of atmospheric instability phenomena. To the best of our knowledge, these experiments uniquely combine the remotely sensed layers, waves, thermal plumes, etc., and the sophisticated meteorological measurements necessary to physically or quantitatively describe such events. Having done so one may then have more confidence in interpreting the meteorological conditions around similar appearing phenomena sensed only remotely.

20.1 The FM-CW Radar

The introduction of the 10 cm FM-CW radar technique by Richter (1969) for studies of clear-air fluctuations of refractivity has rapidly led to studies of elevated inversions (Gossard and Richter, 1970), and detailed discussion by Atlas et al., (1970), on the possibility that breaking of these waves is the origin of clear-air turbulence (CAT). The details of Richter's technique are given in Appendix I, a reprint of his basic paper.

Richter's technique was adapted to probe the clear air over the high plains of eastern Colorado. The vertically pointing radar used here differs insignificantly, for the present study, from that of Richter. The experiment was performed near a 150 m micrometeorological tower at Haswell, Colorado. The site is sparsely covered with 15 cm high clumps of buffalo grass for a minimum of 3 km in any direction from the tower. The site is at an elevation of 1307 m above mean sea level and displays a characteristic high plains meteorological regime with strong solar heating during the day and radiative cooling at night. The radar was placed 240 m from the base of the tower in the direction of the prevailing wind. The 150 m tower is instrumented to measure fluctuations in wind, temperature, and radio refractive index utilizing a microwave refractometer at several fixed levels as well as a similarly equipped carriage that travels the height of the tower in ten minutes. The carriage was also equipped to measure the difference in radio refractive index over a 20 cm vertical spacing. Thus, the carriage allowed the placing of the meteorological probes within the region of radar returns. Further, the presence of the tower at 240 m away from the radar results in a uniform return which is confined between 255 and 295 m (see F20.1) on the radar records and serves not only as a reference but also allows one to track the motion of the carriage within the tower echo from approximately 65 m to the top of the tower. A reference line is placed on the radar returns at 150 m to indicate the tower height.

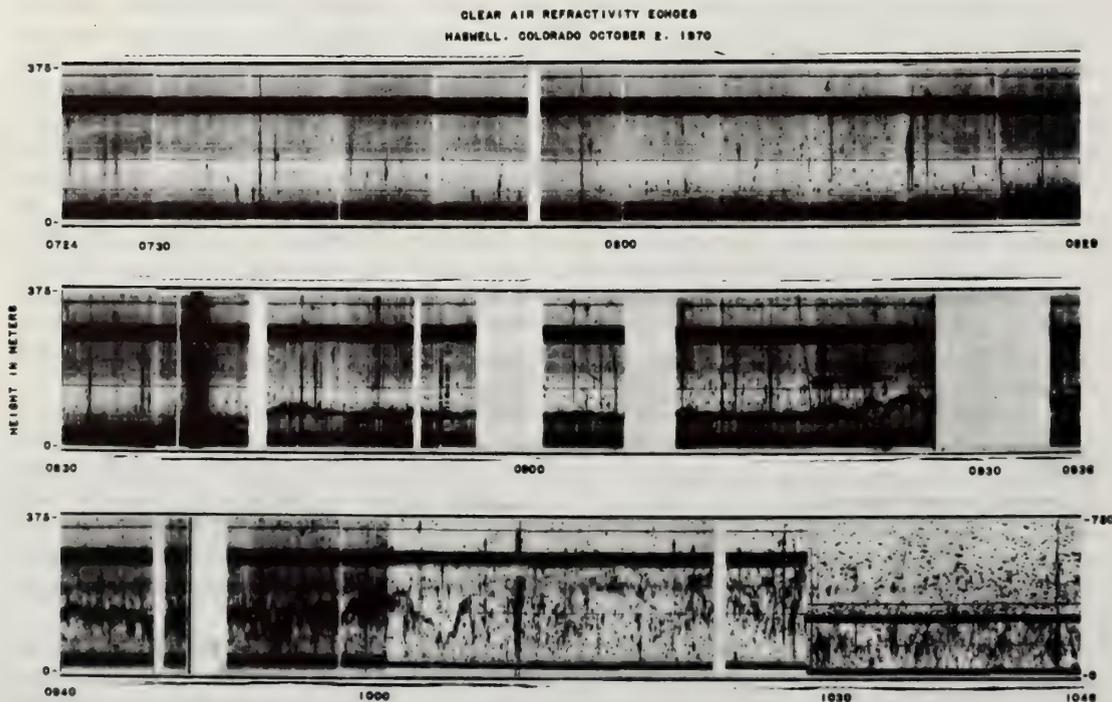


Figure 20.1 Radar record of clear air returns of October 2. The $(range)^{-2}$ variation of radar return for volume scattering has been removed electronically. The record was obtained from photographs of an intensity-modulated oscilloscope face.

20.2 Comparison of Tower and Radar Returns

The data from the morning of October 2, 1970, was utilized to compare radar reflectivities from the tower and radar. As can be seen on (F20.1), the region of strong radar returns lifted slowly and nearly uniformly from 0800 through approximately 0930 corresponding to the growth of a well-mixed layer below the inversion. The dots that appear throughout the record are attributed to insects. The blanks in the record correspond to times when the reflectivity of the atmospheric returns were monitored. The tower data were converted to radar reflectivity by assuming that the scales of interest corresponding to the radar wavelength (10 cm) lie within the “-5/3” region of the turbulence spectrum. The refractive index difference over 20 cm was measured so that when the structure function is given by (Ottersten, 1969)

$$\Delta n_{20}^2 = C_n^2 (0.2)^{2/3} \quad , \quad (20:1)$$

one may obtain the radar reflectivity in (m⁻¹) by substitution into

$$\eta = 0.38\lambda^{-1/3} C_n^2 \quad , \quad (20:2)$$

or

$$\eta = 2.4 \Delta n_{20}^2 \quad . \quad (20:2a)$$

Examination of both sets of data where the stationary meteorological carriage lies at a height equal to the layer examined yields:

<u>Time</u>	<u>$\eta(m^{-1})$ radar</u>	<u>$\eta(m^{-1})$ tower</u>
0844	1×10^{-5}	1.1×10^{-5}
0907	3×10^{-3}	0.7×10^{-3}
0930	1×10^{-3}	0.7×10^{-3}
0948	1×10^{-4}	2.8×10^{-4}

The agreement of the reflectivities measured by the radar and computed from the tower data is excellent considering the several unknowns involved in this kind of comparison.

The major unknown in the numerical comparison is the effect of non-stationary conditions across the 240-meter separation of the radar and tower sites. This separation was chosen to keep the tower echo from appearing within the height interval equal to the tower above the radar. Therefore, it was impossible to obtain colocated measurements of radar reflectivity, refractivity, temperature, and wind. Every effort was made during data reduction to obtain either time or height coincidence between the two sets of data, depending upon whether or not the tower carriage was stationary.

The vertically-pointed radar antennas were set into separate earth excavations two meters deep and no subsequent measurements were made of the resultant antenna patterns. This type of antenna configuration is required to keep the leakage power from the transmitter below the saturation level of the receiver. A recent study of similarly configured acoustic antennas (Strand, 1971) has shown that this shielding technique may be utilized to reduce the side lobes of the antennas in the 30° to 90° zenith angle region with little noticeable effect upon the main lobe pattern. Hence, it was assumed that the main lobe pattern of the sunken antennas was substantially unchanged from the free-space pattern.

The system performance of the radar was continuously monitored during the course of reflectivity measurements. The transmitted power was measured with a thermistor element and did not vary by more than ±0.5 dB. Frequent calibration of the power meter indicated most of this ±0.5 dB variation is the effect of ambient temperature changes upon the thermistor element, implying the transmitted power is sensibly constant. In addition, some of the transmitted power was continuously leaked into the receiver through a 2.5 μs delay line providing a continuous system echo at 375 meters range. Thus, the total electronic system is essentially self-monitored and self-calibrated with reference to a single power meter.

The correspondence between the profiles of the meteorological parameters and the radar returns were examined in detail, particularly the period 0900-0907 MDT. The radar return shown on the side of (F20.2) displays two peaks at 60 and 75 meters which closely correspond to the occurrence of inversions in the temperature profile. The return at 60 meters is not reflected in the wind profile while a wind maximum is observed above the 75 m inversion.

The stability of an atmospheric layer is commonly described by the Richardson number, R_i , which is defined as the ratio of the buoyant damping force to the shear production of energy, i.e.,

$$R_i = \frac{g}{T} \frac{\left(\frac{\partial \bar{T}}{\partial z} + \Gamma \right)}{\left(\frac{\partial \bar{u}}{\partial z} \right)^2}$$

where T is the temperature, g the acceleration due to gravity, Γ the adiabatic lapse rate of temperature, and u the horizontal wind. The infinite values of R_i correspond to regions of zero wind shear. Although there has been much speculation that regions of large ΔN and strong radar returns correspond to regions where $R_i = 1/4$ (Atlas et al., 1970), our data do not show this. Note, however, that the radar return region is one of large changes of R_i with no obvious critical value of R_i . Figure(20.3) compares the profiles of N , ΔN_{20}^2 with the radar returns.

The radio refractive index is given by (Bean and Dutton, 1968)

$$N = \frac{77.6}{T} \left(P + \frac{4810 e}{T} \right) ,$$

where P is the total pressure, e is the partial pressure of water vapor, and T the temperature. Pressure effects are normally negligible, and for all practical purposes changes in refractivity are wholly due to changes in temperature and water vapor content of the air. Consequently, we may take $\Delta N = -1.08\Delta T + 4.75\Delta e$.

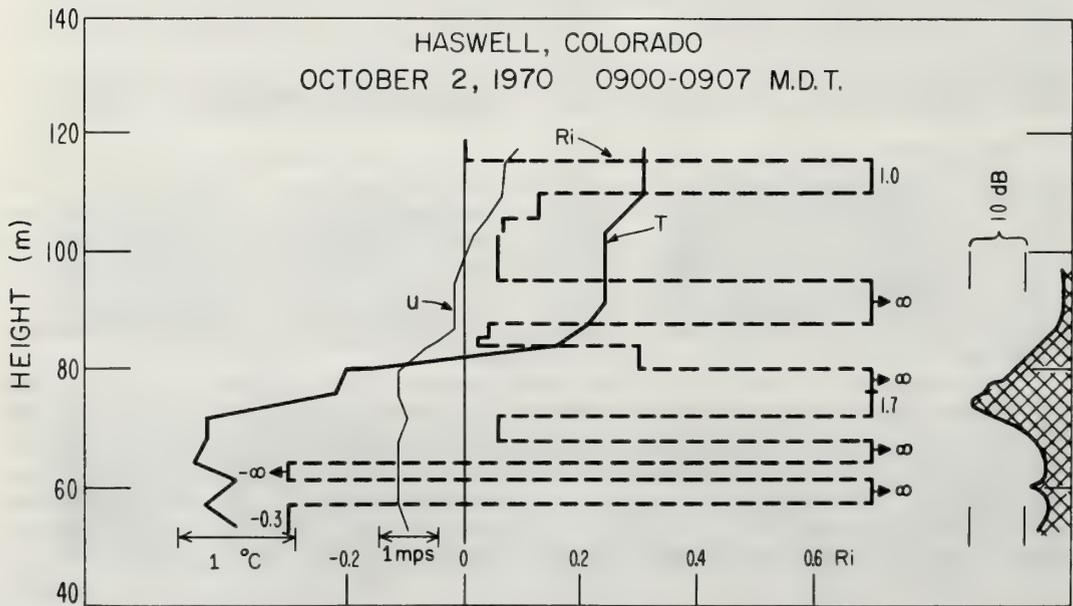


Figure 20.2 Comparison of temperature and horizontal wind speed profiles taken 0900-0907, the resulting Richardson number and an average of radar return taken at 0907.

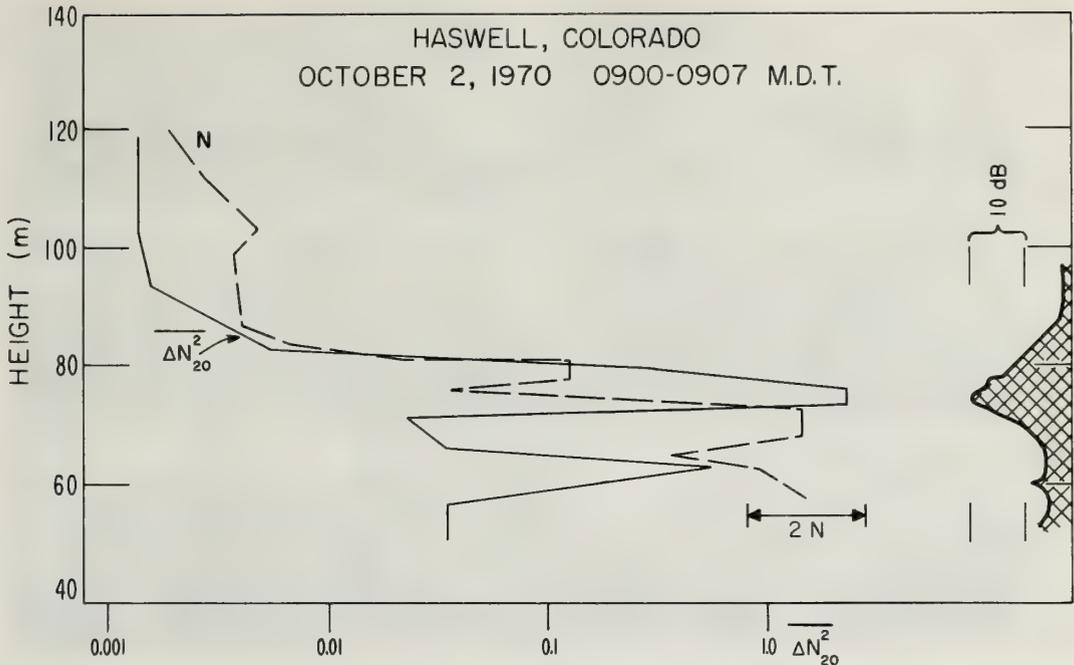


Figure 20.3 Comparison of refractivity and refractivity fluctuation profiles taken 0900-0907 and a 3.2 second average of radar return (represented by the crosshatched curve) taken at 0907. The refractivity, N , is defined by $N = (n-1) \times 10^6$, where n is the refractive index.

Figure (20.3) compares the profile of refractivity, N , the refractivity change (squared and averaged) over a 20 cm spacing, $\overline{\Delta N_{20}^2}$, with the radar returns. One notes maxima in $\overline{\Delta N_{20}^2}$ corresponding to those in the radar returns as well as the edges of the temperature inversions. The difference in the observed maxima is 11 dB while that calculated from the ΔN data is 8 dB, i.e., within a factor of two. This agreement is taken to be good considering the time and space separation of the measurements.

For the case under consideration, the N profile is mostly controlled by the T profile, especially across the main inversion at 80 m. From (F20.2) and (F20.3), we take $\Delta T = 3.25^\circ\text{C}$ and $\Delta N = -4.75$, then from above

$$\Delta e = \frac{\Delta N + 1.08\Delta T}{4.75}$$

or

$$\Delta e = -0.26 \text{ mb.}$$

This represents 26% of the total change in N across the inversion indicating the dominant role of temperature in this case. Normally fluctuations of humidity will dominate those of N , which is why the microwave refractometer is often used as a high speed humidity probe.

20.3 Other Examples of Radar Returns (Insects, Convective Elements, Inversion and Waves)

A large number of insects were observed in the radar returns during this experiment. This was concluded from the presence of feeding gulls during these periods and from the reports of pilots who had difficulty in keeping their windshields clear of insects while flying past the tower. The insects become quite visible to an observer standing in the shadow of an object which just shields the sun's disk. At 10 cm wavelength a radar is quite sensitive to insects. For example, the system m.d.b. (minimum-discernable-bug) is 0.2mm in diameter at a range of one kilometer. Insects were observed (F20.4) to a range of 1500 m during the entire period of 1227 to 1517 MDT on October 2, 1970. By this time the morning inversion is

Application of FM-CW Radar

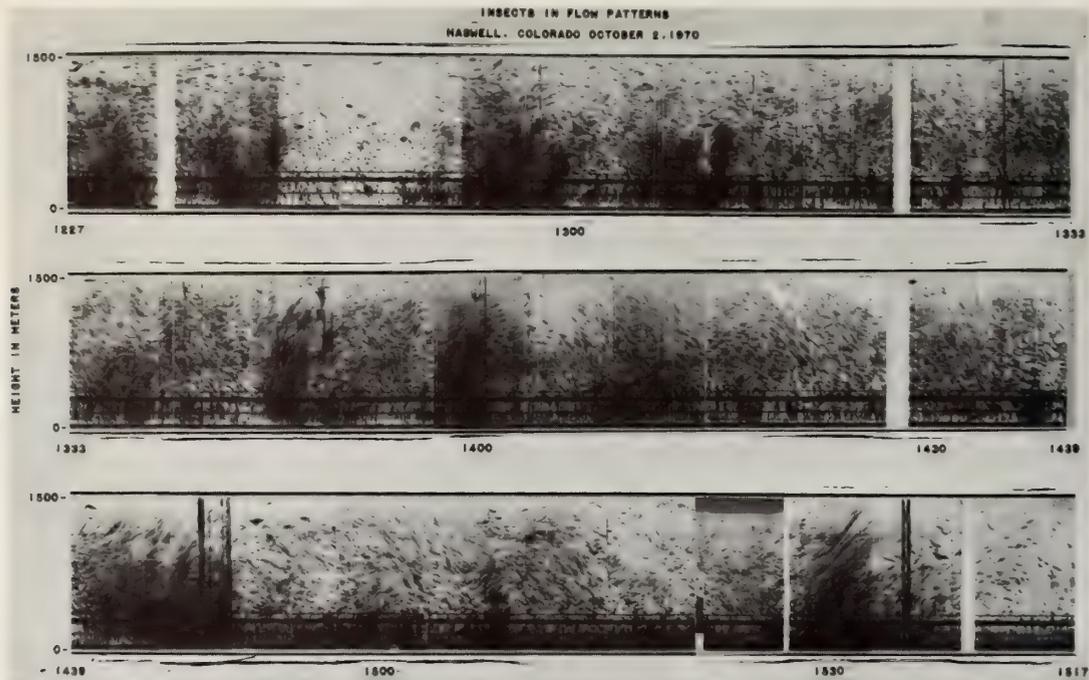


Figure 20.4 Radar record of insect returns of October 2.



Figure 20.5 Radar record of clear air and insect returns of October 12.

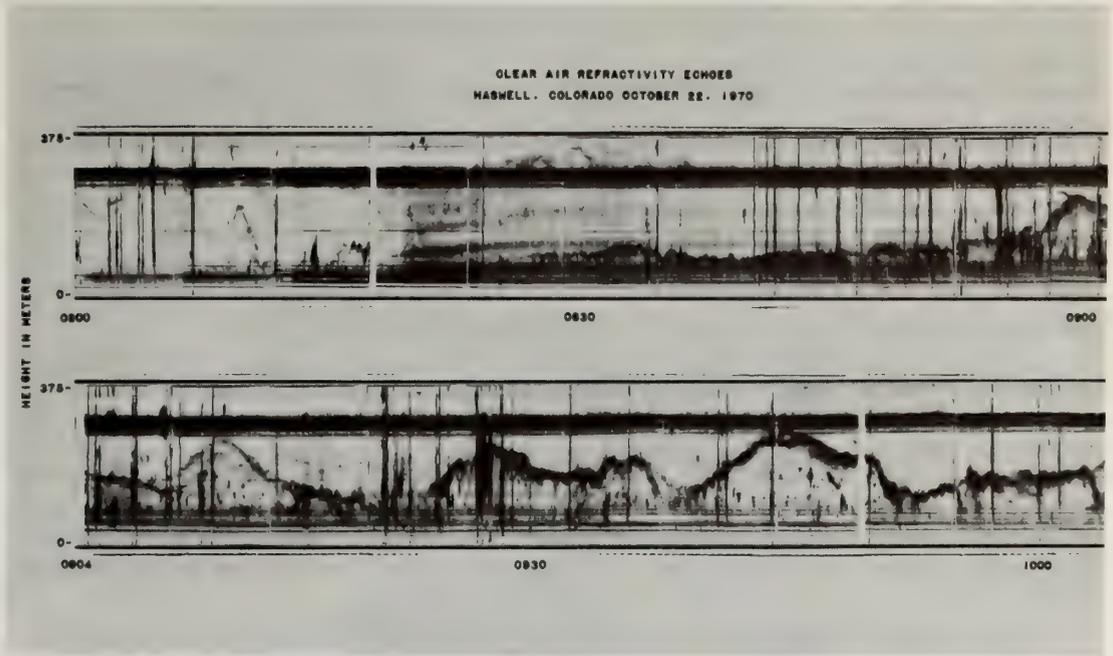


Figure 20.6 Radar record of clear air returns of October 22.

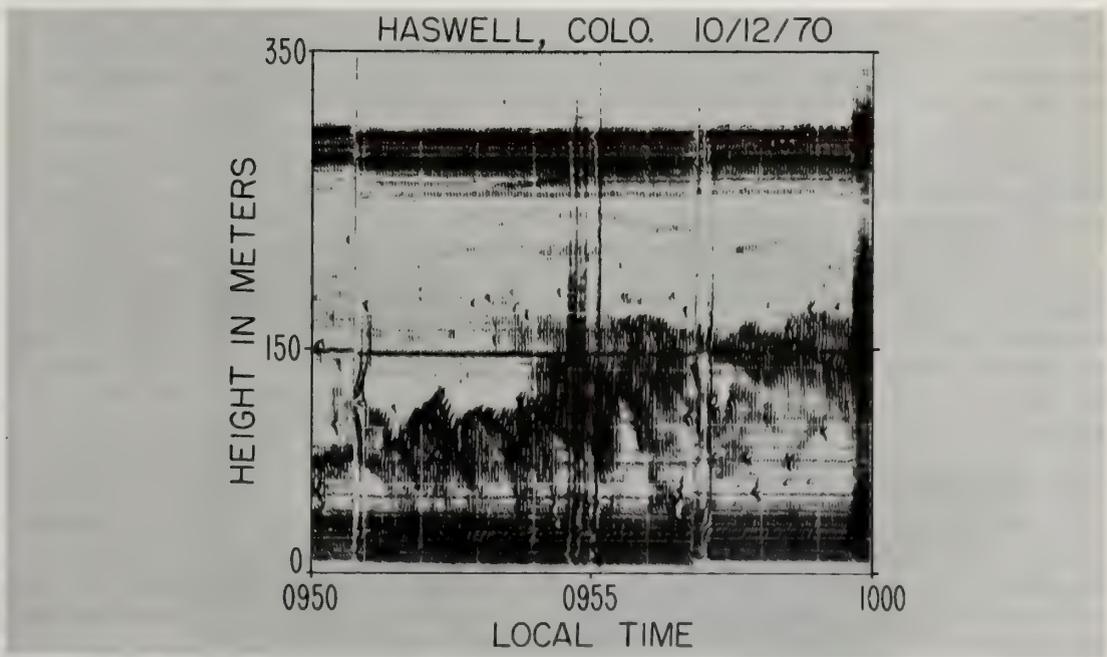


Figure 20.7 An enlargement of the period 0945-1000 of (F20.5). Note the wide, permanent return from the meteorological tower and the record of the carriage within it.

completely destroyed. The insects appear to act as tracers of an oscillatory flow. One presumes that the "rising" portions of the record are convective elements since the air in the boundary layer is slightly unstable at this time due to solar heating of the ground. The rising portion appears darker, indicating a filling in of the background by mixing of the leading edge of the element with the ambient air while the descending portion displays only the insects, indicating well-mixed air. This is in keeping with the convective plume model of Kaimal and Businger (1970) determined from data in the first 30 meters of the atmosphere. *Note well*, that the plume rises with time, indicating, since we only see that which the wind carries past us, that the plume is bent *backwards* in space relative to the air flow. This is just the opposite of Kaimal and Businger's observations. This discrepancy could be explained by the presence of a wind maximum above tower height. Verification of this conjecture must await subsequent observations. The vertical velocities of these insects are also quite large. For example, at 1530 it is quite easy to follow a single insect rising over 450 meters in 3 minutes with an average vertical velocity of 2.5 ms^{-1} .

A further example (F20.5) illustrates thermal plume modulation of an elevated inversion during the late morning of October 12, 1970. With few exceptions, the insects are confined under the inversion and within the mixing layer. Again, where a convective plume is clearly delineated, such as near 1320, the insect returns indicates that it is bent backwards in space relative to the air flow. The period before 1235 does not show such clear cut convective activity but rather undulations on the inversion with some mixing of insects. The period from 1200 to 1235 shows structures reaching to the ground that are spatially tilted *forward* with the wind. These are assumed to be instability waves on the inversion that break to produce trailing wakes. Although they are not Kelvin-Helmholtz waves, their physical cause has not yet been explained.

Waves of the Kelvin-Helmholtz (K-H) variety are shown on (F20.6) for the morning of October 22, 1970. The record starts at 0800 with faint echoes of high amplitude internal gravity waves of the characteristic shape that Gossard and Richter (1970) attribute to non-linear third order Stokes waves. The inversion then rises slowly until just before 0900 when a large amplitude wave appears. The characteristic braided return of K-H instability appears from 0901-0904, mixing presumably occurs thus weakening both the contrast in N across the layer and the resultant radar returns. The layer remains relatively weak until about 0925 when wave action again commences. The long wave at 0945 crests with a roll that captures clear ambient air in its center. Some instability is indicated by the diffuse returns on the downward side of the wave. Note that the returns from this wave are as strong as those from the tower.

Such waves are by no means a rarity. The author's colleague, B. D. Warner, informs him that on clear nights with calm surface conditions and strong winds aloft, wind-driven waves are often produced on elevated inversions. He notes the persistence of these waves for several hours and a strong tendency towards a 6 to 8 minute period with a variation in wave surface as much as 400 m in a few minutes.

It is this rapid variation in the height of such inversions that makes difficult the intercomparison of such remote and in-situ measurements. For example, (F20.7) illustrates a nocturnal radiation inversion being lifted by solar heating on the morning of October 12, 1970. Examine the record just prior to 1000. One notes that the layer lifts from 60 to 152 m in the course of 5 minutes. Notice that imbedded within the permanent return from the tower one can see the carriage moving from 60 to 152 m. The motion of the carriage and the layer are so similar that one would draw a completely different view of the atmosphere's structure than if a straight traverse of the layer had been made as in (F20.1).

20.4 Comparison of Acoustic Sounder and Tower Data

One year earlier, a similar experiment was conducted at Haswell, Colorado, utilizing the Australian acoustic sounder under the direction of L. McAllister. The sounding system has been described elsewhere (McAllister, 1969), as well as in earlier chapters of this lecture series. In this instance, however, the tower was instrumented in greater detail to obtain full turbulent exchange measurements of heat, momentum, and water vapor.

A comparison of the sounder return with data is given on (F20.8) for the period 0030-0130 MST on September 16, 1969. One notes that at the start of the record, the 16°C isotherm parallels the top of the strong returns near the ground. Shortly after 0100, this region lifts suddenly with the 16°C isotherm following. One notes that the low level inversion has now been extended greatly in height and, in fact, becomes nearly isothermal. The 6 ms^{-1} isopleth follows a similar pattern, showing a region of lower wind

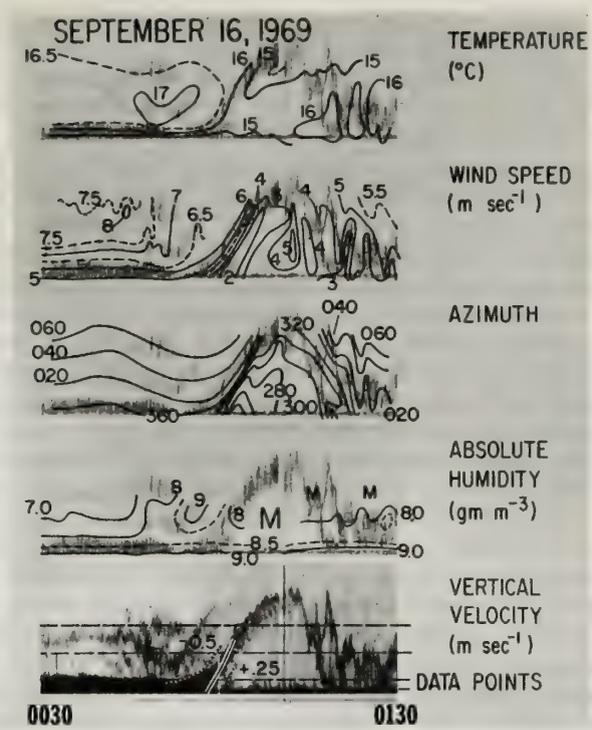


Figure 20.8 Comparison of acoustic sounder return with meteorological tower data for 0030-0130 MST, September 16, 1969.

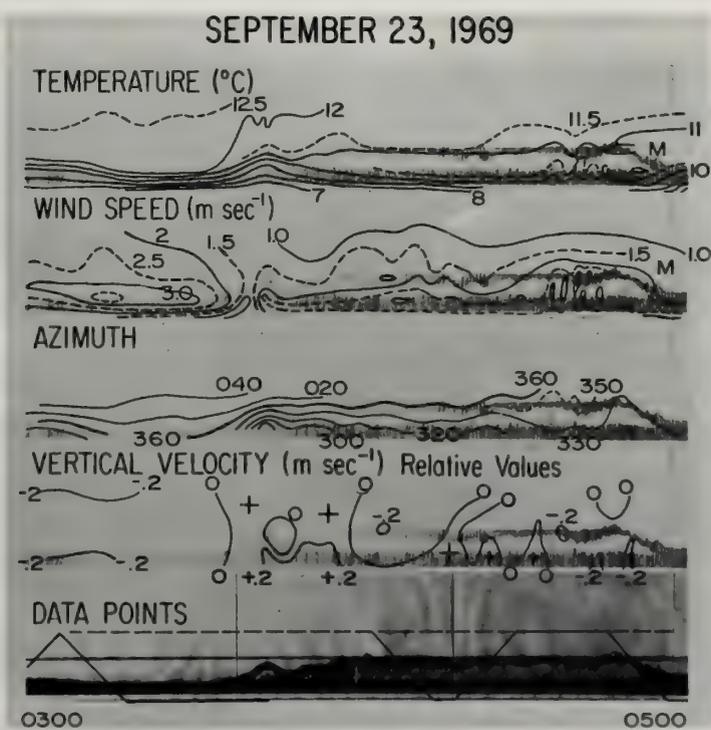


Figure 20.9 Comparison of acoustic sounder return with meteorological tower data for 0300-0500 MST, September 23, 1969.

velocity under the lifted inversion. The most dramatic change comes in wind direction where, with the lifting of the inversion, the azimuth shifts from 20 to 40 degrees to 280-300°, a change of 90 degrees. As the acoustic pattern tends back to its previous pattern, so do the meteorological fields.

A somewhat similar case is presented in (F20.9) for the period 0300-0400 on the morning of September 23, 1969. The top of the sounder record is rather closely followed by the 11°C isotherm even though the top of the returns nearly double in height just before 0400. The clear region below the top of the returns again is a region of decreased temperature gradient. The rise in return level is again reflected in decreased winds and a change in azimuth from 360° to 300-330°.

More detailed examination of the meteorological conditions associated with sounder characteristics are revealed by examination of carriage traverses of the event. For example, (F20.10) compares the wind, temperature, Richardson's number as well as mechanical energy production, $M = \overline{u'w'}(\partial\bar{u}/\partial z)$ and thermal buoyancy, $H = -\overline{w'\theta'}(\partial\bar{\theta}/\partial z)$, with the sounder record of October 9, 1969 (see Chapter 1 for a discussion of these terms). This record shows saw-toothed waves descending after a nocturnal inversion has been lifted in the same fashion as that of (F20.9). There is a slow increase in temperature from the ground to the base of the saw-toothed structure and then a sudden increase of about 2°C accompanied by an increase in wind of 3 mps. At this point, the record is typical of traversing a nocturnal temperature inversion. But note that when the carriage enters the clear area, the temperature and wind *decrease*. This pattern is then repeated for each traverse of the saw-toothed structure. This at first appears to be a rather violent event due to the very nature of the acoustic sounder presentation. Note, however, that the approximately three minutes that the carriage requires to traverse the structure permits a kilometer of atmosphere to advect past the tower while the layer oscillates at most 100 meters. If the sounder record is converted into geometric space by assuming a mean wind, one sees a series of waves with nearly vertical fronts and more gently sloping rear portions, reminiscent of waves coming onto a beach.

The second illustration (F20.11) is chosen from the morning of October 7, 1969, when braided or "rope-like" returns often associated with Kelvin-Helmholtz (K-H) instability (Atlas et al., 1970) were between 46 and 92 m. The temperature structure shows a 5°C ground-based inversion up to 30 m with light winds of about 1 ms⁻¹. It is then isothermal to the inversion base and about one-third into the K-H region. The temperature then increased by 2°C to the top of the structure and then much more slowly.

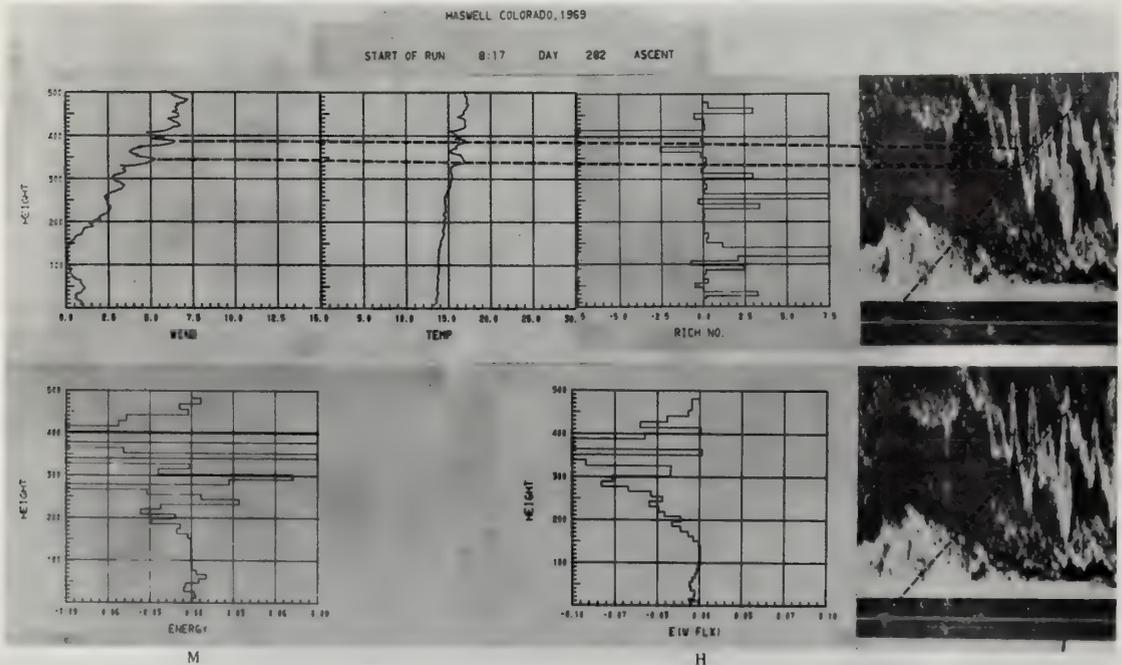


Figure 20:10 Comparison of acoustic sounder return with meteorological tower data for the morning of October 9, 1969.

The wind shows a similar increase within the layer. Richardson's number indicates stable conditions (positive values) below the layer, unstable (negative values) for the bottom third of the layer and stable in the top third of the layer and above. It is interesting to note that mechanical turbulence (M negative) is extracting energy from the mean flow in the upper third of the layer and above while buoyancy forces (H negative) are also doing the same. Normally these terms tend to offset each other.

20.5 Observations of Helmholtz Waves in the Lower Atmosphere With an Acoustic Sounder

The theoretical treatment of the stability of an infinitesimal wave disturbance of the form $e^{i(\alpha x + \beta z)}$ introduced at an internal surface of density as well as velocity discontinuity was first done by Helmholtz (1868). For our present consideration the surface of discontinuity is the temperature inversion as often present in the lower atmosphere. The inversion is produced by either warm air being advected over cold air or the air near the surface of the earth being cooled below that of the overlying air. In either case, less dense air overlies more dense air. The normal winds then effectively "substitute" the "smooth" air interface for the "rough" ground. This allows the wind to increase above the inversion. This increase in wind continues until waves are produced on the interface by the drag of the wind. These waves then may build until they become unstable and break. It is during this breaking process that CAT is believed to occur (Atlas, 1970). Theoretical considerations lead to the following expression for the phase velocity, c , of the wave

$$c = \frac{\rho u + \rho' u'}{\rho + \rho'} \pm \left[\frac{g\lambda}{2\pi} \left(\frac{\rho - \rho'}{\rho + \rho'} \right) - \frac{\rho\rho'}{(\rho + \rho')^2} (u - u')^2 \right]^{1/2} \tag{20:3}$$

where ρ , ρ' and u , u' represent, respectively, the densities and velocities of the lower and upper layers. The first term on the right side of (20:3) represents a weighted (by the density) average of the basic currents in the layers, the second term involves the density and velocity discontinuities and is referred to as the dynamic term. Normally, the density discontinuity term has a stabilizing effect while the velocity discontinuity term has a destabilizing effect on the wave perturbation. The waves whose phase velocity is

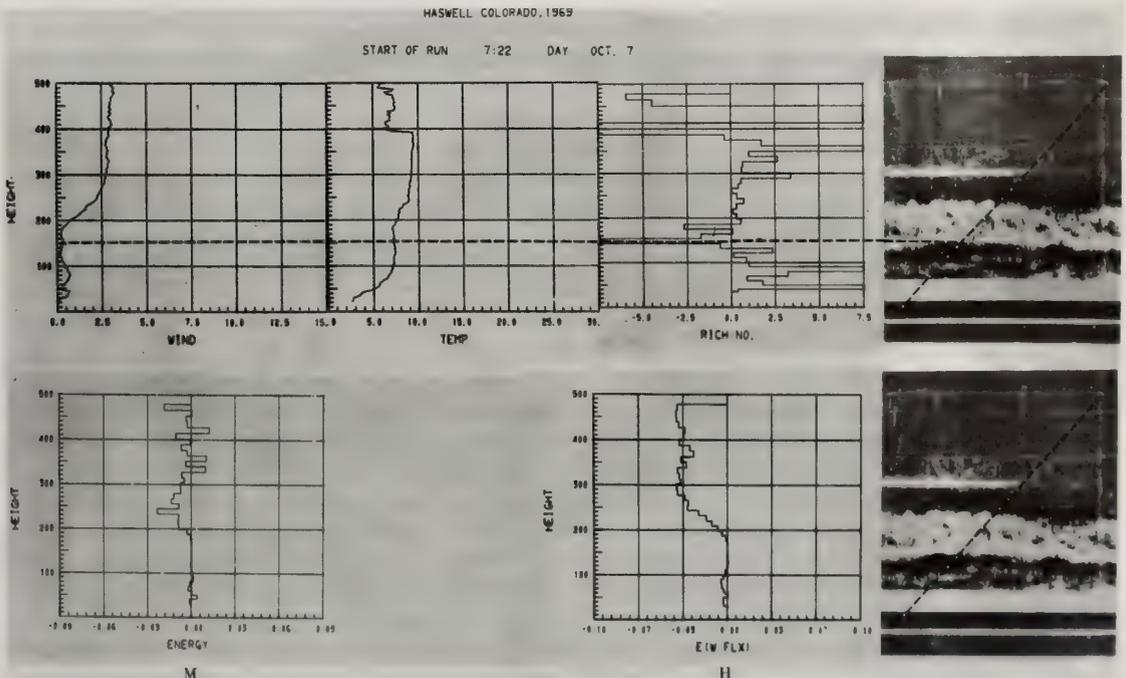


Figure 20.11 Comparison of acoustic sounder return with meteorological tower data for the morning of October 7, 1969.

given by (F20.3) are therefore of a combined gravitational and shear type and are referred to as shearing-gravitational waves, or Helmholtz waves. Specifically, the phase velocity c becomes complex and hence the waves become unstable if the wavelength λ is less than some critical wavelength λ_c , where λ_c is obtained by setting the bracket in (20:3) equal to zero, i.e.,

$$\lambda_c = \frac{2\pi}{g} \frac{\rho\rho'(u-u')^2}{(\rho-\rho')(\rho+\rho')} \quad (20:4)$$

Elimination of the density via the equation of state, using the common notation $\Delta u = u - u'$, $\Delta T = T'$, $\bar{T} = \frac{1}{2}(T + T')$, and assuming $TT' \cong \bar{T}^2$, then (20:4) becomes

$$\lambda_c = \frac{\pi}{g} \frac{(\Delta u)^2}{(\Delta T/\bar{T})} \quad (20:5)$$

If we let $\rho \rightarrow \bar{\rho}$ and use the thickness Δz of the layer of discontinuity as some reference length, we find that the bracket in (20:3) may be written as (see for example, Reiter, 1963)

$$\frac{1}{k\Delta z} \frac{g}{2\rho} (\Delta z)^2 \frac{d\bar{\rho}}{dz} - \frac{(\Delta z)^2}{4} \left(\frac{du}{dz}\right)^2 = \frac{(\Delta z)^2}{4} \left(\frac{du}{dz}\right)^2 \left[\frac{2}{k\Delta z} R_i - 1 \right] \quad (20:6)$$

where $\bar{\rho}$ represents the density of the undisturbed fluid, $k\Delta z$ represents the dimensionless wave number, and R_i denotes the Richardson number

$$R_i = \frac{g}{\rho} \frac{\partial \bar{\rho}}{\partial z} \left(\frac{\partial u}{\partial z}\right)^{-2} \quad (20:7)$$

Since the stability of the Helmholtz model is decided by the sign of the bracket in (20:3), the wave number $k\Delta z$ and the Richardson number affect the stability of the flow. From (20:6) we note that the only portion that determines the sign is that given in the brackets, as a result we find

$$R_i = \pi \frac{\Delta z}{\lambda_c} \quad (20:8)$$

as the condition on the stability boundary.

On the basis of some recent clear-air turbulence studies, Sekioka (1970) finds that the critical layer Richardson number is of the order of 0.5 and that the layer thickness Δz is related to the wavelength as

$$\Delta z = \frac{\lambda}{2\pi} \quad (20:9)$$

In the analysis that follows we have calculated the Richardson number from

$$R_i = \frac{g}{\bar{T}} \frac{\left(\frac{\partial \bar{T}}{\partial z} + \Gamma\right)}{\left(\frac{\partial u}{\partial z}\right)^2} \quad (20:10)$$

where Γ is the adiabatic lapse rate of temperature. Upon equating (20:8) and (20:10) we find that the critical wavelength is now given by

$$\lambda_c = \frac{\pi}{g} \frac{(\Delta u)^2 \bar{T}}{(\Delta T + \Gamma \Delta z)} \quad (20:11)$$

as the more appropriate expression for the atmospheric case.

The experimental work was carried out at Haswell, Colorado, during the latter part of 1969 utilizing the acoustic sounder.

The acoustic sounder, an array of 49, 20 cm loud speakers, was located 250 m away from the meteorological tower along a line in the northwest direction. It operated at a frequency of 950 Hz, 20 ms pulse duration at 2 sec intervals with a power output of eight acoustic watts (McAllister, 1968).

From the many acoustic sounder records available we utilize the few time periods for which there are carriage traverses through the layered structures. The carriage records were checked for consistency against the fixed level measurements and our attention in this study shall be confined to the carriage data. The time periods chosen for analysis are: 0513-0533, 0717-0734, 7 October 1969; 0104-0134, 8 October 1969; 0812-0832, 9 October 1969.

For each carriage traverse the following parameters were obtained:

- (i) the mean wind profile,
- (ii) the mean temperature profile, and
- (iii) the layer Richardson number.

The carriage profiles were subjected to a digital filter (Ormsby, 1961) to separate the mean profile from the fluctuations having dimensions less than 3 m. The resulting profiles were used to determine \bar{u} , \bar{T} , and R_i .

Figure (20.12), Case I, presents the acoustic sounder record for the period 0513-0533, 7 October 1969, as well as the mean horizontal wind and temperature profiles measured at the tower. Below the layer indicated by the strong acoustic returns, the wind speed is constant and approximately zero. Through the layer in the region of instability waves as indicated by the acoustic sounder record, the wind speed increases with height to a maximum of approximately 3 m sec⁻¹. Immediately above the layer the wind speed decreases, then attains a constant value of approximately 2.75 m sec⁻¹ at about 120 m. A temperature inversion is indicated at the base of the layer, an increase of 1°C through the layer (~0.03°C m⁻¹), i.e., a second inversion is indicated at approximately the top of the layer. The temperature then becomes constant above 120 m.

Figure (20.13), Case II, presents the same information as Case I, but for the period 0717-0734, 7 October 1969. A wind speed of approximately 0.4 m sec⁻¹ prevails below the layer. Through the layer the wind gradient is approximately 0.08 sec⁻¹, then the wind attains a constant value with height. The temperature shows an increase with height below 30 m, it remains approximately constant between 30- and 57-m. Through the layer the temperature shows an increase of 1.65°C, then undergoes a slight increase of (~0.5°C) to about 115 m.

The acoustic sounder record and the prevailing wind and temperature profiles for the time period 0104-0134, 8 October 1969, Case III, are shown (F20.14). In this case sharp gradients in both wind and temperature are evident in the tower profiles. The thickness of the layer at the height where the carriage passes through is only about 10 m. The thickness, however, is not constant with time; the layer thickness appears to more than double shortly after 0111 and indications are present of the existence of two layers from about 0116 onwards. During and immediately prior to the carriage traverse through the layer itself, small scale instabilities are shown. These have an average wavelength of approximately 30 m, well below the maximum unstable wavelength of some 90 m as predicted by the Helmholtz model. In this case, therefore, we have small scale wave instabilities imbedded in a large scale wave motion (wavelength ~ 350 m) which appears to be stable.

Figure (20.15), Case IV, presents the pertinent information for the time period 0812-0832, 9 October 1969. The marked difference between this record and those shown in the previous figures is evident. Although this "jagged" structure might initially be interpreted as an extremely violent event, it merely represents an undulating layer. Superposed on this layer, small scale instabilities may be seen on the upward side of some of the undulations. These instabilities have an average wavelength of 30 m, well below the maximum unstable wavelength as predicted by the Helmholtz model, i.e., (20:11). The undulating layer itself as depicted by the acoustic sounder echoes has an average wavelength of 120 m.

For the cases presented, the Richardson number was calculated from the measured wind and temperature profiles over a vertical spacing equivalent to the depth of the layer as determined from the acoustic sounder returns. In addition, the maximum unstable wavelength, λ_c , was also calculated. Since the phase velocity of the Helmholtz waves is equal to the mean wind speed of the layer, we were able to calculate the wavelengths of the "breaking" waves from the discernible "rolls" as indicated by the acoustic

sounder returns. Accordingly, we may define a stability criterion in terms of the average measured wavelength λ , and the maximum unstable wavelength $\bar{\lambda}_c$, i.e.,

$$w_s = \frac{\bar{\lambda}}{\lambda_c} > 1 \text{ stable}$$

$$w_s = \frac{\bar{\lambda}}{\lambda_c} < 1 \text{ unstable}$$

Table(20.1) summarizes the pertinent results of the present study.

The results presented for Cases I and II represent the "breaking" phenomenon as predicted by the Helmholtz model. The wavelengths λ given in (T20.1) do fall within the instability range as predicted by theory. For each case, the average wavelength was calculated from the measured wavelengths and was used in the evaluation of w_s . Values of w_s so obtained are shown in (T20.1) and depict the Helmholtz instability rather clearly.

The Richardson number of the layer for Cases I and II is of the order of 1/2. The layer thickness was determined from the acoustic sounder records, although this determination could well have been made from the carriage profiles as well.

Table 20.1. Critical wavelength as determined from the Helmholtz theory and wavelengths as determined from the acoustic sounder returns. The layer Richardson number and the stability parameter w_s are also given.

Case	$\Delta z(\text{m})$	$R_{i,L}$	w_s	$\bar{\lambda}(\text{m})$: Acoustic Sounder	$\lambda_c(\text{m})$ (20:9)	$\lambda = 2\pi\Delta z$	Remarks
I	30	0.50	0.63	155	188	188	$\bar{\lambda}$ was calculated from peak-to-peak measurements of the discernible "rolls" shown on the acoustic sounder records. See text for additional comments.
II	27	0.51	0.38	100	166	170	
III	5	0.17	0.33	30	90	31	
IV	6	0.10	0.48	30	57	38	

One notes in (T20.2) that the stable wave behaves as one might expect with relatively small energy change and small dissipation. The same is true of the breaking wave case *before* the wave breaks, but not so after breaking when mechanical production and dissipation increase by 10 to 20 times. It is of interest to note that the thermal plume case exceeds the breaking wave case in all terms but dissipation. Our preliminary conclusion from these observations is that K-H instabilities are indeed sources of CAT, particularly in regions well above the region of convective mixing.

Table 20.2 Variation of terms in the kinetic energy equation for stable layer, or breaking waves, and thermal plumes. All terms are in $(\text{m}^2 \text{sec}^{-2}) \text{sec}^{-1}$, Haswell, Colorado, October 1969.

Energy Term	Stable Wave	Breaking Wave		Thermal Plume
		Before	After	
Mechanical Production	~0	5	50	+400
Thermal Buoyancy	~0	5	-30 to +30	100
Vertical Divergence	~0	-3	-20	100
Dissipation	10	10	190	70

The acoustic sounder records presented for Cases III and IV show undulating layers upon which smaller scale instabilities may be seen. For these smaller scale instabilities the layer Richardson number is found to be less than $1/4$, the critical Richardson number as found from theoretical considerations by Taylor (1931), Miles (1961). Strata of $R_i \leq 1/4$ are interpreted as regions capable of generating Helmholtz waves, these grow in time until finally they "roll" and "break" as clearly shown on the first two cases.

The acoustic sounder records obtained in this experimental study point to the fact that the sounder cannot only map regions of instability waves, but can also delineate such time periods most suitable for analysis of particular atmospheric phenomena.

The independent and concurrent meteorological measurements taken at the tower support well the acoustic sounder records thus lending credit to the use of acoustic sounder techniques in low level atmospheric studies.

The Richardson number calculated from the tower wind and temperature profiles for the layer thickness as indicated by the acoustic sounder records is found to be approximately 0.5. Consequently, we interpret this to be the critical layer Richardson number for the Helmholtz instability.

Recent work utilizing powerful 10 cm wavelength pulsed radars in both the United States and England yield similar conclusions. For example, Boucher (1970) utilizing the Wallops Island radar, observed a case of "rope-like" radar returns that corresponded to a region of moderate turbulence as detected by an aircraft dispatched to the region. The region also exhibited strong wind shear ($\geq 2 \times 10^2 \text{ sec}^{-1}$) and a temperature inversion that combined to yield Richardson's numbers of 0.16 and 0.26 at the base of the inversion. In a somewhat similar study Browning (1971) observed some 17 different "rope-like" or Kelvin-Helmholtz instabilities between 3 February and 10 July 1970 at Defford, Worcestershire, U.K. These events were also observed at a stable interface with wind shear above to yield Richardson's numbers of 0.1 and 0.3. This is quite consistent with the conclusions reached from the Haswell data since these types of radar returns represent the wave *after* it has broken and this one would expect $R_i < 0.5$.

In checking the critical wavelengths from the data of these papers, one finds that $\lambda \leq \lambda_c$ in every case of "rope-like" radar returns. A pattern is thus emerging which tells us that the large-scale Kelvin-Helmholtz instabilities as observed by various radars appear to be reasonably described by the relatively simple theory of the preceding sections. A nagging thought does persist, however. The radars utilized for the above studies obtain returns from regions of intense *small-scale* (\sim one-half radar wavelength) fluctuations of *refractive index*, not wind or energy fluctuations of the magnitude to produce CAT. One concludes, then, that the radar returns are from regions of intense mixing of air across the temperature inversion and *outline the lower boundary of the wave or instability event*.

As a final observation, some preliminary results are presented of the energy balance as calculated from the 300 foot level of the Haswell tower during stable waves, breaking waves and thermal plumes. These data represent tentative conclusions as of this writing. We do, however, believe they will serve to delineate the trend of our final conclusions.

20.6 A Note of Caution

The above observations involve either fixed towers or vertically pointing radars. These, then, are observations of atmospheric structure as presented to the observer by the prevailing wind. This fact alone may bias the view of the atmosphere. For example, the author's airborne measurements of vertical flux of water vapor were made both down- and cross-wind at various heights within the BOMEX† array. Although the over-all flux was the same ($\sim 0.5 \text{ cm/day}$), the spectral distribution is quite different for the two aspects (F20.16). For example, the cross-wind spectra display sharp, well-defined peaks that show a tendency to move to lower frequency with increasing elevation while the down-wind spectra show broad maxima at lower frequencies yet. This suggests that the mechanism of transporting water vapor vertically is quite aspect sensitive, perhaps helical rolls elongated in the direction of the wind.

One would not be able to discern this difference with tower or sounder measurements and would thus reach quite different conclusions than if the cross-wind observation were available. One also concludes that an effective remote sensing technique must be able to scan in azimuth as well as height. Such instruments, capable of probing to the detail of the above observations are not as yet available.

† Barbados Oceanographic and Meteorological Experiment

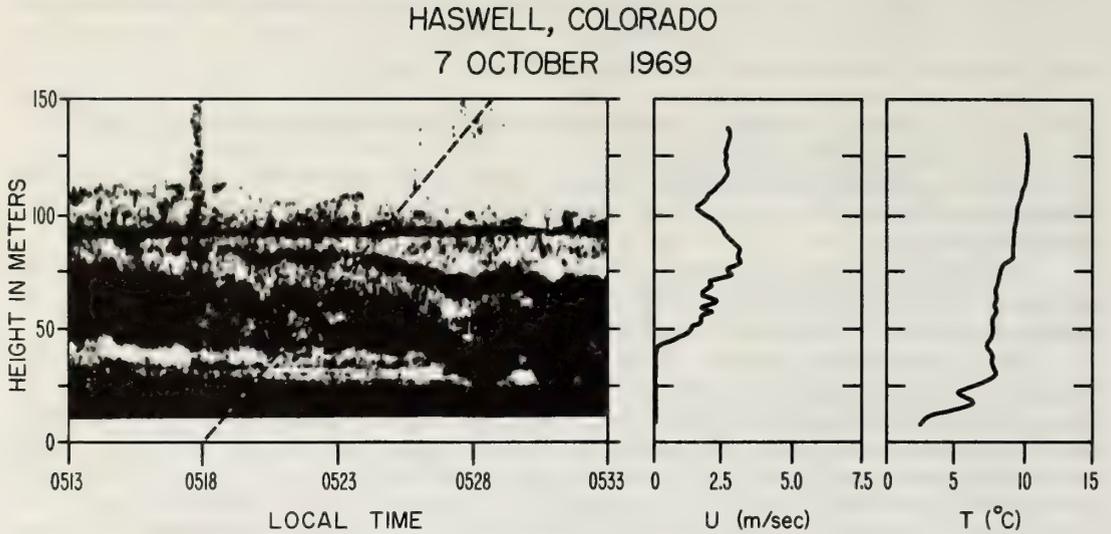


Figure 20.12 *Acoustic sounder record for 0513-0533, 7 October 1969, and the wind and temperature profiles as measured at the meteorological tower. The dashed line on the acoustic sounder record indicates the carriage traverse.*

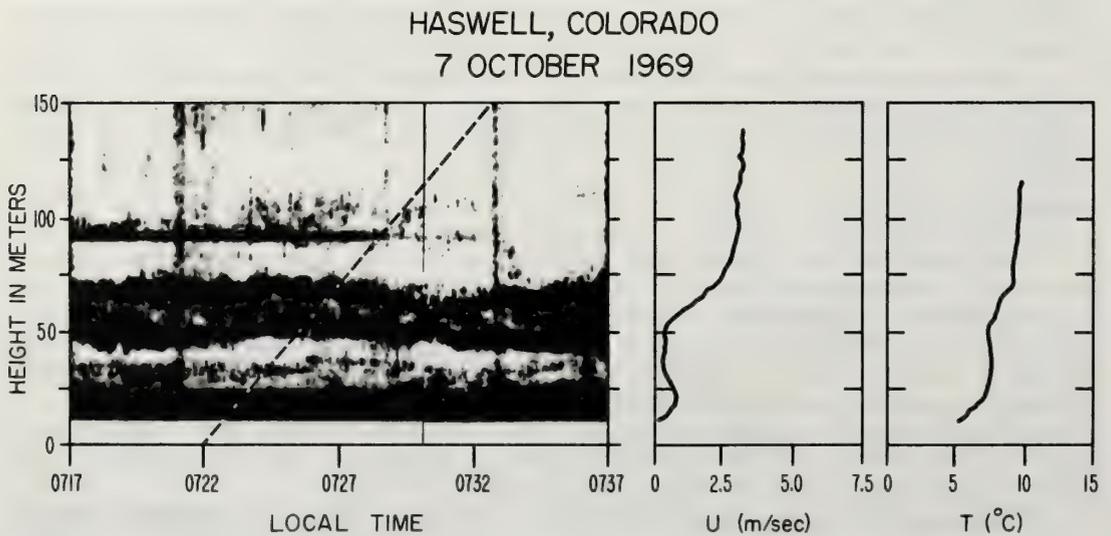


Figure 20.13 *Acoustic sounder record for 0717-0734, 7 October 1969, and the wind and temperature profiles as measured at the meteorological tower.*

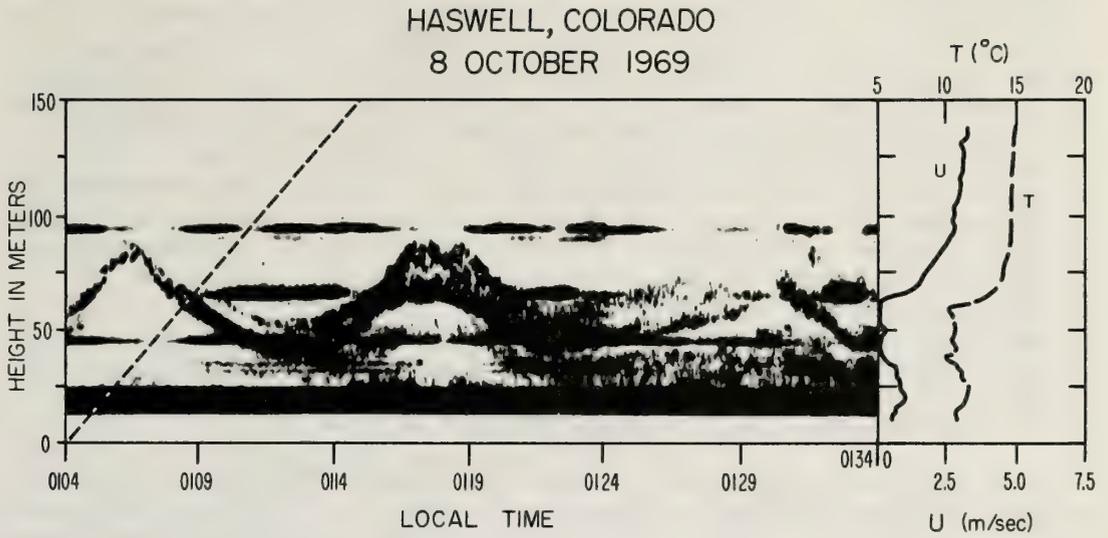


Figure 20:14 Acoustic sounder record for 0104-0134, 8 October 1969, and the wind and temperature profiles as measured at the meteorological tower.

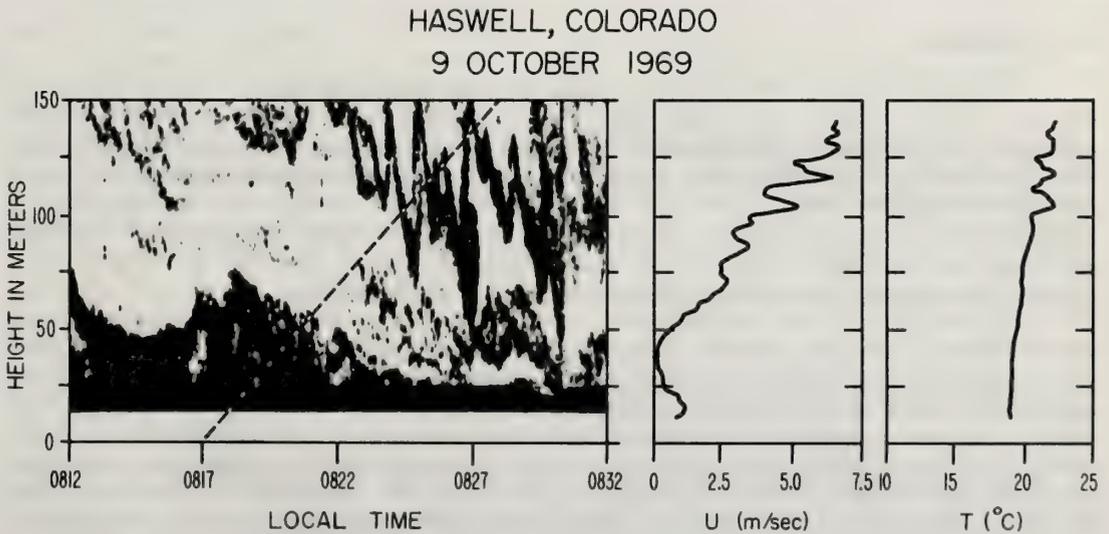


Figure 20.15 Acoustic sounder record for 0812-0832, 9 October 1969, and the wind and temperature profiles as measured at the meteorological tower.

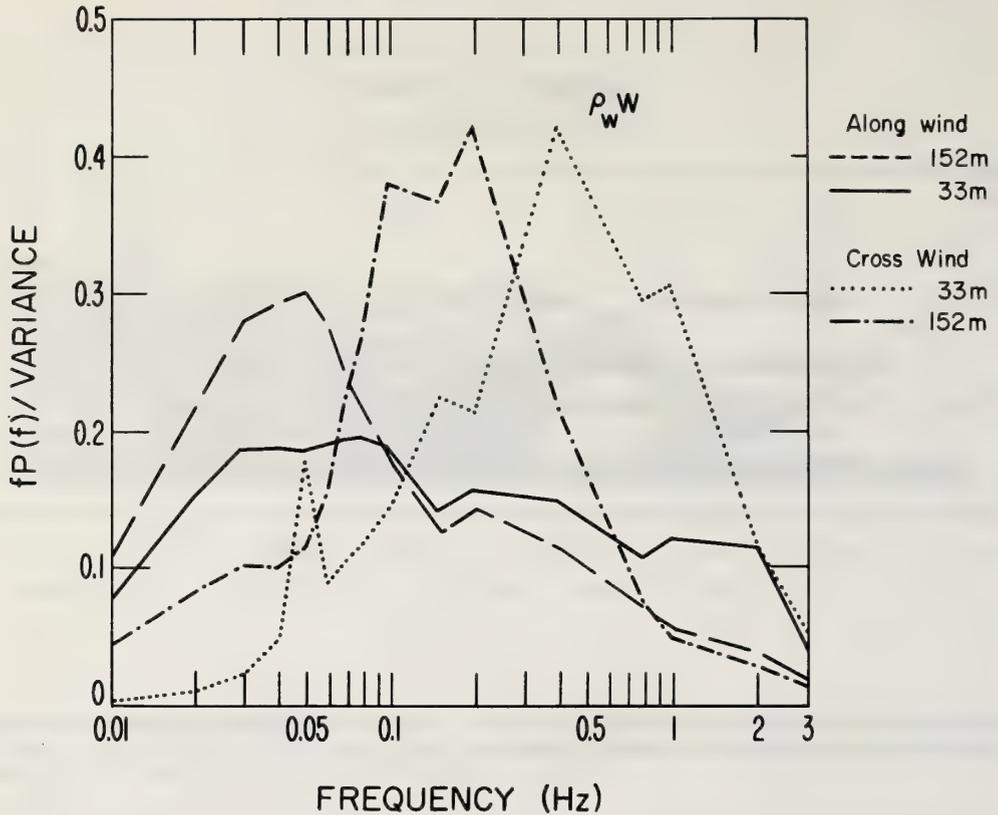


Figure 20.16 Along and cross wind spectra of the vertical transport of water vapor ($\rho_w w$) during BOMEX.

20.7 Conclusions

Perhaps the most important conclusion is that many boundary layer studies would be enhanced by the use of a clear air radar. For example, the author is at present concerned with the budgets of heat, momentum, and water vapor as determined from the 150 m tower. In such budgets, one must determine the flux quantities as well as the profile characteristics, i.e., averages and covariances must be determined over some representative averaging time. The radar indicates that many times are not as stationary as they appear to the eye and can serve to delineate those periods most suitable for analysis of particular phenomena such as instability waves and convective processes. One may also make the complementary conclusion that the in-situ measurements aid in interpreting the sounder records.

The physical model that unfolds is that of an atmosphere comprised of shallow strata which are characterized by a temperature inversion and enhanced wind shear. These strata are accompanied by small scale temperature irregularities which acoustic sounding techniques are able to map via the echo returns which are obtained from such regions. The accentuation of the wind shear across a statically stable layer leads to a dynamic instability of the flow and intermittent generation of more random turbulence. The instability appears initially as bursts of successive waves which are attributed to a cycling process of the interaction between the wave and the background wind field. The amplification of these waves then depends on whether there is a continued excess of shear kinetic energy over buoyancy energy. For a horizontally stratified atmosphere, the source of the instability must necessarily lie in the energy stored in the kinetic energy of relative motion of the different layers. Consequently, the greater the prevailing wind shear, the greater will be the tendency towards mixing and instability. The instability, whenever it occurs at the interface of two such layers, is referred to as the Helmholtz instability. Furthermore, the only counteracting force which is capable of damping the instability is derived from buoyancy; and so long as this force can suppress the mixing, the Helmholtz instability will not occur.

From the results presented, it appears that the Helmholtz instability occurs in a shallow stratum in which the local Richardson number is less than or near 0.25. Within such strata, small wave perturbations will have a tendency to amplify and, as a result, there will be a net kinetic energy released into turbulence.

The critical Richardson number, $R_i = 0.25$, obtained by Taylor was for particular type flows. The work of Miles, however, is of a more general nature and he, too, obtains $R_i = 0.25$ as the critical value. The Richardson number profiles, which have been presented for each carriage traverse, certainly show shallow strata within the layer specified by the Helmholtz waves where the local Richardson number attains the value $R_i \leq 0.25$. These regions represent, then, strata within which small perturbations may amplify until ultimately they "roll" and "break" over a layer thickness which is several times as large as these generating regions. In addition, we have found that the Richardson number over the layer thickness occupied by the Helmholtz waves is of the order of 0.50. Consequently, one concludes that although $R_i = 0.25$ is a necessary condition for the onset of the Helmholtz instability, once such waves begin to grow, they have the effect of redistributing the shear and temperature fields. The growth of these waves ceases when the layer Richardson number becomes 0.50, at which point all Helmholtz waves "roll" and "break." This also implies that all Helmholtz waves will attain the same amplitude within the layer provided the dynamically unstable layer is horizontally homogeneous and steady.

Measurements of the wavelengths associated with the "breaking" phenomenon conform to the theoretically predicted range of values.

From the fixed level measurements of the pertinent meteorological parameters, the intensity of turbulence has been found to be several orders of magnitude greater within the regions occupied by instability waves than it is in clear, or no echo return, regions.

20.9 References

- Atlas, D., J. I. Metcalf, J. H. Richter, and E. E. Gossard (1970), The Birth of "CAT" and Microscale Turbulence, *J. Atmos. Sci.* 27, 903-913.
- Bean, B. R. and E. J. Dutton, 1968: *Radio Meteorology* (Dover Publ., Inc., New York)
- Bean, B. R. and B. D. Warner, 1967: Some Radio-Physical Considerations in Studies of the Fine Scale Structure of the Atmosphere, *Atmospheric Turbulence and Radio Wave Propagation*, Editors A. M. Yaglom and V. I. Tatarsky, Publishing House, NAUKA, Moscow.
- Gossard, E. E. and J. H. Richter (1970), The Shape of Internal Waves of Finite Amplitude from High-Resolution Radar Sounding of the Lower Atmosphere, *J. Atmos. Sci.* 27, 971-973.
- Helmholtz, H. (1868), *Über Discontinuirliche Flüssigkeitsbewegungen*, *Mber. Preuss. Akad. Wiss.*, 215.
- Kaimal, J. C. and J. A. Businger (1970), Case Studies of a Convective Plume and a Dust Devil, *J. Appl. Meteorol.* 9, 612-620.
- Little, C. G. (1969), Acoustic Methods for the Remote Probing of the Lower Atmosphere, *Proc. IEEE* 57, No. 4, 571.
- McAllister, L. G. (1968), Acoustic Sounding of the Lower Troposphere, *J. Atmos. Terrest. Phys.* 30, 1439.
- McAllister, L. G., J. R. Pollard, A. R. Mahoney, and P. J. R. Shaw (1968), Acoustic Sounding—A New Approach to the Study of Atmospheric Structure, *Proc. IEEE* 57, 579.
- Miles, J. W. (1961), On the Stability of Heterogeneous Shear Flow, *J. Fluid Mech.* 10, 496.
- Ormsby, J. F. A. (1961), Design of Numerical Filters with Applications to Missile Data Processing, *J.A.C.M.*, July, 440.

- Ottersten, H. (1969), Radar Backscattering from the Turbulent Clear Atmosphere, *Radio Sci.*, 4, 12, 1251-1255.
- Reiter, E. E., 1963: *Jet-Stream Meteorology*, (The University of Chicago Press.)
- Richter, J. H. (1969), High Resolution Tropospheric Radar Sounding, *Radio Sci.*, 4, 12, 1260-1268.
- Sekioka, M. (1970), Application of Kelvin-Helmholtz Instability to Clear Air Turbulence, *J. Appl. Met.* 9, No. 6, 896.
- Strand, O. N. (1971), Numerical Study of the Gain Pattern of a Shielded Acoustic Antenna, *J. Acoustic Soc. Amer.* 49, 6, Pt. 1, 1698-1703.

List of Symbols

- FM/CW – Frequency Modulated Continuous Wave
- Δn_{20} – Difference in refractive index, n measured over 20 cm
- C_n^2 – refractivity structure function
- η – radar cross section
- R_i – Richardson number
- g – acceleration of gravity
- T – atmospheric temperature
- Γ – adiabatic lapse rate
- u – horizontal wind speed
- N – radio refractivity = $(n-1)10^6$
- P – total atmospheric pressure
- e – partial pressure of water vapor

APPENDIX A20 Reprint of
Richter, J. H., 1969: High Resolution Tropospheric
Radar Sounding, *Radio Science* 12, pp. 1261-1268.

Radio Science, Volume 4, Number 12, pages 1261-1268, December 1969

High resolution tropospheric radar sounding

Juergen H. Richter

Radio Physics Division, Naval Electronics Laboratory Center
San Diego, California 92152

(Received August 4, 1969.)

A radar sounder for the study of the refractive index structure in the troposphere is described. The specially built, high-sensitivity FM-CW radar has a maximum range resolution of one meter. Design considerations and performance characteristics are given, and recordings obtained with the radar sounder are presented and discussed. The predominant layers are very well correlated with the major inversions at which we find either sharp vertical gradients of refractive index or changes in the gradient. The base of the lowest inversion is almost always so marked. A second echo layer frequently appears some tens of meters above this base level. Although the echo strata may sometimes be 30 to 40 meters in depth, they are often only a meter or less thick. Their reflectivities may be 10 to 100 times larger than those previously reported with radars of coarser resolution. The layers also commonly appear in the form of breaking Kelvin-Helmholtz waves with periods of 2 to 4 minutes and peak-to-peak amplitudes as small as 10 to 15 meters.

INTRODUCTION

The effects of the troposphere on microwave radio propagation have been studied extensively, but in spite of theoretical and experimental efforts some basic questions still remain unresolved. These questions concern the temporal and spatial behavior of the refractive index structure in the troposphere. For example, we would like to know the true thickness of the thin radar backscatter layers that *Hardy et al.* [1966] have observed by ultra-sensitive radar and the mechanisms responsible for the development of the fine-scale refractivity inhomogeneities which must be present to explain those echoes.

To study these phenomena we need a radar sounder that combines the properties of high-range resolution, high sensitivity, and close minimum range without clutter. A radar sounder with these characteristics has been designed and built at the Naval Electronics Laboratory Center in San Diego. This paper describes the radar and presents some of the observations of clear-air-scatter layers along with a preliminary analysis.

RADAR SYSTEM

The requirement of high-range resolution capability would suggest the use of pulse compression techniques. However, any pulse radar imposes the

restriction of a minimum range below which no return signals can be observed. This restriction is due to the fact that during the transmission of the pulse (and usually several pulse widths after) no signal can be received. The reason that the minimum range is such an important factor for the sounder is that low-level inversions frequently produce scatter layers in the lowest 300 meters of the atmosphere, and it is these which can be studied most effectively by simultaneous meteorological observations by tower and tethered balloon instruments.

FM-CW radar. A radar that does not have the minimum range restriction and that allows high-range resolution is an FM-CW radar. It has the additional advantage of a peak-power-to-average-power ratio of unity that is very desirable from a technological viewpoint. It also permits a change of resolution without affecting the transmitted energy (which is not the case for pulse radars). To discuss the features and requirements of this radar, the principle of its operation is given briefly in Figure 1. The top portion shows a linear ramp (solid line) which is transmitted during T_M (modulation time) and a reflected ramp (dashed line) which is delayed by Δt according to the distance of the reflecting object. These two ramps are mixed instantaneously, which yields a difference or beat frequency f_b . The middle portion of Figure 1 shows this beat frequency and the lower part its frequency spectrum. In the

JUERGEN H. RICHTER

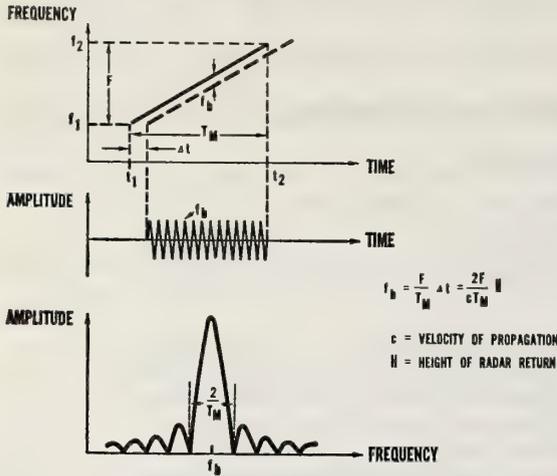


Fig. 1. Principle of FM-CW radar using linear frequency modulation. The sinusoidal beat frequency f_b is generated by instantaneous mixing of the transmitted and the received ramp.

case of multiple targets, reflected waves will arrive at different time intervals and cause different beat frequencies that are superimposed. A spectrum analysis of the beat frequencies allows the different targets to be resolved according to their range. The amplitudes of the different beat frequencies are a measure of the reflection coefficients of the targets. For this kind of radar, range resolution means the ability to separate adjacent spectra of the kind shown in the lower portion of Figure 1. The minimum resolvable distance h is given by

$$h = c/kF \quad (1)$$

where F is the frequency excursion, c is the velocity of propagation, and k is a constant whose value depends on the criterion used to define 'separation' of two spectra. (It is, e.g., $k = 1$ for $\Delta f = 2/T_M$ and $k = 2$ for $\Delta f = 1/T_M$, where Δf is the frequency separation between two spectra.) Equation 1 is essentially the same formula as that used for pulse radars and shows that resolution is only dependent on the frequency excursion or bandwidth of the transmitted signal.

Radar design. The frequency of the radar was chosen to be around 3 GHz as a compromise between higher and lower frequencies each of which would have been desirable for different reasons. Thus, to achieve a narrow antenna beam with a small antenna diameter, it would have been better

to select a higher frequency. However, we must also consider the reflectivity-wavelength dependence of the scattering volume. For specular surfaces, the power reflection coefficient increases as the electrical thickness of the transition zone decreases, and we would thus want to use the largest possible wavelength [Atlas and Hardy, 1966]. On the other hand, the predominate evidence available suggests that clear-air-scatter layers are comprised of inhomogeneities in refractivity, producing a reflectivity (η)-wavelength dependence of about $\lambda^{-1/3}$ [Atlas and Hardy, 1966; Hardy et al., 1966; Hardy and Katz, 1969]. Although this wavelength dependence is small, it suggests using the shortest wavelength provided that it does not approach the limiting microscale (l_m) of the refractivity perturbations. Atlas et al. [1966] therefore suggest an optimum wavelength of about five times the limiting microscale. In weakly turbulent conditions, l_m is of the order of 1 to 2 cm, and so a λ of 5 to 10 cm seemed optimum.

By means of theoretical calculations and experimental data [Saxton et al., 1964; Atlas and Hardy, 1966], it was concluded that with an equivalent pulse length $h = 1$ meter and an antenna gain of 35 decibels, a ratio of 190 decibels between average transmitted and received power should be adequate to detect most scatter layers in the lowest 2 km. The actual sensitivity will be discussed later.

With the above radar parameters determined, our search was then directed toward a suitable FM generator. It is possible to achieve a 3-meter resolution at 100 MHz frequency excursion with a commercially available microwave sweep generator that uses a backward wave oscillator. But the difficulty with this kind of an FM generator is that even after compensating for the nonlinear modulation characteristic, fine-grain variations exist that are due to imperfections of the helix. We then had to find a suitable part of the frequency range of an individual tube. During the course of these investigations a new device became available that consists of a transistor oscillator that is tuned in frequency by changing the magnetic field of a YIG sphere. The tuning linearity of this device is much higher than that previously attainable, and it also does not exhibit the fine-grain variations of backward wave oscillators. The Watkins-Johnson YIG-tuned transistor oscillator that we used in the radar sounder has a deviation of less than 0.001% from straight-line tuning over a frequency range of 200 MHz (2.9 GHz center frequency). The power output is 12 mw and the tuning

TROPOSPHERIC RADAR SOUNDING

sensitivity 3.6 MHz/ma. The high-tuning sensitivity necessitates a very linear, low noise, low ripple sweeping current supply.

The spectrum analysis of the beat frequencies has to be performed in real time, and this is done with a Ubiquitous spectrum analyzer from Federal Scientific Corporation. The analyzer 'time compresses' the signal and performs a scanning analysis on the speeded-up and stored signal. The 10-kHz analysis band has a 20-Hz bandwidth and can be positioned anywhere from 0 to 1 MHz. It takes 50 msec to fill the memory, and this time determines the modulation time T_M . During T_M the memory is updated, and the analysis starts at the end of T_M when the memory is put into 'hold.' The input signal is time-compressed by a factor of 500, which enables the scanning spectrum analysis to be completed in 50 msec. After this time, a new modulation cycle starts. In other words, the entire cycle of frequency sweeping, signal storage, and frequency-spectrum analysis is repeated at a rate of 10 per sec. The output of the spectrum analyzer is multiplied by a sine-squared weighting function which suppresses the side lobes of the $|\sin x/x|$ spectrum shown in Figure 1. The resulting spectrum is somewhat broadened, and the resolution bandwidth is about 30 Hz.

The mixer used for the radar sounder is a balanced mixer using backward diodes with a matched low-noise preamplifier. The noise figure of this combination is 13 decibels at 10 kHz. The incoming RF signal passes through a low-noise traveling wave tube amplifier with a noise figure of 4.8 decibels and 30

decibels gain. The over-all noise figure of the entire receiver is, therefore, only about 5 decibels. A simplified block diagram of the sounder is shown in Figure 2. The linear FM is generated with the YIG-tuned transistor oscillator, amplified to a maximum power of 150 watts and radiated by the transmitting antenna. Part of the transmitted power is delayed and coupled into the backward diode mixer where it is mixed instantaneously with the signal that is received by the receiving antenna and amplified by a low-noise traveling wave amplifier. The reason for the delay of the transmitted portion of the signal is to compensate for delays created by the preamplifiers, the power amplifier, and the cables to and from the antennas. The delay line is exactly long enough to produce a zero beat frequency with the signal that is created by coupling between the transmitting and receiving antenna. The second delay line that couples part of the transmitted energy into the low-noise preamplifier serves as a monitor for the entire system. The length is equivalent to a radar return from a height of 167 meters. The reference signal amplitude created in this manner is a check of all the active components of the system, as well as giving a height calibration and providing a check of the range resolution in connection with another short delay line. The output of the spectrum analyzer is further processed in a signal analyzer that averages consecutive sweeps. The number of sweeps averaged can be selected, thus determining the time constant introduced to the signal. The signals are then displayed on cathode ray tubes in an A-scope presenta-

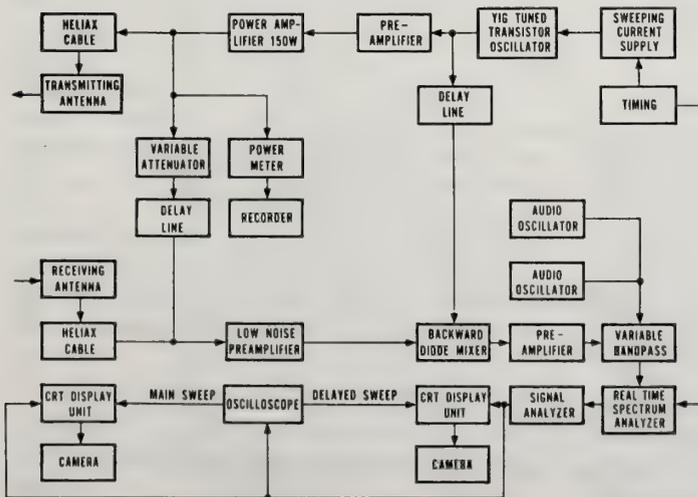


Fig. 2. Block diagram of the radar sounder.

JUERGEN H. RICHTER

tion, or an intensity modulation, or in a combination of both amplitude and intensity modulation (or A-I scope), which will be described below. The horizontal sweep for the display units is provided by an oscilloscope having a delayed sweep, which allows one to view simultaneously the entire analysis band and any enlarged portion. The two audio oscillators provide variable height markings. The data are recorded by 35-mm, shutterless movie cameras with film speeds ranging from 1 cm/hr to 0.5 cm/sec.

The antennas consist of two parabolic dishes, 3 meters in diameter, with waveguide feeds. To achieve good isolation, the antennas were located in pits, an arrangement that has been used successfully before [Fehlhaber and Grosskopf, 1965]. At least 60 decibels isolation are necessary to avoid saturation of the preamplifier. The insides of the pits are lined with microwave absorbers to suppress reflections. The absorbing screens are also required to suppress echoes from nearby ground targets. The antennas are steerable within a limited degree ($\pm 2.5^\circ$) to optimize the common volume for a given height.

Performance characteristics. The performance characteristics of the radar are compiled in Table 1. The basic components of the radar system allow operation in the entire 2-4 GHz range. The restriction to the narrower range around 2.9 GHz is determined by the standing wave ratio of the antenna feed and the balanced mixer. The variable range resolution is an important feature of the radar. In the case of volume scattering, the signal amplitude changes continuously with the resolution, thereby permitting significant conclusions to be drawn. The

maximum range resolution of about 1 meter was determined experimentally. The spectra of two targets 93 cm apart (simulated by delay lines) were separable by using $F = 219$ MHz frequency excursion.

The values for computing the minimum detectable cross section and the minimum detectable reflectivity are based on a minimum detectable signal of -150 dbm. This minimum detectable signal was determined by using a delay line with a delay time equivalent to that of a target at a height of 167 meters. The receiving antenna was disconnected for this measurement. When connected, the baseline noise on the A scope, presumably due to nearby ground-clutter echoes, increases the minimum discernible signal (measured with the delay line target at the fixed range of 167 meters) by some 10 decibels. However, typical atmospheric echoes undulate in range and so can commonly be discerned at lower signal levels than indicated by calibration measurements with the antenna connected. Thus, a minimum detectable signal of -150 dbm is still thought to be a reasonable estimate.

The sensitivity of the radar for a point target is given by

$$\sigma_{\min} = 3.7 \cdot 10^{-6} r^4 \quad (2)$$

where σ_{\min} is the minimum detectable cross section in cm^2 , and r is the distance of the target in kilometers. Typical insects with cross sections of 10^{-8} cm^2 [Hardy and Katz, 1969] can, therefore, cause dot echoes 24 decibels above noise at 1-km height.

For distributed targets, the minimum detectable

TABLE 1. Performance characteristics of the radar sounder

Parameter	Value	Remarks
Power	150 watts	
Center frequency	2.8 to 3.1 GHz	
Frequency excursion	Variable	Linear modulation
Range resolution	Variable	Maximum resolution 1 meter for 200 MHz frequency excursion
Sweep duration	50 msec	Ten sweeps/sec
Receiver noise figure	5 db	
Minimum detectable signal	-150 dbm	
Antenna gain	35 db	
Antenna beam width	2.3 deg	
Isolation between antennas	105 db	
Minimum detectable cross section at 1 km	$\approx 3.7 \cdot 10^{-6} \text{ cm}^2$	
Minimum detectable reflectivity at 1 km	$\approx 4.2 \cdot 10^{-14} \text{ cm}^{-1}$	For 1-meter range resolution

TROPOSPHERIC RADAR SOUNDING

reflectivity [*Atlas*, 1964] η_{\min} in cm^{-1} is

$$\eta_{\min} = 4.2 \cdot 10^{-15} r^2/h \quad (3)$$

where r is the distance in kilometers, and h is the range resolution in meters. Typical values of η for a 10-cm wavelength are $10^{-16} - 10^{-14} \text{ cm}^{-1}$ as given by *Hardy et al.* [1966], *Atlas and Hardy* [1966], and *Hardy and Katz* [1969]. It can be seen from (3) that even for the highest range resolution of 1 meter, returns can be expected from as high as 1 km. To detect the weaker scattering regions at greater ranges, we would have to increase the effective pulse depth or resolution, provided of course that the scatter regions are thicker than the effective pulse.

By way of comparison, the powerful 10.7-cm radar used by *Hardy et al.* [1966] and *Atlas and Hardy* [1966] at Wallops Island has a sensitivity equation of $\eta_{\min} \simeq 6 \times 10^{-17} r^2/h$ (same units as equation 3) for beam filling distributed targets. Although the Wallops 10.7-cm radar thus appears to be 18 decibels more sensitive than our radar, this is true only for scattering regions having a range extent 300 times as great (at maximum resolution) because their pulse length in space is 300 meters. Since we shall see that some scatter layers have a thickness of only a meter or less, the Wallops radar would produce signals comparable to those of the FM-CW radar for such thin layers. Moreover, the absence of appreciable ground clutter permits us to measure at ranges 20 to 100 times as small as those at Wallops Island, thus providing greatly enhanced signals.

Even though there is no basic minimum range restriction inherent in this kind of radar, some practical limitations (which could be overcome if necessary) reduce the sensitivity of the radar for targets closer than 50 meters to the antennas. Reflections from an adjacent building give permanent returns for the first 50 meters in height; therefore targets must be capable of providing stronger returns or must change their position in order to be discernible. Also, the separation of the two antennas imposes a minimum height range for beam intersection. The antennas are separated by 5 meters, and their beams intersect within 50 meters above the ground.

SOME PRELIMINARY RESULTS

Figures 3, 4, and 5 illustrate some of the features of thin scatter layers that are depicted by this radar system.

Figure 3 is a 45-minute section of an intensity modulated trace recorded on January 7, 1969, during a Santa Ana-type occurrence which was observed continuously for three days. (The Santa Ana is a downslope wind flow from the higher desert area east of San Diego. It produces a subsidence inversion leading to dry, warm air aloft and providing a large contrast in humidity and temperature with the marine air next to the sea surface.) The refractive index profile computed from a radiosonde flight at the radar site is shown at the right. (A modified radiosonde system is employed that uses two transmitters on one balloon to provide continuous, fast-response temperature and humidity readings.) Note that the echo from the sonde is seen on the radar record, thus permitting close correlation of the two records. Clearly, there are two predominant echo layers, although three are noticeable at about 0935 PST (Pacific Standard Time). Both the echo layers and the atmospheric strata undulate. At the time of the radiosonde transmissions the bottom echo layer was very close to the height of the sharpest refractive (N) gradient at 136 meters, while the upper thick layer was co-located with the top of the N gradient stratum centered at 167 meters. While the lower echo occurred at the base of the major temperature inversion, the upper one occurred at the point of changing slope in the temperature trace below the top of the inversion. This seems to be a characteristic feature. In other words, the base of the inversion is detected routinely, whereas the top is not. Clearly, there is a close association between the heights at which the vertical N gradient changes sharply and the echo layers. This will be seen in Figure 5 as well and has been observed in many other cases.

The thickness of the echo strata is seen in Figure 3 to vary from about 1 meter, the maximum resolution (by comparison with the thickness of the 167-meter height marker), to about 30 meters. But the 30-meter thickness of the upper layer at 0955 PST is due to a combination of blooming of the display and rapid oscillations in layer height which are apparent under close scrutiny at about 0944 PST. A 10 to 15 meter thickness is more representative of both layers.

Figure 4 illustrates the combined amplitude-intensity (A-I) display recorded on its typically expanded time basis. Each vertical trace on the record is a running average of the signals on the previous 32 frequency sweeps with an exponential weighting

JUERGEN H. RICHTER

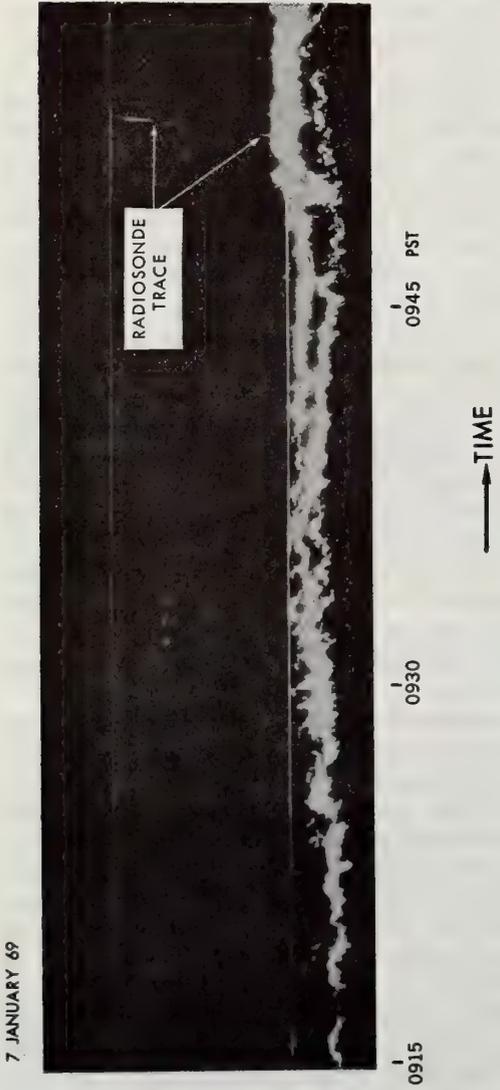
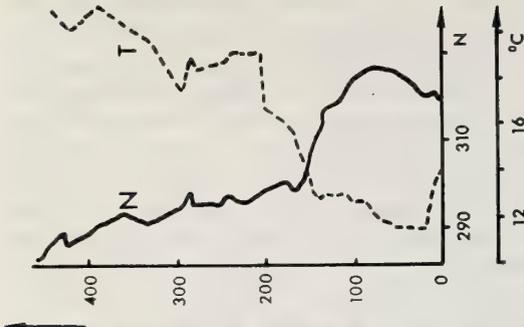


Fig. 3. Typical echo layers as seen with low time resolution and corresponding temperature and modified refractive index profiles. $N = (n - 1) 10^6$, where n is refractive index and T is temperature. (Height in meters)

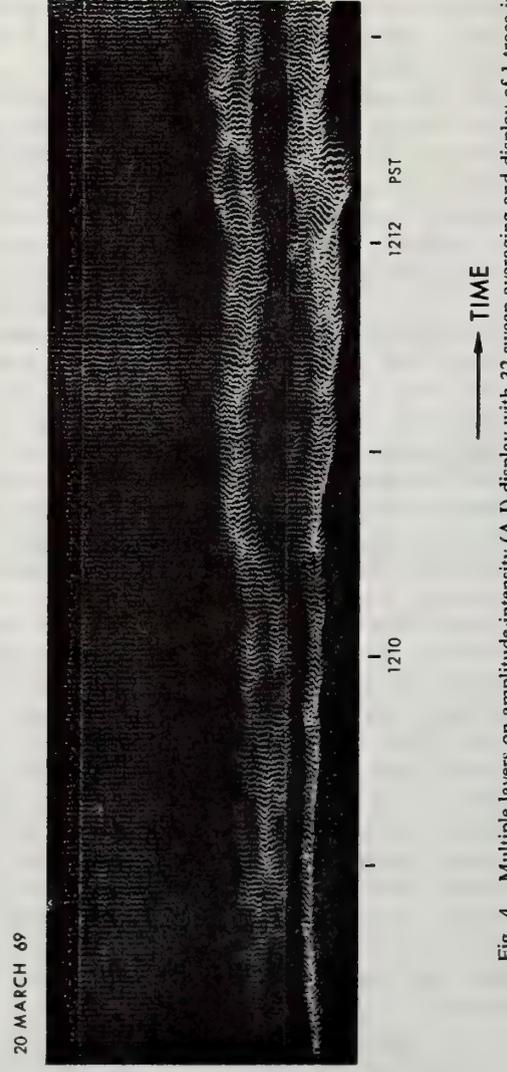
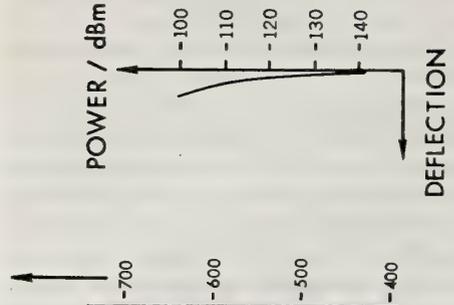


Fig. 4. Multiple layers on amplitude-intensity (A-I) display with 32-sweep averaging and display of 1 trace in 16. (Height in meters.)

7 JANUARY 69

20 MARCH 69

TROPOSPHERIC RADAR SOUNDING

function and a time constant of 3.2 sec. However, only every sixteenth trace is displayed so that the record shows 37.5 sweeps per minute. It is only in this way that the signal amplitudes can be read quantitatively without excessive overlap of adjacent traces. By using the signals themselves to provide simultaneous intensity modulation, the record is shaded so that the brightest echoes are also the strongest. The signal intensities can be scaled off quantitatively from the calibration curve on the right. Because both the fine structure of the layers and their intensity can be obtained from the A-I display, it is especially valuable for this kind of work.

Although no radiosonde record is shown to correspond with Figure 4, the two predominant layers were found close to the base and the top of the major lapse of refractivity. Close examination of the record shows that the bottom layer thickens progressively with time from about 5 meters at the start to a maximum of about 35 meters at about 1213 PST. In fact, the apparent broadening of the stratum just after 1212 PST appears to be due to a jump in the layer height. Careful scrutiny shows that the lower layer has two well-defined intensity peaks at about 1209 PST that are separated by about 5 to 10 meters. Thus, the lower layer may actually be comprised of two undulating bands also, each having a thickness of 10 to 25 meters and occasionally blending to form one apparent, thick layer.

It will be seen that the signal intensity of the lower layer increases to a maximum of about -102 dbm just after 1212 PST. At the corresponding range of 430 meters, this signal corresponds to a volume reflectivity of about $5 \times 10^{-11} \text{ cm}^{-1}$. Although the accuracy is no better than about ± 5 decibels at this stage, the indicated reflectivity is more than 1000 times the maximum reported by *Atlas and Hardy* [1966] and *Hardy and Katz* [1969] at a 10-cm wavelength. If we can assume that the scattering is from refractive inhomogeneities, this suggests that the inhomogeneities observed with our high resolution are indeed much stronger than previously observed either by radar or by refractometers. In fact, this would have to be the case; if the Wallops 10-cm radar reports a reflectivity of 10^{-14} cm^{-1} when only 1% of their scatter volume is filled, then the actual reflectivity would be 100 times as great. Therefore, the reality of such high localized reflectivities can hardly be questioned.

Figure 5 shows an even more interesting case that is typical of many, if not most of our observations.

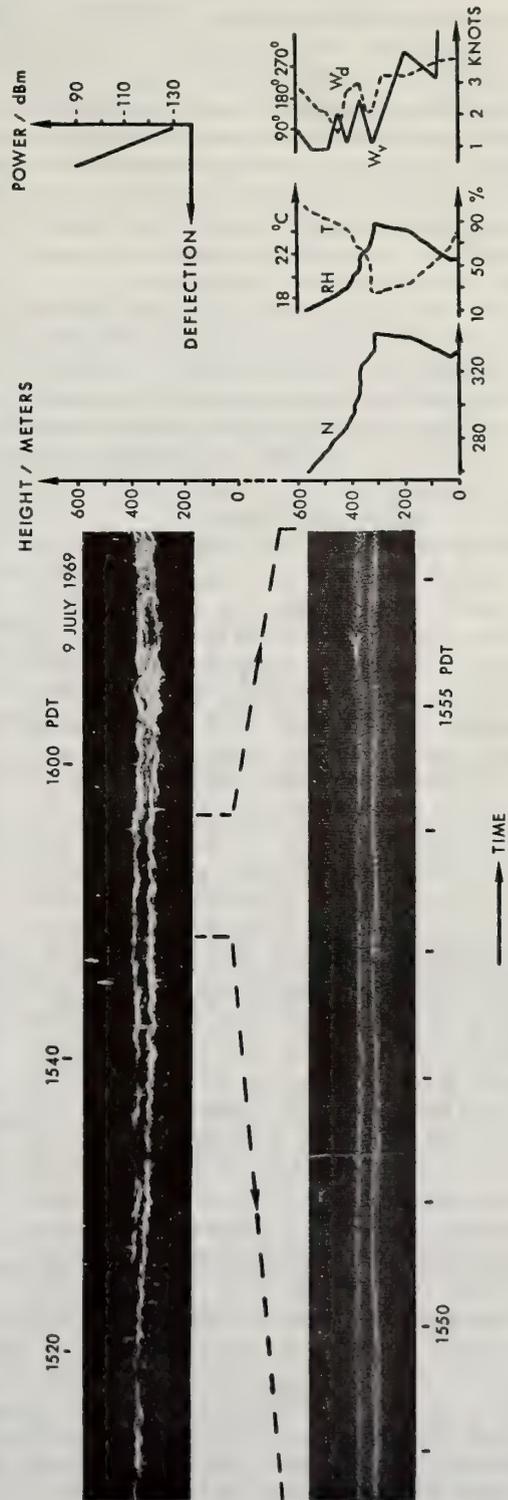


Fig. 5. Layer structure with unstable Kelvin-Helmholtz waves and associated meteorological profiles. The upper photograph is recorded with intensity modulation; the lower with A-I scope presentation. In both cases 16 sweeps are averaged. The power-deflection chart applies to the A-I scope presentation only. $N = (n - 1) 10^6$, where n is refractive index. T is temperature, RH is relative humidity, W_v is wind velocity, W_d is wind direction, and PDT is Pacific Daylight Time.

JUERGEN H. RICHTER

The upper record is the compressed intensity-modulated trace; the lower one is the time-expanded A-I scope trace for the period indicated. Again, there are two major undulating echo layers: one centered close to 310 meters, the other close to 390 meters. Both layers correspond identically to strata of sharpest N lapse, the lower is at the base of the major temperature inversion and the upper at a second, more stable interface within the inversion. Note also that these surfaces are marked by wind shear.

However, the dominant feature of these observations is the wavelike nature of both echo layers, with the lower one showing clear-cut evidence of breaking so that the peak of one wave seems to have overtaken the trough of the next. The periods of these waves are about 2 to 3.5 min. Their spatial wavelength is unknown. Their amplitudes, however, are only 10 to 14 meters peak-to-peak.

These 'breaking' waves are almost identical to those reported by Hicks and Angell [1968] and Hicks [1968]. They are also very similar to those reported by Woods [1968] at the thermocline in the Mediterranean Sea and by Thorpe [1968] in laboratory fluid experiments. Woods and Thorpe attribute the formation of these waves to Kelvin-Helmholtz instabilities that occur when the interfacial Richardson number becomes less than a critical value of $1/4$. Calculation of the Richardson numbers (R_i) in the layers in question shows $R_i = 9$ between 310 and 330 meters (the lower echo layer) and 0.9 between 370 and 395 meters (the upper echo layer). Both values are larger than critical, but they are undoubtedly overestimated because of the inadequacy of the balloon-measured winds when used for shear calculations in such thin strata. Thus, we can only use the winds to indicate that the layer of K-H (Kelvin-Helmholtz) waves is marked by both large thermal stability and large shear. The presence of the breaking K-H waves themselves is probably evidence of the existence of the critical R_i , although we would like to determine this directly. Our observations suggest, however, that associated meteorological measurements would have to be made with exceedingly high vertical resolution in order to evaluate the vertical gradients and R_i at the heights of the thin echo layers.

CONCLUSIONS

Although this is but a preliminary report, it is clear that the FM-CW radar described has an excellent capability for detecting the fine-scale refrac-

tivity inhomogeneities that occur at atmospheric interfaces showing either sharp refractivity gradients or changes in gradient. The high resolution of the radar has been demonstrated; the scatter layers often have a thickness of only 1 meter or less, although they may extend to 20 to 30 meters in thickness. The echo layers also frequently come in pairs spaced in height by some tens of meters. Previous radar observations made with much coarser-resolution equipment have not been capable of displaying the occurrence of multiple line layers. Moreover, the echo layers are commonly marked by K-H waves of 2- to 4-min periods, and peak-to-peak amplitudes of but 10 to 15 meters. The layer reflectivities measured with a 1-meter resolution may also be higher than previously observed by a factor of 10 to 100.

Acknowledgments. The author thanks Dr. E. E. Gossard for the encouragement of this work and valuable suggestions and D. W. Truax and W. K. Horner for their assistance. The author is particularly indebted to Prof. D. Atlas of the University of Chicago who provided stimulating discussions and pointed out some of the features indicative of unstable K-H waves.

REFERENCES

- Atlas, D. (1964), Advances in radar meteorology, *Advan. Geophys.*, 10, 317-478.
- Atlas, D., and K. R. Hardy (1966), Radar analysis of the clear atmosphere: Angels, in *Proceedings of the 15th General Assembly of URSI*, pp. 401-469, International Scientific Radio Union, Munich, Germany.
- Atlas, D., K. R. Hardy, and K. Naito (1966), Optimizing the radar detection of clear air turbulence, *J. Appl. Meteorol.*, 5, 450-460.
- Fehlhaber, L., and J. Grosskopf (1965), An investigation of the structure of the troposphere by means of a vertically pointed radar, *Nachricht. Z. Commun. J.*, No. 4, 185-190.
- Hardy, K. R., D. Atlas, and K. H. Glover (1966), Multi-wavelength backscatter from the clear atmosphere, *J. Geophys. Res.*, 71(6), 1537-1552.
- Hardy, K., and I. Katz (1969), Probing the clear atmosphere with high power, high resolution radars, *Proc. IEEE*, 57(4), 468-480.
- Hicks, J. J. (1968), Radar observations of gravitational waves in a clear atmosphere, in *Proceedings of the 13th Radar Meteorology Conference*, pp. 258-261, American Meteorological Society, Boston, Mass.
- Hicks, J. J., and J. K. Angell (1968), Radar observations of breaking gravitational waves in the visually clear atmosphere, *J. Appl. Meteorol.*, 7, 114-121.
- Saxton, J. A., J. A. Lane, R. W. Meadows, and P. A. Matthews (1964), Layer structure of the troposphere, *Proc. IEE, London*, 111(2), 275-283.
- Thorpe, S. A. (1968), A method of producing a shear flow in a stratified fluid, *J. Fluid Mech.*, 32, 693-704.
- Woods, J. D. (1968), Wave-induced shear instability in the summer thermocline, *J. Fluid Mech.*, 32, 791-800.

Chapter 21 PASSIVE SENSING OF NATURAL ACOUSTIC-GRAVITY WAVES AT THE EARTH'S SURFACE

T. M. Georges and J. M. Young
Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

A tutorial account is given of efforts to observe and interpret natural acoustic-gravity waves in the atmosphere. Emphasis is on how the waves are passively detected and analyzed, the kinds of wave phenomena that are actually observed, and what is now known (and what remains unknown) about the origins of the waves. Present and proposed joint observational programs with other sensing techniques promise new insight into the role of waves in atmospheric dynamics.

21.0 Introduction

Man's needs for ever-increasing quantities of environmental information have driven him to extend his senses across both the electromagnetic and acoustic spectrums, devising active as well as passive "remote sensing" abilities, the most recent of which are the subject of this lecture series. This lecture deals with efforts to extend our passive acoustic-sensing capabilities to frequencies far below the threshold of natural hearing, where wave periods are often measured in seconds, minutes, and even hours.

The kind and amount of environmental information to be extracted by "listening" to these waves depends on our understanding (1) which atmospheric or other processes generate the waves, (2) the nature of the waves and how they propagate through and interact with the atmosphere and (3) the relationship between the waves themselves and what our sensors measure.

In this lecture, we will concentrate on:

1. how natural acoustic-gravity waves are detected with ground-based sensors;
2. the kinds of wave motion that are actually observed;
3. what is known (and what remains unknown) about wave origins; and
4. what useful things such observations can tell us about atmospheric dynamics and other geophysical processes.

In assessing the current state of knowledge, we also identify some outstanding problems where future research should concentrate.

The basic properties of the waves themselves, particularly in the atmospheric boundary layer, have been discussed in the lectures of Hines and Hooke and will not be dwelt upon here, except to illustrate, with some specific examples, the effects of the real atmosphere on the propagation of waves from particular kinds of sources.

Those wishing to study the acoustic-gravity wave literature may find the identification of a few starting points helpful: Many of the theoretical and observational advances in the last decade have been motivated by Defense Department interests in the atmospheric waves launched by nuclear explosions. A recent unclassified review of this work is given by Donn and Shaw (1967). Lucid treatments of basic acoustic-gravity wave theory are given by Hines (1960) and Tolstoy (1963). The proceedings of a recent symposium on acoustic-gravity waves in the atmosphere (Georges, 1968c), contains 40 papers and ample references to previous work. The most comprehensive bibliographies of acoustic-gravity-wave generation and propagation seem to be those of Pierce and Posey (1970) and Thomas et al.,(1972).

21.1 Some Definitions

A few words about nomenclature for various portions of the atmospheric-wave spectrum seem warranted here; a consistent system has not yet become widely accepted, and some ambiguity exists in the atmospheric-wave literature.

The various sub-species of atmospheric waves are most consistently defined and classified according to the degree of simplification of the *mathematical model* describing them, i.e., according to the particular terms that are retained in the equations of motion. Such a scheme applies whether one is concerned with simple plane-wave solutions or the solutions of complex boundary-value problems.

Here, the term "atmospheric wave" refers to any and all wave motions permitted by the complete hydrodynamic equations applied to planetary atmospheres; for example, planetary or Rossby waves and

atmospheric tides, as well as ordinary sound waves.

Acoustic or sound waves may be defined as those waves satisfying the equations of motion in which air inertia is balanced only by pressure-gradient forces. Acoustic-gravity waves may be defined as the waves that satisfy the same equations as acoustic waves with only the force of gravity (and the resulting atmospheric density stratification) added. The name "atmospheric gravity waves" may be given to the wave solutions of the equations for incompressible motion which include gravity (or more correctly, solutions in which compressibility is retained only in the terms containing the gravitational force, thus permitting buoyancy forces, i.e., the "Boussinesq approximation").

Thus, atmospheric gravity waves and acoustic waves may be regarded as sub-classes of acoustic-gravity waves, which, in turn, are a sub-class of atmospheric waves.

Other means of wave classification are used: Hines (1960), for example, identifies "acoustic" and "internal gravity" waves as the two wave solutions of the linearized equations of motion in an isothermal atmosphere including gravity and compressibility. The two solutions represent distinct spectral sequences, separated by a spectral gap in which only evanescent waves exist. The word "internal" is used to distinguish waves with real vertical wave numbers from those having imaginary vertical wave numbers, e.g., surface or evanescent waves. The name "acoustic-gravity waves" has sometimes been used to refer only to waves with frequencies higher than the Brunt frequency (the frequency of buoyancy oscillations), but only up to frequencies where gravity becomes "negligible." Acoustic-gravity waves have also been defined as waves in which neither the force of gravity nor compressibility is negligible. This definition is based on the properties of real waves, not a model, but is too vague to be useful in any quantitative sense.

Still other schemes are based on various *observed* wave properties; audible sound waves are sometimes called "acoustic," while those with higher frequencies are called "ultrasonic" and those with lower frequencies are called "infrasonic" or "infrasound." Some observers have called atmospheric waves that travel at speeds slower than the classical speed of sound, "subsonic" waves (cf. Bowman, 1968). A more specific classification of the various species of observed acoustic-gravity waves, according to their measurable parameters, is attempted in (21.4).

Clearly, all such models and distinctions are only artifices constructed for our own convenience; nature, of course, never "neglects" any forces and constructs no boundaries between wave types. The particular model (and hence the wave name) that one chooses to represent a given physical situation is usually the simplest one that he thinks contains the essential features of reality. But when we give observed physical phenomena the same names as our simplified models of them, we often forget that, in a sense, we are forcing nature to fit our conceptual molds.

The wave nomenclature in this lecture is based on particular forms of the hydrodynamic equations; thus, "acoustic-gravity waves" represents wave solutions of the equations of motion including gravity and compressibility. The real atmospheric waves of concern here are those whose properties are adequately modeled by these solutions. In practice, this restriction limits our scope to atmospheric waves with periods shorter than 2 or 3 hours.

21.2 Sensing Techniques

Acoustic-gravity wave sensors have historically been of the microbarograph type. Most modern microbarographs record changes in atmospheric pressure (p) relative to the pressure in a "reference volume," a stable pressure chamber which is normally vented to the atmosphere through a "slow leak." This leak provides a high-pass filtering action, which causes the instrument output to represent dp/dt at frequencies below a "rolloff frequency," determined by the leak's time constant, and p above that frequency. Instruments presently used by NOAA and a number of cooperating groups (Matheson, 1964; Cook, 1962; Chrzanowski, et al., 1961) are adjusted for a time constant of about 7.5 sec, giving a 3 dB response point at 50-sec period. Another result of high-pass filtering is to "pre-whiten," or compensate the background atmospheric pressure fluctuations, which, as we will see, increase in intensity at longer periods. In many operational systems, it is desired to record $p(t)$ rather than dp/dt , so that the microphone output is filtered in such a way that the entire system response is relatively flat over some desired band.

Microbarographs are usually connected to "noise-reducing pipes," typically about 300 m long with small openings every 5 feet or so which serve to average the sensor's response over a specified spatial area, and thus filter out smaller-scale pressure fluctuations, which are presently regarded as "noise." Such filters also reduce the sensor response (above about 1 Hz) to ordinary acoustic noises. (To be sure, the definitions

of "signal" and "noise" are arbitrary and depend on where one focuses his attention at any given time; thus, one designs spatial and temporal filters to pass certain fluctuations and to reject those not of immediate interest, regarding them as "noise.") A simplified schematic diagram of a modern microbarograph system is shown in (F21.1). Instrumentation details may be found in the technical report by Matheson (1964) and the paper by Cook and Bedard (1972).

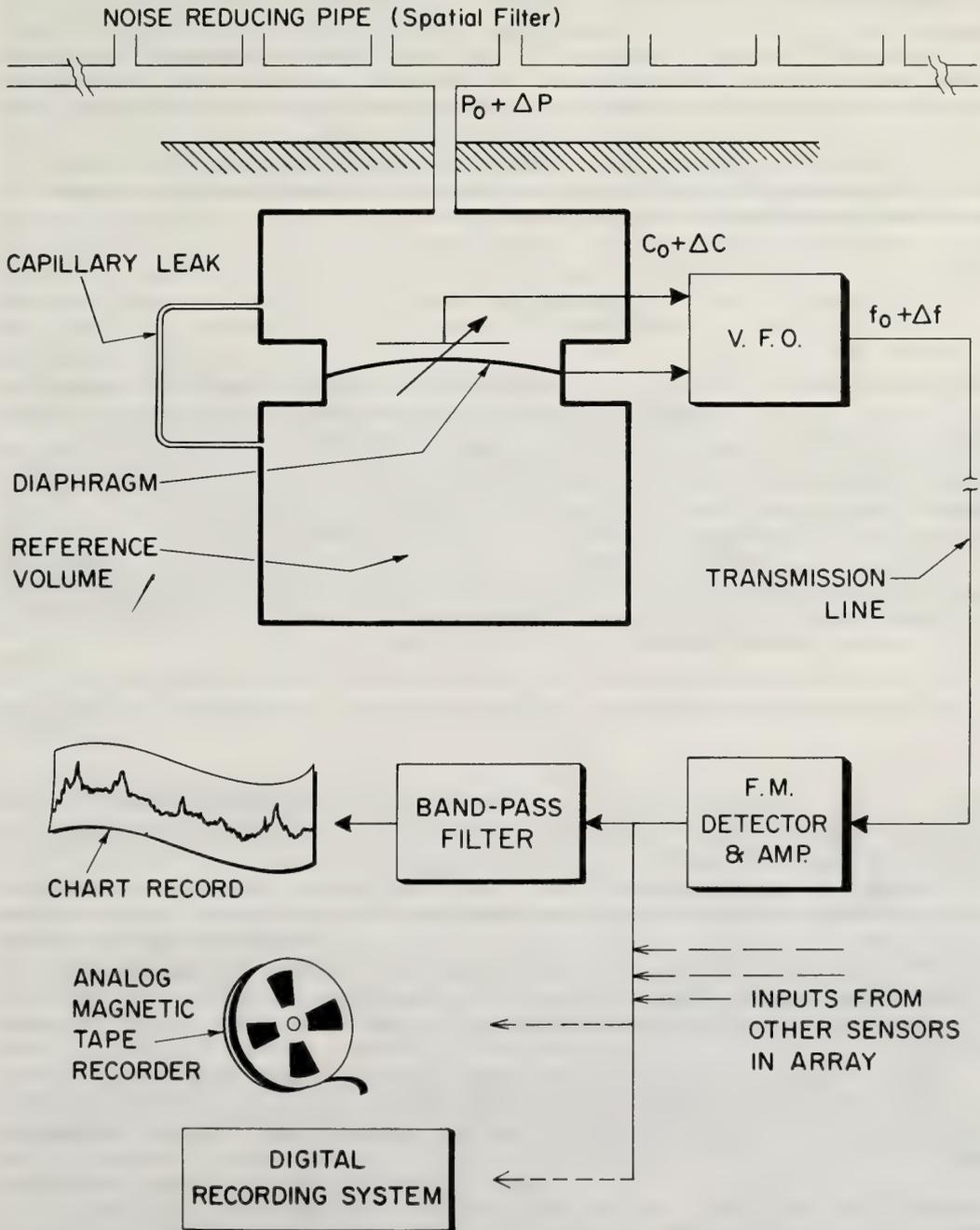


Figure 21.1. A system for recording microbarometric pressure fluctuations.

One might legitimately ask why pressure is the quantity most wave sensors have been designed to sense, when fluctuations in temperature, air density and air velocity also accompany the passage of atmospheric waves. The reason seems to be partly historical and basically a question of instrument sensitivity. Pressure sensors have now been developed that can measure pressure changes of about 1 part in 10^8 , or $0.01 \mu\text{b}$ at standard pressure. An acoustic wave with pressure fluctuations of 1 part in 10^8 is accompanied by temperature and density fluctuations of the same order of magnitude. To match microbarograph sensitivity in terms of detectable sound waves, a temperature sensor would have to measure changes of the order of $8 \times 10^{-7} \text{ }^\circ\text{K}$; a density-measuring sensor would have to measure fluctuations of about 1 part in 10^8 ; and a wind sensor would have to measure speeds of about $2 \times 10^{-6} \text{ m/s}$, all well beyond the range of today's field instruments. Microbarographs thus appear to be the most sensitive instruments presently available for passive sensing of the higher-frequency portion of the acoustic-gravity wave spectrum (wave periods between about 1 second and a few minutes).

At high (acoustic) frequencies, Hines' "polarization relations" (Chapter 5, 5:17-5:20) show that the ratios of wave-associated fluctuations in pressure, temperature, density and velocity remain constant. The general wave "impedance relation" shows, for example, that

$$\frac{p'}{u'} = \rho_0 \frac{\omega}{k} \cong \rho_0 C \quad (21:1)$$

where ω/k is the wave phase speed, and C is the speed of sound. This fact allowed us to make the above estimates of wave-associated density, temperature and velocity fluctuations without specifying the wave frequency or wave vector magnitude or direction. At longer wave periods, however, gravitational anisotropy strongly influences wave propagation, so that ω/k , the phase velocity, no longer equals C , but depends in a complicated way on wave frequency and wave vector magnitude and direction. (See Chapter 5 for the dispersion relation for acoustic-gravity waves in an isothermal atmosphere.) Therefore, the wave-variable ratios exhibit the same complicated behavior. In general, gravity wave phase speeds are smaller than acoustic wave speeds, and so are the wave-associated pressure fluctuations relative to velocity fluctuations. In a particular asymptotic limit corresponding to long-period gravity waves whose vertical wavelengths are much less than an atmospheric scale height, $\omega \cong \omega_g k_x/k_z$, so that

$$\frac{p'}{u'_x} = \frac{\rho_0 \omega_g}{2\pi} \lambda_z \quad (21:2)$$

where ω_g is the Brunt-Väisälä frequency. This shows that pressure fluctuations (relative to velocity fluctuations) decrease as vertical wavelength becomes smaller. For a given horizontal wavelength (or horizontal phase speed), decreasing λ_z corresponds to lengthening wave periods.

Therefore, for some internal gravity waves, it would seem profitable to explore the capabilities of temperature, density or air-speed sensors in the observation of natural atmospheric waves. Air-velocity sensors of sufficient sensitivity and frequency selectivity could provide information on wave travel direction as well. It should be realized, of course, that the frequency range in which auxiliary sensors might provide increased wave-observation capabilities is determined not only by instrument sensitivity but also by the levels of extraneous fluctuations in the quantities being measured. These "noise levels" are determined experimentally.

Earlier in this section, we said that microbarographs measure changes in "atmospheric pressure." A firm concept of what "atmospheric pressure" means comes to mind only for the case of a static atmosphere, where it is normally defined as the weight of a vertical column of air on a unit horizontal surface:

$$p_0 = \int_0^{\infty} \rho_0 g dz \quad , \quad (21:3)$$

where $\rho_0(z)$ is the atmospheric density and g is the acceleration due to gravity. When vertical accelerations are present, i.e., in the nonstatic case, a term $-\rho_0(du_z/dt)$ must be added to the integrand, where $u_z(z)$ is the vertical air velocity. Thus a pressure sensor can be regarded as having the ability to sense (remotely, if you will) wave-associated air motions (with a phase shift of $\pi/2$ and weighted by $\rho_0(z)$) throughout the atmosphere above it. This is not to say that, in the case of atmospheric waves, for example, a true pressure sensor responds to anything except the local wave-associated perturbation pressure. An inspection of the development of the theory of plane waves in an isothermal atmosphere will show that integrated wave density fluctuations are completely accounted for in the formula for wave pressure. In the case of the other types of atmospheric motions, however, (such as convective or "turbulent" motions) it must be remembered that pressure sensors respond not only to local air motions, but integrate density and velocity fluctuations throughout the columns above them. Thus the sensors are not constrained to respond only to scales of motion that exist at the earth's surface.

The problem of signal detection in a background of pressure fluctuations, or "wind noise," caused by turbulent wind fluctuations near the sensor has been considered by Morse and Ingard (1968, 1.11). They discuss the theory of acoustic transducer response as it depends on orientation with respect to the mean fluid flow. For example, if the sensor moves with the mean flow, or if its aperture plane is approximately parallel to the mean flow, then the turbulent velocity fluctuations u_t' cause pressure fluctuations ("noise") of order $\frac{1}{2}\rho_0 u_t'^2$. If, on the other hand, the sensor aperture is located normal to the mean flow, i.e., at a stagnation point, the pressure fluctuations due to u_t' are enhanced by a factor of order $2U_0/u_t'$, where U_0 is the velocity of the mean flow. Pressure sensors are thus least influenced by turbulence when their aperture planes are oriented parallel to the mean flow. These considerations strongly influence the actual design of microbarographs and their spatial filtering arrays.

Because microbarographs are usually located at the earth's surface, one has also to understand possible boundary-layer and other surface effects on pressure-sensor response. It is clear, for example, that no vertical air motions can exist at the surface. This fact alone requires that plane-wave-associated pressure perturbations (due to an incident and a reflected wave) double, compared to the magnitude of an incident plane wave alone. But, near the surface, eddy viscosity plays a complicated role in determining the wave-associated horizontal air motions (and possibly the existence of boundary waves), and hence the pressure fluctuations there. The details of this role, especially as it depends on surface roughness, are not yet as well understood as they should be if we are to understand exactly what surface-pressure sensors measure.

As Hooke mentioned in his lecture (Chapter 7), one expects the boundary layer to select certain wave modes that propagate readily in its temperature and wind structure, and these modes should be preferentially observed.

21.3 Identifying Atmospheric Wave Motions

Since we are interested in observing atmospheric waves, we first have to know how to identify "waves" in the presence of "non-waves," i.e., pressure fluctuations caused by non-propagating atmospheric motions, for example, advected "turbulence." Clearly, such distinctions must rely on spatial as well as temporal pressure sampling, and criteria for assessing the degree of confidence one places on wave identification must be established.

In systems designed to measure pressure-wave properties, microbarographs are usually positioned on the ground to form a two-dimensional spatial array, in which individual sensors may be located from hundreds of meters to tens of kilometers apart, depending on the wavelengths and spatial coherence of the waves to be detected. (Although pressure sensors located above the earth's surface, to form a three-dimensional array, would provide valuable additional wave information, the requirement of very high supporting structures and mechanical problems associated with wind-noise reduction have retarded their development.) The relative times of arrival of an identifiable pressure event at multiple sensors can then be used to determine the speed and direction of travel of that event past the array, through the use of visual superposition or more sophisticated correlation techniques.

Correlation of pressure signatures from spaced sensors does not, of course, automatically distinguish "wave" from "turbulence" or "noise." If relative arrival times of an event indicate a travel speed greater than the local sound speed, it is unlikely that the event could be anything but an acoustic wave, and such events are usually interpreted in terms of the "trace" of an acoustic wave arriving obliquely at the earth's

surface. But, in the case of more slowly moving events, the distinction between “wave” and “noise” is readily made only when spatial coherence is so high and the waveform so periodic as to make identification at successive sensors unmistakable. As spatial coherence deteriorates, however, one is soon faced with conceptual and semantic problems related to the definitions of “wave” and “turbulence,” and ultimately, with the task of imposing arbitrary criteria for distinguishing the two. Analytical tools that permit the choice of numerical criteria are provided by cross-correlation techniques, including cross-power spectral analysis.

Cross-correlation analysis (Briggs, et al., 1950) is perhaps the most commonly used technique for studying motion fields sampled at spaced sensors. The cross-correlation (sometimes called the cross-covariance) function for two time series $x(t)$ and $y(t)$ is

$$\phi_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) y(t + \tau) dt \quad (21:4)$$

Here the two functions are multiplied and averaged over a sample length $2T$, while the time shift τ is varied. The higher the degree of correspondence of the two functions, the larger their averaged product ϕ_{xy} will be. If $x(t)$ and $y(t)$ are data from spaced sensors, the value of τ for which $\phi_{xy}(\tau)$ maximizes is taken as the time displacement, or the signal travel time between the two sensors. The value of τ for each pair of sensors in a network can be interpreted in terms of a trace velocity of a signal along the line joining the pair. To the degree that, when each trace velocity is plotted as a vector from a common origin, the tips of all the vectors align, the analysis is said to yield a single-plane-wave solution with a high degree of confidence (see F21.2).

However, Gossard (1967) has shown that cross-correlation analysis alone can yield grossly misleading information about the patterns of movement when records contain several frequency components of differing velocities. The various frequency components may be separated by taking the Fourier transform of the cross-correlation function, which is complex if the cross-correlation function is asymmetrical about $\tau = 0$. This new function is called the cross-power spectral density, or simply the cross spectrum:

$$\phi_{xy}(\omega) \equiv C_{xy}(\omega) + iQ_{xy}(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \phi_{xy}(\tau) \exp(i\omega\tau) d\tau, \quad (21:5)$$

where $C_{xy}(\omega)$ and $Q_{xy}(\omega)$ represent, respectively, the in-phase (co-spectrum) and quadrature components of the complex spectrum of $\phi_{xy}(\tau)$. A quantity called “coherence” is simply the normalized magnitude of the squared cross-spectrum:

$$\text{coh}(\omega) \equiv \frac{C^2 + Q^2}{E_x E_y} \quad (21:6)$$

where E_x and E_y are the power spectra of the x and y functions above. The phase angle of the cross-spectrum yields the time lag τ_0 between the two records, as a function of ω :

$$\omega\tau_0 = \text{Tan}^{-1} \frac{Q_{xy}(\omega)}{C_{xy}(\omega)}, \quad (21:7)$$

permitting the computation of velocity and angle of arrival *as a function of frequency*.

Put as simply as possible, the squared magnitude of the cross-spectrum (the coherence) describes the accuracy with which a given set of multisensor data can be represented by the superposition of a number of pure plane waves; the phase angle of the cross-spectrum permits the computation of velocity and angle of arrival of each spectral component.

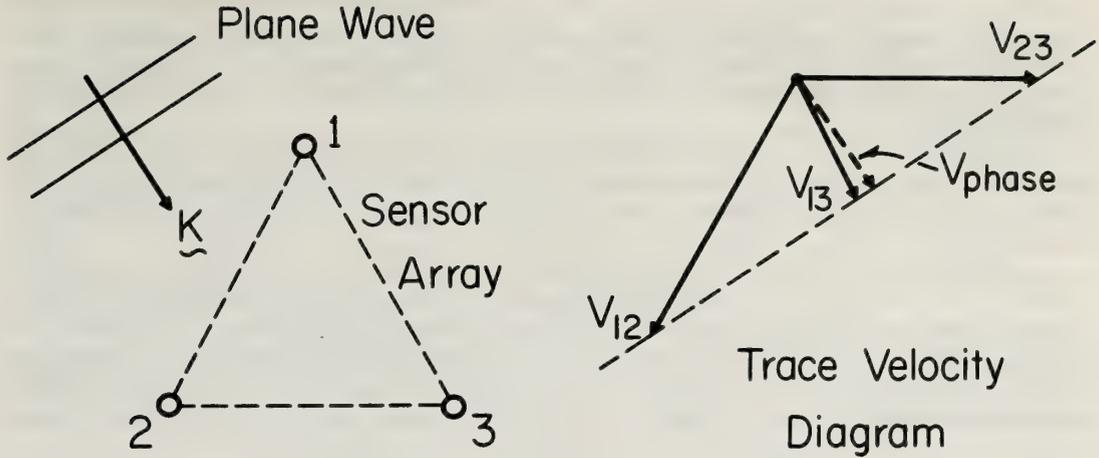


Figure 21.2. Construction of the wave phase velocity vector from trace-velocity vectors measured across sensor pairs. The degree to which the tips of the trace-velocity vectors align indicates the degree of confidence one places on a plane-wave fit to the data.

Analyzing the cross spectra from sensor pairs can, in principle, help distinguish fluctuations caused by superimposed wave motions from those due to drifting or wind-carried eddies (Gossard, 1967). If the fluctuations are due entirely to wave motions, one would expect to observe different frequency components in the records moving with different speeds and in different directions; if eddies are “frozen in” the mean flow, on the other hand, one expects all spectral components to have the same velocity, i.e., that of the mean flow. In the case of atmospheric gravity waves, however, wind speeds may often approach (or even exceed) wave propagation speeds, so that observed wave propagation speeds and directions may be closely related to local wind conditions, and therefore, may be interpreted as advected turbulence. It is not certain whether such waves can be (or even should be) distinguished from other fluctuations; it may be possible to utilize the theoretically predicted phase relationships among pressure, density, temperature and wave-associated air velocity to identify them.

Criteria for judging the significance levels of coherence in terms of sampling parameters are beyond the scope of this introductory discussion but are covered by Gossard and Noonkester (1967). Gossard (1969) and Gossard and Sailors (1970) also show that the behavior of coherence as sensor spacing varies can be used to deduce (a) temporal variability of the dispersive properties of the source or the propagation medium, in the case of stations separated in the direction of propagation, and (b) effective beamwidths or angular dispersion of arriving waves, in the case of stations separated perpendicular to the direction of wave travel.

The coherence properties of spaced-sensor data are readily determined by digital techniques that are presumably familiar, and such techniques are commonly employed in the analysis of atmospheric-wave data. However, several geoacoustic observatories have for some years employed an ingenious analog-correlating device, developed by the NOAA Geoacoustics Group, which automatically performs cross-correlation operations on data in analog form, and displays cross-correlation coefficients as a function of azimuth and speed of travel (Brown, 1963). The device records a short time interval of multisensor data in analog form on a multitrack magnetic tape loop, then systematically varies the relative leads and lags of movable tape heads in a pattern programmable for any sensor configuration. Tape-head outputs are combined to represent a generalized correlation coefficient of the form (Brown, 1963):

$$R_g = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n R_{ij} \quad (21:8)$$

where R_{ij} is a normalized form of the cross-covariance given by (21:4) and n is the number of sensors. One form of the output of this correlator is shown in (F21.3) for waves called "mountain associated." Wave motions are identified by the distinctive appearance of the correlator output (top of the figure) when it responds to a traveling wave, compared to the response to incoherent fluctuations (bottom), which are thus regarded as "noise."

When pressure vs. time records of traveling waves are displayed (as they are later in this paper), the multisensor outputs are normally superimposed with time delays of maximum correlation. It is this correlation, not any distinctive "signature" on an individual pressure record, that permits identification of a traveling wave.

An alternative to the spaced-sensor-correlation-analysis method for determining wave properties is currently being explored. It involves modifying present pressure sensors to measure *spatial pressure gradients*, as well as dp/dt . Point measurements of both spatial and temporal pressure gradients could, in principle, provide a direct indication of the direction of the three-dimensional wave vector and wave phase velocity. Such an instrument should be compact enough to permit installation on towers, to measure vertical wavenumber.

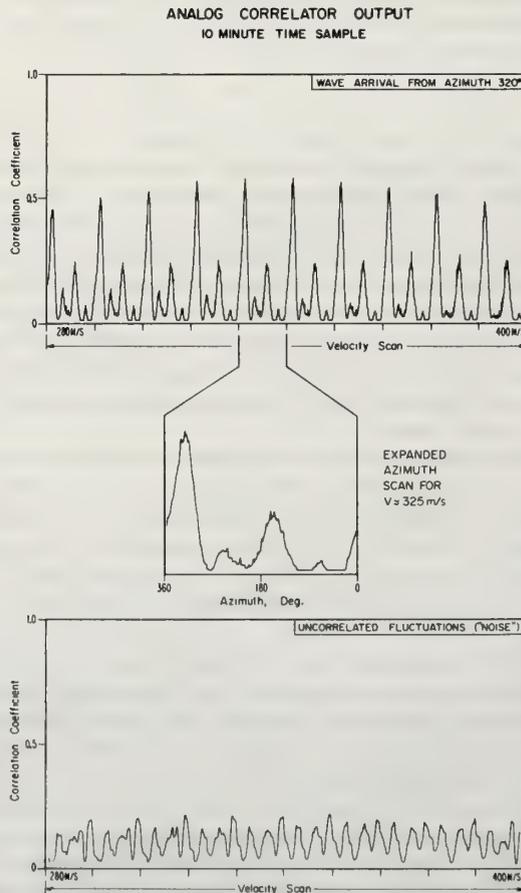


Figure 21.3. Examples of the output of an analog correlator for a wave classified as "mountain-associated" (top), and for uncorrelated "noise" (bottom). Azimuth scans are nested within the velocity scan, which samples a moving 10-min. time window.

21.4 The Atmospheric-Wave Environment

When the output of a single microbarograph is examined, one finds that, under a wide variety of conditions, the power spectral density decreases approximately as the inverse square of the frequency, over the frequency range 10^4 to 10^2 Hz (cf. Kimball and Lemon, 1970, and references therein). Since time differentiation is equivalent to multiplication by frequency, an f^2 dependence of the power spectrum (p^2) implies a flat or uniform spectrum (per unit bandwidth) of dp/dt . This is just the quantity that microbarographs measure for wave frequencies below the rolloff frequency determined by the capillary leak. This is in rough agreement with an $f^{7/3}$ dependence for pressure fluctuations predicted (Lumley and Panofsky, 1964) for an "inertial subrange" in which there is a transfer or "cascade" of energy to smaller and smaller eddies, but no production or dissipation in that portion of the spectrum. In many experimental spectra it is not possible to distinguish between f^2 and $f^{7/3}$ behavior.

Because of the apparent constancy of dp/dt across the spectrum of interest, we can characterize the atmospheric background pressure fluctuations by values of dp/dt (per 1 Hz band) representing upper ("noisy") and lower ("quiet") bounds, without specifying frequency. Interpreting the measurements cited earlier, these numbers would be approximately 0.1 to 10 $\mu\text{bar}/\text{sec}/\text{Hz}$; a level of 1 $\mu\text{bar}/\text{sec}/\text{Hz}$ thus serves as a "rule-of-thumb" average background level. (A microbar is 10^{-6} atmospheres and equals 1 dyne/cm^2 .)

At the present time, it is not known what fraction of these fluctuations are caused by "local" atmospheric motions, i.e., "turbulence," and what fraction represents the composite of many atmospheric waves from distant sources. Indeed, a meaningful distinction between the results of the two processes may not be possible; i.e., both may be viewed as random processes. Whenever the number of superimposed waves is small, or only a few dominate, the analysis techniques described in the previous section can be expected to reveal their presence.

A few wavelike events have sufficiently distinctive properties to permit them to be identified as a particular class of wave without reference to other simultaneous geophysical data. Such events naturally receive the most attention, and are most readily associated with other geophysical events. The reviews by Cook (1962), and Cook and Young (1962) show microbarograph records representing many different but distinctive wave signatures.

For example, "microbaroms" are nearly monochromatic fluctuations, usually of about 0.5 μbar amplitude. Wave period is in the 5-sec range, and they often last for several hours. Spatial coherence is poor; correlation transverse to the direction of travel usually falls to zero in about two wavelengths, and propagation speeds are nearly sonic. Figure (21.4) illustrates composite pressure traces of a microbarom

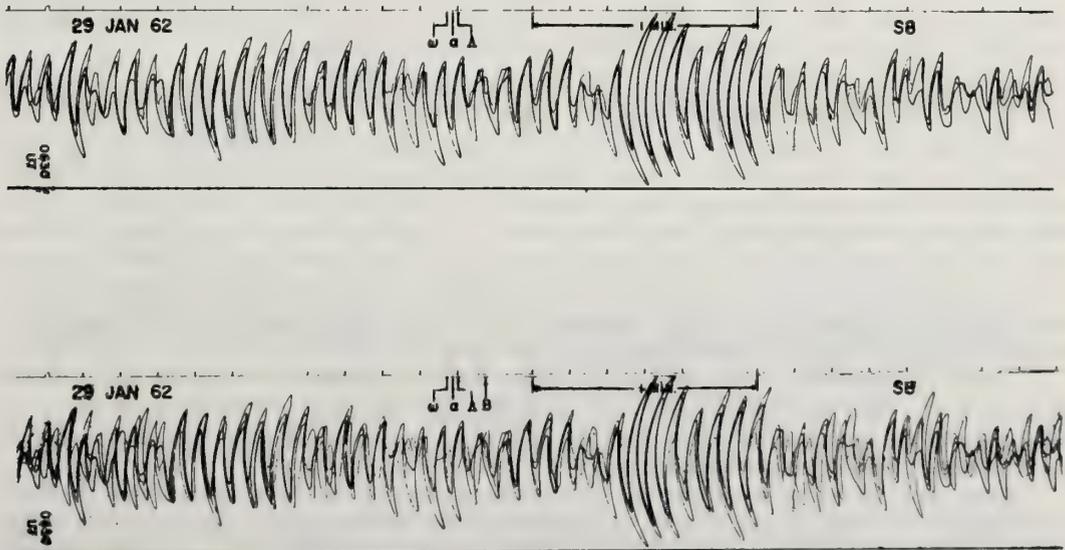


Figure 21.4. Examples of pressure waves known as "microbaroms". Good frequency coherence and low spatial coherence characterize these waves. The upper record shows the superimposed pressure traces from three microphones spaced about 1 km apart, while in the bottom record, a fourth trace from another microphone about 3.5 km away is added.

event. Past research (Benioff and Gutenberg, 1939; Saxer, 1945, 1953) has established a close relationship between microbarom activity and sea-wave height; but although generation processes related to sea-wave motion have been suggested, none is yet generally accepted.

Another distinctive signature of a natural event is shown in (F21.5). The waves are sound waves radiated locally by seismic waves passing the microbarograph sites; in the case illustrated, the epicenter was in Montana and the microbarographs were located near Washington, D. C. (Cook and Young, 1962). The distinctive feature of the waves is a large horizontal trace speed ranging from 6.0 to 3.8 km/sec for the several wave groups. Acoustic waves traveling nearly vertically correspond to those launched most efficiently by the shear and Rayleigh waves traveling along the earth's crust.

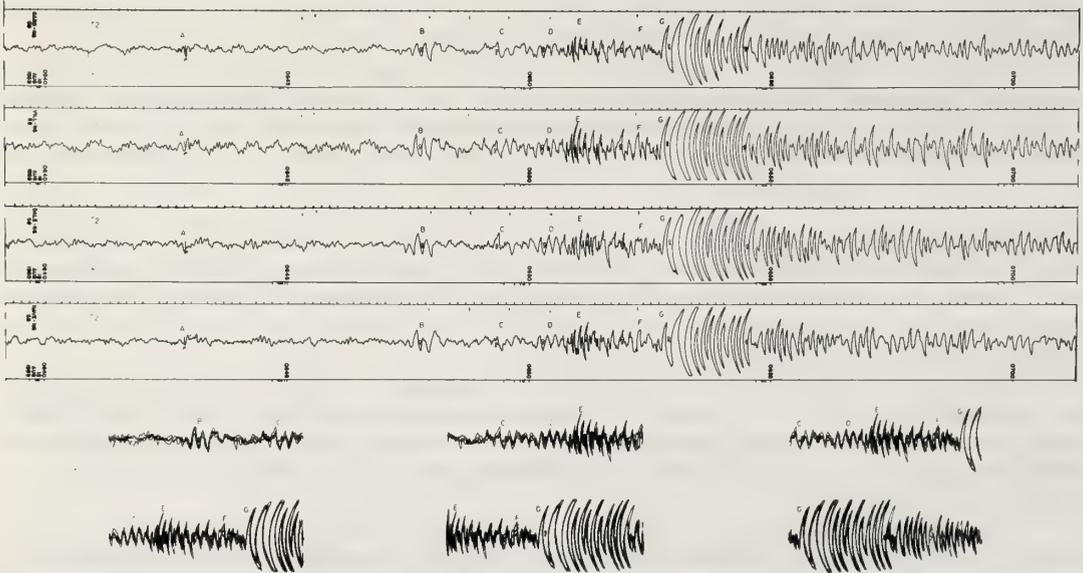


Figure 21.5. Pressure traces from four microphones separated by about 7 km at Washington, D.C., following an earthquake in Montana. The arrival of several wave groups with varying trace velocities characterizes these waves and is illustrated by superposed sections at the bottom. These waves are generated in the vicinity of the microphones by coupling with seismic waves.

A third wave type can often be identified without reference to other information; it comes from the azimuth quadrant centered on north, to the east of north before midnight and to the west of north after midnight. They usually travel at trace speeds greater than 400 m/s, but not as great as the earthquake-associated waves. It has long been known that these waves tend to occur when magnetic activity indices are high (Chrzanowski, et al., 1961); more recently, a theory has been advanced relating the waves to the bow shocks of auroral forms in supersonic equatorward motion (Wilson, 1969). Figure (21.6) shows microbarograph traces of an auroral "event" observed at Washington, D.C. Midlatitude observations of these events tend to be greatly "smoothed" by propagation effects, whereas the signatures of the same events observed near the source tend to be more impulsive in appearance.

Characteristic pressure fluctuations from volcanic explosions (Goerke, et al., 1965) and meteorite impacts (Jordan, 1967) have been recorded on microbarographs thousands of miles from the source. At least five instances have been recorded of pressure pulses being received following the sighting of prominent (and sometimes explosive) meteor trails at night.

Other types of clearly wavelike events are commonly observed, but their characteristics are either not sufficiently consistent or not sufficiently distinctive to permit unambiguous association with certain geophysical phenomena without the aid of supplementary information. However, considerable evidence exists, on the basis of supplementary observation during specific wave events, that acoustic-gravity waves radiate from severe convective storms, that others are apparently related to the proximity of the jet stream (Goerke and Woodward, 1966; Herron and Tolstoy, 1969) or the passage of weather fronts (Donn, et al.,

1954). Still others appear to emanate from certain relatively constant and localized geographic locations; Goerke (personal communication) and Larson (1971) have suggested an association with topographical features, for example, mountain ranges, and have called such waves "mountain-associated" although some of the events arrive from directions where no mountain ranges can be found. These waves have periods from 10 to 100 sec (Larson, 1971) and maintain a relatively constant bearing from hours to days. Most common trace velocities are between 300 and 400 m/s, and the direction of the apparent source often veers toward the east. Observations in North America report that the most common apparent origin of these waves is the coastal region of British Columbia, Canada, with a seasonal occurrence peak in the winter.

Table (21.1) is an attempt to characterize wave properties according to the several wave types that are presently distinguished. The name identifying each wave type corresponds to current usage, even though (particularly in the case of the "mountain-associated" waves) the association with the indicated mechanism may be quite tenuous. These names should, therefore, be regarded as tentative. The order of the wave type in the table (from left to right) is intended to indicate our assessment of the degree of certainty of the association and the degree of development of an adequate generation model.

As space does not permit detailed discussions of the phenomenology of each of the many species in the atmospheric-wave "zoo" that have so far been classified, our attention will focus on one particular class of waves now being studied, but by no means well understood, in an attempt to illustrate how such an analysis proceeds toward the formulation of a plausible physical model.

21.5 Acoustic-Gravity Waves From Convective Storms: A Case Study

Substantial evidence now exists, though mostly in unpublished form, that some severe convective storms radiate atmospheric waves that are often detectable thousands of kilometers away. The wave signatures themselves do not appear to be sufficiently distinctive or unique to permit unambiguous identification without supplementary storm information; in some cases, resemblance to other classes of waves is strong and identification is more difficult.

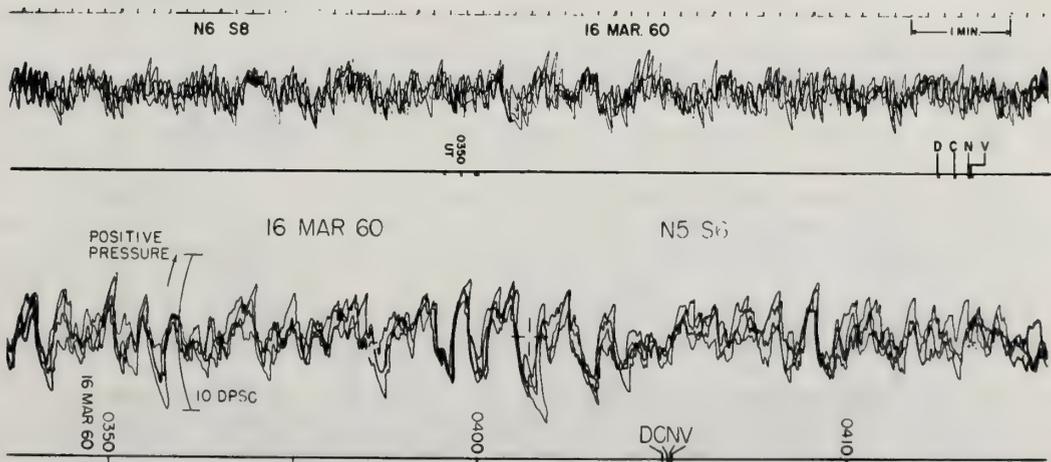


Figure 21.6. Waves observed by a 7 km array of 4 microphones at Washington, D.C., and believed to originate in the aurora borealis during magnetic disturbances. The upper trace represents a passband with half-power points at periods of 1 and 22 sec, while the bottom trace displays the same signal through a 10-60 sec. passband.

Table 21.1 Observed Characteristics of Natural Acoustic-Gravity Waves

OBSERVED CHARACTERISTICS	WAVE TYPE		VOLCANO and METEOR IMPACT	"AURORAL"	"MICROBAROMS"	"FRONTAL PASSAGE"	"SEVERE WEATHER"	"JET STREAM"	"MOUNTAIN ASSOCIATED"	"BOUNDARY-LAYER GRAVITY WAVES"
	EARTHQUAKES	IMPULSIVE								
NATURE OF SOURCE	via		EXPLOSIVE	SERIES OF IMPULSES (Bow Shock?)	CONTINUOUS (Air-Sea-Wave interaction?)	PRESSURE JUMP & BUOYANCY OSCILLATIONS?	UNKNOWN: 1. Penetrative Convection? 2. Lightning? 3. Turbulence?	UNKNOWN (Shear Instabilities?)	OROGRAPHIC?	AERODYNAMIC: SHEAR INSTABILITIES?
	Seismic	Air Wave								
WAVE PERIOD (Seconds)	12-25 (0.5-80)*	12-25 (1-30)	1-300 (0.5-1500)	10-400	5-7 (2-8)	180-1500 (180-1800)	12-60 (6-300)	(240-7200)	20-70 (10-120)	300-900 (250-1500)
WAVE AMPLITUDE (μbar 0-peak)	1-10 (0.1-20)	1-2 (0.1-20)	up to 5000	1-5 (0.2-20)	0.1-1 (0.1-5)	(50-2000)	0.5-3 (0.1-20)	50-200 (10-1000)	0.5-3 (0.1-10)	25-200 (10-1000)
DURATION	up to 1hr. Depends mainly on distance and dispersive properties of media.			Seconds to hours	Hours to Days (Noise Limited)	Minutes to hours	Minutes to Hours	Hours to Days	Hours to Days	Up to 1 Hr.
HORIZONTAL TRACE SPEED (m/sec)	3k-15k	340 (320-360)	320-500	400-1200	340	2-20	330-360 (330-450)	15-75	340 (330-360)	5-20 (2-20?)
DIRECTION OF TRAVEL	Vertical	from Epicenter ±6°	from Source	from intersection of Midnight Meridian with Auroral Zone	from certain Oceanic "Centers of Activity," e.g., North Atlantic Gulf of Alaska	Direction of Frontal Travel	from Centers of Strong Convective Activity; Mainly Midwest	Flow Direction	Coastal/Mountain locations; e.g., West Coast of Canada, Argentina/Chile.	Mean Wind
TRANSVERSE COHERENCE DISTANCE (to zero correlation)	not measured		~ 14 λ	Large at Low Latitudes; 10° Azimuth & Elevation Fluctuations	~ 2 λ	not measured	~ 14 λ	not measured	~ 14 λ	1.35 λ
DISTINGUISHING PROPERTIES	Seismic Dispersion	Long Periods precede Short Periods	Long Periods precede Short Periods	High Speeds; Short Impulse precedes Long Periods; Observed at Magnetic Latitudes >30°	Nearly Monochromatic; Beaded Wave Envelope; Low Coherence	Large Amplitude; Often Wavelike	Azimuth Change with Time; Spring / Summer Peak	Low Phase Velocities	Fixed Azimuth for up to Days; Local Winter Occurrence	Probably Exist down to Smaller Scales than Presently Studied; Usually Advection by Mean Wind.

* Figures indicate nominal values; those in parentheses indicate extremes.

Figure (21.7) illustrates the spectral variety of waves believed to arise from severe storms. This figure also illustrates, through the superimposed traces, the degree of coherence among signals received at sensors spaced about 7 km apart.

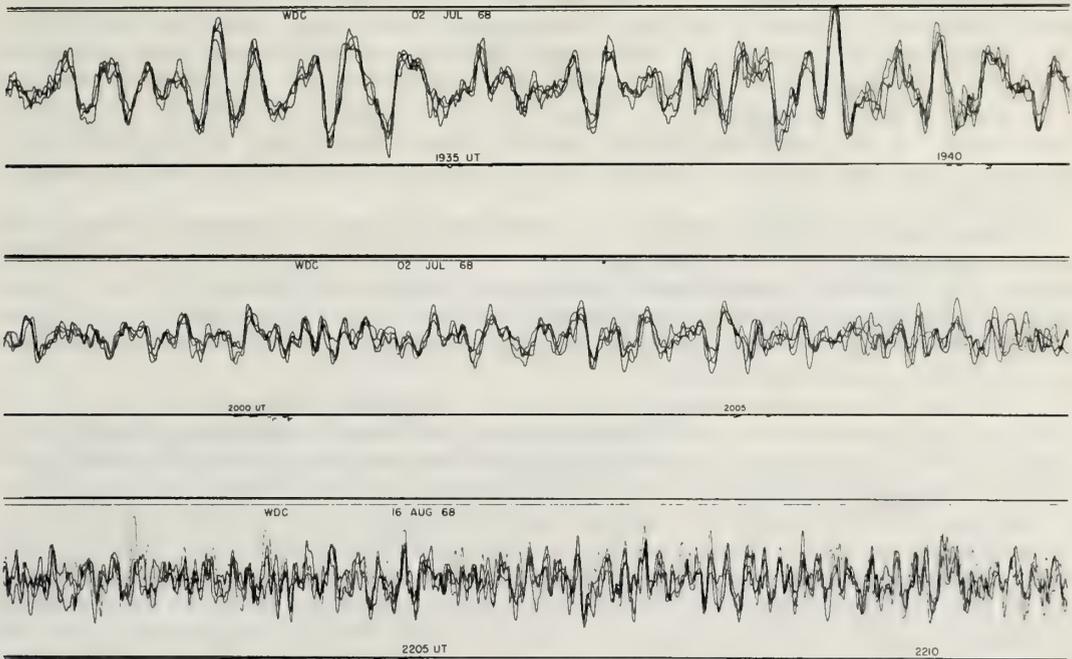


Figure 21.7. Three groups of waves observed by a 7 km array of 4 microphones at Washington, D.C., and believed to arise in severe convective storms within 100 km of the microphones.

Observed properties of wavelike events that appear to be severe-storm related may be summarized as follows:

1. Duration is typically several hours, i.e., much longer than the lifetime of a single convection cell. Occasionally very short-duration events are recorded, but their apparent brevity may result from masking by background noise.
2. The amplitude distribution of several hundred events recorded at Boulder has a median at $0.6 \mu\text{bar}$ and 10- and 90-percentile values of 1.25 and $0.3 \mu\text{bar}$ (Goerke, private communication).
3. Spatial coherence is usually sufficient to permit azimuth-of-arrival determinations to within about three degrees. The signals are apparently generated in regions whose dimensions are much smaller than those of storm groupings indicated on weather radars. Such groupings typically subtend 20 degrees or more at the sensors.
4. Azimuth of arrival often shifts abruptly during an event, indicating that the instruments may be dominated successively by different emitting centers within a storm complex (Goerke and Woodward, 1966).
5. During observations of waves attributed to nearby storms (within about 100 km), horizontal trace speeds appear to increase with storm's approach (suggesting an increase in vertical arrival angle).
6. Some storms over 1000 km from the sensors have been identified as sources by triangulation and tracking from several observatories.
7. A few instances have been recorded of simultaneous observation of surface pressure fluctuations and ionospheric electron density fluctuations (Georges, 1968a, 1968b; Davies and Jones, 1971; Baker and Davies, 1969) attributable to the same storm system. The characteristics of the waves observed by the two methods usually bear little resemblance to each other in spectral content, however.
8. Observed wave periods range from about 6 to 300 seconds with 12-60 seconds being "typical"; there is some evidence linking longer period waves with more severe and larger storms.

A recent survey of observed properties of storm-related atmospheric waves is given by Bowman and Bedard (1972).

In practice, microbarograph wave signatures are associated with convective storms by establishing: (a) coincidence in time (allowing for propagation time) and duration with those of a known storm system; (b) coincidence in azimuth angle of arrival (and often in variation of angle with time) with that of a known storm system; and, if available, (c) triangulation from several microbarograph observatories coinciding with known storm location. Coincidence has been studied in the details of selected events during which strong atmospheric waves were detected and good weather and radar information were available. Figure (21.8) shows several examples of coincidence of the angle of arrival of atmospheric waves with radar-indicated storm activity. A few observations (Goerke, personal communication) suggest that the generation of atmospheric waves may precede, by an hour or more, the appearance of severe-weather indications on weather radar.

Coincidence can also be established on a statistical basis by comparing independent searches of both microbarograph and weather (principally radar) data during a specified sample period, perhaps several months long. A thorough statistical study, indicating the frequency of coincidence of severe weather indicators with wave observations and the frequency of severe storms without signals has not yet been completed, but an indication of the statistical connection between the two can be seen in the azimuth-time plot of (F21.9), for a single 24-hour period.

The strong experimental evidence connecting severe storms with acoustic radiation has stimulated some theoretical speculation on possible acoustic-gravity wave generation mechanisms. Speculation has centered on three basic processes: (a) sound generation in convection cells by turbulence according to the theory formulated by Lighthill (1952, 1954), refined by Stein (1967) and Lighthill (1967) and discussed by Meecham (1971); (b) buoyancy oscillations and perturbations of the mean flow associated with "penetrative convection" of a cell above the tropopause (Pierce and Coroniti, 1966; Townsend, 1966, 1968); and (c) acoustic radiation (thunder) from lightning discharges. Although none of these speculations has yet given rise to a complete model of storm radiation, some preliminary tests of the speculations thus far advanced suggest themselves: A test of the theory that the emissions are associated directly with electrical activity seems immediately possible, using radio-direction-sensing equipment to locate the sources of "atmospherics" emitted by lightning strokes. Direct comparisons between the time, azimuth and duration of individual electromagnetic and acoustic emissions should then be possible. If buoyancy oscillations play an important role in atmospheric wave generation, it should be possible to observe the oscillations in the heights of cloud tops, particularly those penetrating the tropopause. Photographic (time-lapse) and radar techniques should be useful here. Clear-air radars should search for evidence of wavelike perturbations of the mean flow in the lee of cumulus tops, as possible sources of gravity-wave radiation.

Rough energy comparisons can be made to see if proposed source mechanisms provide sufficient energy to explain observed wave intensities. A 1μ bar wave pressure amplitude corresponds to a plane-wave energy density of $p'^2/\rho_0 C^2 = 7.2 \times 10^8$ joules/m³, or a wave energy flux of $p'^2/\rho_0 C = 2.4 \times 10^5$ watts/m² for values of atmospheric density ρ_0 and sound speed C at the earth's surface. This can be regarded as the energy flux of a "typical" storm-associated wave event from a storm that may be some 1000 km distant. If the energy spreads cylindrically inside a duct effectively 200 km high, energy is emitted by the source at an average rate of 3×10^7 watts.

First, let us see how much kinetic energy is contained in a simple model of a cumulus updraft, and how the required acoustic energy flux compares with it. The kinetic energy of a cylindrical updraft 10 km in diameter and 10 km high with an average draft of 10 m/s (Newton, 1967) is $1/2 \rho_0 u^2 V$ (where ρ_0 is taken at a height of 5 km and V is the cylinder volume) - 2.8×10^{13} joules. Taking a cell lifetime of 1000 sec (the time for the kinetic energy to be converted into other forms) gives an energy-conversion rate of 2.8×10^{10} watts. Our 3×10^7 watts of acoustic energy is only 0.1% of this total energy. Such an efficiency is compatible with estimates of the acoustic efficiency of explosive sources.

Next let us see how much energy is available from lightning flashes, and estimate the flash rate required to produce the observed acoustic energy flux. A typical lightning "flash" (i.e., usually several strokes) is said to contain about 2 to 5×10^{10} joules, of which 4×10^7 joules is converted to thunder (Remillard, 1960). Such flashes would have to occur about once every 1.3 seconds to generate acoustic energy at an average rate of 3×10^7 watts. Cloud-to-ground strokes are reported from severe storms as often

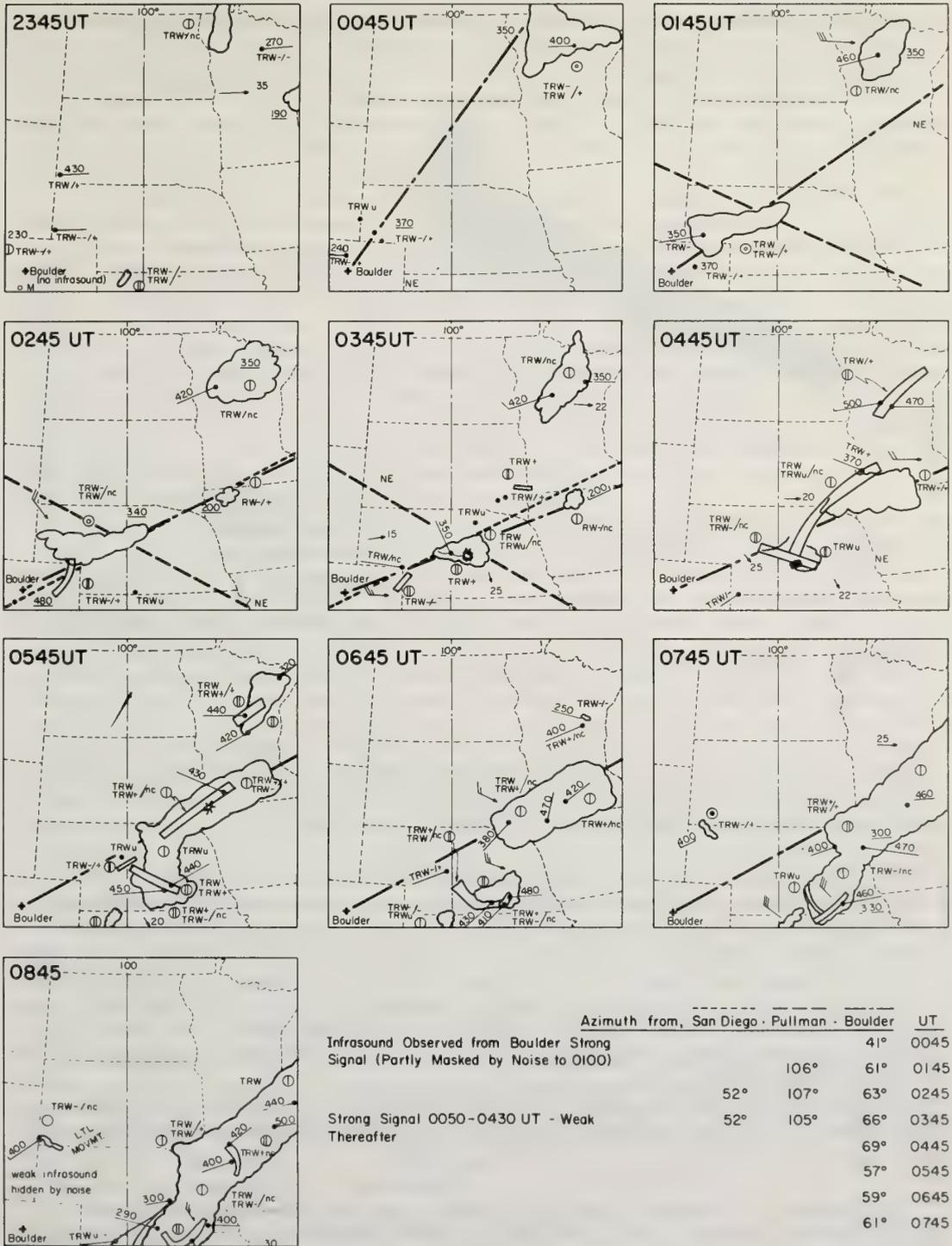


Figure 21.8. A sequence of tracings from 19 June 1968 weather radar charts upon which are superimposed dashed lines that indicate the bearings of pressure wave arrivals at various geoaoustic observatories. The fluffy outlines indicate areas of diffuse radar returns, while the dots indicate strong single-cell echoes that reach altitudes indicated (in hundreds of feet) by the adjoining numbers. (Figure courtesy of V. H. Goerke.)

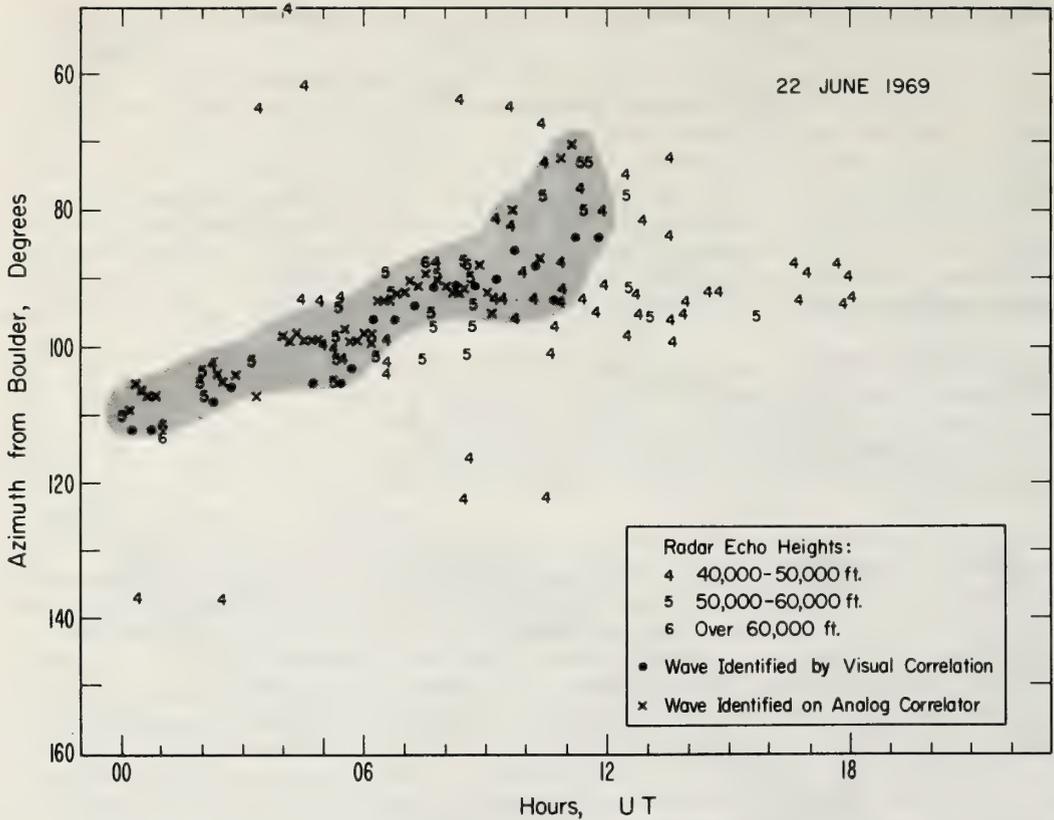


Figure 21.9. An azimuth-time plot combining observations of weather radars and pressure-wave arrivals. Single-cell radar echoes 40,000 ft. or higher and within 1500 km of the microphones are plotted as numbers that indicate echo (and presumably cloud) height. Wave arrivals are indicated by x's or dots and the general portion of the plane occupied by waves has been shaded.

as every 6 to 10 seconds, not an unreasonable disagreement, considering that some of our numbers may be in doubt by an order of magnitude or more. Conflicting estimates of the acoustic spectrum of thunder are reviewed by Uman (1969); the most reliable estimates indicate little energy radiated in the infrasonic portion of the spectrum. If so, the acoustic energy radiated to long distances may be considerably diminished; however, the relevance of most acoustic measurements of thunder, usually made within a few kilometers of the source, to long-distance propagation and the spectral content of the source can and should be questioned. Furthermore, nonlinear effects in the upper atmosphere may account for the required spectral shift in wave energy.

Theoretical bases exist in the works of Lighthill and Stein, for example, for calculating the acoustic energy emitted by turbulence associated with convective cells. Formulas for radiated acoustic power usually take the form

$$P = K \frac{\rho_0 u_t^3}{\ell_0} VM^5 \text{ ergs/sec} \quad (21:9)$$

where K is a "constant" that varies from unity to 10^3 , depending on the author, ρ_0 is the ambient air density, u_t is the rms turbulent velocity fluctuation, ℓ_0 is the length scale of the turbulence, M is the turbulence Mach number, u_t/c , and V is the volume extent of the turbulent source.

Clearly, this formula is extremely sensitive to the value of u_t one selects (the eighth power!), and accurate experimental data on what values represent severe storm conditions are lacking. Values between 1 and 10 m/s are suggested by Doppler radar observations and will be used in a sample calculation. A value of 100 is chosen for K , l_0 will be taken as 10 m, and ρ_0 and C appropriate to an altitude of 5 km will be used. The volume extent of the turbulence is assumed to be $5 \times 5 \times 5$ km. The 1 m/s value of u_t yields an acoustic power output of 0.27 watts, while the 10 m/s value yields 2.7×10^7 watts, very near our estimate of 3×10^7 watts, based on acoustic observations.

The distinguishing characteristic of sound generated by turbulence would thus be a strong "threshold effect," with sound seemingly emitted by only the strongest convection cells.

Present observations suggest that the frequency spectra of waves from convective storms are not clearly related to cell size, distance, or other storm parameters. However, filtering of the source wave spectra by propagation effects has not been studied in detail. For example, it is known that prevailing winds in the upper atmosphere prevent long-distance detection at the surface of some frequencies when they blow predominantly against the direction of wave travel, but the exact filtering function of the winds is not known. This point illustrates the importance of understanding how acoustic-gravity waves propagate through the atmosphere; i.e., when performing tests of various source mechanisms one must know how the emissions are altered by the propagation medium. For example, if sound is generated by turbulence, a broad spectrum of radiation would be expected, unless the atmosphere severely band-limits the waves that can reach significant distances. Similarly, waves from an impulse source are dispersed, or spread out in time, by the frequency-dispersion characteristics of the atmosphere.

Acoustic-gravity-wave propagation can be studied by modeling the atmospheric wind and temperature fields and solving the appropriate wave equation with suitable boundary conditions. Unfortunately, wave solutions in closed form have been found for only the simplest of atmospheric structures, e.g., an isothermal atmosphere with constant wind (Hines, 1960) and for linear and exponential temperature profiles (Pitteway and Hines, 1965). Numerical schemes have been devised for seeking "modal" solutions of coupled wave equations in multilayered atmospheres, but these so far appear to be limited to horizontally stratified temperature and wind fields with no cross winds (Pfeffer and Zarichny, 1962; Press and Harkrider, 1962). These, and many more recent analyses, show marked atmospheric dispersion of the various frequencies present in, say, an impulsive source, and the resultant time spreading of signals traveling long distances.

Under conditions in which the "geometric-acoustics" approximation is justified (small changes in refractive index within a wavelength) various "ray-tracing" procedures reveal the paths that wave energy follows through model atmospheres that can have quite complicated three-dimensional wind and temperature fields. Ray approaches tend to be the most appropriate way to describe propagation to short distances and at high frequencies, while modal descriptions more readily describe propagation to long distances and at low frequencies.

The significance of acoustic-gravity wave "rayplots" is readily grasped intuitively because of the close analogy with geometric optics. A sample ray plot is shown in (F21.10) for a wave source at a height of 5 km, representing the mean height of a thunderstorm cell. The atmospheric model is a temperature profile resembling the 1962 U.S. Standard Atmosphere with a logarithmic increase of eastward wind with height.

This figure immediately tells us two things: (a) significant acoustic energy reaches ionospheric heights (above 200 km) only within about 200 km (ground range) of the source, and (b) virtually no acoustic energy reaches the ground within a "shadow zone" extending from about 50 to 300 km from the source, except when winds are present. The ground is directly illuminated only out to about 25 km from the source, a number often given as the zone of audibility of thunder. We also see how winds prevent long distance detection "upstream" by surface sensors.

Upper atmosphere winds also cause frequency and angular dispersion; for example, when cross winds vary along the propagation path, or when horizontal wind shears are present, azimuthal refraction occurs and complicated relationships between wavefront and group packet (energy) propagation direction must be reckoned with. Atmospheric wave raypaths can now be calculated numerically in atmospheres with three-dimensional wind and temperature structure as well as the anisotropy introduced by gravity (Georges, 1971). Such programs greatly enhance our ability to simulate and understand atmospheric-wave propagation in realistic atmospheres.

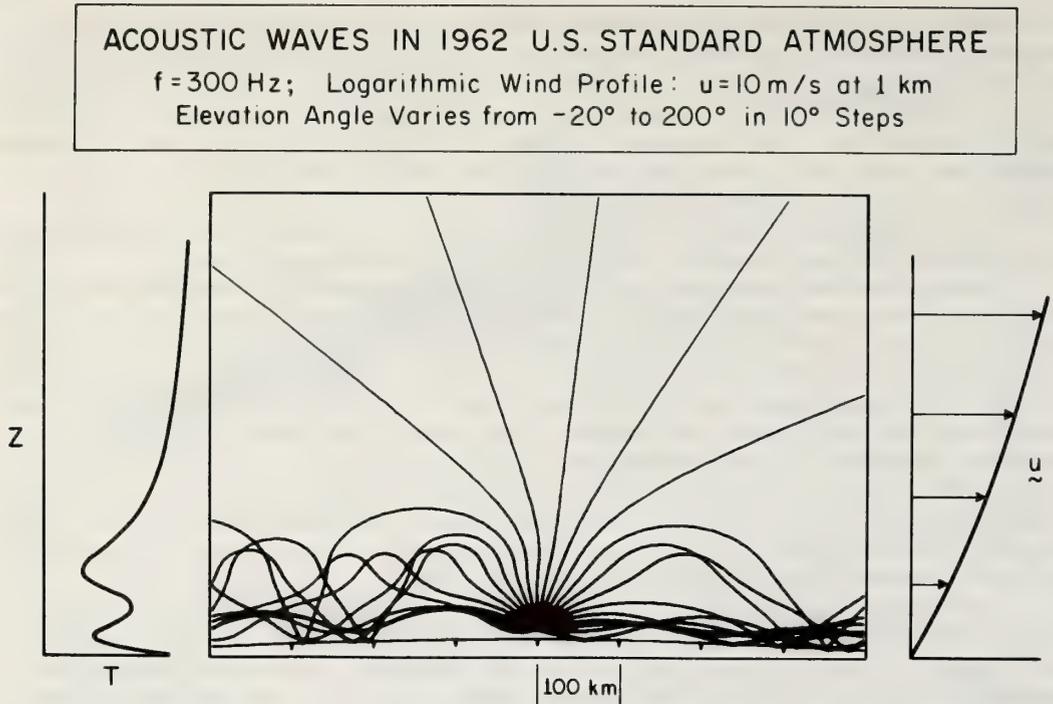


Figure 21.10. A computer-generated plot of acoustic ray paths in a model atmosphere with a temperature profile resembling USSA 1962 and a logarithmic wind profile.

Another type of wavelike pressure fluctuation has been linked with severe-weather outbreaks, but should be distinguished from those just discussed because they travel across the sensor arrays with speeds much less than acoustic. They are thus sometimes called “subsonic” pressure fluctuations (Bowman, 1968; Bowman and Bedard, 1972) and are most often associated with the passage of convection cells or frontal systems. Donn, et al. (1954) reviewed in detail the essential features of these waves; not much more is known about them today. The largest fluctuations are probably the same as “pressure jumps” often observed on ordinary recording barometers.

Refined pressure-measurement capabilities are now indicated for closer study of severe-weather-related waves: specifically, a pair of microbarograph observatories about 100 km apart are being implemented near Boulder, Colorado, for the purpose of triangulating and tracking severe weather in the midwest. Data from these stations will be augmented with passive electromagnetic direction-sensing stations and with data from the NOAA operational weather radar network. Elevated pressure sensors are being designed to permit direct measurement of k_z as well as the height dependence of p in the atmospheric boundary layer.

Joint efforts with meteorological Doppler radars could help test models of turbulent wave generation, by establishing the velocity fields within cells emitting detectable acoustic-gravity waves.

21.6 Manmade and Controlled Sources

Although we are concerned here principally with atmospheric waves of natural origin, the characteristics of some signals thought to be of manmade origin should be mentioned briefly. In some cases, (e.g., nuclear and non-nuclear explosions) they provide (roughly) known sources by which atmospheric propagation can be studied (Donn and Shaw, 1967); in others (e.g., supersonic aircraft shocks) they merely leave signatures that need be recognized only to eliminate them from analysis.

Donn and Shaw (1967) show several microbarograph signatures of large nuclear explosions detonated many thousands of kilometers from the sensor arrays.

Figure (21.11) shows signatures believed to be caused by "sonic booms," the shock wave generated by supersonic aircraft. Although most of the sonic-boom signatures are caused by aircraft too remote to produce audible effects, the identity of the source mechanisms for this characteristic signature has been fairly well established by association with audible booms.

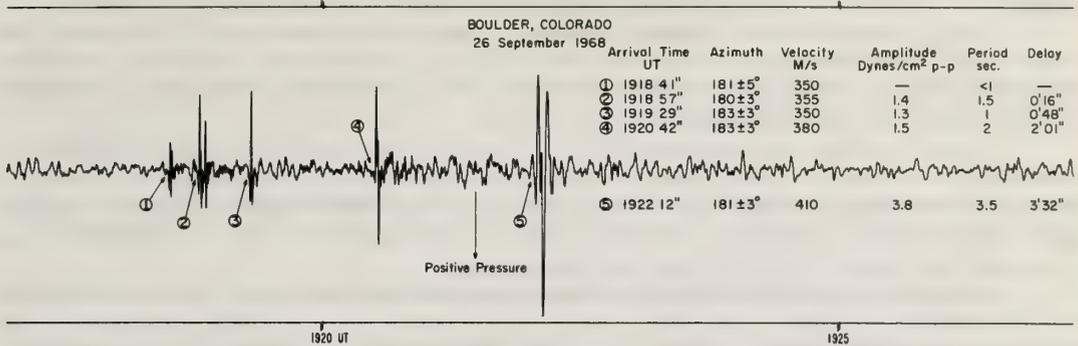


Figure 21.11. A series of pressure signatures believed to be caused by the sonic boom of a supersonic aircraft. Multiple arrivals characterize propagation over distances greater than about 100 km and probably represent propagation paths via the upper atmosphere. Note pulse lengthening of later arrivals. (Figure courtesy of V. H. Goerke.)

Explosive sources provide broadband energy that, in principle, can yield a kind of atmospheric "impulse response," which in turn can be transformed into an atmospheric transfer function in the frequency domain. Unfortunately, atmospheric time variability appears to exert an important effect on wave propagation that cannot be diagnosed with impulse-testing methods. This has been demonstrated in several instances where sequential explosions produce markedly different signatures on distant microbarographs. Consequently, some interest has been expressed in the development of artificial monochromatic sources of acoustic-gravity waves to study propagation in the time domain. Because the wavelengths of interest are exceptionally long (a wave with a period of 10 seconds is 3.4 km long under standard conditions), generators that could be constructed are quite inefficient. Several generating schemes have been proposed, but none, as far as we know, has yet been implemented.

21.7 The Role of Passive Acoustic-Gravity Wave Observations in Studies of Atmospheric Dynamics

Apart from whatever interest acoustic-gravity waves attract in themselves or as remote indicators of the other important geophysical phenomena, increasing attention is being given the possibility that atmospheric waves are efficient transporters of energy and momentum over long distances in the atmosphere, and thus may play an important role in atmospheric dynamics. The key to understanding this role is developing an ability to measure and model how much energy and momentum are transported by waves from one part of the atmosphere to another and how wave energy and momentum are converted into other forms of atmospheric motions.

If pressure fluctuations, p' , the vertical velocity fluctuations, u_z' , and the horizontal velocity fluctuations, u_x' can be measured directly at some point in the atmosphere, one can estimate energy and momentum fluxes through, say, a unit vertical or horizontal surface in the vicinity of the measurement point. Instantaneous vertical energy flux, for example, is proportional to the product $p'u_z'$; vertical flux of horizontal momentum is proportional to $u_x'u_z'$. These relations do not depend on the measured quantities being wave-associated.

Suppose one is interested in the wave fluxes that enter and leave a cubical volume of atmosphere, and also in the origins of waves that enter the volume from outside or that are generated within the volume. One could consider placing pressure and velocity sensors over the faces of the volume, with sufficient sensor density to record fluctuations down to some small scale determined by his interests. Usually, however, the atmosphere is physically and economically inaccessible to anything like the required

sensor density. If distinct spectral components can be identified (i.e., wave vectors \underline{k} and frequencies Ω with respect to the background air, determined) one can assume the presence of superimposed plane waves and use the theoretical wave-polarization relations (see Chapter 5) to provide the ratios among p' , u_x' , u_z' . One then needs to measure only one indicator of wave amplitude at each point in the atmosphere where a flux is desired. Furthermore, if the presence of plane waves is assumed, one can replace the need for densely spaced sensors with inferences about the spatial properties of the wave-associated motions.

Usually, one is interested in average fluxes over temporal and spatial intervals. In the case of wave motions, one usually averages over at least the wave period of the wave with the longest period of interest. The relevant spatial averaging surface is determined by the coherence distance of the wave, i.e., the distance over which the wave maintains its identity.

In principle, then, one can measure vertical fluxes for each spectral component of interest, by measuring one indicator of wave amplitude at each height of interest, along with each wave vector and wave period. Caution is obviously called for in assuming that plane-wave relations adequately describe a given observed wavelike motion; if a plane wave model is not warranted, then direct measurements of both wave quantities required for the flux calculation are necessary.

Atmospheric waves deposit (dissipate) their energy in at least two ways. One is by viscous (or thermal) dissipation, in which wave energy is converted into random motions, directly on a molecular scale (heat) where molecular viscosity dominates as in the upper atmosphere. They may also dissipate via an eddy scale where eddy viscosity dominates, as in the lower atmosphere. The other wave-dissipation process is the interchange of momentum between wave and the mean flow at "critical layers," i.e., where wave phase velocity closely matches the mean air velocity. The longer-period acoustic-gravity waves, in particular internal-gravity waves, couple most effectively with the atmospheric mean flow because their phase velocities can be arbitrarily small with respect to mean flow. Models of atmospheric dynamics (e.g., global circulation) are able to account for viscous dissipation of waves by just adding terms in the equations of motion and parameterizing the appropriate atmospheric quantities. Critical-layer phenomena, on the other hand, are not so readily parameterized, and indeed are not yet fully understood. It is thus of interest to measure wave fluxes in the vicinity of critical layers in the atmosphere.

21.8 Conclusion

Various lectures in this series have shown how active and passive sensors can be used to measure meteorological variables remotely in the atmosphere. Several of these new sensors, most notably acoustic-echo sounding and FM/CW electromagnetic sounding, are beginning to reveal the importance—perhaps even dominance—of atmospheric-wave motions in tropospheric dynamics. This lecture has been about techniques for passively observing these waves with relatively inexpensive sensors that can be deployed in spatial arrays. Such a spatial-resolution capability seems to be the ideal companion to more expensive tools that yield height-time details of atmospheric motions from a single ground location. Hooke (Chapter 7), for example, has shown how spaced microbarographs can supplement active sounding techniques and provide wave-velocity information not obtainable with a single sounder.

Passive atmospheric-wave sensors also reveal signatures of distant atmospheric events, many of which remain unidentified. Some of these signatures seem useful as remote indicators of important atmospheric events, either natural or manmade, even though the waves themselves may not play important roles in atmospheric dynamics. Others may indicate important vertical energy and momentum transfer processes, as seems likely in the case of jet-stream associated waves.

This review has attempted not only to provide a cross section of current research and an indication of the present state of knowledge in the field, but also to reveal the state of our ignorance about the sources, propagation and dissipation of the atmospheric waves whose importance in atmospheric dynamics is just beginning to be appreciated. This review will have served its purpose if it stimulates the interest of both theorist and experimentalist in the imposing array of unsolved theoretical and observational problems that remain.

21.9 References

Baker, D. M. and K. Davies (1969), F2-region acoustic waves from severe weather, *J. Atmosph. Terr. Phys.* *31*, 1345-1352.

- Benioff, H. and B. Gutenberg (1939), Waves and air currents recorded by electromagnetic microbarographs, *Bull. Am. Meteorol. Soc.* 20, 42.
- Bowman, H. S. (1968), Subsonic waves and severe weather phenomena in Acoustic gravity waves in the atmosphere—Symposium Proceedings, T. M. Georges, ed., U. S. Govt. Printing Office, Washington, D. C., 215-222.
- Bowman, H. and A. J. Bedard (1972), Observations of infrasound and subsonic pressure disturbances related to severe weather, *Geophys. J. of Roy. Astron. Soc.* (special issue on Infrasound), (to be published).
- Briggs, B. H., G. S. Phillips and D. H. Shinn (1950), The analysis of observations on spaced receivers of fading radio signals, *Proc. Phys. Soc. London B* 63, 106-121.
- Brown, R. F., Jr. (1963), An automatic multi-channel correlator, *J. Res. NBS* 67C, 1, 33-38.
- Chrzanowski, P., G. Greene, K. T. Lemmon and J. M. Young (1961), Traveling pressure waves associated with geomagnetic activity, *J. Geophys. Res.* 66, 3727-3733.
- Cook, R. K. (1962), Strange sounds in the atmosphere, I, *Sound I*, 12-16.
- Cook, R. K. and A. J. Bedard (1972), On the measurement of infrasound, *Geophys. J. Roy. Astron. Soc.* (to be published).
- Cook, R. K. and J. M. Young (1962), Strange sounds in the atmosphere, II, *Sound I*, 25-33.
- Davies, K. and J. E. Jones (1971), Three-dimensional observations of traveling ionospheric disturbances, *J. Atmos. Terrest. Phys.* 33, 39-46.
- Donn, W. L., R. Rommer, F. Press and M. Ewing (1954), Atmospheric oscillations and related synoptic patterns, *Bull. Am. Meteorol. Soc.* 35, 301-309.
- Donn, W. L., and D. M. Shaw (1967), Exploring the atmosphere with nuclear explosions, *Rev. Geophys.* 5, 53-82.
- Georges, T. M. (1968a), Short-period ionospheric oscillations associated with severe weather in Acoustic gravity waves in the atmosphere—Symposium Proceedings, T. M. Georges, ed., U. S. Govt. Print. Office, Washington, D. C.
- Georges, T. M. (1968b), HF Doppler studies of traveling ionospheric irregularities, *J. Atmos. Terrest. Phys.* 30, 735-746.
- Georges, T. M. (ed) (1968c), Acoustic-gravity waves in the atmosphere—Symposium Proceedings, U. S. Govt. Print. Office, Washington, D. C.
- Georges, T. M. (1971), A program for calculating acoustic-gravity ray paths in the atmosphere, NOAA Tech. Report ERL 212-WPL 16, U. S. Govt. Print. Office, Washington, D. C.
- Goerke, V. H. and M. W. Woodward (1966), Infrasonic observations of a severe weather system, *Monthly Weather Review* 94, 395.
- Goerke, V. H., J. M. Young and R. K. Cook (1965), Infrasonic observations of the May 16, 1963, volcanic explosion on the island of Bali, *J. Geophys. Res.* 70, 6017-6022.

- Gossard, E. E. (1967), The apparent movement of the spectral components in fading records of ionospherically reflected radio waves, *J. Geophys. Res.* **73**, 1563-1569.
- Gossard, E. E. (1969), The effect on bandwidth on the interpretation of the cross spectra of wave recordings from spatially separated sites, *J. Geophys. Res.* **74**, 325-337.
- Gossard, E. E. and V. F. Noonkester (1967), A guide to digital computation and use of power spectra and cross-power spectra, NELC Tech. Doc. 20, Naval Electronics Laboratory Center, San Diego, Calif.
- Gossard, E. E. and D. B. Sailors (1970), Dispersion bandwidth deduced from coherency of wave recordings from spatially separated sites, *J. Geophys. Res.* **75**, 1324-1329.
- Herron, T. J. and I. Tolstoy (1969), Tracking jet stream winds from ground level pressure signals, *J. Atmos. Sci.* **26**, 266-269.
- Hines, C. O. (1960), Internal atmospheric gravity waves at ionospheric heights, *Can. J. Phys.* **38**, 1441-1481.
- Jordan, J. L. and K. C. Bayer (1967), Exploding meteor located by seismographs and microbarographs, *Natl. Earthquake Info. Bull.* **1**, May, 1967.
- Kimball, B. A. and E. R. Lemon (1970), Spectra of air pressure fluctuations at the soil surface, *J. Geophys. Res.* **75**, 6771-6777.
- Larson, R. J. (1971), The correlation of winds and geographical features with the production of certain infrasonic signals in the atmosphere, Research Report No. 71/16-1, College of Eng., Washington State Univ., Pullman, Wash.
- Lighthill, M. J. (1952), On sound generated aerodynamically, Part I, General Theory, *Proc. Roy. Soc. A*, **211**, 564-587.
- Lighthill, M. J. (1954), On sound generated aerodynamically, Part II, Turbulence as a source of sound, *Proc. Roy. Soc. A*, **222** 1-32.
- Lighthill, M. J. (1967), Predictions on the velocity field coming from acoustic noise and a generalized turbulence in a layer overlaying a convectively unstable atmospheric region, *Proc. I.A.U. Symposium* no. 28, 429-453.
- Lumley, J. L. and H. A. Panofsky, 1964: *Structure of atmospheric turbulence*, John Wiley, New York.
- Matheson, H. (1964), Instructions for the operation of NBS infrasonic equipment, NBS Report 8519, U. S. Dept. of Commerce, Washington, D. C.
- Meecham, W. C. (1971), On aerodynamic infrasound, *J. Atmos. Terrest. Phys.* **33**, 149-155.
- Morse, P. and K. Ingard, 1968: *Theoretical Acoustics*, McGraw-Hill Publishing Co. (New York).
- Newton, C. W. (1967), Severe convective storms, *Advances in Geophy.* **12**, 257-308.
- Pfeffer, R. L. and J. Zarichny (1962), Acoustic-gravity wave propagation from nuclear explosions in the earth's atmosphere, *J. Atmos. Sci.* **19**, 175-199.
- Pierce, A. D., and J. W. Posey (1970), Theoretical prediction of acoustic gravity pressure waveforms generated by large explosions in the atmosphere, Final Report AFCRL-70-0134, Dept. of Mech. Eng., Mass. Inst. Tech., Cambridge, Mass.

- Pierce, A. D. and S. C. Coroniti (1966), A mechanism for the generation of acoustic-gravity waves during thunderstorm formation, *Nature* 210, 1209-1210.
- Pitteway, M. L. V. and C. O. Hines (1965), The reflection and ducting of atmospheric acoustic-gravity waves, *Can. J. Phys.* 43, 2222-2243.
- Press, F. and D. G. Harkrider (1962), Propagation of acoustic-gravity waves in the atmosphere, *J. Geophys. Res.* 67, 3889-3908.
- Remillard, W. J. (1960), The acoustics of thunder, Technical Memo. No. 44, Acoustics Research Laboratory, Harvard Univ., Cambridge, Mass.
- Saxer, L. (1945), Electrical measurement of small atmospheric pressure oscillations, *Helv. Phys. Acta.* 18, 527-550.
- Saxer, L. (1953), *Arch. Meteorolo. Geophys. U. Bioklimatol.* A6, 451.
- Stein, R. F. (1967), Generation of acoustic-gravity waves by turbulence in an isothermal stratified atmosphere, *Solar Physics* 2, 385-432.
- Thomas, J. E., A. D. Pierce, E. A. Flinn and L. B. Craine (1972), Bibliography on infrasonic waves, *Geophys. J. Roy. Astron. Soc.* (to be published).
- Tolstoy, I. (1963), The theory of waves in stratified fluids including the effects of gravity and rotation, *Rev. Mod. Phys.* 35, 207-230.
- Townsend, A. A. (1966), Internal waves produced by a convective layer, *J. Fluid Mech.* 24, 307-319.
- Townsend, A. A. (1968), Excitation of internal waves in a stably-stratified atmosphere with considerable wind shear, *J. Fluid Mech.* 32, 145-171.
- Uman, M. A., 1969: *Lightning*, McGraw-Hill, New York.
- Wilson, C. R. (1969), Auroral infrasonic waves, *J. Geophys. Res.* 74, 1817-1836.

List of Symbols

C	The classical speed of sound	Φ_{xy}	Cross-correlation function of time series $x(t), y(t)$
C_{xy}, Q_{xy}	Co- and quadrature spectra of $x(t), y(t)$	ρ	Atmospheric density (= $\rho_0 + \rho'$, a static plus a perturbation part)
f	Wave frequency	R_{ij}	Generalized correlation coefficient
g	Acceleration of gravity	\underline{u}	Vector velocity (= $\underline{u}_0 + \underline{u}'$, a static plus a perturbation part; Cartesian components: u_x, u_y, u_z ; u_t is a turbulent velocity component; u is the magnitude of \underline{u})
\underline{K}	Vector wave number (Cartesian components: k_x, k_y, k_z ; magnitude k)	τ	Time lag in correlation analysis
λ	Wavelength (Cartesian components: $\lambda_x, \lambda_y, \lambda_z$)	ω	Angular wave frequency
M	Mach number u/C	ω_g	The Brunt-Väisälä frequency
p	Pressure (= $p_0 + p'$, a static plus a perturbation part)	x, y, z	Cartesian coordinates
P	Acoustic power		

Chapter 22 REMOTE SENSING OF THE OCEAN

Kirby J. Hanson

National Oceanic and Atmospheric Administration
Atlantic Oceanographic and Meteorological Laboratories
Sea-Air Interaction Laboratory
Miami, Florida

22.0 Introduction

Progress in remote sensing of the oceans has been slow since the satellite era began more than a decade ago. The reasons for this are numerous. However, I believe the over-riding reason is that, although satellite experiments were designed to study both meteorological and ocean variables, the atmospheric variables (such as cloudiness) were clearly distinguishable and had great utility for meteorologists. For the oceanographer, the satellite provided useful sea surface temperature measurement only over a small fraction of the world's oceans where upwelling and western boundary currents cause horizontal temperature gradients which were large enough to be detectable above atmospheric and measurement noise. It is little wonder the oceanographic community has been slow to respond to the potential offered by satellite platforms.

In spite of the difficulty in interpreting sea surface temperature over broad oceanic areas, considerable progress has been made in the past five years. That progress is discussed in the first section of this chapter.

Studies of the possibility of global measurement of other oceanic variables have continued, mainly supported by NASA. These studies have led to a great deal of optimism about the possibility of remote sensing of salinity, temperature, surface roughness, foam, sea spray, sea ice coverage, and phytoplankton in the surface mixed layer of the ocean. Passive microwave and ocean color measurements may make these global measurements possible. They are discussed in the second and third sections of this chapter. A shortcoming of this chapter is that it does not include a discussion of the potential of active microwave sensing, or remote sensing of heat flux from the ocean surface as exemplified by the work of McAlister and his colleagues of Scripps, University of California, San Diego.

Because of the rapid development in these three areas, there is no single document which gives students a statement of present knowledge. To attempt to fill this need, this chapter gives the reader the parametric relationship between oceanic variables and the electromagnetic energy either emitted or reflected from the ocean surface. It also discusses what the major uncertainties are today.

The atmosphere is not transparent to all electromagnetic energy and measurement systems for remote sensing of the ocean must make use of windows in the absorption spectra of the active atmospheric gases (mainly H_2O , CO_2 , O_3 , O_2) in order to detect energy from the ocean. The principal atmospheric windows for visible, IR and microwave radiation are listed in (T22.1).[†] The boundaries of the windows are difficult to define because the transmissivity depends both on concentration of the active absorbing gases, and on path length of the radiation through the atmosphere.

22.1 Infrared Detection—Sea Surface Temperature

22.1.1 Satellite Experiments

When the first infrared (IR) window data were obtained from TIROS II in November, 1960, the interest of the investigators was meteorological; they examined the space and time variation in surface temperature over continental regions (e.g., Fritz, 1963, Rao and Winston, 1963). They also investigated the remote detection of cloud top heights for use in synoptic meteorology. Small attention was given to oceanography in the analysis of these early TIROS Medium Resolution IR (MRIR) data.

In 1964 the High Resolution Infrared Radiometer (HRIR) on NIMBUS-I provided an improved resolution of 5 n mi at the sub-satellite point. This was a great step forward for oceanography because it gave better resolution of horizontal temperature gradients along ocean current boundaries and upwelling regions than was possible with TIROS MRIR data. Table (22.2) (see page 22-13) shows the improvement of resolution in the IR window sensors as the satellite program evolved.

[†]The radiation windows can also be seen in the atmospheric transmissivity curve in the top of (F22.24).

Table 22.1 Principal Atmospheric Windows of the Electromagnetic Spectrum

Window	Wavelength	Frequency (GHz)	Absorbing Gas	
			Lower Boundary†	Upper Boundary†
Visible	0.3 to 0.9 μm	—	O ₃	H ₂ O
Near IR	1.5 to 1.6 μm	—	H ₂ O	H ₂ O
Near IR	2.0 to 2.3 μm	—	H ₂ O	H ₂ O
Intermediate IR	3.5 to 4.5 μm	—	H ₂ O	H ₂ O
Far IR	8.0 to 9.0 μm	—	H ₂ O	O ₃
Far IR	10.0 to 12.5 μm	—	O ₃	CO ₂
Microwave	0.3 to 0.36 cm	100 to 80	H ₂ O	O ₂
Microwave	0.7 to 1.0 cm	45 to 30	O ₂	H ₂ O
Microwave	2.0 to > 30 cm	15 to < 1	H ₂ O	—

† Refers to wavelength

An analysis of surface temperatures of the Gulf Stream, shown in (F22.3), is another example of significant oceanic temperature gradients (Allison, et al., 1967). These data are also from the NIMBUS II HRIR experiment. Studies of ocean current boundary detection and movement have been done by Warnecke, et al. (1967), Hansen and Maul (1970), and Maul and Hansen (1971).

These and other data from NIMBUS II provided hope for the remote detection of sea-surface temperature on a global scale to delineate major current boundaries and upwelling regions. But they also introduced the difficult task of properly interpreting sea-surface temperature from IR window measurements. This requires an understanding of how active atmospheric gases, clouds, and aerosols modify radiation emitted by the sea surface. The major progress of a number of investigators on these problems is reported in subsequent sections of this chapter.

In the period from 1966 to 1970 there was great improvement in displaying, mapping, and brightness enhancing IR satellite data. In 1970 ITOS-I provided many spectacular views of horizontal gradients in sea temperature which mark current bodies. Figure (22.4) shows the position of the north wall of the Gulf Stream on February 15, 1971. Relatively cold water outlines the Bahama Banks and coastal water in the Gulf of Mexico.

In (F22.5) cold air is moving off the east coast from New York to Florida causing cloudiness to develop over the Gulf Stream. The colder surface water south of Cape Hatteras appears to be moving eastward into the warm Gulf Stream in two large swirls under the influence of this offshore flow.

The scanning radiometers (SR) of ITOS-I and NOAA-I have 4 n mi resolution at nadir.† Future NOAA satellites, beginning with ITOS-D, are expected to have improved resolution with the addition of a Very High Resolution Radiometer (VHRR) which has resolution of 0.5 n mi at nadir (McClain, 1970).

Data from the NIMBUS-I HRIR experiment were used by many investigators for oceanographic studies. The radiometer contained a lead selenide photo-conductive cell which was radiatively cooled to -75°C and responsive to radiation in the 3.4-4.2 μm region. The radiometer field of view was 0.5 degrees. A scanning mirror, rotating in front of the detector, scanned normal to the orbital path. A sample of the detector output over two complete scans is shown in (F22.1). During each complete scan, the detector viewed space, the satellite housing, and the earth scene below. Data from space and the radiometer housing are useful in earth locating and evaluating the data. A complete description of the experiment is given in NASA, (1966).

NIMBUS satellites are in a noon/midnight local sun time orbit. This provides global coverage twice a day. However, only night-time coverage is useful for surface temperature determination with HRIR data, because a sufficient amount of reflected solar radiance is present at 3.4-4.2 μm during the daytime to contaminate the IR measurements.

In using NIMBUS HRIR data, it was natural for investigators to look for ocean temperature features which have large horizontal temperature gradients. One example is that due to upwelling along the Somali coast, as shown in (F22.2) from Allison and Kennedy (1967). Here the temperature varies by 25°F over only a few degrees latitude. Similar studies have recently been reported by Warnecke, et al. (1971).

†Nadir: the point of the earth's surface intersected by a line between the satellite and the earth's center.

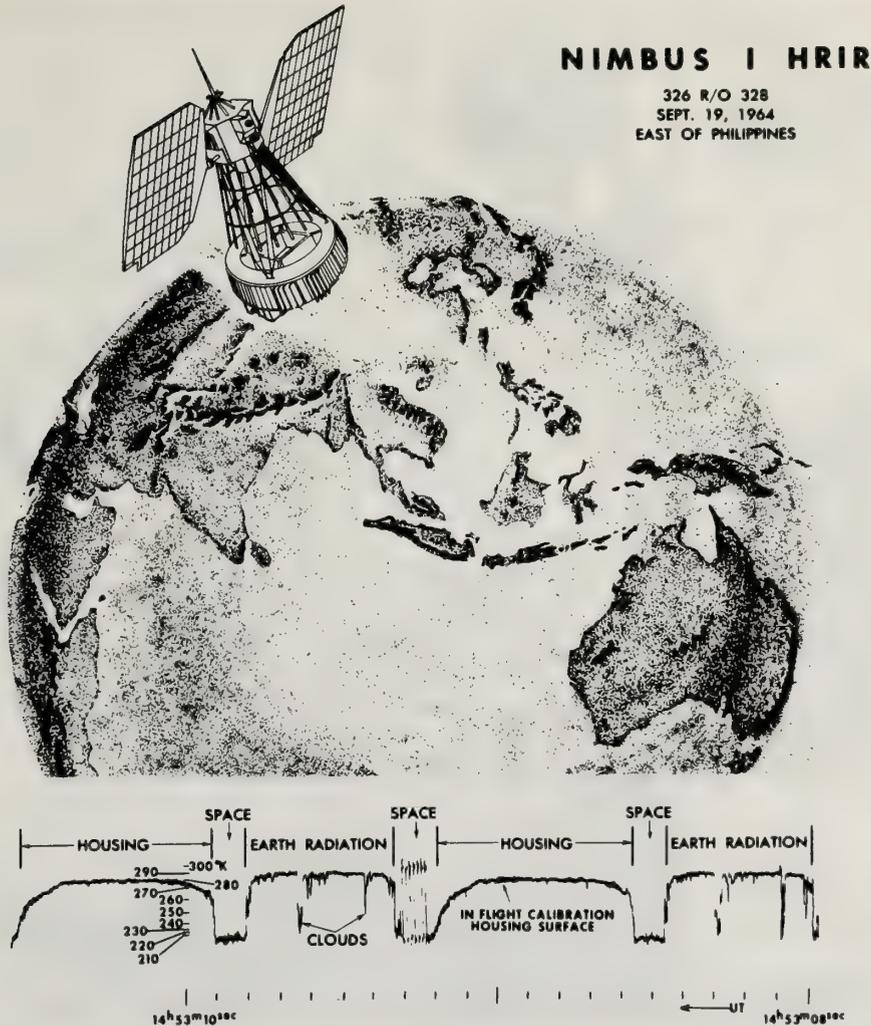


Figure 22.1 Portion of an HRIR analog record of NIMBUS I. In the example, the mirror made two complete revolutions, scanning the effective radiation temperature of space and the radiometer housing as well as the earth. The space and radiometer housing data are useful in positioning the satellite and calibrating the radiometer (after LaViolette and Chabot, 1968).

An example of sea surface temperature mapping, in (F22.6), is from a computer automated technique for global monitoring of sea surface temperature (Leese, 1971b) being developed at the National Environmental Satellite Service, NOAA. The two left polar projections show mean temperatures on a grey scale from 270-304 °K. The two right polar projections indicate the number of days which have elapsed since new satellite data have been added to the mean temperature value for each location. The grey scale for the two right projections ranges from (a) greater than 5 days (light grey); (b) 2-5 days (medium grey); (c) 1 day (black). Complete coverage from each satellite orbit is not possible because the computer program has certain rejection criteria when the effect of clouds is present in the data. A description of the experimental model for obtaining these temperature maps from ITOS-I is given by Leese, et al., (1971a). In (F22.6) there are numerous grey areas which indicate no satellite coverage was possible for more than 5 days due to clouds. This difficulty of cloud contamination of the data should improve as the very high resolution IR sensors are added in the NOAA satellite series and 0.5 n mi radar resolution is obtained with these sensors.

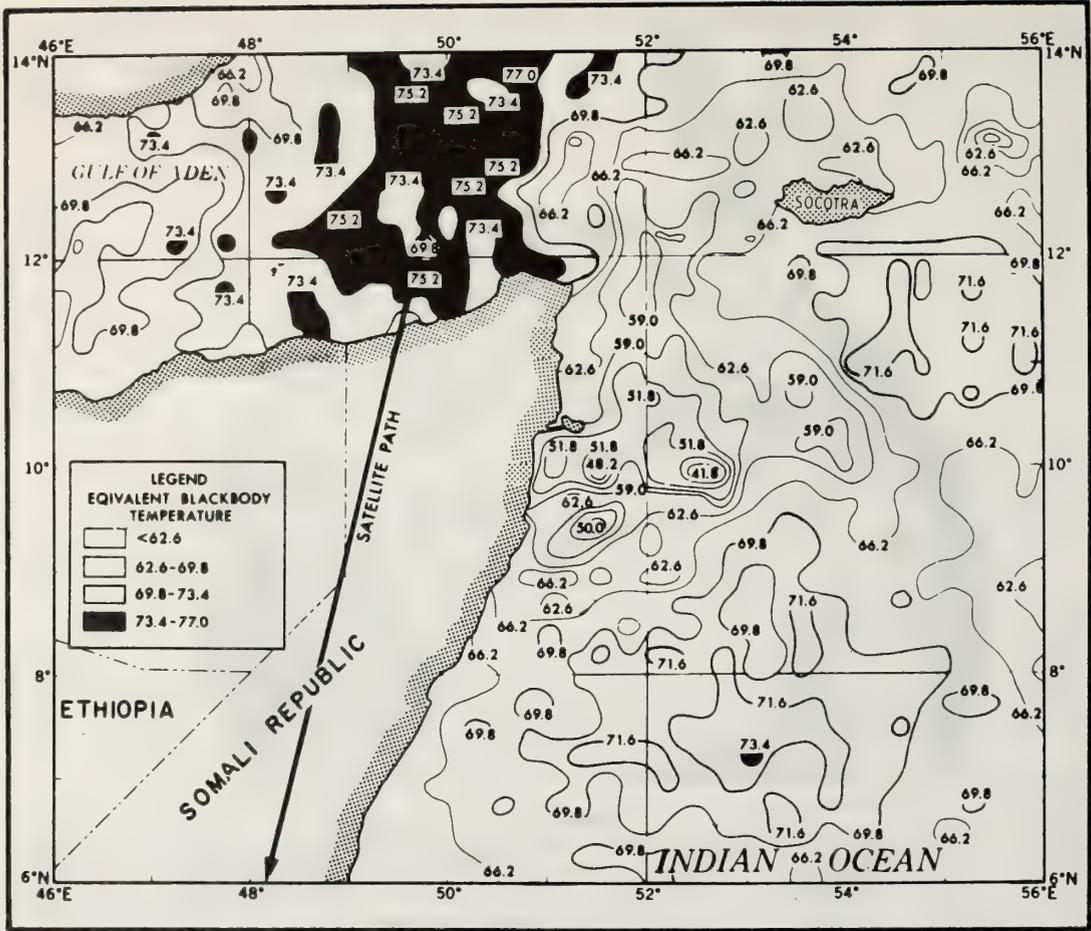


Figure 22.2 NIMBUS I determination of sea surface isotherms ($^{\circ}\text{F}$) along the Somali coast related to upwelling of cold water in that region (after Allison and Kennedy, 1967).

22.1.2 Theory

The notation and some of the discussion in this section follows that of Holter (1970). The fundamental radiation equation of Plank (1900) gives the spectral distribution of radiation from a self-emitting perfect radiator, called a blackbody, with uniform temperature. This theoretical law agrees with experimental results and applies to all regions of the electromagnetic spectrum. Plank's law, stated in terms of spectral radiant emittance, is

$$W_{B\lambda} = C_1 \lambda^{-5} [\exp(C_2/\lambda T) - 1]^{-1} \tag{22:1}$$

and has units of $\text{erg}/\text{cm}^2 \text{ sec } \mu\text{m}$. C_1 and C_2 are well known radiation constants; λ is wavelength and T is the temperature of the emitting surface.

Because most objects are not perfect radiators, a radiative efficiency factor, ϵ_{λ} , termed spectral emissivity, is defined as

$$\epsilon_{\lambda} = W_{\lambda}/W_{B\lambda} \tag{22:2}$$

where W_{λ} is the spectral radiant emittance of an imperfect radiator, called a greybody. Depending on the



Figure 22.3 Northern boundary of the Gulf Stream as observed by NIMBUS II (after Allison, et al., 1967).

radiative efficiency of a surface, the spectral emissivity ranges from near zero for poor radiators to unity for perfect radiators. Using (22:1) and (22:2) calculations can be obtained of the spectral radiant emittance of any surface of known temperature and spectral emissivity.

However, the spectral radiant emittance of a surface does not describe completely the electromagnetic energy emanating from that surface. The reason for this is that greybody surfaces reflect as well as emit energy. This is shown by a form of Kirchhoff's law which states that if a surface is optically opaque, then the spectral reflectivity, ρ_λ , and spectral emissivity, ϵ_λ , of that surface have the simple relation of

$$\rho_\lambda = 1 - \epsilon_\lambda \tag{22:3}$$

Thus, infrared radiation from a surface is due both to self-emission of the surface and reflection of a portion of the radiation incident on the surface.



Figure 22.4 *ITOS-1 scanning radiometer IR picture over the East coast of the United States, February 15, 1971, orbit No. 4853. The northern boundary of the Gulf Stream is clearly visible. Temperature gradients over the Bahama Banks and in the Gulf of Mexico are also visible (after Watson, 1971).*

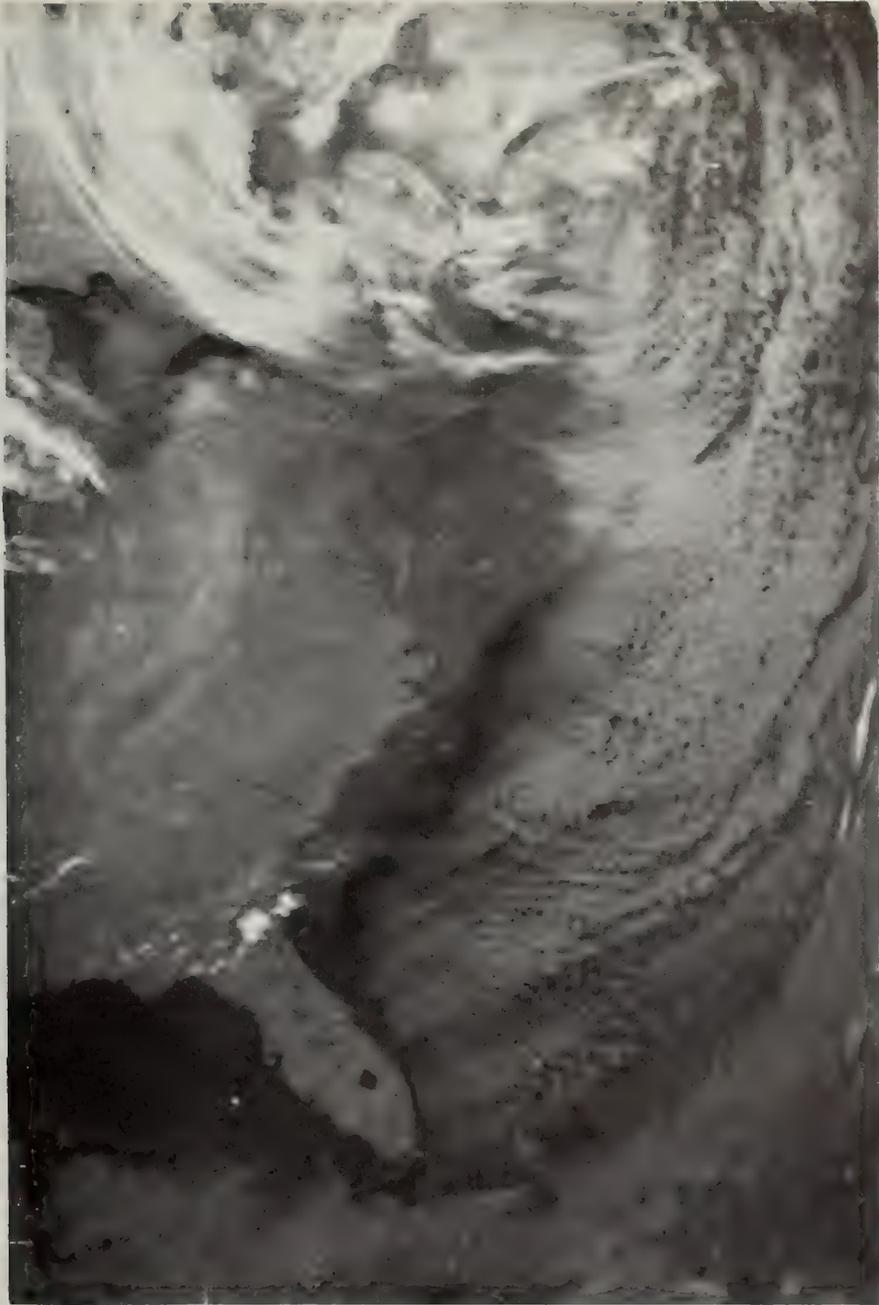


Figure 22.5 *ITOS-1 scanning radiometer IR picture over the East coast of the United States, March 5, 1971, orbit No. 5078. Cold water eddies appear to penetrate the northern boundary of the warm (dark) Gulf Stream east of North Carolina, as cold polar air moves off the entire east coast of the United States (after Watson, 1971).*

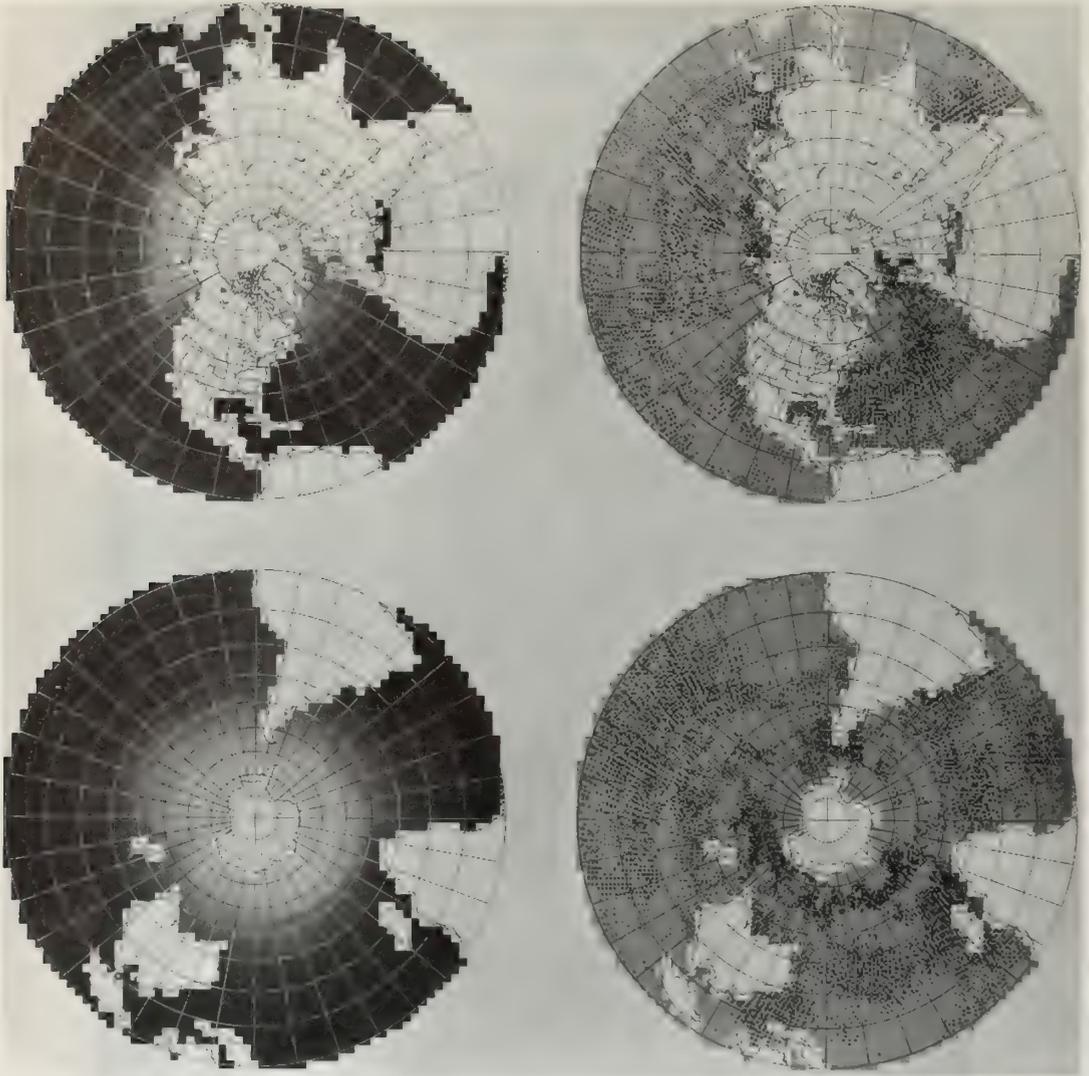


Figure 22.6 Average sea-surface temperature from ITOS-1 obtained by computer automated technique developed at the National Environmental Satellite Service, NOAA (Leese, 1971b). Temperature maps (left) indicate temperatures from 270-304°K. Maps (right) indicate the frequency with which new satellite data have been used to update the mean temperature (Leese, 1971a).

If these radiation principles are applied to an ocean surface, it follows from the previous statement that infrared radiation emanating from the ocean surface is due both to self-emission of the top few micrometers of the ocean and to reflection by the ocean surface of infrared radiation from such natural atmospheric sources as clouds, water vapor, carbon dioxide, and aerosols. The sum of these two radiation components is defined as the effective spectral radiance emittance of the sea surface, $W_{e\lambda}$.

$$W_{e\lambda} = \epsilon_{\lambda} W_{B\lambda} + \rho_{\lambda} W_{a\lambda} \quad (22:4)$$

where the self-emission component is $\epsilon_{\lambda} W_{B\lambda}$ and the surface reflection component is $\rho_{\lambda} W_{a\lambda}$.

If a determination of the effective spectral radiant emittance of the sea surface is to be obtained from satellite measurement, then account must be made of infrared absorption and emission by atmospheric water vapor, carbon dioxide, and ozone. At satellite height, the upward spectral radiance at a given zenith angle, θ , and wavelength, λ , is given by

$$N_{\lambda}(\theta) = W_{e\lambda}(T_s) \tau_{\lambda a}(\theta) + \int W_{a\lambda}(T_a) d\tau_{\lambda}(\theta) \quad (22:5)$$

Where: $W_{e\lambda}$, the effective spectral radiant emittance of the sea surface at temperature T_s ; $\tau_{\lambda a}$, the spectral transmissivity through the entire atmosphere; τ_{λ} , the spectral transmissivity through a layer which extends from any height in the atmosphere to the top of the atmosphere; $W_{a\lambda}$, the spectral radiant emittance of the atmosphere at temperature, T_a . The first term on the right side of (22.5) is the radiant emittance of the sea surface to space, and the second term is the emittance of a cloudless atmosphere to space. The integration is carried out from the ocean surface to satellite height.

In the real atmosphere, clouds and aerosols attenuate the radiant emittance of the ocean surface and emit radiation as well. Realistic radiative transfer models of the combined effects of clouds, aerosols and gaseous absorbers are needed, because at present experimental results do not agree with theoretical transfer calculations based on pure gaseous atmospheres (Kuhn, 1963, 1972) (Kuhn and Stearns, 1971, 1972).

At infrared wavelengths, molecular absorption by atmospheric gases is important. The active absorbing gases in the infrared are H_2O , CO_2 , N_2O , and O_3 and other minor absorbers. Where one or more of these gases are strongly absorbing in the infrared, the atmosphere is essentially opaque. A region of weak absorption by these gases is called an atmospheric window.

As shown in (F22.7) these regions are difficult to define because the wings of the absorption bands cause partial absorption in the upper and lower wavelength portion of the window. Satellite systems for sensing sea-surface temperature make use of these atmospheric windows. The wavelength region of the window is of primary consideration in selecting the wavelength region of satellite IR detectors of sea-surface temperature. Table (22.2) shows that the 3.4-4.2, 8-12, and 10.5-12.5 μm regions have been employed successfully in satellite IR sensors. A second consideration in selecting sensor wavelength is the amount of radiant energy in that portion of the spectrum. As previously discussed, the amount of energy radiated by a blackbody surface is given by Planck's function (22:1) and is shown in (F22.8). Clearly, the 8-12 μm window region has the advantage of maximum emitted energy at temperatures near 300 °K. In the 3.4-4.2 μm window, roughly comparable amounts of reflected sunlight and emitted radiation are

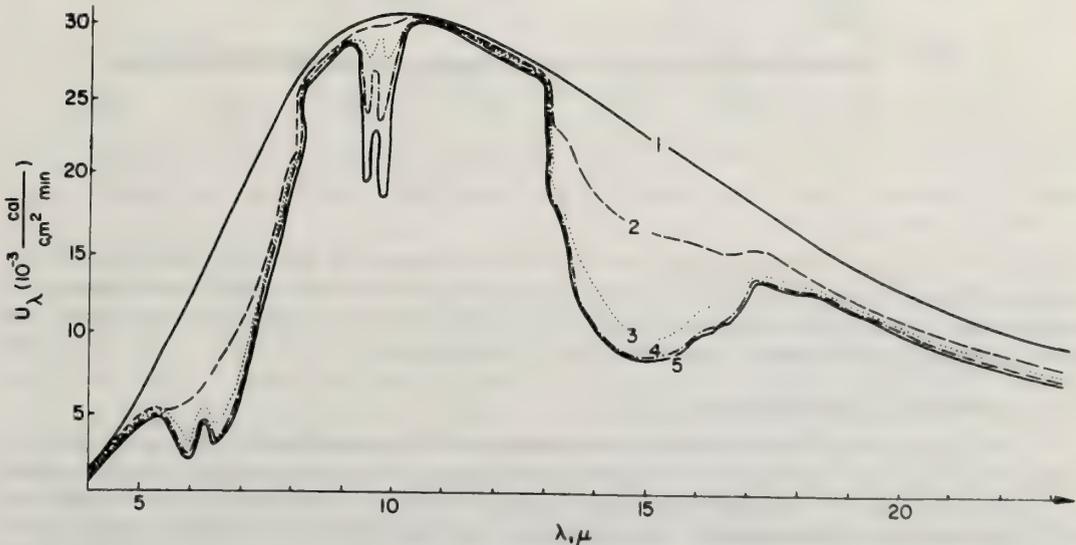


Figure 22.7 Spectral intensities of the up going radiation from zenith at different levels in the atmosphere at 2000 hr., Nov. 14, 1963, Rostov-on-Don. (1) at the level 1,000 mb (100 m); (2) 575 mb (4, 600 m); (3) 275 mb (10,000 m); (4) 75 mb (18,000 m); (5) 17.5 mb (28,000 m) (after Kondratyev, 1969).

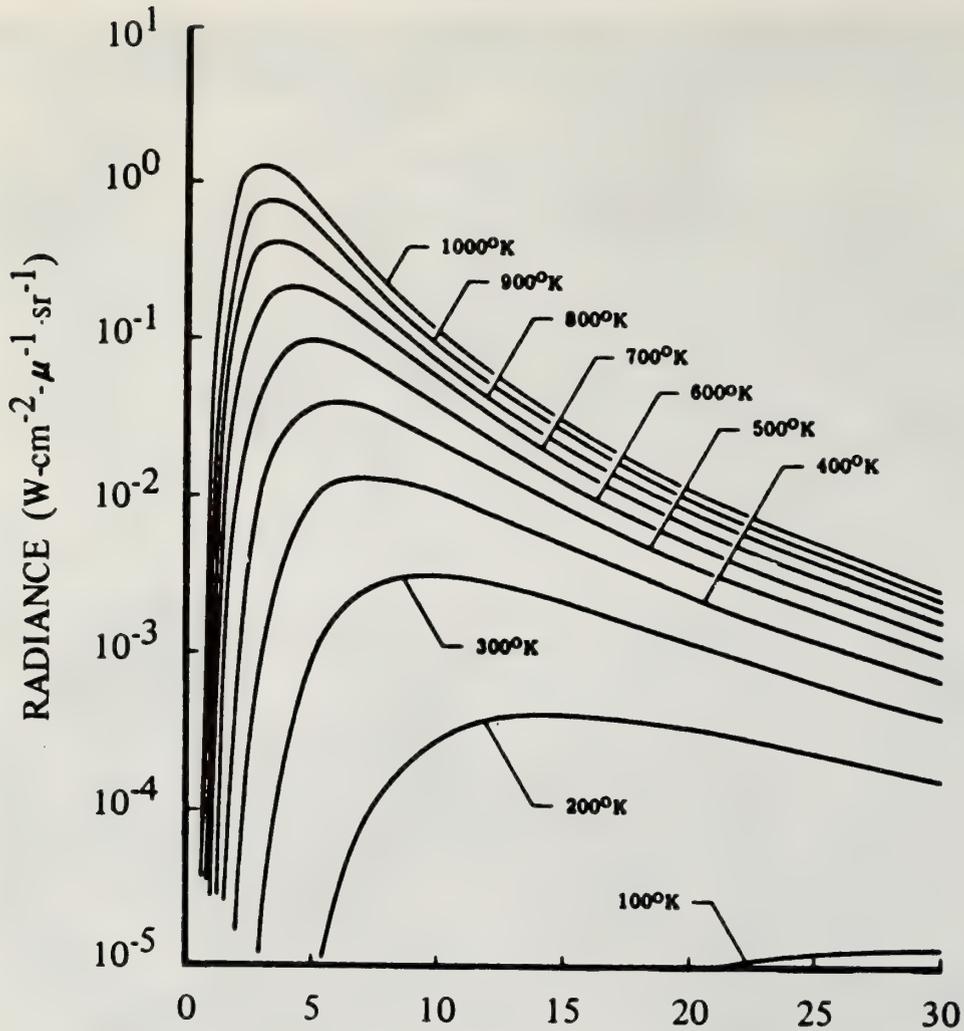


Figure 22.8 Spectral radiance from a blackbody versus wavelength, in micrometers, with temperature as a parameter (after Wolfe, 1965).

present in the daytime. This is apparent from (F22.9) which compares these two radiation sources for various reflectivity and emissivity values. Because of this reason the NIMBUS HRIR 3.4-4.2 μm data are contaminated with sunlight during the day-side of the orbit. ITOS-I and NOAA-I have side stepped this problem of solar contamination by using the 10.5-12.5 μm wavelengths.

22.1.3 Atmospheric Correction Models

In order to obtain accurate sea-surface temperature information from satellite IR window data, it is necessary to correct the measurement for the effect of the atmosphere on the upward IR radiance from the sea surface. That effect is to absorb and re-emit a portion of the sea surface radiance. To determine the error introduced by the atmosphere, some investigators have made theoretical calculations with radiative transfer models of the upward spectral radiance in the atmosphere. They have examined the error under various (a) assumed model atmospheres with varying temperature and water vapor and (b) zenith angles. They have incorporated the significant absorbing gases. Two such studies have been reported in the past year; these will be discussed here. To my knowledge, they are the most useful attempts at solution of the atmospheric problem which have been reported to date.

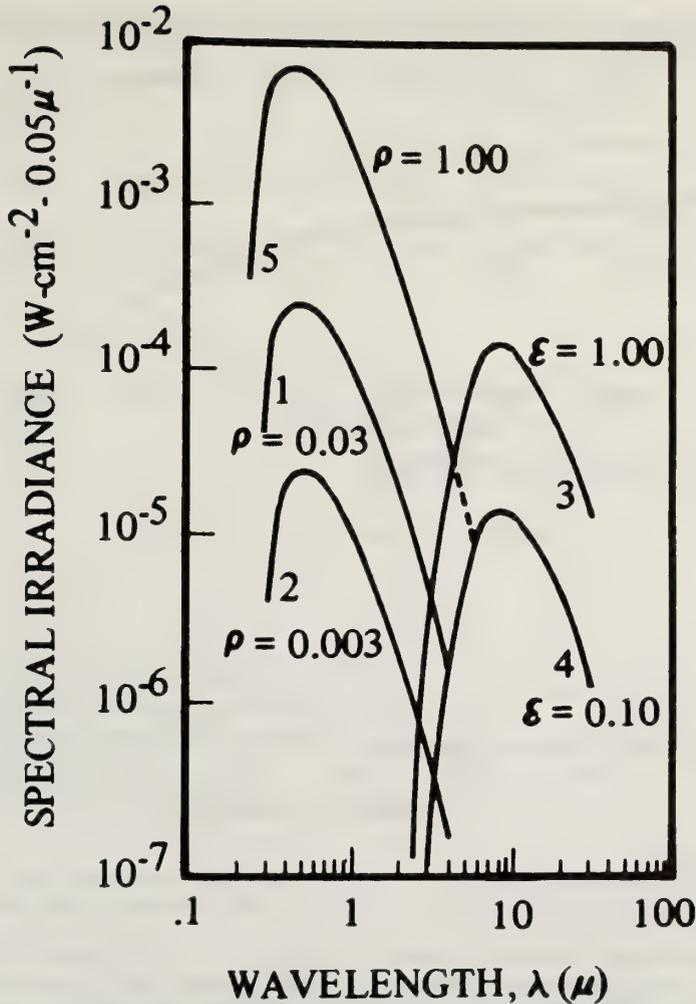


Figure 22.9 Comparison of reflected sunlight and emitted radiation for various emissivity, ϵ , and reflectivity, ρ , values and a surface temperature of 300°K (after Holter, 1970).

22.1.3.1 Single Band Model

Smith, et al. (1970) have used a radiative transfer model to calculate the upward radiance at $3.8\ \mu\text{m}$. The model accounts for H_2O and CO_2 absorption and emission. They have utilized three model atmospheres (polar, midlatitude, and tropical) with total water vapor amounts 0.15, 1.2, and 4.7 cm., respectively. Using this model they have calculated the temperature difference between the surface and the effective radiative temperature at satellite height. The resulting temperature differences are shown in (F22.10) as a function of zenith angle (θ) for each model atmosphere. It is clear that the temperature error increases with increasing zenith angle. The tropical atmosphere has a greater error than the polar atmosphere. For these calculations the atmosphere was assumed to be cloud-free.

Smith et al. (1970) also obtained calculations in which the radiating surface was raised from 1000 to 300 mb for the three model atmospheres. His calculations simulate what may be expected with clouds (ground) which have an emissivity of unity and completely fill a radiometer field of view. The results of these calculations with clouds which vary in height and latitude are shown in (F22.11). The temperature differences (surface minus satellite) are plotted as a function of satellite observed temperature. It is apparent that a high correlation exists between the temperature difference and satellite observed

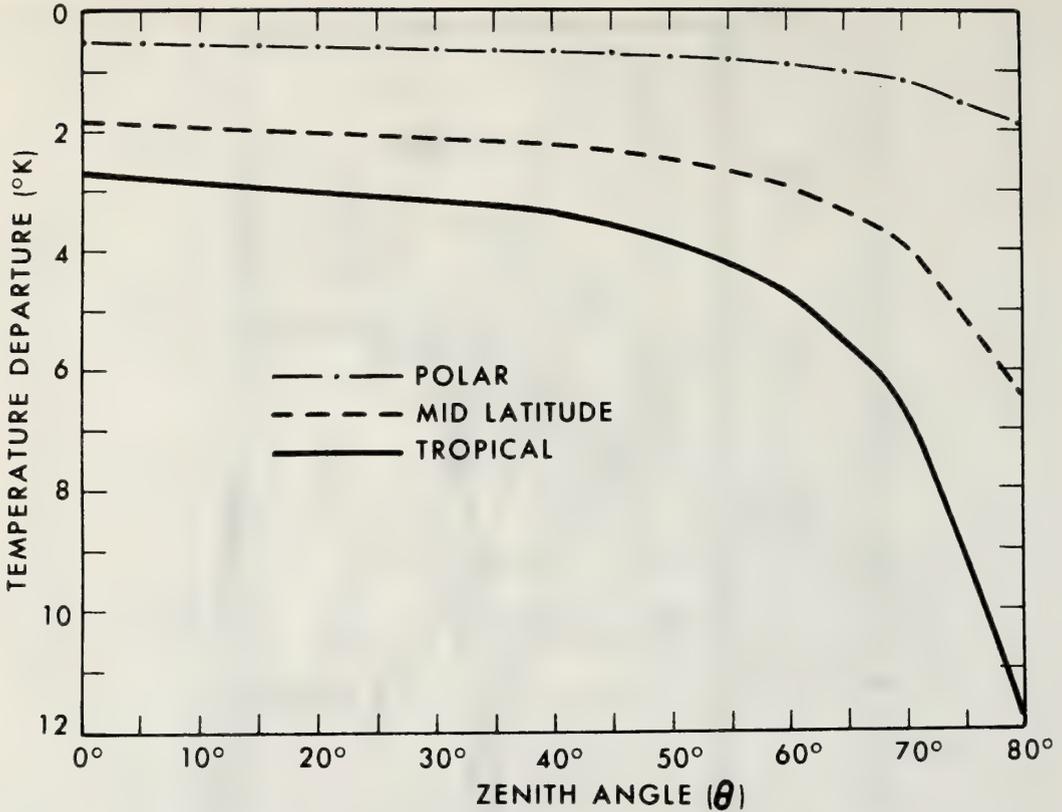


Figure 22.10 Calculated departure of NIMBUS HRIR ($3.8 \mu\text{m}$) temperature from ground surface temperature as a function of local zenith angle for three model atmospheres (after Smith, et al., 1970).

temperature, when the height of the emitting clouds is varied from 1000 to 300 mb and the latitude is varied by the use of these 3 model atmospheres. Smith, et al., (1970), indicate "this is due to the fact that water vapor and carbon dioxide attenuation is naturally correlated with the surface radiating temperatures since it decreases with increasing cloud height and increasing latitude." They point out that this characteristic of the atmosphere makes it possible to predict the surface to satellite temperature difference from only two parameters: the satellite observed temperature; and the zenith angle of the measurement. They have presented this relationship for satellite observed temperatures from 210 to 300 °K and for zenith angles of less than 60° as

$$\Delta T = [a_0 + a_1(\theta/60^\circ)^{a_2} \ln(100/300 - T_{s0})] \quad (22:6)$$

where: $a_0 = 1.13$, $a_1 = 0.82$, $a_2 = 2.48$, θ is the local zenith angle of measurement, and T_{s0} is the satellite observed temperature in °K.

Although it would have been possible to estimate probably accuracy of this single band model using dependent data, the authors have not given an error figure. A summary of the important features of the single band model is given in (T22.3).

22.1.3.2 Double Band Model

The other recent work with an atmospheric correction model is that of Anding and Kauth (1970). They have used a radiative transfer model, described by Anding and Kauth (1969), capable of including

Table 22.2 U.S. Satellites with IR Window Sensor Experiments

Satellite	TIROS II, III,IV,VII	NIMBUS I	NIMBUS II	NIMBUS III	ITOS-1 (TIROS-M)	NIMBUS IV	NOAA1 (ITOS-A)	ITOS-D
Launch	1960-63	8/28/64	5/15/66	4/14/69	1/23/70	4/8/70	12/11/70	Planned 7/72
Orbit Inclina- tion and Altitude (n. mi.)	48-58° 350	Sun Syn. 400	Sun Syn. 600	Sun Syn. 600	Sun Syn. 790	Sun Syn. 600	Sun Syn. 790	Sun Syn. 790
IR Window Sensors	MRIR ²	HRIR ¹	HRIR ¹ MRIR ²	HRIR ¹ MRIR ²	SR ⁴	THIR ³	SR ⁴	SR ⁴ VHRR ⁵
Resolution (n. mi.)	20	5	5 30	5 30	4	4	4	4 0.5
Sensor Spectral Response (μm)	8-12	3.4-4.2	3.4-4.2 10-11	3.4-4.2 10-11	10.5-12.5	10.5-12.5	10.5-12.5	10.5-12.5

- 1 - High Resolution Infrared Radiometer
- 2 - Medium Resolution Infrared Radiometer
- 3 - Temperature-Humidity Infrared Radiometer
- 4 - Scanning Radiometer
- 5 - Very High Resolution Radiometer

Table 22.3 Summary of Atmospheric Correction Models

Model	Band	Theoretical Radiative Transfer Model			Model Atmosphere			Error
		Assumes	Accounts for	Zenith Angle(°)	Temperature Profile	Water Vapor Profile	Range of Cloud Height	
Single band (Smith et al., 1970)	3.8 μm	Surface Emissivity of Unity	H ₂ O CO ₂	0-60	Polar Midlatitude Tropical	0.15 cm 1.2 cm 4.7 cm	Surface to 300 mb.	Not given
Two band (Anding and Kauth, 1970)	9.1 and 11.0 μm	Cloudless	H ₂ O CO ₂ O ₃ NO ₂	0 60 75	Summer Winter Mean 30°N	Wet Dry Mean 30° N	Clouds not included	±0.15°C (on depen- dent data)

water vapor, carbon dioxide, ozone, methane, nitrous oxide, and haze. In the computations, atmospheric gas concentrations were held constant and only water vapor and temperature were varied through the use of five model atmospheres, indicated in (T22.4). The temperature and water vapor profiles for these model atmospheres are shown in (F22.12) and (F22.13).

Calculations were carried out for all combinations of (a) five model atmospheres, (b) three zenith angles (0, 60, and 75°), and (c) five ocean surface temperatures (280, 285, 290, 295, and 300 °K). A total of 75 spectra were calculated. An example of the spectral radiance calculation for the summer wet model atmosphere and zenith angle of 60° is shown in (F22.14). Each spectral radiance curve in (F22.14) is that of a greybody water surface, at the temperature indicated, modified by the summer wet model atmosphere. Emissivities were based on Buettner and Kern (1965). The 0 °K surface temperature was included by Anding and Kauth (1970) to show the emission of the atmosphere alone. In the intense absorbing regions, the spectra are the same regardless of the surface emission because the atmosphere is opaque in these regions.

Atmospheric attenuation affects the radiance in both the 9.1 and 11.0 μm channels, although the effect on the radiance at 9.1 μm is greater. Since a high correlation exists between the two spectral radiance values, Anding and Kauth (1970) have plotted the 9.1 μm radiance versus the 11.0 μm radiance for various ocean surface temperatures (F22.15). The results show that given the radiance in these two spectral bands, the surface temperature can be predicted with high accuracy, providing the viewing angle, sea temperature, and atmospheric conditions are similar to one of the 75 combinations tested. The surface temperature is simply interpolated between the isothermal lines in (F22.15).

An estimate of error in ocean temperature determination has been obtained using the 75 radiance pairs (which are dependent data, i.e., correlated) to predict the surface temperature. The rms error quoted by Anding and Kauth (1970) is ±0.15 °C, which is very encouraging for oceanography.† It should be remembered, however, that this two channel model does not include the effect of clouds. Anding and Kauth (1970) have indicated they are working on the addition of clouds in their model; this will require the addition of one or more spectral bands. They have indicated also the method of regression coefficients is unworkable if the cloudiness exceeds 25 percent.

Table 22.4 Summary of Model Atmospheres

(after Anding and Kauth, 1970)

No.	Model Atmosphere Name	Temperature Profile	Water Vapor Profile
1.	Summer Wet	Summer	Summer Wet
2.	Summer Dry	Summer	Summer Dry
3.	Winter Wet	Winter	Winter Wet
4.	Winter Dry	Winter	Winter Dry
5.	Mean, 30° N.	Mean	Mean

22.1.4 Satellite Experimental Methods

In attempting to remove the effects of clouds and the atmosphere in obtaining satisfactory data on sea surface temperatures, investigators have employed several simple statistical techniques on the data. In some cases the investigators have not evaluated the error of the technique, perhaps because of lack of suitable ground truth data. In these cases, the results are presented here to show the history of these attempts at obtaining useful data.

†An error of 0.15°C is reasonable only if the absorption coefficients for water vapor are known accurately. A realistic error is undoubtedly higher than this figure.

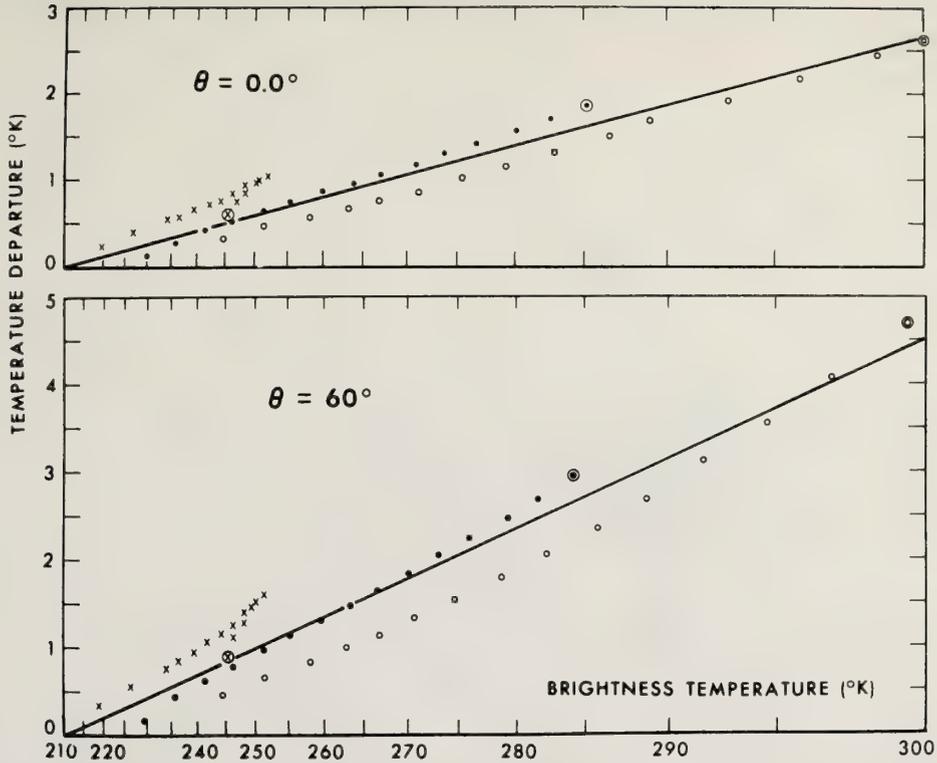


Figure 22.11 Departure of NIMBUS HRIR brightness temperature from ground and cloud radiating temperature as a function of "observed" brightness temperature. The crosses, dots, and open circles pertain to the polar, midlatitude, and tropical model atmospheres, respectively (after Smith et al., 1970).

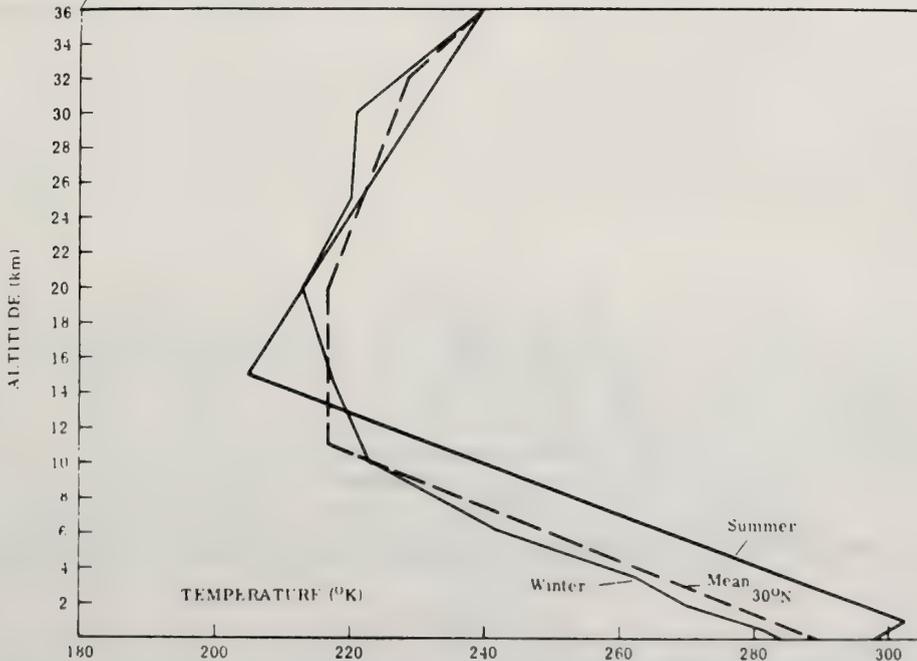


Figure 22.12 Temperature profiles for model atmospheres used for double band model radiance calculations (after Anding and Kauth, 1970).

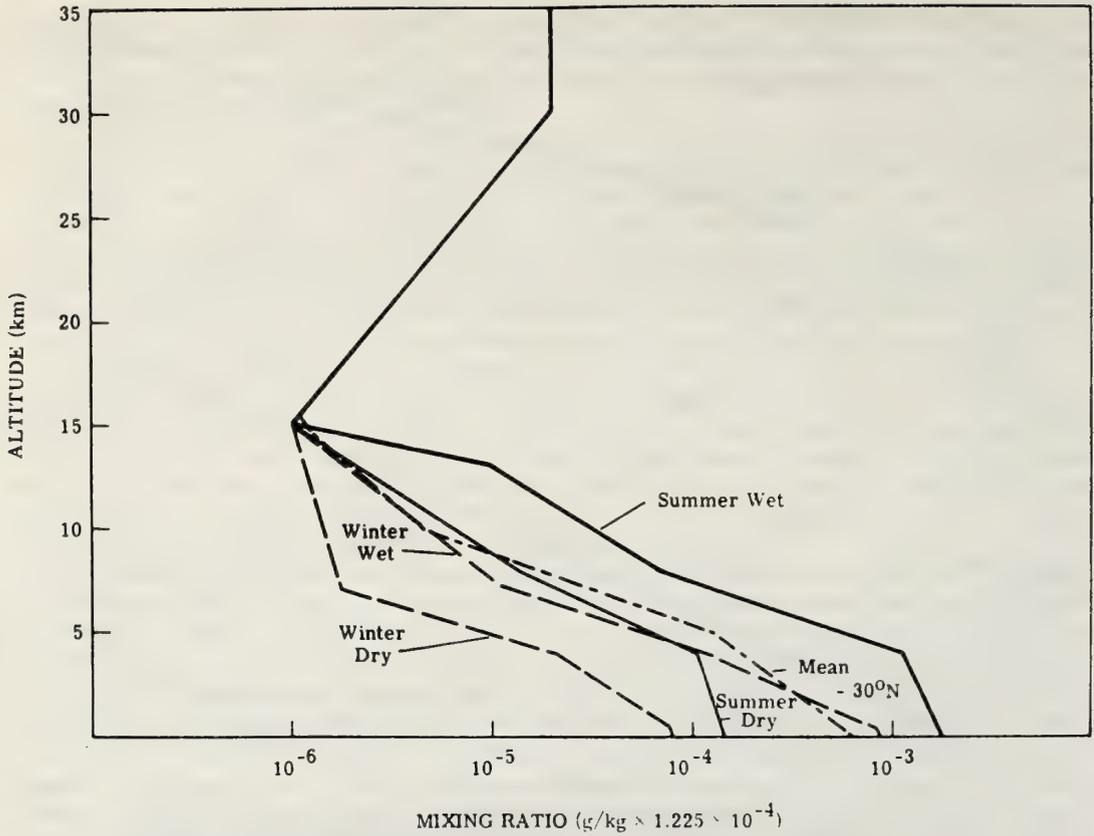


Figure 22.13 Water vapor profiles for model atmospheres used for double band model radiance calculations (after Anding and Kauth, 1970).

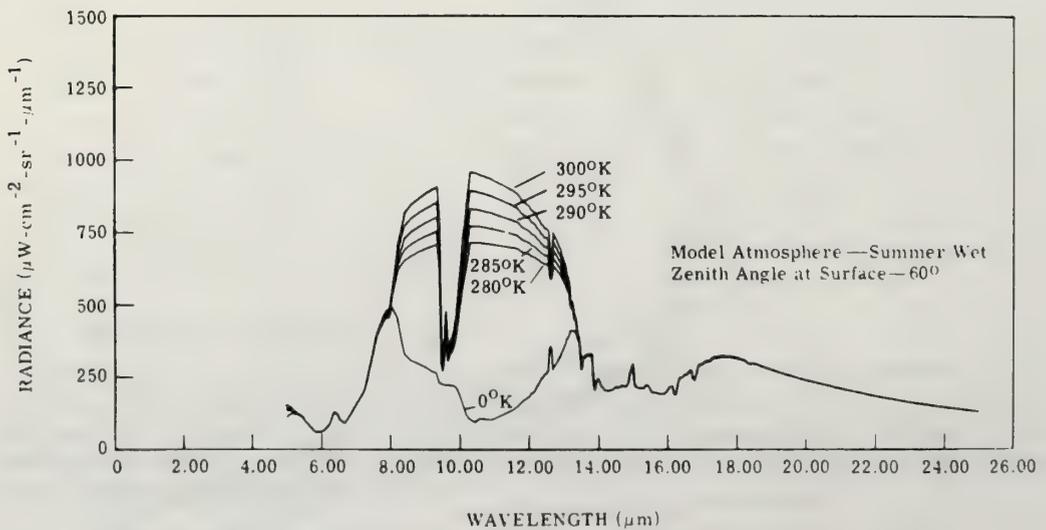


Figure 22.14 Upward spectral radiance at 100 Km with sea temperature as a parameter (after Anding and Kauth, 1970).

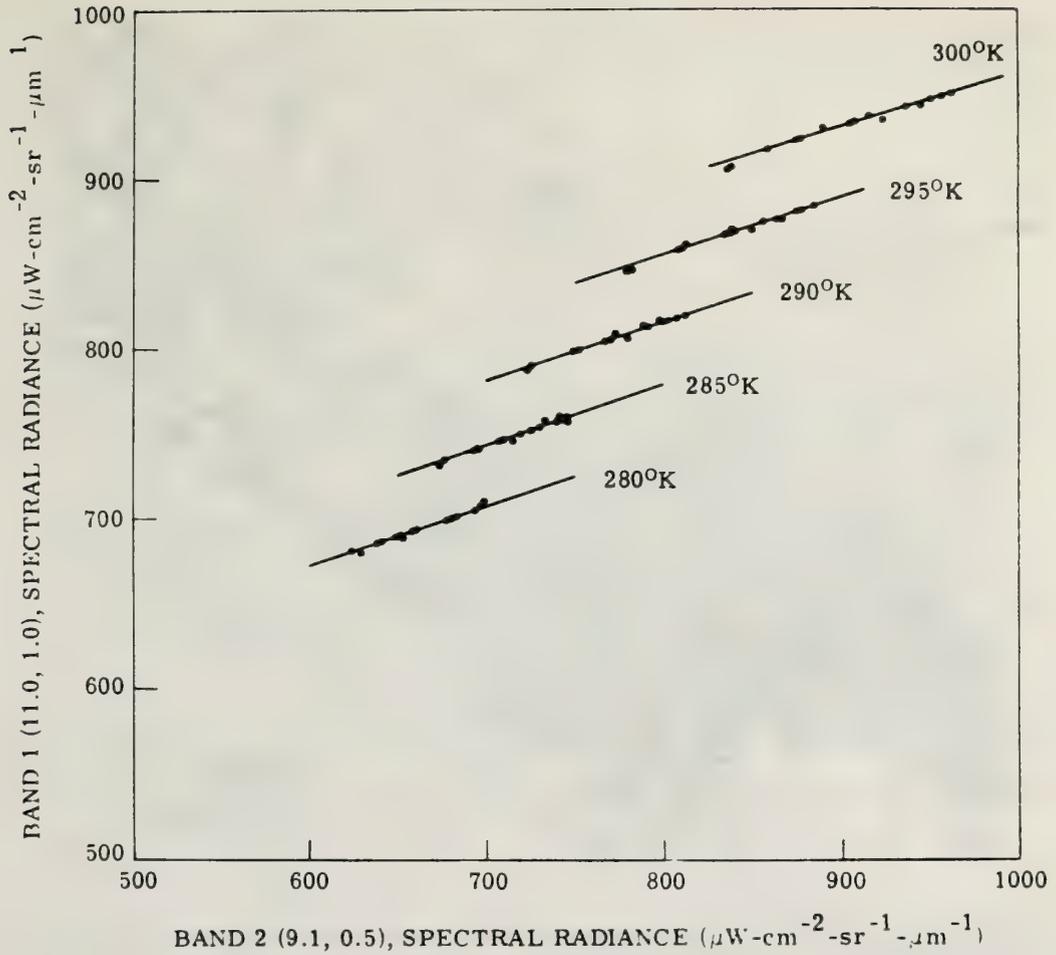


Figure 22.15 Spectral radiance at $9.1 \mu\text{m}$ for various ocean surface temperatures (after Anding and Kauth, 1970).

22.1.4.1 Data Averaging Techniques

Some of the early experimental data have shown much detailed structure in the temperature pattern; this was believed to result from sensor noise.† Some examples of this fine scale structure are shown in (F22.3) and (F22.16). In an attempt to remove this noise, Curtis and Rao (1969) have averaged NIMBUS II HRIR data over one-half degree latitude/longitude rectangles. Beginning with the data shown in (F22.16), Curtis and Rao (1969) have applied an averaging technique and arrived at the smoothed temperature field shown in (F22.17). The Gulf Stream boundary is preserved in the smoothed analysis as is the strong gradient between the cold, cloudy region and the warm Gulf Stream. The influence of clouds is not removed by simple data averaging, although sensor noise is suppressed. Others have also used simple averaging techniques.

†E. J. Williamson (unpublished manuscript, 1968) has shown from NIMBUS II HRIR housing scan data that the noise-equivalent temperature is about 2°K (at $280\text{-}300^\circ\text{K}$) (cited by Curtis and Rao, 1969).

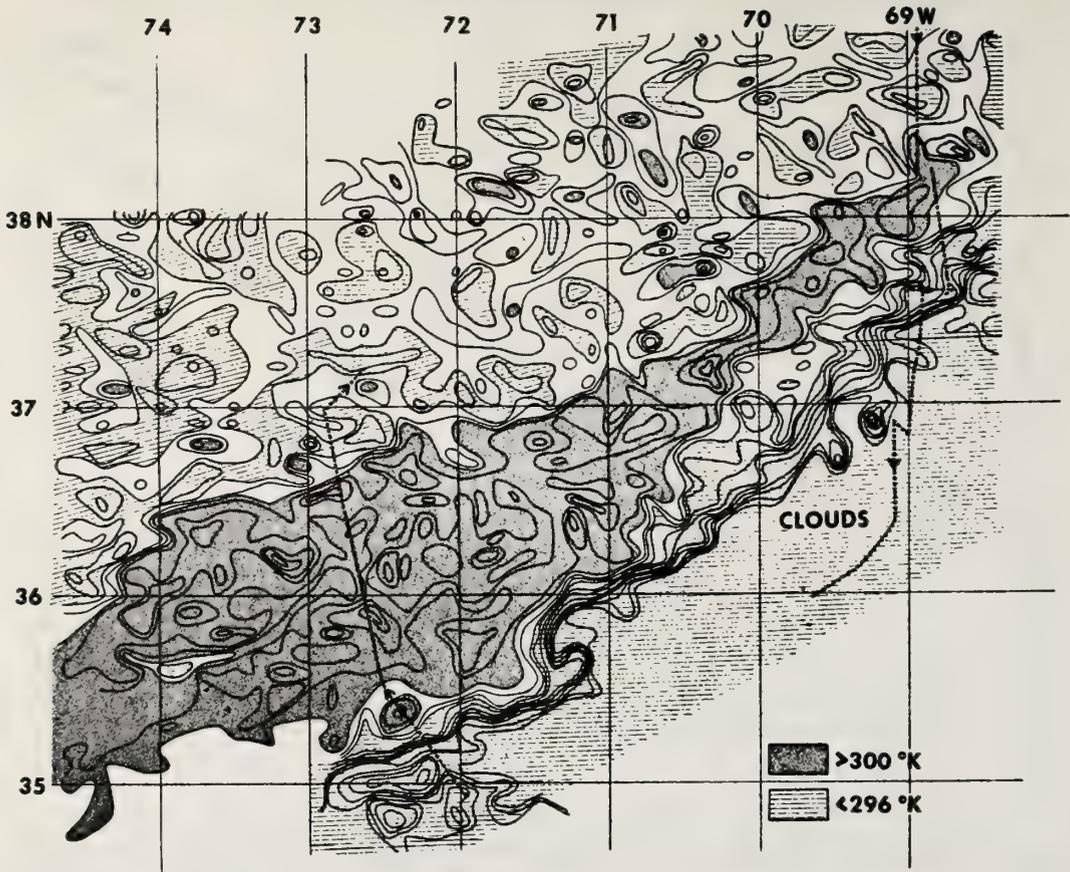


Figure 22.16 The Gulf Stream from the analyses of digitized HRIR data from NIMBUS 2, orbit 504, June 22, 1966, 0430 UT (after Curtis and Rao, 1969).

22.1.4.2 Highest Daily Value Technique

Because clouds tend to reduce satellite-observed sea temperatures, a technique has been used by a number of investigators to select only the highest temperature in a given area as representative of the surface temperature. Curtis and Rao (1969) have used this method on the data set in (F22.16); the results are illustrated in (F22.18). Temperatures are 2-4 °K higher than with the averaging technique (F22.17) and probably indicate a sensor noise of this magnitude. In the boundary of the cloudy region of (F22.17), the temperature has increased as much as 13 °K by using the highest value technique. Apparently most measurements representing clouds are eliminated by this method, depending on the degree of cloudiness. A part of the cloudy area in (F22.17) still remains in (F22.18).

Selecting the highest reading in a region practically guarantees that detector noise will make the reading higher than the true temperature. In addition, if the detector noise is 2-4 °K, the sea surface temperatures are of little value other than for climatological studies.

LaViolette and Chabot (1969)†, used a similar highest daily value technique. HRIR scan spot data within the range of climatological sea surface temperature were averaged over 2,205 square mile areas. Data outside the temperature range were excluded from the average. For each day this included about 10 scan spots per area. A five-day composite was obtained by selecting the highest daily average for each area. The result is shown in (F22.19). The basic idea of this technique is that clouds are transient and sea temperature is horizontally homogeneous and relatively constant with time. Thus, the likelihood is small that clouds would cover the same area for many days. This region, Baja California, is one of the most cloud-free regions on earth—so one would expect optimum results there.

†Cited by LaViolette and Seim (1969).

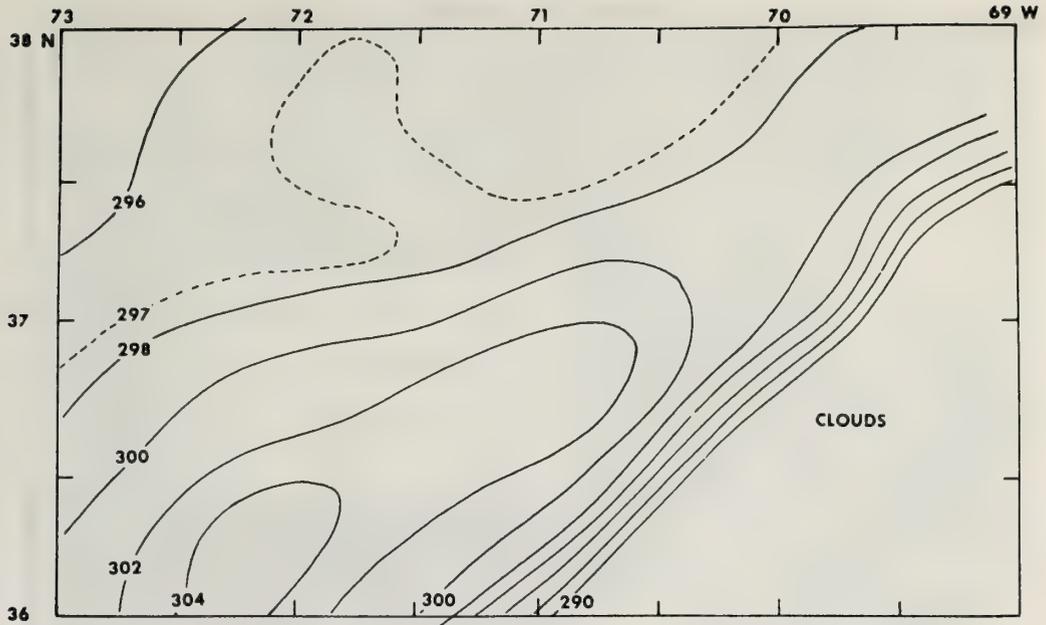


Figure 22.17 Sea-surface temperature analysis made by using the average values in each half-degree latitude/longitude box from NIMBUS 2 HRIR data, orbit 504, June 22, 1966, 0430 UT (after Curtis and Rao, 1969).

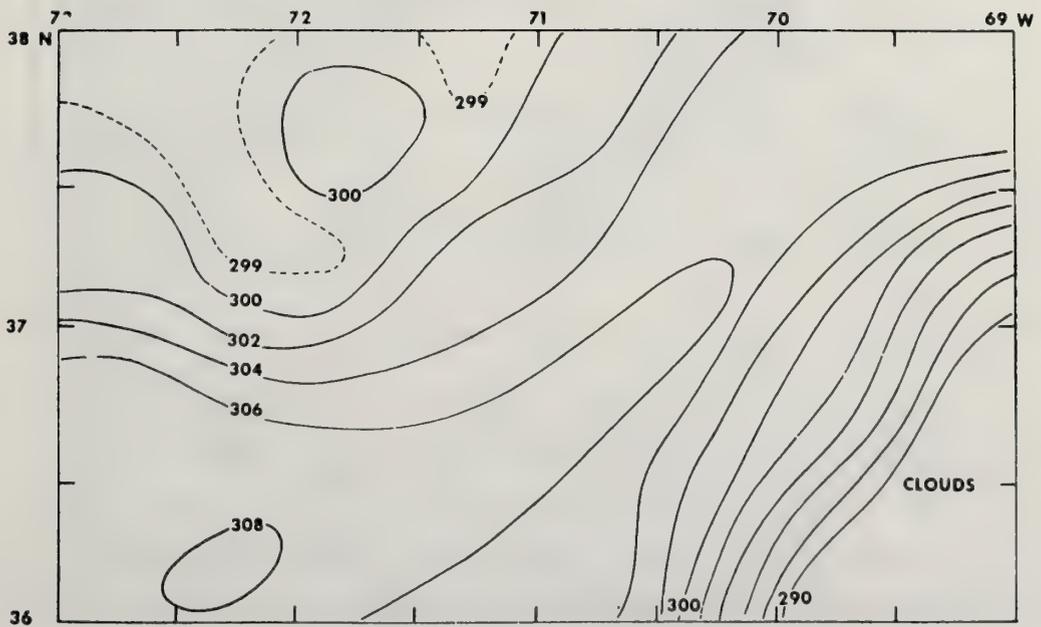


Figure 22.18 Sea-surface temperature analysis made by using only the highest value in each half-degree latitude/longitude box from NIMBUS 2 HRIR data orbit 504, June 22, 1966, 0430 UT (after Curtis and Rao, 1969).

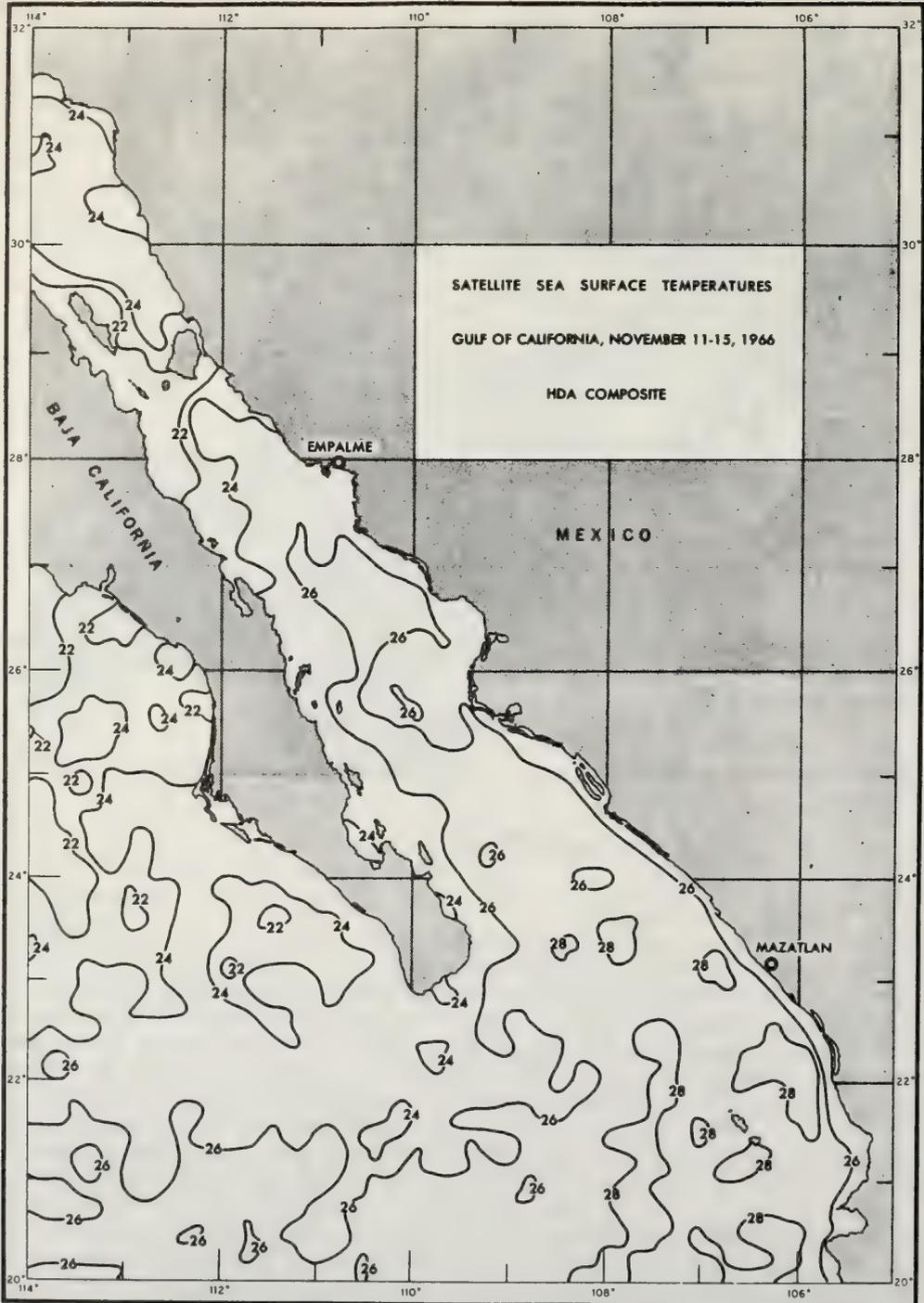


Figure 22.19 Highest temperatures of a 5-day period of sea-surface temperatures derived from NIMBUS II HRIR data, Nov. 11-15, 1966 (after LaViollette and Chabot, 1969).

Smith, et al. (1970) have used a similar maximum temperature technique on NIMBUS III HRIR data over the BOMEX area near Barbados. In that study the highest satellite observed temperature was selected in each one degree latitude/longitude square. Results are shown in (F 22.20). Ship ground truth sea surface temperature data are given by an asterisk. The investigators have commented that "the highest temperatures are extremely spatially unstable due to cloud and instrument noise. As a result, they do not provide any useful information on the true sea-surface temperature distribution."

22:1.4.3 Histogram Technique

The basis of the histogram technique lies in the fact that a HRIR sensor has random noise which will give an otherwise constant signal a Gaussian distribution on a frequency of occurrence diagram. An example of the use of this technique is given in (F22.21) by Smith, et al. (1970) based on NIMBUS HRIR data. These diagrams show the frequency of occurrence of satellite observed temperature for two areas, each with dimensions of 2.5 x 2.5° latitude/longitude. The warmest side of the distribution curve is least affected by clouds and is somewhat Gaussian in shape because of the random sensor noise. The investigators have indicated that at one standard deviation the standard error is 1.5 °K for the NIMBUS HRIR sensor at these temperatures. The most probable sea-surface temperature for both histograms (F22.21) is 302.5-1.5 = 301.0 °K. When this technique is applied to the previously mentioned NIMBUS data for the BOMEX area, the set of temperatures in (F22.22) is produced. As can be seen by comparing the 5 BOMEX ship observations with the satellite derived surface temperatures, these data are in much closer agreement than resulted from the high temperature technique illustrated in (F22.20).

	60W				55			52W
18N	299	302 301(RAI)	301	301	302	*301(OCE) 302	301	299
	* 300	303	302	300	300	300	299	296
	300	300	300	301	294	302	301	305
15N	303	304 *301(ROC)	303	307	305	304	306	303
	304	306	306	304	303	304 301(DIS)	*304	303
	303	309 *302(MIT)	310	302	304	306	305	304
	305	305	304	296	298	305	300	304
	304	302	303	302	267	269	300	306
10N								

Figure 22.20 Highest brightness temperatures in each latitude/longitude box from NIMBUS 3, HRIR, nighttime, orbit 681, on June 4, 1969 (after Smith et al., 1970).

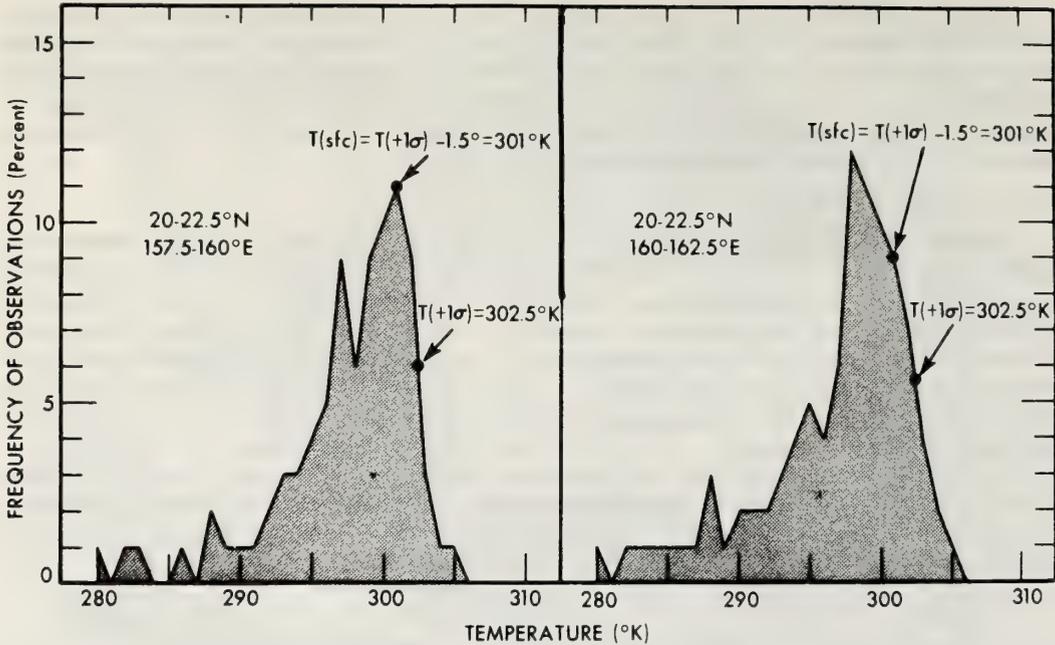


Figure 22.21 Sample histograms of corrected NIMBUS HRIR data showing their relation to sea-surface temperature (after Smith et al., 1970).

To test the histogram technique on additional data, Smith, et al. (1970) have derived sea temperature data for the North Pacific from HRIR data and compared them with ship reported temperatures. The rms difference of 97 observations was 1.71°K . The scatter diagram of the comparison is shown in (F22.23). The relative discrepancy between satellite and ship data is less than 1°K .

22.1.5 Cloud Problem

Although the data averaging technique does not remove the influence of clouds on the upwelling spectral radiance, both the high temperature technique and the histogram technique remove the influence of clouds to some extent, depending on the nature of the cloudiness.

In addition, it was mentioned in the discussion on the two-band model 22.1.3.2 for correcting atmospheric effects that there is a possibility of correcting for certain cloudiness with the addition of one or more spectral bands, perhaps in the visible portion of the spectrum. This is probably the best hope for removing the influence of clouds to the extent that absolute sea temperature can be obtained to $\pm 0.5^\circ\text{K}$.

This section was written while at sea in the sub-tropics on a cruise of three weeks. During that time I made a special effort to observe the horizontal spacing of trade cumulus, and smaller fragments of these clouds, relative to the scan spot size of IR sensors. It is my personal feeling that the concept of "looking between the clouds" with a 4 n mi resolution satellite radiometer is close to pure fiction for many areas of the tropics; even with a 0.5 n mi resolution sensor the occurrence of truly cloud-free data will be extremely rare. Perhaps there are regions over the earth's oceans which are systematically cloud free, but it is my impression that over most of the tropical and sub-tropical oceans trade cumulus and small cloud fragments exist a high percentage of the time. With present techniques it is not possible to remove the effect of these clouds which are much smaller than the field of view of the satellite sensor. Their effect must be evaluated.

Many of the initial problems of remote sensing of sea surface temperature have been solved by the investigators whose work has been summarized here. The cloud problem has been partially solved but is still a serious limitation on achieving the accuracies needed for oceanography. There is a continuing need for theoretical studies of radiative transfer in realistic atmospheres including both clouds and aerosols, and for experimental studies to evaluate the influence of clouds, particularly those smaller than the radiometer field of view, on the spectral radiant emittance of the earth and atmosphere.

	60W				55			52W
18N	296	298	298	298	299	*301 (OCE) 298	297	C
	301 (RAI) *298		298		298		C	
	297	298	299	297	297	296	C	C
	C	C	296	297	C	C	C	C
15N	C		299		C		298	
	C	299	299	300	300	C	301	299
	*301 (ROC)							
	300	303	302	303	301	301	*301 301 (DIS)	300
	300		300		301		300	
	300	302	C	C	C	C	301	C
	*302 (MIT)							
	C	C	C	C	C	301	C	C
	C		C		C		C	
10N	299	300	C	C	C	C	C	C

Figure 22.22 Surface temperatures, °K, derived from NIMBUS 3 HRIR nighttime data orbit 681 on June 4, 1969. Centered values in each box were derived from 1° latitude/longitude histograms, underlined values from 3° histograms. Asterisks are BOMEX ship observations and C denotes indeterminate temperatures due to excessive cloudiness (after Smith et al., 1970).

22.2 Passive Microwave Detection

Passive microwave detection is clearly distinguished from active microwave detection. An active microwave system, such as a radar, emits microwave frequency energy which is reflected from natural surfaces; the characteristics of this reflected energy can be used to identify the properties of the surfaces from which it was reflected or backscattered. Passive microwave detectors measure the intensity of microwave energy emitted by natural surfaces.

Present microwave technology evolved with the science of radio astronomy. The first successful measurement of radio frequency thermal radiation from an extra-terrestrial source (the Milky Way Galaxy) was obtained in 1931 using a radio receiver at 20.5 MHZ and 30 MHZ. Radar research during and after World War II has provided many studies of the propagation of radio waves, microwave absorption and scattering in the atmosphere, emission of the sea surface, and improved technology of microwave radiometers. An excellent summary of the historical development of microwave radiometry is given by Paris (1969).

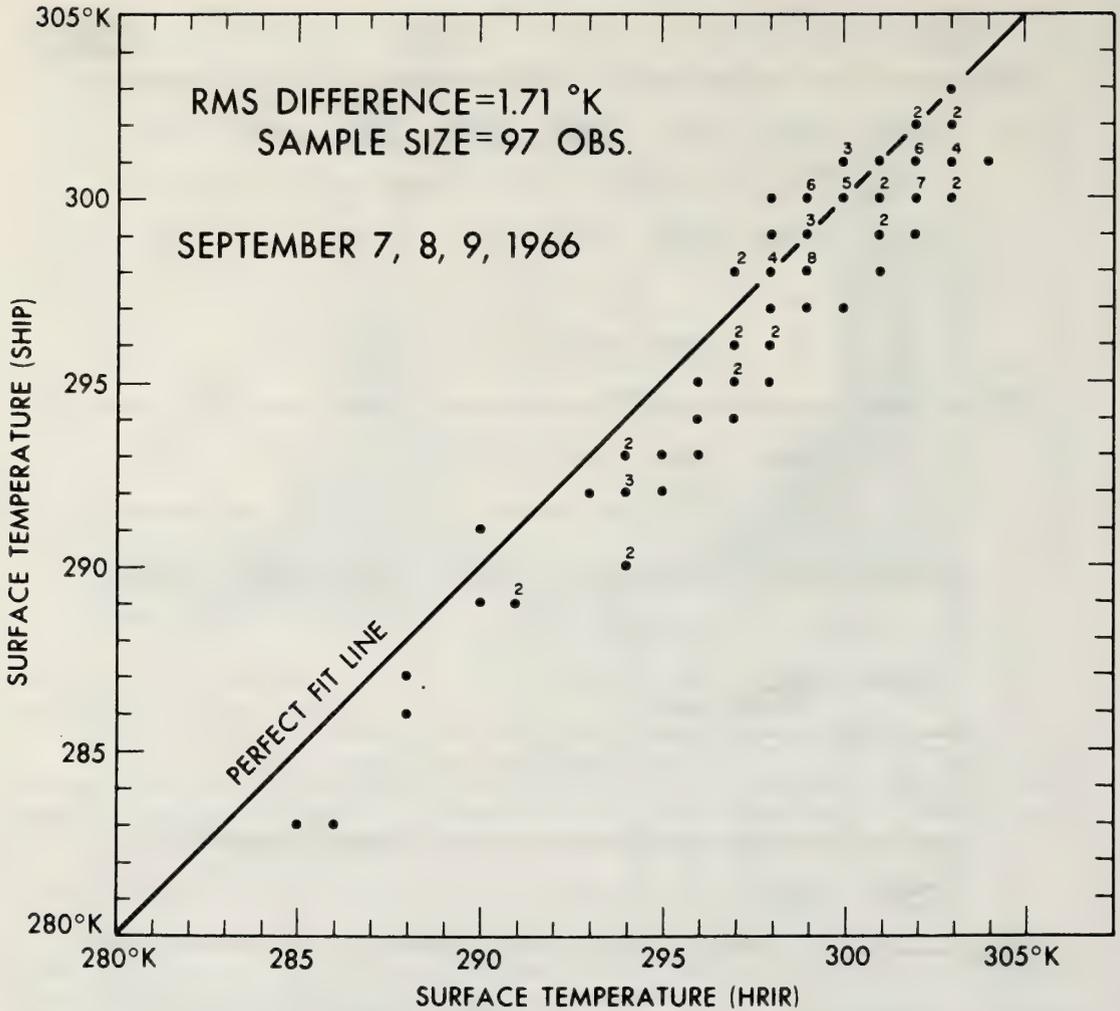


Figure 22.23 Scatter diagram of surface temperatures obtained from ships and inferred from NIMBUS II HRIR data (after Smith et al., 1970).

The basic problem in passive microwave remote sensing is to measure the microwave radiance of the ocean and atmosphere (N) and to infer from it the physical state of these media. In addition to temporal variation, there are three useful detection parameters which provide data for interpretation. These are microwave frequency, ν , both horizontal and vertical polarizations, P , and nadir angle, Ψ .

A number of oceanic and atmospheric variables control the upward microwave radiance, $N(\nu, P, \Psi)$; they are sea-surface temperature, T_s , salinity, S_s , roughness, R_s , and foam, F_s . When present, sea ice provides a different microwave emittance than that of the sea surface. Atmospheric water vapor, WV_a , liquid water, LW_a , and molecular oxygen, MO_a , also modulate the upwelling microwave emission from the sea and atmosphere. In parametric form this may be written as,

$$N(\nu, P, \Psi) = f(T_s, S_s, R_s, F_s, LW_a, WV_a, MO_a) \tag{22:7}$$

Figure 22.24 shows how the oceanic and atmospheric variables depend on frequency, ν , alone. Since P and Ψ have not been specified in this figure, the curves are qualitative and are intended only to give the reader an indication of the microwave frequencies at which oceanic variables have the greatest effect on microwave emission. The temperature sensitivity curve, for example, is defined as the ratio of the change in microwave brightness temperature to the change in molecular temperature of the ocean surface, $\Delta T_B/\Delta T_S$. Because of the emissive properties of water, maximum temperature sensitivity occurs at a frequency of 5-6 GHz.†

In the case of salinity sensitivity, (F22.24), maximum sensitivity occurs at a frequency of about 1 GHz. At higher frequencies the microwave brightness temperature is practically insensitive to salinity variations.

Wind increases both the roughness and extent of foam coverage on the ocean surface, and both of these variables have an effect on the emitted microwave brightness temperature. (F22.24) shows that the greatest combined effect of roughness and foam occurs at the higher frequencies indicated, about 10 GHz and above, and that roughness and foam have a smaller effect at frequencies near 1 GHz. The effect of roughness and foam is one of the most difficult areas to interpret; some of the complications are discussed in 22.2.4.

Microwave emittance of water is much less than of thick ice, and the difference is only slightly dependent on frequency (F22.24). Studies of sea ice have shown that two characteristic types of sea ice have different emittances (Wilheit, et al., 1971). One type of relatively smooth, thin ice termed “first year ice” has a lower emittance than the second type described as “multi-year, heavily hummocked ice.” This may lead to some ambiguity in determining the percentage of ice cover by passive microwave detection, unless polarization measurements are obtained.

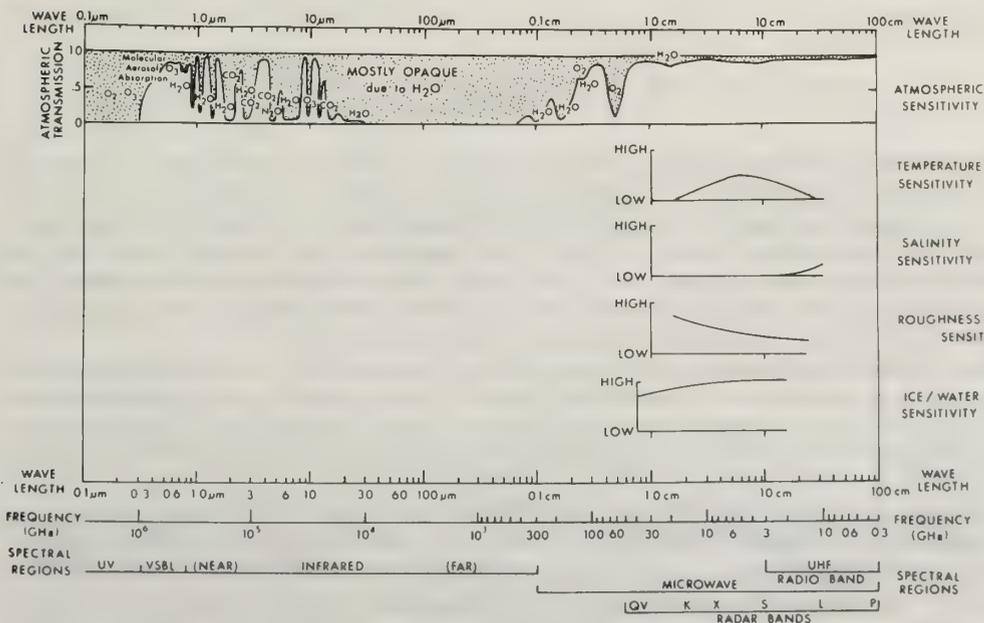


Figure 22.24 Quantitative atmospheric transmission curve, adapted from Huston, (1971), and qualitative relationships between microwave brightness temperature and various oceanic variables.

†GHz = 10⁹ cycles per second (one giga Hertz).

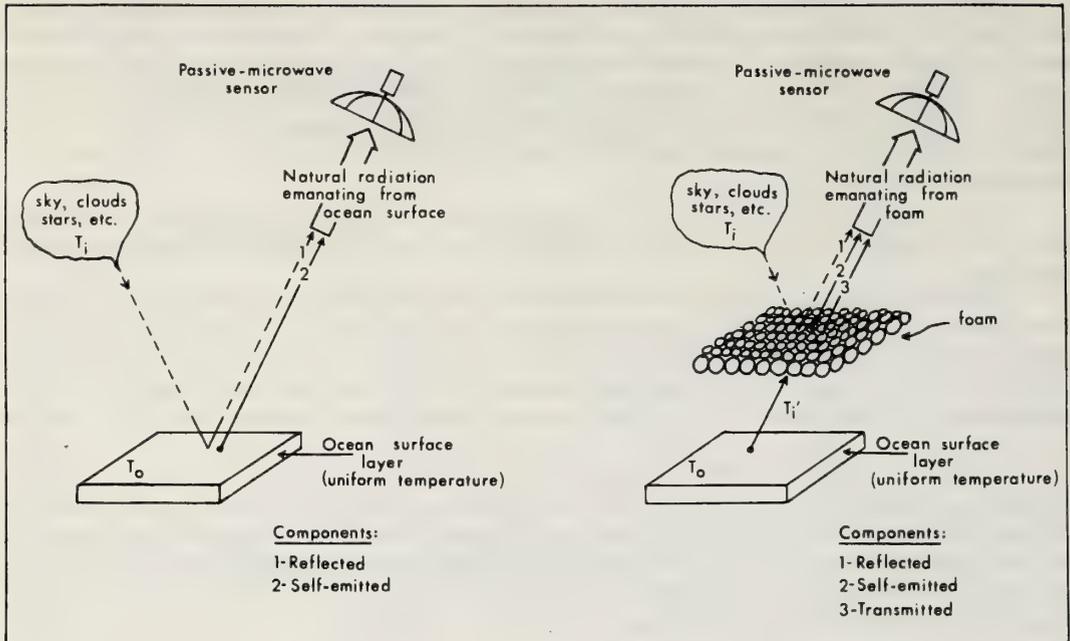


Figure 22.25 Natural microwave radiation emanating from a smooth ocean surface (left) and a foam covered ocean surface (right) (adapted from Holter, 1970).

The atmospheric transmission curve, (F22.24), shows the effects of water vapor and molecular oxygen on microwave transmission. An atmospheric window is located at 80-100 GHz and a strong molecular oxygen absorption band at 60 GHz. The highest frequency window in which satellite microwave measurements can be obtained presently is 30-45 GHz, because at 80-100 GHz flyable low-noise radiometers are not yet available (Huston, 1971). The window at 30-45 GHz is somewhat affected by the water vapor absorption band centered at 22.235 GHz. The effect of this band must be considered to frequencies as low as 10 GHz, below which water vapor is relatively transparent.

Water droplets in the atmosphere, in clouds and rain, also have an effect on microwave propagation. Both absorption and emission due to cloud water droplets increase as the square of the microwave frequency. The result is that the presence of water droplets increases atmospheric absorption and emission in the 35 to 90 GHz windows, but has a much less significant effect at frequencies in the window below 10 GHz.

The curves presented in (F22.24) are intended to give the reader a simple interpretation of the frequency dependence of changes in temperature, salinity, and other variables of the ocean. Subsequent sections will provide further discussion on these variables and will show that in addition to being frequency dependent, they are dependent on polarization, P , and viewing angle, Ψ . So the reader should be aware that the magnitude of the sensitivity functions in (F22.24) will change for other combination of P and Ψ .

22.2.1 Theory

As in 22.1.2, the notation and some of the discussion in this section follows that of Holter (1970) and Simon (1966).

Planck's fundamental radiation equation, presented in (22:1) gives the spectral distribution of radiation from a perfect radiator, called a blackbody, with uniform temperature. An approximation valid for long wavelengths is obtained from (22:1) by expanding the exponential function and retaining only the

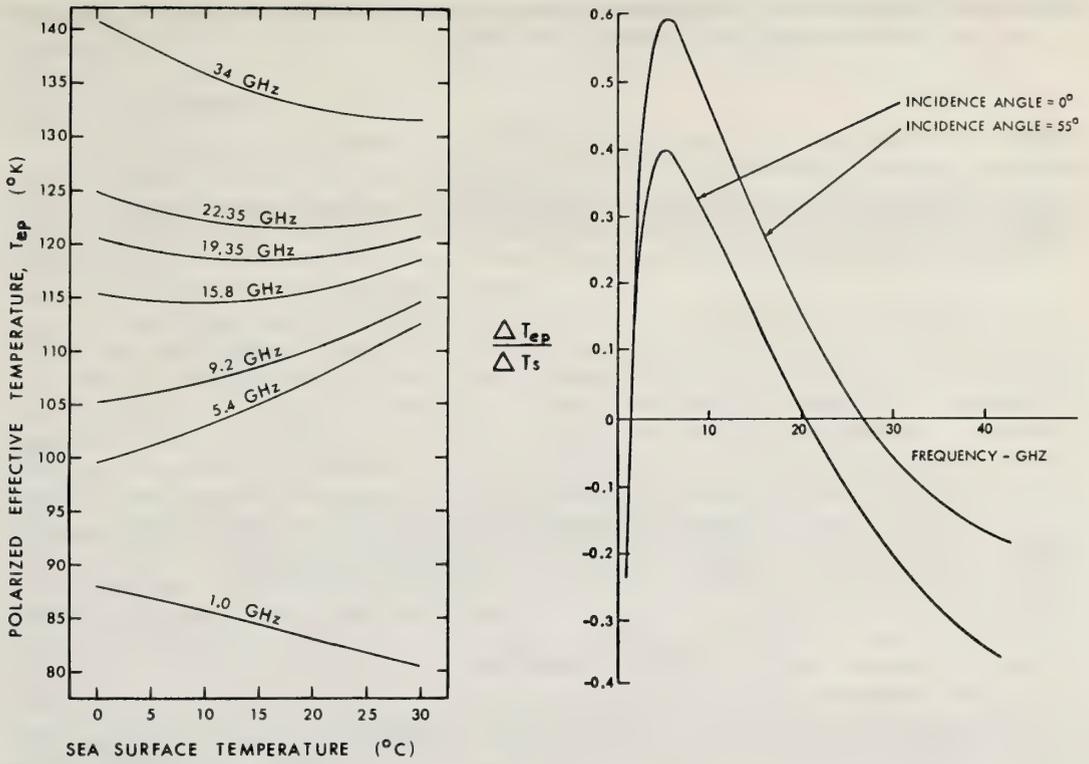


Figure 22.26 The polarized emissive temperature of a flat sea surface versus the thermometric temperature of the sea surface for frequencies of 1, 5.4, 9.2, 15.8, 19.35, 22.235, and 34 GHz and for a salinity of 35 ‰ (after Paris, 1969).

Figure 22.27 Radiometric temperature sensitivity to changes in molecular temperature, for calm sea, vertical polarization, 270 < T_s < 303 °K (after Huston, 1971).

first-order term:

$$W_{B\lambda} = \frac{C_1}{C_2} \frac{T_s}{\lambda^4} \tag{22:8}$$

This is the *Rayleigh-Jeans Law*† which is a good approximation when the product λT_s is greater than 10^5 micron degrees, and is valid for microwave emission from the earth's surface which is the order of 10^7 micron degrees. The importance of this approximation in microwave work is that the spectral radiant emittance of a surface, W_{λ} , is proportional to the first power of the absolute temperature. This simplifies the notation and many calculations.

22.2.1.1 Brightness Temperature

A commonly used term in microwave work is brightness temperature, T_B . This is often termed effective target temperature by investigators to refer to the case where a target does not completely fill the field of view of the sensor. The definition of brightness temperature evolves from basic radiation principles, discussed in 22.1.2, that radiation emanating from a surface is made up of three parts: a self-emitted component, a reflected component, and a transmitted component. These are shown schematically in

†This approximation and Planck's law are given in most basic textbooks on modern physics.

(F22.25). By (22:8) the self-emitted microwave component is proportional to the molecular temperature, T_o . Thus, an effective temperature of a surface can be defined as

$$T_e = \epsilon_\lambda T_o \quad (22:9)$$

Similarly, effective temperature contribution, T_ρ and T_τ , due to the surface spectral reflectance, ρ_λ , and transmittance, τ_λ , can be written as,

$$T_\rho = \rho_\lambda T_i \quad (22:10)$$

$$T_\tau = \tau_\lambda T_i' \quad (22:11)$$

where T_i and T_i' are terms proportional to the radiation incident on the surface from above and below, respectively. The brightness temperature, is thus the sum of these three terms, i.e.,

$$T_B = \epsilon_\lambda T_o + \rho_\lambda T_i + \tau_\lambda T_i' \quad (22:12)$$

This is the surface temperature which would be measured with a remote microwave sensor having an ideal antenna, assuming no atmospheric attenuation. It would agree with the molecular temperature only if the surface was a blackbody, in which case $\rho_\lambda = \tau_\lambda = 0$ and $\epsilon_\lambda = 1$.

Applying these principles to the ocean surface, one can visualize from (F22.25) that if the ocean has a smooth surface, without foam, and if the molecular temperature is constant in the near surface (active) layer, then the brightness temperature has only two components: the self-emitted and reflected components. In this special case, $\tau_\lambda = 0$. However with foam, all three components are present: the self-emitted component is radiation from the foam; the reflected component is microwave radiation from the atmosphere and cosmic sources reflected by the foam; and the transmitted component is radiation from the near-surface active layer of the ocean transmitted by the foam. Second order terms are present but are not shown here for simplicity.

Theoretical radiative transfer models have been developed by a number of investigators in order to calculate the three components of brightness temperature. Particularly valuable work has been done by Paris (1969, and 1971), Hidy, et al., (1971), Stogryn (1967), Lewis, et al., (1954), and McMahon (1950), among others.

22.2.1.2 Polarization

Because the emitted microwave radiation is polarized, it is convenient to divide the radiance field into two components—the vertical and horizontal polarized radiance. Then the total spectral radiance is the sum of the two polarized spectral radiance components, i.e.,

$$N_\lambda = N_{\lambda h} + N_{\lambda v} \quad (22:13)$$

where N_λ has units of ergs/cm² sec sr μ m. In general, $N_{\lambda h}$ and $N_{\lambda v}$ are not equal.

For purposes of notation the subscript P will be used whenever either horizontal or vertical polarization is applicable. For example, the effective temperature, T_e can be expressed as the polarized effective temperature, T_{eP} , if it represents either of the two polarized components. Similarly, the brightness

temperature, T_B , can be termed polarized brightness temperature T_{BP} if it expresses either polarized component.

22.2.2 Ocean Temperature

Some investigators have developed radiative transfer models in order to study how the microwave radiant emittance of an active surface layer varies with temperature and dielectric properties of the layer. With such models, these investigators are able to determine how the microwave component varies with the detection parameters of frequency, ν , polarization, P, and nadir angle, Ψ . Such models are particularly useful because they allow one to simulate a wider range of conditions that occurs in nature or to examine problems which cannot be investigated easily in nature.

A simple model used by Paris (1969) and others is that of a smooth sea-air interface and an active surface layer which reflects, absorbs, and emits microwave energy and is homogeneous in both temperature and dielectric properties. This active layer has a thickness of only a few wavelengths. For the case of a smooth interface, one may use Fresnel's Law to predict the polarized components of reflectance and emittance, which leads directly to a calculation of the polarized effective temperature, T_{ep} . Calculations with this model have been carried out by Paris (1969) for molecular temperatures ranging from 0 to 30 °C and for various microwave frequencies. The results, shown in (F22.26), indicate that at some frequencies the sea surface temperature does not uniquely determine the polarized effective temperature. At 15.8, 19.35, and 22.235 GHz, for example, two sea surface temperatures yield the same polarized effective temperature value. Another interesting feature of (F22.26) is that at 5.4 GHz the polarized effective temperature increases with surface temperature, but at 1.0 GHz it decreases with surface temperature. Obviously, the emissivity of a flat sea surface is dependent on both microwave frequency and molecular temperature. These computations also show that maximum sensitivity to surface temperature change occurs near 5 GHz. This is shown more clearly in (F22.27) which indicates a sharp maximum in the ratio of the change of polarized effective temperature to that of molecular temperature. The maximum is located at 5-6 GHz. However, above 20 and below 2 GHz the ratio becomes negative, indicating that warmer water appears to be colder and vice-versa.

Additional measurement information can be obtained by utilizing the detection parameters of polarization and nadir angle. (F22.28), which is derived from the calculations of Lewis, et al. (1954) shows that the spectral emittance of a smooth ocean surface depends on both these detection parameters. For these calculations $T_s = 293$ °K and $\nu = 23$ GHz. Maximum emittance is obtained with a nadir angle of about 80 degrees with vertical polarization.

The calculations in this section assume a perfectly smooth surface, however the reader should keep in mind that surface roughness will affect the polarized microwave emission from sea water. This is described in detail in 22.2.4.

22.2.3 Salinity

Salinity is an important oceanic parameter which is studied on a global scale and affects microwave emission at certain frequencies. To illustrate this frequency dependence of brightness temperature on salinity, Paris (1969) has made calculations of polarized brightness temperature assuming: (1) $T_s = 10$ °C; (2) a smooth sea surface; (3) $\Psi = 0$; (4) a standard, cloudless atmosphere; and (5) a measurement height of 1 Km. His results are shown in (F22.29). It is apparent that salinity variation affects the polarized brightness temperature for a frequency of 1 GHz, but the affect is insignificant at higher frequencies. This suggests that multifrequency techniques may be employed for measurement of sea surface molecular temperature and other parameters, with one channel near 1 GHz in order to asses the salinity of the active layer.

Other studies have shown that the maximum effect of salinity on polarized brightness temperature occurs when $\Psi = 0$ (Sirounian, 1968), and that this effect increases with sea surface temperature (Paris, 1969).

Some investigators have proposed using the S-band frequency of 2.69 GHz to determine molecular temperature of the ocean surface. They must therefore consider the possible uncertainty due to salinity. A set of such uncertainty curves has been derived by Hidy et al. (1971) and is presented in (F22.30). This shows the uncertainty in measured molecular temperature due to variation in salinity for certain water

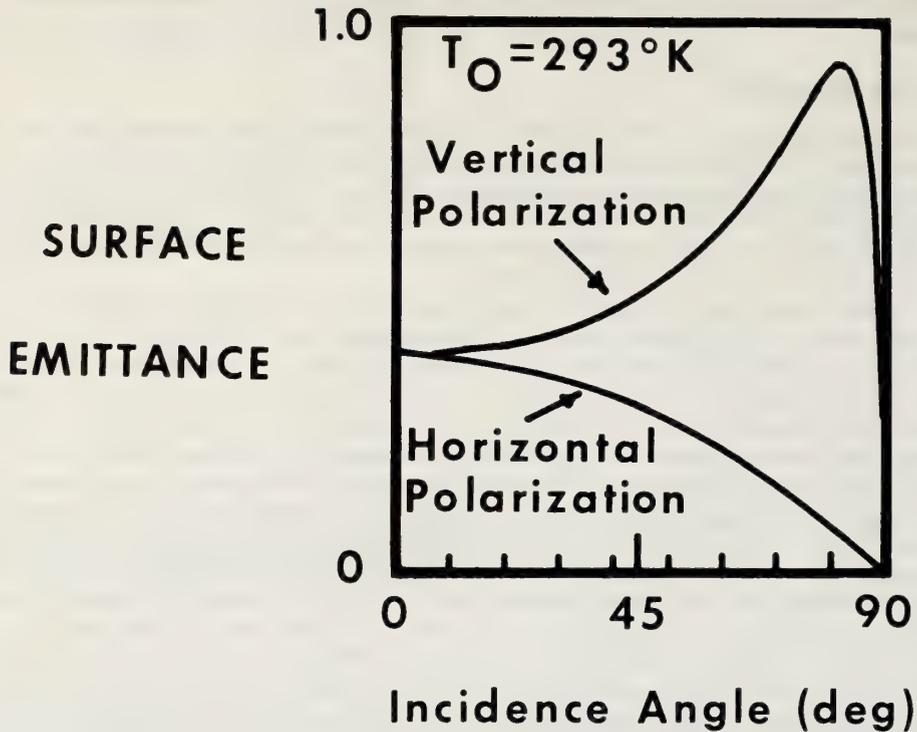


Figure 22.28 Calculated emittance of a water surface, $T_s = 293^\circ\text{K}$ and $\nu = 23\text{ GHz}$ (after Lewis et al., 1954).

temperatures and a frequency of 2.65 GHz. They have also indicated that the uncertainty is inversely proportional to frequency, so by doubling the frequency, the uncertainty is reduced by approximately a factor of 2.

There is geographical variation in salinity as shown in F22.31. The change in salinity from equator to pole is about 10 percent which would cause an orbiting microwave 2.65 GHz radiometer to observe a molecular temperature change of 3°K . If this is compared to the actual sea surface temperature change of 35°K between the equator and pole, it is apparent that first order corrections for the geographical variation of salinity must be applied.

Hidy, et al., (1971) have also studied the temporal variations of salinity and find them large enough to be significant. Daily variations of 0.5 percent in mid-oceans surface salinity lead to an uncertainty in molecular temperature of 0.1-0.2 $^\circ\text{K}$. Seasonal and year-to-year variations produce an uncertainty in surface temperature of 0.5°K .

Thus, geographical and temporal salinity variation in the world's oceans represents a small but significant uncertainty in interpreting S-Band 2.65 GHz radiometer observations for sea surface temperature. However, these variations are correctable to first order with known geographical and temporal variation (Hidy, et al., 1971).

22.2.4 Sea State—Roughness and Foam

If either sea surface temperature or wind speed is to be inferred from microwave brightness temperature, it is essential that the effect of sea state on microwave emission be adequately understood for the range of environmental conditions of the world's oceans. At present only first order theoretical calculations have been carried out on the effect of surface roughness, i.e., wave slope, and there is no physical theory that is adequate to describe the effective dielectric constant of an ocean surface covered with foam. Theoretical and experimental data which are available suggest that the brightness temperature

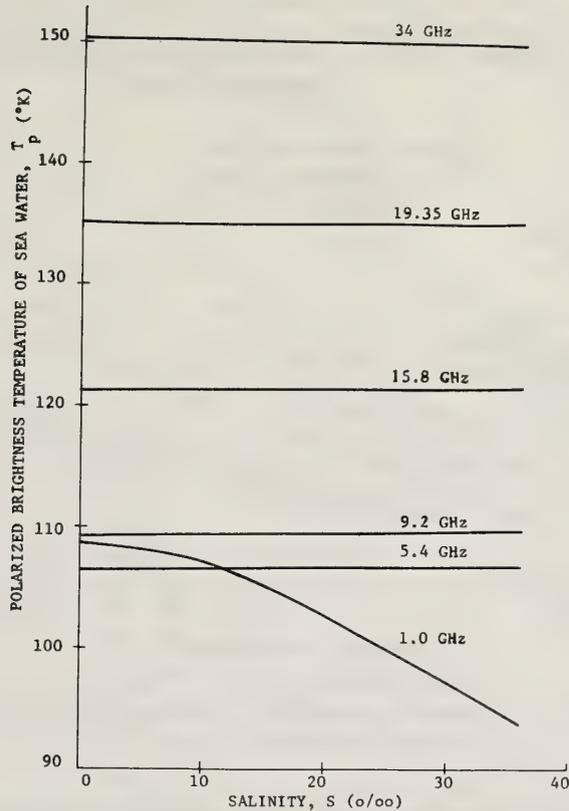


Figure 22.29 The polarized brightness temperature of sea water versus salinity for a water temperature of 283 °K, for an incidence angle of 0°, for frequencies of 1, 5.4, 9.2, 15.8, 19.35, and 34 GHz and for an altitude of 1 km (after Paris, 1969).

increases with increasing wind speed, although this relationship is critically dependent on polarization and viewing angle. At lower wind speeds, without foam, the brightness temperature depends on the RMS wave slope, whereas when foam is present the brightness temperature increases appear to be influenced primarily by foam.

22.2.4.1 Roughness

Investigators, such as Stogryn (1967) and Hidy, et al. (1971), have developed geometric optics models which include the effect on microwave emission of roughening of the sea surface by wind. The procedure is to assume Fresnel's Law of reflection is valid at a particular point on the surface, and to compute mean polarized emissivities by integration of the wave slope distribution, as described by Buettner et al. (1968). The wave slope distribution given by Cox and Munk (1954a, 1954b) is used for expressing the wave slope as a function of wind speed.

The potential use of detection parameters, P and Ψ , in determining the distribution of wave slopes was demonstrated first by Stogryn (1967). He calculated the brightness temperature versus incidence angle for both polarizations at 19 GHz, and for the following conditions:

1. An active ocean layer with homogeneous temperature and salinity of 17 °C and 36‰ (which are representative of midlatitude conditions).
2. An ARDC standard atmosphere,
3. A sensor height of 1 km, and,
4. Surface wind speeds of 0, 4, 8, and 14 ms^{-1} .

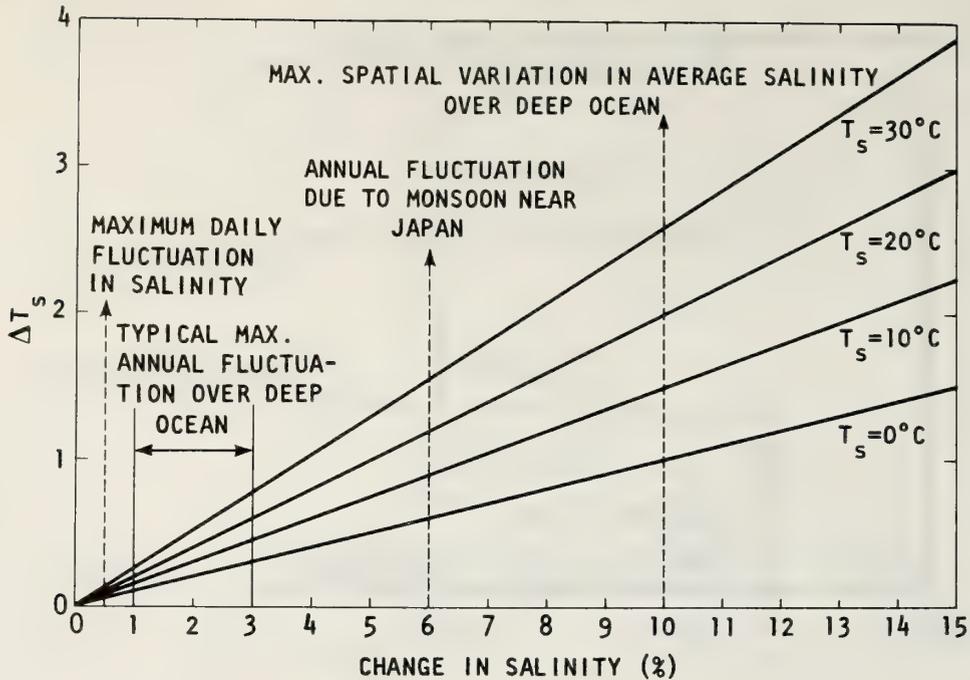


Figure 22.30 Uncertainty in molecular temperature T_s , measured by a 2.65 GHz radiometer, due to changes in salinity of the world's oceans (after Hidy *et al.*, 1971).

Stogryn's calculations, illustrated in (F22.32), show a number of significant features. Brightness temperature strongly depends on polarization, as previously shown for the smooth ocean case in (F22.28) (Lewis *et al.*, 1954). For vertical polarization, roughness of the ocean tends to increase the brightness temperature at $\Psi < 55^\circ$, but decreases it at $\Psi > 55^\circ$. The fact that vertically polarized brightness temperature is independent of wind speed at $\Psi = 55^\circ$ may be used to advantage in multi-frequency scanning from satellites. For horizontal polarization, ocean roughness produces insignificant changes in brightness temperature at $\Psi = 0$ and has maximum change at Ψ values near 60° . The principle cause of this phenomenon, appearing in both vertically and horizontally polarized brightness temperature, is the change in the source of reflected sky radiation which results from wind roughening of the surface.

The theoretical calculations of Stogryn have been partially substantiated in experimental studies by Hollinger (1970). Using an 8.36 and 19.34 GHz radiometer from an ocean tower near Bermuda, Hollinger obtained polarized brightness temperatures for nadir angles from $0-75^\circ$ and for wind speeds of 1.6 and 7.4 m s^{-1} . His results, in (T22.33), show no significant difference between the measurements at these two frequencies. They verify the reversal in wind effect on vertically polarized brightness temperatures at $\Psi = 55^\circ$ as calculated by Stogryn. They are in excellent quantitative agreement with the same effect on horizontally polarized brightness temperature at $\Psi = 55^\circ$, as is evident by comparison of the appropriate $\Delta T_{\text{Bh}}/\Delta v$ values in (F22.5). The only significant difference is the wind effect on horizontally polarized temperatures at small values of Ψ . Stogryn's theoretical calculations showed minimal wind effect as Ψ approached 0° ; whereas Hollinger's experimental data show a diminishing but yet significant wind effect at $15^\circ < \Psi < 55^\circ$. The data at $\Psi \leq 15^\circ$ may be anomalous due to the presence of the tower in Hollinger's data (Hollinger, 1970).

The theoretical calculations also predict very little dependence of the wind effect on microwave frequency. However, other experimental work by Hollinger (1971) shows that the wind effect at 1.41 GHz is only about 1/3 of that at 19.34 GHz, based on wind speeds in the range 0.5-13.5 m/sec. The appropriate numerical values are given in (T22.5). Thus, there is an important discrepancy between theory and experiment on the degree to which ocean roughness effects are frequency dependent.

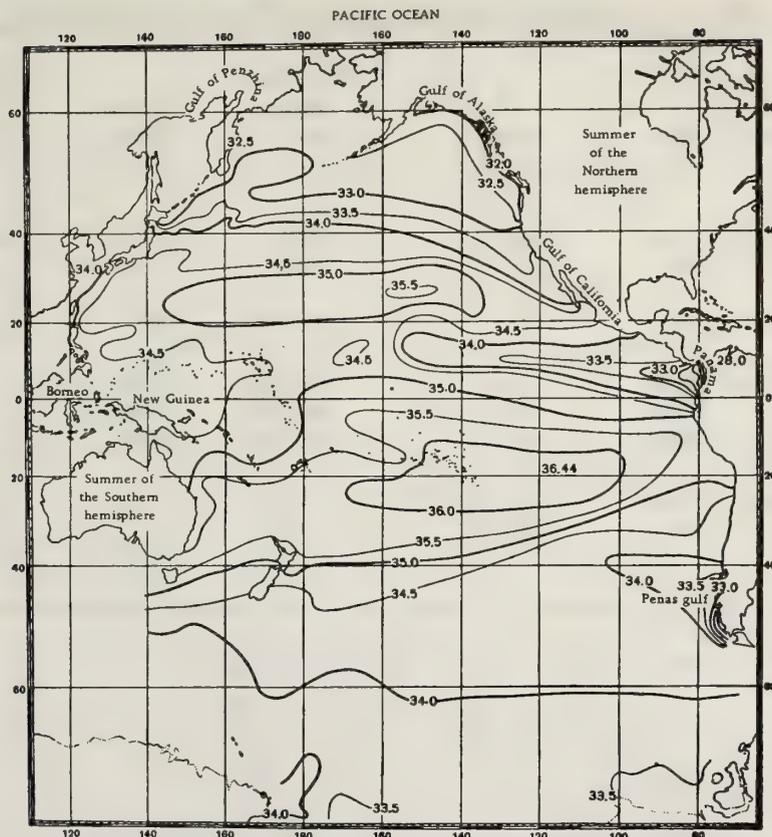


Figure 22.31 Average distribution of sea-surface salinity for the Pacific Ocean in parts per thousand (after Muromtsev, 1963).

22.2.4.2 Foam

The emissivity of foam is much higher than that of the sea surface. This was first suggested by Williams (1969) based on experimental data, and was supported theoretically on the basis of a physical model for foam by Droppleman (1970). Recent experimental data, reported by Ross et al. (1970) and Nordberg et al. (1971), indicate that over typical oceanic whitecaps the brightness temperature may be 100°K higher than over adjacent foam-free ocean areas. Such a temperature anomaly indicates that foam may act as both an error source for molecular temperature measurement and a possible useful indicator of wind speed.

Both theoretical and experimental studies show there is a significant increase in whitecap coverage and spray density, beginning at $6\text{-}7\text{ ms}^{-1}$ (Monahan 1969, and Cardone, 1969). Similarly, the aircraft microwave radiometer data by Nordberg, et al. (1971) suggest a significant increase in microwave brightness temperature begins at about $6\text{-}7\text{ ms}^{-1}$. They used a horizontally polarized 19.34 GHz radiometer and from a time series of measurements were able to show an increase in brightness temperature as a function of wind speed for speeds of <5 to 25 ms^{-1} with the sensor looking directly at the nadir. Their results in (F22.34) suggest that at wind speeds $>7\text{ ms}^{-1}$ the increase in brightness temperature is due to an increase in foam coverage. It is apparent from their study that a determination of foam coverage is critical because only a few percent foam coverage can cause brightness temperature anomalies of about 2°K .

The natural variation in foam coverage on a global scale has been investigated by Blanchard (1963) and is shown in (F22.35). In the tropics and mid-latitudes one would expect 2 percent or more foam coverage in both summer and winter, but as much as 8-9 percent at higher latitudes. A first order correction in brightness temperature will be necessary at certain angles and polarization in order to correct for foam.

Table 22.5 Summary of Brightness Temperature Sensitivity to Wind Speed

INVESTIGATOR	ν (GHz)	ψ (deg.)	Wind Speed Range (m s^{-1})	$\Delta T_B / \Delta v$	
				Vertical ($^{\circ}\text{K}/\text{m s}^{-1}$)	Horizontal ($^{\circ}\text{K}/\text{m s}^{-1}$)
Theoretical					
Stogryn (1967)	19.34	30	0-14	+0.4	+0.4
	19.34	55	0-14	0	+1.4
Experimental					
Hollinger (1970)	19.34	55	1.6-7.4	-0.1	+1.3
Hollinger (1971)	1.41	55	0.5-13.5	+0.02	+0.34
	8.36	55	0.5-13.5	+0.04	+0.60
	19.34	55	0.5-13.5	-0.06	+1.06
Nordberg et al. (1971)	19.34	0	6-25	—	+1.2
	19.34	70	6-25	—	+1.8

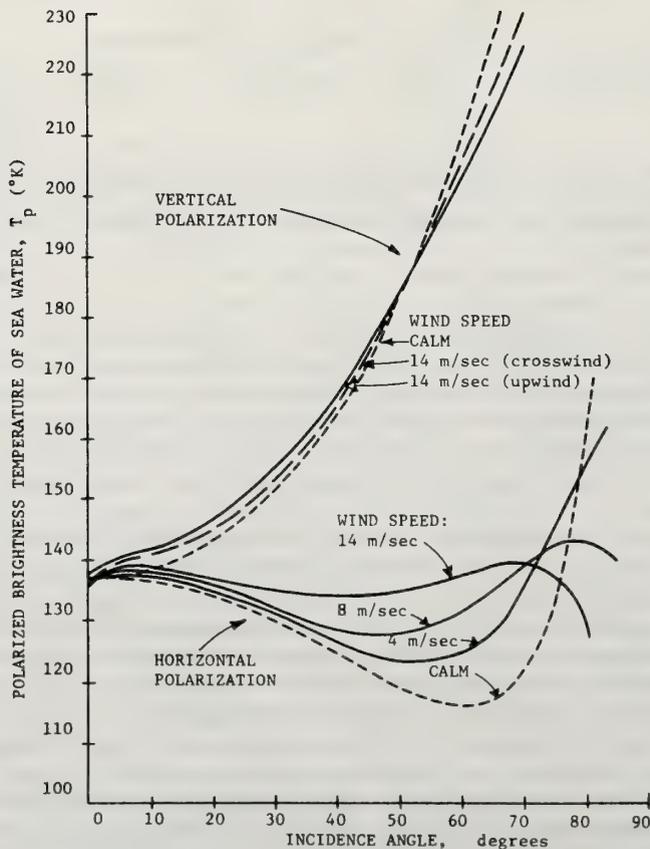


Figure 22.32 The polarized brightness temperatures of sea water versus incidence angle for horizontal and vertical polarization, for fully developed sea driven by surface winds of 0, 4, and 14 m/sec, an altitude of 1 km, and a frequency of 19.4 GHz (after Stogryn, 1967).

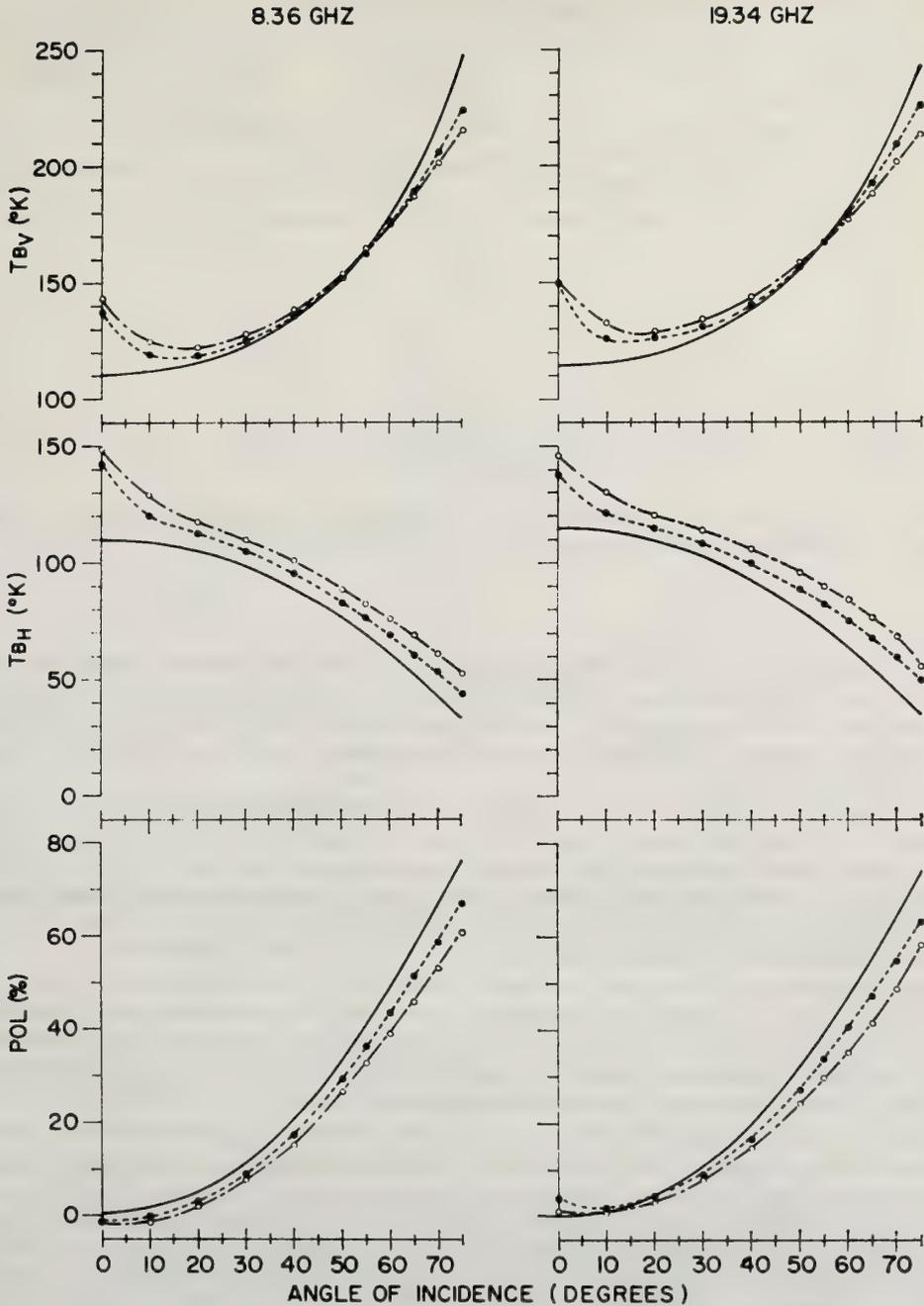


Figure 22.33 The vertical and horizontal components of brightness temperature of the sea and the percentage of polarization (defined as the ratio of the difference of these two components of their sum) at 8.36 and 19.34 GHz are plotted as a function of incidence angle for two groups of measurements comprising the lowest and highest sea conditions encountered. The solid circles represent average sea conditions of 2.6 m/sec winds with 1.6 meter waves; the open circles, 7.4 m/sec winds with 1.9 meter waves. The solid line has been calculated for a specular surface (after Hollinger, 1970).

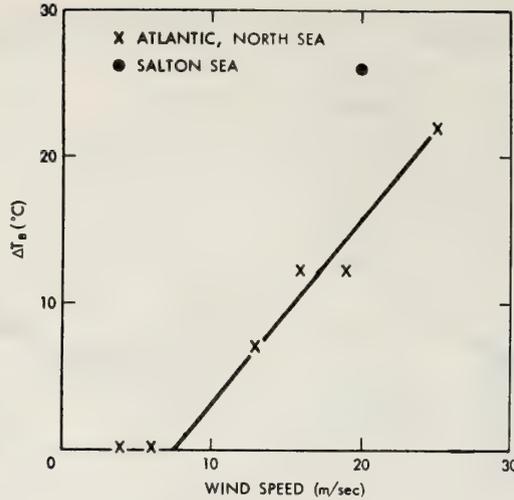


Figure 22.34 Brightness temperature differences between observations made at wind speeds $< 5 \text{ m sec}^{-1}$ and higher wind speeds as a function of wind speed (after Nordberg et al., 1971).

22.2.5 Ice

There have been many excellent studies, such as that of Pascalar and Sakamoto (1965), which show that the brightness temperature contrast between ice and water is quite large because the emissivity of optically thick ice is in the range 0.8-1.0, whereas that of water is only about 0.4-0.5. This natural difference in their emissivities might possibly be employed in remote sensing to determine the extent of open water in the arctic ice pack.

Studies by Badgley (1966) of polar ocean areas have suggested that the heat flux from the ocean to the atmosphere is at least two orders of magnitude greater in open water areas of the ice pack than through the ice pack itself. Other studies have indicated that at any time of the year as much as 10 percent of the Arctic Ocean may be open water, resulting from polynyas (large cracks) and leads (small cracks) in the ice pack (Wittman and Schule, 1966).[†] Hence the ocean to air heat flux in ice covered ocean regions may be monitored through microwave sensing by measurement of the relative amount of open water.

Because the open water leads may be relatively narrow, antenna beamwidths must be small enough to resolve them. With the restrictions imposed by space craft on antenna sizes, this requires high frequencies. However the attenuation due to water vapor and oxygen limit the choices. The window centered near 37 GHz is the highest frequency range which can be used successfully because of the attenuating and emitting effects at higher frequencies of molecular oxygen and water droplets in the atmosphere.

Recently in experimental studies, Wilheit, et al., (1971) discovered that year-old, smooth ice has a lower emissivity than older, more heavily hummocked ice. This would lead to an ambiguity in determining the relative amount of open water by observation of emissivity contrast. Fortunately, the difference in emissivity between the two ice types is relatively independent of polarization and is proportional to frequency (up to 37 GHz), whereas the emissivity contrast between water and ice is strongly dependent on polarization. Thus, with a measurement of both polarizations, the environmental variables can be resolved.

The large emissivity contrast between ice and water has been used by the U. S. Coast Guard to locate and track icebergs in the North Atlantic Ocean (Roeder, 1967). Such an iceberg target must be sufficiently large within the field of view to give a detectable emissivity contrast from the ocean background.

[†]cited by Wilheit et al., (1971).

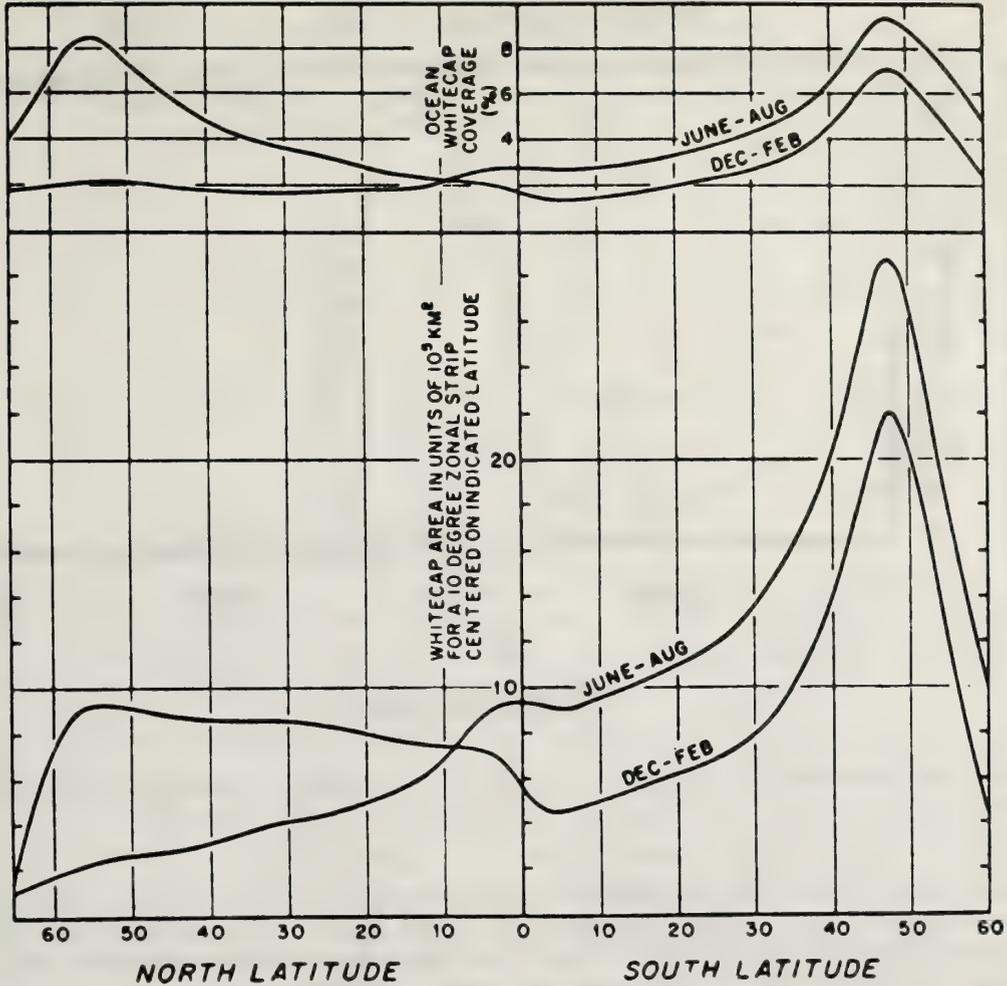


Figure 22.35 The latitudinal variation of the world ocean whitecap distribution. The upper set of curves give the percentage coverage during the winter and summer seasons. The lower set gives the area of the sea surface that is covered with whitecaps. The bump in the June-August curves near the equator reflects the monsoons in the Indian Ocean (after Blanchard, 1963).

22.2.6 Atmospheric Extinction and Emission

Microwave emission by the atmosphere affects remote sensing of oceanographic variables in two ways. One is that the atmosphere itself is a source of microwave illumination for the ocean surface, and because of the relatively high reflectivity of the surface, a remote detector would receive the combined microwave emission of the ocean and atmosphere. Clearly the atmospheric portion must be evaluated in order to interpret the emission from the ocean alone. Secondly, the atmosphere attenuates microwave radiation from the ocean surface, tending to limit the information available for detecting oceanographic variables. Only a limited discussion of these atmospheric effects will be presented here because they are discussed in other chapters of this volume.

The reader will recall from (F22.24) there are two relatively broad atmospheric windows on either side of the 60 GHz oxygen absorption band; these windows are shown in greater detail in (F22.36). Although the 80-90 GHz window has a high transmissivity, it has only limited usefulness because of the relatively large influence of liquid water drops at those frequencies. Thus, the window on the lower

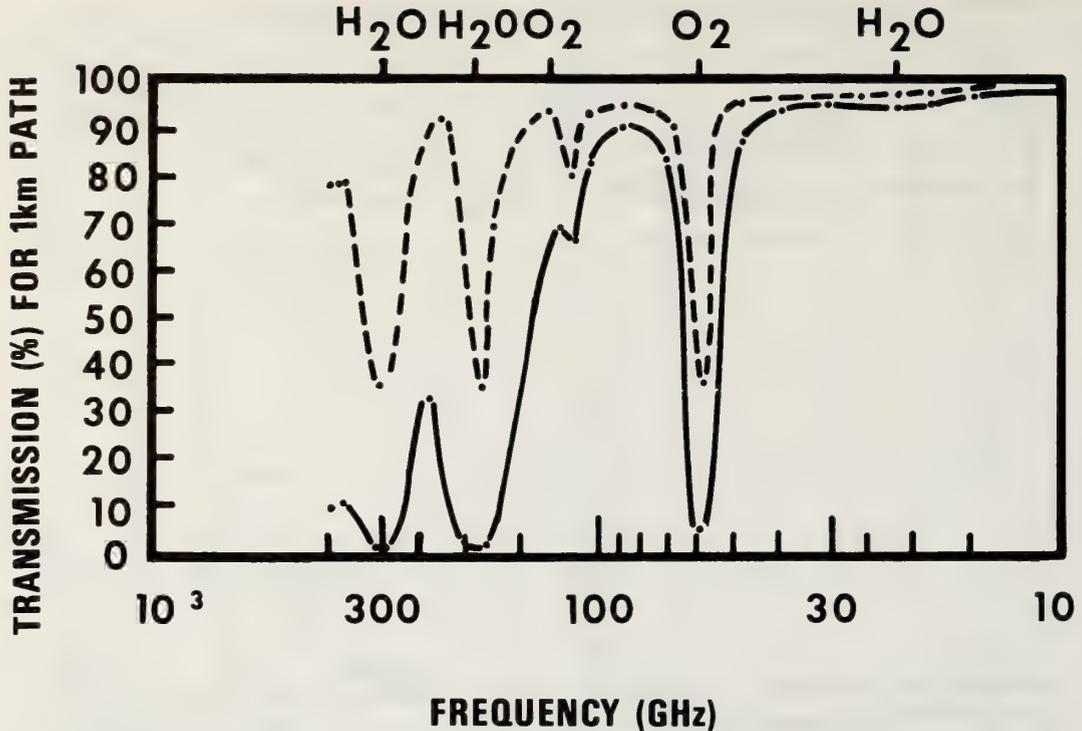


Figure 22.36 *Transmission for 1 Km path. Attenuation due to atmospheric oxygen and water vapor. Solid curve: $p = 760$ mm Hg.; $T = 20^{\circ}\text{C}$; water vapor = 7.5 gm/ M^3 (adapted from Meyer, 1966).*

frequency side of the 60 GHz absorption band is most promising for passive microwave sensing. Here transmissivities are relatively high at frequencies below 40 GHz. However, a water vapor absorption line is centered at 22.235 GHz, as shown in (F22.37), and it has a significant influence on atmospheric absorption at frequencies as low as 5-10 GHz, depending on the amount of water vapor.

Galactic and cosmic background radiation becomes relatively large at frequencies below 1 GHz and can usually be ignored at frequencies above 2 GHz (Hidy, et al., 1971).

As long as liquid water droplets in clouds are much smaller than microwave wavelengths, scattering is assumed negligible and the absorption of cloud water is (1) approximately proportional to the square of the frequency, (2) directly proportional to liquid water content in the path, and (3) independent of the drop size distribution (Huston, 1971). Calculations by Gunn and East (1954) indicate the absorption of ice in clouds is roughly an order of magnitude less than that for liquid water.

It is considerably more difficult to evaluate the influence of hydrometeors in microwave attenuation than to evaluate the influence of cloud droplets because hydrometeors are large enough to approach microwave wavelengths. Thus, both scattering and absorption must be considered. Gunn and East (1954) obtained calculations of radiation attenuation with precipitation and found that (1) attenuation was roughly proportional to precipitation rate, and (2) snow fall effects are at least two orders of magnitude less than liquid water.

22.3 Ocean Color Detection—Chlorophyll

A discussion of ocean color introduces a new aspect to remote sensing beyond the physics of light alone—that is, use of color stimulation of the human eye for discrimination of differences in scene illumination. Probably the first remote sensing of the sea was by early sailors who used color to discriminate ocean depth. This technique is still used today.

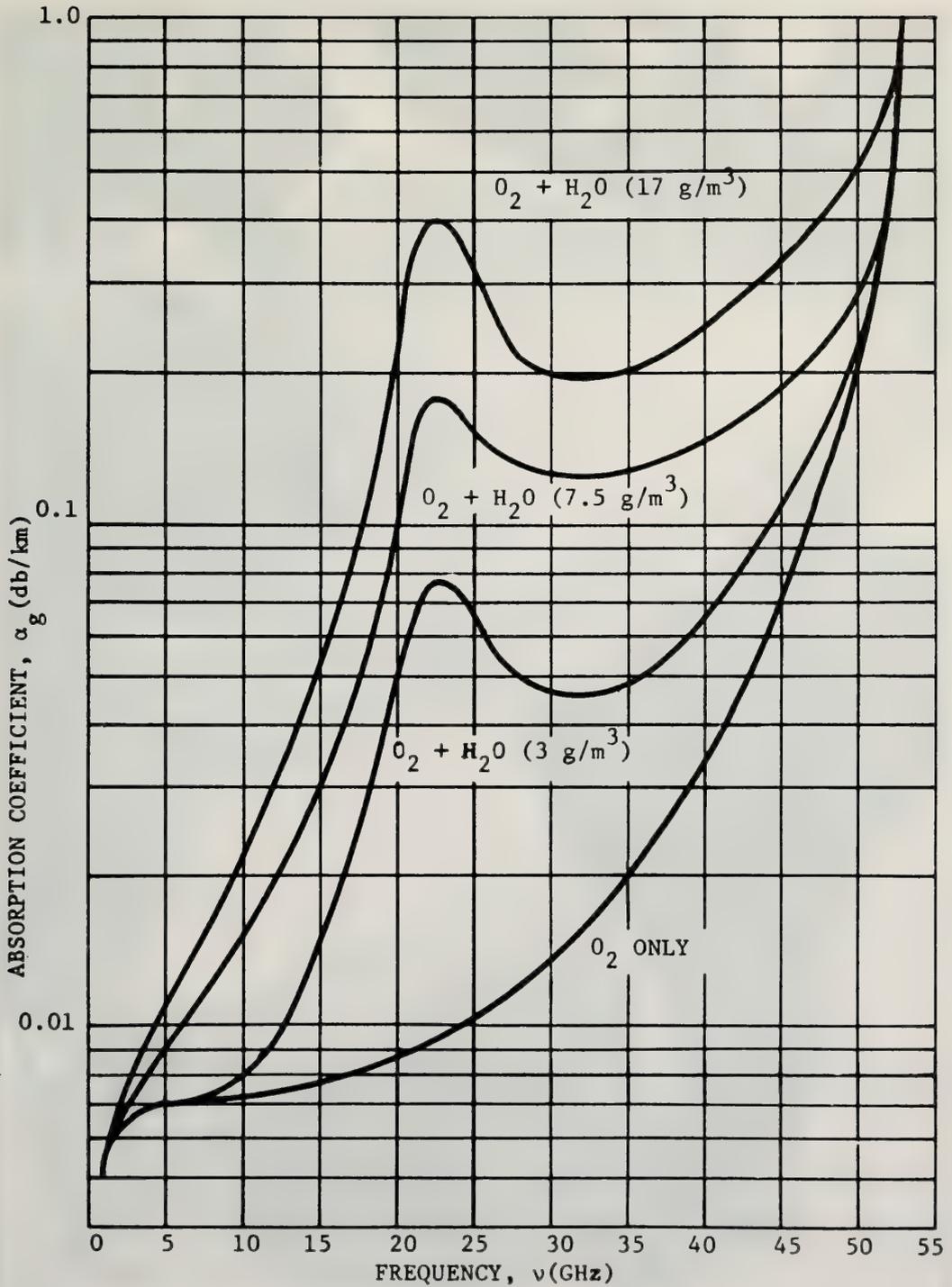


Figure 22.37 The absorption coefficient of molecular oxygen and water vapor at sea level versus frequency for a temperature of 293°K and for water vapor densities of 0, 3, 7.5, and 17 grams per cubic meter (after Paris, 1969).



Figure 22.38 Color of the sea as indicated in percent of yellow according to the Forel scale (according to Schott (1942) from Sverdrup, et al., (1942)).

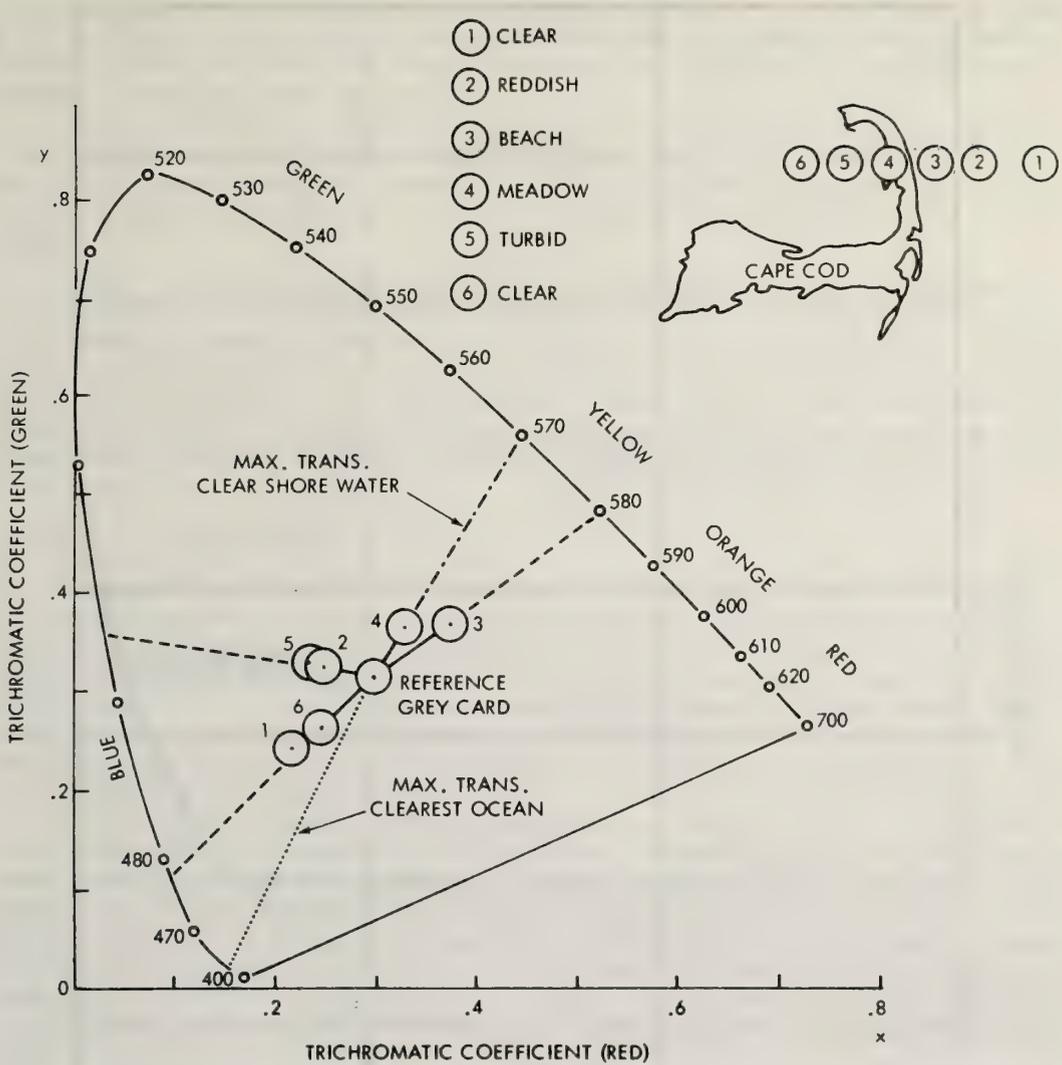


Figure 22.39 Chromaticity Diagram (after Ewing, 1969).

Solar radiation that penetrates the sea makes possible photosynthesis in green plants which are primarily phytoplankton. Because the production of all marine creatures depends on the growth of phytoplankton, it is important to assess the possibility of measuring the concentration of phytoplankton in the world's oceans by remote sensing techniques.

Although direct measurement cannot be made, it is known that light which penetrates the sea is absorbed and scattered by the particulates and dissolved materials, of which chlorophyll is an active absorber. Because high concentrations of chlorophyll are known to contain large populations of phytoplankton, color of light emerging from the ocean may be used to discriminate concentration of phytoplankton, and hence regions of high productivity (Clarke, et al., 1969).

22.3.1 Color Perception Methods

Oceanographers, over the past decades, have obtained measurements of ocean color and clarity for use in ocean watermass analysis. These color determinations have been obtained by lowering a Secchi disc one meter below the surface and comparing the apparent color of the disc with each of eleven tubes

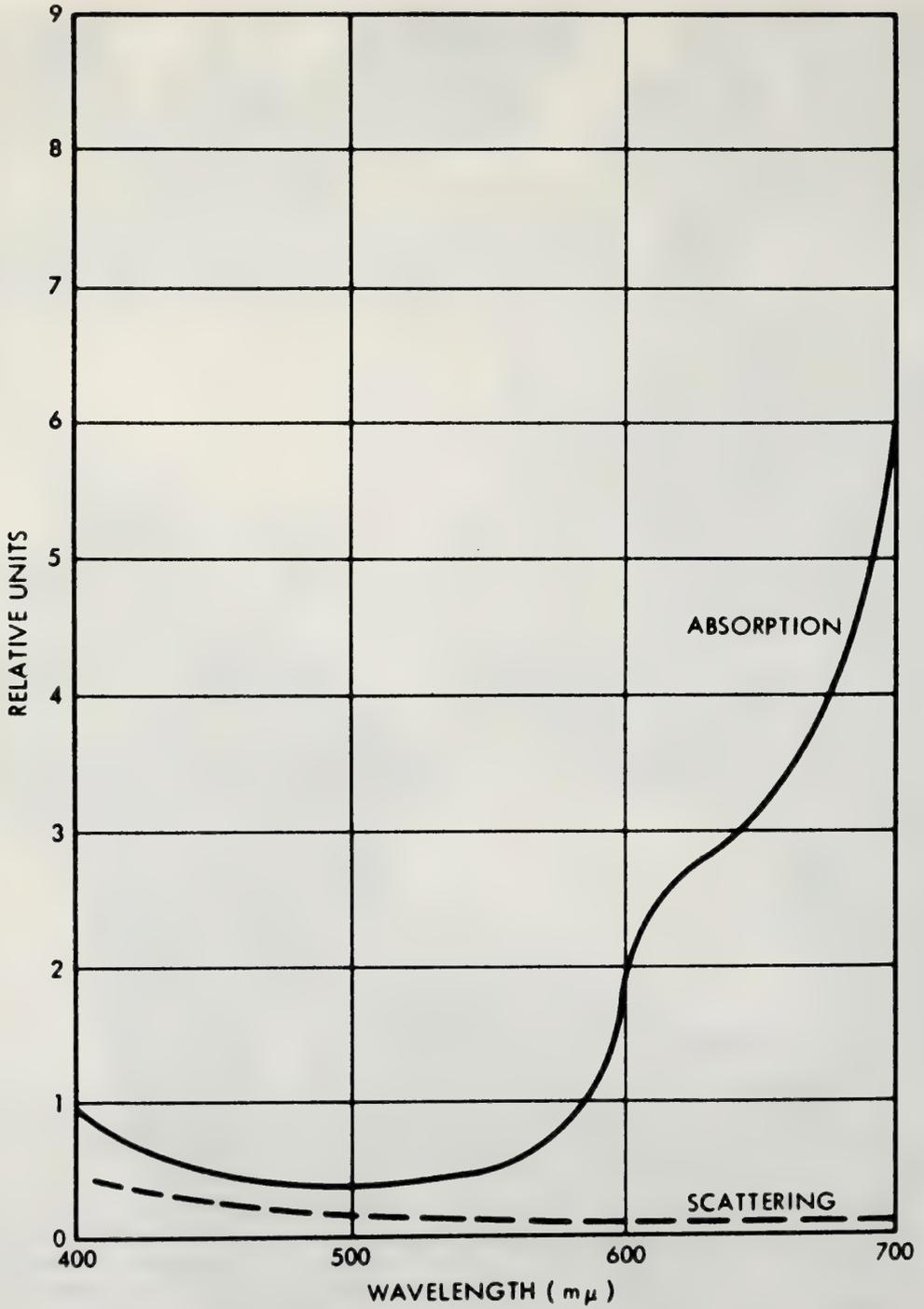


Figure 22.40 Absorption and Scattering of Distilled Water (after Hulbert, 1945).

containing colored water samples ranging from blue to green. The advantage of such a technique is that it is affected to only a minor degree by the atmosphere. However, it is deficient because it is based on a subjective judgment, has a color resolution of only one part in eleven, presents serious global sampling problems, and varies as a function of cloud cover.

The Forel Scale† color determinations were begun in the late 1800's and were improved by Kalle (1938) with the development of a comparator which allowed the observer to produce a color with red, green, and blue filters matching that observed in the sea. The three colors represent the primary colors of the color triangle of the Young-Helmholtz theory of color vision (Neumann and Pierson, 1966 and Jerlov, 1968).

In spite of the areal sampling problem with ocean color measurements, Schott (1942) has shown that the distribution of color over the Atlantic Ocean ranges from deep blue, Forel Scale 0, in the Sargasso Sea to greenish or yellow-green colors in equatorial regions, near shore, and at high latitudes. Schott's map is shown in (F22.38). Neumann and Pierson (1966) point out that this map reveals a close relationship with a map (not shown here) of the vertical extinction coefficient of surface water in the Atlantic Ocean obtained by Joseph and Wattenberg (1944). This provides evidence of an apparent large-scale relationship between oceanic particulates and ocean color, and also provides some hope for satellite remote detection on this spatial scale.

22.3.1.1 Chromaticity Coordinates

A completely objective method of determining color is by color analysis in C.I.E. chromaticity coordinates (Anonymous, 1957). This numerical definition of color is based on tristimulus values of the spectrum colors, red, green, and blue. Each color has standardly defined spectral response values. The green values are identical with the standard luminosity curve for photopic vision of the human eye (Committee on Colorimetry, 1966). A plot of pure color in chromaticity coordinates is called the spectrum locus and defines a horseshoe-shaped boundary of the color space (see F22.39). The chromaticity coordinates are taken nominally as the red and green trichromatic coefficients. By definition white light is located only at the central coordinate point (.333, .333).

Ewing (1969) has presented on a chromaticity diagram an example of the color variation of backscattered light from the ocean and from the land at Cape Cod, Mass., obtained by aircraft measurements at 500 ft. altitude. The diagram is shown in (F22.39). Clearly there is a distinct color separation between the measurements over and adjacent to Cape Cod, 2-5, as compared to those over clear ocean water, 1 and 6. The dominant wavelength is another term associated with the chromaticity diagram and is defined by extending a line from the central (white light) point, through the point in question to the spectrum locus. In (F22.39) the clear ocean points, 1 and 6, have a dominant wavelength of .478 μm . This is slightly higher than the wavelength of maximum transmission for the clearest oceans, about .470 μm (Jerlov, 1968).

The chromaticity coordinate technique appears to be a useful but laborious method for discriminating color information from either properly obtained photographic data or spectral radiance data.

22.3.2 Spectral Radiance

Apart from human color perception, it is possible to specify color objectively as the spectral radiance of a particular scene. In this definition, color need not be restricted to the visible portion of the spectrum and could include, for example, spectral radiance in the near IR as well. This section will discuss the use of spectral radiance in remote detection of ocean chlorophyll and productivity. The established convention of relating color to wavelength is given in (T22.6).

It is of interest to consider the path of solar radiation before it reaches a remote sensor. Of the solar illumination impinging on the upper atmosphere, about 75 percent of this broad-band radiation reaches and is transmitted downward through the ocean surface. Once in the ocean photic zone†† the

†Forel (1895 and 1901).

††Photic zone is defined as the oceanic layer with upper limit at the surface and lower limit at a level where the irradiance is reduced to 1 percent of the surface value.

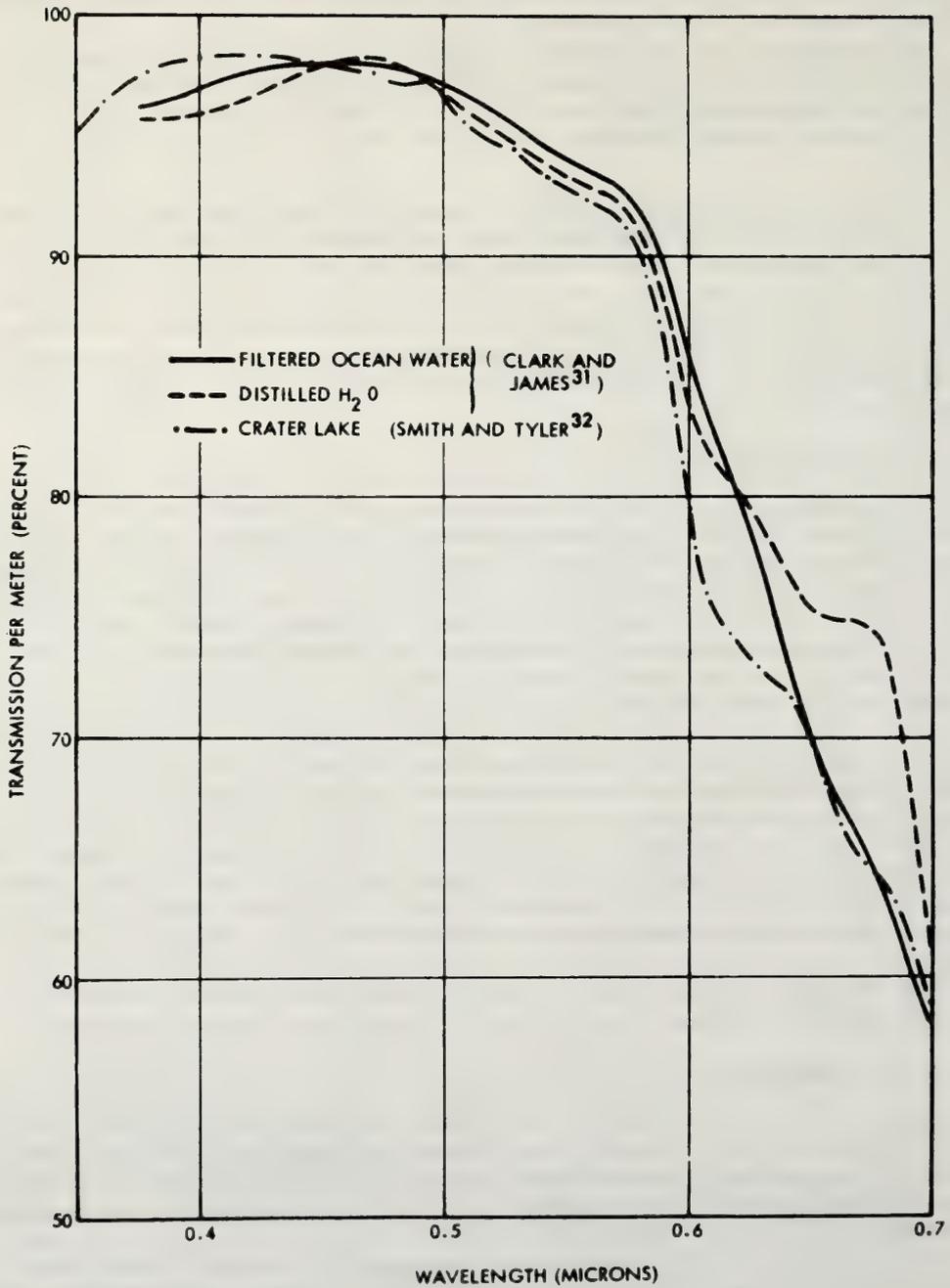


Figure 22.41 Percent Transmission of Clear Waters (according to Clarke and James (1939) and Smith and Tyler (1967), after Ramsey, 1968).

radiation is scattered and absorbed by particles, solutes, and the water itself in a complex way. A large portion of the radiation is scattered forward; a small but spectrally significant portion is back-scattered to the surface. The color of this backscattered light is critically dependent on the concentration of particles and solutes in the water. Pure water backscatters predominantly blue light with a dominant wavelength near $.470 \mu\text{m}$; however, particulates including chlorophyll, and solutes including yellow substance, in some ocean waters, tend to shift the dominant wavelength toward higher values causing the ocean to appear green, yellow-green, or brown. This relatively small back-scattered light which emerged from the sea surface traverses the atmosphere with some attenuation, finally reaching the remote aircraft or satellite detector. Also reaching the detector is light from two other sources; (1) that reflected by the sea surface, and (2) that scattered upward by atmospheric molecules and aerosols. Thus, remote sensing of oceanic chlorophyll requires the interpretation of upwelling spectral radiance which is the sum of three components, one of which varies, in part, because of variations in chlorophyll.

Only the backscattered spectral radiance of the photic zone will be discussed in the following sections. For a discussion of the sea surface and atmospheric radiation components, the reader is referred to an excellent review of the subject by Ramsey (1968).

22.3.2.1 Absorption and Scattering of Pure Ocean Water

Progress in the investigation of the attenuation of pure water was relatively slow, because the measurements were handicapped by the difficulty of preparing pure water. However, it is generally thought that Clarke and James (1939) have been successful in preparing pure water, and their values of attenuation (beam attenuation) for the visible region are representative (Jerlov, 1968).

A comparison of the attenuation data by Clarke and James and the scattering data for pure water by LeGrand (1939) indicates that the attenuation is primarily due to absorption rather than scattering. This is also verified by the data of Hulbert (1945) shown in (F22.40). Pure water has highly selective absorption properties and acts essentially as a monochromator for blue light near $.470 \mu\text{m}$, as shown in the transmittance curve, (F22.41), for distilled water from Clarke and James (1939).

The question arises whether the addition of sea salts changes the attenuation of pure water. To answer this, Clarke and James filtered clear ocean water from the Sargasso Sea and compared its transmittance with that of pure water. They found, (F22.41), no significant difference between the pure water and filtered Sargasso Sea water. Their finding was verified by Sullivan (1963). The conclusion is that sea salts exert little or no influence on visible light attenuation. Slight effects have been studied by Morel (1966).

22.3.2.2 Absorption by Yellow Substance and Chlorophyll

Natural ocean water contains both solutes and particles which alter significantly the radiative properties of pure water. This became apparent from the work of Clarke and James who obtain transmittance spectra on filtered water samples from turbid, coastal areas as well as from Sargasso Sea water. They found (F22.42) that even after careful filtration and removal of all suspended matter, the extinction of light remained much greater in filtered coastal water than for both pure water and filtered Sargasso Sea water. This indicates that turbid matter, remaining in the coastal water after filtering, is the major cause of selective extinction at short wavelengths (blue).

Evidence suggests that filter-passing turbidity matter is probably due to dissolved organic substances. Kalle (1938) has shown that sea water, especially in coastal areas, contains soluble pigments of yellow color which are chemically related to humic acids. This yellow substance represents fairly stable metabolic products, mainly carried into the sea by rivers but probably also produced in ocean water as a result of plankton metabolism (Neumann and Pierson, 1966). Yellow substance plays an important part in explaining the transition of the color of sea water from blue to green. The spectral absorption data for yellow substance (F22.43) show a strong increase in absorption at shorter wavelengths, which tends to shift the wavelength of maximum transmission of light in water from near $0.470 \mu\text{m}$, blue, to the longer yellow or green wavelengths.

Although the absorption spectrum of plant leaves is relatively well known, that for living phytoplankton has had much less attention. Yentsch (1960, 1962) has provided particularly useful information in this area. He has examined the absorption of various living phytoplankton groups and compared them with a natural population of phytoplankton sampled from ocean water near Woods Hole,

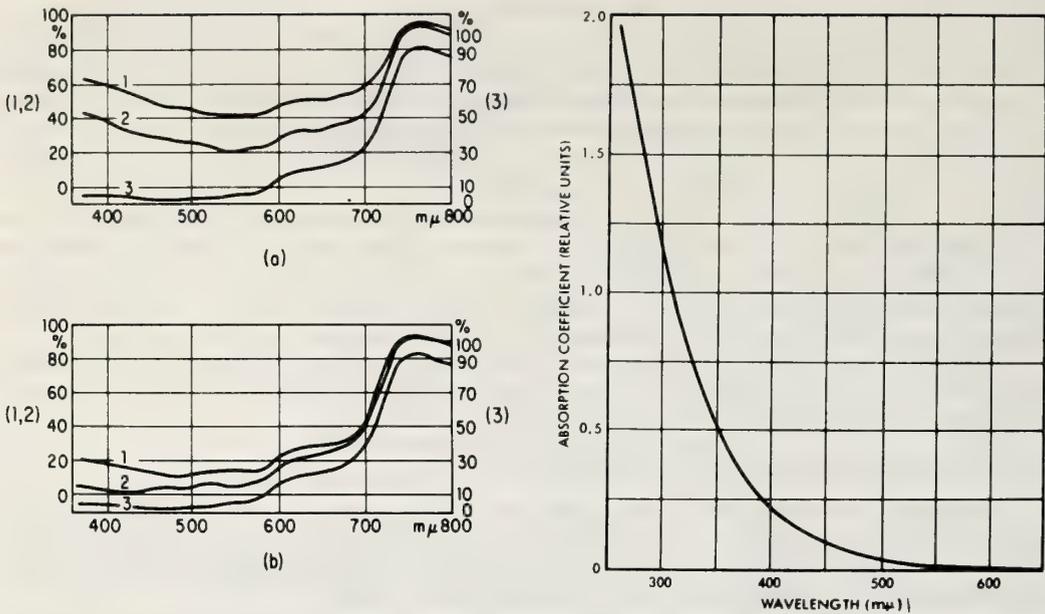


Figure 22.42 (a) Spectral extinction of coastal water (1) before and (2) after filtering, in comparison with double-distilled water (3). (b) Spectral extinction of Sargasso Sea water (1) before and (2) after filtering, in comparison with double-distilled water (3). The scale for curve 3 (double-distilled water) is on the right-hand side of the figures and vertically displaced with respect to the scale for curves 1 and 2 (according to Clark and James (1939), after Newmann and Pierson (1966)).

Figure 22.43 Absorption curve of yellow substance (after Jerlov, 1948).

Massachusetts (Yentsch, 1960). His results, in (F22.44) show that all of the populations exhibit (1) a strong, broad absorption maximum in the blue related to chlorophyll and carotenoid absorption, (2) an absorption minimum in the green, and (3) a minor absorption maximum at $.680 \mu\text{m}$ due to chlorophyll *a* (Yentsch, 1960). This suggests that naturally occurring phytoplankton populations in the oceans may have sufficiently uniform absorption properties that their concentration can be detected and quantified by spectral signature or radiance integral inversion techniques.

In order to improve the computational utility of his data, Yentsch(1960) determined the combined absorption coefficient of pure water and chlorophyll in varying concentration from $0.5\text{-}200 \text{ mg/m}^3$. His spectral absorption curves, in (F22.45), show clearly the strong, broad absorption band in the blue and the sharp absorption peak at $.680 \mu\text{m}$. Of particular interest is the shift in minimum absorption from the blue-green to the green-yellow part of the spectrum. It may be of interest to the reader mentally to combine the pure water absorption in (F22.40) with the phytoplankton absorption in (F22.44) in order to see the similarity of the result of this combination with Yentsch's combined absorption in (F22.45).

22.3.2.3 Calculations of Backscattered Light

Hulbert (1943) derived a general expression for back-scattered light from the sea, including also the effect of surface reflection. Ramsey (1968) has used this expression together with Yentsch's absorption coefficient for pure water and plant pigments, (F22.45), in order to calculate the spectral reflectance of the deep ocean for a given solar zenith angle and atmospheric turbidity. His results, in (F22.46), show the change from blue color with low chlorophyll concentrations to the predominant green and yellow colors for heavy chlorophyll concentrations.

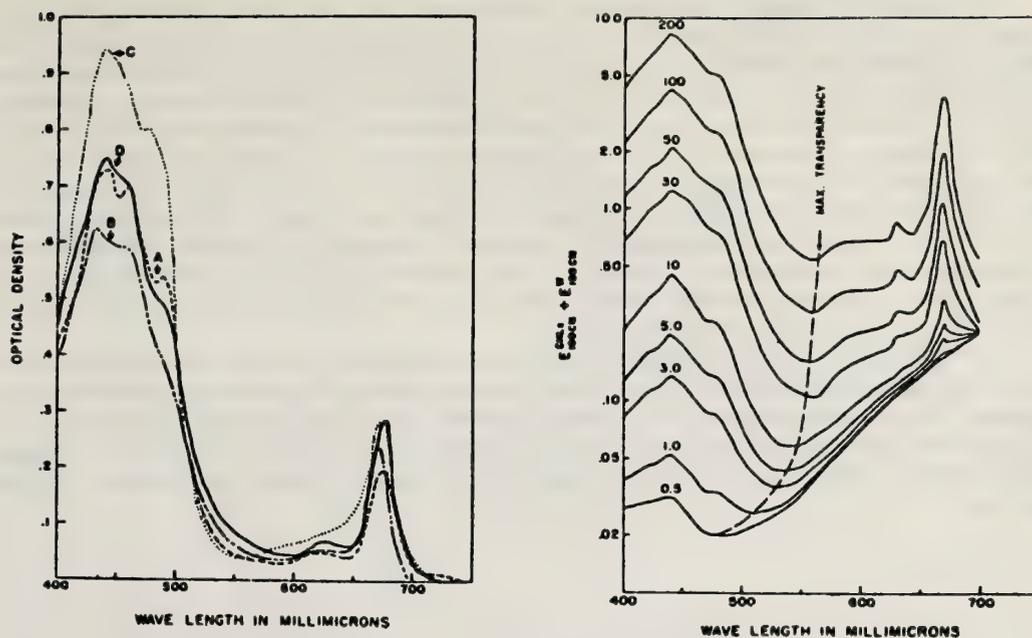


Figure 22.44 Pigment Spectra of Living Phytoplankton, (a) Diatom *Cyclotella* sp., (b) Dinoflagellate, *amphidium* sp., (c) green Flagellate *chlamydomonas*, (d) Natural population sampled from Woods Hole waters (after Yentsch, 1960).

Figure 22.45 Combined absorption coefficient for pure water and plant pigments. Numbers adjacent to the curves indicate the chlorophyll concentration in mg/m^3 (after Yentsch, 1960).

In the blue-green spectrum, 0.4-0.5 μm , of F22.46, increasing chlorophyll concentration from 0 to 10 mg/m^3 causes a systematic reduction in backscattered light. However, for concentrations $> 10 \text{ mg}/\text{m}^3$, the amount of backscattered light appears to be relatively constant. This suggests there is little or no information content in this part of the spectrum for chlorophyll concentrations of 10 to 100 mg/m^3 , although the backscattered light is particularly sensitive to changes of low chlorophyll concentrations.

In the green-yellow-red spectrum, 0.5-0.7 μm , however, the reflectance systematically increases for concentrations from 0 to 100 mg/m^3 . This portion of the spectrum appears to be useful for remote detection of all concentration levels.

22.3.2.4 Experimental Measurement of Backscattered Light

An important question is whether the calculations of Ramsey (1968), using Hulbert's theoretical expression for upwelling light from the sea and Yentsch's (1960) absorption data for pure water and phytoplankton, can be verified with experimental data.

In order to attempt such a verification, investigators at Woods Hole Oceanographic Institution (WHOI) have obtained measurements of upwelling spectral radiance by aircraft and also obtained simultaneous *in situ* measurements of chlorophyll content over the range 0.1 to 3.0 mg/m^3 . Their results have been published by Clarke, et al. (1970 and 1971) and Ewing (1971) and are shown in (F22.47). The locations of the samples, chlorophyll concentrations, and mean square slopes of the spectral reflectance curves are also given in that figure. Clarke and Ewing's data show that with increasing concentrations of chlorophyll the energy in the blue region of the spectrum decreased markedly, whereas the energy at longer wavelengths, green, tends to increase a small amount. Ewing (1971) says that "these effects can be explained by the well-known absorption of chlorophyll *a* which is particularly large in the blue (Yentsch, 1960)."

To examine the question of experimental verification of Ramsey's calculations, it is of interest to compare (F22.46) and (F22.47). One sees the slopes of the reflectance curves are in general agreement but differ in some detail. For example, the relatively strong absorption band for plant pigments near $.450 \mu\text{m}$ in Ramsey's calculations do not show up as strongly in the aircraft data. Another difference in the curves is the point of reversal in reflectance which, in Ramsey's calculations, is centered near $0.500 \mu\text{m}$ but is somewhat uncertain in the aircraft data. Perhaps this could be resolved if it were true that the A station water was deficient in yellow substance compared to stations B-D, (which is reasonable considering the source and distribution of yellow substance). If true, the A station curve would be displaced toward shorter wavelengths which would allow the aircraft data to support a reversal in reflectance at about $0.500 \mu\text{m}$. The aircraft data also suggest that at stations B-D water is relatively high in yellow substance, because the measured values of backscattered light are considerably lower than those of Ramsey's calculations which did not include the absorption of yellow substance.

The high negative correlation between chlorophyll concentration and mean squared slope of the spectral radiance curve (Table in F22.47) is extremely promising for those interested in remote detection of chlorophyll. These data also suggest that other natural variables which scatter and absorb light in the ocean, such as yellow substance, must be better understood before either theoretical calculations or empirical relations between spectral radiance and chlorophyll can be used to support global surveys of ocean chlorophyll.

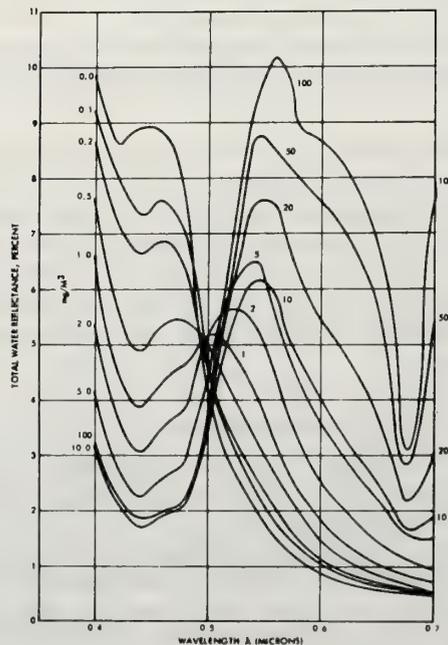


Figure 22.46 *Calculated spectral reflectance of deep ocean with varying amounts of chlorophyll, clear sunny day - 45° solar zenith angle, includes surface reflection from smooth ocean with turbid atmosphere (after Ramsey, 1968).*

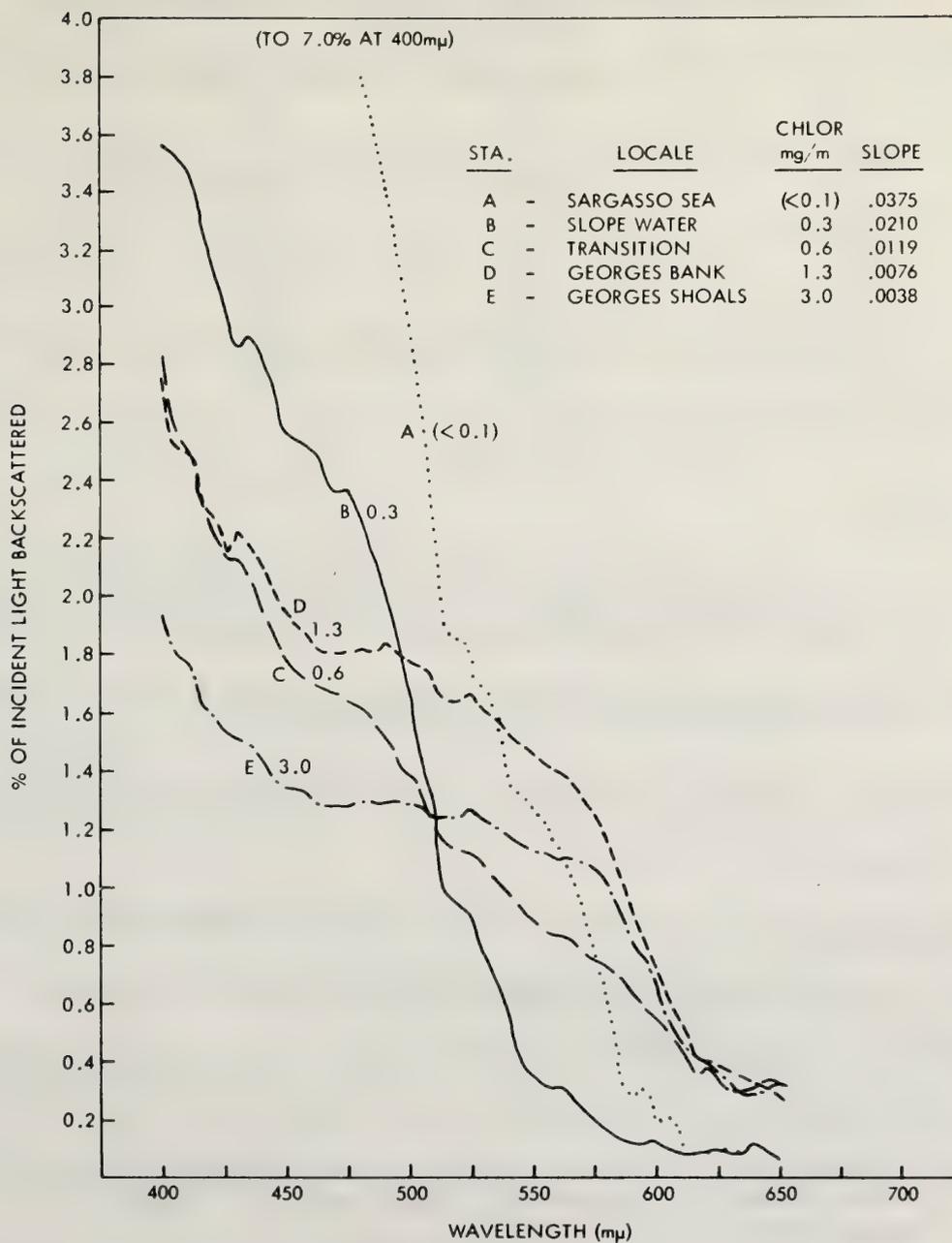


Figure 22.47 Spectra of backscattered light measured from the aircraft at 305 m on 27 Aug. 1968 at the following stations and times (all E.D.T.): Station A, 1238 hours; Station B, 1421 hours; Station C, 1428.5 hours; Station D, 1445 hours; Station E, 1315 hours. The spectrometer with polarizing filter was mounted at 53° tilt and directed away from the sun. Concentrations of chlorophyll a were measured from shipboard as follows: on 27 Aug., Station A, 1238 hours; on 28 Aug., Station B, 0600 hours; Station C, 0730 hours; Station D, 1230 hours (after Ewing, 1971).

Table 22.6 Colors and Corresponding Electromagnetic Wavelengths

(After Kondratyev, 1969)

Color	Wavelength Interval	Typical Wavelength
	$m\mu$	$m\mu$
Violet	390-455	430
Dark blue	455-485	470
Light blue	485-505	495
Green	505-550	530
Yellow-green	550-575	560
Yellow	575-585	580
Orange	585-620	600
Red	620-760	640

22.4 References

- Allison, L. J., L. L. Foshee, G. Warnecke and J. C. Wilkerson (1967), An analysis of the north wall of the Gulf Stream utilizing NIMBUS II High resolution infrared measurements, (paper presented at the 48th meeting of the American Geophysical Union, April 17, 1967, Washington, D. C.).
- Allison, L. J. and J. S. Kennedy (1967), An evaluation of sea surface temperature as measured by NIMBUS I high resolution infrared radiometer, TN D-4078, NASA, Washington, D. C.
- Anding, D., and R. Kauth (1969), Atmospheric modeling in the infrared spectral region: Atmospheric effects on multispectral sensing of sea-surface temperature from space, Report 2676-1-P, Willow Run Laboratories, Institute of Science and Technology, The University of Michigan, Ann Arbor, Michigan.
- Anding, D. and R. Kauth (1970), Estimation of sea surface temperature from space, *Remote Sensing of Environment*, 1, 217-220.
- Anonymous, 1957: *International Lighting Vocabulary – Comm. Intern. Eclairage*, Publ. 1(1): 136 pp.
- Aukland, G. C., P. J. Caruso, Jr., and W. H. Conway (1969), Remote sensing of the sea conditions with microwave radiometer systems, *Proceedings of the Sixth International Symposium on Remote Sensing of Environment, II*, Univ. of Michigan, Ann Arbor, 709-720.
- Badgley, F. J. (1966), Heat budget of the surface of the Arctic Ocean, *Proceedings of the Symposium on the Arctic Heat Budget and Atmospheric Circulation*, Edited by J. O. Fletcher, The Rand Corporation, (RM-5233-NST).
- Blanchard, D., 1963: *Electrification of the atmosphere, Progress in Oceanography*, Vol. I, M. Sears (Ed.) Pergmon Press, N. Y., p. 151.
- Bristor, C. L. (1968), Computer processing of satellite cloud pictures, NESC TM3, NESC, ESSA, Washington, D. C.
- Buettner, K. J. K., (1968), On the uses of intermediate infrared and microwave infrared in meteorological satellite, Semi-annual report, Contact NASA NSG-632, 25 pp., Univ. of Washington, Seattle.

- Buettner, K. J. and C. D. Kern (1965), The determination of infrared emissivities of terrestrial surfaces, *J. Geophys. Res.*, 70(6).
- Cardone, V. J. (1969), Specification of the wind field distribution in the marine boundary layer for wave forecasting, Geophysical Sciences Lab., New York University, Report TR 69-1.
- Clarke, G. L., 1961: The significance of spectral changes in light scattered by the sea, *Remote Sensing in Ecology*, Ed. P. J. Johnson, Univ. of Georgia Press, Athens.
- Clarke, G. L. and H. R. James (1939), Laboratory analysis of the selective absorption of light by sea water, *J. Opt. Soc. Amer.*, 29, 43-55.
- Clarke, G. L. and G. Ewing (1971), Remote sensing of ocean color from aircraft, Third annual Earth Resources Program Review, Vol. III, Section 59, NASA, MSC, Houston, Texas.
- Clarke, G. L., Ewing, G. C., and Lorenzen, C. J. (1969), Remote measurement of ocean color as an index of biological productivity, Proc. of the Sixth International Symp. of Remote Sensing of the Environment, University of Michigan, Ann Arbor, Michigan, 2, 991-1001.
- Clarke, G. L., G. Ewing and C. Lorenzen (1970), Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration, *Science*, 167, 3921.
- Committee on Colorimetry, 1966: *The Science of Color*, 6th Ed. Opt. Soc. Amer., Washington, D. C. 385 pp.
- Cox, C. and W. Munk (1954a), Measurements of the roughness of the sea surface from photographs of the sun's glitter, *J. Optical, Soc. Amer.*, 44, 838-850.
- Cox, C. and W. Munk (1954b), Statistics of the sea surface derived from sun glitter, *J. Marine Res.*, 13, 198-227.
- Curtis, W. R. and P. K. Rao (1969), Gulf Stream thermal gradients from satellite, ship and aircraft observations, *J. Geophys. Res.*, 74, 28, 6984-6990.
- Droppleman, J. D., Mennella, R. A., and Evans, D. E. (1970), An airborne measurement of the salinity variations of the Mississippi River outflow, *J. Geophysical Res.*, 75, 30, 5909-5913.
- Ewing, G. (1969), The color of the ocean, Report of the Conference, 5-6 Aug. 1969, Woods Hole Oceanographic Institution, Cambridge, Mass.
- Ewing, G. (1971), Remote spectrography of ocean color as an index of biological activity, Proceedings of the Symposium on Remote Sensing in Marine Biology and Fisheries Resources, Remote Sensing Center, Texas A&M University, College Station, Texas.
- Forel, F. A. (1895), LeLeman: Monographic Limnologique, Vol. II, Lausanne, F. Ronge, 651 pp.
- Forel, F. A., 1901: *Handbook der seenkunde: Allegemeine Limnologie*, Stuttgart, 249 pp.
- Frederick, M. A. (1970), An atlas of secchi disc transparency measurements and forel-ule color codes for the oceans of the world, Naval Postgraduate School Thesis, 188 pp., AD-718 333.
- Fritz, S. (1963), The diurnal variation of ground temperature as measured from TIROS II, *J. Appl. Meteorol.* 2:5, 645-648.

- Gunn, K. L. and T. W. R. East (1954), Microwave properties of precipitation particles, Quarterly J. Royal Meteorol. Society, 80, 522-545.
- Hansen, D. and G. Maul, (1970), A note on the use of sea surface temperature for observing ocean currents, Remote Sensing of Environment, 1, 161-164.
- Hidy, G. M., W. F. Hall, W. N. Hardy, W. W. Ho, A. C. Jones, A. W. Love, M. J. VanMelle, H. H. Wang and A. E. Wheeler (1971), Development of a satellite microwave radiometer to sense the surface temperature of the worlds oceans, S-BAND Radiometer, Final Report, Contract No. NAS 1-10106, Space Division, North American Rockwell.
- Hollinger, J. P. (1970), Microwave measurements of sea surface, J. of Geophys. Res. 75:27, 5201-5214.
- Hollinger, J. (1971), Passive microwave studies, Third annual Earth Resources Program Reviewer, Vol. III, Sect. 69, NASA, MSC, Houston, Texas.
- Holmes, R. W. (1971), The secchi disc in turbid coastal waters, Limn. and Ocean., to be published.
- Holter, M. (1970), Chapter 3, Imaging with non-photographic sensors, in *Remote Sensing with Special Reference to Agriculture and Forestry*, National Academy of Sciences, Washington, D. C.
- Hulbert, E. O. (1943), Propagation of radiation in a scattering and absorbing medium, J. Opt. Soc. Amer., 33, 42-45.
- Hulbert, E. O. (1945), Optics of distilled and natural water, J. Optical Soc. Amer., 35:11, 698-705.
- Huston, W. B. (study manager) (1971), Earth observatory satellite (EOS), Definition Phase Report, Vol. 1, Goddard Space Flight Center, Greenbelt, Md.
- Jerlov, N. G. (1948), Optical studies of ocean water, Reports of the Swedish Deep Sea Expedition, Vol. III, Physics and Chemistry, No. 1.
- Jerlov, N., 1968: *Optical Oceanography*, Elsevier Publishing Company. Amsterdam, London, New York.
- Joseph, J. and H. Wattenberg, 1944: Untersuchungen uber die optischen verhältnisse inn Meere. Mitteilungen Chef Hydr. Dienstes, Oberkomm, d. Kriegsmarine (Berlin).
- Kalle, K. (1938), Zum problem der meerliasserfarbe, Ann. d. Hydr. u. Maril, Meteorol., 66, 1.
- Kondratyev, K. Ya., 1969: *Radiation in the Atmosphere*, International Geophysical Series, Vol. 12, Academic Press, New York and London.
- Kuhn, P. (1963), Soundings of observed and computed IR flux, Journal of Geophysical Research, 68:5, 1415-1420.
- Kuhn, P., 1972: Parameterizing continuum radiation to remove the calculated-observed radiative transfer discrepancy, in Sea Surface Topography From Space, Key Biscayne, Florida, Oct., 1971.
- Kuhn, P. and L. Stearns (1971), Radiative transfer observations and calculations during BOMEX, NOAA Technical Report, ERL 203- APCL 19, April, 1971, 171 p.
- Kuhn, P. and L. Stearns (1972), Radiosonde humidity retrieval by simultaneous radiation measurements, Journal of Applied Meteorology, 11, 1.

- LaViolette, P. E. and P. L. Chabot (1968), NIMBUS II satellite sea surface temperature versus historical data in a selected region: A comparative study, *Deep Sea Research*, 15, 617-622.
- LaViolette, P. E. and P. L. Chabor (1969), A method of eliminating cloud interference in satellite studies of sea-surface temperature, (unpublished).
- LaViolette, P. E. and S. E. Seim (1969), Satellites capable of oceanographic data acquisition – a review, TR-215, U. S. Naval Oceanographic Office, Washington, D. C.
- LeGrand, Y. (1939), The penetration of light into the sea. *Annales l'Inst. Oceanogr. (Monaco)*, 19, 393-436.
- Leese, J. (1971), National Environmental Satellite Center, NOAA, Wash. D. C. (private communication).
- Leese, J., W. Pichel, B. Goddard and R. Brower (1971), An experimental model for automated detection, measurement and quality control of sea-surface temperatures from ITOS IR data. Proc. of the Intl. Symposium on Remote Sensing of the Environment, Univ. of Michigan, Ann Arbor.
- Lewis, E. A., J. P. Casey, and A. J. Vaccaro (1954), Polarized radiation from certain thermal emitters, Electronics Research Directorate, Air Force Cambridge Research Center, AFCRL Tech. Rep. No. 54-G.
- Maul, G. and D. Hansen (1971), An observation of the Gulf Stream surface front structure by ship, aircraft and satellite, *Remote Sensing of Environment*, 2.
- McClain, P. (1970), Applications of environmental satellite data to Oceanography and Hydrology, ESSA Technical Memorandum NESCTM 19, U. S. Dept. of Commerce, ESSA, NESS, Washington D. C.
- McMahon, H. O. (1950), Thermal radiation from partially transparent reflecting bodies, *J. Opt. Soc. Amer.* 40, 376-380.
- Meyer, J. W. (1966), Radio astronomy at millimeter and submillimeter wavelength, *Proc. IEEE.* 54, 484-492.
- Monahan, E. C. (1969), Freshwater whitecaps, *J. Atmos. Sci.* 26, 1026-1029.
- Morel, A. (1966), Étude expérimentale de la lumière par l'eau les solutions de chlorure de sodium et l'eau de mer optiquement pures, *J. Chim. Phys.*, 10, 1359-1366.
- Mount, Wayne, D. (1969), Use of the millimeter wave radiometry to remotely measure atmospheric stability, *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, 1, Univ. of Michigan, Ann Arbor, 469-491.
- Muromtsev, A. M. (1963), The principal hydrological features of the Pacific Ocean (Trans. of Osnovnye Cherty gidroogii Tikhogo Okeana, Gidrometeor – ologicheskoe Iz datel'stoe, Leningrad 1958) Israel Program of Sci. Trans, Jerusalem.
- NASA (1966), NIMBUS II User's Guide, Aracon Geophysics Company, Goddard Space Flight Center, Greenbelt, Maryland.
- Neumann, G. and W. Pierson, Jr., 1966: *Principles of Physical Oceanography*, Prentice-Hall, Inc., Englewood Cliffs, N. J.
- Nordberg, W., Conaway, J. W., Ross, D. B., and Wilheit, T. (1971), Measurement of microwave emission from a foam covered, wind driven sea, *J. Atmos. Sci.*, 28, 3, 429-435.

- Paris, Jack F. (1969), Microwave radiometry and its application to marine meteorology and oceanography, Ref. No. 69-IT, Dept. of Oceanography, Texas A&M University, College Station, Texas, January.
- Paris, J. F. (1971), Transfer of thermal microwaves in the atmosphere, Vols. I and II, Dept. of Meteorology, Texas A&M University, College Station, Texas, Contract Final Report, NASA Grant NGR 44-001-098, NASA, GSFC, 211 pp.
- Pascalar, H. G. and R. T. Sakamoto (1965), Microwave radiometric measurements of ice and water, Proc. 3rd Symp. Remote Sensing of Environ., Ann Arbor, Mich., 803-812.
- Plank, M. (1900), *Ann. Phys.* 4:533.
- Polaym, F. (1971), Measurements of water depth by multispectral ratio techniques, Third Annual Earth Resources Program Review, Vol. III, Section 61, NASA, MSC, Houston, Texas.
- Porter, Ronald A. and Frank J. Wentz III. (1971), Microwave radiometric study of ocean surface characteristics, Contract 1-35140, Radiometric Technology, Inc., Wakefield, Mass. 01880.
- Ramsey, R. C. (1968), Study of the remote measurement of ocean color, Final Report, Prepared by TRW for NASA Headquarters, Washington, D. C., Contract NASW-1658.
- Rao, P. K. and J. S. Winston (1963), An investigation of some synoptic capabilities of atmospheric "window" measurements from satellite TIROS II, *J. Applied Meteorol.* 2, 1, 12-23.
- Roeder, R. S. (1967), Airborne measurements with an AN/AAR-33 radiometric search set, Sperry Microwave Electronics Company, A Division of Sperry Rand Corporation.
- Ross, D. B., V. J. Cardone, J. W. Conaway, Jr., (1970), Laser and microwave observations of sea-surface conditions for fetch-limited 17-25-M/s winds, *IEEE Transactions on Geoscience Electronics*, GE-8, 4, 326-336.
- Ross, D. S. (1969), Enhanced oceanographic imagery, Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Vol. II, Univ. of Michigan, Ann Arbor, 1029-1044.
- Ross, D. S., and Jensen, R. C. (1969), Experiments in oceanographic aerospace photography: Ben Franklin special filter tests: NASA, 2nd Annual Earth Resources Aircraft Program Status Review, Vol. III, Houston, Tex., 51-1 to 51-32.
- Schott, G., 1942: *Geographic des Atlantischen Azeans*, C. Boysen, Hamburg, 3rd Edition (1942; reprinted in 1944).
- Sherman, J. W. III, 1969: Passive microwave sensors for satellites, Proceedings of the Sixth International Symposium on Remote Sensing of Environment, II, Univ. of Michigan, Ann Arbor, 651-670.
- Simon, I., 1966: *Infrared Radiation*, Van Nostrand Company, Inc., Princeton, New Jersey.
- Sirounian, V. (1968), The effect of temperature, angle of observation, salinity and thin ice on microwave emission of water, *J. Geophys. Res.*, 73, 14, 4481-4486.
- Smith, R. C. and J. E. Tyler (1967), Optical properties of clear natural water, *J. Opt. Soc. Amer.*, 57, 5, 589-595.
- Smith, W. L., P. K. Rao, R. Koffler, and W. R. Curtis, (1970), The determination of sea-surface temperature from satellite high resolution infrared window radiation measurements, *Monthly Weather Review*, 98, 8, 604-611.

- Stogryn, A. (1967), The apparent temperature of the sea at microwave frequencies, *IEEE Trans. Ant. Prop. AP-15*, 2, 278-286.
- Sullivan, S. A. (1963), Experimental study of the absorption in distilled water, artificial sea water, and heavy water in the visible region of the spectrum, *J. Opt. Soc. Amer.* 53, 963-967.
- Sverdrup, H. U., M. W. Johnson and R. H. Fleming, 1942, 1st edition; 1946 2nd edition: *The Ocean, Their Physics, chemistry and General Biology*. Prentice Hall, Englewood Cliffs.
- Tyler, J. E. (1968), The secchi disc, *Limn. and Ocean.*, 13, 1, 1-6.
- Vary, W. (1969), Remote Sensing by aerial color photography for water depth penetration and ocean bottom detail, in, *Proceedings of the Sixth International Symposium on Remote Sensing of Environment, II*, Univ. of Michigan, Ann Arbor, 1045-1060.
- Warnecke, G., L. J. Allison and L. L. Foshee (1967), Observation of sea surface temperatures and ocean currents from NIMBUS II, NASA Report X-622-67-435, Greenbelt, Maryland.
- Warnecke, G., L. Allison, L. McMillin, K. H. Szekulda (1971), Remote sensing of ocean currents and sea surface temperature changes derived from the NIMBUS II satellite, *J. of Phys. Oceanogr.*, 1, 1, 45-60.
- Watson, L. A. (1971), National Environmental Satellite Center, NOAA, Washington, D. C. (private communication).
- Wilheit, T., J. Blim, W. Campbell, A. Edgerton, N. Nordberg (1971), Aircraft measurements of microwave emission from arctic sea ice, X651-71-417 (preprint) NASA, Goddard Space Flight Center, Greenbelt, Md.
- Williams, G. F., Jr. (1969), Microwave radiometry of the ocean and the possibility of marine wind velocity determination from satellite observations, *J. Geophys. Res.*, 74, 4591-4594.
- Wittman, W. and J. Schule (1966), Comments on the mass budget of Arctic pack ice, *Symposium on the Arctic Heat budget and atmosphere circulation*, edited by J. O. Gletcher. The Rand Corporation (RM-5233-NSF).
- Wolfe, W. L. (ed.) (1965), *Handbook of Military infrared technology*, Office of Naval Research, Washington, D. C.
- Yentsch, C. S. (1960), The influence of phytoplankton segments on the color of sea water, *Deep Sea Res.*, 7, 1-9.
- Yentsch, C. S. (1962), Measurement of visible light absorption of particulate matter in the ocean, *Deep Sea Res.*, 6, 207-217.
- Yost, Edward and Sondva Wenderoth (1969), Agricultural and oceanographic applications of multispectral color photography, *Proceedings of the Sixth International Symposium on Remote Sensing of Environment, I*, Univ. of Mich., Ann Arbor, 145-173.

LIST OF SYMBOLS

Symbol	Name	Dimension, Units or Value
c_1	radiation constant	$3.7415 \times 10^{-5} \text{ erg cm}^2 \text{ s}^{-1}$
c_2	radiation constant	1.4388 cm K
N_λ	spectral radiance	$\text{erg cm}^{-1} \text{ s}^{-1} \text{ sr}^{-1} \mu\text{m}^{-1}$
P	polarization	
S_s	salinity of active layer	‰
T	molecular temperature	C, K
T_B	brightness temperature	K
T_{BP}	polarized brightness temperature	K
T_e	effective temperature	K
T_{eP}	polarized effective temperature	K
T_o	molecular temperature of active layer	C, K
T_a	temperature of the atmosphere	C, K
T_s	sea surface molecular temperature	C, K
T_{so}	satellite observed temperature	K
v	wind velocity	ms^{-1}
$W_{a\lambda}$	spectral radiant emittance of the atmosphere	$\text{erg cm}^{-2} \text{ s}^{-1} \mu\text{m}^{-1}$
$W_{B\lambda}$	spectral radiant emittance of a blackbody	$\text{erg cm}^{-2} \text{ s}^{-1} \mu\text{m}^{-1}$
$W_{e\lambda}$	effective spectral radiant emittance	$\text{erg cm}^{-2} \text{ s}^{-1} \mu\text{m}^{-1}$
W_λ	spectral radiant emittance of a greybody	$\text{erg cm}^{-2} \text{ s}^{-1} \mu\text{m}^{-1}$
ϵ_λ	spectral emissivity	—
θ	zenith angle	deg.
λ	wavelength	μm
ν	frequency	GHz
ρ_λ	spectral reflectivity	—
τ_λ	spectral transmissivity through a layer which extends from a given level in the atmosphere to the top of the atmosphere	—
$\tau_{\lambda a}$	spectral transmissivity of the entire atmosphere	—
Ψ	nadir angle	deg.

Chapter 23 ATMOSPHERIC REMOTE SENSING WITH LASER RADAR

R. G. Strauch and A. Cohen†

Wave Propagation Laboratory

Environmental Research Laboratories

National Oceanic and Atmospheric Administration

This chapter discusses the methods used in laser radar probing of the atmosphere. The systems equations used in computing signals and noise are presented and results of Raman, Rayleigh, and Mie scattering experiments are used to illustrate how laser radars can contribute to remote sensing of the atmosphere.

23.0 Introduction

Numerous experiments with laser radar, sometimes called “lidar,” have been conducted in the atmosphere to demonstrate their capability to remotely measure atmospheric conditions and properties of atmospheric scatterers. These experimental devices are now approaching a stage of maturity that will enable quantitative measurements of meteorological parameters from ground stations. Specialized laser devices for cloud observation and visibility measurements are in operation. Meteorological laser radar systems capable of measuring vertical profiles of water vapor, temperature, aerosols, wind and turbulence are not operational but substantial progress has been made in demonstrating their feasibility.

The purpose of this chapter is to review the status of meteorological laser probing of the lower atmosphere, to indicate what measurements can now be made, to indicate what capabilities should soon be possible and to discuss future directions for atmospheric probing.

23.1 Optical Interactions in the Lower Atmosphere

Meteorological information obtained by laser probes depends on the interaction of the laser (transmitter) energy with the atmosphere. Atmospheric interactions attenuate the transmitted laser energy and introduce random fluctuations of amplitude and phase. Some of the attenuated laser energy appears as scattering in the direction of the receiver. It is this scattered light that must be used to obtain the information concerning the atmosphere. The important atmospheric effects for laser radar are therefore described by the extinction coefficient and by scattering cross sections. The reader is referred to Chapter 10 for a detailed discussion of scattering theory.

23.1.1 Extinction Coefficient and Cross Sections

The attenuation of a laser beam transmitted through a homogeneous layer of air of thickness x is described by the relationship

$$I = I_0 e^{-\beta x} \quad , \quad (23:1)$$

sometimes known as Bouguer’s law. I_0 is the intensity incident on the layer, I is the intensity leaving the layer and β is called the extinction coefficient (m^{-1}) and accounts for all the interactions that extract energy from the laser beam. In general, we have an inhomogeneous path with $\beta = \beta(x)$, and the extinction law is $I(x) = I_0 \exp\left[\int_0^x \beta(s) ds\right]$. For atmospheric work it is usual to consider β as a sum of three major components,

$\beta = \beta_R + \beta_M + \beta_A$, where β_R is the extinction caused by molecular scattering (Rayleigh scattering), β_M is the extinction caused by aerosol particles (Mie scattering) and β_A is the extinction from absorption by molecules or aerosols. Other scattering processes such as Raman scattering are usually much weaker and are negligible

†National Research Council–NOAA Resident Research Associate.

when considering the extinction coefficient. The Rayleigh extinction coefficient can be readily calculated as a function of altitude either for a standard atmosphere or if the pressure profile is known. The earliest uses of laser radar were for studying aerosol density profiles by subtracting the expected Rayleigh contribution from the measured backscatter. Standard tables of Mie extinction coefficients have been prepared but the aerosol content is highly variable so these tables must be used with discretion.

Each extinction process is proportional to the density of interacting particles. The strength of the interaction with each particle can be described by a cross sectional area that interferes with the beam. (Atmospheric absorption is not usually treated in terms of cross sections). For identical particles, we have $\beta = \rho\sigma$ where ρ is the number density (m^{-3}) of particles and σ is the cross section (m^2) of a single scatterer.

23.1.2 Differential Cross Section

When the scattering is isotropic, the scattering cross section per unit solid angle is $\frac{\sigma}{4\pi}$ (per scatterer). In general the angular dependence is not isotropic and must be specified by using a differential cross section $\sigma(\theta, \phi) = \frac{d\sigma}{d\Omega}$, where θ is the scattering angle, ϕ is the angle of polarization, and Ω is the solid angle (F23.1). For unpolarized light, $\sigma(\theta, \phi) = \sigma(\theta)$. The quantity $\rho\sigma(\pi)$ is called the volume backscatter coefficient ($\text{m}^{-1} \text{ster}^{-1}$).

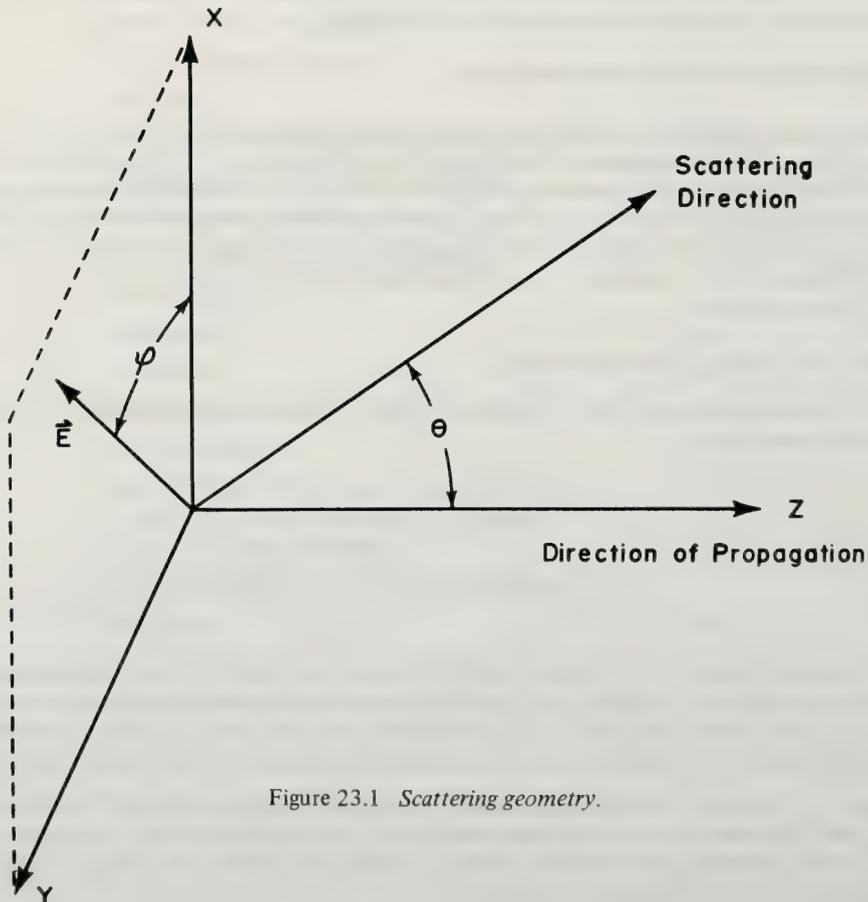


Figure 23.1 Scattering geometry.

23.1.3 Rayleigh Scattering

The Rayleigh scattering cross section (σ_R) can be expressed on a per molecule basis as (Penndorf, 1957)

$$\sigma_R = \frac{8\pi^3 (m_s^2 - 1)^2}{3\lambda^4 N_S^2} \frac{6 + 3\delta}{6 - 7\delta}, \quad (23:2)$$

where N_S is the molecular number density and m_s is the refractive index for a standard atmosphere, λ is the wavelength of the scattered light and δ is a depolarization factor to account for the anisotropy of the polarizability of air molecules ($\delta \approx 0.035$). The Rayleigh extinction coefficient is given by $\beta_R(h) = \rho(h)\sigma_R$, where $\rho(h)$ is the number density of air molecules at height h . The extinction due to Rayleigh scattering in length L is $\exp \int_0^L \beta_R(s) ds$. The scattering is not isotropic and the angular scattering function (assuming isotropic air molecules) is given by (Panofsky and Phillips, 1955)

$$\Psi_R = \frac{3}{8\pi} (\cos^2 \theta \cos^2 \phi + \sin^2 \phi). \quad (23:3)$$

Note that $\int_{4\pi} \Psi_R d\Omega = 1$. For unpolarized light, Ψ_R is averaged over the angle ϕ and the expression reduces to $3/8\pi \left(\frac{\cos^2 \theta + 1}{2} \right)$. This expression must be corrected for the molecular anisotropy and the correct angular function for unpolarized light is $0.06071(1 + 0.932 \cos^2 \theta)$ (Penndorf, 1957). The scattering cross section per unit solid angle is given by $\sigma(\theta, \phi) = \sigma_R \Psi_R$.

It is important to note that for Rayleigh backscatter ($\theta = \pi$), the extinction coefficient and the volume backscatter coefficient, are uniquely related,

$$\rho \sigma_R(\pi) = \frac{3}{8\pi} \rho \sigma_R = \frac{3}{8\pi} \beta_R \quad (23:4)$$

because only for pure Rayleigh scattering ($\beta = \beta_R$) is it possible to determine ρ uniquely from monostatic laser radar measurements.

23.1.4 Mie Scattering in the Atmosphere

The density and size distribution of aerosols in the atmosphere vary widely from hour to hour in many locations so that care must be taken in using a unique extinction coefficient (β_M) for Mie scattering. The angular scattering patterns for Mie scattering are not as simply described as those for Rayleigh scattering.

A summary of the major points that allow Mie scattering to be considered in terms of a cross section, σ_M , and angular distribution Ψ_M are included here (Call, Palmer and Grow, 1967). For a single spherical homogeneous particle, the complex refractive index and the radius are needed to describe the scattering properties of the particle. The complex refractive index is $m = n - ik$, where n is typically 1.5 and the absorption loss factor k is of the order of .001. The particle radius is described by the "size parameter" $\alpha = \frac{2\pi r}{\lambda}$

where r is the radius and λ the wavelength of the scattered light. The ratio of the scattering cross section, $\sigma_M(m, \alpha)$, to the geometric cross section is

$$k_M = \frac{\sigma_M(m, \alpha)}{\pi r^2} \quad (23:5)$$

For a concentration of uniform particles, $\beta_M = \pi r^2 k_M(m, \alpha) \rho$. The factor $k_M(m, \alpha)$ represents the total scattering (4π steradians) and can be calculated in terms of infinite series that were first obtained by Mie (1908). The sizes of atmospheric particles will not be uniform and a distribution $\rho(r)$ is required to describe the aerosol content of the atmosphere. Junge's observations (Junge, 1964) led to the formulation

$d\rho(r) = Cr^{-(v+1)} dr$ for aerosol distributions in the atmosphere with $v \approx 3$ giving a model that best fits the naturally occurring distributions. Thus, the aerosol extinction coefficient can be written (for particles with the same complex refractive index)

$$\beta_M = \int \pi r^2 k_M(m, \alpha) Cr^{-(v+1)} dr \quad (23:6)$$

The angular dependent Mie scattering functions are obtained by integration over the size distribution of the angular dependent scattering function for each particle. The total scattering cross section for a single particle is related to the angular scattering by $k_M(m, \alpha) = \int_{4\pi} g(\theta, \phi, m, \alpha) d\theta d\phi$. The angular dependent scattering function is found by integrating over all particle sizes.

$$\beta_M \Psi_M = g(\theta, \phi, m, \alpha) \pi r^2 Cr^{-(v+1)} dr \quad (23:7)$$

The volume backscatter coefficient is given by $\beta_M \Psi_M(\pi)$ for unpolarized light. Computer solutions have been generated for Mie coefficients for polarized and unpolarized light (Call, Palmer and Grow, 1967). Since the scattering and the extinction are expressed by integrals over the size distribution of series solutions, it is not possible to obtain unique aerosol content from a measurement of monostatic laser radar backscatter.

23.1.5 Raman Scattering

In addition to the molecule scattering that occurs at the same wavelength as the irradiating wavelength, a weak scattering occurs at a shifted wavelength. This frequency-shifted scattering, called Raman scattering, is about two orders of magnitude weaker than Rayleigh scattering, and therefore can be neglected when considering extinction coefficients. Each molecular species exhibits a characteristic frequency shift, so that in principle, a frequency selective receiver can be used to observe the molecular density of each species. In practice, Raman scattering cross sections of atmospheric molecules are so small that few successful atmospheric Raman experiments have been conducted.

The frequency displacement of the Raman scattering is related to the rotation-vibration energy states of the molecule. Detailed calculation of Raman lines can be found in Herzberg (1960), and Mizushima (1958). The angular dependence of the scattering is similar to Rayleigh scattering. The intensity of Raman scattering varies as the inverse 4th power of the scattered wavelength so that most atmospheric Raman experiments have been conducted at ultra violet wavelengths. Raman scattering cross sections have been derived for atmospheric molecules and some laboratory measurements have been made. The volume backscatter coefficient and extinction coefficients are sufficiently well known for system analysis calculations. The reader is referred to Chapter 10 for additional discussion.

23.1.6 Atmospheric Absorption

Most laser radar systems have operated in wavelength regions where atmospheric attenuation by absorption is negligible. There are methods that rely on selective absorption by atmospheric constituents to measure the concentration of the absorber. Ozone absorption occurs throughout the visible spectrum and becomes sufficiently intense near 2800 Å so that operation without background interference from the sun is possible. At the same time, however, the horizontal extinction coefficient is increased and the maximum range of the system becomes reduced by the attenuation. Broadband attenuation has only served to reduce signal intensity in laser radar experiments. Utilization of the ozone attenuation for "solar blind" operation has been proposed but no suitable lasers are available as yet.

Selective absorption, that is absorption by a molecule over a narrow wavelength band, will become more useful as the development of tunable dye lasers yields peak powers suitable for laser radar transmitters and their wavelength control becomes more exact. The only selective absorption that has been used for atmospheric studies employs frequency shifting of a ruby laser near a water vapor absorption line (Shotland, 1965).

A laser operating on an absorption line will have an extinction coefficient of $\beta_{M_1} + \beta_{R_1} + \beta_{A_1} + \beta_{A'}$, where $\beta_{A'}$ is the resonant absorption, while a laser operating just off the resonance will have an extinction of

$\beta_{M_2} + \beta_{R_2} + \beta_{A_2}$. The wavelength dependence of $\beta_{A'}$ will be very much greater than the other factors so that the extinction due to Rayleigh and Mie scattering and the broadband absorption will be approximately equal for the two lasers. Hence, the transmission ratio of the two laser signals will be

$$\frac{I_{O_1}}{I_{O_2}} \exp \left[-\int_0^R \beta_{A'}(s) ds \right] \quad (23:8)$$

The ratio of the returned signal for a monostatic laser radar measuring scattering from aerosols, molecules, or both would be

$$\frac{I_{O_1}}{I_{O_2}} \exp \left[-2 \int_0^R \beta_{A'}(s) ds \right] \quad (23:9)$$

A determination of $\beta_{A'}(R)$ is possible from the laser measurements and one can thereby calculate a density distribution. Resonant absorption by H_2O and O_2 can be found in the wavelength range of present lasers.

23.1.7 Resonant Scattering

Enhanced scattering can occur if the frequency of the incident radiation coincides with a resonant transition of the scattering molecule. Resonant scattering from the sodium layer at 90 km altitude has been observed with a tunable laser radar system (Bowman, Gibson, and Sanford, 1969). The dye laser will enable more precise spectroscopic studies of molecular absorptions and resonance Raman or fluorescence effects. Fluorescence in NO_2 and SO_2 has been observed at atmospheric pressures (Sakurai and Broida, 1969) although the molecular collisions at atmospheric pressure usually quench or perturb fluorescence. No lower atmosphere laser radar experiments have been conducted utilizing resonance effects.

Typical scattering cross sections per particle for various scattering mechanisms are shown below:

Process	Cross Section (cm ² /ster)
Mie	10 ⁻²⁷ to 10 ⁻⁸
Rayleigh	10 ⁻²⁷
Raman	10 ⁻²⁹ to 10 ⁻³⁰
Resonance Raman	10 ⁻²³
Fluorescence	10 ⁻¹⁶ or less

23.2 Laser Radar For Atmospheric Backscatter

23.2.1 Laser Radar Equation

In order to understand how laser probes obtain information about the atmosphere, it is necessary to understand how laser radar works and to understand the mechanisms of interaction between the laser energy and the atmosphere. We will consider in this chapter pulsed laser radar systems operating in a monostatic mode so that the atmospheric interactions of interest are backscatter and extinction.

The laser radar equation tells what system parameters and what atmospheric parameters will contribute to the signal measured by a laser radar system. For Raman scattering, the laser radar equation is most useful in a slightly different form because the signal return is extremely small. Instead of considering the power transmitted and received, we consider the number of photons involved. A few preliminary remarks will help to clarify the laser radar equation and the relationship to other forms of the equation will be clear.

(1) We are dealing only in pulsed laser radar systems, with N_t photons transmitted per laser pulse. The energy per photon is $h\nu_t$ where h is Planck's constant (6.6×10^{-34} Joule-sec), and ν_t is the transmitter

frequency (sec^{-1}). The energy per pulse is $N_t h \nu_t$. The energy per pulse is also given by $P_t \tau_t$, where P_t is the laser power (watts) and τ_t is the pulse width. Therefore we have

$$N_t = \frac{P_t \tau_t}{h \nu_t} \quad (23:10)$$

(2) Figure (23.2) shows the monostatic laser radar configuration. The laser transmitter and receiver

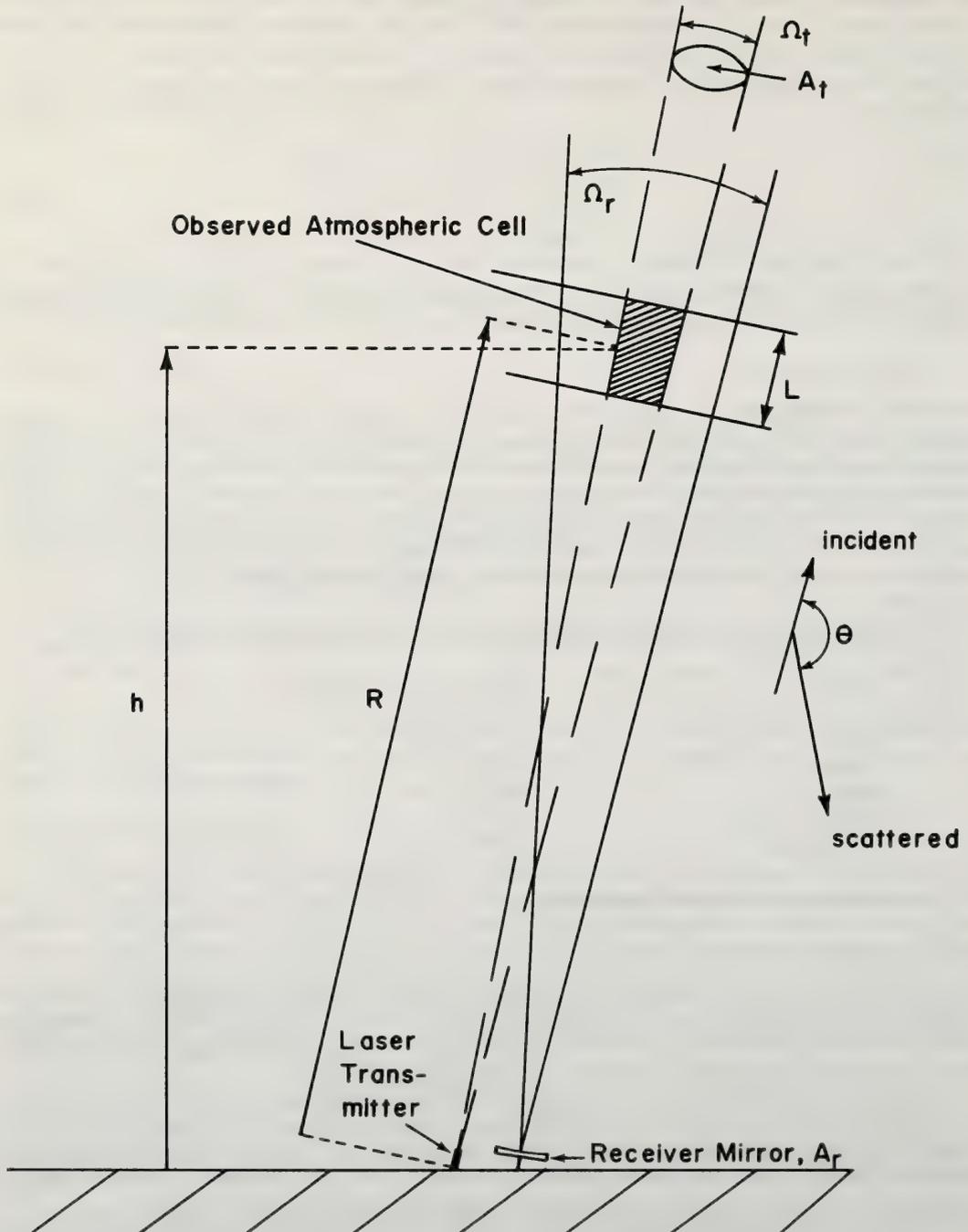


Figure 23.2 Monostatic laser radar.

mirror are shown separated but their separation is very small compared to the altitude (h) or the slant range R. We are therefore considering the monostatic backscatter case. We assume that the receiver solid angle, Ω_r , is at least as large or larger than the transmitter solid angle, Ω_t . Then the receiver will receive backscattered photons from the entire "gated" volume of atmosphere that is illuminated by the transmitter. This volume is $LA_t = L\Omega_t R^2$. The system is gated in time so that the receiver collects photons only during the time corresponding to backscatter from this volume. Note that L can be as large as we please, by simply keeping the receiver gate open longer, but can only (usefully) be as small as $c\tau_t/2$, the maximum radar range resolution, where c is the velocity of light.

(3) The number of photons incident on the volume of interest will be $N_t \eta_t \exp[-\int_0^R \beta_t(s) ds]$ where η_t is an optical efficiency factor for the optical system required to collimate the laser and β_t is the atmospheric extinction coefficient for the transmitted wavelength.

(4) The signal returned to the receiver from the cell of interest is described in terms of the scattering cross section. The volume backscatter coefficient is the backscatter cross section per meter per steradian. The total backscatter per steradian (for identical scatters) is proportional to $\rho\sigma(\pi)L$ where ρ is the number density of scatterers (m^{-3}). (Note that $\rho\sigma(\pi)$ is analogous to microwave radar cross sections used in weather radar studies, with units of m^2 per m^3 per steradian). We have assumed that $\beta_t L \ll 1$, so that the decrease of laser intensity for L meters is $(1 - e^{-\beta_t L}) \approx \beta_t L$. (If L is very large then the number of photons incident on the volume decreases with increasing R.)

(5) The fraction of scattered photons that reach the receiving mirror is given by

$$\frac{A_r}{R^2} \exp\left[-\int_0^R \beta_r(s) ds\right], \tag{23:11}$$

where $\frac{A_r}{R^2}$ is the solid angle at the receiver as seen from the atmospheric cell and β_r is the extinction coefficient of the return signal. The frequency of the backscattered signal may be different than the transmitted frequency.

(6) The number of signal photons counted by the laser radar receiver will be the number incident on the receiver collector times an optical efficiency η_r and the quantum efficiency of the detector η_q . η_q is the ratio of photons counted by the detector system to the number of photons incident on the detector.

Putting all these factors together we have the laser radar equation for the case we are considering,

$$N_r = N_t \eta_t \rho \sigma(\pi) LA_r \eta_r \eta_q \exp\left[-\int_0^R (\beta_t + \beta_r) ds\right] / R^2, \tag{23:12}$$

where N_r is the number of photons received per transmitted pulse due to backscatter from the atmospheric cell of interest.

The extinction coefficients, β_r and β_t , are due to molecular (Rayleigh) scattering, aerosol (Mie) scattering, and absorption. Their numerical values depend on the transmitted and scattered wavelengths, the aerosol content of the atmosphere and the density of the atmosphere. For ranges of interest to lower atmosphere studies (<10 km), the extinction term is small and can be approximated from standard tables if the aerosol content is low or is known. The absorption can be neglected unless the wavelength coincides with a molecular absorption line. The important parameters for meteorological laser probing are the scattering cross section and the number density of the scatterers.

23.2.2 Noise

In any measurement the signal to noise ratio is the important consideration. We have seen how the signal can be calculated. The receiver not only counts signal photons, but also counts photons received from sky background and from noise of the detector. The detector noise is called "dark noise" or "dark counts" and is due to photoelectrons emitted from the photocathode in the absence of illumination. We have computed N_r as the number of signal counts per transmitted laser pulse due to backscatter from a length L of

the atmosphere. The time that the receiver gate is open is $2L/c$. We now calculate the noise counts during this same time.

(1) Dark Counts. If there are N_D dark counts per second, then there will be $N_D(2L/c)$ dark counts per laser pulse. Some detectors are designed especially for low noise detection and N_D is specified. Usually the dark current, I_D , is specified and the dark count can be estimated as I_D/eG where e is the charge of the electron and G is the detector gain.

(2) Background Counts. As above, $N_B(2L/c)$, where N_B is the background counts per second, will be the number of background counts per laser pulse. The background counts are from natural or artificial light incident on the detector.

N_B is given by the same type of derivation used for the laser radar equation,

$$N_B = \eta_q \eta_r \frac{B_\lambda \Omega \Delta \lambda A_r}{h\nu} \quad (23:13)$$

where $\Delta \lambda$ is the spectral width of the receiver, (\AA) centered at the wavelength $\lambda = c/\nu$, Ω_r is the solid angle of the receiver and B_λ is the background brightness, usually tabulated in watts $m^{-2} \text{ster}^{-1} \text{\AA}^{-1}$.

(3) Signal-to-Noise Ratio. The total number of noise photons counted per transmitted laser pulse is therefore $N = (2L/c)(N_D + N_B)$, and the number of noise counts per second is given by ZN where Z is the pulse repetition rate of the laser (sec^{-1}). The effective rms count due to statistical fluctuations of the noise is \sqrt{ZN} . The rms count due to fluctuations of the total signal is $\sqrt{Z(N_r + N)}$, and the signal to noise ratio for a 1 second averaging will be

$$\text{Signal/noise} = \frac{ZN_r}{\sqrt{Z(N_r + 2N)}} = \frac{\sqrt{ZN_r}}{\sqrt{N_r + 2N}}$$

Since ZT will be the number of pulses in T seconds, the signal to noise ratio for T seconds of signal averaging is given by

$$\text{Signal/noise} = \frac{\sqrt{ZTN_r}}{\sqrt{(N_r + 2N)}} \quad (23:14)$$

Note that the factor of 2 in the denominator arises because the noise count is not accurately measured. If the noise is observed for a long time, then it can be accurately determined and the signal/noise becomes

$$\frac{\sqrt{ZTN_r}}{\sqrt{N_r + N}}$$

23.2.3 Noise Considerations

(1) A detector with a low dark count (current) and high quantum efficiency is desired.

(2) The spectral bandwidth of the transmitter and receiver should be small so $\Delta \lambda$ can be as small as possible.

(3) The signal received during time T , is ZTN_r and depends on the average laser power. For the same average power, a high peak power is desired so that the receiver gate is open for as short a period of time as possible. The higher power and lower repetition rate laser yields a better signal/noise than a lower power and higher repetition rate.

(4) The solid angle (field of view) of the receiver should be as small as possible. Therefore, the transmitter beam divergence should be low and the receiver optics should match the transmitted beam pattern.

(5) Raman backscatter is so weak that Raman atmospheric experiments are conducted at night to prevent the background noise from completely obscuring the signal. At night, with a low noise detector, both background and dark counts can be negligible. The signal to noise ratio then becomes

$$\frac{\sqrt{ZTN_r}}{\sqrt{N_r + 2N}} = \sqrt{ZTN_r}$$

This is the limiting case of the signal to noise ratio whenever the signal is much greater than the background and dark current. The "noise" is then due to the statistical nature of the entire process. That is, the noise is the "shot noise" of the signal. It should be noted that all the physical processes involved are statistical and that the laser radar equation for N_r uses mean values. For example, the number of photons transmitted per laser pulse varies from pulse-to-pulse with a distribution whose mean value is N_r .

(6) The signal to noise ratio varies directly with N_r when $N_r \ll N$ and varies as $\sqrt{N_r}$ when $N_r \geq N$.

23.2.4 Example of Laser Radar System Parameters

Assume a pulsed nitrogen laser transmitter is used. Peak power $P_t = 100$ kW, pulse width $\tau_t = 10^{-8}$ sec., the transmitted wavelength $\lambda_t = 3371 \text{ \AA}$ and the repetition rate is $Z = 100 \text{ sec}^{-1}$.

Then

$$N_t = \frac{P_t \tau_t}{h\nu_t} = \frac{P_t \tau_t \lambda_t}{hc} = 1.69 \times 10^{15}$$

photons transmitted per pulse.

The transmitter efficiency for on frequency return could be as high as 90% since only a single mirror is needed for collimation and steering. For Raman experiments η_t would be about 40% because the transmitter output would have to be filtered for this laser.

Suppose we have a range $R = 1$ km, and we examine the S/N ratio for Rayleigh and Raman backscatter. The attenuation factor

$$\exp \left[-\int_0^R \beta(s) ds \right]$$

will be about 0.9 for clear air for both the transmitted and returned energy. Assume a 30 cm diameter mirror for the receiver, $A_r = 0.07 \text{ m}^2$, and assume the desired altitude resolution $L = 50 \text{ m}$. The Raman experiment would be a night experiment with an interference filter to pass the Raman wavelength and a filter to block the strong on-frequency energy. A daylight Rayleigh experiment would require an interference filter for background suppression. The overall optical efficiency of the receiver, η_r would be only about 0.1 because UV filters do not have the quality of visible filters. The quantum efficiency η_q would be about 20% for either case.

Putting these factors together we have

$$N_r = [\eta_t \rho \sigma_D(\pi)] (9.6 \times 10^7)$$

photons received per pulse. For a Rayleigh experiment, $\rho \sigma_R(\pi) = 3/8 \pi \beta_r = (3/8 \pi)(8 \times 10^{-5}) \text{ m}^{-1} \text{ ster}^{-1}$, and $N_r = 825$ photons per pulse. For a Raman experiment using backscatter from nitrogen

$$\begin{aligned} \rho &\approx 2 \times 10^{25} \text{ m}^{-3} \\ \sigma_{Ra}(\pi) &\approx 1.86 \times 10^{-33} \text{ m}^2 \text{ ster}^{-1} \\ \rho \sigma_{Ra}(\pi) &\approx 3.72 \times 10^{-8} \text{ m}^{-1} \text{ ster}^{-1}. \end{aligned}$$

The return signal for the Raman experiment would be $N_r = 1.5$ photons per pulse. For the Raman experiment assume a 10-second observation is used, ($N_r Z T = 1500$) and the system has an uncooled large aperture photomultiplier with a dark count $N_D = 10^4$ counts per second. The 50-meter range gate is open for 0.33×10^{-6} seconds. Hence, the total dark count in 10 seconds would be 3.3 counts. The night background is also negligible compared to the signal count, so the signal to noise ratio for the Raman experiment would be about 38.7.

Suppose the Rayleigh experiment is conducted during daylight. Assume a sky brightness of 10^{-2} watts $\text{m}^{-2} \text{ ster}^{-1} \text{ \AA}^{-1}$, and a 10 \AA filter bandwidth. The receiver solid angle has to be relatively large, 5×10^{-5} steradians, because of the nitrogen laser beam divergence. The number of background photons counted per laser pulse would be about 4000. The signal to noise ratio with a 1-second integration time should be approximately 88.

Additional examples of laser radar system calculations are shown in (T23.1) and (T23.2). These calculations were made by G. T. McNice of WPL, NOAA.

23.3 Results of Meteorological Laser Radar Experiments

In this section we will discuss and summarize the results of numerous meteorological laser radar experiments conducted by various workers. Only papers showing actual atmospheric results will be mentioned. The papers cited are by no means a complete list but rather an attempt to indicate the types of experiments that have been conducted.

Table 23.1 Laser System Calculations

Type	Laser Parameters				Max Range (KM)					
	Energy (Joules)	Wave length (Å)	Beam width (mr)	Line width (Å)	H ₂ O RAMAN		N ₂ RAMAN		AEROSOLS	
					Day	Night	Day	Night	Day	Night
Ruby	15	6943	5	5	.1	0.9	5.5	6.9	>10	9.7
Doubled Nd	1	5300	5	50	.1	0.3	4.3	2.3	>10	6.7
Doubled Ruby	1	3472	5	5	.1	1.6	7.3	7.9	>10	7.3
N ₂	.001	3371	10	30	500	0.2	4.5	1.3	>10	4.5
Quadrupled Nd	.1	2650	5	25	.1	4.5	4.5	7.3	7.3	5.1

Telescope Diameter = 1 meter
 Integration Time = 100 sec
 Instrument Transmission = .5

$$\sigma (\text{H}_2\text{O RAMAN}) = 8.6 \times 10^{-31} \left(\frac{5145}{\lambda_0} \right)^4 \text{ cm}^2/\text{ster}$$

$$\sigma (\text{N}_2 \text{ RAMAN}) = 2.6 \times 10^{-31} \left(\frac{5145}{\lambda_0} \right)^4 \text{ cm}^2/\text{ster}$$

$$\sigma (\text{Aerosols}) = 1.5 \times 10^{-14} \text{ cm}^2/\text{ster}$$

Range Increment = $\frac{R}{10}$ for $R \leq 100$ meters
 $\frac{R}{10}$ for $R > 100$ meters

$$S/N = 3$$

Density Profiles:

$$n(\text{H}_2\text{O}) = 3.1 \times 10^{17} (17 - .3R) \text{ particles/cm}^3$$

$$n(\text{N}_2) = 1.9 \times 10^{19} (19 - .064R) \text{ particles/cm}^3$$

$$n(\text{Aerosol}) = 3.0 \times 10^{18} (18 - .833R) \text{ particles/cm}^3$$

Table 23.2 Laser System Calculations

Type	Laser Parameters				Max Range (KM)						
	Energy (Joules)	Wave length (Å)	Beam width (mr)	Line width (Å)	PPS	H ₂ O RAMAN		N ₂ RAMAN		AEROSOLS	
						Day	Night	Day	Night	Day	Night
Ruby	15	6943	1	5	2	4.5	>10	>10	>10	>10	>10
Doubled Nd	1	5300	1	50	2	2.3	>10	>10	>10	>10	>10
Doubled Ruby	1	3472	1	5	2	5.5	>10	>10	>10	>10	>10
N ₂	.01	3371	10	1	500	2.6	>10	>10	>10	7.8	>10
Quadrupled Nd	.1	2650	1	25	2	9.8	9.8	>10	>10	7.1	7.1

Telescope Diameter = 2 meters
 Integration Time = 300 sec
 Instrument Transmission = .5

$$\sigma (\text{H}_2\text{O RAMAN}) = 8.6 \times 10^{-31} \left(\frac{5145}{\lambda_0}\right)^4 \text{ cm}^2/\text{ster}$$

$$\sigma (\text{N}_2 \text{ RAMAN}) = 2.6 \times 10^{-31} \left(\frac{5145}{\lambda_0}\right)^4 \text{ cm}^2/\text{ster}$$

$$\sigma (\text{Aerosols}) = 1.5 \times 10^{-14} \text{ cm}^2/\text{ster}$$

Range Increment = $\frac{R}{4}$ for $R \leq 100$ meters
 $\frac{R}{3}$ for $R > 100$ meters

Density Profiles:

$$n (\text{H}_2\text{O}) = 3.1 \times 10^{17} (17 - .3R) \text{ particles/cm}^3$$

$$n (\text{N}_2) = 1.9 \times 10^{19} (19 - .064R) \text{ particles/cm}^3$$

$$n (\text{Aerosol}) = 3.0 \times 10^{18} (.18 - .833R) \text{ particles/cm}^3$$

23.3.1 Atmospheric Raman Experiments

As noted in 23.1, few successful atmospheric Raman experiments have been conducted with laser radars. The difficulties are apparent if one examines the scattering cross sections and the resultant backscattered signal for reasonable laser radar and receiver parameters. In spite of the difficulties, substantial progress has been made in demonstrating feasibility of Raman measurement of meteorological variables.

(1) Temperature Measurement Using Raman Backscatter (Strauch, et al., 1971). The most widely used method of obtaining a remote measurement of the vertical temperature profile with a ground-based instrument utilizes the absorption line of oxygen near 60 GHz to vary the transmission distance for received energy in a radiometer. The radiometer temperature measurements require an integral inversion to determine an atmospheric temperature profile. A radar-type determination offers improved resolution, simplicity of data analysis, and allows the measurement of temperature fluctuations at known altitudes of specific interest as well as a determination of a long term average profile. In addition to water vapor profile measurement and pollution detection, temperature information can be obtained directly from the Raman signal intensity.

The concept of using molecular density measurements to derive temperature profiles is not new. Elterman (1953), using searchlights, measured temperature profiles from 10 to 67.6 km by assuming that the scattering from above 10 km is due only to molecular scattering. Accurate temperature and pressure measurements were made at the lowest altitude that the scattering could be assumed to be Rayleigh. Below about 10 km the scattering is from molecules and aerosols and very accurate measurements of particulate scattering would have to be made to separate out the molecular scattering component. Methods of separating the scattering components have not yet been perfected. Aerosol density measurements rely on using an assumed Rayleigh density. Raman backscatter can be used to measure molecular density because the Raman component of backscatter is shifted in wavelength from aerosol or Rayleigh scattering. The relatively weak Raman backscatter can be used for density measurements in the lower atmosphere to complement density measurements made by Rayleigh scattering from the higher clean air.

A laser backscatter measurement provides data for a measurement of the product of the volume backscatter and the two-way extinction. In the case of the on-frequency backscatter, both the volume backscatter and the extinction are the sum of molecular and aerosol terms whereas the volume backscatter for the Raman component involves only the molecular density. The total extinction must be measured for the Raman determination of density. For example, suppose a measurement of the backscatter at 3400Å from 1 km altitude is being made. Suppose the total extinction from aerosol scattering is the same as that from Rayleigh scattering and assume that the total extinction can be measured to within 5%. Then the Raman backscatter can be used to determine a molecular density to within 1.4%, whereas the on-frequency backscatter would not give a useful density measurement.

Theory

The ideal gas law in the form

$$P(h) = \rho(h) T(h) R/M \quad (1)$$

where h is the altitude, P is the pressure, ρ is the density, R is the universal gas constant and M is the molecular weight leads to

$$\frac{\Delta P(h)}{P(h)} = \frac{\Delta \rho(h)}{\rho(h)} + \frac{\Delta T(h)}{T(h)}$$

For a fixed altitude, h , where the pressure can usually be taken as constant for the duration of a measurement,

$$\frac{\Delta \rho(h)}{\rho(h)} = \frac{-\Delta T(h)}{T(h)}$$

The signal received by a laser radar system measuring Raman backscatter from nitrogen would indicate, for constant pressure (altitude)

$$\frac{\Delta S_{N_2}(h)}{S_{N_2}(h)} = \frac{-\Delta T(h)}{T(h)} \quad \text{where}$$

$S_{N_2}(h)$ is the backscattered Raman signal and is proportional to $\rho_{N_2}(h)$. Thus, at any altitude where the partial pressure of nitrogen is approximately constant, the Raman backscatter gives a direct indication of temperature fluctuations.

The measurement of a temperature profile using nitrogen density measurements is not as direct as the measurement of temperature fluctuations. From the ideal gas law and the hydrostatic relationship, it follows that

$$T(h) = \frac{P_{N_2}(h)}{\rho_{N_2}(h)R/M} = \frac{P_{N_2}(h_0) - g \int_{h_0}^h \rho_{N_2}(z) dz}{\rho_{N_2}(h)R/M}$$

where $P_{N_2}(h_0)$ is the partial pressure of N_2 for a reference altitude and g is the acceleration of gravity. The integral in the numerator indicates that in order to measure temperature at any altitude, using density measurements, the density must be measured as a function of altitude from h_0 to h . $\rho_{N_2}(h)$ can be determined from Raman backscatter measurements using the laser radar equation.

The Rayleigh extinction coefficient (neglecting aerosol attenuation) is proportional to ρ_{N_2} , so the measurement of $\rho_{N_2}(h)$ as well as $P_{N_2}(h)$ requires the evaluation of $\int_{h_0}^h \rho_{N_2}(z) dz$. There are two ways to approach the evaluation of the integral; an *a priori* statistical model of the atmosphere can be used, or the integration can be performed using the laser radar measurements.

An experiment was devised to check the feasibility of using Raman backscatter to observe temperature fluctuations. A system with parameters similar to these in (T23.3) was used. The range was reduced to tower height (30.5 meters) so the temperature could be precisely monitored. The length of the backscatter volume was 5 meters and was determined by the geometry of the transmitting and receiving beams. Thermistors were placed near the center and near both ends of the 5-meter observation length. The laser beam passed vertically approximately 2 meters from the thermistors. A pressure indicator located at ground level continuously monitored total pressure. The atmospheric pressure change during any observation time was less than 0.2 mb so that $\frac{\Delta P}{P}$ is negligible compared with $\frac{\Delta T}{T}$ for temperature changes 1°C or greater. Since $\frac{\Delta T}{T} \approx 0.35\%$ per °C, the stability of the laser output power becomes important for monitoring fluctuations of atmospheric N_2 density. (For profile measurements where the backscatter from each range increment is measured for each laser pulse, the laser power stability is not as important.) During these experiments the S/N ratio was sufficient to allow observations of Raman signal fluctuations of 0.35% with a 3-second time constant, but the long term record contained random fluctuations of approximately 1%. These changes were traced to laser power variations, temperature effects on the transmission of optical filters and mechanical instability.

The Raman backscatter signal was recorded using a fast sample and hold gate to measure a voltage level proportional to the number of backscattered photons received from the 5-meter interaction region after each pulse. The step-like signal was filtered and recorded on a zero-suppressed strip chart.

Figures (23.3) to (23.6) show the examples of data in which temperature changes of several degrees were recorded by the thermistors. Only the center thermistor temperature is shown; the other two thermistors indicated the same temperatures. In (F23.3 to F23.5), data points were ready every 20 seconds and a plot of relative Raman backscatter (S) vs. T is shown with the data. The theoretical curve (S)(T)= constant is shown for reference. This curve is approximately a straight line over the range of values measured in these experiments. Although fluctuations other than temperature are recorded on the Raman channel, temperature changes are clearly recorded by the Raman measurements. In (F23.5), compensation for the change in laser power was made in the S-T graph. Figure(23.6) shows a 23-minute segment of data during which T changed by approximately 2 degrees. Data points were read every 4 seconds and a cross correlation of the S-T data was performed. No corrections were made to the raw data. The peak correlation at zero time lag, was 0.76 and the correlation decreases to .2 with about 1-minute time lag.

The results of this preliminary investigation indicate that Raman backscatter from N_2 can be used to measure atmospheric temperature profiles. These experimental results also indicate that it should be possible to measure low altitude temperature inversions. Temperature profiles to 10 km should be possible with present laser systems. Note however, that the total attenuation must be measured.

Table 23.3 Laser Radar System Parameters

N_t ,	photon transmitted per pulse	$= 1.7 \times 10^{15}$
η_t ,	transmitter efficiency	$= 80\%$
σ_{N_2} ,	Raman cross section for N_2	$= 1.86 \times 10^{-33} \text{ m}^2/\text{ster}$
L ,	range resolution	$= 50 \text{ m}$
A_r ,	receiver aperture	$= .36 \text{ m}^2$
η_r ,	receiver optical efficiency	$= 25\%$
η_q ,	detector quantum efficiency	$= 25\%$
Z ,	pulse repetition rate	$= 100 \text{ sec}^{-1}$
λ_t ,	transmitter wavelength	$= 3371 \text{ \AA}$

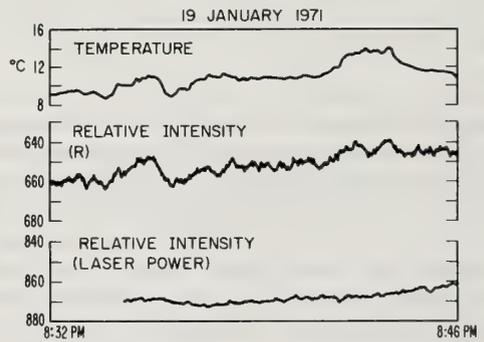
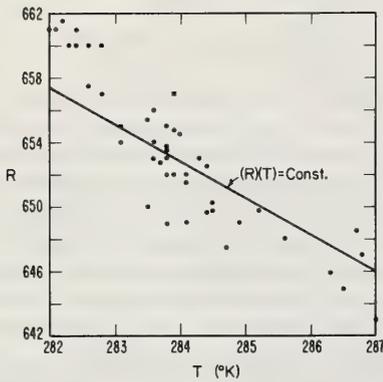


Figure 23.3 Raman backscatter intensity versus temperature, 19 January, 1971.

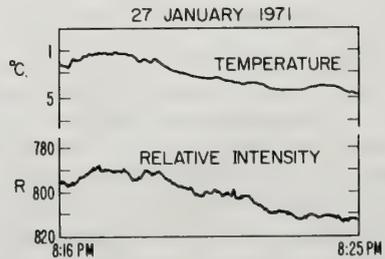
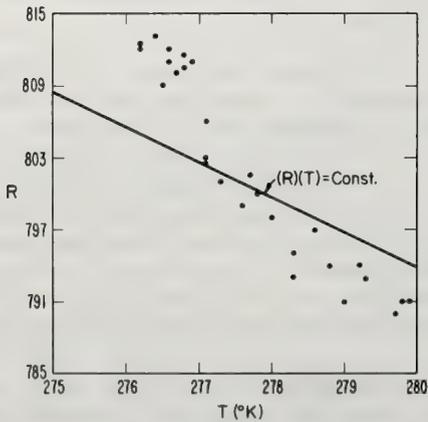


Figure 23.4 Raman backscatter intensity versus temperature, 27 January, 1971.

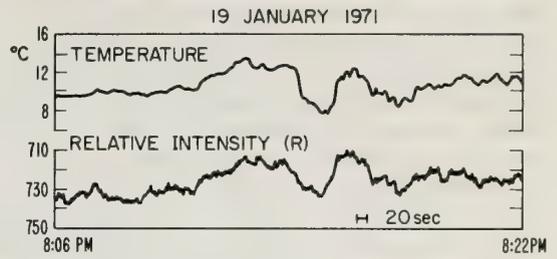
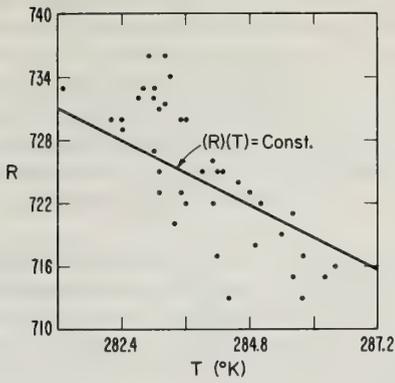


Figure 23.5 Raman backscatter intensity versus temperature, 19 January, 1971.

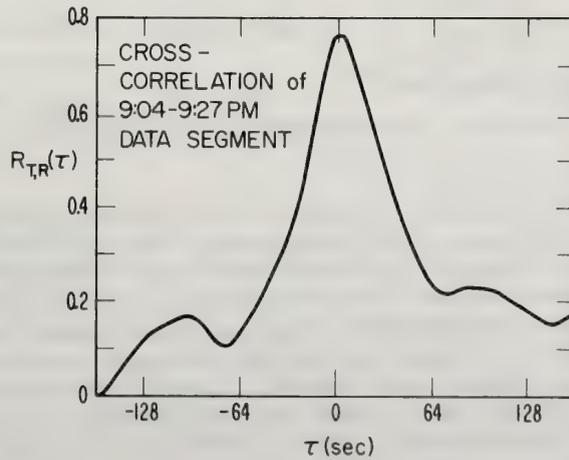
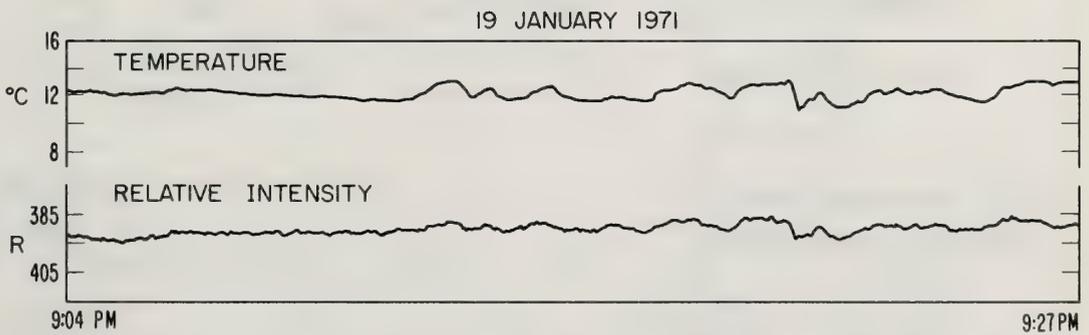


Figure 23.6 Cross-correlation of Raman backscatter intensity and temperature.

(2) Atmospheric Water Vapor Measurements by Raman Lidar (to be published in the Journal of Remote Sensing of the Environment, Strauch, Derr, and Cupp). An experiment was performed to compare Raman backscatter measurements with measurements made with a meteorological instrument, the microwave refractometer. The refractometer was located at the 30.5 m level of a meteorological tower and the laser beam passed approximately 1 meter from the microwave cavity. The backscattered laser signal was observed from a region approximately 20 cm in diameter and 5 meters in length centered at the height of the microwave cavity. The length of the backscattering region was determined by the geometry of the crossing of the transmitter and receiver beams. Laser parameters are listed in (T23.3). The power and repetition rate provide signal returns that allow electronic processing of analog signal levels rather than oscilloscope photographing and photon counting used with low pulse repetition rates.

The electronic processor is shown in (F23.7). The video level from the photomultiplier preamplifier is sampled after each pulse and a voltage proportional to the peak signal level during the sampling period is held until the next pulse. The step-like signal level is smoothed with a 3-second time constant and recorded on one channel of a Sanborn strip chart recorder. This type of processing is useful for signals consisting of 1 or more photons received per pulse as in this experiment. A low-frequency counter on the sampled video level can be used to photon count if the signal level is less than 1 photon per pulse.

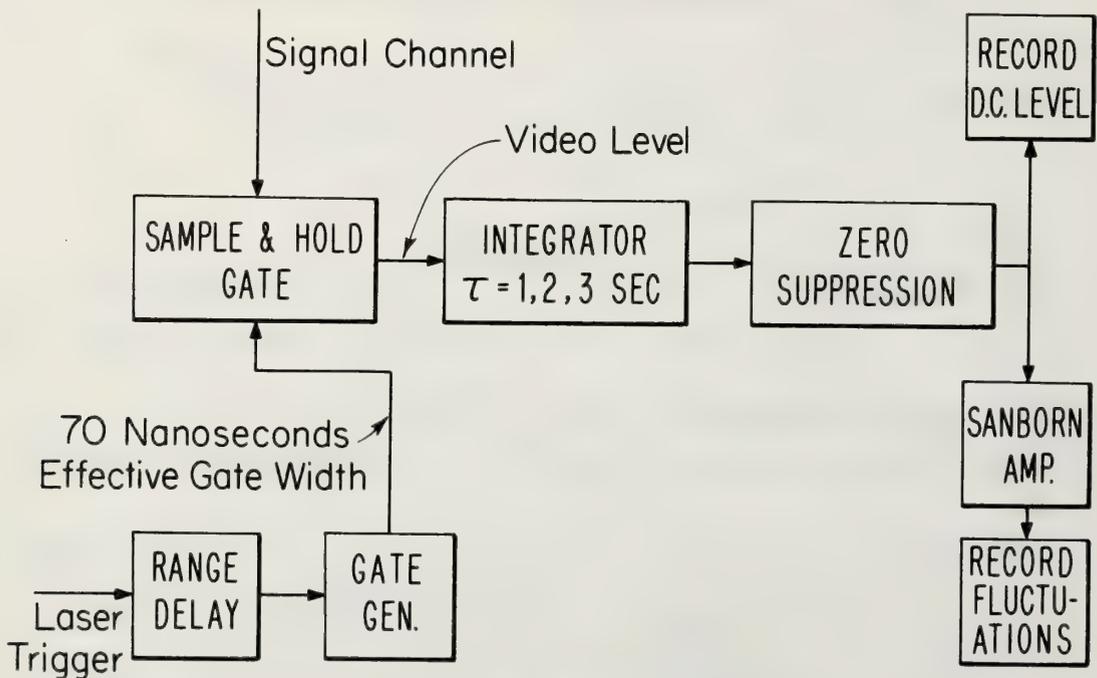


Figure 23.7 Raman lidar electronic data processing.

The experiments were conducted in November of 1970 at Gunbarrel Hill, about 10 miles east of Boulder, Colorado. The water vapor content averaged about 4 mb (total pressure about 830 mb). The temperature range was 0 to 15°C. Dry bulb and wet bulb temperatures were measured about every 30 minutes. The pressure was read at ground level and a 30.5 meter altitude correction was made. Pressure was read about every 30 minutes or whenever a measurable (0.1 mb) pressure change occurred. Thermistor temperature and microwave refractive index were continuously recorded with calibration checks about every 30 minutes. The experiments were performed at night, and the background and dark current were low enough to make the Poisson noise of the signal the dominant system noise. The signal level was continuously recorded with frequent zero level checks made by blocking the transmitter beam. Experiments were conducted during nights when relatively large fluctuations of refractive index and Raman signals indicated that volumes of air with varying water vapor content were passing the laser beam.

Data points for analysis were chosen by selecting points every 400 seconds and by arbitrarily selecting points where large fluctuations occurred. About 1/2 the data points were chosen by each criteria. A typical record of raw data is shown in (F23.8). After selecting the data points, the water vapor E was computed and the Raman signal level was read. The data points were plotted and a least squares fitting of the data to the equation $S = AE + S_0$ was made, where A is the slope and S_0 is the Raman signal intercept at $E = 0$. The section of record shown in (F23.8) was analyzed by reading data points every 4 seconds (237 Points) and performing a cross correlation of the E and S time records. The correlation is shown in (F23.9). The peak correlation is 0.8 and falls to 0.4 at about a 1-minute delay. The decay time is indicative of the scale size of the atmospheric fluctuations and of the wind speed. Figure (23.10) shows the data from one night, and the mean and rms deviation of the experimental S values from the calculated straight line are listed. The data from each night were normalized to make the equation of the fitted straight line have a slope of 20 and zero intercept. Composite data from 10 different nights are shown in (F23.10-a).

Although in the present experiment only a single atmospheric region was probed, this equipment can be expanded to perform average H_2O/N_2 profile measurements by the addition of sequentially delayed sampling gates. Transmitter-receiver alignment, beam overlap, aerosol attenuation and $1/R^2$ attenuation could be programmed into the gates to give a straight line profile for N_2 Raman return. The H_2O profile would then be observed directly by substituting the appropriate interference filter. Using the measured values obtained in this experiment for signal returns, 50-meter range resolution, a 5-minute observation time, a constant H_2O/N_2 mixing ratio, 50% relative humidity and $20^\circ C$, the maximum altitude for observation of the H_2O/N_2 ratio with a S/N ratio of 10 would be 4.25 km.

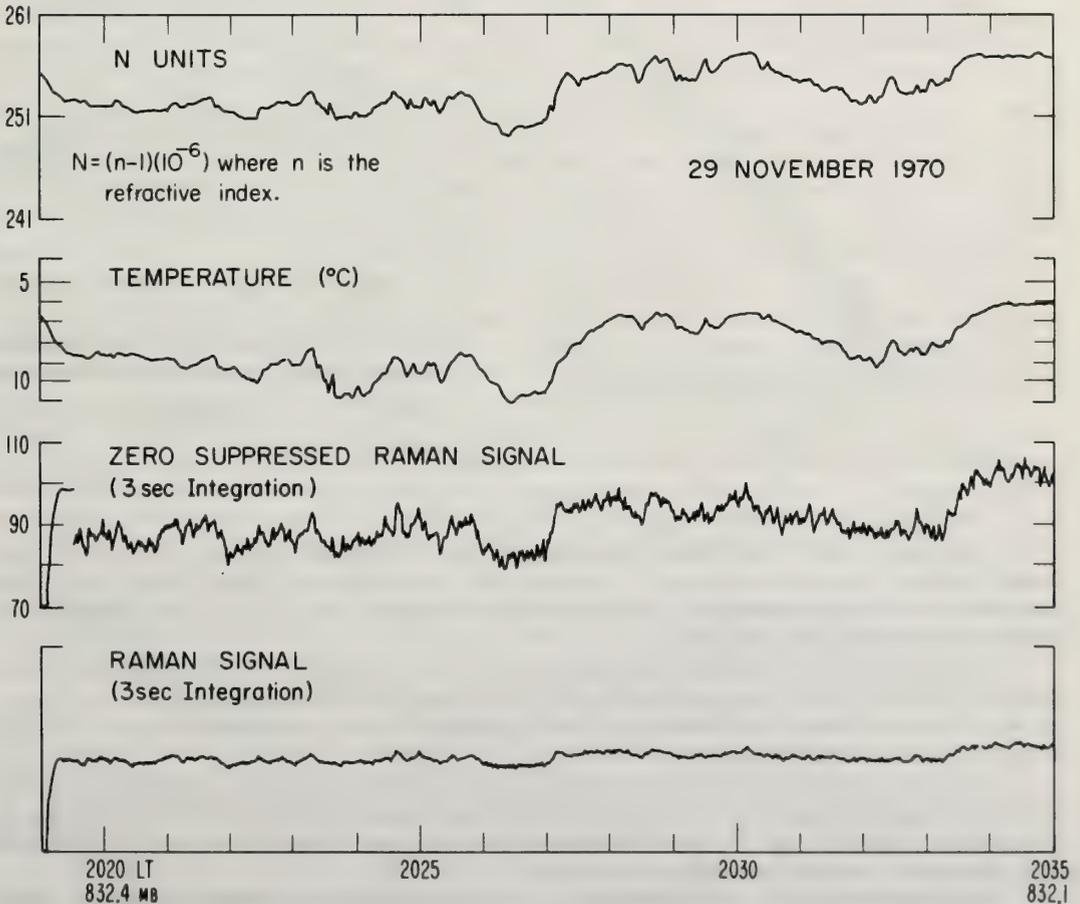


Figure 23.8 Raw data sample. * $N = (n-1)(10^{-6})$ where n is the refractive index.

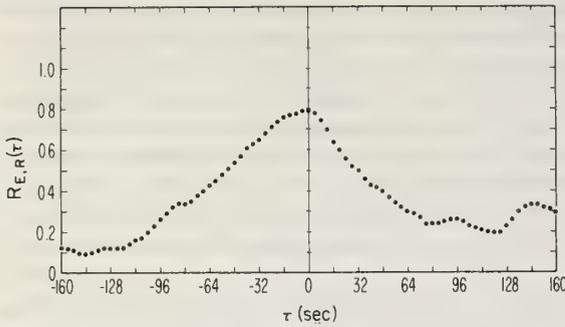


Figure 23.9 Cross-correlation of Nov. 29 data segment.

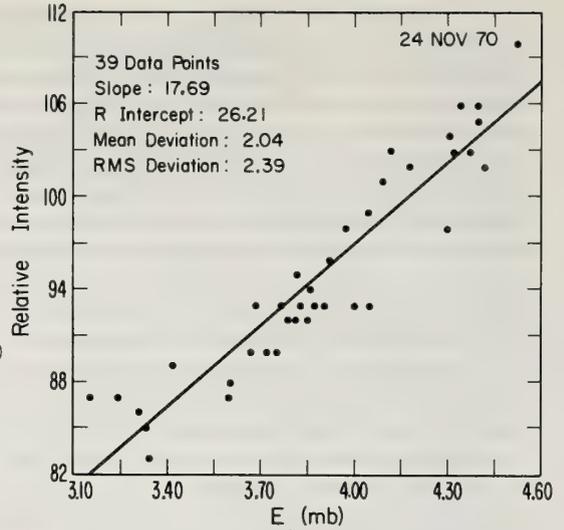


Figure 23.10 Experimental data points.

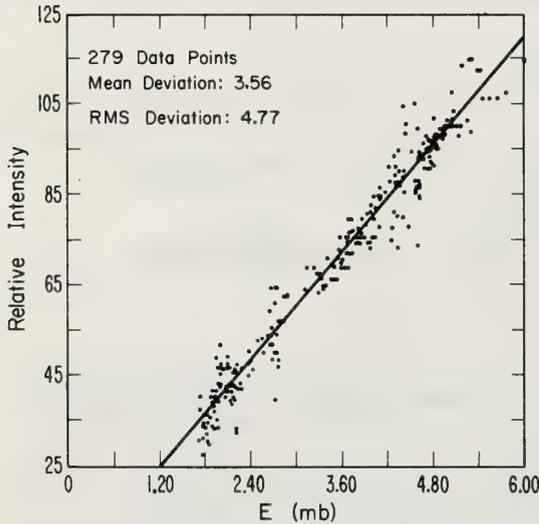


Figure 23.10a Composite data - 10 nights.

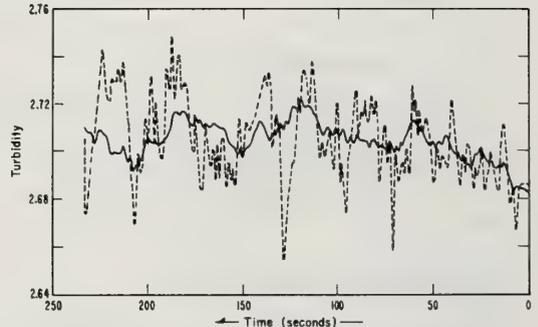


Figure 23.11 A typical turbidity variation as a function of time (dashed line) superimposed on a turbidity curve calculated for a constant air density (solid line).

(3) Measurements on the Raman Component of Laser Atmospheric Backscatter (Cooney 1968). Density profiles of atmospheric nitrogen to 3 km were measured using Raman shifted return from a pulsed ruby laser. The ruby laser wavelength is 6943Å and the N₂ backscatter occurs at 8285Å. Measurements were taken at night under clear sky conditions.

(4) Observation of Raman Scattering from the Atmosphere Using a Pulsed Nitrogen Ultraviolet Laser (Leonard 1967). This is the first published atmospheric Raman laser radar observation. Scattering from O₂ and N₂ is reported using a 100 kw laser at 3371Å. A range of 1.2 km was obtained for the backscattered returns, using oscilloscope photographs.

(5) Observation of Raman Scattering by Water Vapor in the Atmosphere (Melfi, et al., 1969). A frequency doubled ruby laser was used to observe the Raman backscatter from H₂O. The N₂ return was also observed and a H₂O/N₂ mixing ratio was calculated. Data was taken on the oscilloscope photograph and comparisons with radiosonde measurements were made.

(6) Spectral Analysis of Atmospheric Nitrogen by Laser Raman Radar (Kobayasi and Inaba, 1970). The ratio of Raman shifted power return from a laser radar observing atmospheric nitrogen to the same system

observing clear atmosphere returns was $4.3 \pm 0.8 \times 10^{-4}$, leading to a Raman volume backscattering coefficient of $8.6 \pm 1.6 \times 10^{-12} \text{ cm}^{-1} \text{ ster}^{-1}$. Stray light rejection of 10^7 to 10^9 is needed in fog or rain to eliminate spurious contributions to Raman results.

(7) Spectroscopic Detection of SO_2 and CO_2 Molecules in Polluted Atmosphere by Laser-Raman Radar Technique, (Kobayasi and Inaba, 1970). A ruby laser was used to observe SO_2 and CO_2 in a smokestack at a distance of 20 meters. A grating monochromator and filters were used to sort out the Raman backscatter from various molecules. An oil smoke plume burning fuel oil was used for the test.

(8) Remote Measurement of Atmospheric Water Vapor Profiles Using the Raman Component of Laser Backscatter (Cooney 1970). H_2O profiles to 2.5 km are reported. Four wavelength regions were monitored by the receiver; the N_2 Raman wavelength, the H_2O Raman wavelength, the transmitted wavelength and a wavelength where no signal is expected. A frequency doubled ruby laser was employed.

23.4 Measurement of Tropospheric Aerosols by Laser Radar

The optical scattering properties of atmospheric aerosols are described in Chapter 10 and in 23.4.2. The Mie theory of aerosol scattering predicts, and experiment confirms, that the optical scattering cross section of a single water droplet, $1 \mu\text{m}$ in radius, is approximately 10^{-8} cm^2 . In the lower atmosphere, there are usually a number of haze particles in this size range found in each cm^3 of the atmosphere.

A nitrogen molecule and an oxygen molecule have a scattering cross section of nearly 10^{-28} cm^2 . Since at standard temperature and pressure there are approximately 10^{19} molecules in 1 cm^3 of air, it can be seen that an important part of observed atmospheric scatter may be due to aerosols, and that they are detectable by laser radar whenever it is capable of detecting molecular scattering.

Thus, any attempted quantitative measurement of the absolute air density by on-frequency laser radar backscattering is strongly affected by the presence of aerosols in the lower troposphere. At visible wavelengths, in the lower troposphere, the contribution to atmospheric scattering by aerosols is usually greater than that due to molecular scattering; see, for example, McClatchey, et al. (1971).

Hence, air density measurement requires knowledge of the ratio of molecular and aerosol scattering to molecular scattering. Moreover, when laser radar measurements are being used for the exploration of the aerosol content in the air, this ratio must be known to deduce the molecular scattering from a measurement of the total scattering. This ratio, known as the *turbidity*†, is, therefore, one of the very first properties of the atmosphere that must be determined to evaluate quantitatively any laser echo. Theoretical and experimental methods of measuring the turbidity will be described in 23.4.1.1 and 23.4.1.2.

Air density and total aerosol content may be obtained to a good approximation solely from atmospheric turbidity and total scattering. More information is required when detailed properties of the aerosol content such as size distributions, indices of refraction, and shapes are of interest. In 23.4.2 the dependence of the scattering intensities on these properties will be discussed and the necessity of a number of independent measurements (such as varying the incident wavelength or changing the scattering angle) will be shown. Since the number of the possible unknown parameters is, in principle, very large (i.e., equivalent particle radii r_1, \dots, r_N ; number densities ρ_1, \dots, ρ_N ; indices of refraction m_1, \dots, m_N ; shapes and orientations of nonspherical particles) several simplifying *a priori* assumptions are often introduced which affect the accuracy of determining specific aerosol properties from measurements. These assumptions will be described in section 23.4.2.1.

Section 23.4.2.2 will be devoted to examples of theoretical inversion procedure for a few sets of independent measurements using a series of assumptions discussed in 23.4.2.1, as well as one example involving no *a priori* assumption.

Scattering methods of measuring aerosol properties which are not based on remote laser-radar measurements in the troposphere are not discussed here. Readers interested in other methods should refer to:

A. Laboratory or airborne laser aerosol scattering measurements— see, for example, Quenzel (1965); Eiden (1966); Blau, et al. (1970); Ferrara, et al. (1970).

B. Aerosol size distribution derived from solar extinction measurements—see, for example, Yamamoto, et al. (1969).

† It should be noted that there are other definitions of turbidity; i.e., the usual definition as used by AFCRL in their tables of standard atmospheres is the ratio of aerosol attenuation to molecular attenuation. In chemistry measurements, turbidity is usually defined as the extinction coefficient.

C. Solar scattering and polarization measurements for aerosol size and refractive index determination—see, for example, Hansen, et al. (1971).

D. Qualitative aerosol lidar scattering measurements as tracers of atmospheric phenomena were the main achievements of the tropospheric laser radar measurements in the last decade. Fiocco, et al. (1964) were the first to report the observation of aerosol layers by laser radar. Several groups have contributed since then to the qualitative laser radar tropospheric aerosol measurement (see, for example, a review article by Collis 1970). Recent measurements are reported, for example, by Collis (1971) (turbulence and diffusion studies); Schotland, et al. (1971) (depolarization measurements for distinguishing between water and ice clouds); and Derr (1972), who describes a technique for horizontal wind velocity measurements using laser backscatter from spatially separated volumes. If an atmospheric region of varying aerosol density passes one of the volumes illuminated by the laser, the backscatter from this volume will have a component that fluctuates with the aerosol density. If the mean wind carries this atmospheric cell past the second laser-illuminated volume, the backscattered signal can be recognized and correlated with the signal from the first volume (in reality cross correlation coefficients considerably less than 1.0 are expected). The horizontal distance and the delay time determine the mean wind. The atmospheric conditions that allow the method to be used successfully are being explored. The major difficulty has been to find sufficiently large fluctuations that can be measured without elaborate data processing techniques.

Upper atmospheric (and stratospheric) aerosol measurements are not discussed here, although the methods described are also being used for the study of the aerosol layers detected in the stratosphere. Various laser radar studies of the atmosphere above 50 km were summarized by Grams (1970). The 18-23 km aerosol layer had also been measured by most laser radar groups mentioned in Grams' work. The technique is described, for example, in the work by Bartusek, et al. (1970).

23.4.1 Turbidity and Its Measurement

The Rayleigh atmosphere, that is a pure molecular atmosphere, is very seldom found in the lower troposphere. Although the number densities of particulate matter (aerosols) have an average value of ~ 1000 per cubic centimeter and rarely exceed a value of 10^{15} cm^{-3} (in the urban environment), in comparison with 10^{19} air molecules, tropospheric scattering will be composed of a molecular contribution and a significant aerosol contribution. (A useful method to verify the presence of aerosols by polarization measurements is described by Cohen, et al., 1969).

The ratio of total (molecular and aerosol) scattering to molecular scattering (I_T/I_M) is the atmospheric turbidity (τ_λ) and is a function of the scattering wavelength (and angle).

Conventional laser radar techniques provide a measurement of the total scattering I_T . Fluctuations of the total scattering radiance as a function of time were studied by many laser radar groups (i.e., McCormick 1971), and turbidity fluctuations were derived from these measurements using the assumption that the atmospheric molecular density does not vary during the measurements and that the absolute air density value can be based on the U.S. Standard Atmosphere Tables. Obviously, these approximations cannot be introduced when average air density is to be measured. Moreover, temperature measurements based on the nitrogen Raman scattering made by Strauch, et al. (1971) indicate that the air density may fluctuate by $\pm 1\%$ of the average air density, in a time scale comparable to the time between two consecutive laser measurements (of the order of a few seconds). Cohen, et al. (1972) measured simultaneously the total scattering $I_{T(t)}$ and the molecular scattering (nitrogen Raman) $I_M^{(t)}$. A typical turbidity variation as a function of time is shown in (F23.11), superimposed on a turbidity curve using the constant density assumption. It follows that except for clouds or very high particle density, where $\tau_\lambda \gg 1$ (and therefore, 1% variation in the air density can be neglected in the separation of the signal from the molecular scattering), turbidity values have to be measured or calculated whenever accurate air density values are desired or accurate aerosol parameters are sought.

23.4.1.1 Theoretical Method for Turbidity Measurements

The laser radar backscattering intensity $I_{k,j}$ from an altitude Z_k , obeys the following equation

$$Z_k^2 \cdot I_{k,j} = (\sigma_{R_j} \cdot \rho_k + b_{k,j}) \cdot \exp\left(-2 \int_0^{Z_k} (\beta_{j,R} + \beta_{j,M}) dr\right) \cdot A_j = (\sigma_{R_j} \rho_k + b_{k,j}) A_{k,j}^* \quad (23:15)$$

where j stands for the wavelength or polarization (or, possibly, scattering angle), ρ_k = molecular number density at Z_k , σ_{R_j} = molecular backscattering cross section, $b_{k,j}$ = aerosol scattering coefficient, $\beta_{j,R} + \beta_{j,M}(=\beta_j)$ = molecular and aerosol extinction coefficients, and A_j is a system calibration coefficient. If the laser radar is operating at some angle Φ from the zenith and horizontal homogeneity is assumed (Schuster, et al., 1972), Z_k^2 is replaced by $Z_k^2 / \cos^2 \Phi$ and β_j is replaced by $\beta_j / \cos \Phi$.

The coefficients σ_{R_j} and β_j are known as a function of j . For example, if j stands for wavelength, then $\sigma_{j=1} / \sigma_{j=2} = (\lambda_{j=2} / \lambda_{j=1})^4$. Equation (23:15) for the altitude Z_{k+1} becomes (Cohen, et al., 1972):

$$Z_{k+1}^2 \cdot I_{k+1,j} = (\sigma_{R_j} \cdot \rho_{k+1} + b_{k+1,j}) \cdot \exp\left(-2 \int_0^{Z_{k+1}} \beta_j dr\right) \cdot A_j \simeq (\sigma_{R_j} \cdot \rho_{k+1} + C_{k+1} \cdot b_{k,j}) \cdot A_{k,j}^* \quad (23:16)$$

where C_{k+1} is constant for any j (wavelength, polarization or scattering angle).

Equation (23:16) suggests that the following assumptions should be accepted:

1. If $f_k(r,m)$ is the size distribution function describing the aerosols in altitude Z_k , then

$$f_{k+1}(r,m) = C_{k+1} \cdot f_k(r,m).$$

This assumes that the aerosol size distribution as well as the aerosol type and indices of refraction are the same in the two layers. The only difference in the aerosols in the two layers is the number density. This assumption is used for $\Delta Z_{k \rightarrow k+1} \simeq 300$ m. As a consequence, for most practical uses, the scattering intensities from the aerosols in the two layers will have the same ratio C_{k+1} independent of j . In case the turbidity values are much higher than unity, this result will not hold since multiple scattering effects have to be taken into account (see 10.5.2).

2. $\exp(-\int_{Z_k}^{Z_{k+1}} \beta_j dr) \simeq 1$. Since $\Delta Z_{k \rightarrow k+1} \simeq 300$ m this assumption holds for most practical uses (see,

for example, Derr et al., 1970 – Fig. 3).

Four measured values of $I_{k,j}$ ($k = 1,2; j = 1,2$) provide the value of C_{k+1} without the necessity of calibrating the laser radar ($I_{k,j}$ are in relative units), nor measuring the extinction term $\exp\left(-2 \int_0^{Z_k} \beta_j dr\right)$

when j stands for polarization:

$$C_{k+1} = \frac{\sigma_{R_1} \cdot I_{k+1,2} - \sigma_{R_2} \cdot I_{k+1,1}}{\sigma_{R_1} \cdot I_{k,2} - \sigma_{R_2} \cdot I_{k,1}} \quad (23:18)$$

The air density profile calculation requires calibrated values of $D_{k,j}$ and two initial values of the density at altitudes Z_0 and Z_1 :

$$\left. \begin{aligned} \rho_{k+1} &= C_{k+1} \cdot \rho_k + d_{k+1} \\ d_{k+1} &= \frac{I_{k,2} \cdot I_{k+1,1} - I_{k,1} \cdot I_{k+1,2}}{\sigma_{R_1} \cdot I_{k,2} - \sigma_{R_2} \cdot I_{k,1}} \end{aligned} \right\} \quad (23:19)$$

where

Table (23.4) presents an example of the calculation of the aerosol and air density profiles based on an atmospheric model suggested by McCormick (1971).

If no information on the air density at altitudes Z_0 and Z_1 is available an additional assumption is needed:

$$\rho_{k+1} / \rho_k = \rho_k / \rho_{k-1} = R_k.$$

Table 23.4. The separation of the Rayleigh (molecular) and the Mie (aerosol) scattering from total scattering measurements (Cohen and Derr, 1972). The $D_{k,j}$ values were calculated by McCormick (1971).

Altitude† (km)	Total "Measured" Scattering Intensity – Relative Units		Aerosol Density Ratio		Air Density – Relative Units	
	$D_{k,1}$	$D_{k,2}$	C_{k+1}		ρ_k	
Z_k	$(\tau_{j=1} = 6943\text{Å})$	$(\tau_{j=2} = 3472\text{Å})$	from McCormick's model	(23:18)	U.S. Standard Atmosphere	(23:19)
0	5.89	13.0	0.255	0.256	2.54	
1	1.86	10.5	0.256	0.256	2.31	
2	0.792	7.74	0.266	0.267	2.09	2.09
3	0.493	6.51	0.279	0.279	1.89	1.90
4	0.385	5.79	0.367	0.368	1.70	1.70
5	0.333	5.19	0.736	0.736	1.53	1.53
6	0.297	4.65	1.104	1.104	1.37	1.37
7	0.267	4.16			1.23	1.23

† Since $f(r,m,Z)/c_{k+1}$ was assumed to be constant in this atmospheric model, ΔZ is allowed to be 1 km.

Table 23.5 The turbidity calculation by use of an exponential air density profile assumption. $D_{k,j}$ values as a function of height for two polarizations were derived by the use of a constant air density ratio value (=0.986) and a random contribution of $b_{k,j}$ (generated by the computer). The ratio $b_{k,1}/b_{k,2}$ was kept constant.

Altitude (Meters)	"Measured" backscattering in two polarizations		Aerosol Density Profile (28:18)	Mean Air Density††	Calculated Mean Density Ratio	Mean Turbidity	Values††
	Parallel†	Normal					
Z_k	$D_{k,1}$	$D_{k,2}$	C_{k+1}	ρ_k	R_k	$\tau_{j=1}$	$\tau_{j=2}$
125	5.413	0.244	1.57	1.277	0.989	2.01	6.04
250	6.917	0.359	1.10	1.263	.986	2.60	(±0.02)†† 8.38
375	7.300	0.390	0.72	1.246	.991	2.78	(±0.03) 9.89
500	5.970	0.292	0.06	1.234	.985	2.29	(±0.01) 7.47
625	2.759	0.053	1.25	1.216	.983	1.0753	(±0.0002) 1.38
750	2.765	0.056	7.04	1.195	.978	1.096	(±0.002) 1.48
875	4.183	0.165	2.56	1.169	0.979	1.696	(±0.005) 4.46
1000	6.820	0.364		1.144		2.82	(±0.04) 10.05

† Parallel - the component parallel to the linearly polarized incident laser light.

†† Three air density and turbidity values can be calculated for each layer. (For example, for Z_5 three sets of $T_6 D_{k,j}$ values can be used for the air density and turbidity calculations T: 1. $k=3,4,5$; 2. $k=4,5,6$; and $k=5,6,7$). Values of ρ_k and τ are the arithmetic mean of the three possible values. The brackets in $\tau_{j=1}$ indicate the range within which the three possible values were calculated for each layer.

This is approximately true for most polytropic atmospheres, [where $P_k/P_0 = (\rho_k/\rho_0)^Q$], and suggests an average exponential air density profile. In that case, six measured $I_{k,j}$ values ($k = 1, 2, 3; j = 1, 2$) lead to the molecular density profile as well as the turbidity values. The molecular density ratio R_k can be calculated from uncalibrated $I_{k,j}$ values—see Table 23.5. It should be noted that the air density and turbidity values calculated in the Table are based on 4 significant digits in $I_{k,j}$. If the air density only is of interest, Schuster, et al. (1972) suggest that an assumption could be made on the values of $b_{k,1}$ and $b_{k,2}$, namely $b_{k,1} = b_{k,2}$ where 1 and 2 stand for two wavelengths. In this method the laser radar elevation angle Φ (from the zenith), is changed and backscattering is measured for two Φ values and 2 wavelengths. The resulting four $I_{k,j}^*$ values together with the above mentioned assumptions (1. $b_{k,1} = b_{k,2}$; 2. horizontally homogeneous atmosphere) may then be used for determining pressure as a function of altitude.

For laboratory or airborne measurements the study of the laser line broadening due to molecular and aerosol motions may result in the separation of the molecular from the aerosol scattering. The line broadening due to molecular Brownian motion is approximately 2 orders of magnitude larger than that expected due to aerosol motion. The technique for the application of this method is being developed, for example, by Grams (1972, private communication).

23.4.1.2 Experimental Methods for Turbidity Measurements

Most turbidity measurements involve an assumption that the aerosol concentration is very low at some altitudes, Z_{k_0} (e.g., ~ 30 km) and in consequence, the aerosol scattering can be neglected. The measured scattering value I_{k_0} is made to correspond to the U.S. Standard Atmosphere air density value for that height.† Thus, I_k values for a Rayleigh atmosphere can be estimated for various heights Z_k and the corresponding U.S. Standard Atmosphere air density. The turbidity value is then found by calculating the ratio between the actual measured I'_k and the estimated-expected molecular backscattering I'_k .

Schotland and Reiss (1972) and Cohen, Derr, Cupp and Barth (1972) suggest combining Raman scattering measurements with on-frequency scattering for the remote measurement of atmospheric turbidity. Schotland and Reiss point out that the development of highly monochromatic light-sources has made possible examination of the spectral composition of the scattered light within a few tens of Angströms of the transmitted wavelength, and as a result they suggest the observation of the pure rotational Raman scattering due to N_2 and O_2 molecules. The advantage of the rotational Raman scattering measurements is that extinction is practically the same for both the aerosol and Raman scattering and the measurements can be interpreted for the exact turbidity without an absolute scattering measurement.

Actual measurements of turbidity were performed by Cohen et al. (1972) using the system shown in (F23.12). The *top* photomultiplier detected the vibrational N_2 Raman scattering while the *side* photomultiplier detected the on-frequency scattering. I_k 's have to be corrected by the corresponding extinctions for each wavelength whenever the extinction is not negligible. Methods for the remote determination of atmospheric transmissivity had been suggested by several authors; e.g., Barret et al. (1967) who assumed that the extinction to scattering ratio is constant at all heights and is known; and Melfi (1972), who assumed a U.S. standard atmosphere for the evaluation of extinction. This last method is therefore not adequate for exact turbidity measurements although such an assumption for the extinction estimation has a much weaker effect on the accuracy of the results than that used for the scattering approximations.

On the other hand, vibrational Raman scattering techniques are already operational and for several practical aerosol size distributions, the extinction can be assumed to be the same for both the laser and the N_2 Raman wavelengths.

23.4.1.3 Horizontal Extinction and Visibility Measurements

While the turbidity (as defined here) is a measure of the aerosol content in the air from the scattering point of view, the "visibility distance" provides additional information as to the *extinction* (absorption and

† This approximation is being widely used for the *calibration* of the lidar system into absolute units (the coefficient A_j or A_j^* in (23.15)). Since a "Rayleigh-Atmosphere" assumption cannot be applied *a priori* whenever low altitude tropospheric measurements are concerned, other calibration methods were developed (e.g., Hall et al., 1970). It should be noted that N_2 Raman scattering measurements provide a direct laser radar calibration, since an additional independent measurement of air density (for example, pressure and temperature measurements) is all that is needed for the calibration.

total scattering) properties of the aerosols. The visibility distance V and the average extinction coefficient $\bar{\beta}$ obey the following relation

$$V = \frac{\text{const.}}{\bar{\beta}} \quad (23:20)$$

where the averaging is over the visible range and over the path V . The laser wavelength visibility distance, V_λ , is inversely proportional to the extinction coefficient β_λ , and, therefore, the measurement of β_λ can be directly interpreted as the definition of the visibility distance (for a given wavelength). Several methods for the measurement of β_λ were suggested and the reader should refer to a review article by Grimes (1969). One method will be briefly mentioned here. The backscattering from a distance R_i can be written as (see 23:12)

$$I(R_i) = A \cdot \frac{\sigma(\pi)}{R_i^2} \cdot \exp\left(-2 \int_0^{R_i} \beta \, ds\right) \quad (23:21)$$

If the backscattering from two very close distances, R_i and R_{i+1} (where $\Delta R_{i \rightarrow i+1} < 100$ m,) is recorded, the ratio $I(R_i)/I(R_{i+1})$ is given by

$$I(R_i)/I(R_{i+1}) \cong (R_{i+1}/R_i)^2 \cdot \exp\left(2 \int_{R_i}^{R_{i+1}} \beta \, ds\right) \cong (R_{i+1}/R_i)^2 \exp(2\Delta R \beta_i) \quad (23:22)$$

where only β_i is unknown. Equation(23:22) is based on the assumption that for $\Delta R < 100$ m, $\sigma_i(\pi) \cong \sigma_{i+1}(\pi)$.

23.4.2 Aerosol Properties and Their Effect on Scattering Intensities†

Turbidity and extinction determination provide a means to separate aerosol scattering from total scattering. The resulted aerosol scattering cross sections are dependent on the aerosol properties, thus providing an indirect way, which requires an inversion procedure to calculate those properties. The aerosol properties can be divided into two categories:

1. Properties that appear as parameters in the exact theoretical scattering functions: sizes of spherical aerosols, (low) number densities, relative indices of refraction.
2. Properties that involve an approximation in the theoretical scattering functions: non-spherical shapes; orientations relative to the scattering plane; high-number densities (multiple scattering).

Most of the theoretical work dealing with the inversion of laser radar aerosol scattering introduces the assumption that the particles are either spherical or can be considered spherical (Holland, et al., 1967; Eiden 1966) when calculating scattering properties. This assumption has been up to the present time, a *requirement* for any quantitative approach in determining the size-distribution function of aerosols derived from laser radar scattering information.

A few theoretical papers have considered the complicated interpretation of multiple-scattering of lasers by clouds. Fortunately, in most practical cases (in respect to aerosol number density and the common field of view of lidar systems), the scattering from aerosols (except for water and ice clouds) is mainly (>90%) single scattering.

One technique of measuring qualitatively to what degree the sphericity and the single-scattering assumptions are valid, is provided by depolarization measurements of the linearly polarized incident laser beam. According to theory and experiment, a depolarization of 1.5% (Cohen, et al., 1969) or less is expected when these assumptions are valid. Experimental values of the depolarization factor were recorded by a few laser radar groups and in some cases depolarization of 10% and higher were detected. Under average aerosol density conditions, exact depolarization values are difficult to measure accurately and therefore are assumed not to exceed 2%.

† Readers interested in specified properties of atmospheric aerosol including composition, size-distribution and tabulated optical properties should refer for example, to: Elterman, et al. (1969) and McClatchey, et al. (1971).

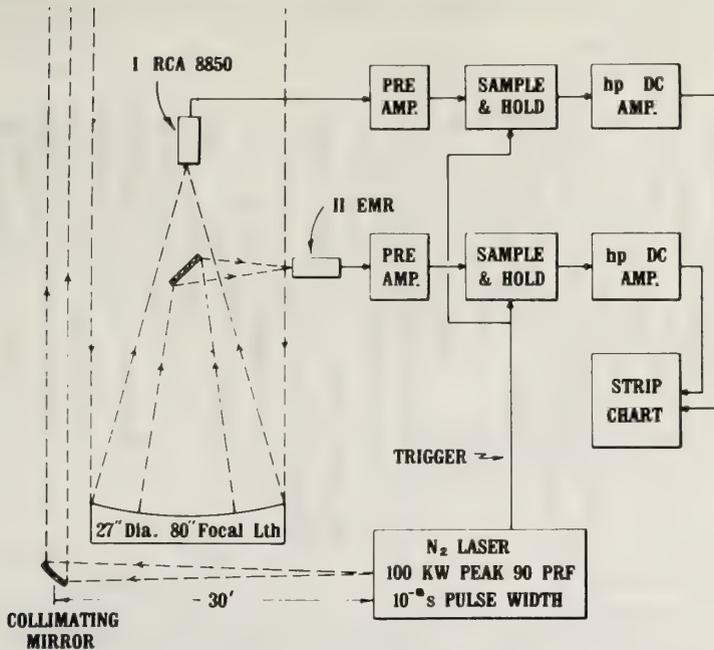


Figure 23.12 Schematic of the experimental system used for the turbidity measurement.

Although these favorable atmospheric conditions (e.g. single scattering and sphericity) reduce the number of the parameters, $I_{k,M}$ is still a function of several variables:

$$I_{k,M} = f(r_1, \dots, r_N, m_1, \dots, m_k, \rho_1, \dots, \rho_N, \theta, \lambda) \tag{23:23}$$

where θ (the scattering angle) and λ are known, whereas the other variables have to be determined (ρ_j is the number density of the particles with radius r_j and refractive index m_j). Still, the conditions were described above as being “favorable”, because $I_{k,M}$ can now be calculated for any given set of parameters. Indeed, one of the first widely used approaches of relating scattering information to aerosol properties had been to calculate $I_{k,m}$ values for a number of atmospheric aerosol models (see, for example, Diermendjian 1964) and compare the results with the laser radar measurements (Barrett, et al., 1967; see 23.4.2.2). Calculations were performed for a constant index of refraction (for each model) and a continuous particle size distribution function. Naturally this method is very limited in its applicability to the real atmosphere and different general methods had to be developed. (A discussion on the applicability of theoretical tables to the real atmosphere can be found in the work by Dave, 1969).

In order to better understand the basic requirements for any method of determining aerosol properties, let us now examine briefly the behavior of the Mie scattering intensities (10.4) as a function of the parameters r, λ, θ and m :

A. The dependence on the size parameter $\alpha = 2\pi r/\lambda$ (see F23:13). The main characteristic of this dependence is the nonmonotonic fluctuations of the scattering efficiency as a function of the aerosol size; for example, three periodicities can be found: $\sim 0.8 \cdot \alpha$; $\sim 2(m - 1) \cdot \alpha$ and $\sim 0.277 \cdot \alpha$. The fluctuations in the scattering efficiency can be reduced either by increasing drastically the laser radar field of view or by using a wideband light source for the transmitter. Neither possibility is applicable in the remote lidar measurements.

B. The dependence on very small variations of the refractive index m . In (T23.6) the dependence of the refractive index on wavelength is shown and the resulting variations in the scattering intensities are shown in (F23.14). Figure (23.15) shows the dependence of the scattering intensity from a monodispersed cloud of

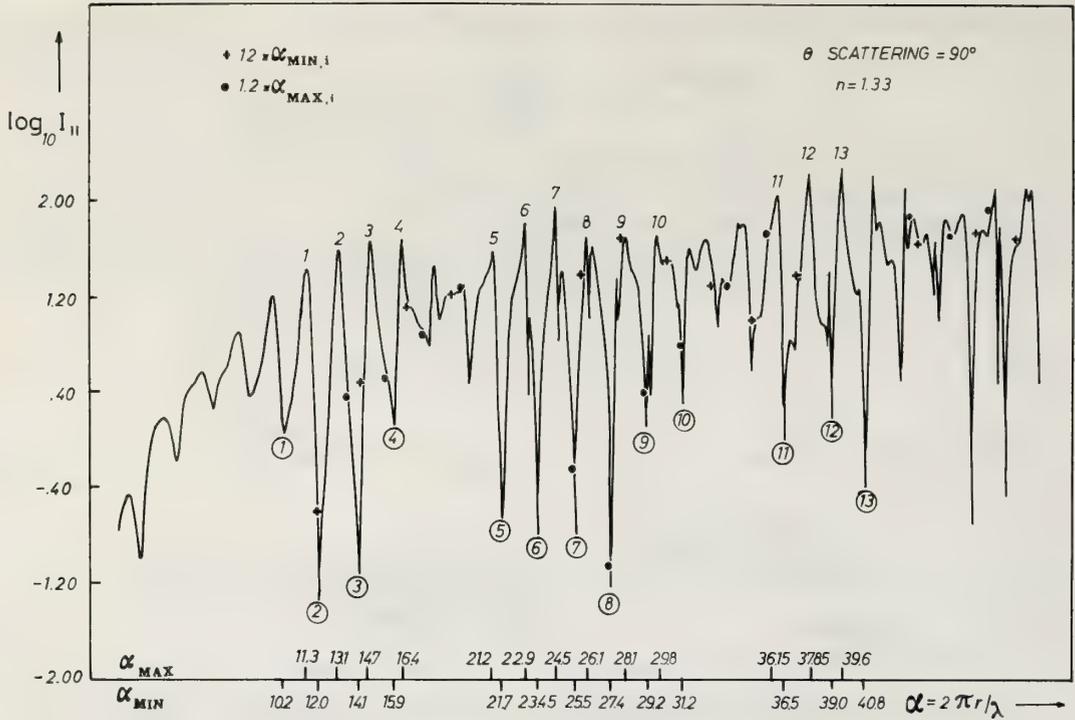


Figure 23.13 Mie scattering intensity as a function the size parameter $\alpha = 2\pi r/\lambda$. The scattering angle θ is 90° and the refractive index of the spherical scatterers is 1.33 (water droplets).

Table 23.6 Refractive index variations associated with wavelength and temperature changes.

I. Refractive index of liquid water ($t=20^\circ\text{C}$) as a function of wavelength.

Wavelength (Å)	4046.6	4471.5	5015.7	5460.7	5892.6	6562.8	7065.2
Refractive Index	1.342724	1.339423	1.336363	1.334466	1.332988	1.331151	1.330019

II. Refractive index of liquid water ($\lambda=5892.6\text{Å}$) as a function of temperature.

Temperature ($^\circ\text{C}$)	0	14	15	16	20
Refractive Index Relative to Air ($\rho = 1.225 \times 10^{-3} \text{g/cm}^3$)	1.3339493	1.33348	1.33341	1.33333	1.33299
Refractive Index in Vacuum			1.33377		1.33335

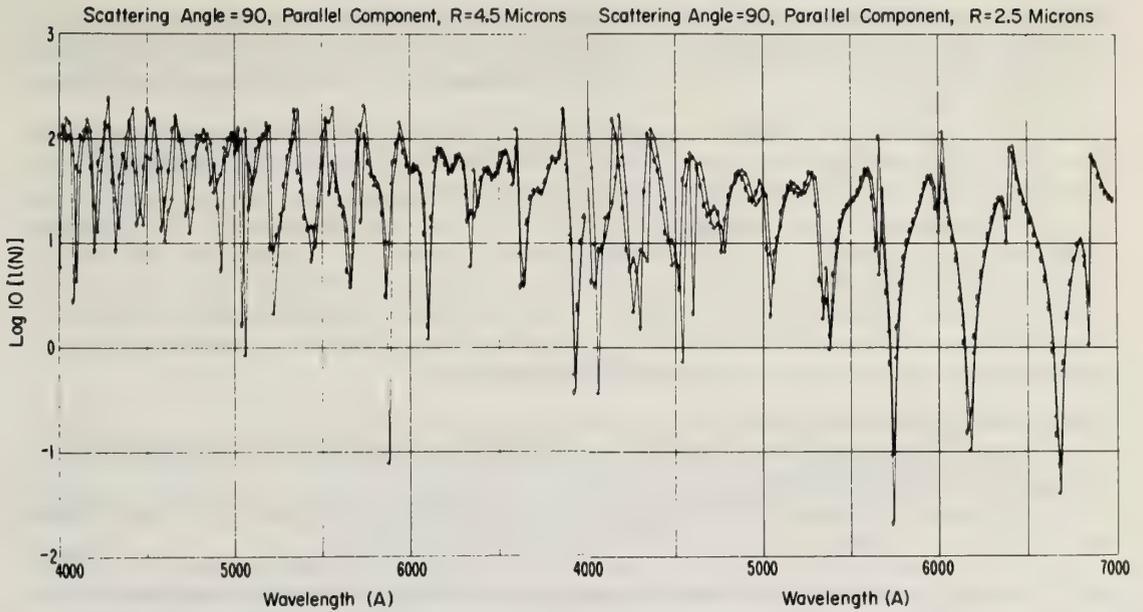


Figure 23.14 The log of the scattering intensities of the component normal to the scattering plane as a function of wavelength; the scattering angle $\theta = 90^\circ$, $\rightarrow m = 1.33$; $\circ \Rightarrow m = m(\lambda)$. Left: R (= radius of scattering sphere) = $4.5\mu\text{m}$ and $4000\text{\AA} \leq \lambda \leq 7000\text{\AA}$; Right: $R = 2.5\mu\text{m}$.

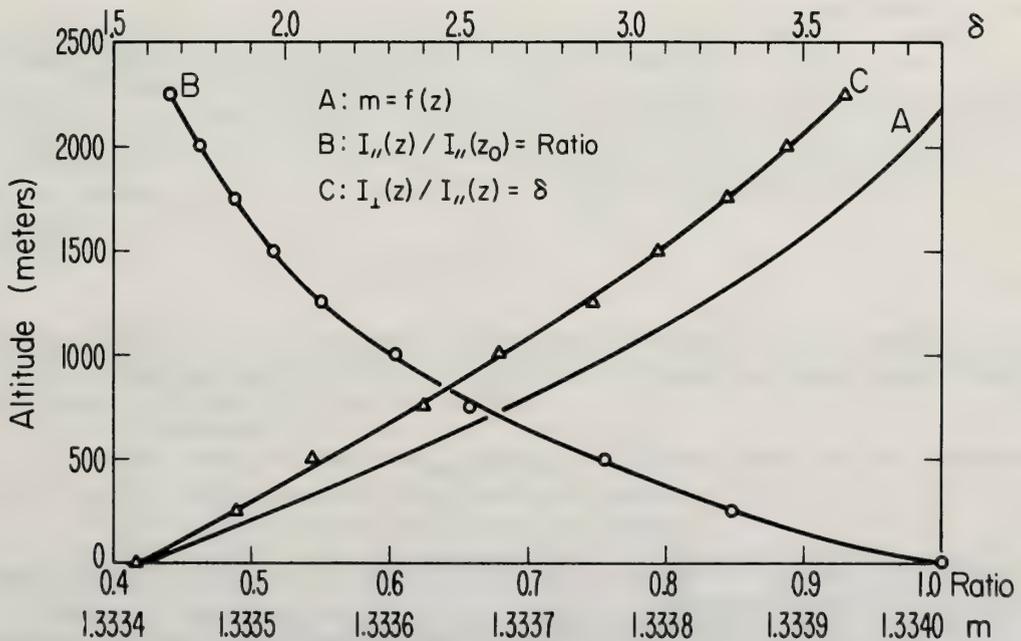


Figure 23.15 The dependence of scattering intensity as a function of height: A - $m = f(z)$; B - intensity ratio of the components parallel to the scattering plane $I(z)/I(z_0)$ as a function of height ($\theta=90^\circ$; $\lambda=5892.6\text{\AA}$; $R = 2.6\mu\text{m}$); C - depolarization (δ) of the scattered light as a function of height: $\delta = I_{\perp}(z)/I_{\parallel}(z)$. I is the component parallel to the scattering plane.

particles in a standard atmosphere, as a function of height, due to changes in the relative index of refraction. Fortunately, many of the actual atmospheric aerosol size distributions compensate for the refractive index dependence, although a careful study has not yet been made to determine the conditions under which the refractive index effect is important.

C. The dependence on the scattering angle θ ; see, for example, Eiden (1966). This dependence also shows a fluctuating scattering intensity as a function of θ . The envelope of the fluctuations has a minimum in the neighborhood of $\theta = 90^\circ$ and a high maximum for forward scattering.

As a consequence, if no *a priori* assumptions are made and one is left with all possible sizes and indices of refraction, the exact solution requires more independent measurements with a common laser radar than the number of unknowns since there is no one to one correspondence between scattering intensity and size, index of refraction, or scattering angle. A high degree of *a priori* information is, therefore, essential in order to reduce the number of the required independent measurements, which are provided by multi-wavelength laser radars, bistatic laser radars (θ changes) and polarization measurements.

23.4.2.1 Basic Characteristics of Atmospheric Aerosols

Samples of aerosol particles collected from the atmosphere have been studied in the laboratory and the results can be summarized as follows:

a. The real part of the index of refraction varies from 1.33 to 1.59 (relative to the air), and for a large class of aerosols can be taken as 1.5 or 1.54, Eiden (1966), Herman, et al. (1971).

b. Most aerosols can be described by a size-distribution function with a few parameters, Diermendjian (1969). A typical, nonprecipitating, terrestrial atmosphere contains particles whose radii vary between 0.01μ and 15μ (Dave, 1969). Based on these results most of the authors suggesting mathematical methods for the inversion of a few independent measurements used the following assumptions:

1. The index of refraction is known and *unique*. Thus particles may differ only in size.

2. A "best fit" to one of the suggested possible size distribution functions should be the target of the inversion procedure, or in some cases, a close resemblance to an initial estimate of the particle size spectrum. Single scattering of spherical particles is assumed and weighting functions (scattering as a function of size) for spherical particles are used.

A theoretical discussion of various inversion procedures, including the solution to Fredholm's integral equation of the first kind, are presented in Chapter 16, although their application to the aerosol size distribution is not emphasized. One example of the theoretical approach to the inversion method based on the above assumptions for the determination of the size spectrum function, $N(r)$, is presented by Dave (1971). We shall outline the main equations used in an inversion procedure for aerosol size distribution determination:

The range of particle sizes is divided into J sub-intervals:

$$N_j = \int_{r_j}^{r_{j+1}} N(r) dr / (r_{j+1} - r_j) \quad (23:24)$$

where N_j is the total number of particles per unit volume per unit size increment with radii between r_j and r_{j+1} . Similar expressions can be calculated for the initial estimate of the size spectrum function, $N^*(r)$:

$$N_j^* = \int_{r_j}^{r_{j+1}} N^*(r) dr / (r_{j+1} - r_j)$$

This division reduces the number of unknowns. It is based on the approximation that the scattering intensity variations are small within each subinterval.

Suppose we have M independent observations (scattering measurements in various wavelengths, scattering angles, or polarizations). Then for each observation, m , the following matrix elements can be defined:

$$a_{mj} = \underbrace{\text{const.}(\lambda)}_{C_\lambda} \cdot \int_{r_j}^{r_{j+1}} I(\alpha_m, \theta_m, P_m) dr \quad (23:25)$$

where α_m , θ_m and P_m define the wavelength, scattering angle and polarization of the m^{th} observation. I is the Mie scattering intensity for the assumed index of refraction, and C_λ is constant for a fixed λ . If D_m is the scattering intensity measured for the m^{th} observation, then:

$$D_m = \sum_{j=1}^J N_j \cdot a_{mj} \quad (23:26)$$

Thus, we have a system of linear equations that have the matrix form:

$$A \cdot N = D \quad (23:27)$$

The least-square solution is:

$$N = (A^*A)^{-1} A^*D \quad (23:28)$$

where a unique solution exists only if A^*A is nonsingular. In many cases a smoothed solution is required to avoid the appearance of possible negative particle number density in the mathematical solution. (For the smoothed solution see, for example, Twomey 1962). When only one measurement is used, only one equation can be written and, therefore, the size distribution must be assumed and the aerosol concentration is the only unknown.

23.4.2.2 Examples of Inverting Laser Radar Measurements to Obtain Aerosol Properties

Example 1: Barret, et al. (1967):

1a. Assumed known properties: The size distribution function has the form:

$$N = c \int_{r_1}^{r_2} r^{-\nu} d(\ln r)$$

where $r_1 = 0.04\mu$; $r_2 = 10\mu$ and $\nu = 3$; the aerosol refractive index is constant, $m = 1.5$.

1b. Unknowns: Aerosol concentration.

1c. Number of required independent measurements: one, obtained by laser radar backscattering measurements at 6943\AA .

1d. Method: Mie scattering tables are used for the definition of the expected scattering from a unit mass of the aerosol and if the measured aerosol backscattering b_k is expressed in units of $10^{-7}\text{m}^{-1}\text{sr}^{-1}$ the total aerosol mass is given by:

$$Q(\mu\text{g m}^{-3}) = 4.96(b_k - 5.8) \quad (23:29)$$

Expressions for the visibility distance and the turbidity index (\equiv total scattering at $\lambda = 5500\text{\AA}$) are also suggested. This method for the estimation of the total aerosol mass, has also been used for stratospheric measurements (see, for example, Bartusek, et al. 1970).

Example 2: Herman, et al. (1971).

2a. Assumed known properties: first, estimate for the size distribution—the Junge distribution, $dN = ar^{-(\nu-1)}$ with an unknown value for ν ; index of refraction = 1.54.

2b. Unknown: The exact size-distribution function of all particles.

2c. Number of required independent measurements: 20—the four Stokes' polarization parameters at five scattering angles θ .

2d. Method: Although the only unknowns are the number densities of one type of spherical particle with $.025 \leq r \leq 10\mu$ an exact inversion solution requires the division of the size range into at least 1000 subdivisions (Dave, 1969) and a corresponding set of 1000 independent observations (different scattering angles and polarization parameters). Therefore, an iteration technique is suggested in which an "intelligent" trial solution—an assumed size distribution—is made and the resulting size distribution function (based on the observations and on a calculated inversion matrix), is used as a input for the second iteration step.

It should be pointed out that the resulting solution will depend on the initial trial solution.

Example 3. Cohen (1971).

3a. Known properties: Upper limit of size range.

3b. Unknowns: Content and size distribution in the atmospheric scattering volume of spherical particles with known indices of refraction.

3c. Number of required independent measurements: Several scattering intensities as a function of wavelength (dye-laser radar). The number of observations is not required to correspond to the number of unknowns.

3d. Method: To the particles having a radius r_0 and refractive index m_0 , two series of λ values can be fitted ($\lambda_{i0}, \lambda'_{i0}$), so that;

$$\lambda_{i0} = 2\pi r_0 / \alpha_{\text{MAX},i}; \lambda'_{i0} = 2\pi r_0 / \alpha_{\text{MIN},i} \quad (23:30)$$

where $\alpha_{\text{MAX},i}$ ($\alpha_{\text{MIN},i}$) is the value of the size parameter α corresponding to i -th maximum (minimum) point of the theoretical $I(\alpha, \theta, m_0)$ —see (F23.13).

If we define:

$$f(r_j) = \sum_{i=1}^N [I(\lambda_{ij}) - I(\lambda'_{ij})], \quad (23:31)$$

the function $f(r_j)$ describes the size distribution with a degree of accuracy which increases with N . The N value, representing the “available” number of maxima (and minima, respectively), can be increased to a practically desired value by recording $I(\lambda)$ at various scattering angles. Since the accuracy of this technique is better for the large particles, $f(r_j)$ is calculated starting from the largest assumed r value so that the error to the function $f(r_j)$ introduced by the bigger particles can be estimated and taken into account.

23.5 Future Directions for Laser-Probing of the Atmosphere

The feasibility of water vapor profile measurements by laser radar using Raman backscatter of selective absorption has been demonstrated. Raman profiles of O_2 and N_2 have also been obtained and from these measurements a temperature profile should be possible although this is not yet demonstrated in the lower atmosphere. Measurement of cloud height should be trivial and measurement of cloud velocity should be possible. Measurement of aerosol profiles has been demonstrated.

New capabilities for laser radar meteorology will be demonstrated when the dye laser becomes suitable for field laser radar probes. Resonant absorption experiments should allow profile measurements of additional molecules. Resonant Raman scattering or fluorescence scattering may also become feasible for some molecules. Advances in peak power, repetition rate, and beam divergence will make Raman laser measurements more attractive. Daylight operation of Raman laser radar will probably require lasers in the 2700 to 2900Å region to take advantage of the ozone attenuation of solar radiation. Improved filters are needed at ultraviolet wavelengths; the Fabry-Perot interferometer may be useful for field work. The dye laser will allow improved techniques for separating aerosol and molecular backscatter by multi-wavelength measurements. If these methods are perfected, the density, pressure, and temperature profiles could be measured in the lower atmosphere.

The methods described in 23.4 for determining aerosol properties from laser radar scattering measurement illustrate that the problem is still far from being solved for most practical applications. Monochromatic-monostatic laser radar measurements seem to be of little practical use for determining aerosol properties. Future laser radars should provide polarization measurements in addition to fast wavelength and/or scattering angle scanning. More independent observations will increase the accuracy of the result achieved by an inversion procedure and/or decrease the amount of *a priori* information needed concerning the aerosol properties. Application of improved inversion procedures (see Chapter 16) will also lead to improved aerosol measurements. On the other hand, for turbidity and extinction measurements, the monostatic laser radar will play a significant role in the future.

Another possible new development for meteorology is the Doppler laser probe. The power spectrum of the Doppler return from molecules will be temperature dependent and may be a suitable method for measuring temperature. Wind velocity from the mean Doppler shift of backscattered energy from aerosols is

another possible use. The problems of coherence of lasers with high power has prevented atmospheric experiments. Problems of optical doppler have been discussed (Schotland, 1968; Owens, 1969) and appreciable laser advances will be needed for practical optical Doppler systems for atmospheric probing. However, the optical Doppler problem has thus far been approached from the standpoint of coherent optical systems, where the severe restrictions of optical heterodyning are encountered. Another approach, using coherent microwave modulation of the laser transmitter may be a solution to the Doppler measurement problem. This approach has been used in precise distance measuring devices.

23.6 References

- Barrett, E. W. and O. Ben-Dov (1967), Application of the Lidar to Air Pollution Measurements, *J of Appl. Meteor.* 6, 500-515.
- Bartusek, K., D. J. Gambling, and W. G. Elford (1970), Stratospheric aerosol measurements by optical radar, *J. Atmos. Terr. Phys.*, 32, 1535-1544.
- Blau, H. H., Jr., M. L. Cohen, L. B. Lapson, P. von Thuna, R. T. Ryan, and D. Watson (1970), A prototype cloud physics laser nephelometer, *Appl. Opt.* 9, 1798-1803.
- Bowman, G., A. J. Gibson, and M. C. W. Sandford (1969), Atmospheric Sodium Measured by a Tuned Laser Radar, *Nature*, 221, 456-457.
- Call, R. W., E. Palmer, and R. Grow (1967), Light Scattering, Univ. of Utah Tech. Rept. NSF-11.
- Cohen, A., J. Neumann, and W. Low (1969), An experimental determination of the depolarization of scattered laser light by atmospheric air, *J. Appl. Meteor.*, 8, 952-954.
- Cohen, A. (1971), The size spectrum determination of spherical aerosols by light scattering—Part I: Method, *Proc. of the Air Pollution, Turbulence and Diffusion Symposium, Las Cruces, New Mexico* (in press).
- Cohen, A., and V. E. Derr (1972), A calibration technique for lidar measurements and average air density definition by a multiple wavelength lidar, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.
- Cohen, A., V. E. Derr, R. E. Cupp, and J. J. Barth (1972), On the correlation of continuous lidar echos from separated altitudes in the lower troposphere, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.
- Collis, R. T. H. (1970), Lidar, *Appl. Opt.*, 9, 1782-1788.
- Collis, R. T. H. (1971), Lidar observations in turbulence and diffusion studies, *Proc. of the Air Pollution, Turbulence and Diffusion Symposium, Las Cruces, New Mexico* (in press).
- Cooney, J. (1968), Measurements on the Raman Component of Laser Atmospheric Backscatter, *Applied Phys. Lett.*, 12, No. 2, 42-44.
- Cooney, J. (1970), Remote Measurements of Atmospheric Water Vapor Profile Using the Raman Component of Laser Backscatter, *J. of Appl. Meteor.*, 9, 182-184.
- Dave, J. V. (1969), Effect of coarseness of the integration increment on the calculation of the radiation scattered by polydispersed aerosols, *Appl. Opt.*, 8, 1161-1167.
- Dave, J. V. (1971), Determination of size distribution of spherical polydispersions using scattered radiation data, *Appl. Opt.*, 10, 2035-2044.

- Deirmendjian, D. (1964), Scattering and polarization properties of water clouds and hazes in the visible and IR. *Appl. Opt.*, *3*, 187-196.
- Deirmendjian, D. 1969: *Electromagnetic Scattering on Spherical Polydispersions*, American Elsevier, New York.
- Derr, V. E. and C. G. Little (1970), A Comparison of Remote Sensing of the Clear Atmosphere by Optical, Radio, and Acoustic Radar Techniques, *Appl. Opt.*, *9*, No. 9, 1976-1992.
- Derr, V. E., R. E. Cupp, and A. Cohen (1972), Active wind measurements by lidar, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.
- Eiden, R. (1966), The elliptical polarization of light scattered by a volume of atmospheric air, *Appl. Opt.*, *5*, 569-575.
- Elterman, L. (1953), A Series of Atmospheric Temperature Profiles Obtained with the Searchlight Technique, *J. Geophys. Res.*, *58*, 519-530.
- Elterman, L., R. Wexler, and D. T. Chang (1969), Features of tropospheric and stratospheric dust, *Appl. Opt.*, *8*, 893-903.
- Ferrara, R., G. Fiocco, and G. Tonna (1970), Evolution of the fog droplet size distribution observed by laser scattering, *Appl. Opt.*, *9*, 2517-2521.
- Fiocco, G., and G. Grams (1964), Observation of the aerosol layer at 20 km by optical radar, *J. Atmos. Sci.*, *21*, 323-324.
- Grams, G. W. (1970), Laser radar studies of the atmosphere above 50 km, *J. Atmos. Terr. Phys.* *32*, 729-736.
- Grimes, A. E., 1969: *An Annotated Bibliography on Methods of Visibility Measurement, 1950-1969*, ESSA Tech Memo ATSTM, Lib 2 PB-188652, Springfield, Virginia.
- Hall, F. F., Jr., and H. Y. Ageo (1970), Absolute calibration of a laser system for atmospheric probing, *Appl. Opt.*, *9*, 1820-1824.
- Hansen, J. E., and A. Arking (1971), Clouds of Venus: evidence for their nature, *Science*, *171*, 669-672.
- Herman, B. M., S. R. Browning, and J. A. Reagan (1971), Determination of aerosol size distributions from lidar measurements, *J. Atmos. Sci.*, *28*, 763-771.
- Herzberg, G., 1960: *Molecular Spectra and Molecular Structure, 2, Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, Princeton, N.J.
- Holland, A. C., and G. Gagne (1970), The scattering of polarized light by polydisperse systems of irregular particles, *Appl. Opt.*, *9*, 1113-1121.
- Junge, C. E., 1964: *Air Chemistry and Radio Activity*, Academic Press, New York.
- Kent, G. S. and R. W. H. Wright (1969), A Review of Laser Radar Measurements of Atmospheric Properties, *J. Atmosph. Terr. Phys.*, *32*, 917-943.
- Kobayasi, T. and H. Inaba (1970), Spectral Analysis of Atmospheric Nitrogen by Laser-Raman Radar, *Opto.-Electronics*, *2*, No. 1.

- Kobayasi, T. and H. Inaba (1970), Spectroscopic Detection of SO₂ and CO₂ Molecules In Polluted Atmosphere by Laser-Raman Radar Technique, *Appl. Phys. Lett.*, *17*, No. 4, 139-141.
- Leonard, D. (1967), Observation of Raman Scattering from the Atmosphere Using a Pulsed Nitrogen Ultraviolet Laser, *Nature*, *216*, 142-143.
- McClatchey, R. A., R. W. Fenn, J. E. A. Selby, F. E. Voltz, and J. S. Garing, 1971: *Optical Properties of the Atmosphere* (revised). AFCRL-71-0279, Bedford, Massachusetts.
- McCormick, M. P. (1970), Simultaneous multiple wavelength laser radar measurements of the lower atmosphere, presented at the Electro-Optical Systems Design Conference, New York.
- Melfi, S., J. Lawrence, and M. McCormick (1969), Observations of Raman Scattering by Water Vapor in the Atmosphere, *Appl. Phys. Lett.* *15*, No. 9, 295-297.
- Melfi, S. H. (1972), Remote measurement of atmospheric transmissivity, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.
- Mie, G. (1908), Beitrage zur optik truber medien, *Ann. Physik*, *25*.
- Mizushima, S., 1958: *Handboch der Physik*, S. Flugge, Ed., Springer-Verlag, Berlin.
- Owens, J. (1969), Optical Doppler Measurement of Microscale Wind Velocity, *Proc. of IEEE*, *57*, No. 4, 530-536.
- Panofsky, W. and M. Phillips, 1955: *Classical Electricity and Magnetism*, Addison-Wesley, Reading, Mass.
- Penndorf, R. (1957), Tables of Refractive Index for Standard Air and the Rayleigh Scattering Coefficient for the Spectral Region Between 0.2 and 20.0 μ and Their Application to Atmospheric Optics, *J. Opt. Soc. Am.*, *47*, 176.
- Quenzel, H. (1965), Influence of refractive index on the accuracy of size determination of aerosol particles with light-scattering aerosol counters, *Appl. Opt.*, *8*, 165-169.
- Sakurai, K. and H. J. Broida (1969), Spectral Study of NO₂ Fluorescence Excited by 11 Lines of Argon and Krypton Ion Lasers, *J. Chem. Phys.*, *50*, 2404-2410.
- Schotland, R. M. (1965), Study of Active Probing of Water Vapor Profiles and Results of Experiments, New York University Dept. of Meteorology and Oceanography, Geophysical Sciences Laboratory Report No. 65-6.
- Schotland, R. M. (1968), Some Aspects of Remote Atmospheric Sensing by Laser Radar, Contribution No. 62 of the Geophysical Sciences Laboratory, New York University.
- Schotland, R. M., K. Sassen, and R. Stone (1971), Observations by lidar of linear depolarization ratios for hydrometeors, *J. Appl. Meteor.*, *10*, 1011-1017.
- Schotland, R. M., and N. M. Reiss (1972), Remote measurement of atmospheric turbidity by means of rotational Raman scattering of laser light, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.
- Schuster, B. G., J. J. de Luisi, G. W. Grams, and R. K. Sato (1972), A technique for separating the Mie and Rayleigh contributions from lidar backscattering signatures, presented at the Fourth Conference on Laser Radar, Tucson, Arizona.

Strauch, R. G., V. E. Derr, and R. E. Cupp (1971), Atmospheric temperature measurement using Raman backscatter, *Appl. Opt.*, 10, 2665-2669.

Twomey, S. (1963), On the numerical solution of fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, *J. Assoc. Comput. Mach.*, 10, 97-101.

Yamamoto, G., and M. Tanaka (1969), Determination of aerosol size distribution from spectral attenuation measurements, *Appl. Opt.*, 8, 447-453.

List of Symbols

a_{mj}	a matrix element which determined the scattering intensity for the m^{th} wavelength, scattering angle, or polarization, and the j^{th} sub-size-range	I_R	Rayleigh scattering intensity
		I_M	Mie (aerosol) scattering intensity
		I_D	photodetector dark current
A	matrix containing the elements a_{mj}	k	imaginary part of complex refractive index
A_t	cross sectional area of transmitted beam	$k_M(m, \alpha)$	ratio of scattering cross section to geometric cross section for Mie scatterer
A_r	receiver aperture	L	length of scattering region
A_j, A_j^*	system calibration coefficients	m	complex refractive index
$b_{k,j}$	aerosol scattering coefficient at Z_k for wavelength or polarization j	m_s	refractive index for a standard atmosphere
b_k	measured aerosol backscattering	M	molecular weight
c	velocity of light	n	real part of complex refractive index
c_k	aerosol number density ratio	N	noise photons counted per second
dN	number density of particles with radii $r \rightarrow r + dr$	N	vector of $J N_j$ values
D	vector of $M D_m$ values	N_j	number density of particles
D_m	scattering intensity recorded at the m^{th} observation	N_r	number of photons received per laser pulse
e	charge of electron	N_s	molecular number density for a standard atmosphere
E	partial pressure of water vapor	N_t	number of photons transmitted per laser pulse
$f(\vec{r}, m)$	aerosol size distribution function	N_B	photons counted because of background noise
g	acceleration of gravity	N_D	photodetector dark counts per second
G	photodetector gain	P_t	peak transmitted power
$g(\theta, \phi, m, \alpha)$	angular dependent scattering function for Mie scatterer	P(h)	pressure at height h
h	Planck's constant	q	a parameter used in the definition of a specific polytropic atmosphere
h	height		
$I, (I_0)$	radiation intensity (incident)		
I_T	total scattering intensity		

Q	total aerosol mass	η_q	quantum efficiency of photodetector
r	particle radius	η_r	efficiency of receiver optics
R	universal gas constant	η_t	efficiency of transmitter optics
R, R_i	range	θ	scattering angle
R_k	molecular number density ratio	λ	wavelength
T	averaging time	ν_t	transmitted wavelength
T(h)	temperature at height h	ρ	number density of particles
v	Junge size distribution parameter	$\rho_k, \rho(h)$	density of air at height h
V	visibility distance	$\sigma, (\sigma_i)$	scattering cross section of single scatterer (for i-type scattering)
x	distance	$\sigma(\theta, \phi)$	differential cross section
Z, Z_k	height	$\sigma(\pi)$	backscattering cross section
Z	laser pulse repetition rate	$\sigma_M(m, \alpha)$	scattering cross section for Mie scatterer
α	particle size parameter	σ_R	scattering cross section for Rayleigh scatterer
$\alpha_{\max(\min)}$	size parameter for which the scattering intensity has a maximum (minimum) value	τ_t	pulse width of transmitter
$\beta, (\beta_i)$	extinction coefficient (for i-type extinction)	τ_λ	turbidity for a specific wavelength λ
β^1_A	extinction coefficient due to resonant absorption	ϕ	angle of polarization
δ	depolarization factor	Φ	laser-radar operating angle from the zenith
ΔP	pressure change	$\Psi, (\Psi_i)$	angular scattering function (for i-type scattering)
ΔT	temperature change	Ω	solid angle
$\Delta \lambda$	spectral bandwidth of the receiver	Ω_r	receiver solid angle

Chapter 24 REMOTE SPECTRAL SENSING OF POLLUTANTS

R. L. Schwiesow

Wave Propagation Laboratory

Environmental Research Laboratory

National Oceanic and Atmospheric Administration

Various spectroscopic techniques in the infrared, visible and ultraviolet regions of the spectrum are useful for the remote measurement of pollutant concentrations. Building on the fundamental theory of chapters 9 and 10, this chapter discusses the monitoring of ambient air quality by means of long path optical absorption and discusses pollutant distribution mapping by means of scattering spectroscopy. For each of the two measurement problems, techniques, examples and current research problems are considered. The information presented is designed to outline the potential and limitations of each technique for practical pollutant sensing.

24.0 Introduction

Remote sensing of pollutants implies the utilization of some interaction between the pollutant and the probing electromagnetic radiation. We limit this discussion to the propagation of electromagnetic radiation as a carrier of the information concerning pollutant concentration. This limitation is because other methods of information transmission such as acoustic waves have not yet demonstrated a capability or significant promise of sensing pollutant concentrations directly.

Two classes of pollution sensing problems are particularly suited to remote sensing. One class concerns the long path average of ambient pollutant concentration. The other general problem is the determination of a three-dimensional map of pollutant concentration with a spatial resolution appropriate to a given study. These two types of problems are treated in separate sections. Each section discusses the problem, detection principles and examples of actual remote sensing techniques and equipment. Information in chapters 9 and 10 is necessary to the understanding of detection principles involved in the methods of sensing discussed in this chapter.

While this chapter considers only remote sensing systems, the alert reader interested in additional information will also find useful ideas in articles (not referenced here) devoted to point (or sample cell) sensor spectroscopic instrumentation. In many cases, instrumentation designed for remote sensing is applicable with minor modification to point sampling. The converse is often also true.

The purpose of this chapter is to provide a conceptual framework in which to fit various sensing results as they become available. A reader interested in an engineering handbook for the selection of the proper remote sensing technique for a given problem will be disappointed. Unfortunately, the field of remote sensing of pollutants is not well enough developed to provide such a handbook. All published results of remote pollutant measurements and descriptions of working instruments, to the writer's knowledge, are included in the references to this chapter. The lack of published information dictates that we discuss some of the topics in generalities. There is no lack of proposed systems analyses (without results or new information) in the literature, but only a representative sampling of some of the more useful analyses have been included in the references to this chapter.

Only with further research will it be possible to be specific about a quantitative comparison (as a function of each pollutant type) of the different sensing methods discussed here. Summaries of current research problems are included in each section. Technical details, used to more fully explain the conceptual framework outlined here, are limited by available published information and by size constraints on the chapter.

24.1 Measurement of Ambient Air Quality

By the term ambient air quality we imply that the pollution sensing problem is concerned with spatial averages of pollutant concentration over distances of at least the order of hundreds of meters. Such averages can be determined by measuring the concentration at a number of evenly spaced sensors along a path. An average of the numerous point concentrations yields an approximation to the ambient air mean pollutant concentration.

Although the point averaging technique is useful, an ambient air quality measurement is better made using a technique which inherently integrates the pollutant concentration along the chosen path. The measurement of molecular absorption of electromagnetic radiation is such a technique. In principle, remote spectral sensing of molecular pollutants requires only a source, a detector and knowledge of the absorption spectrum of the molecules present in the sample.

We discuss variations on this principle in two spectral regions, the infrared and visible-ultraviolet. The physical reasons for this spectral limitation are explained more fully in 10.0. Techniques and examples are restricted by the physical limitations of the interaction mechanism. Microwave spectra and broadband spectra of particulate pollutants are mentioned at the end of this section.

24.1.1 Infrared Absorption

Characteristic vibration-rotation energy level transitions for molecules occur primarily in the infrared from 500 to 10,000 cm^{-1} . The basic explanation for these absorption transitions is given in chapter 9. For present purposes it is sufficient to observe that the presence of a given molecular species in an electromagnetic beam allows us to observe a characteristic spectrum; that is, a variation of transmission of radiant energy with photon energy, or equivalently, with photon wavenumber (given in cm^{-1} or waves per centimeter). The existence of the characteristic spectrum is quantitative evidence of the existence of the associated molecular species in the path.

Techniques

The basic spectroscopic technique for infrared measurement of ambient air quality involves a source and detector separated by a few hundred meters in typical cases. The path must be long enough for the desired spatial average and for an adequately strong absorption by the species of interest, and short enough for an adequate signal to noise ratio (S/N) in the spectrum of the molecular species. For convenience, a retroreflector is often used at the midpoint of the path so that source and detector may be physically (but not optically) close to each other.

A Globar is most often used as a continuum source of infrared radiation, although Wolfe (1965) discusses many alternative sources as well. The sun provides an intense but somewhat inconvenient infrared continuum for atmospheric studies. Applications of both continuum and line emission infrared sources to air pollution problems are discussed in more detail than is appropriate here by Hanst (1970). Continuum sources have been traditionally used for classical laboratory work. This background of experience and development makes the continuum sources practical for pollutant sensing, even though the spectral radiance (power per unit wavenumber interval) is lower than discrete frequency sources discussed below.

Tunable lasers provide an interesting bridge between continuum and discrete sources. Such lasers combine the extended spectral coverage (potentially over the range 500 to 2500 cm^{-1}) of the Globar with the high spectral radiance of line sources. Two types of tunable lasers are discussed by Hinkley and Kelley (1971) and by Mooradian et al., (1970). These tunable sources hold exciting promise for remote detection of atmospheric pollutants, but are still in the early stages of development. If problems of cost, tuning range, stability and portability can be solved, these lasers are ideal sources for ambient air monitoring.

Infrared sources which emit at a few discrete frequencies, such as CO_2 , Xe and other types of conventional lasers, have also been suggested as suitable sources for long path infrared spectroscopy. Discussions by Hanst (1970) and Long (1966) mention the possibility of using these sources, and give preliminary measurements of appropriate absorption coefficients of some molecular species at these laser frequencies. Unfortunately, coincidences between laser lines and strong absorption lines of molecules of interest are rather rare in practice, especially when one considers the absorption spectra of the gases in high (pressure broadened limit) resolution. Zeeman shifting a laser line to provide slight tuning (of the order of 1 cm^{-1}) has been suggested by Hanst (1970) and others to increase the possibility of a usable coincidence between laser emission and molecular absorption.

A variety of detectors can be used with each of the infrared sources considered. With the continuum sources, a detector with frequency (i.e. wavenumber or wavelength) discrimination is required. The simplest frequency discrimination involves the use of interference filters with band passes of tens of cm^{-1} at 1000 cm^{-1} (10 μm wavelength in the infrared). Either spectrally fixed filters may be used for selected spectral regions or wedge filters may be used to scan the spectrum. While filters are rugged, the extremely broad band passes mean that little discrete information is available. This is, many fewer pairs of frequency-transmission data points are measured by filters than are measured by higher resolution techniques used over the same spectral interval.

Traditionally, prism or grating spectrometers are used for photon frequency discrimination. Instruments suitable for field use exhibit spectral resolutions typically 10 to 1 cm^{-1} with the higher resolution associated only with grating instruments. The required entrance slit limits the collected power from large low-radiance sources and makes such detectors sensitive to atmospheric induced scintillation patterns because the fluctuating scintillation-induced intensity variations are the same spatial scale as the slit width dimensions.

Small interferometric spectrometers have also been proposed for long path infrared pollution monitors. Michelson interferometers compare favorably with portable grating spectrometers in resolution and ruggedness, and are outstanding in energy throughput and signal-to-noise ratios for weak sources. However, because the interferometer output must be processed via a Fourier transform to yield a conventional Transmission vs. Wavenumber spectra, data processing and equipment costs are much higher than for other spectrometers.

These three basic types of frequency discrimination all require photon detectors with broadband spectral response suited to the spectral region under study. The fixed or tunable line emission sources which do not require frequency discrimination (other than some sort of filter to reduce background radiation) also require detectors of the same type used with the spectrometers. Wolfe (1965) discusses the traditional detectors such as thermocouples and Golay pneumatic cells. Thermistor bolometers and pyroelectric detectors are more recently developed broadband devices which are commercially available. For cases of extremely long path length or high resolution, the source signal may be weak enough to require photovoltaic or photoconductive infrared detectors. These semiconductor devices reduce detector noise far below values typical for the traditional thermal detectors. Cooling the detector element is required for minimization of semiconductor detector noise levels.

Nondispersive infrared analyzers using the absorption properties of some molecular species for frequency discrimination are potentially useful for the remote spectral sensing of pollutants. Although available for years in a point sampling instrument for CO_2 , the nondispersive technique has only recently been suggested for remote measurements (see for example Hanst, 1970). In one example of the method, a cell containing the gas one wishes to measure is placed in the long path infrared beam alternately with a cell containing only clear air or perhaps nitrogen. If the same cell transmission is observed for both samples, obviously the gas of interest is strongly in the beam. If there is a difference in transmission between the cells, a lesser amount of the particular gas is present. The difference can be normalized and quantified in terms of the amount of gas present by various proprietary schemes involving filters, concentrations and the like. Another version of the principle uses the pollutant gas as the absorber in an otherwise nonabsorbing version of the Golay pneumatic cell detector. This makes the detector sensitive only in the regions where the gas of interest absorbs.

Examples

Many possible combinations of the sources, analyzers and detectors are currently being tried. A Globar source with variable wedge interference filter and a cooled photoconductive detector is being used near 1050 cm^{-1} for the detection of ozone. A similar setup with a grating monochromator scanned over wide ranges of 740 to 1450 cm^{-1} and 1800 to 3300 cm^{-1} is undergoing field trials for the Environmental Protection Agency near Los Angeles. Block Engineering is also utilizing a Globar and cooled detector, but with an interferometer, for long path infrared sensing. At NOAA and other laboratories, the sun in conjunction with a grating monochromator and thermocouple detector is used for remote spectral analysis of the atmosphere. Four lines of the CO_2 laser in combination with a retroreflector and pyroelectric detector are being applied by General Electric Co., Inc., to the measurement of ambient air quality in the field.

Preliminary results for all these techniques are promising but not conclusive. Need for further research is evident in the next section. None of these examples has yet produced results in the journal literature, to this writer's knowledge.

Current Research Problems: Reference Spectra

The instrument development outlined above is relatively well established; long path infrared spectra for the atmosphere under a variety of conditions can be produced by many different techniques. In order to extract useful information from the measured atmospheric spectra, a number of questions require further study.

Complex infrared vibration-rotation spectra are often very difficult to interpret. In principle, if the resolution of the instrument does not affect the spectrum, the transmission of radiant energy is given by

$$T(\nu) = \exp(-k(\nu)W) \quad (24:1)$$

where $k(\nu)$ is the absorption coefficient and W is the gaseous concentration times path length product in molecules/cm². At atmospheric pressure, linewidths of spectral lines are typically the order of 0.05 to 0.08 cm⁻¹. This means that instrumental resolution typically 1 cm⁻¹ or wider strongly affects the spectrum. One could formally replace $k(\nu)$ by $k(\nu, W)$ in (24:1) to mathematically model the resolution effects, but this is of little use since the variation of the resulting empirical absorption coefficient with concentration is different for every line and every instrumental resolution.

A related problem is the fact that the overall pattern and structure of a molecular spectrum changes with changes in resolution. Spectra of the same gas with different resolutions can be qualitatively unrecognizable. In practice, these two difficulties of interpretation demand that reference spectra be available in order to analyze the measured atmospheric spectra. The reference spectra must be taken under the same conditions of resolution and buffer gas pressure as are used for the ambient air measurements. They must be taken for a range of concentration times path length products. They must be quantitatively correct in terms of absolute (not relative) transmission values and gas concentrations.

Some useful spectra at approximately 10 cm⁻¹ resolution have been reported by Pierson et al., (1956). Sufficient information on molecular structure constants exists for H₂O, CO₂, CO, O₃, N₂O, CH₄ and HCl to allow calculation of spectra for any desired resolution. These calculations have been done at 1 cm⁻¹ resolution at NOAA and will be published soon. Experimental measurements of SO₂, HNO₃ and NH₃ at 1 cm⁻¹ resolution are underway at NOAA and other laboratories. Hinkley (1971) is making very high resolution measurements on SO₂, NH₃ and C₂H₄. Extremely high resolution measurements are required to give useful absorption coefficients for calibration of laser line absorption systems. Taylor and Yates (1957) have made a useful measurement of the transmission of a normal (i.e. unpolluted) tropospheric atmosphere with moderate spectral resolution. Clearly much additional research is required on suitable reference spectra. For example, quantitative, high resolution experimental or calculated spectra of H₂S, PAN, NO₂ and various hydrocarbons are needed.

Current Research Problems: Atmospheric Noise

In addition to source and detector noise which is familiar from laboratory work, the atmosphere introduces noise. The magnitude and spectral effect of this noise is not well known.

Atmospheric scintillation (11.3) introduces fluctuation of received signal strength. The power spectrum of this noise must be accurately known in order to determine what time filtering of the infrared signal will be most effective in limiting the noise from scintillation. This filtering will affect the maximum data rate at which the pollutant spectrum can be scanned, and may limit the ultimate accuracy of a practical remote sensing device.

Water vapor has significant continuum absorption even in the nominal infrared transmission windows of the atmosphere. Like atmospheric refractive index fluctuations, variation in water vapor content along the optical path introduce a noise signal on the spectrum. The behavior of water vapor concentration fluctuations in the atmosphere is not yet well documented. Such atmospheric knowledge is required to evaluate the potential difficulties of spectral remote sensing.

Changes in attenuation by aerosol scattering out of the infrared beam can also introduce unwanted spectral noise. In addition to a better understanding of atmospheric fluctuations and signal filtering, ratioing techniques involving comparison of broad band and narrow band signals are being tried to reduce the effect of atmospheric noise signals on the spectrum. The combination of atmospheric knowledge and instrumental development may reduce the atmospheric noise problem to the extent that it does not seriously limit the spectral measurements. Inaccuracies in concentration determination resulting from residual noise are best evaluated by means of controlled field experiments. Insufficient information is now available to assess the magnitude of the inaccuracy or the noise to be expected.

Current Research Problems: Data Analysis

The spectrum of a multicomponent gas mixture is at best complicated (9.6.1). Serious overlapping of the apparent spectral lines (especially at moderate or low resolution) precludes any simple few-point comparison of spectral transmission with a reference spectrum to determine the concentration of any particular molecular species such as CO. It is generally extremely difficult to recognize even qualitatively the spectrum of a desired gas in the spectrum of an unknown mixture.

Some proponents of the laser line absorption technique suggest that accurate concentration measurements are possible using a single, fixed frequency of infrared radiation for each gas to be determined. Such a simplistic approach has never been successfully demonstrated in a mixture of real gases such as exists in

the atmosphere. Interference or overlapping between the spectra of the various gases is probably the reason that the single frequency approach does not work well.

Transmission measurements at a few discrete frequencies can be made with multiline lasers (such as CO₂ or CO lasers) or with an assembly of interference filters. Moderate success with four lines of a CO₂ laser, analyzing for C₂H₄ in the absence of interferants, has been obtained by analyzing the one dimensional array of measured transmissions with an empirically determined 4 by 4 matrix of weighting factors (General Electric Co., Inc.). The result of this matrix correction yields the constituent concentrations. Clearly the accuracy and freedom from interference of such a scheme improves as the number of dimensions of the array increases. The applicability of a given correction matrix to be a range of ambient condition is uncertain. Interferences from new or unconsidered constituents are not yet evaluated for any particular example of this general technique.

It is possible to consider a large number of spectral points to analyze a spectrum for the concentration of pollutant gases. At NOAA an analysis program is underway which utilizes essentially all the data points in the 740 to 1450 cm⁻¹ and 1900 to 3300 cm⁻¹ infrared atmospheric window regions. The analysis involves the calculation of a trial spectrum using the reference spectrum library and a best estimate of gas concentrations. From the differences between the calculated and measured atmospheric spectra, correction factors are calculated to correct the original concentration estimates. The process is iterated until the gas concentration values are chosen to minimize, in a least squares sense, the difference between measured and computed spectra. In some senses this analysis procedure is an extension of the matrix weighting scheme discussed earlier, with an adaptive matrix and a large number of data points. The method works well on an idealized input spectrum with only 15 reference spectra, but the effects of noise and of constituents for which no reference spectrum exists have not been evaluated. The number of data points (approximately 8000 data points for the unknown spectrum and typically 6000 for each of 40 reference spectra) involved in the more complete analysis system makes the computational and storage problems significant.

Preliminary work has been done on a correlation analysis technique which handles a single constituent at a time. The concentration (choice of reference spectrum) is determined by maximizing the correlation coefficient between the reference and measured spectra. Problems of interference and normalization remain.

24.1.2 Visible and Ultraviolet Absorption

The spectral principles and many of the techniques involved in utilizing the visible and ultraviolet regions of the spectrum for measurement of ambient air quality are similar to those discussed for the infrared. In this spectral region, loosely 12,500 to 50,000 cm⁻¹, electronic rather than vibration-rotation transitions dominate the spectrum of a given atmospheric constituent. Chapter 9 discusses the physics involved. Here we are concerned with obtaining and analyzing the resulting spectrum.

Techniques

Only continuum sources have been applied for Vis-UV spectral measurements in the atmosphere. The source may be a hydrogen discharge lamp, a tungsten-halogen high temperature incandescent lamp, a xenon or mercury high pressure discharge lamp or the sun. Because of the " ν^4 law" (10.3) which expresses the fact that high UV photon frequencies are strongly scattered, scattered solar radiation is also a practical source for long path absorption spectroscopy in the near ultraviolet.

Tunable dye lasers or parametric oscillators are potentially useful line emission sources in the visible and are commercially available. However only a few pollutant molecules such as NO₂ have absorptions in this region. For these molecules, coherent sources are applicable in the same way as they are applicable to the infrared. Traditional absorption work in the Vis-UV has used continuum sources. Significant development is required yet on tunable UV sources to adapt them for useful long path measurements.

No fixed line emission sources appear useful in the Vis-UV region. The electronic spectra of molecules are generally less sharp and distinct than infrared spectra, causing serious overlap of pollutant spectra. In special cases, such as O₃ in an atmosphere of pure O₂ and N₂, line emission sources are useful because in this case no spectral overlap occurs in the O₃ absorption region. The general ambient air problem is much more difficult to analyze by fixed line sources.

For the continuum illumination sources, grating monochromators are most widely used for Vis-UV frequency discrimination. Typical resolutions for field instruments are in the range 2 to 20 cm⁻¹ at 20,000 cm⁻¹. Matching the spatially extended source image to the slit results in loss of signal but this signal loss is less than that experienced with other methods of spectral discrimination in the UV. In contrast to a monochromator which uses fixed entrance and exit slits, a polychromator configuration is often used. For the

polychromator arrangement, the exit slit is replaced by an assembly of slits. This assembly is spaced to correspond to the spectrum of the gas one wishes to detect. In effect the slit assembly acts as a simplified multipoint correlator. Kay (1967) discusses this type of instrument in more detail.

Interference filters (either fixed or wedge variable) are useful photon frequency discriminators in the visible. Fractional resolutions ($\Delta\nu/\nu$) are 0.1% at best, or about a 20 cm^{-1} or larger bandpass at $20,000\text{ cm}^{-1}$. Peak transmissions are approximately 50%. In the ultraviolet, filters are more difficult to construct. Bandpasses are no narrower than 0.5% and peak transmissions are typically 20%. Such poor performance suggests the use of grating instruments in the UV.

Scanning Michelson interferometric spectrometers have not been successful for field use in the Vis-UV. The short wavelengths of the light impose impractical mechanical tolerances on moving mirror devices and alignment systems. Fabry Perot interferometers have proven particularly valuable for some purposes, however, such as studies of Na and O in the airflow.

Photon detection in this spectral region is usually done with photomultipliers. These devices introduce little detector noise, and are sensitive and convenient. For applications where the received power level is high, such as the case for short paths or very low resolution systems, solid state photodiodes are ideal. In higher power situations, the dark current of the photodiodes (considerably higher current than that of a photomultiplier) is not a limitation.

Examples

Compared to the infrared, little ambient air quality measuring is currently done in the Vis-UV. Unpublished studies of NO_2 concentration in Europe using visible absorption involve a tungsten source and a small monochromator-photomultiplier detector.

Newcomb and Millan (1970) discuss atmospheric absorption studies in the ultraviolet. The spectra are analyzed using the grating polychromator correlation detection instrument. These workers use both high pressure gas discharge lamp sources for horizontal paths and scattered solar radiation for integrated absorption in a vertical direction. While questions can be raised concerning the quantitative accuracy of the pollutant measurements by the UV technique, the data do provide useful qualitative pictures of spatial and temporal variations in pollutant concentration. The UV correlation methods have been most useful for SO_2 and O_3 . Green (1966) discusses other molecules with UV absorption spectra suitable for possible analysis by this technique. The correlation approach may also be valuable for NO_2 absorption in the visible.

Current Research Problems

Most research in ambient air quality monitoring by means of Vis-UV absorption spectroscopy is directed toward refinements of existing instruments and procedures. Many fundamental spectral measurements are also to be made.

In contrast to the well-explored infrared, spectra of only a few molecules have been reported that are suited to monitoring by long path Vis-UV methods. Green (1966) provides an introduction to the source literature. Quantitative spectra of SO_2 and O_3 are rather well known (Griggs, 1963). Since the reported electronic spectra of these molecules is typically of lower resolution than the highest resolution infrared spectra reported, the dependence of the absorption coefficient $k(\nu)$ on concentration W is harder to evaluate than it is in the case of relatively sharper infrared lines. Therefore the requirements for reference spectra research are less certain in the Vis-UV than in the infrared. Further study is required on the quantitative spectra of NO_2 in the visible. Fink (1971) discusses some of the considerable uncertainties in the NO_2 spectrum.

Current development problems also include field trials of the Vis-UV correlation instruments to verify calibration assumptions and to determine the effects of various molecular interferences. Calibration cells with a known gas concentration are an effective method of frequent calibration, especially required when source stability (in the case of scattered solar radiation) is a problem.

The work of Curcio et al. (1964) is important in providing a reference spectrum for the normal (i.e. nominally unpolluted) atmosphere in the visible region. The work must be extended by further research in the ultraviolet region.

24.1.3 Other Long Path Pollution Sensors

We will not discuss in detail either the long path absorption of microwave radiation or the spectrally broad band attenuation caused by aerosols in the atmosphere.

Microwave spectroscopy produces very high resolution spectra of the rotational energy level transitions

in molecules. The levels involved in rotational transitions are much more closely spaced than are the levels involved in infrared vibrational or Vis-UV electronic absorptions. Therefore many different energy levels are populated, resulting in an extremely rich spectrum. Isotope effects are also clearly resolvable, adding additional lines. Pressure broadening of the multitude of lines is much larger, as a fraction of the absorbed photon frequency, than it is for other spectral regions. Therefore the interline spacing, on the average, is less than in the infrared. This results in a larger probability of spectral overlap. The net result of these effects is that at the microwave frequencies where molecules absorb, the ambient atmospheric absorption spectrum has a strong, quasi-continuous nature which is ill suited to any sort of constituent sensing.

Although microwave spectroscopy of pollutant gases can be done in a low pressure sample cell, spectroscopy in this region at 760 torr is neither useful nor promising as a long path ambient air quality monitoring technique.

Attenuation of optical radiation by means of aerosol scattering is a useful indicator of aerosol content. The entire study is better considered under the heading of visibility rather than spectroscopy. However some broad band spectral features can be distinguished in aerosol attenuation. Randerson (1970) has done some preliminary work in this area.

Unfortunately, no way of identifying aerosols from only absorption information is known. It is not possible to differentiate between natural and pollutant aerosols. There is no simple relation between attenuation and either number or mass density of aerosol concentration because of uncertainties in particle composition, shape and size.

Understanding details of aerosol attenuation (10.4.2) is important in order to evaluate the effect of this interference on infrared and visible-ultraviolet determinations of gas phase molecular concentration. Partial, but not definitive, information on aerosol concentration may also result from these aerosol attenuation studies.

24.2 Pollutant Mapping

In addition to ambient air quality (spatial average) pollutant concentration information, three dimensional mappings of pollutant molecule distribution are important for source monitoring and control. The ability to map these distributions, combined with information on wind velocities, is also useful in studying the movement, interaction and decay of various types of atmospheric contamination.

The operational elements of spectrally mapping pollutants involve constituent identification via spectroscopy and spatial resolution via range, elevation and azimuth information. By implication, only scattering spectroscopy will be useful where range information is required. Chapter 10 discusses in detail spectral scattering processes. For the purpose of this section one can examine the spectral variation in scattered light with a fixed pump frequency (Raman scattering), the spectral variation for a particular pump frequency (fluorescence), or the sharp variation in on-frequency return as the pump frequency is scanned (resonance reradiation).

Since the previous chapter as well as Derr and Little (1970) have given quantitative information on the lidar equation and other system variables, this section will be limited to spectroscopic and pollutant aspects of the lidar problem. We discuss general techniques and then examine examples of and research on the various applicable scattering mechanisms.

24.2.1 Lidar Principles

General optical techniques for pollutant mapping are common to the application of Raman, fluorescent, resonance and Mie scattering effects. Each type of scattering has special requirements which favor one of the many alternative sources and detectors available. Pollutants amenable to study by each type of scattering are discussed later under the separate headings.

Technique

Any source for scattering spectroscopy must have high spectral radiance (powerful, monochromatic and well collimated). A laser is the universal choice for lidar measurements in order to achieve the best possible S/N. Range information is most easily available if the laser is pulsed or frequency modulated, allowing a single ended experimental arrangement. A CW source may be used in a source beam crossed with receiver beam arrangement, but this is less convenient. Both the ν^4 dependence of scattering intensity on pump photon frequency (10.3) and receiver noise and quantum efficiency consideration favor a laser output as far toward

the ultraviolet as possible, although atmospheric attenuation due to O_2 prohibits work above $50,000\text{ cm}^{-1}$ and aerosol attenuation favors lower photon frequencies. The stratospheric ozone layer provides almost complete attenuation of solar radiation at energies above $33,000\text{ cm}^{-1}$. Therefore the background noise in the lidar receiver is strongly reduced if the laser source can be chosen so that the receiver operates above $33,000\text{ cm}^{-1}$. This optimization requirement is consistent with the ν^4 law. Actually, atmospheric background brightness decreases with increasing frequency from approximately $22,000\text{ cm}^{-1}$ upward, so that laser pump frequencies toward the ultraviolet are favored. Any spurious off frequency source radiation must of course be suppressed.

Unless an accidental coincidence occurs between a fixed laser line and a molecular or atomic absorption, use of fluorescence and resonance scattering requires a tunable laser. Both dye lasers and parametric oscillators are possibly useful, although dye lasers appear at the moment to be most highly developed for tunable pollutant lidar. NO_2 and Na are detectable in this manner, although insufficient information exists to evaluate detection limits.

Apparatus for frequency discrimination and detection is similar to that discussed previously for the visible and ultraviolet regions (24.1.2). Grating monochromators provide fractional resolutions ($\Delta\nu/\nu$) as good as 10^{-5} for narrow slits and 0.75 meter instrument focal lengths. High resolution in scattering spectroscopy gives an advantage in reducing background signal rather than an advantage in increasing information density as in the case of absorption spectroscopy. Hence, detectivity is improved. Up to the limits of the laser linewidth, the received background power is directly proportional to the receiver optical bandwidth. Two monochromators in series are often used for Raman spectroscopy. This reduces spurious contributions (as off band transmission) to the Raman signal from instrumental scattering of the much stronger on frequency scattering return. Off band rejections are typically 10^4 to 10^5 for each element of a monochromator. The principal disadvantage of the grating instrument is loss of signal in matching the image of the scattering region to the slit. This signal loss can be as large as orders of magnitude. In addition, the grating instruments exhibit reflective losses and are bulky and complex in comparison to filter systems.

Interference filters are simple and compact. The great advantage of filters for lidar applications, especially where scattering returns are weak, is that the wide filter apertures are easily matched to the image of the scattering region. In other respects the filters do not equal grating instruments. Off band rejections are typically 10^3 and fractional bandwidths are 10^{-3} (visible) to 5×10^{-2} (ultraviolet). The off band rejection characteristic of such narrow band filters is often improved by use of a sharp cut band-pass filter consisting of a color glass or liquid dye element. Filters, dye cells and all optical components intended for use in the ultraviolet must be carefully selected to avoid fluorescence problems.

Detection of the scattered light is done using photomultipliers. For strong signal returns (photon arrival rate greater than frequency response of photomultiplier) pulse height analysis of the signal yields scattered intensity. For weak signal returns (arrival rate less than frequency response) photon counting is appropriate. Either the photomultiplier tube itself may be gated by controlling the voltage on the first dynode, or the analyzing electronics only may be gated. The former technique is useful if photomultiplier saturation or overload would result from large returns at short ranges.

More than one spectral element may be examined in parallel if a spectrograph is used at the frequency analyzer and the photomultiplier is replaced with an image tube. Such a device is useful to speed data taking and to reduce detector noise, since the effective photocathode area for each frequency is small. These optical spectrum analyzers have recently become commercially available and should be useful for multicomponent remote sensing by scattering spectroscopy.

A receiving telescope of some sort is required to collect the scattered photons. Scattering irradiances typically a few photons per m^2 per second from 100 meters (23.1.4) for a 100 mW laser of course require large apertures of the order of 50 cm or more. For energy collection purposes a spherical mirror is adequate in a reflective system. Even at a relative aperture of $f/4$, the blur angle is 0.12 mrad, which is less than typical laser beam divergences.

Current Research Problems: Techniques

More powerful lasers than those currently available, especially tunable and ultraviolet types, are under development. Increases in the average power of dye lasers over present levels are resulting from dye material and cooling research. Effort is also directed toward making optical harmonic generation devices efficient and reliable. Doubled ruby lasers and quadrupled Nd:YAG lasers are especially promising for lidar work because they operate above $33,000\text{ cm}^{-1}$ where ozone attenuates the solar background. Commercial pulsed nitrogen

lasers have recently shown a five-fold average power increase over previously available levels. Research into methods for reducing beam divergence of such superradiant devices as the nitrogen laser will increase spectral radiance and allow lidar ranges to be extended by the square root of the increase in radiance for signal limited (shot noise) cases (23.1.3). Direct laser action, as opposed to harmonic generation, in the desired 33,000 to 50,000 cm^{-1} range has been demonstrated but not practically developed.

Fabry-Perot interferometers are being developed as spectral frequency discriminators in the ultraviolet. As a specialized type of efficient, often tunable interference filter, these devices overcome many of the bandwidth and flexibility limitations of conventional interference filters when applied to scattering. Further research on the image tube method of processing parallel channels of spectral information will increase the practical applicability of such spectral discrimination to lidar problems by reducing the dark current levels and increasing the quantum efficiency of the tubes.

Few new results of importance to lidar pollution sensing are expected from developments in electronics, photomultipliers or telescopes since present applications are not limited by these instrument elements.

24.2.2 Raman Lidar

Chapter 10 shows that Raman scattered photons are shifted in frequency from the exciting radiation by an amount characteristic of the scattering molecule. This shift uniquely identifies the molecule and allows quantitative determination of molecular concentration by measurement of the intensity of the shifted radiation. This true scattering process is not pressure quenched and so is pressure independent (see 10.3).

Example

Although many Raman lidar pollution measurements are in process, the work by Kobayasi and Inaba (1970a, 1970b) has yielded the only results so far in the technical literature. These workers have used a pulsed ruby laser and a single grating monochromator. Usable ranges are of the order of tens of meters for measurement of perhaps 300 ppm SO_2 in an artificial smoke plume, for example. Kobayasi and Inaba are doing further work with a pulsed nitrogen laser, a doubled ruby laser and filters for spectral discrimination.

Other work, which has not yet resulted in reported success for remote pollution measurements, uses equipment similar to that mentioned above. Doubled and even quadrupled Nd:YAG lasers are used for sources, as well as pulsed nitrogen, ruby and doubled ruby illuminators. Besides interference filters coupled with bandpass dye filters, image tube spectrographs, tandem (double) 0.25 meter monochromators, and echelle (higher order diffraction) spectrometers are being tried. S-20 type response photomultipliers are most common. SO_2 , NO and CO_2 are the principal pollutants studied in these experiments.

Achieved range and concentration limits are certain to be extended as further technique developments are put into practice. It is fairly obvious that pollutant mapping via Raman lidar will be limited to short range (up to 100 m) and near source concentrations (of 500 parts per million or more) for some time. In as far as the measurement accuracy is shot noise (photon statistic) limited, increasing range or concentration sensitivity by a factor of two requires a factor of four increase in the laser power times receiver sensitivity product.

The most convenient method of calibrating a lidar measurement is to compare the signal at the frequency shift of the molecule of interest to the signal at the nitrogen Q branch frequency. Signal ratios provide concentration ratios if the Raman scattering cross sections of the molecules (and ratio of atmospheric transmissions at the two frequencies) are known.

Current Research Problems: Scattering Cross Sections of Pollutant Gases

To calibrate field measurements, at least relative cross sections of pollutant molecules to nitrogen are required. Fouche and Chang (1971) report preliminary measurements at a 514.5 nm laser pump wavelength, without polarization analysis. Our preliminary results are shown in (T24.1). Measured quantities are defined in (10.3.2).

Depolarization ratios are required to accurately determine molecular scattering characteristics. Cross sections for all pollutants not represented in these references, at all available laser wavelengths, should be measured to at least a 10% uncertainty in order to support Raman lidar work.

Exciting research involves documenting and understanding the changes in the scattering cross section of a molecule with changes in incident laser frequency. Chapter 10 outlines the expected behaviour, and (T24.1) shows some results for water vapor. The larger scattering cross sections of this resonance Raman effect mean that by proper selection of source laser frequency it is possible to achieve significant increases in range and concentration limits in practical pollution lidars. A significant example of this resonance Raman scattering is

Table 24.1 Measured Relative Raman Cross Sections ($\sigma_{N_2} = 1$)

Gas	Laser Wavenumber, cm^{-1}	Relative Cross Section	Depolarization
SO ₂	20 981 (476.5nm)	5.2	.04
	20 487 (488.0nm)	5.5	.05
	20 135 (496.5nm)	5.5	.04
	19 430 (514.5nm)	5.6	.05
NO	20 981	.36	.11
	20 487	.36	.10
	20 135	.34	.11
	19 430	.39	.08
H ₂ O	20 981	2.5	
	20 487	1.9	
	19 430	3.3	

provided by mixed hydrocarbons as gasoline vapor. These hydrocarbons scatter on the average 115 times as strongly as nitrogen, on a per molecule basis, at a total pressure of 750 torr. For the experiment a 20, 487 cm^{-1} pump was used. Cross section variation with incident frequency for these molecules has not yet been checked.

Non-molecular off-frequency scattering characteristics must also be examined. For example, at NOAA we are examining in the laboratory the spectral scattering behaviour of water droplet and ice crystal aerosols. If the droplet and crystal spectra differ, this would indicate promise that constituent identification of pollutant aerosols, such as dusts or sulfuric acid droplets, would be possible using Raman lidar. The need for exploratory research in this area clearly exists.

Current Research Problems: Other Applications

In addition to mapping molecular and perhaps particulate pollutant concentrations in the atmosphere, Raman lidar, or more generally, scattering spectroscopy, has application to water pollution research at the lower limit of the troposphere. We have documented (Schwiesow, 1971) some aspects of the scattering from air to water interfaces. For example the shape of the principle spectral feature of a synthetic sea water interface with air is strongly temperature dependent. This provides a way of sensing thermal water pollution. Similarly, a gasoline film on water exhibits a characteristic backscatter spectrum which may be used to identify this water-borne pollutant. Research on other aspects of air to water interface scattering is a continuing effort.

Current Research Problems: Field Tests

In addition to the laboratory research outlined, field tests are being made (Kobayasi and Inaba, 1970a and 1970b). Instrument tests were described earlier. Field spectroscopic questions such as the effect of sky brightness background noise sources on the pollutant measurement need to be answered. Differential absorption of the pollutant and reference gas Raman returns affects the measurement. While it is estimated that this effect gives less than 25% error, only carefully instrumented field trials can check the estimate. Other unexpected research questions will probably be exposed by extended field tests of Raman lidar systems.

24.2.3 Fluorescence and Resonance Reradiation Lidar

Fluorescence and resonance reradiation are scattering processes which occur in two separated steps (10.2). No obvious examples of these scattering mechanism exist for atoms and molecules at total pressures near 760 torr. For completeness, the possibility of fluorescence and resonance reradiation lidar is considered here.

Example

Bowman et al., (1969) report measurements of stratospheric sodium using a dye laser to excite resonance reradiation from sodium atoms in a low pressure environment. The lidar gives a definite, though noisy, profile of sodium concentration near 90 km. An interference filter provides adequate spectral discrimination for night measurements.

Current Research Problems

To apply fluorescence and resonance reradiation to pollutant mapping, one must find molecules or atoms of interest which exhibit these types of scattering interactions at tropospheric pressure (or in the liquid state at an air to water interface). This search is the first research problem.

Pressure quenching of fluorescence at 760 torr depends on the excited molecule and the quenching molecule. For example, NO_2 in N_2 is partially (not completely) quenched at 760 torr. Measurements to determine the quenching coefficient are in process but not reported. Practically, there is no observational difference between fluorescence and resonance Raman scattering except the pressure dependence of the former. The scattering of gasoline vapor, earlier called resonance Raman, may be fluorescence. Only careful studies of scattering at different pressures (which have not yet been done) will properly classify the experimental observation.

The reason for a scattering survey of pollutant molecules likely to exhibit fluorescence or resonance reradiation effects is that such scattering, if it exists, will result in much stronger scattering than results from the Raman effect. Fluorescent scattering lidar will have much greater range and sensitivity than do existing examples of Raman lidar.

Fluorescence studies of water borne pollutants have been described in general in the latter part of 24.2.2. For liquids, the distinction between resonance Raman and fluorescence is even less clear than it is for gases. Broad absorption bands in liquids mean that neither effect will be as frequency dependent as fluorescence in gases.

24.2.4 Mie Lidar

On-frequency scattering from particulates and aerosols (Mie scattering) is not, strictly speaking, a spectral technique for pollutant sensing. Mie scattering is a highly useful remote sensing tool for the determination of aerosol concentration, which is closely related to the quantum scattering tools discussed earlier. Since atmospheric targets are spectrally indistinguishable in Mie (or even Rayleigh) scattering, Mie lidar does not have the ability to distinguish between pollutants, or even between pollutants and natural aerosols.

In contrast to the striking disadvantage of failure to identify scatterers, Mie lidar has the major advantage of many orders of magnitude stronger scattering signal when compared to Raman, or even fluorescence and resonance reradiation lidars. That is, with Mie one is not sure what one is seeing, but one sees it well. For some purpose where the scatterer is somewhat known, as in emissions from a plume or when a tracer aerosol is injected into an otherwise clear atmosphere, Mie lidar can provide very useful information on pollutant intensity and dispersion paths. Other pollution related applications, such as detecting an inversion layer, are discussed in the previous chapter.

Example

Barrett and Oded (1967), among others, have demonstrated the application of Mie lidar to pollution sensing problems. Because of the intense return and nonselective target characteristics, Mie lidar has been much more widely applied than any other sort of scattering lidar. Collis (1970) as well as Kent and Wright (1970) review some of the extensive research and applications related to the Mie lidar technique.

In general, almost any pulsed laser is suitable as a transmitter. Commercial interference filters provide adequate spectral selection for the on-frequency signal. Typical scattering returns are so strong that daytime operations, with good signal to noise, are often possible.

Current Research Problems

One line of research effort is devoted to examining the scattering intensity of real aerosol distributions as a function of pump frequency and scattering angle. For a bistatic lidar arrangement, where scattering angles other than π rad can be observed, angular information can be obtained. Angle and perhaps frequency dependence information enhances somewhat the information available about the scatterer. However, even for spherical particles, angular dependence of the scattering gives neither size or refractive index of the scatterer. A continuum of size and refractive index pairs satisfy a given angular dependence. The addition of frequency dependent effects helps the situation for an ideal scatterer. For practical cases of nonspherical particles with an inhomogeneous index distribution, identification of the scatterer is even more difficult.

Instrument oriented research is designed to display the lidar information in a highly usable format. Scattering diagrams of intensity as a function of height and horizontal distance, as well as intensity as a function of 3-dimensional coordinates, are examples of a well developed field. Improvement in laser power and beam pattern would be helpful for Mie lidar studies, but is not as crucial as for other types of scattering

lidars. High repetition rate lasers and data collection systems will make possible the study of tropospheric transients.

Considerable recent research has been devoted to the identification of aerosol size distribution and composition by means of multiple measurements. The parameters used to characterize an aerosol sample can include the scattered intensity and scattered polarization as a function of scattering angle and of incident photon frequency.

The depolarization of light backscattered from water clouds has been studied (Liou and Schotland, 1971) in an attempt to eventually differentiate between water droplet and ice crystal clouds. A bistatic lidar (23.1.4) can yield the angular scattering parameters of a sufficiently extended (assumed homogeneous) aerosol sample. Lidar studies of Herman, et al. (1971) measure the four Stokes polarization parameters of the scattering. If one assumes spherical particles with a known index of refraction, Herman, et al. (1971) show how the aerosol size distribution may be inferred. Cohen (1972) and Grassl (1971) have discussed an alternative approach to the determination of an aerosol size distribution by means of a multiple wavelength measurement of Mie scattering intensity. The analysis of this multicolor scattering experiment can proceed via iterative calculations to improve a first estimate of the aerosol size distribution (Grassl, 1971) or via an *a priori* analysis technique (Cohen, 1972). The latter technique is essentially a correlation search for the intensity minima spacing in an intensity vs. scattering parameter α plot (10.4.2) corresponding to a particular size parameter.

Unfortunately, these referenced methods all require some significant assumptions about the sample aerosols. Typically, one must assume that the refractive index of the aerosols is known, that the particles lie in some predetermined size range and that all particles are spherical. Except for water droplet clouds, it has not been demonstrated that these assumptions are valid for real-atmosphere aerosol distributions.

24.3 Summary

Ambient air quality can be measured by long path optical absorption of ultraviolet, visible or infrared radiation. These absorption techniques inherently provide the desired spatial average of pollutant concentration. Many uncertainties in the absorption technique still exist, however. The effect of spectral interference between molecules is unclear. Efficient data analysis techniques are required to analyze the spectra for the many interfering components present. Reference absorption spectra of pollutants at high resolution must still be studied.

Scattering techniques provide a method of three dimensionally mapping pollutant concentrations. However, spectral scattering has only demonstrated a limited range and pollutant concentration sensitivity while on-frequency scattering maps only unidentified aerosols. Further studies of molecular scattering cross sections are required. Practical remote mapping devices depend on improvements in laser transmitters and in the spectral discrimination and sensitivity of receivers.

24.4 References

- Barrett, E. W. and B. Oded (1967), Application of the Lidar to Air Pollution Measurements, *J. Appl. Meteorology* 6, 500-515.
- Bowman, M. R., A. J. Gibson and M. C. W. Sandford (1969), Atmospheric Sodium Measured by a Tuned Laser Radar, *Nature* 221, 456-457.
- Cohen, A. (1972), The Size Spectrum Determination of Spherical Aerosols by Light Scattering. Part I: Method (to be published in) *Proceedings of the Air Pollution, Turbulence and Diffusion Symposium* (Dec. 1971).
- Collis, R. T. M. (1970), Lidar, *Appl. Opt.* 9, 1782-1788.
- Curcio, J. A., L. F. Drummeter, and G. L. Knestrick (1964), An Atlas of the Absorption Spectrum of the Lower Atmosphere from 5400Å to 8520Å, *Appl. Opt.* 3, 1401-1409.
- Derr, V. E. and C. G. Little (1970), A Comparison of Remote Sensing of the Clear Atmosphere by Optical, Radio and Acoustic Radar Techniques, *Appl. Opt.* 9, 1976-1992.

- Fink, W. H. (1971), Calculations of the Ground and Excited State Energies of NO_2 , *J. Chem. Phys.* *54*, 2911-2914.
- Fouche, D. G. and R. K. Change (1971), Relative Raman Cross Section for N_2 , O_2 , CO , CO_2 , SO_2 and H_2S , *Appl. Phys. Letters* *18*, 579-580.
- Grassl, H. (1971), Determination of Aerosol Size Distributions from Spectral Attenuation Measurements, *Appl. Opt.* *10*, 2534-2538.
- Green, A. E. S. (Editor), 1966: *The Middle Ultraviolet, Its Science and Technology*. Wiley and Sons, Inc., New York.
- Griggs, M. (1968), Absorption Coefficients of Ozone in the Ultraviolet and Visible Regions, *J. Chem. Phys.* *49*, 857-859.
- Hanst, P. L. (1970), Infrared Spectroscopy and Infrared Lasers in Air Pollution Research and Monitoring, *Appl. Spectroscopy* *24*, 161-174.
- Herman, B. M., S. R. Browning and J. A. Reagan (1971), Determination of Aerosol Size Distributions from Lidar Measurements, *J. Atmos. Sci.*, *28*, 763-771.
- Hinkley, E. D. and P. L. Kelly (1971), Detection of Air Pollutants with Tunable Diode Lasers, *Science* *171*, 635-639.
- Kay, R. B. (1967), Absorption Spectra Apparatus Using Optical Correlation for the Detection of Trace Amounts of SO_2 , *Appl. Opt.* *6*, 776-778.
- Kent, G. S. and R. W. H. Wright (1970), A Review of Laser Radar Measurements of Atmospheric Properties, *J. Atmos. Terr. Phys.* *32*, 917-943.
- Kobayasi, T. and H. Inaba (1970a), Spectroscopic Detection of SO_2 and CO_2 Molecules in Polluted Atmosphere by Laser-Raman Radar Technique, *Appl. Phys. Letters* *17*, 139-141.
- Kobayasi, T. and H. Inaba (1970b), Laser-Raman Radar for Air Pollution Probe, *Proceedings of the IEEE* *58*, 1568-1571.
- Liou, K., and R. M. Schotland (1971), Multiple Backscattering and Depolarization from Water Clouds for a Pulsed Lidar System, *J. Atmos. Sci.*, *28*, 772-784.
- Long, R. K., 1966: *Atmospheric Absorption and Laser Radiation* (Engineering Experiment Station Bulletin 199). Ohio State University, Columbus, Ohio.
- Mooradian, A., S. R. J. Brueck and F. A. Blum (1970), Continuous Stimulated Spin-flip Raman Scattering in InSb, *Appl. Phys. Letters* *17*, 481-483.
- Newcomb, G. S. and M. M. Millan (1970), Theory, Applications and Results of the Long-Line Correlation Spectrometer, *IEEE Trans. Geosci. Elect.* *GE-8*, 149-157.
- Pierson, R. H., A. N. Fletcher and E. S. Gantz (1956), Catalog of Infrared Spectra for Qualitative Analysis of Gases, *Anal. Chem.* *28*, 1218-1239.
- Randerson, D. (1970), A Comparison of the Spectral Distribution of Solar Radiation in a Polluted and a Clear Air Mass, *J. Air Pollution Control Assn.* *20*, 546-548.
- Schwiesow, R. L. (1971), Raman Scattering from Pollutant Gases and Air-Water Interfaces, submitted to *AIAA J.*

Taylor, J. M. and H. W. Yates (1957), Atmospheric Transmission in the Infrared, *J. Opt. Soc. Am.* 47, 223-226.

Wolfe, W. L. (editor), 1965: *Handbook of Military Infrared Technology*. Office of Naval Research, Department of the Navy, Washington, D.C.

24.A Appendix of Infrared Spectral Data

Figure 24.1, taken from work by Calfee and Derr of NOAA's Wave Propagation Laboratory, shows the regions of the infrared where various molecules that are of interest to remote sensing of the troposphere exhibit significant absorption features. In addition, (F24.1) also indicates something about the state of spectral knowledge for each of these molecules.

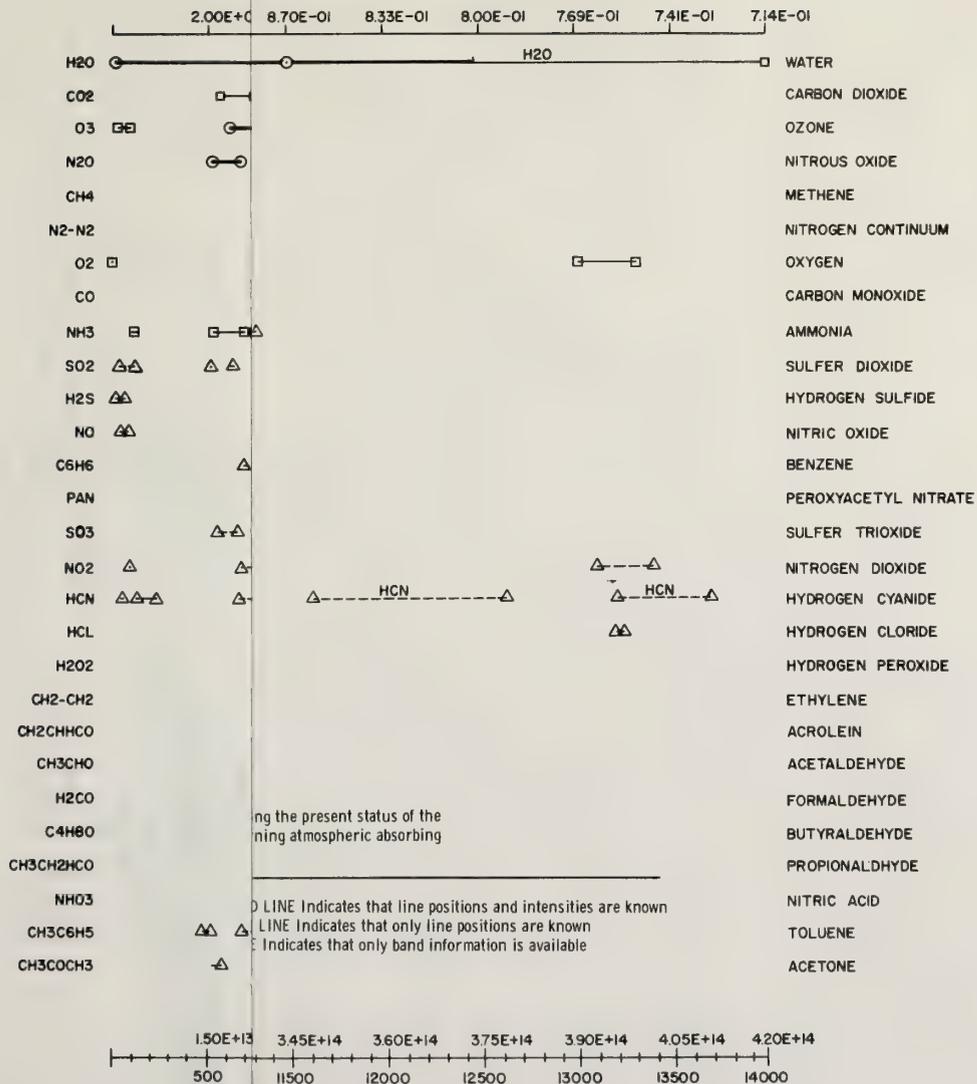
Section 9.7 gives the individual spectra of the molecules N_2 , O_2 , H_2O , CO_2 , O_3 , CH_4 , CO , HCl , H_2 and N_2O . Some of these molecules are pollutants when present in higher than normal concentrations. In addition to the spectra in (9.7), the following spectra give an indication of the infrared transmission of various other pollutant gases. Figure 24.2 shows previously unpublished data on NH_3 and SO_2 at a resolution of 1 cm^{-1} below 1600 cm^{-1} and of 3 cm^{-1} above 1600 cm^{-1} . Appendix (24.A) is a compilation of spectra from various sources. The intent of (24.A) is merely to introduce sources of quantitative infrared spectra of some pollutant gases. From the spectra it is possible to abstract some notion of the complexity and spectral overlap inherent in any multicomponent sample.

24.A References

Pierson, R. H., A. N. Fletcher, E. St. Clair Gantz (1956), Catalog of infrared spectra for qualitative analysis of gases, *Analytical Chemistry*, Vol. 28, No. 8, pp. 1218-1239.

Rhine, P. E., L. D. Tubbs and D. Williams, (1969), Nitric acid vapor above 19 km in the earth's atmosphere, *Applied Optics*, Vol. 8, No. 7, p. 1500.

Stephens, E. R., E. F. Darley, O. C. Taylor and W. E. Scott, (1960), Photo chemical reaction products in air pollution, 25th Midyear Meeting of the American Petroleum Institute's Division of Refining, Detroit, Michigan, p. 6.



F

Taylor, J. M. and H. W. Yates (1957), Atmospheric Transmission in the Infrared, *J. Opt. Soc. Am.* 47, 223-226.

Wolfe, W. L. (editor), 1965: *Handbook of Military Infrared Technology*. Office of Naval Research, Department of the Navy, Washington, D.C.

24.A Appendix of Infrared Spectral Data

Figure 24.1, taken from work by Calfee and Derr of NOAA's Wave Propagation Laboratory, shows the regions of the infrared where various molecules that are of interest to remote sensing of the troposphere exhibit significant absorption features. In addition, (F24.1) also indicates something about the state of spectral knowledge for each of these molecules.

Section 9.7 gives the individual spectra of the molecules N_2 , O_2 , H_2O , CO_2 , O_3 , CH_4 , CO , HCl , H_2 and N_2O . Some of these molecules are pollutants when present in higher than normal concentrations. In addition to the spectra in (9.7), the following spectra give an indication of the infrared transmission of various other pollutant gases. Figure 24.2 shows previously unpublished data on NH_3 and SO_2 at a resolution of 1 cm^{-1} below 1600 cm^{-1} and of 3 cm^{-1} above 1600 cm^{-1} . Appendix (24.A) is a compilation of spectra from various sources. The intent of (24.A) is merely to introduce sources of quantitative infrared spectra of some pollutant gases. From the spectra it is possible to abstract some notion of the complexity and spectral overlap inherent in any multicomponent sample.

24.A References

Pierson, R. H., A. N. Fletcher, E. St. Clair Gantz (1956), Catalog of infrared spectra for qualitative analysis of gases, *Analytical Chemistry*, Vol. 28, No. 8, pp. 1218-1239.

Rhine, P. E., L. D. Tubbs and D. Williams, (1969), Nitric acid vapor above 19 km in the earth's atmosphere, *Applied Optics*, Vol. 8, No. 7, p. 1500.

Stephens, E. R., E. F. Darley, O. C. Taylor and W. E. Scott, (1960), Photo chemical reaction products in air pollution, 25th Midyear Meeting of the American Petroleum Institute's Division of Refining, Detroit, Michigan, p. 6.

Remote Spectral Sensing of Pollutants

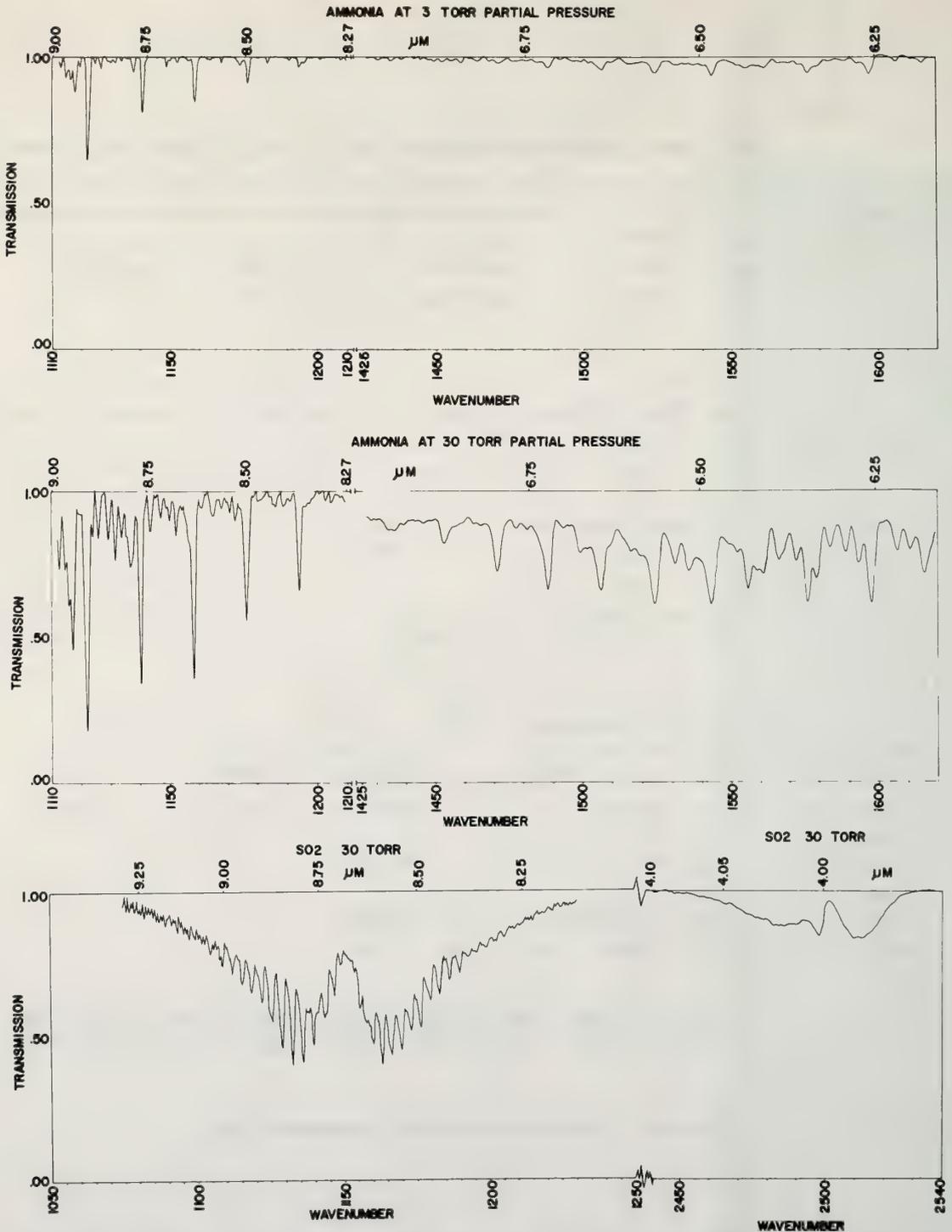
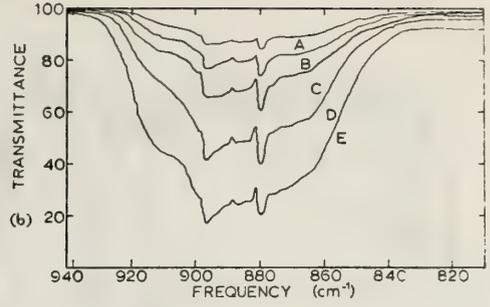
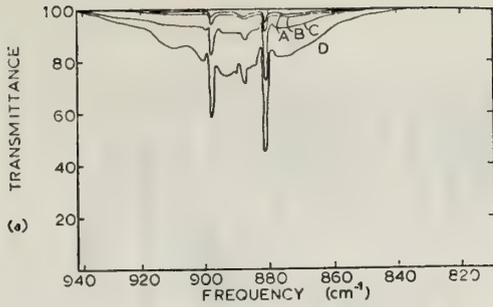
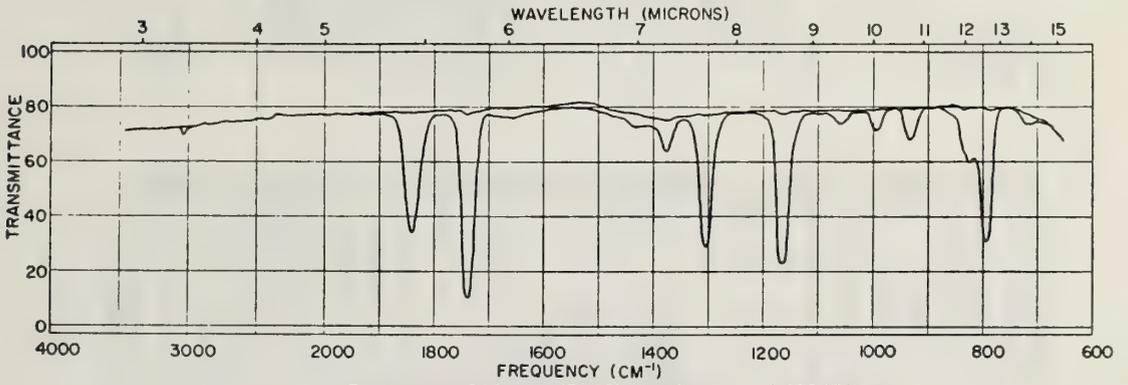


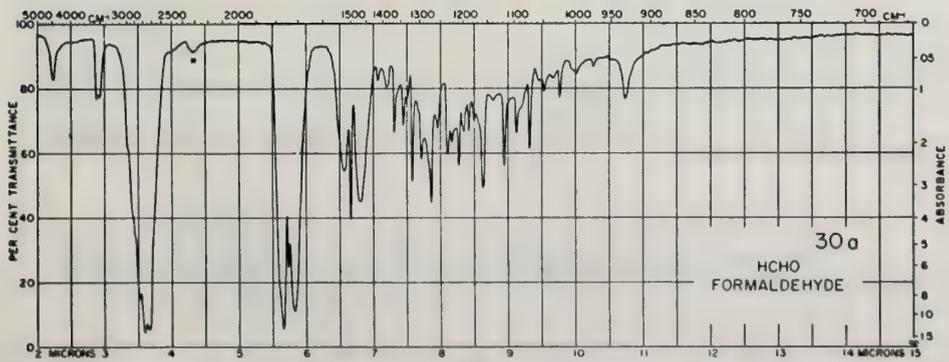
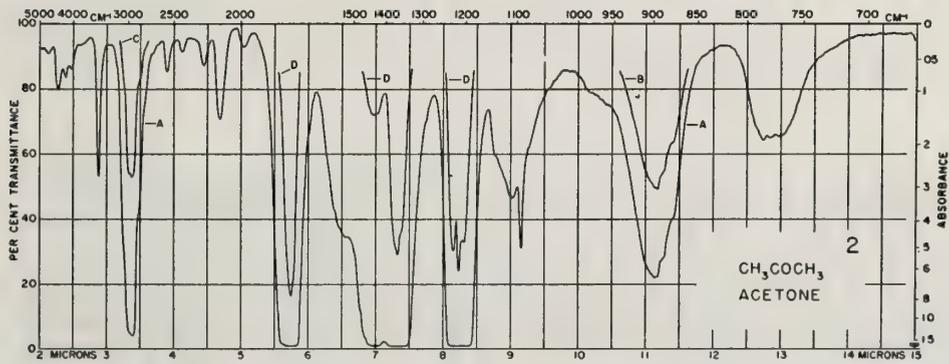
Figure 24.2 Infrared spectra of NH_3 and SO_2 . Resolution is 1 cm^{-1} below 1600 cm^{-1} and 3 cm^{-1} for spectra above 1600 cm^{-1} . All samples were buffered with N_2 to 760 torr total pressure. The concentration conversion for the cell used is: 3 torr sample corresponds to 1.06×10^{18} molecules/ cm^2 , 10 torr sample corresponds to 3.54×10^{18} molecules/ cm^2 , and similarly for other powers of ten.



Nitric Acid, P. E. Rhine, et al (1969).

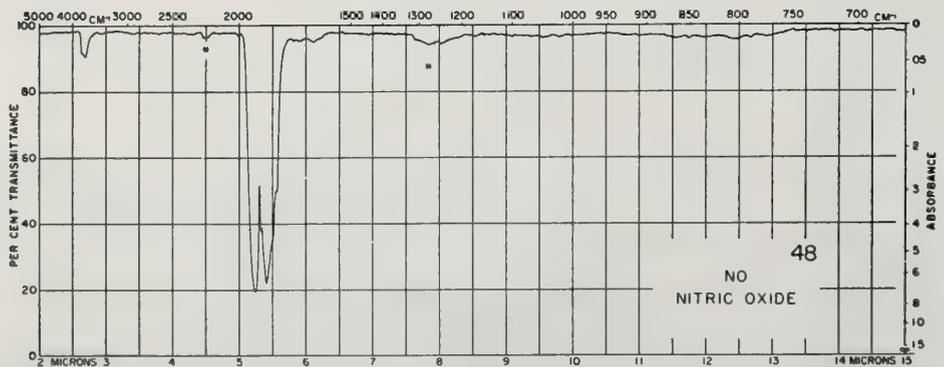
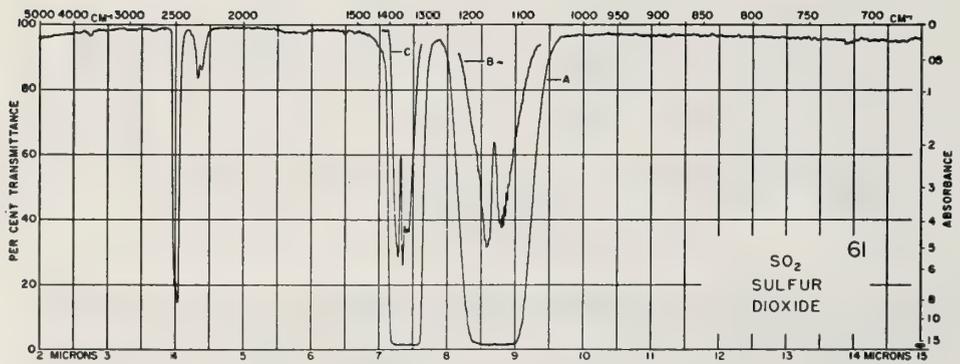
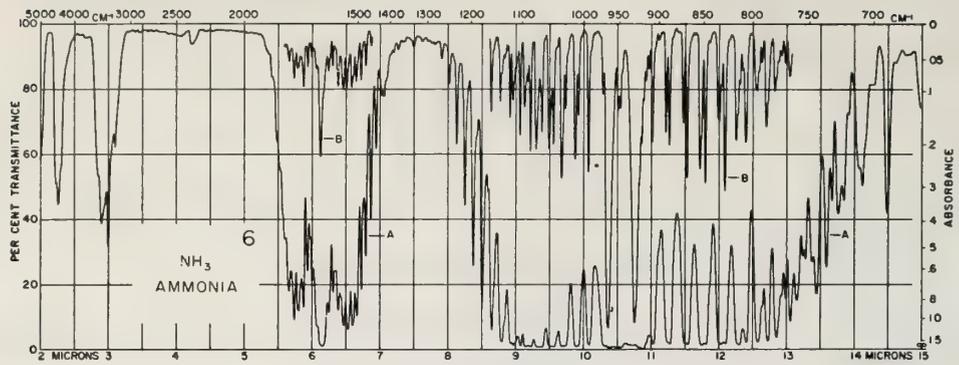


Peroxyacetyl Nitrate (PAN), Stephens, et al (1960).

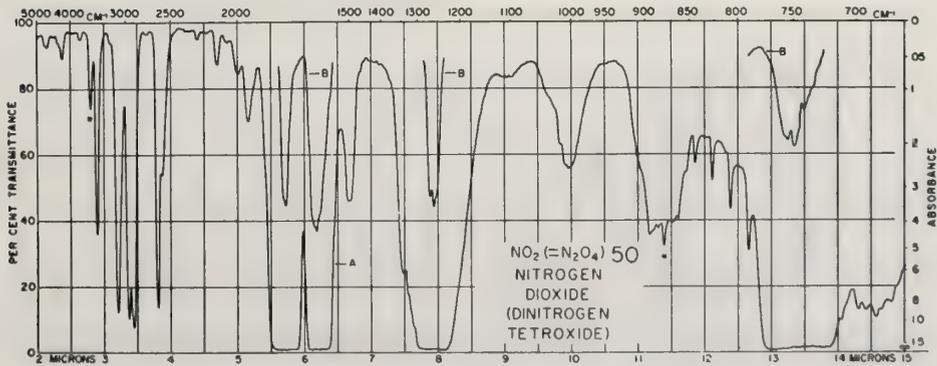
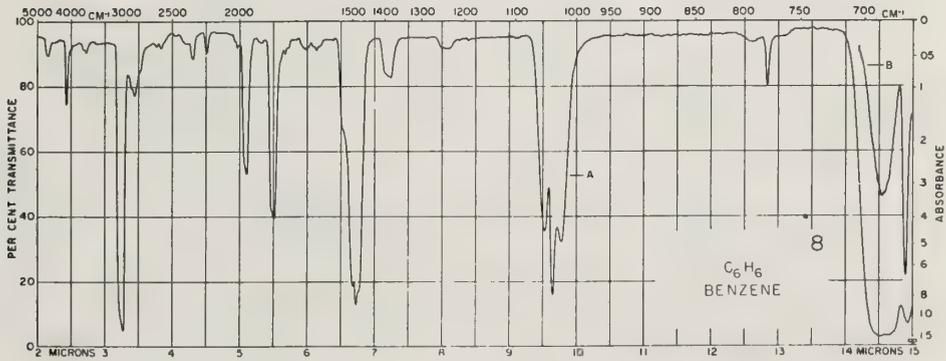
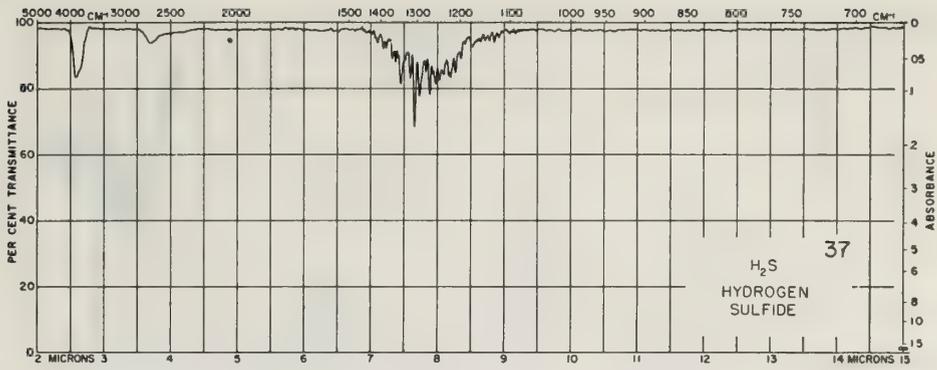


CH₃COCH₃ Acetone, HCHO Formaldehyde, Pierson, et al (1956).

Remote Spectral Sensing of Pollutants

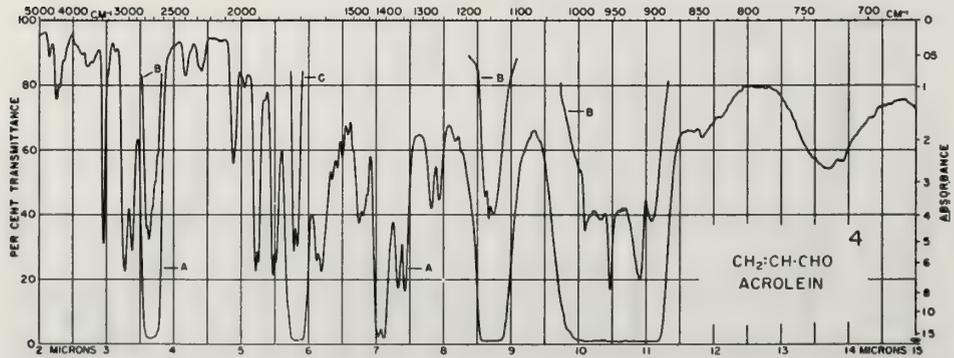
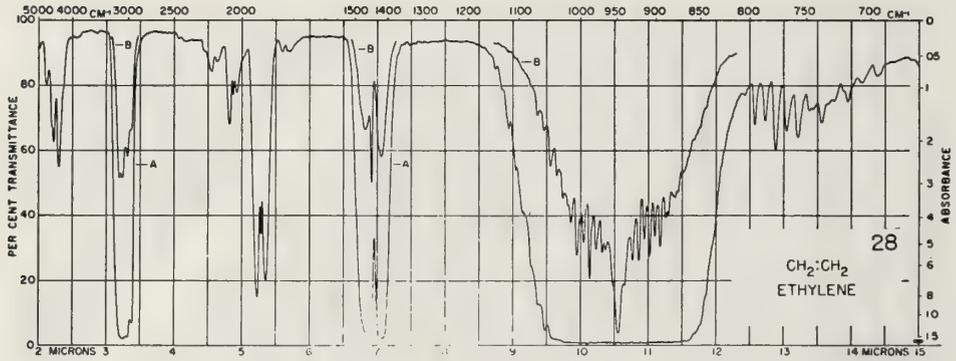
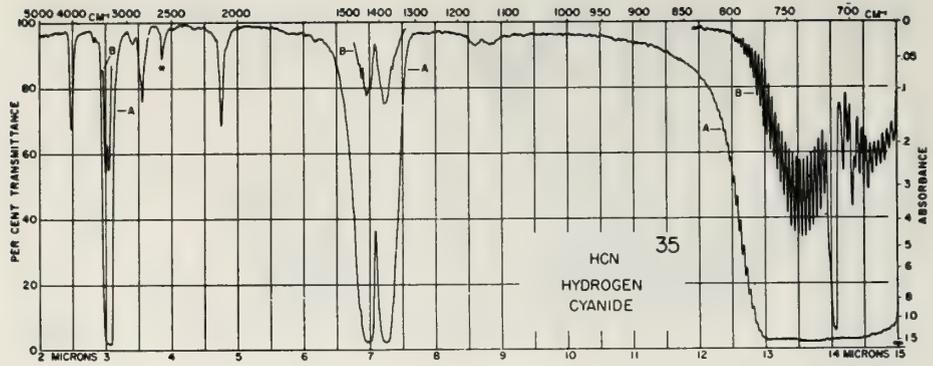


NH_3 Ammonia, SO_2 Sulfur Dioxide, NO Nitric Oxide, Pierson, et al (1956).



H₂S Hydrogen Sulfide, *C₆H₆* Benzene, *NO₂(=N₂O₄)* Nitrogen Dioxide (Dinitrogen Tetroxide), Pierson, et al. (1956).

Remote Spectral Sensing of Pollutants



HCN Hydrogen Cyanide, $\text{CH}_2:\text{CH}_2$ Ethylene, $\text{CH}_2:\text{CH}\cdot\text{CHO}$ Acrolein, Pierson, et al (1956).

List of Symbols

a_{mj}	a matrix element which determines the scattering intensity for the m th wavelength, scattering angle, or polarization, and sub-size range j	n	real part of complex refractive index
A	matrix containing the elements a_{mj}	N	noise photons counted per second
A_t	cross sectional area of transmitted beam	N_j	vector of N_j value
A_r	receiver aperture	N_r	number of photons received per laser pulse
$A_j A_j^*$	system calibration coefficients	N_s	molecular number density for a standard atmosphere
$b_{k,j}$	aerosol scattering coefficient of z_k for wavelength or polarization j	N_t	number of photons transmitted per laser pulse
b_k	measured aerosol backscattering	N_B	photons counted because of background noise
c	velocity of light	N_D	photodetector dark counts per sec.
c_k	aerosol number density ratio	P_t	peak transmitted power
dN	number density of particles with radii $r \rightarrow r+dr$	$P(h)$	pressure at height h
D_m	scattering intensity recorded at the m th observation	q	a parameter used in the definition of a specific polytropic atmosphere
D	vector of $M D_m$ values	Q	total aerosol mass
e	charge of electron	r	particle radius
E	partial pressure of water vapor	R	universal gas constant
$f(\bar{r}, m)$	aerosol size distribution function	R, R_i	range
g	acceleration of gravity	R_k	molecular number density ratio
G	photodetector gain	T	averaging time
$g(\theta, \phi, m, \alpha)$	angular dependent scattering function for Mie scatterer	$T(h)$	temperature of height h
h	Planck's constant	V	visibility distance
h	height	x	distance
$I, (I_0)$	radiation intensity (incident)	z, z_k	
I_T	total scattering intensity	Z	laser pulse repetition rate
I_R	Rayleigh scattering intensity	α	particle size parameter
I_M	Mie (aerosol) scattering intensity	$\alpha_{\max(\min)}$	size parameter for which the scattering intensity has a maximum (minimum) value
I_D	photodetector dark current	β, β_i	extinction coefficient (for i -type extinction)
k	imaginary part of complex refractive index	β_A^i	extinction coefficient due to resonant absorption
$k_M(m, \alpha)$	ratio of scattering cross section to geometric cross section for Mie scatterer	δ	depolarization factor
L	length of scattering region	ΔP	pressure change
m	complex refractive index	ΔT	temperature change
m_s	refractive index	$\Delta \lambda$	spectral bandwidth of the receiver
M	molecular weight	ζ	laser-radar operating angle from the zenith

η_q	quantum efficiency of photodetector
η_r	efficiency of receiver optics
η_t	efficiency of transmitter optics
θ	scattering angle
λ	wavelength
ν	Junge size distribution parameter
ν_t	transmitted wavelength
ρ	number density of particles
ρ_k	
$\rho(h)$	density of air at height h
$\sigma_i(\sigma_i)$	scattering cross section of single scatterer (for i-type scattering)
$\sigma(\theta, \phi)$	differential cross section
$\sigma(\pi)$	backscattering cross section
$\sigma_{M(m, \alpha)}$	scattering cross section for Mie scatterer
σ_R	scattering cross section for Rayleigh scatterer
τ_t	pulse width of transmitter
τ_λ	turbidity for a specific wavelength λ
ϕ	angle of polarization
$\Psi_i(\Psi_i)$	angular scattering function (for i-type scattering)
Ω	solid angle
Ω_r	receiver solid angle
Ω_t	transmitter solid angle

Chapter 25 REMOTE SENSING BY OPTICAL LINE-OF-SIGHT PROPAGATION

Robert S. Lawrence

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

Turbulence in the clear atmosphere causes fluctuations in both phase and amplitude of an optical wave. The fluctuations can be used to measure remotely the strength of the turbulence. In addition, motion of the turbulent eddies produces effects that permit measurement of the transverse component of wind velocity including, for space-to-earth paths, winds aloft. The average density of the air along a light beam affects the phase of the optical wave and of any microwave modulation. These effects permit the measurement of average air temperature along the beam, and through beam curvature, the vertical temperature gradient.

25.1 Review of Selected Aspects of Propagation through Turbulence

The scintillations and phase fluctuations imposed upon an optical wave by propagation through a turbulent medium have been described in Chapter 11. The use of those fluctuations for remotely sensing the properties of the clear atmosphere requires appreciation of several points that we review in this section.

25.1.1 Refractive-index Variations in the Turbulent Atmosphere

Whenever we refer to “turbulence” in this chapter, we shall mean refractive-index (or temperature) turbulence rather than the velocity turbulence commonly measured with hot-wire probes. The distinction is important because, when the atmosphere is in neutral thermal stability, i.e., when the temperature lapse rate is adiabatic, strong mechanical turbulence may exist with little or no optical effect.

The refractivity of air at optical frequencies is, to a good approximation, simply

$$N \approx 79 \frac{P}{T} \quad (25:1)$$

Here, P is the atmospheric pressure in mb, and T is the temperature in K. The deviation of the refractive index from unity in parts per million is $N = (n-1)10^6$. The value of N is roughly 290 at sea level. The approximation (25:1) neglects the variation of N with wavelength and the minor effect of the variation of air density caused by the presence of water vapor. The wavelength variation is about 10 percent over the optical range; the humidity variation is less than 1 percent. These matters have been reviewed in detail by Owens (1967).

The most widely accepted description of the structure of turbulence appeared when Kolmogorov (1941) considered the structure function between two components of velocity; call them α, β where $\alpha, \beta = x, y, z$,

$$D_{\alpha\beta}(r) = |v_{\alpha}(\xi) - v_{\beta}(\xi+r)|^2 \quad (25:2)$$

for a span, r , along coordinate ξ . Purely from dimensional analysis, he found $D_{\alpha\beta}$ to be proportional to the two-thirds power of the separation, i.e.,

$$D_{\alpha\beta}(r) = C_{\alpha\beta} |r|^{2/3} \quad (25:3)$$

where $C_{\alpha\beta}$ is a parameter depending on the components and the energy involved.

An advance of direct interest to optical propagation came when Obukhov (1949) and Corrsin (1951) used similar dimensional analysis to consider the temperature fluctuations in turbulence. The structure function of this scalar parameter turns out also to obey the two-thirds law, i.e.,

$$D_T(r) = C_T^2 |r|^{2/3} \quad (25:4)$$

where C_T^2 , the "temperature structure parameter," depends on the energy involved in the turbulence.

The structure of the turbulence has been treated in detail in a number of places, e.g., Lumley and Panofsky (1964).

The small scale sizes and short lifetime of the temperature fluctuations ensure that they are adiabatic so that

$$\frac{\delta P}{P} = \frac{\nu}{\nu-1} \frac{\delta T}{T} \quad (25:5)$$

where the ratio of specific heats, $\nu = c_p/c_v \approx 1.4$ for air. Differentiating (25:1) and eliminating δP , we see that the refractive-index fluctuations are proportional to the temperature fluctuations, viz.

$$\delta N = \frac{79}{\nu-1} \frac{P}{T^2} \delta T \quad (25:6)$$

Accordingly, from (25:4), the refractive-index structure function,

$$D_N(r) = C_N^2 |r|^{2/3} \quad (25:7)$$

where C_N^2 is the refractive-index structure parameter. This parameter, C_N^2 , decreases rapidly with height above the ground. The variation has been measured by Tsvang (1963) and discussed by Hufnagel (1966) and by Fried (1966).

25.1.2 The Production of Scintillations by Turbulence

Let us consider the behavior of a light wave as it travels outward from a point source through the turbulent atmosphere (F25.1). The wave front is initially spherical, as at A. Upon passing through irregularities to reach position B, it becomes distorted. Since absorption and wide-angle scattering are negligible, the energy density of the wave front B is still uniform and equal to its free-space value. Thus an ordinary square-law detector located at B would be unaffected by the irregularities and incapable of measuring them. The irregularities in the wave front can, of course, be measured by a phase-sensitive detector such as an interferometer.

As the wave progresses from B toward C, the various portions of the distorted wave front travel in slightly different directions and eventually begin to interfere. The interference is equivalent to a redistribution of energy in the wave and causes intensity fluctuations (scintillations) which can be detected by a square-law detector. On the way from B to C, the wave front passes through additional refractive-index irregularities and so suffers additional phase perturbations. These new irregularities are, however, relatively ineffective in producing intensity fluctuations.

Let us examine the criteria that determine which of the turbulence irregularities along a line of sight are most effective in producing intensity fluctuations. In (F25.2) consider an irregularity of diameter r at an arbitrary point, A, on the line of sight between the source, S, and the receiver, R. That irregularity can be fully effective in producing intensity variations only if the extreme ray paths, SAR and SBR, involving it differ in length by at least half a wavelength, i.e., the irregularity must be at least equal in size to the first zone of a

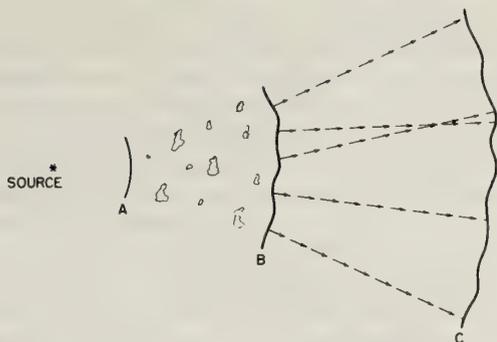


Figure 25.1 Schematic diagram of the propagation of a spherical light wave through a turbulent atmosphere. Phase fluctuations at B develop into phase and intensity fluctuations at C.

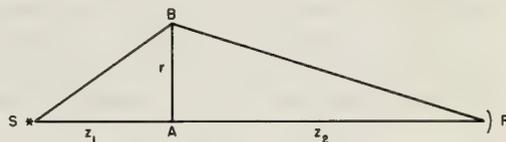


Figure 25.2 The geometry involved in determining the irregularity size most effective in producing scintillations.

Fresnel zone plate situated at A. This minimum effective size is, in fact, the optimum size for the irregularity. Larger irregularities at the same point are rendered ineffective by the smaller ones just as a lens is made ineffective by a ground glass surface. While it is true that smaller irregularities produce intensity fluctuations at points closer than R, and that these fluctuations persist in a modified form until the wave reaches R, such smaller irregularities are relatively ineffective because of the steep increase in the spectrum of atmospheric turbulence with irregularity size.

Working out the geometry of (F25.2), we find that the diameter of the most effective irregularity is $r \approx \sqrt{q\lambda}$, where λ is the wavelength and $q = z_1 z_2 / (z_1 + z_2)$ depends upon the position of A. This diameter is plotted in (F25.3) for a wavelength of 633 nm and a path length of 10 km.

If we assume, for the moment, that the turbulence is uniformly distributed along the path and has a Kolmogorov spectrum, it is clear that the mean-square fluctuation of refractive index attributable to irregularities of optimum size varies systematically along the path. There is, therefore, a weighting function that expresses the relative effectiveness of turbulence in producing intensity fluctuations as a function of position along the path. From (F25.3) it is clear that this weighting function must reach a maximum at the midpoint of the path and must drop symmetrically to zero at the ends. The expression comes directly from the integrals given in 11.4. It is

$$E = (z_1 z_2)^{5/6} \tag{25:8}$$

Figure (25.4) compares this function of position along the path with the best-fitting parabola. In summary, the relative effectiveness of a uniformly turbulent atmosphere in producing intensity scintillations is approximately a parabolic function of position along the path, being a maximum at the midpoint and zero at the ends.

Next, let us examine the scale sizes of the intensity patterns at the receiver due to the optimum-sized refractive-index irregularities located at various points along the path. Referring to (F25.5) and recalling that the diameter r of the optimum-sized irregularity at A was such that SBR exceeded SAR by a half wavelength, we can see that the radius p of the pattern at the receiver is determined by the requirement that SBR' must equal SAR'. Then, when destructive interference is present at R, constructive interference will occur at R'. Working out the geometry, we find that the pattern is larger than the turbulent irregularity by the factor $p/r = \frac{1}{2}(1 + z_2/z_1)$. The pattern radius p is shown in (F25.3) for a 10 km path.

We have seen that the diffraction process that produces intensity fluctuations in the light wave selects only certain optimum sizes from the broad spectrum of irregularities available in Kolmogorov turbulence. The optimum size selected depends on the position along the path, and each position produces a predominant, and

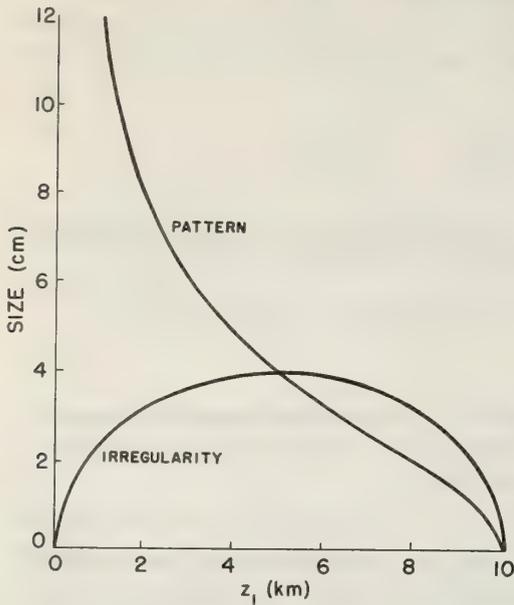


Figure 25.3 The diameter of the most effective irregularity at various points along a 10 km path, and the resulting pattern size.

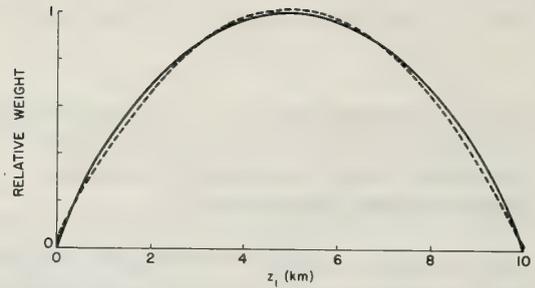


Figure 25.4 The relative effectiveness of Kolmogorov turbulence at various points along a 10 km path in producing intensity fluctuations. The dashed curve is the best-fit parabolic approximation.

unique, pattern size at the receiver. When the weighting function shown in (F25.4) is combined with the pattern-size function of (F25.3), there results the composite spectrum of sizes observable in the intensity pattern. Notice that, for turbulence distributed uniformly along the path, this composite spectrum of sizes depends only on the wavelength and the path length; it is not indicative of any predominant size of turbulent eddies in the atmosphere. Expressions for this spectrum and for its Fourier transform, the log-intensity covariance function, have been given in Chapter 11. The theoretical covariance function is compared in (F25.6) with observations.

25.1.3 Additional Effects of Turbulence

The phase fluctuations caused by turbulence eventually produce scintillation but, even after this diffraction process is completed, the phase fluctuations persist in the wavefront. They are not useful in themselves for remote-sensing purposes because phase-sensitive optical detectors, such as interferometers, tend to be expensive and difficult to use. However, when the source of the optical wave is a laser, the energy can be confined to a beam having, in the absence of atmospheric turbulence, a divergence angle of only λ/d , where d is the diameter of the transmitter aperture. Turbulence-induced phase fluctuations increase the divergence angle, and they do so with a weighting function that is very different from that of (25:8). Rigorous analysis of beam spreading is surprisingly difficult, though an approximate geometrical solution has been presented by Beckmann (1965). Simple arguments similar to those used by Beckmann indicate that the weighting function for beam diameter is linear. It has a maximum at the laser and drops to zero at the receiver.

25.2 Remote Measurement of Turbulence

Beam spreading and scintillations are both produced by turbulence along the path. Therefore, with certain limitations, observation of these phenomena can be used to infer the presence and strength of refractive-index turbulence. We must emphasize, again, the distinction between refractive-index turbulence, caused by thermal irregularities, and dynamic turbulence involving velocity fluctuations.

An important limitation to the use of optical scintillations for remote measurement of turbulence is the saturation effect described in Chapter 11. Beyond a certain point, the log-intensity variance of the light (i.e., the strength of scintillations) no longer increases with an increase in C_N^2 (i.e., strength of turbulence) integrated over the path. For strong, naturally occurring turbulence within a meter or two of the ground, saturation arises for paths as short as 0.5 or 1 km. At greater heights, much longer paths are free from saturation. Presumably, beam spread is not affected by saturation and so continues to increase indefinitely. However, neither adequate theory nor observations exist to confirm this presumption. Nevertheless, Hogge and Visinsky (1971) have used beam spread with apparent success to measure the strength of turbulence in the exhaust of a jet engine.

The "inner scale" of turbulence (i.e., the irregularity size below which viscous damping steepens the spectrum) presents a limitation to the use of scintillations for the measurement of turbulence on short paths. Under normal conditions, the inner scale is a few millimeters and so equals or exceeds the size of the most important irregularity for producing scintillations on paths shorter than a few hundred meters. Short paths, then, suffer less scintillation than might otherwise be expected.

25.2.1 Uniform Path, Kolmogorov Spectrum

With the exceptions noted above, the log-intensity variance of an optical wave depends upon path length and strength of turbulence as follows:

$$\sigma_{\ln I}^2 = 0.50 (2\pi/\lambda)^{7/6} L^{11/6} C_N^2 \quad (25:9)$$

Here, the intensity (more properly, the irradiance) is assumed to be measured with a receiving aperture small compared to $\sqrt{\lambda L}$, and the light is assumed to originate from a point source at distance L from the receiver. If, instead, the source is a plane wave entering the turbulence at a distance L from the receiver, the numerical coefficient is increased by a factor of about 2.5.

25.2.2 Non-uniform Path, Kolmogorov Spectrum

We pointed out in 25.1 that the turbulence along the path must be weighted linearly to account for its effect upon beamwidth, while it has a symmetrical, nearly parabolic, weight in its effect upon intensity fluctuations. If the turbulence is uniformly distributed along the path there will be a fixed relationship between beamwidth and log-intensity variance, at least until the turbulence becomes sufficiently strong to saturate the scintillations. If the turbulence is not uniformly distributed, this relationship will not hold in general. Thus, the simultaneous observation of beam spread and log-intensity variance can be used to check the uniformity of the turbulence along a path. There is not, however, enough information in such a pair of measurements to invert the integral and learn any appreciable details of how the turbulence is distributed along the path.

The covariance function of log-intensity scintillations was shown in (F25.6). The discussion in 25.1.2 showed how this function (or its Fourier transform, the spatial spectrum of the fluctuations) results from the combined effects of the turbulence along the path. The turbulence at each point contributes to the spectrum a limited range of pattern sizes distributed closely around the sizes shown, for example, in (F25.3). Accordingly, if turbulence were absent over a portion of the path the spectrum of intensity fluctuations would be distorted in a characteristic way.

Thus, we have a possible method for determining the distribution of turbulence along the path. Like so many remote-sensing techniques, it involves the inversion of an integral to obtain the desired answer. The measurements, themselves, are statistical in nature and so contain a random element, i.e., they are inherently noisy. The noise cannot be reduced indefinitely by extending the observation period because the open atmosphere is notorious for yielding non-stationary time series. Although guesses can be made at the present time, it remains for experiment to disclose how accurate the method can be. Such experiments are in progress at Boulder. Our expectation is that the turbulence distribution can be represented in terms of a third or fourth order polynomial, but probably not in much more detail. Inclusion of beam-spread measurements might yield a slight improvement.

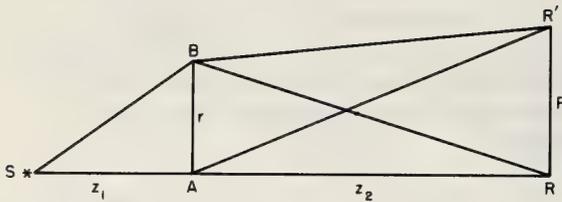


Figure 25.5 The geometry involved in determining the pattern size produced by the most effective irregularities.

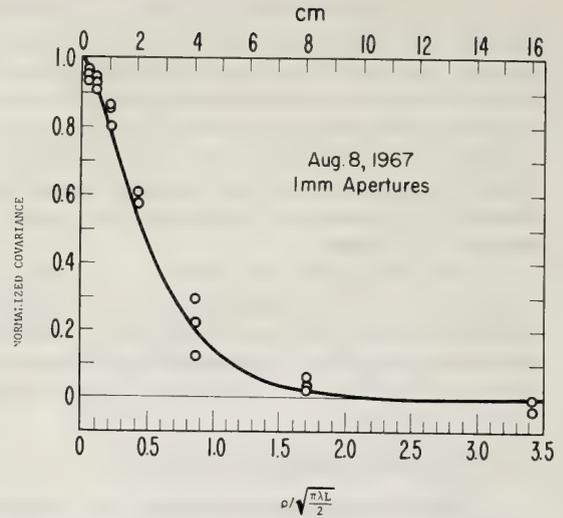


Figure 25.6 The theoretical normalized covariance of log-intensity fluctuations, compared with observations made over a path length $L = 5.5$ km. The detector spacing is ρ .

25.2.3 The Effect of Spectrum Variations

While the Kolmogorov spectrum is widely used to represent the turbulent atmosphere, and while, on the average, it seems to agree well with measurements in the inertial subrange well above the ground, individual measurements, extending over periods of one or more hours, often show significant deviations from the ideal spectrum. The effect on scintillations of these spectrum variations is similar to the effect of non-uniformity of turbulence along the path. Thus, scintillation remote-sensing techniques intended to measure either of these atmospheric parameters must, in fact, determine them simultaneously.

An additional complication arises for path lengths shorter than about 100 m as the size of the most effective irregularity approximates the inner scale of turbulence. Gray and Waterman (1970) have used this fact to measure, with short-path scintillations, the size of the inner scale and the slope of the spectrum in the dissipation subrange.

Strohbehn (1970a) has calculated a number of scintillation covariance functions corresponding to various models of the turbulent spectrum, and he discusses the sensitivity of scintillation covariance measurements for the determination of spectral parameters of turbulence, particularly the slope of the spectrum in the inertial subrange. While his work is useful in showing the magnitude and nature of the effects, it cannot be applied directly to remote sensing because it ignores the highly coupled effects of non-uniformity of turbulence along the path.

25.2.4 The Effect of Aperture Size on the Weighting Function

We saw in (F25.3) that turbulence near the receiver produces fine-scale scintillations, while near the light source it produces large patterns. A very small receiver will be sensitive to all these scales as the scintillation pattern drifts by and so will measure turbulence over the entire path. As the size of the receiving aperture increases, the receiver becomes less sensitive to the fine-scale structures in the pattern because both bright and dark portions of the wave enter the aperture simultaneously and their variations tend to cancel. Such aperture averaging modifies the weighting function along the path, larger receiving apertures being relatively more sensitive to turbulence farther from the receiver. In principle, then, we have another method of probing the distribution of turbulence along the path. The disclaimers mentioned at the end of 25.2.2 still apply, and it should be noted that the aperture-averaging method is not independent of the scintillation-covariance method. They are really slightly different ways of looking at the same phenomenon.

A related, but more complicated, situation has been considered by Carlson (1969) in which he shows that the use of sizeable transmitting apertures and collimated or focused laser beams produces sharper weighting functions that should be correspondingly more powerful for the determination of turbulence profiles and spectral form. This method has apparently not yet been tried in practice and it will, of course, be limited to conditions of moderate turbulence where the concept of focused beams is meaningful.

25.2.5 Practical Applications

Here, we try to assess the probable success of scintillation-sensing of turbulence in typical practical situations, keeping in mind the fact that the turbulence being measured is temperature turbulence, not dynamic turbulence. We can expect in each case, as mentioned in 25.2.2, that the distribution of the turbulence along the line of sight will be determined only roughly, perhaps in terms of a best-fitting 3-parameter function.

The wake turbulence produced by large aircraft presents a serious safety problem for smaller aircraft in the vicinity of airports, and it is important for traffic controllers or pilots to have a simple, immediate indication that turbulence from preceding traffic has cleared the immediate vicinity of the runway. For the region near the ground where both ends of the line of sight are accessible, a scintillation monitor can clearly give the desired information, assuming that it is not necessary to pinpoint the exact location along the path of the turbulent vortex.

High-level, clear-air turbulence (CAT) is also important to aviation interests. Many methods have been suggested for sensing the presence of CAT before an airplane enters it. Airborne lidar and infrared radiometers have been flown extensively, but with little or no success (Veazey, 1970). Peskoff (1968) and Fried (1969) have suggested that ground-based measurements of scintillation of starlight could, in principle, indicate the presence of CAT, though Strohhahn (1970b) pointed out that small deviations from the normal Kolmogorov spectrum would affect the covariance function of scintillations in much the same way as would a sizeable CAT layer. Peskoff (1971) disagrees, stating that the extreme tail of the covariance function would be changed in a characteristic way by a CAT layer. The truth probably lies somewhere between the two views, since it is unlikely that the tails of the covariance curve can be measured accurately enough in a real, non-stationary atmosphere to provide unambiguous indication of CAT. The stellar-scintillation method has not been tried, so its possibilities remain uncertain.

A satellite-borne laser would provide a more suitable light source for scintillation detection of CAT. The greater spectral irradiance would permit covariance measurements to be made more quickly, thus helping to eliminate the problems of atmospheric non-stationarity. Also, laser measurements would be feasible in the presence of sunlight. When, in the next few years, lasers become available on geostationary satellites, ground-based detection of CAT will surely be attempted.

25.3 Remote Measurement of Wind

We discussed in 25.1.2 the spatial spectrum of the scintillation pattern and we showed its Fourier transform, the spatial covariance function, in (F25.6). This curve, of course, is only the cross-section in the space direction of the complete time-space covariance function of the scintillations. The complete function depends upon the distribution of turbulence along the path and the distribution of winds, primarily those components normal to the path of the light beam, that move the turbulent eddies through the beam. In principle, a measurement of the complete time-space covariance function provides the information needed to invert the integral equation and determine the distribution of turbulence and of transverse wind along the propagation path. Fried (1969) derived the pertinent equations for the cases where the scintillations have not become saturated. As in all integral-inversion, remote-sensing situations, questions must be raised concerning the detail that can be expected in the answer in light of the unavoidable errors in measuring covariance with the real, non-stationary atmosphere. In this case, other practical difficulties arise. First, the measurement of the complete covariance function would require a large number of spaced sensors and their signals would have to be compared pairwise, each with a large number of different time lags. Second, the integral equation is non-linear and apparently cannot be inverted analytically in a useful form. Because of these difficulties, all attempts to measure wind have involved approximations and simplifications of one kind or another.

25.3.1 The Temporal Power-spectrum Method.

Perhaps the simplest way to estimate the wind velocity from observations of scintillations is to measure the temporal power spectrum of the scintillations with a single detector. Assuming that the turbulence is unchanging and is simply being transported across the line of sight by the mean wind, the temporal spectrum is a direct replica of the spatial spectrum and is related to it by the wind velocity as a simple proportionality constant. The assumption involved is "Taylor's hypothesis," and it permits the peak frequency in the temporal spectrum to be interpreted directly as the frequency with which Fresnel-zone sized irregularities are carried past the line of sight by the wind.

The theoretical shape of the temporal frequency spectrum has been derived by Clifford (1971). The frequency-spectrum method of estimating winds has been checked by Ryznar (1965) and by Höhn (1966), though some of their results may have been confused by aperture averaging. Mandics (1971) demonstrated that the method works well with acoustic waves over a 68 m path. Lee and Harp (1969) briefly discuss and illustrate the same method, described in terms of the width of the autocovariance function.

25.3.2 Use of the Time Delay between Spaced Detectors.

If the scintillation pattern could be assumed not to change in time, but simply to be moved bodily by the wind, it would be a straightforward process to measure the velocity of the pattern and, therefore, of the wind normal to the propagation direction. Two detectors aligned parallel to the wind vector would display identical fluctuations, one lagging the other by the time interval required for the wind to move the scintillation pattern between the detectors. In the more general case where the direction of the wind vector (still assumed normal to the propagation path) is unknown, three sensors are sufficient to define wind speed and direction. This method of similar fades was first used with reflected radio waves to measure ionospheric winds (Mitra, 1949). The method was soon supplanted by a correlation technique (Briggs and Spencer, 1954) that is less affected by random pattern changes that partially destroy the similarity of fades observed at the two detectors. The correlation technique continues to be used to measure ionospheric winds (Fedor, 1967) and has recently been applied to radio-star scintillations to measure the speed of the interplanetary solar wind (Cohen et al. 1967).

The techniques just mentioned are essentially measurements of the time delay required to maximize the time-lagged cross-correlation function, illustrated schematically in (F25.7), of the scintillations observed at the two detectors, corrections being applied for the fact that the correlation is imperfect because of pattern decay. They work well as long as the pattern decay is slight during the time it takes the pattern to move between the detectors. The applications mentioned above meet this criterion: the ionospheric case because the scintillations are all impressed at a single portion of the path (the point of reflection), and the radio-star case because the solar wind is reasonably localized (to the point nearest the sun) along the path and also because the wave is originally a plane rather than a diverging wave. Line-of-sight paths through the troposphere are quite different. We have already seen that, at each point along the path, a specific irregularity size is most effective in producing scintillations, and it produces a unique pattern size. The small-scale irregularities that are effective near the transmitter produce large patterns, while irregularities of the same size near the receiver produce fine-scale patterns. Each of these patterns moves through its own characteristic dimension during the time required for the wind to carry the irregularity through its own diameter. Thus, it is a general property of the scintillation pattern of a diverging wave passing through an extended turbulent medium that the large-scale components of the pattern move more rapidly than do the fine-scale details. As a result, the pattern decays rapidly even in the absence of random motion or eddy decay in the turbulent medium. This means that the correlation sketched in (F25.7) will be progressively reduced as the time delay increases from zero. The effect is not only to lower the curve but to displace the peak toward the vertical axis. The time delay to the peak is systematically reduced and the deduced wind speed is biased toward high values.

25.3.3 The Correlation-slope Method

Measurement of the slope of the correlation function at zero time lag avoids the problems of pattern decay, since it involves comparing features at nearly the same instant (see F25.7). Shen (1970) showed that this slope is proportional to the mean transverse-wind speed. Lawrence, Ochs, and Clifford (1972) demonstrated that it produces a correct result for all profile realizations with the same average wind. It is also relatively insensitive to reasonable variations in the strength of turbulence along the path. As distinct from the time-delay method, the correlation-slope method is unaffected by the presence of a component of transverse wind perpendicular to the line joining the detectors.

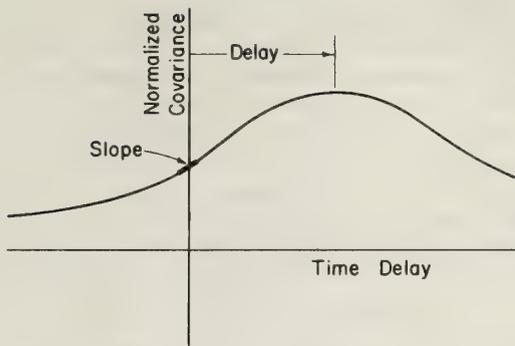


Figure 25.7 A schematic illustration of two methods of using the cross-correlation function of scintillations observed at spaced detectors to measure wind and speed.

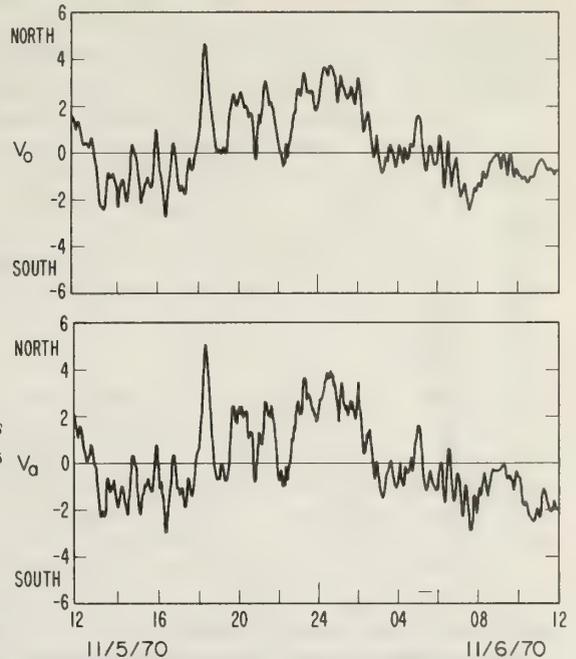


Figure 25.8 A 24-hour comparison between the optically measured wind (upper curve), using the correlation-slope method, and the mean of seven propeller-type anemometers spaced evenly along the 1 km test path.

When evaluating these methods, or any method amenable to simple, on-line calculation and display, it is important to realize that they use only a portion of the information contained in the entire space-time correlation function and must necessarily involve assumptions concerning the over-all form of that function. The extent to which these assumptions limit the usefulness of a particular method can be indicated by computer simulations using a wide range of wind and turbulence profiles, but is finally determined only through extensive quantitative comparison with direct measurements on real paths in the open atmosphere. One very important reason for the need for actual outdoor tests is our present lack of knowledge of the effect of saturation on the correlation function. Until an adequate saturation theory is developed, it will not be possible to make a complete theoretical analysis of the performance of scintillation wind sensors operating over long, low-level paths where saturation is likely to occur.

We have tested the correlation-slope method over a one-kilometer path 1.6 m above flat ground, using seven propeller-type anemometers equally spaced along the path for comparison. Figure (25.8) shows a 24-hour sample of such a comparison, indicating that the correlation-slope method does, indeed, produce very acceptable results for many practical purposes. The apparatus used to produce this record has been described by Lawrence et al. (1972) and in more detail by Ochs (1972).

25.3.4 The Path Weighting Function

We have seen that the center of the propagation path is more effective than the ends in producing scintillations and that the various scale sizes in the scintillation pattern are produced primarily at specific portions of the path. It follows that the slope of the correlation function at zero lag will be most strongly affected by winds near the center of the path, and that this path weighting function can be modified by selective spatial filtering of the received scintillation pattern. For example, large receiving apertures placed before each of the two detectors will tend to average out those scintillations with scale sizes much smaller than the apertures, leaving the large scales in the scintillation pattern to predominate. This aperture-averaging effect will cause the winds near the transmitter end of the path to be relatively more effective than those near the receivers. Similarly, spacing the detectors widely will accentuate the large scales in the pattern because the fine-scale structure will decay during the time required for it to drift between the detectors.

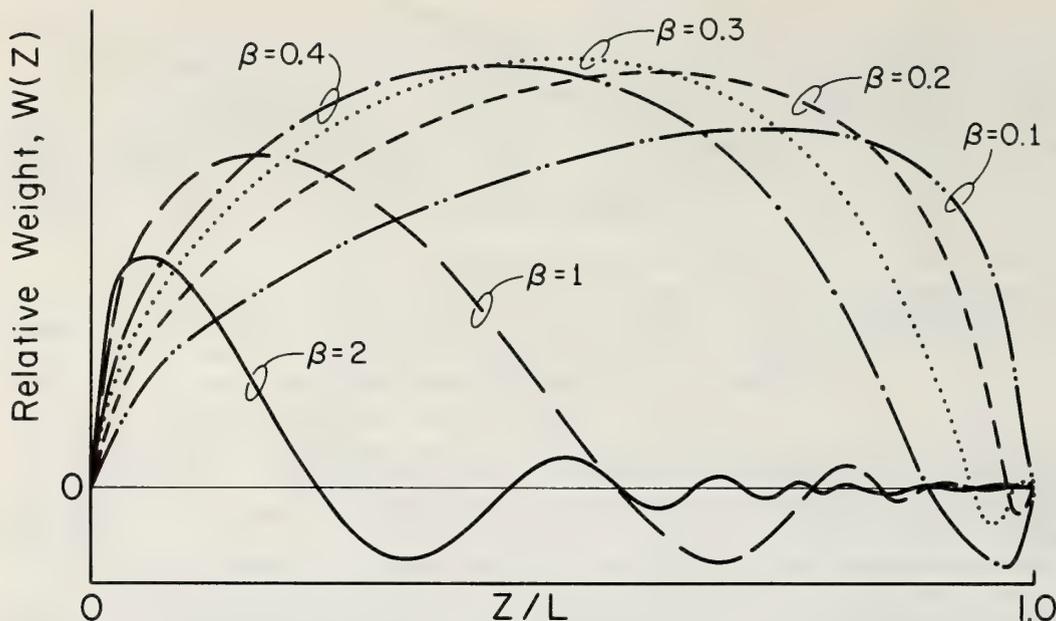


Figure 25.9 The relative weights of different portions of the path in determining winds by the correlation-slope method. The parameter $\beta \equiv \rho(\lambda L)^{-1/2}$ is the separation of the sensors measured in units of the Fresnel distance. These curves are calculated for point sensors; the finite area of a real sensor will tend to remove the negative weights near the receiver.

Figure (25.9) shows the path weighting function for various values of $\beta \equiv \rho(\lambda L)^{-1/2}$ where ρ is the spacing between detectors and L is the length of the propagation path. These calculations are made for point detectors and neglect any modification of the correlation function that may be produced by saturation. The negative weights shown in the figure can be removed in practice by the use of finite apertures.

Ability to manipulate the path weighting function suggests the possibility of determining independently the winds at various portions of the path. In general, the distribution of turbulence along the path must be determined at the same time. Lee and Harp (1969), Fried (1969), and Peskoff (1971) have all presented the integral equation relating the time and space intensity covariance function to turbulence and velocity, at least for non-saturated scintillations. The equation is nonlinear and no attempts have been made to invert actual observations of the integral to obtain turbulence and velocity. All the cautions and difficulties mentioned in 25.2 still apply, but in this case the equation is nonlinear and numerical inversions are likely to be less well behaved. Analytic inversion appears to be impossible.

Lee and Harp (1969) suggested that the correlation-slope method can be used to linearize the equation, though this process certainly involves loss of information and presumably increases the sensitivity of the answer to unknown changes in the spectrum of turbulence. Shen (1970) inverted the linearized equation, using some actual observations but inadequate direct wind measurements did not permit satisfactory verification of the results. Peskoff (1971) presented a formal inversion of the linearized equation.

25.3.5 Practical Applications

Lawrence, Ochs, and Clifford (1972) have shown direct, on-line measurement of the average wind crossing a laser beam to be practical. The method has already been used for observations of total flow over the mountains west of Boulder, Colorado, and for studies of the drifting of air pollutants over Salt Lake City. Measurement of total, or average, wind across a long path has an important advantage for such applications over point measurements with conventional anemometers. Point measurements are affected by variations that occur on scales smaller than the separation of the sensors, and their interpretation suffers from spatial aliasing, an effect in the space domain exactly analogous to the frequency aliasing that occurs when a physical quantity is sampled in the time domain at a rate too slow to reproduce all its frequency components. The optical method of average measurement across a light beam inserts a low-pass spatial filter into the measurement process and so prevents spatial aliasing.

Proposed applications for the laser-beam method include measurement of cross-wind components at airport runways and measurement of total convergence of air into, say, a city or a thunderstorm by surrounding an area with laser beams. This could best be done with mirrors, though the effect of finite mirrors in eliminating large-scale scintillations must be investigated. Mirrors, if they can be used, would permit the use of a zig-zag beam across an airport runway that would permit determination of longitudinal wind.

Fried (1969) has suggested the possibility of analysis of the scintillation of starlight to determine winds aloft. Fried has been very optimistic about the potential accuracy of the measurement, apparently ignoring the severe limitation to observational accuracy that is imposed by the non-stationary statistical nature of the actual atmosphere. Despite this difficulty, some measurement of winds aloft can certainly be made, and the method is presently being tested at our laboratories in Boulder. As in the case of measurement of turbulence aloft, the method will have marginal practical utility with starlight but may be quite useful when lasers have been placed on geostationary satellites.

25.4 Remote Measurement of Temperature

We now leave the effects of turbulence and discuss briefly a remote-sensing method that depends only on the average density of the air along a line of sight. This, of course, determines the optical path length. A change of 1°C in temperature or of $300 \text{ n}\cdot\text{m}^{-2}$ (3 mb) in pressure is sufficient to affect the optical path length by one part in 10^6 . This is several hundred times greater than the second-to-second fluctuations in optical path length caused by turbulence. Figure (25.10) shows an example of path-length changes measured over a 3 km round-trip path. The reference temperature was measured with a thermometer at the center of the path.

If we assume that the barometric pressure is known, the average temperature along a fixed open-air path can be determined by comparing the optical path length with the known true geometrical length. The apparatus for doing this could, in principle, be as simple as that shown in (F25.11). This is a modern version of Fizeau's classical experiment. A laser beam is passed through a modulator and traverses the open-air path to a retro-reflector. It returns through the modulator and is deflected by a beam splitter to a detector. A small adjustment of the modulation frequency is made to minimize the detected signal. The frequency is then measured by the counter. In practice, it would be desirable to provide a means for servo-controlling the oscillator to maintain a null output. The frequency read from the counter provides the optical path length, subject only to an ambiguity that can be removed by using several nearby frequencies that also minimize detector output. Optical path length can be converted to temperature given only the barometric pressure.

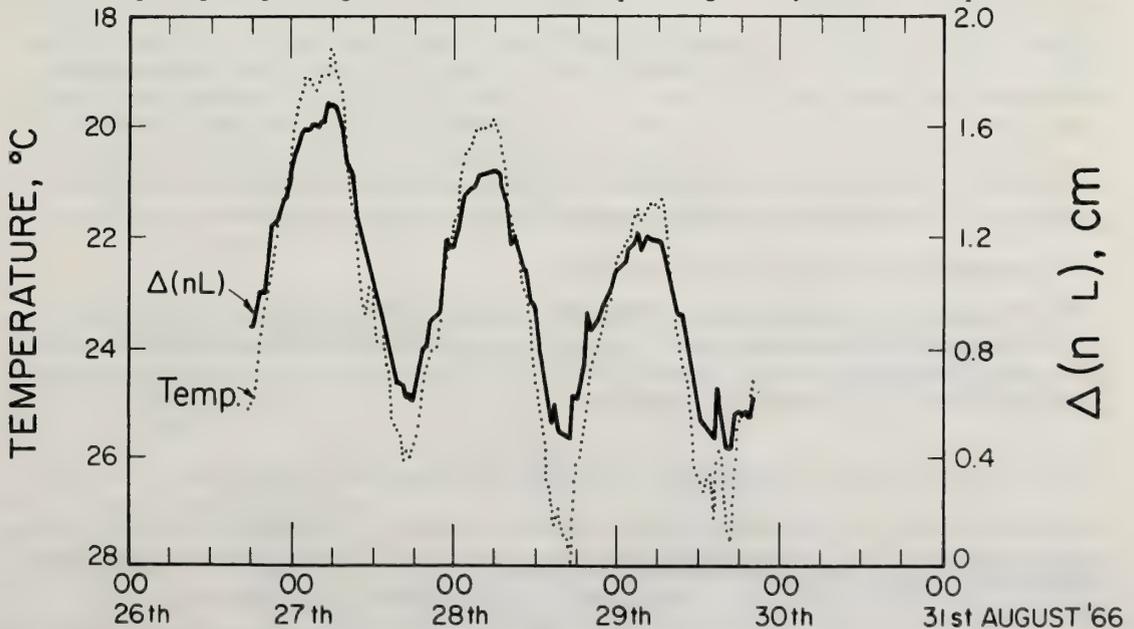


Figure 25.10 The relationship between optical path length over a 1.5 km path and the temperature measured at the midpoint. In this plot the relative scales have not been properly adjusted, so the variation of path length appears too small.

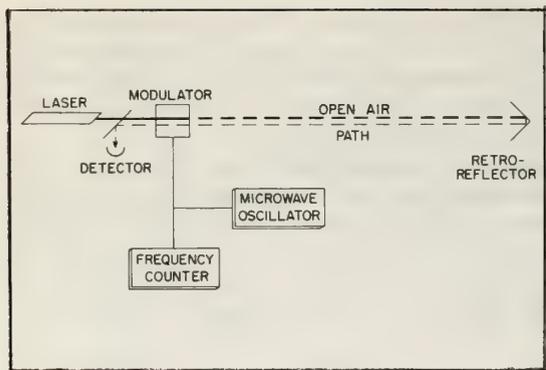


Figure 25.11 A block diagram of a simple device for measuring average temperature over a fixed path.

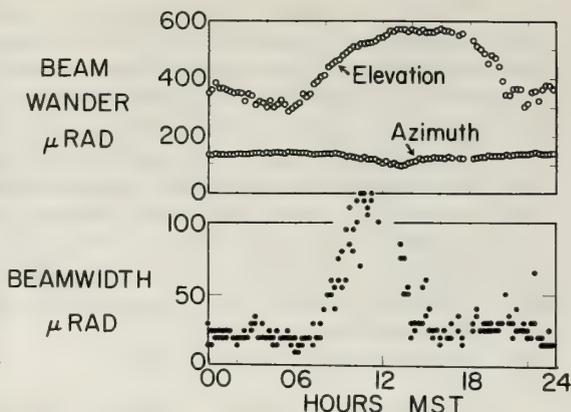


Figure 25.12 A typical day's measurements, made every 15 minutes, of the width and wander of a laser beam propagated over a 15 km path near Boulder, Colorado.

Turbulence will not be a limitation to measurements of average temperature unless the required accuracy is better than 0.1°C. For 0.1°C temperature accuracy, the barometric pressure must be measured to 30 n·m⁻² (0.3 mb) and both the modulation frequency and the fixed geometrical length of the path to 1 part in 10⁷. The effect of varying composition of the air, particularly the effect of water vapor, has been discussed in detail by Owens (1967). For temperatures near 20°C., a relative humidity error of 15 percent results in a temperature error of 0.1°C.

Determination of the geometrical length to one part in 10⁷ is at, or perhaps just beyond, the limit of the state of the art, and requires the averaging of a large number of optical measurements taken under various carefully measured weather conditions.

If the path is not fixed, or if the path length is unknown, a more complicated variant of the optical path-length method can be used, utilizing the dispersion of the atmosphere. In round numbers, the atmosphere reduces the velocity of blue light by 330 parts per million while it reduces the velocity of red light by only 300 ppm. Both these numbers are proportional to atmospheric density and therefore, for a given barometric pressure, inversely proportional to temperature. Simultaneous measurement of optical path length with both red and blue light provides the two equations needed to solve for the path length and temperature. A discussion of this method and its accuracy has been given by Owens and Earnshaw (1968). Briefly, temperature measurements to 1°C. and distance measurements accurate to one part in 10⁶ are feasible.

A further elaboration of the two frequency principle suggests itself and has been investigated by Bean and McGavin (1970). Simultaneous measurement of apparent path length over a fixed path with an optical frequency and a microwave frequency yields the average temperature and the average water-vapor content.

25.5 Measurement of Vertical Temperature Gradient

The curvature C of a ray is given by the transverse gradient of refractive index or, for nearly horizontal rays, by the vertical refractive-index gradient. Adopting a sign convention that downward curvature is positive and expressing refractivity in N units (defined as N = [n-1] 10⁶) so that curvature is measured in μrad·km⁻¹, we can differentiate (25:1) with respect to height to obtain

$$C = \frac{dN}{dh} = - \frac{79}{T} \frac{dP}{dh} + \frac{79P}{T^2} \frac{dT}{dh} \tag{25:10}$$

For sea-level conditions, where the pressure is 1.01325 x 10⁵ n·m⁻² (1013.25 mb), the vertical pressure gradient is -12.1 n·m⁻²·m⁻¹ (-121 mb·km⁻¹), and the temperature is 20°C; the ray curvature is related to the temperature lapse rate (°C·km⁻¹) by

$$C = 32.6 + 0.93 \frac{dT}{dh} \text{ urad} \cdot \text{km}^{-1} \tag{25:11}$$

Under these conditions the temperature gradient needed to prevent bending ($C = 0$) is $-35^{\circ}\text{C}\cdot\text{km}^{-1}$. Although this temperature lapse rate is much greater than that for a normal atmosphere ($-6^{\circ}\text{C}\cdot\text{km}^{-1}$) or an adiabatic atmosphere ($-10^{\circ}\text{C}\cdot\text{km}^{-1}$), even greater lapse rates can occur for a short distance above a hot surface. Such conditions produce upward curvature and cause a *mirage*. On the other hand, if the vertical temperature gradient becomes strongly positive, exceeding about $134^{\circ}\text{C}\cdot\text{km}^{-1}$ as may well happen above a cold surface, the normal downward curvature of the ray may exceed $157\ \mu\text{rad}\cdot\text{km}^{-1}$, the curvature of the earth's surface. This produces another kind of mirage known as *looming*. A review of the theory and observation of mirages, with an extensive bibliography, was recently prepared by Viezee (1968).

Since the curvature of a light ray depends so strongly on the vertical temperature gradient of the atmosphere, the apparent position of (or angle or arrival of the light from) a distant object or, what is equivalent, the spot produced at a distance by a fixed laser, will vary from time to time and can be used to measure changes in the vertical temperature gradient. Figure (25.12) shows measurements made in Boulder of this effect over a 15 km path during a typical day. For this path, a beam deflection of $50\ \mu\text{rad}$ is a sizeable fraction of the turbulent broadening of the beam and is easily measurable. Such a deflection corresponds to a change in curvature of $6.7\ \mu\text{rad}\cdot\text{km}^{-1}$ or a change in vertical temperature gradient of $7^{\circ}\text{C}\cdot\text{km}^{-1}$.

25.6 References

- Bean, B. R. and R. E. McGavin, 1970: Electromagnetic phase variability as a measure of water vapor and temperature variations over extended paths, Chap. 58, pp. 729-743 of K. Davies, ed. *Phase and Frequency Instabilities in Electromagnetic Wave Propagation*, AGARD Conference Proceedings No. 33. Technivision Services, Slough, England.
- Beckmann, P. (1965), Signal degeneration in laser beams propagated through a turbulent atmosphere. *Radio Sci.* 69D, 4, 629-640.
- Briggs, B. H. and M. Spencer (1954), Horizontal movements in the ionosphere. *Reports on Prog. in Phys.* 17, 245-280.
- Carlson, F. P. (1969), Application of optical scintillation measurements to turbulence diagnostics. *J. Opt. Soc. Am.* 59, 10, 1343-1347.
- Clifford, S. F. (1971), Temporal frequency spectra for a spherical wave propagating through atmospheric turbulence. *J. Opt. Soc. Am.* 61, 10, 1285-1292.
- Cohen, M. H., E. J. Gundermann, H. E. Hardebeck, and L. E. Sharp, (1967), Interplanetary scintillations. II. Observations. *Astrophys. J.* 147, 2, 449-466.
- Corrsin, S. (1951), On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *J. Appl. Phys.* 22, 469-473.
- Fedor, L. S. (1967), A statistical approach to the determination of three-dimensional ionospheric drifts. *J. Geophys. Res.* 72, 21, 5401-5415.
- Fried, D. L. (1966), Limiting resolution looking down through the atmosphere. *J. Opt. Soc. Am.* 56, 10, 1380-1384.
- Fried, D. L. (1969), Remote probing of the optical strength of atmospheric turbulence and of wind velocity. *Proc. IEEE* 57, 4, 415-420.
- Gray, D. A. and A. T. Waterman Jr. (1970), Measurement of fine-scale atmospheric structure using an optical propagation technique. *J. Geophys. Res.* 75, 6, 1077-1083.
- Hogge, C. B. and W. L. Visinsky (1971), Laser beam probing of jet exhaust turbulence. *Appl. Opt.* 10, 4, 889-892.

- Höhn, D. H. (1966), Effects of atmospheric turbulence on the transmission of a laser beam at 6328 Å. II – Frequency spectra. *Appl. Opt.* 5, 9, 1433-1436.
- Hufnagel, R. E., 1966: An improved model turbulent atmosphere, NAS Summer Study, *Restoration of atmospherically degraded images*, Appendix 3.
- Kolmogorov, A., 1941: in *Turbulence, Classic papers on statistical theory*, S. K. Friedlander and L. Topper, Eds., New York, Interscience, p. 151 (1961).
- Lawrence, R. S., G. R. Ochs, and S. F. Clifford (1972), The use of scintillations to measure average wind across a light beam. *Appl. Opt.* 11 2, 239-243.
- Lee, R. W. and J. C. Harp (1969), Weak scattering in random media, with applications to remote probing. *Proc. IEEE* 57, 4, 375-406.
- Lumley, J. L. and H. A. Panofsky, 1964: *The structure of atmospheric turbulence*, John Wiley and Sons, New York.
- Mandics, P. A. (1971), Line-of-sight acoustical probing of atmospheric turbulence, Stanford Electronics Laboratories Technical Report No: 4502-1 (SEL-71-002) (March 1971).
- Mitra, S. N. (1949), A radio method of measuring winds in the ionosphere. *Proc. IEE* 441, 96-111.
- Obukhov, A. M. (1949), Structure of the temperature field in a turbulent flow. *Izv. Akad. Nauk, SSSR. Ser. Geograf. Geofiz.* 13, 38.
- Ochs, G. R. 1972: Pattern velocity computers – Two types developed for wind velocity measurement by optical means. *Rev. Sci. Instr.* 43, 6, 879-882.
- Owens, J. C. (1967), Optical refractive index of air: Dependence on pressure, temperature, and composition. *Appl. Opt.* 6, 1, 51-59.
- Owens, J. C. and K. B. Earnshaw, 1968: Development of a microwave modulated, dual optical wavelength geodetic distance-measuring instrument, ESSA Tech. Rept ERL 117-WPL 8, U. S. Government Printing Office.
- Peskoff, A. (1968), Theory of remote sensing of clear-air turbulence profiles, *J. Opt. Soc. Am.* 58, 8, 1032-1040.
- Peskoff, A., (1971), Theory for remote sensing of wind-velocity profiles, *Proc. IEEE* 59, 2, 324-326.
- Ryznar, E. (1965), Dependency of optical scintillation frequency on wind speed. *Appl. Opt.* 4, 11, 1416-1418.
- Shen, L. (1970), Remote probing of atmosphere and wind velocity by millimeter waves. *IEEE Trans. AP-18*, 4, 493-497.
- Strohbehn, J. W. (1970a), The feasibility of laser experiments for measuring the permittivity spectrum of the turbulent atmosphere. *J. Geophys. Res.* 75, 6, 1067-1076.
- Strohbehn, J. W. (1970b), Remote sensing of clear-air turbulence, *J. Opt. Soc. Am.* 60, 7, 948-950.
- Tsvang, L. R. (1963), Some characteristics of the spectra of temperature pulsations in the boundary layer of the atmosphere. *Izv. Geophys. Ser. No.* 10, 1594-1600.

Veazey, D. R. (1970), A literature survey of clear air turbulence, Texas A. and M. University, College Station, Texas 77840, Rept. No. NASA CR-106211.

Viezee, W., 1968: Optical mirage, in *Scientific Study of Unidentified Flying Objects*, E. U. Condon, Ed. New York: Bantam Books, ch. VI-4.

List of Symbols

N	optical refractivity, $(n - 1)10^6$
P	atmospheric pressure, mb.
T	temperature, Kelvin
D	structure function
C_T	temperature structure parameter
D_T	temperature structure function
r	separation of sensors, or diameter of an atmospheric irregularity
ν	ratio of specific heats, c_p/c_v
c_p	specific heat at constant pressure
c_v	specific heat at constant volume
λ	wavelength of light
z	distance along a propagation path
p	size of a scintillation pattern from an irregularity of size r
d	aperture diameter
L	optical path length
C_N	refractive index structure parameter
$\sigma_{\ln I}^2$	log-intensity variance
I	intensity (more properly, irradiance)
β	path weighting function
C	curvature of a ray
h	height above the ground

Chapter 26 REMOTE SENSING FROM SATELLITES

Harold W. Yates

Satellite Experiment Laboratory
National Environmental Satellite Service
National Oceanic and Atmospheric Administration

The Satellite has enormous advantages in both temporal and spatial coverage of the earth for remote sensing applications. It also has limitations imposed by its relatively great distance from its target, its position (it must view the surface of the earth and the lower troposphere through the remainder of the atmosphere) and the weight, size and power capabilities of practical spacecraft. Achievements of the research and operational satellites are briefly described and the directions of current developmental efforts are enumerated. We are entering a phase of satellite sensors when microwave and radar systems are being practically exploited in addition to the traditional passive sensors operating in the infrared and visible regions of the spectrum.

26.1 The Satellite as a Platform for Observation

A principal reason for the current importance of remote sensing is the need for the collection of data on a scale and with a coverage either prohibitively expensive or virtually impossible by conventional or direct means. In terms of accuracy and precision, direct or contact methods are almost always superior to remote techniques.† However, the requirements of modern technology – especially meteorology – cannot be met by any imaginable network of in situ sensors. Thus a weather radar produces an image of precipitation and dense clouds within its operating range which could be equaled only by an enormous network of local sensors and data transmission circuits. Similarly an earth-orbiting satellite covering the entire globe of the earth once every 12 hours in a polar orbit, can provide a coverage which has never before been possible with local sensors because the resources to install and operate such a vast network never have been – and in all probability never will be – available for that purpose.

Remote sensing from a satellite is particularly advantageous in the study of the earth's atmosphere. This is because the atmosphere is an enormous three-dimensional fluid (there are more than 2 million tons of air for each of the 3.5 billion inhabitants of the earth), the majority of which lies over the oceans. An adequate network of implanted sensors would be quite impractical†† and would be unacceptable in any event because of the hazard to air transportation. Aircraft, which today supply a large amount of operational data, cannot provide global coverage because the bulk of commercial traffic is concentrated in areas of dense, affluent population and extending this to all other areas would be prohibitively expensive.

It is interesting to note that the most productive use of the satellite for earth studies has been in meteorology – the observation of the atmosphere rather than the surface of the earth. This is no accident but results from the fact that in looking down from a satellite the first thing one sees is the atmosphere which is the target, or object of study, to the meteorologist. To observe surface features, one must look through the atmosphere and for most such observations our essential mantle of gas, haze and clouds constitutes a nuisance and formidable source of masking and interference. This is also true to a large extent of observing the troposphere where one must look through the overlying atmosphere. Often the success of a promising technique will hinge on the ability to overcome these interferences.

† There are exceptions, particularly where the energy required to drive the contact sensor is large enough to distort the object being observed. An example is thermography of the human body (see Applied Optics, Sept. 1968).

†† There have been effective research programs in implanted sensors, such as the free floating, constant-altitude GHOST balloons or the French EOLE program, but only on a limited scale and for very specific purposes.

26.1.1 Advantages and Disadvantages – General Discussion

Coverage and Resolution in Time and Space

In coverage the satellite has no peer for remote sensing. From an altitude of 1500 km the NOAA satellites can effectively observe instantaneously a circular area of the earth roughly 1300 km in radius, which is about 2% of the surface of the earth. Its orbit is such that it will cover the entire globe in 12 hours (24 hours to observe the full globe under sunlit conditions). Figure (26.1) is a mercator montage of approximately 84 individual TV pictures taken from the ITOS satellite over a 24-hour period. A geostationary satellite will continuously see nearly half of the earth and can effectively observe about 25% of the surface of the earth. Figure (26.2) is a single frame from the ATS III spin-scan camera which requires 20 minutes to produce one full frame. The vantage point of space provides a view of the earth and its atmosphere not possible before the advent of earth orbiting satellites.

Along with this extensive and speedy coverage, however, come the disadvantages of distance from the target and the attendant limitations on resolution and sensitivity. These limitations are important for applications concerned with observations of the surface of the earth but are less restrictive for systems designed to observe the atmosphere of the earth. For atmospheric observations, passive systems – and we consider here only passive systems – are more likely to be energy limited than resolution limited so that design becomes a matter of trading off between the parameter of spatial resolution, temporal resolution and intensity resolution or contrast. As an example of this process in developing a system to meet a specific application, the design of a specific instrument, the VISSR (Visible Infrared Spin Scan Radiometer) will be discussed briefly.

The VISSR is a radiometer being designed to produce, from Geosynchronous altitude of 35,300 km (see 26.1.2), an image of the earth similar to the ATS image shown in Figure (26.2). Whereas the ATS scanner produced only an image in the visible region of the spectrum, the VISSR will produce an image in the 10 to 12 micrometer region of the infrared as well as an image in the visible. A primary use of these images, which are produced at the rate of three per hour, will be to deduce wind vectors from cloud motions (26.2.1, ATS satellite). Both the visible and infrared channels will be used for this purpose and the infrared images are particularly important because they are produced both day and night, whereas the visual images come only from the sunlit portion of the earth. The infrared channel is further valuable in that it can provide a good estimate of the altitude of a cloud under observation, and hence position the derived wind vector vertically. The temperature of the cloud is deduced from the infrared radiance and the



Figure 26.1 *Mercator montage of Vidicon pictures from ITOS satellite.*



Figure 26.2 *ATS III Earth image Nov. 18, 1967.*

altitude is taken as that point at which this temperature equals that of the air in the best available vertical temperature profile for the region of observation. The accuracy of this determination depends upon the opacity (density) of the cloud, whether or not it fills the field of view of the radiometer, the transmission of the atmosphere above the cloud and, of course, on the accuracy of the vertical temperature profile used. A dense cloud with a well defined top which fills the field of view of the radiometer and has clear air above it in a region where there is a good observed temperature profile can be positioned to ± 100 meters in altitude.

Wind vectors are derived from the observed motion of a cloud between two or more successive frames taken 20 minutes apart. The assumption, valid under most circumstances, is that the cloud has both the same vector velocity and the same temperature as the air in which it is entrained. The actual extraction of winds will be discussed in somewhat more detail later (26.2.1, *ATS Satellites*). Here we are concerned with the design of the instrument which will produce the images to be used in these derivations and the trade-offs involved in optimizing performance for the application.

The VISSR works from geosynchronous altitude and from this great distance maximum sensitivity is required, particularly in the infrared channel. Therefore the largest practical diffraction-limited optical system commensurate with spacecraft capabilities and the most sensitive available detector define the basic system. These are a 16' diameter aperture and a HgCdTe detector radiatively cooled to 100 °K or below. Now the designer must adapt to the optimum trade-off between spatial resolution (IFOV or instantaneous field of view), temporal resolution (frame time, or time between consecutive looks at the same spot in the field of view) and temperature resolution, or the ability to discern a cloud against the earth background by virtue of its difference in temperature from that background. A small field of view will enable one to see and follow a small cloud or feature but at a sacrifice in temperature and temporal resolution. Similarly a large field of view will give good temperature resolution, better temporal resolution but at a sacrifice in detail. Taking the fundamental system as a starting point – 16" optical system, HgCdTe detector and 100 rpm for the spacecraft – one computes for a 0.2 mr field of view (4 nautical miles spot size at the earth) a frame time, or time to complete one full scan of the earth, as 20 minutes and the temperature resolution as 1.3 °K for a background at 300 °K. This temperature resolution, or smallest discernable contrast between a feature and its background, is a function of the field of view and of the background temperature. The instrument responds to radiance (watts/cm²) and a 1 K temperature difference represents a larger radiance difference at 300 K than at a lower temperature. The relationships between IFOV, NEDT (noise equivalent temperature difference, or least discernable contrast between an element of the field of view and its surroundings) and temporal resolution can be presented in graphical form to facilitate understanding the trade-offs which are possible.

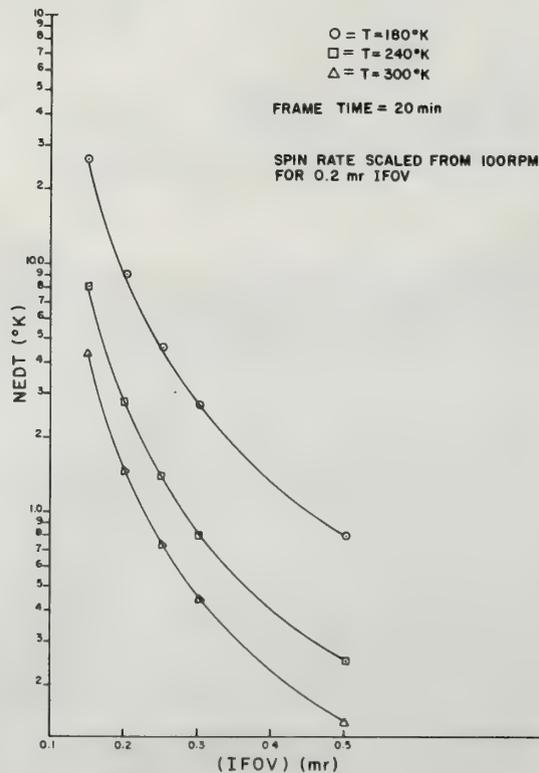


Figure 26.3 Temp. resolution (NEDT) for VISSR as a function of instantaneous field-of-view (IFOV).

Figure (26.3) shows the relationship between IFOV and NEDT for a fixed frame time of 20 minutes. Because NEDT depends also upon the average temperature of the scene, three curves are shown for 180 K, 240 K and 300 K. The VISSR scans one line of the picture for each rotation of the satellite and at 0.2 mr IFOV, the 1800 lines required to cover the full disc of the earth and the retrace time absorb 20 minutes. Therefore the plot is normalized to 100 rpm spin rate of the satellite at 0.2 mr IFOV. To maintain 20 minutes frame time, the satellite rpm would have to be increased if a smaller IFOV were used and decreased for a larger IFOV. The plot shows that if one is willing to accept less spatial resolution (larger IFOV) a much more sensitive temperature resolution is possible. However, the more coarse spatial resolution may result in the loss of discrimination of a small cloud or other feature in the picture.

Figure (26.4) relates NEDT to frame time for a constant field of view of 0.2 mr, again for three different scene temperatures. If one is willing to accept a longer frame time, a higher sensitivity in temperature resolution is possible. The sacrifice, of course, is in the rate at which any one element of the scene is sampled. An hour of frame time, for example, may be too gross for a phenomenon of interest.

Figures (26.3) and (26.4) enable one to relate the parameters of NEDT, IFOV and frame time for a single element of the scene. If the feature of interest in the scene is larger than the IFOV, as it frequently and usually is, these results must be modified to include the increased sampling time afforded by a large target. If the target of interest is a cloud subtending 0.6 mr, for example, and the radiometer has a 0.2 mr IFOV, the output is the result of the integration of 3 samples and the temperature resolution is improved by about $\sqrt{3}$ or 1.732. Figure (26.5) illustrates how the temperature resolution (ΔT , or temperature difference between the cloud and its surround, which is the equivalent of NEDT), varies with the cloud

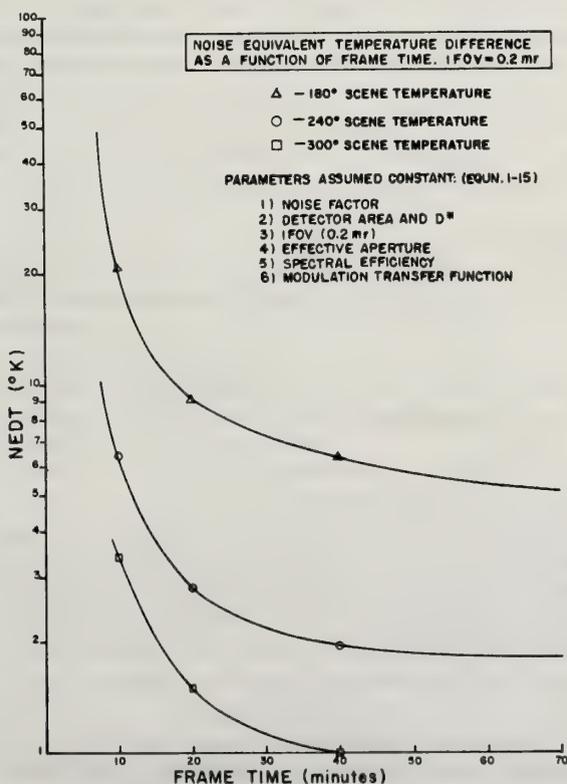


Figure 26.4 Temp. resolution (NEDT) for VISSR as a function of frame time.

dimensions for a fixed satellite rotation rate of 100 rpm (not frame time, note). Observe that ΔT here equals NEDT in the other plots when the cloud subtends the IFOV. If the cloud is larger than the IFOV, it can still be discerned even though its contrast is less than the NEDT. Conversely, a cloud *smaller* than the IFOV can be seen if its contrast is sufficiently large. This relationship between contrast and discernable level can be observed in other optical systems – the human eye for example. A ¼ inch diameter telephone wire can be “seen” from a distance of 200 feet against the sky if it is black, even though it subtends an angle only 1/10 the 1.0 mr resolution limit of the human eye. Reduce the contrast to, say, 50% by painting the wire gray and it can no longer be seen.

The above example is chosen to do two things: (1) illustrate the relationship between spatial, temporal and temperature (contrast) resolution in radiometric sensors and (2) indicate the approximate state-of-the-art of infrared systems today. The situation in the visible region of the spectrum is entirely different, in general being restricted only by the limits imposed by the optical inhomogeneity of the atmosphere of the earth. Since these limits in general range from 1 to 10 μ radian, data handling considerations usually determine system parameters for systems designed to observe the atmosphere. It becomes expensive to process and store the data from very high resolution pictures so that the lowest acceptable resolution is normally employed.

Limitations of Weight, Size and Power

In almost all instances, weight, physical size and power requirements for all spaceborne sensors must be minimized – and in some cases compromised – to fall within the capabilities of the satellites which carry them. In general, power and weight are the more serious restrictions, with size of secondary consideration. The ITOS spacecraft, for example, is a rectangular box whose size is dictated by the area of the solar panels to be deployed (they are stored flush with 3 large sides of the rectangle for launch and fold out in orbit). The resulting volume is several times larger than that required to accommodate all of the ITOS sensors and instruments, and the inside of the spacecraft is mostly empty area.

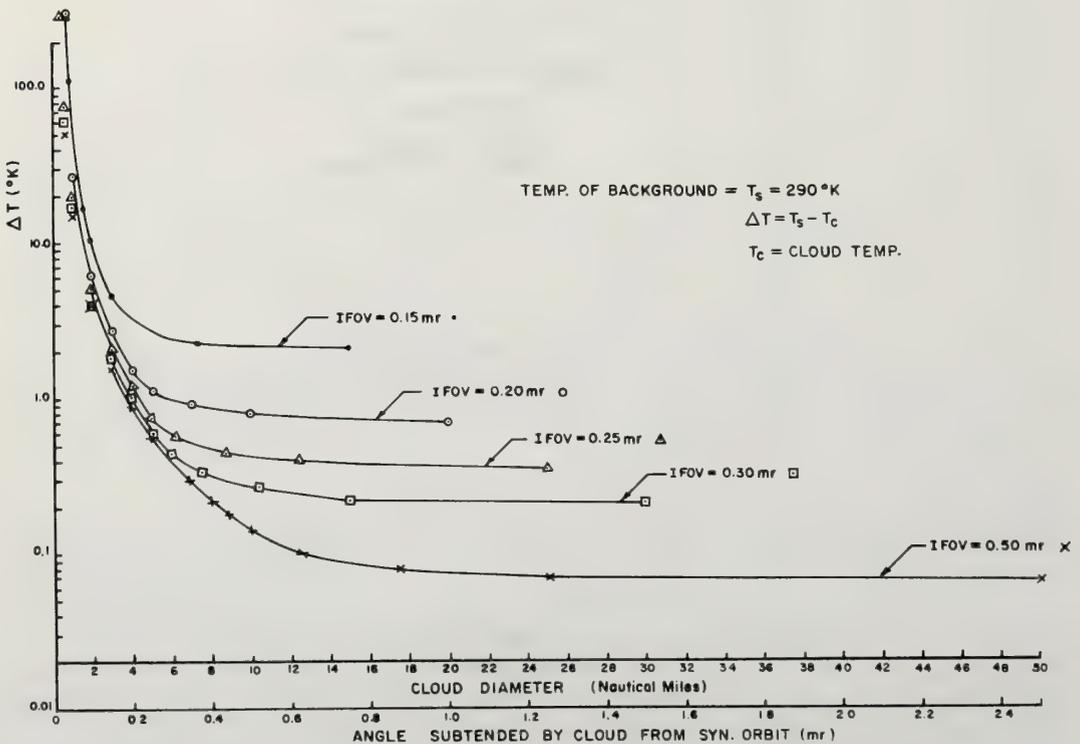


Figure 26.5 Least discernable target contrast (ΔT) for VISSR as a function of target size and IFOV.

Power is the most restrictive commodity. Virtually all spacecraft power is supplied at present by silicon photovoltaic cells which convert sunlight into electric power. A square meter of these cells can produce about 100 watts when in full sunlight at the earth's distance from the sun. The period of power production is determined by the orbit. Polar orbiters, such as Nimbus and ITOS, are in eclipse by the earth nearly half the time while geosynchronous satellites, such as ATS and SMS, are eclipsed only around the time of the equinoxes and then only for about 5% of each orbit. Thus a polar orbiter must have a solar panel capacity at least twice its average power consumption and a battery to sustain it during eclipse, whereas a geosynchronous satellite can operate continuously near peak power and requires only a very small storage capacity. A further consideration is the lifetime of the satellite, since solar cells degrade to about half their initial output after about two years in space.

Table (26.1) summarizes the weight, size and power for spacecraft having missions related to remote sensing. Nimbus, ITOS and ATS are in current use with 4, 2 and 3 respectively of their number having been successfully launched. TIROS-N is a planned follow-on to ITOS as the NOAA operational satellite; ERTS (Earth Resources Technology Satellite) will have the first of two approved spacecraft launched in late 1972 or early 1973; SMS will become the operational version of the ATS research satellites; and SKYLAB is the manned, earth-orbiting follow-on to the Apollo program.

The only practical relief from the rather serious power restrictions of current spacecraft will probably come from small nuclear fission generators. Development is underway and several units have been tested aboard Nimbus and other spacecraft, but they are not yet entirely satisfactory. Practical systems will likely employ solar panels for some time to come and this will mean that active sensor systems, such as LIDAR and RADAR, will be severely limited or in many cases not possible.

Effects of the Atmosphere of the Earth

A cursory examination of the effective application of spaceborne sensors to observation of the earth will show that meteorological observations predominate, particularly where quantitative measurements are required. This is not to say that studies of the surface of the earth have not or will not be effectively pursued, but the atmosphere, the first thing seen on looking down from a satellite, is the target for meteorological observations and is an intervening source of interference for surface observations. This interference is always present and ranges from minor to so severe as to render some desirable observations essentially impossible.

Even in the case of atmospheric studies, atmospheric interference can be serious. Infrared radiance measurements of dense clouds are used to estimate the height of the clouds by computing a temperature from the observed radiance and finding where this temperature prevails in the vertical temperature profile for the subject geographical position and time. (The assumption, normally valid, is that the cloud and its surrounding air are at the same temperature.) If there are thin clouds or even appreciable haze present above the target cloud, the radiance measurements will be distorted by emission and scattering from the intervening materials. The resulting errors in deduced cloud altitude can be minimized by corrections based on multi-frequency observations, but may still be significant.

More serious is the interference of clouds in the derivation of vertical temperature profiles of the atmosphere from infrared radiance measurements (26.2.2, ITOS satellite systems). Successful application of the method depends critically on the degree to which that portion of the observed radiance emanating from clouds and aerosol can be evaluated and subtracted in order to produce a "clear column radiance." If the clouds are solid and opaque, the soundings can be extended only down to the top of the cloud. These limitations have led to an active interest in the microwave region of the electromagnetic spectrum for making such observations. The advantage is that in the microwave region the wavelength – from 0.5 to 10 cm – is large compared to the dimensions of aerosol and cloud particles – from 1 to 20 micrometer in general – so that scattering is orders of magnitude less than in the infrared region. This enhanced transparency of clouds in the microwave region is very important. However, the cloud effects, though small, are still not zero and effective application for quantitative measurements will depend upon their magnitude and the degree to which corrections can be applied. Probably a combination of infrared and microwave systems will have synergistic advantages which will make their joint use almost assured.

Many operational and research techniques of tropospheric measurements involve radiation exchange between the earth, its atmosphere, the sun and space. For earth-atmosphere temperatures, the peak of the Planck blackbody curve lies in the range 10 to 15 μ meters and therefore this region of the infrared is of

particular interest. The earth — both land and water — radiates essentially as a gray body of high emissivity, generally above 0.9. The atmosphere is a mixture of clear (non-absorbing, non-emitting) gasses, gasses which have absorption bands and aerosols (clouds, haze, dust, etc.) which scatter, absorb and emit. In order to understand some of these processes and how they interrelate, consider (F26.6) which shows the radiance of the earth-atmosphere system as observed from the Nimbus IV satellite by the IRIS spectrometer (this instrument will be discussed more fully in 26.2.2, (Hanel, 1970)).

The spectral range is 400 to 1500 cm^{-1} (25 to 6.5μ meters) and three geographical areas are represented under clear, cloudless conditions — the Sahara Desert, the Mediterranean Sea nearby, and finally the Antarctic continent. The Sahara appears as a 40 to 50°C blackbody peaking between 700 and 880 cm^{-1} (bear in mind that the peak of a blackbody curve is displaced toward longer wavelengths when the abscissa units are cm^{-1} rather than μ meters) viewed through the atmosphere. H_2O vapor, CO_2 , O_3 and CH_4 (as well as some minor constituents) have absorption bands in this region and show as such. If the atmosphere were at the same temperature as the desert below, these bands would disappear and a smooth blackbody curve would be seen. The atmosphere, however, is neither isothermal nor homogeneous but has a temperature structure typically like that shown in (F26.12).

Two features of particular interest are the CO_2 band between 600 and 750 cm^{-1} and the O_3 band between 1000 and 1100 cm^{-1} . The CO_2 absorption band has three branches, labeled P, Q, and R respectively, or identifiable regions. A very narrow central Q-branch of extremely high absorbing (and hence emitting) power is bracketed by two broad branches, the P and R, of lower, but still fairly high, absorbing power. The O_3 band, on the other hand, consists of two broad branches separated by a narrow region of fairly low absorptivity. In the P, Q and R branches of the CO_2 band, the atmosphere is essentially opaque and one sees only radiation from the gas integrated over a path which reaches deeper into the atmosphere as the absorptivity of the gas decreases. One sees deeper in the P and R branches than in the Q-branch. The bulk of the radiation for the P and R branches arises in the upper troposphere and lower stratosphere, the region of lowest temperature in the atmosphere. The Q-branch sees mainly the middle stratosphere where the temperature is higher and hence it appears as a warm spike between the P and R branches. Note that there is relatively little difference between the absolute levels in the CO_2 band for all three curves. Over the Mediterranean, the upper troposphere is actually a little warmer than over the Sahara though, of course, the water is much cooler than the desert. The Antarctic continent is colder than the upper troposphere above it and the P and R branches both appear reversed from the other two situations.

The O_3 band behaves in a similar fashion with one notable exception. The central spike over the Sahara and Mediterranean appears warm because one sees the lower troposphere and the surface. Over the Antarctic, where the surface is colder, not warmer than the air, this central spike is reversed. Further differences, which contribute to the larger variations in absolute level for the O_3 band, are that CO_2 is uniformly distributed throughout the atmosphere whereas O_3 has a concentration peaked in the lower stratosphere and that the O_3 layer is not entirely opaque so that the surface of the earth is always contributing.

Figure (26.6) provides a dramatic illustration of radiation exchange in the earth-atmosphere system.

26.1.2 Optimum Orbits

The number of orbits possible for an earth satellite is infinite. Those which have the greatest utility for earth observation, however, fall into three categories: (1) polar orbit, sun-synchronous; (2) Geostationary; and (3) inclined orbit. Each has its unique characteristics — advantages and disadvantages — and no one orbit is optimum for all uses.

Polar Orbit, Sun-Synchronous

A satellite in a polar orbit travels in a plane that contains, or makes a small angle with, the axis of the earth. Its period of revolution is a function of its altitude above the earth, increasing upward from a practical minimum of about 90 minutes. The optimum polar orbit for a meteorological or earth observing satellite is one with the following characteristics, which are essentially those of the NOAA operational satellites:

- (1) As near polar as possible. This is determined by the inclination angle necessary to induce a precession of one rotation of the orbit plane in a year. A perfectly polar orbit would not precess

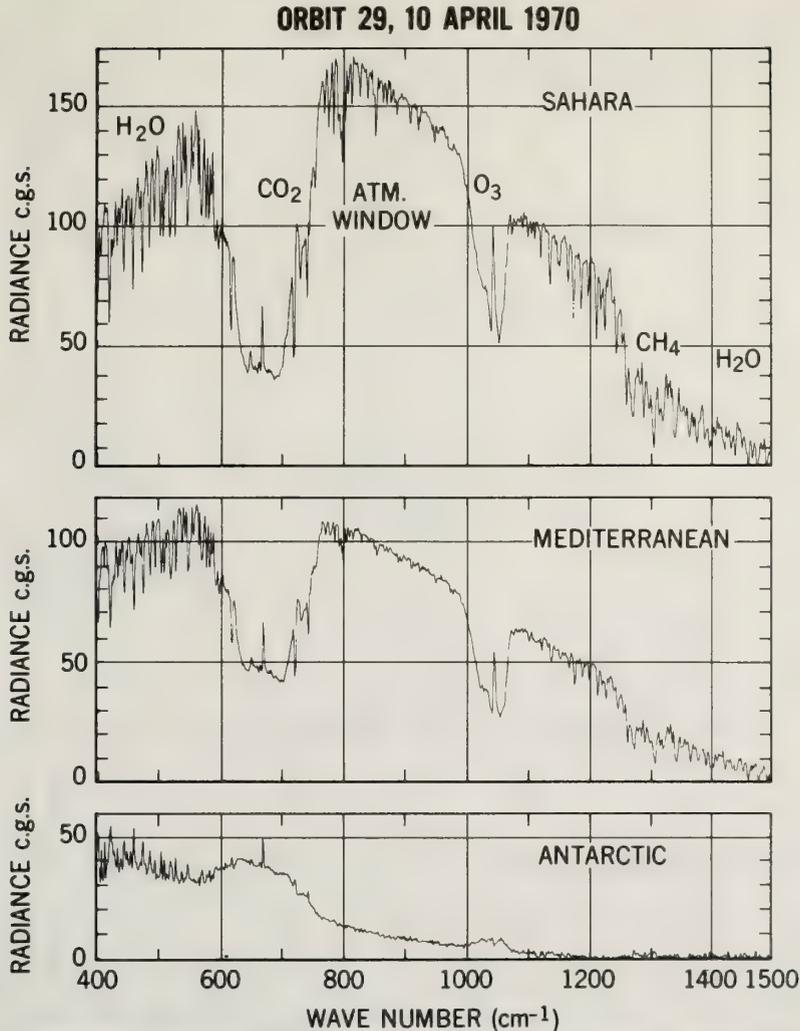


Figure 26.6 *Thermal emission spectra of the earth from IRIS D experiment on NIMBUS IV (from Hanel, R. A. and B. J. Conrath, 1970).*

and hence would not be sun-synchronous (i.e., pass over the equator at the same local time on each pass). An inclined orbit precesses because the earth, which is essentially an oblate spheroid, does not have a spherically symmetrical gravitational field. The angle required to produce sun-synchronous operation is a function of altitude (e.g., as shown in (T26.1)), ERTS at an altitude of 912 km. requires an inclination of 82° , or 8° away from the pole, while ITOS at 1500 km altitude requires a 78° inclination).

- (2) Sun-synchronous and retrograde to the equator (i.e., moving from west to east) with an equator crossing time of 0900 or 1500 local time to provide a 45° sun illumination angle for increased contrast in reflected light.
- (3) A period of 115-120 minutes in a circular orbit at an altitude of about 1500 km. enabling observation contiguity at the equator.

The polar orbit provides full coverage of the earth surface in a 12 hour period from an altitude low enough to provide reasonably good resolution at the surface of the earth. However, since it sees any one point on the major equatorial belt of the earth only twice each day (once under sunlit conditions) it cannot be used to study dynamic phenomena which occur on a shorter time scale. The sub-point (nadir) track of a polar satellite on a Mercator projection is illustrated in (F26.7). Note that the period is adjusted such that each succeeding day has the equator crossings approximately midway between those of

Table 26.1 Physical Characteristics of the Basic Nimbus Satellite

	Orbit	Spacecraft Weight (pounds)	Payload Weight (pounds)	Spacecraft Size (ft ³)	Payload Power (watts)	Communications Power (watts)	Solar Panel Output (watts)
NIMBUS	Circular Near-Polar (81°) Sun-Synchronous Noon Equator Crossing 1000 KM Alt.	1716	490	40	156	104	550
ITOS	Circular Near-Polar (78°) Sun-Synchronous 3 p.m. Equator Crossing 1500 KM Alt.	678	220	50	30	30	400
TIROS-N	Similar to ITOS	1380	580	66	200	45	600
ERTS	Circular Near-Polar (82°) Sun-Synchronous 9:30 a.m. Equator Crossing 912 KM Alt.	2080	450	52	271	104	550
ATS	Geostationary 37,000 KM Alt.	2950	520	48	400	200	600
SMS	Geostationary 37,000 KM Alt.	1256	165	150	33	80	140
SKYLAB	Circular 50° Inclination	200,000	8,700	90,000	500	7000	22,000

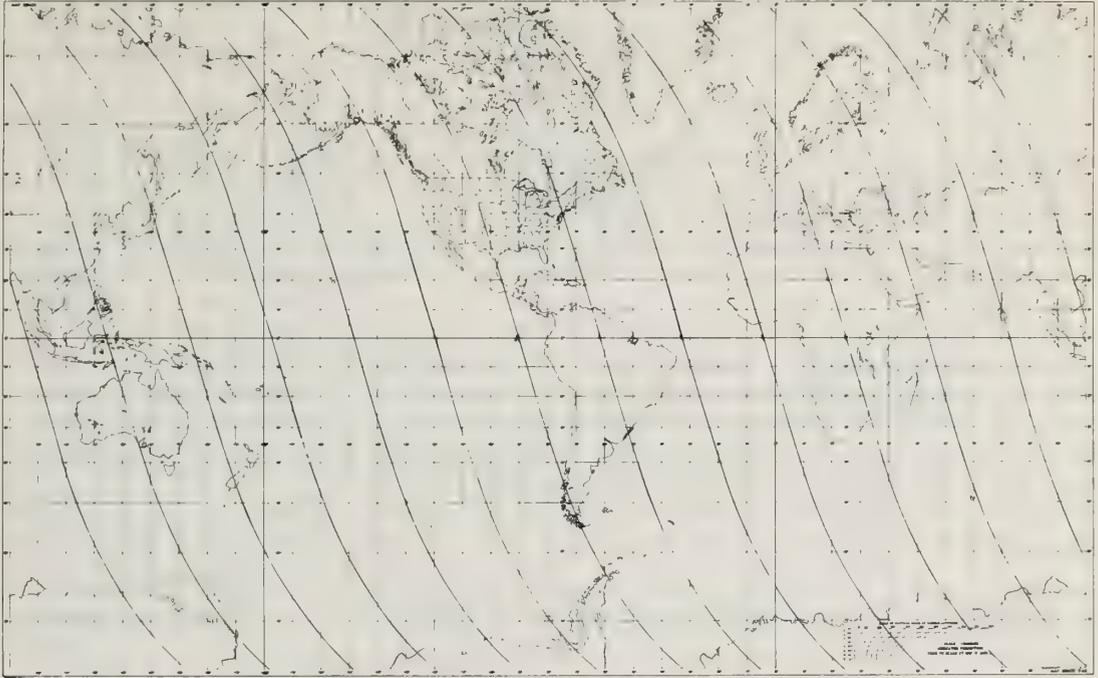


Figure 26.7 *Ground track of subsatellite point of syn-synchronous polar orbit.*

the previous day. This allows a point which was viewed only at an oblique angle between two passes on one day to be viewed from directly above on the next.

Geostationary Orbit

The geostationary or earth-synchronous orbit is a circular one in which the satellite is at such a distance (35,300 km.) that its period of revolution is identical to that of the earth's period of rotation. Thus, it always remains above the same meridian. If the orbital plane coincides with the equatorial plane of the earth, the satellite hovers above the same point on the equator. If it is inclined N° to the equatorial plane, the satellite moves N° north and N° south of the equator with a period very nearly equal to a sidereal day. The great distance of the geostationary satellite makes high spatial resolution much more difficult but it is ideally suited to the study of phenomena varying rapidly in time, since it looks continuously at one section of the earth. The earth coverage of two geostationary satellites is illustrated in (F26.8) where the large circles define the maximum range for communications and the small circles the limits of useful observations at a local observing angle of $\leq 60^\circ$. Full coverage of the equatorial belt of the earth would require four satellites spaced 90° apart.

Inclined Orbit

An inclined orbit is one similar to the sun-synchronous orbit but with a much larger angle between the orbital plane and the axis of the earth. The plane usually precesses fairly rapidly and the orbit drift can be virtually anything desired. A typical inclined orbit is shown in (F26.9) where the ground track of the first five orbits of APOLLO 7 are shown on a Mercator map of half of the equatorial belt of the earth. APOLLO 7 flew at an altitude of 240 km with a period of 100 mins. This orbit provides excellent coverage of the equatorial belt of the earth without the repetitive and redundant coverage of the poles characteristic of a polar orbit. It can be adjusted to observe the sun eclipsed by the horizon of the earth over a wide geographic range and this is an advantage, to be discussed later, which is not available with the other orbits. It does not observe the polar region, however, and cannot be used if such observations are a requirement.

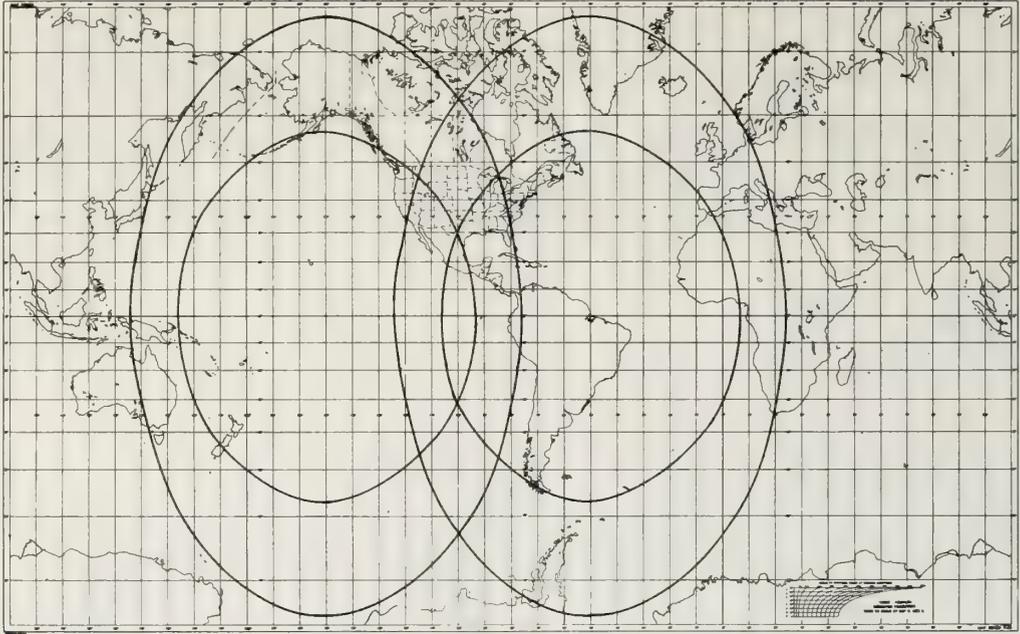


Figure 26.8 Earth coverage of two satellites in geostationary orbit.

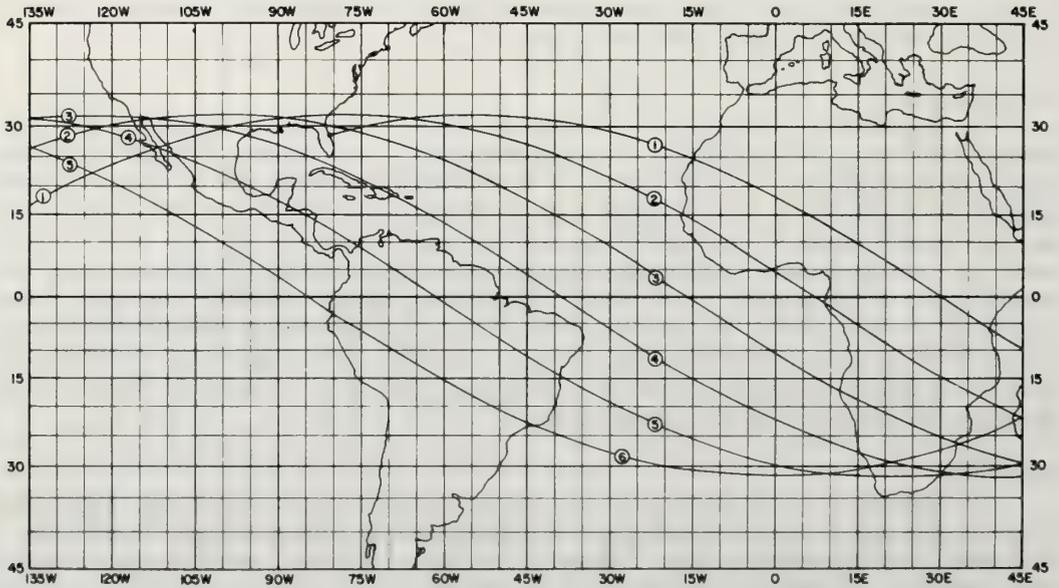


Figure 26.9 Path of subsatellite point of a typical inclined orbit – Apollo 7, first 5 passes.

26.2 Current Operational Satellites and Research Program

26.2.1 Research Satellites

NIMBUS

The Nimbus series of satellites has been devoted to meteorological research. Four members of the series have been successfully launched in the past seven years with some sensors on Nimbus IV, launched in April of 1970, still operational. The orbit and physical characteristics of the basic Nimbus satellite are given in (T26.1) and the sensor complement for Nimbus IV and Nimbus E (the members of a series of spacecraft are designated by a letter prior to launch and by a numeral after launch) are listed in (T26.2) and (T26.3) respectively.

The primary emphasis of Nimbus IV is the test of several instrumental approaches to sounding the atmosphere for temperature and humidity profiles. All of the instruments on this spacecraft function in the infrared and differ primarily in the instrumental approach to measuring infrared radiance. The SIRS (Satellite Infrared Spectrometer) is a conventional Fastie-Ebert spectrometer with 14 detectors whose positions and masks in the image plane determine the center and bandpass of the 14 channels observed (Wark, 1970). IRIS (Infrared Interferometer Spectrometer) is a Michelson interferometer which scans continuously through the spectrum from 6.3 to 50 micrometers with a resolution of 3 cm^{-1} . Both SIRS and IRIS were flown in earlier versions aboard the Nimbus III, which preceded Nimbus IV by about one year.

The SCR (Selective Chopper Radiometer) employs cells containing CO_2 gas – the atmospheric constituent being observed – and modulates by switching between cells of differing concentrations or between space and earth through a fixed cell (Barnett, 1971). This provides a narrower band radiation measurement for temperature determinations in the stratosphere.

The primary mission of the instruments on Nimbus IV is the measurement of temperature and humidity profiles of the atmosphere. Secondary objectives are profiles of other atmospheric constituents, such as O_3 , and cloud observations. The measurement of atmospheric temperature profiles from spacecraft is the most significant of the many contributions of space technology to meteorology. While some parts of the globe – notably North American and Europe – are adequately covered by today's network of radiosonde stations, vast areas, particularly in the Southern Hemisphere, have inadequate coverage. Satellites can provide a uniform and near-synoptic coverage of the globe never before available.

Nimbus E, due to be launched in November, 1972, carries three advanced sounding instruments: (1) the ITPR (Infrared Temperature Profile Radiometer); (2) an expanded and improved SCR; (3) NEMS (Nimbus E Microwave Spectrometer). ITPR measures infrared radiance much as SIRS, but makes a large number of contiguous samplings with a smaller field-of-view in order to operate in the presence of a partial cloud cover, (Smith, 1968). NEMS is the first satellite experiment designed to deduce temperature and humidity profiles from radiance measurements in the microwave region of the spectrum. It has 5 channels; 3 on the edge of the 60 GHz oxygen band, one in the 22 GHz water vapor band and one in the window region between these bands. The principal advantage of the microwave region is that the scattering by clouds and aerosols is many orders of magnitude less than in the infrared region. A passive microwave radiometer can see through most clouds and could provide data in regions where infrared systems could not penetrate. Clouds are not entirely transparent in the microwave region, however, and it is the purpose of the NEMS experiment to evaluate its efficacy in the presence of a wide variety of meteorological conditions.

The deduction of temperature and constituent profiles of a planetary atmosphere from passive radiation measurements was first proposed by Dr. Jean King (1956) and suggested for application to the meteorological problem by Kaplan (1959). The method, as applied to the Nimbus instruments, involves measuring the radiation emanating from different regions of an emission band of an atmospheric gas. The regions are chosen for different opacities so that the received radiation in each channel is integrated over a different region of the vertical atmosphere. The radiance is a function both of the temperature of the atmospheric gas and its emissivity which, in turn, is dependent upon the concentration of the gas. Thus in order to measure temperature one must know the emissivity which can only be accomplished by employing a gas of known and constant concentration. In the infrared region CO_2 is so used and in the microwave region O_2 is used.

Table 26.2 Radiometric Instruments for Remote Atmospheric Probing in the Infrared From the Nimbus IV Satellite

Radiometric Instrument	Channel(s)	Detector(s)	Instantaneous field of view (deg)	Linear resolution at subsatellite point from nominal orbital height (in parentheses) for indicated satellite (km)	Application
IRIS	1) 6.3-50 μ ; Spectrally scanning; resolution $\Delta\nu = 3.0 \text{ cm}^{-1}$ ($\Delta\lambda = 0.07 \mu @ 15 \mu$)	Thermistor bolometer	5	90 (1100)	Vertical profiles (temperature, H ₂ O, O ₃ , etc.)
SIRS	Total of 14 channels 7 in 15 μ CO ₂ band 6 in H ₂ O rotation band 1 in 11 μ window Resolution $\Delta\nu = 5 \text{ cm}^{-1}$	Thermistor bolometers	12	230 (1100)	Vertical profiles (temperature, H ₂ O)
Selective Chopper Radiometer (SCR)	Total of 6 channels 6 in 15 μ CO ₂ band 1 switchable to 11 μ window	Thermistor bolometers	10	190 (1100)	Vertical temperature profile
Filter Wedge Spectrometer (FWS)	1) 1.2-2.4 μ ; 3.2-6.4 μ Spectrally scanning; resolution $\lambda/\Delta\lambda = 100$ ($\Delta\lambda = 0.05 \mu @ 5 \mu$)	PbSe Single-stage, Radiatively-cooled to 160K	2	38 (1100)	Determine liquid water or ice content of clouds (1.2-2.4 μ) Vertical temperature profile (4.3 μ CO ₂ band) Vertical H ₂ O profile (6.3 μ H ₂ O band)
Temperature Humidity Infrared Radiometer (THIR) [Scanning]	1) 10.5-12.5 μ 2) 6.5-7.0 μ	Thermistor bolometers	0.40 1.20	8 23 (1100)	Cloud cover mapping Storm tracking Surface temperature Cloud top heights Air mass discrimination, vertical motion Jet stream location

Table 26.3 Radiometric Instruments for Remote Probing From the Nimbus-E Satellite

Radiometric Instrument	Channel(s)	Detector(s)	Instantaneous field of view (deg)	Linear resolution at subsatellite point from nominal orbital height (110 km) for indicated satellite (km)	Application
Infrared Temperature Profile Radiometer (ITPR)	Total of 7 channels 4 in 15 μ CO ₂ band 1 in H ₂ O rotation band 1 in 11 μ window 1 in 3.8 μ window	Pyroelectric (Tryglycine Sulfate)	1.5 (scanning $\pm 35^\circ$ each side nadir)	30.6	Operational temperature soundings to 20km with cluster sampling
Selective Chopper Radiometer (SCR)	Total of 16 channels 8 in 15 μ CO ₂ band 1 in 11 μ window 3 in pure rotation H ₂ O band at 18.6, 46.5 and 100 μ	Pyroelectric (Tryglycine Sulfate)	1.5 and 4.0	30.6 and 4.0	Temperature profiles to 60 km under clear and partially cloudy conditions, and cloud density and distribution
Surface Composition Mapping Radiometer (SCMR)	Total of 3 channels 2 IR(8.5, 10.7 μ) 1 visible (1.0 μ)	Mercury Cadmium Telluride (cooled photoconductive)	.03 ± 60 (scanning)	.545	Surface mineral composition, surface temperature with very high spatial resolution; selected coverage
Temperature Humidity Infrared Radiometer (THIR)	Total of 2 channels 6.7 μ water vapor 10.5-12.5 μ window	Germanium immersed thermistor bolometer	1.2 0.3 ± 60 (scanning)	22 for H ₂ O channel 8 for window channel	Day and night cloud top or surface temperatures moisture content of upper troposphere and stratosphere
NIMBUS E Microwave Spectrometer (NEMS)	Total of 5 channels 22.235 GHz 31.400 GHz 53.650 GHz 54.900 GHz 58.800 GHz	RCVR (Super Heterodyne)	10 (antenna beamwidth)	200	Tropospheric temperature soundings to 20km in presence of clouds
Electrically Scanning MW Radiometer (ESMR)	1.55cm (19.35 GHz)	RCVR (Super Heterodyne)	1.4, ± 50 (scanning)	27	Cloud liquid water content and ice mapping
Tracking and Data Relay Experiment (T&DRE)	2.063 GHz RCVR 2.253 GHz XMTR	(Transponder)	10 $^\circ$, 2 axis gimbal	not applicable	Track NIMBUS Satellite via ATS satellite data relay

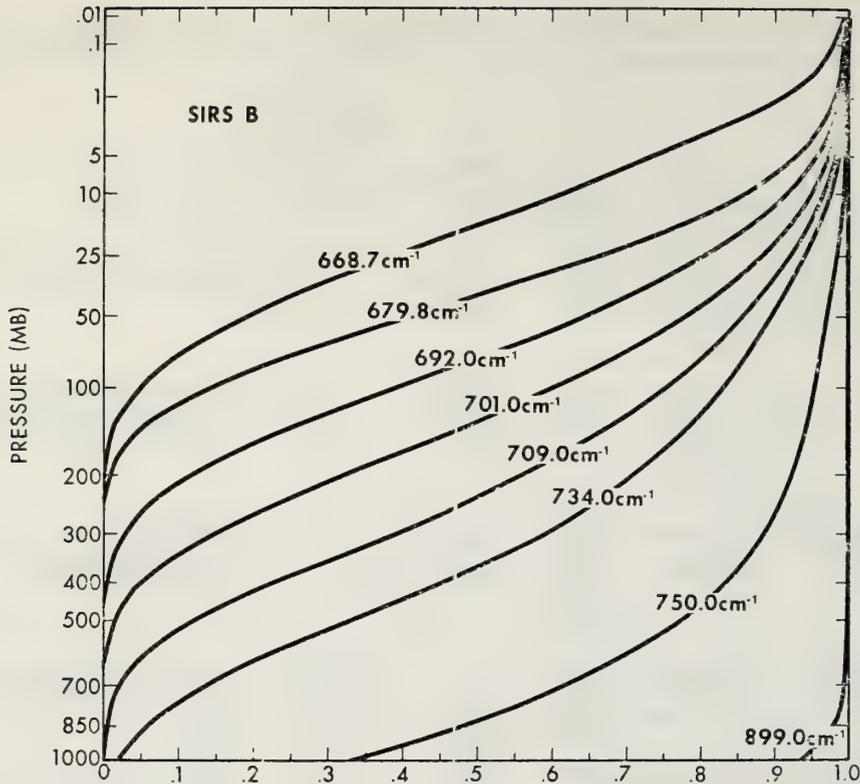


Figure 26.10 *Transmittance of SIRS B CO₂ channels.*

In order to understand the concept, consider the transmissivities of the atmosphere in the 8 channels of the SIRS B (Nimbus IV) instrument. The transmission from any given altitude through to the top of the atmosphere is plotted in (F26.10). Channel 8 (668.7 cm^{-1}) is chosen in the most strongly absorbing region of the CO_2 band and hence the atmosphere becomes opaque at about 200 mb (about 13 km). The remaining channels are selected for steadily increasing transmission so that received radiation comes from lower and lower regions of the atmosphere. The relative contribution of the atmosphere as a function of height is given by the derivative, $d\tau/dx$, of the curves of (F26.10) and these are plotted in (F26.11). Channel 7 receives more radiation from the vicinity of 20 km in the atmosphere than from the remainder, but it receives radiation from a wide range and overlaps all but one of the other channels. This overlapping, which is determined by the exponential structure of the atmosphere and hence unavoidable, means that an exact solution for temperature derived from these radiance measurements is not possible.

The weighting functions for two other Nimbus experiments, SCR and NEMS, are shown in (F26.12) and (F26.13). Figure (26.12) illustrates the higher altitude objective of SCR.

There are a number of mathematical approaches to computing temperature from the observed radiances, and these are treated in detail in Chapter 16, "Inversion Techniques," of this book. All of these solutions are approximations whose quality (accuracy and precision) are determined by the number of channels, the noise level on these channels, and the degree of contamination by radiation from clouds and aerosol within the field of view of the radiometer.

Performance of the instruments aboard Nimbus are illustrated in (F26.14), (F26.15), and (F26.16) where derived temperature profiles are compared with data from radiosondes where the two observations were geographically and temporally close enough to assure comparability. SIRS, SCR, and IRIS show very similar capabilities (except that SCR reaches to higher altitudes) averaging an rms deviation from radiosonde curves between $\pm 0.5 \text{ K}$ and $\pm 1.0 \text{ K}$ for clear (cloudless) atmospheric conditions. Under

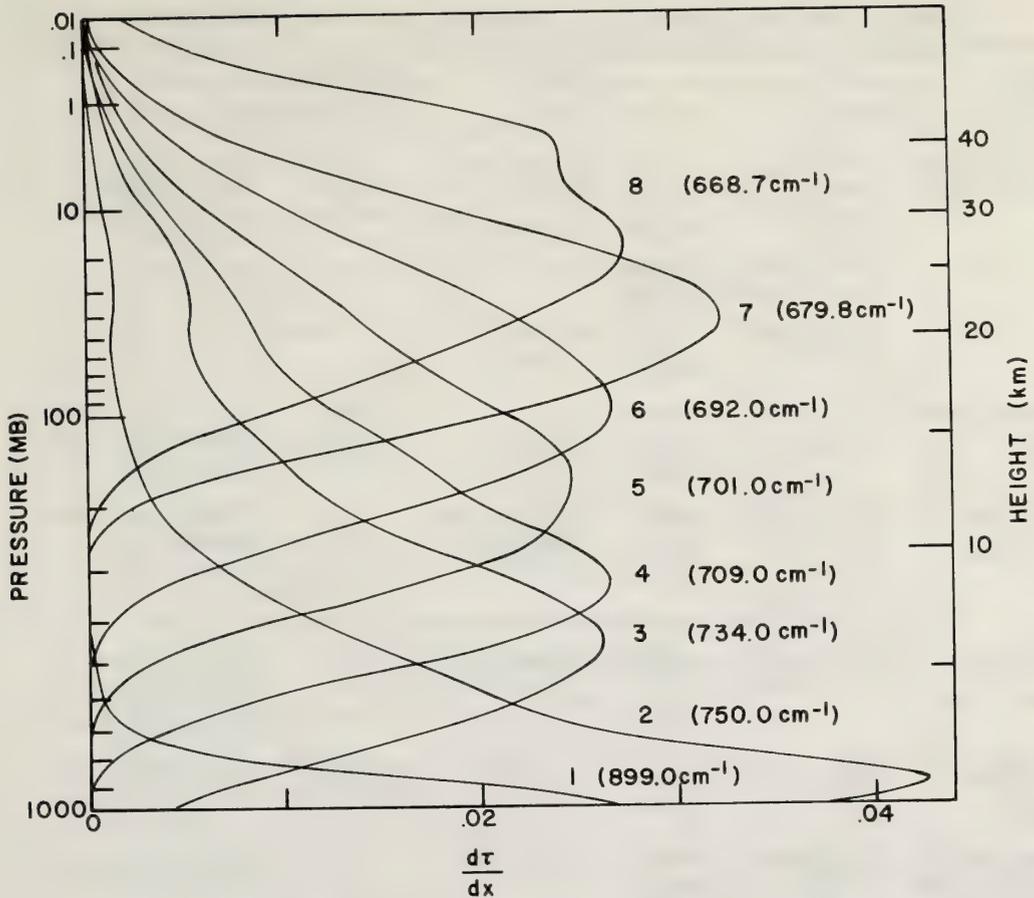


Figure 26.11 Weighting functions, $d\tau/dx$, for SIRS B CO_2 channels.

partly cloudy conditions the techniques for removing the cloud contributions to the radiance are still less than perfect and rms deviations up to ± 3.5 K are typical. Considerable effort, both analytical and instrumental, is being put into improving this capability because a very large part of the most interesting part of the earth is normally partly cloudy.

Nimbus F is the last of the approved Nimbus series. NASA plans beyond Nimbus F, call for combining the missions of Nimbus and ERTS (see following section) into a single spacecraft called EOS (Earth Observation Satellite). Nimbus F is heavily oriented toward experimental investigations in the microwave region of the spectrum. Table (26.4) lists the approved experiments and the purpose and characteristics.

Aside from the microwave experiments, the ERB (Earth Radiation Balance) and the LRI (Limb Radiance Inversion) experiments are of particular interest for possible operational use. The ERB is aimed at a very accurate measurement (meteorologists maintain that $\pm 1\%$ is required in order for the data to be truly meaningful) of the earth radiation to space with particular emphasis on the angular variation (i.e., is it Lambertian, as has been assumed). The LRI has important applications in the measurement of stratospheric constituents and may be adaptable to pollution monitoring should such measurements in the stratosphere become important in the monitoring of long range changes.

ERTS (Earth Resources Technology Satellites)

There are two approved ERTS satellites, the first of which is scheduled to be launched in 1973. They carry two primary sensors: (1) a Return Beam Vidicon (RBV) and (2) a Multispectral Scanner (MSS). Both instruments are designed to study the surface of the earth rather than the troposphere so that

Table 26.4 Radiometric Instruments for Remote Probing From the Nimbus-F Satellite

Radiometric Instrument	Channel(s)	Detector(s)	Instantaneous field of view (deg.)	Linear resolution (NM) sub-satel. point nom. orb. hgt. (110nm)	Objective
Earth Radiation Budget Radiometer (ERB)	Total of 22 channels 10 solar flux, 0.2 through $50 + \mu\text{m}$ 4 earth flux, 0.2 through $50 + \mu\text{m}$ 4 short-wave scanning channels, 0.3 to $4.0 \mu\text{m}$ 4 long-wave scanning channels, 4.0 to $> 50 \mu\text{m}$	Thermopile	Ch 1S-8S Max.FOV $<26^\circ$ Ch 9S-10S " " $<28^\circ$	Approx. 50 miles at sub-satellite point	Determine the earth's radiation budget including effects of angular dependence of outgoing radiation.
		Thermopile	Ch 11E-14E Max.FOV $<33.3^\circ$		
		Pyroelectric or thin film thermopile	$0.25^\circ \times 5.14^\circ$		
		Pyroelectric or thin film thermopile.	$0.25^\circ \times 5.14^\circ$		
Temperature Humidity Infrared Radiometer (THIR)	Total of 2 channels 6.7μ water vapor $10.5-12.5\mu$ window	Germanium immersed thermistor bolometer	1.2° 0.3° $\pm 60^\circ$ (scanning)	22 for H_2O channel 8 for window channel	Day and night cloud top or surface temperature moisture content of upper troposphere and stratosphere
High Resolution Infrared Radiation Sounder (HIRS)	Total of 17 channels 668,680,690,703, 716,733,749,900, 1225,1490,2190, 2210,2240, 2270 2700 and $14,500 \text{ cm}^{-1}$	Mercury cadmium and indium antimonide.	1.5° step scan complete contiguous coverage $\pm 38^\circ$ right left of nadir.	15	Global soundings up to 40 km obtained.
Pressure Modulated Carbon Dioxide Radiometer (PMR)	Total of 2 channels in $15\mu\text{m}$ CO_2 band.	Pyroelectric and mercury cadmium telluride.	10° straight down and $\pm 15^\circ$ along the orbital track.		Radiometric measurements of the temperature of selected layers at altitudes between 40 and 65 km on a global scale.
Electronically Scanning Microwave Radiometer (ESMR)	1 channel 37 GHz	RCVR (superheterodynes) dual.	$\pm 35^\circ$ conical scan at constant nadir angle at 50° .	About 1° width using stepped scan.	Map the cloud liquid water content and ice mapping.
Scanning Microwave Spectrometer (SCAMS) (same channels as Nimbus E NEMS)	Total of 5 channels 22.235 GHz 31.400 GHz 53.650 GHz 54.900 GHz 58.800 GHz	RCVR (superheterodyne).	$\pm 45^\circ$ in 7° increments.	200	Map tropospheric temperature profiles, water vapor content
Limb Radiance Inversion	Total of 4 channels $9.4-10\mu\text{m}$ $14.0-16.3\mu\text{m}$ $14.0-16.7\mu\text{m}$ $20-40\mu\text{m}$	Mercury cadmium telluride and thermistor bolometer.	$.03^\circ \times 0.3^\circ$ and $0.1^\circ \times 0.7^\circ$	2 km vertical resolution at the limb from 15-60 km (5 km for $20-40\mu\text{m}$ channel.	Limb radiance to determine global stratospheric temperature structure, and ozone distribution.

Table 26.5 Design Summary

Design Parameters	Visible Channel	Thermal Channel
Number of Channels	8	1 plus 1 redundant channel
Wavelength Band of Operation, Half-Power Points	0.54-0.7 microns	10.5-12.6 microns
Instantaneous Geometric Field of View (IGFOV)	0.025 x 0.021 mr	0.2 x 0.2 mr
Collecting Aperture	1090 cm ²	1090 cm ²
Detector	PMT	HgCdTe
Size	--	0.105 x 0.105 mm
Response	S-20 (enhanced)	--
Scan Period	0.6 sec	0.6 sec
Dwell Time	2.4 x 10 ⁻⁶ sec	1.9 x 10 ⁻⁵ sec
Information Bandwidth ^(a)	210 kHz	26 kHz
Dynamic Range, Albedo (%); Target Temperature (°K)	0 to 100	0 to 315
Performance Characteristics	Visible Channel	Thermal Channel
Noise Equivalent Radiance, NEN for an Extended Source (watt-cm ⁻² -sterad ⁻¹)	--	1 x 10 ⁻⁵
Noise Equivalent Differential Temperature for an Extended Source	--	1.5 °K at 200 °K 0.4 °K at 300 °K
S/N at 0.5% Albedo, for an Extended Source	3:1	
Modulation Transfer Function	0.34 at 2 x 10 ⁴ cycles/rad	0.42 at 2.5 x 10 ³ cycles/rad
V _p /V _{ss} , Target Size Equal to IGFOV's, Includes Optical Response Factors (Approximate)	0.45	0.55
Physical Characteristics	Scanner	Electronics Module
Estimated Weight	124 lb	12.5 lb
Size (b)	60 x 25.5 x 25.5	450 in. ³
Power Requirements	20 watts	
Inflight Calibration Provisions	Visible Channel 1. Sun and Space 2. Electronics Gain and Linearity Thermal Channel 1. Sun and Space 2. Calibration Blackbody and Space 3. Electronics Gain and Linearity	

(a) Physical location of information bandwidth filters is in the spacecraft.

(b) Excluding entrance aperture sunshade.

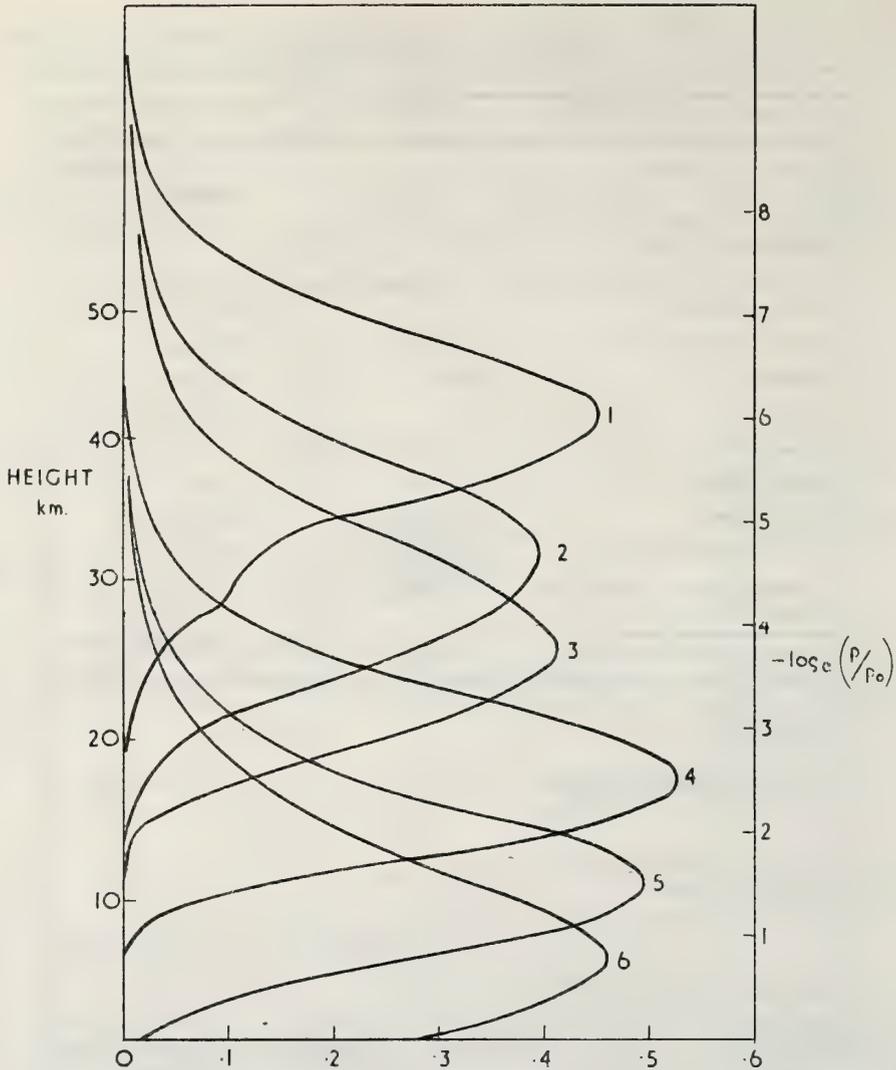


Figure 26.12 *Weighting functions for SCR (from Barnett, J. J., et al., 1972).*

they are not of direct application to this study. However, because of interaction between the surface and the troposphere, the ERTS data is expected to be useful for tropospheric studies.

The RBV consists of three vidicon cameras filtered to the spectral bands 0.475 to 0.575, 0.580 to 0.680, and 0.698 to 0.830 micrometers. The three are boresighted and have an instantaneous field of view on the ground of 185 km square. Resolution within this square is approximately 45 meters on the ground with a 33 db S/N ratio and about 10 gray scale steps for the first two wavelength regions. The 0.698 to 0.830 micrometer channel has a 25 db S/N at 8 gray scale steps.

The MSS is a 4 channel mechanical scanner which scans across the track below the spacecraft with six detectors in each channel. From nominal altitude the resolution element at nadir for each channel is 79 meters square and the spectral bands covered are: 0.5 to 0.6, 0.6 to 0.7, 0.7 to 0.8, and 0.8 to 1.1 micrometers. Bands 1 through 3 use photomultiplier tubes while band 4 uses silicon photodiodes. For the ERTS B mission a fifth band will be added in the 10.4 to 12.6 micrometer region. The ground resolution element for this added band will be 240 meters or about 3 times larger than bands 1 through 4.

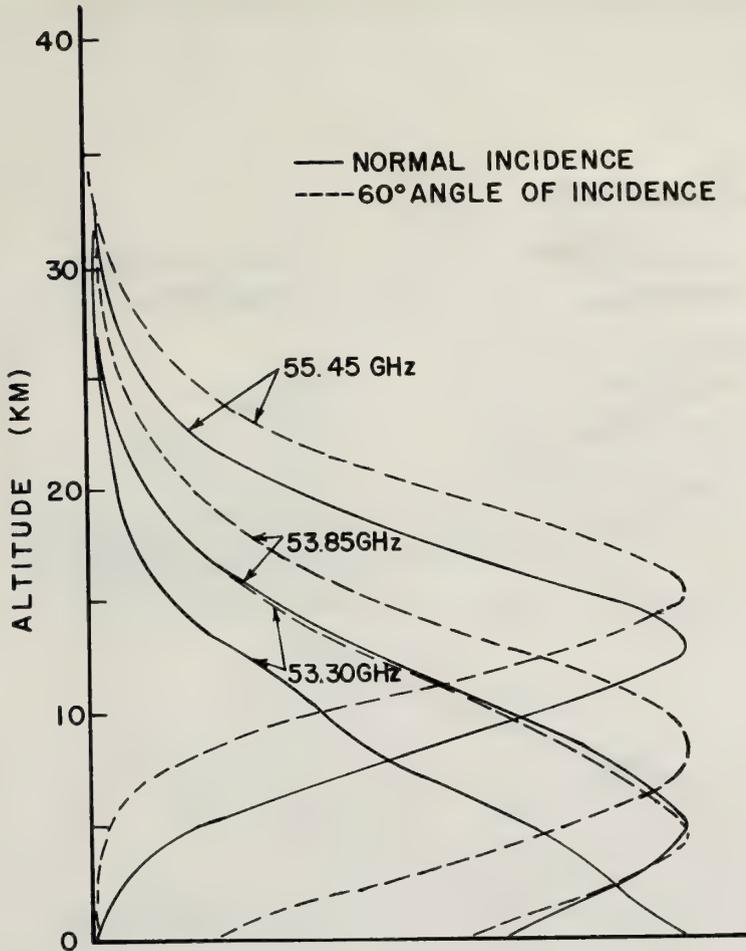


Figure 26.13 *Weighting functions for NEMS (from Nimbus User's Guide).*

ATS (Application Technology Satellite)

The NASA Goddard Space Flight Center has flown a series of satellites in Geostationary orbit (26.1.2) for the primary purpose of research in communications systems. Three of the series, ATS 1 through 3, have been launched with 1 and 3 achieving successful orbits in June, 1966 and May, 1967, respectively. Two additional members of the series, ATS F and G, are scheduled for launch in 1973 and 1975.

In addition to the communications experiments, ATS 1 and 3 carry optical scanners which produce an 1800 line image of the full earth disc (see (F26.2)) every 20 minutes. ATS-1 has a single channel in the visible region while ATS-3 has three channels which combine to produce a color image of the earth.

These ATS images provide a unique and valuable means of studying phenomena which vary fairly rapidly in time. Whereas the polar orbiting satellites see a given point on the earth only once every 12 hours (once every 24 hours under sunlit conditions) the geostationary satellite hangs continuously above the same point on the equator and the ATS cameras give three pictures per hour. A very fruitful application of these images is the derivation of wind vectors from cloud motions. By measuring the velocity vector of a given cloud or cloud feature as it moves from one frame to the next, the wind velocity is measured. One must, of course, assume that the cloud is moving at the same speed as the air in which it is entrained, and this is not always a valid assumption. However, it is so for most cases and questionable conditions can be identified and rejected.

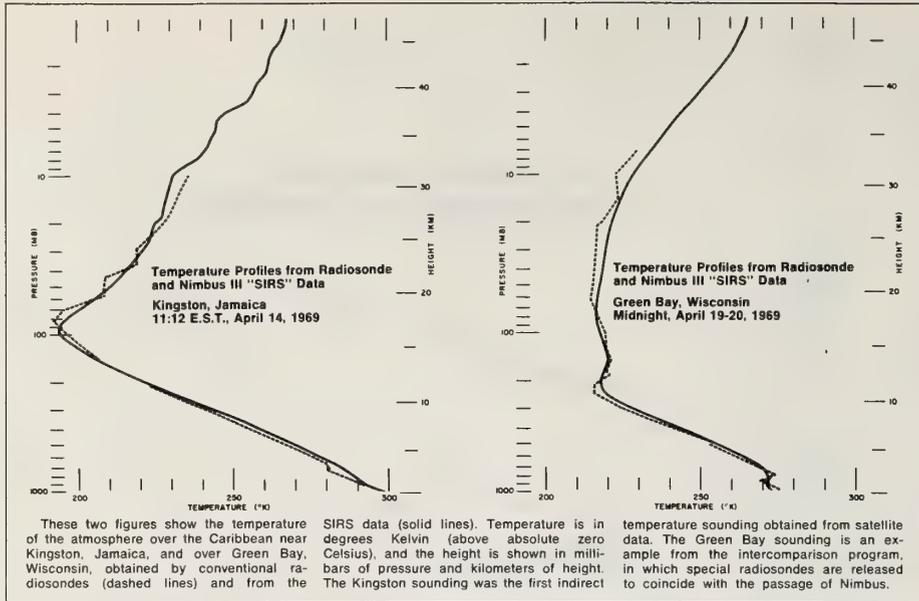


Figure 26.14 Temperature profiles for Kingston, Jamaica and Green Bay, Wisconsin, from SIRS data.

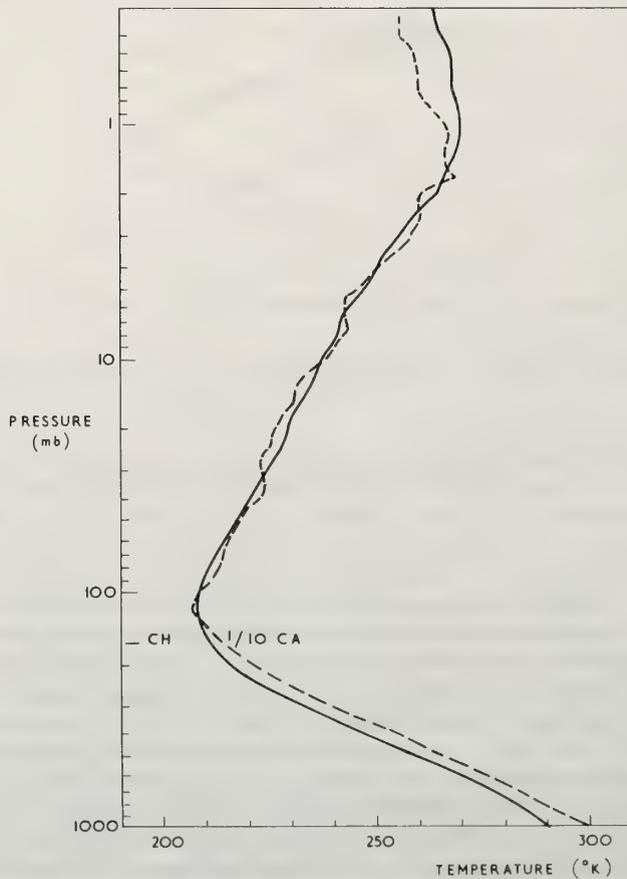


Figure 26.15 Temperature profile for Wallops Island, Va. from SCR data (from Barnett, J. J., et al., 1972).

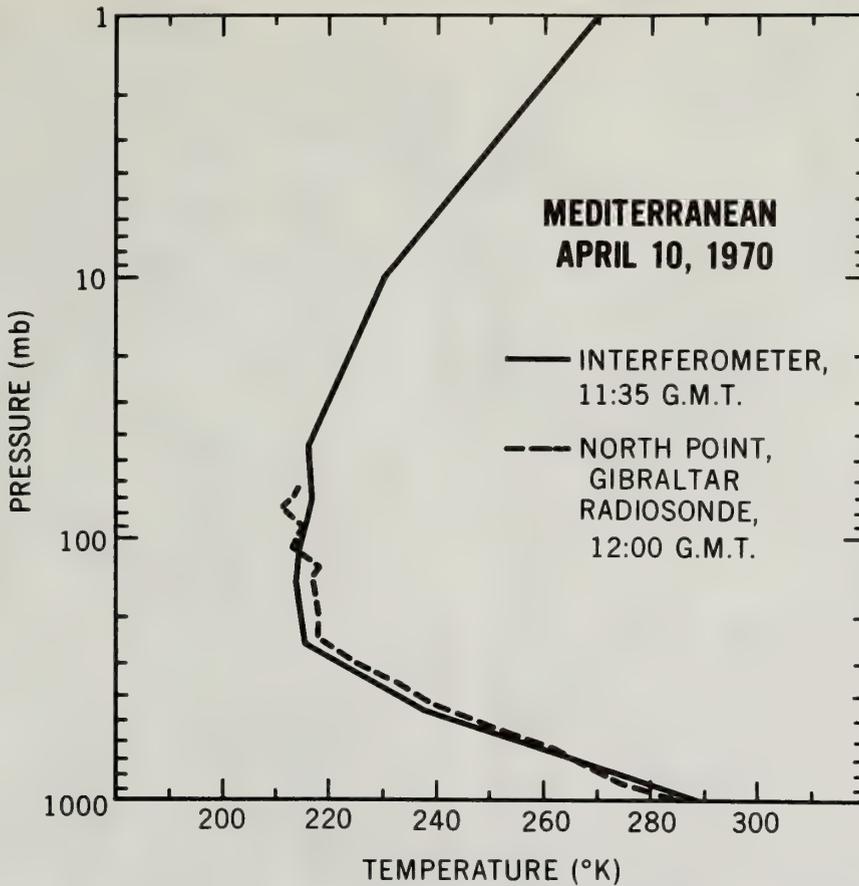


Figure 26.16 *Temperature profile for Gibraltar, from IRIS data (from Hanel, R. A. and B. J. Conrath, 1970).*

Wind fields from ATS cloud motions have been the subject of major research efforts at the University of Wisconsin, the University of Chicago and in the research activities at the National Environmental Satellite Service. At NESS the operational accuracy has been estimated at ± 3 to 6 knots in displacement routinely with ± 2 knots achievable with careful procedures. A more serious problem is in the identification of the height of the cloud being studied. The chief limitation to accuracy of wind estimate is due to uncertainty in cloud height. In general, clouds at 150 mb are indistinguishable in appearance from those at 450 mb, but in that layer winds can change by 180° in direction and upward of 35 knots in velocity. Placed at the correct elevation, a wind estimate may be accurate to 3 knots. But placed at the wrong elevation, the same measurement may imply a wind vector in error by 180° in direction and 35 kts in speed.

The addition of infrared images will significantly improve the height determination by relating indicated cloud top temperatures (for dense clouds) to specific altitudes in the temperature profile. In the near future infrared channels will be added for this purpose, and for the further purpose of providing nighttime coverage as well. Figure (26.17) illustrates the wind vectors derived for a specific time along with available wind reports from aircraft for comparison.

26.2.2 Operational Satellites

The Operational Satellites are those managed and controlled by the Environmental Satellite Service of NOAA. They differ from the research satellites of NASA in that they provide a continuous daily flow of data to specified users. Their operation must be assured by the availability of a replacement for expeditious launch in the event of a failure in orbit and each spacecraft carries redundant primary sensors so that the

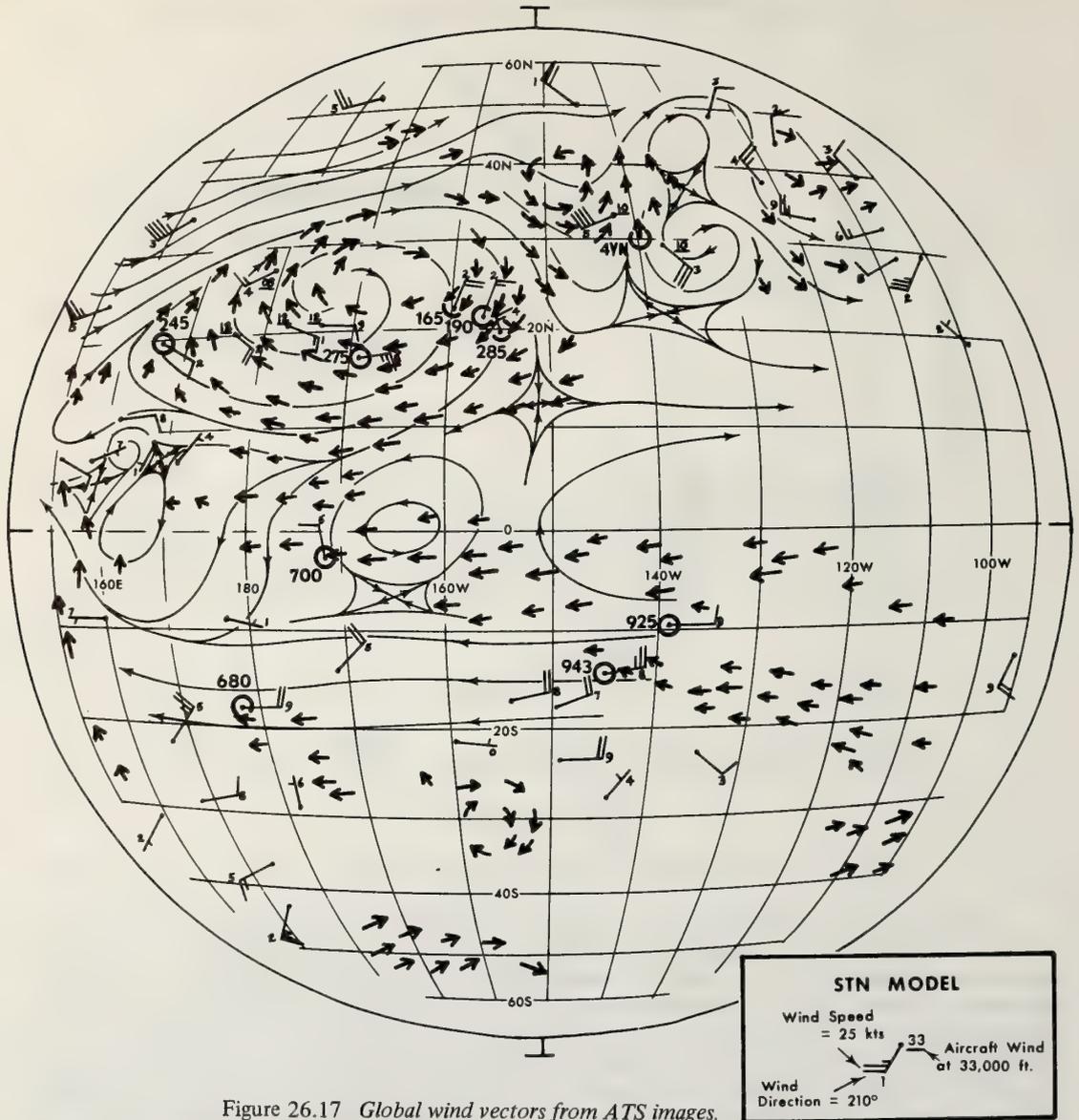


Figure 26.17 Global wind vectors from ATS images.

failure of one instrument will not constitute a failure. There are two prerequisites to the inclusion of a new sensor on the operational satellite: (1) the concept and the instrumentation must have been proven by research and development programs aboard research satellites and (2) there must be a demonstrated requirement for the data by qualified users.

The Operational Satellites started with the TIROS (Television and Infrared Observation Satellites) series early in 1960. They, and the follow on TOS (TIROS Operational Satellites)/ESSA continued until early 1970, when the first of the current generation of ITOS (Improved TIROS Operational Satellite) was launched. The TIROS and ESSA spacecraft carried, as primary sensors, vidicon cameras which took a series of overlapping snapshots which could be assembled and registered to provide a full-earth picture in the visible region (see (F26.1)). Alternate members of the ESSA series stored the pictures from each orbit on a tape recorder to be read out when passing over the NESS ground Station at Wallops Island, Virginia and Gilmore Creek, Alaska. The other members carried similar vidicon cameras but broadcast each picture in real time so that a ground station in view of the satellite could receive as transmitted. The stored

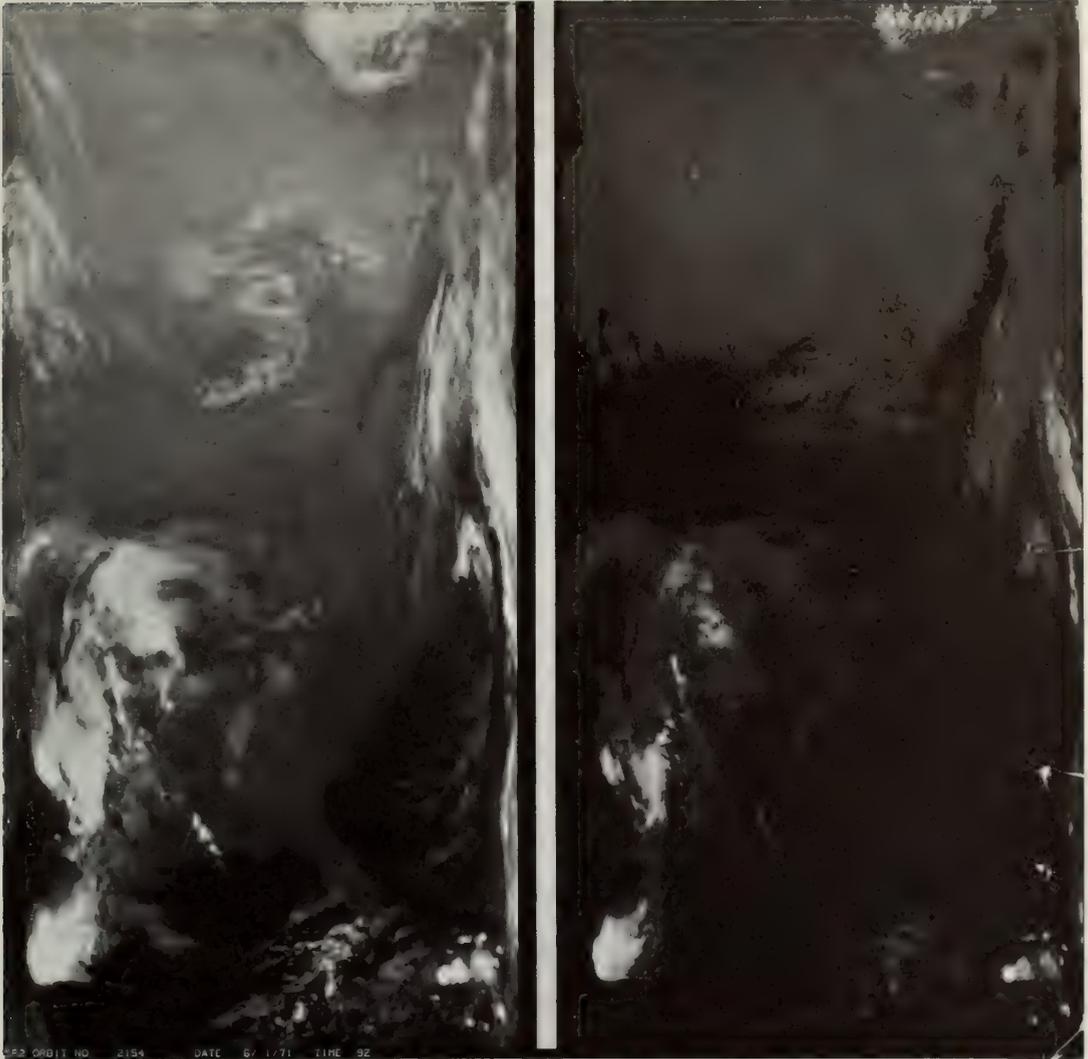


Figure 26.18 *ITOS scanning radiometer infrared image of eastern North America, June 1, 1971.*

pictures were used by NESS to provide the full-earth cloud cover pictures; the real time transmissions, called APT (Automatic Picture Transmission), were and are received by hundreds of ground stations, both professional and private, throughout the world. They provide the receiver with a real-time view of the world around him (each picture covers a square roughly 1400 km on a side) as seen from the satellite and have been of enormous value to such users as meteorologists lacking an adequate observational network and ships at sea.

ITOS

The current operational satellites are the ITOS, the first of which was launched in January of 1970. The physical characteristics and orbit parameters are given in (T26.1). The spacecraft is stabilized in roll, pitch and yaw to $\pm 1^\circ$ and carries two vidicons for stored TV pictures, two vidicons for APT (Automatic Picture Transmission), two Scanning Radiometers, one flat-plate radiometer and one solar proton monitor. The principal advance of the early ITOS satellites has been the provision of infrared images from the Scanning Radiometer.

The Scanning Radiometer (SR) is a new system based on the technology of previously proven instruments. In this device, a mirror, rotating at 48 rpm, provides an optical scan perpendicular to the direction of spacecraft motion. Energy from the single spot, which is the elemental field of view at any

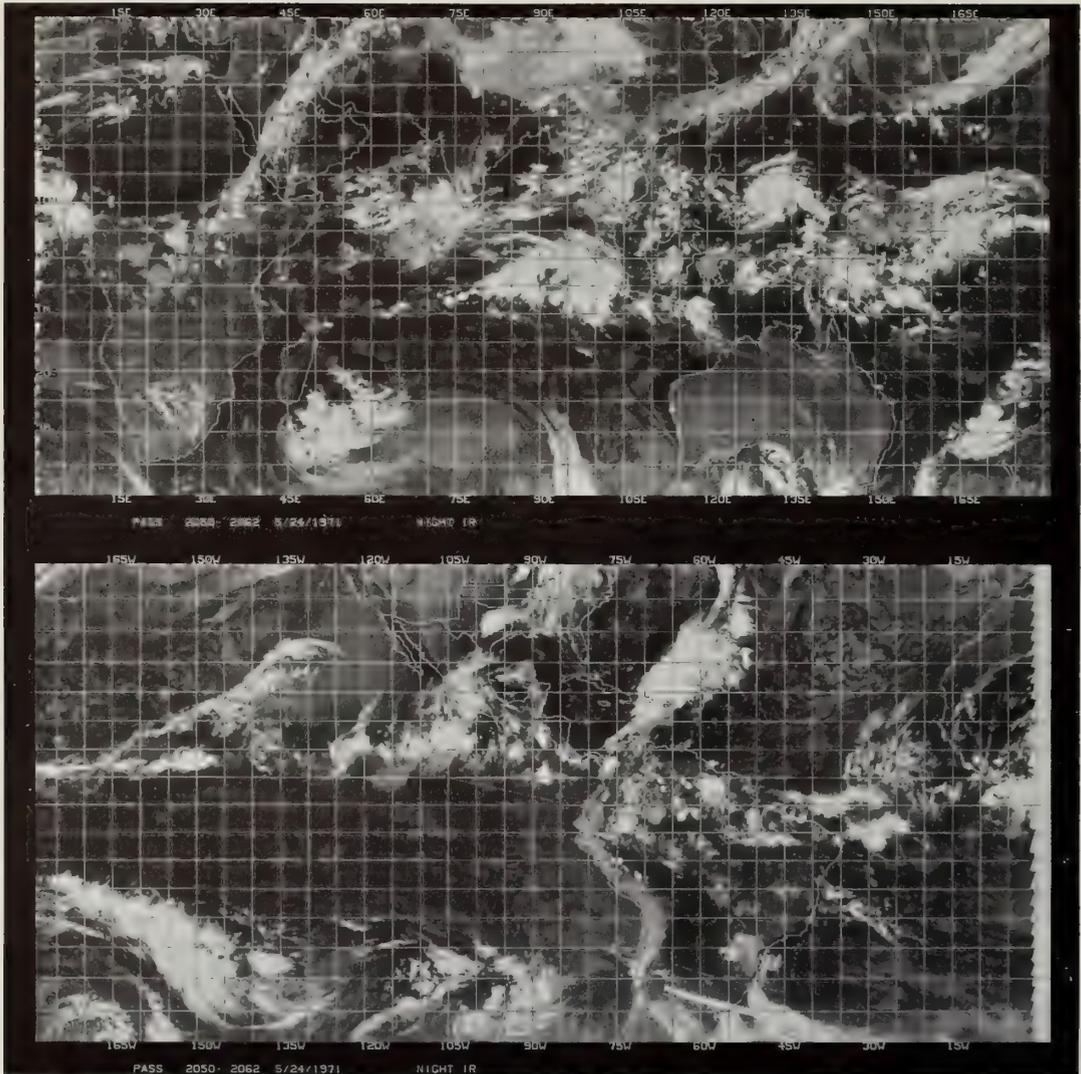


Figure 26.19 Mercator montage of ITOS infrared images, May 24, 1971.

instant, passes through a beam splitter and spectral filters to provide both a $10.5\text{-}12.5\ \mu\text{m}$ IR channel and a $0.52\text{-}0.73\ \mu\text{m}$ visible channel.

System optics provide a 5.3 mrad field of view for the IR detector. It will therefore “see” a 4 n mi spot at the nadir. As the scan mirror rotates and the “spot” moves toward the horizon, its size increases. In equatorial regions, contiguous data from successive orbits will occur at points about 900 n mi from the subsatellite point where the data zenith angle is 60° and the spot covers an area of about 8 by 12 n mi.

If data from this IR channel are interpreted in terms of the temperature of an equivalent blackbody radiator, the noise equivalent temperature difference ($NE\Delta T$), which is the smallest temperature difference the instrument can discern, is expected to range from 1K for a 300K scene to 4K for a 185K scene. Instrument stability, combined with pre-launch calibration and post-launch in-flight calibration and analysis, is expected to provide absolute temperature measurement accuracy (for an equivalent blackbody) equal to about 2K.

The visible channel optics are designed for a 2.7 mrad field of view or a 2-mi “spot” at the nadir. Since the scan advances about 4 mi per mirror rotation, visible scan lines are separated at the nadir. However, they expand to fill the scene at the sides where orbit-to-orbit contiguity occurs. This sensor geometry allows higher effective resolution to be obtained than would otherwise be possible. Due to its

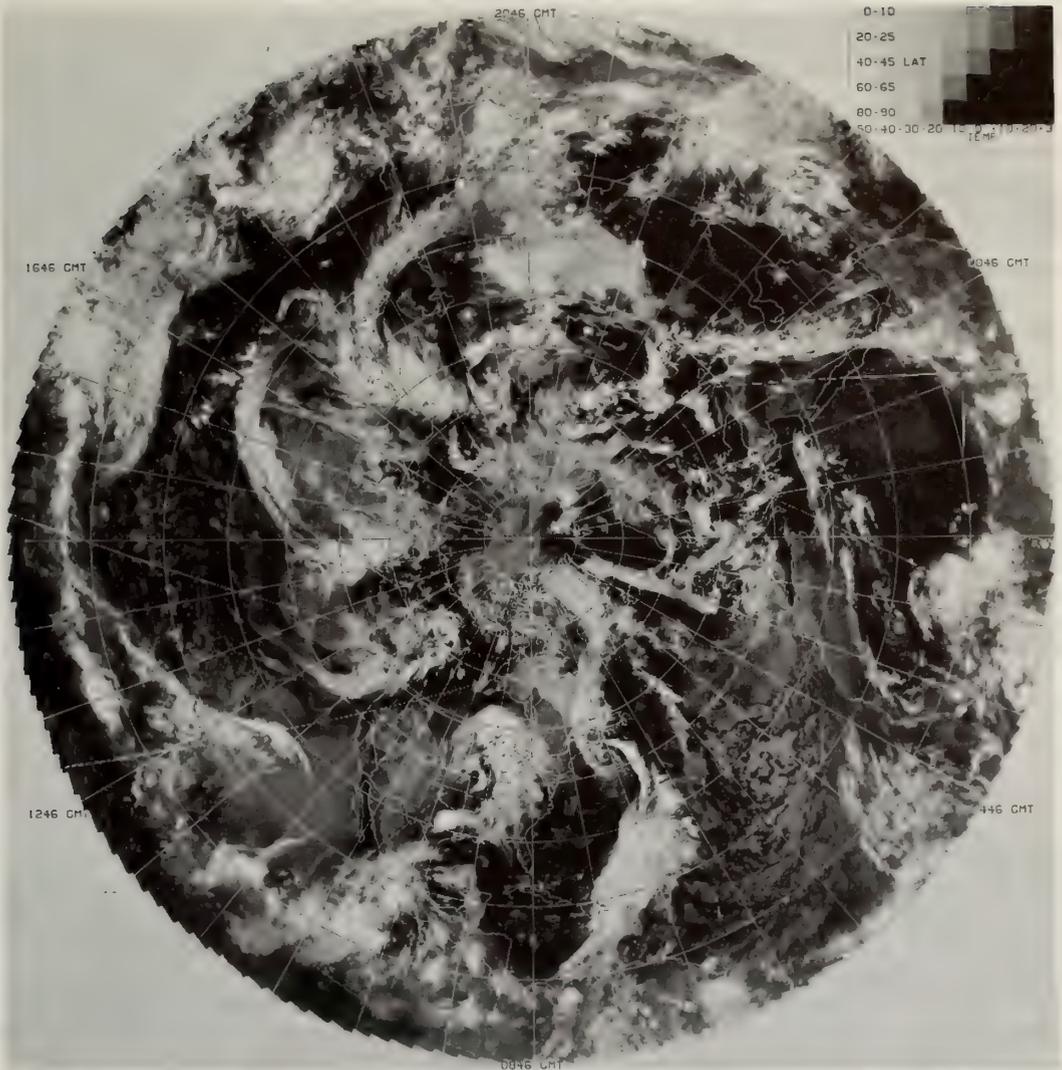


Figure 26.20 Northern hemisphere, polar stereographic montage of ITOS infrared, May 24, 1971.

more favorable modulation transfer function and dynamic range, SR visible data are expected to be of better quality than APT or AVCS pictures.

The scanning radiometer IR channel provides the only ITOS primary sensor data that will be usable at night. It is equally usable for daytime observations as is the visible channel. This instrument versatility is combined in the ITOS with a flexible data programming system so that the SR becomes a nearly universal sensory subsystem. Since the SR will be capable of providing completely usable data over the daylight side of the earth, it will function as an operational replacement for both camera systems in the event of their total failure. Figures (26.18), (26.19) and (26.20) illustrate the data being provided from the infrared channel of the scanning radiometer. Figure (26.18) shows the direct analog read out over the eastern part of the North American continent on the left and as seen sliced to 5 gray scale levels on the right. Figure (26.19) is a montage similar to (F26.1) and (F26.20) is a polar stereographic presentation of the Northern Hemisphere. These figures are examples of the product of ITOS infrared images as prepared by NESS.

For direct readout to local APT stations, IR data may be broadcast continuously during the 70-min nighttime period, and either IR or visible channel data may be broadcast continuously in place of APT video during the 45-min day portion of the orbit. For global observations (data recorded on magnetic tape and replayed through CDA stations to global weather analysis centers), IR data may be recorded during the

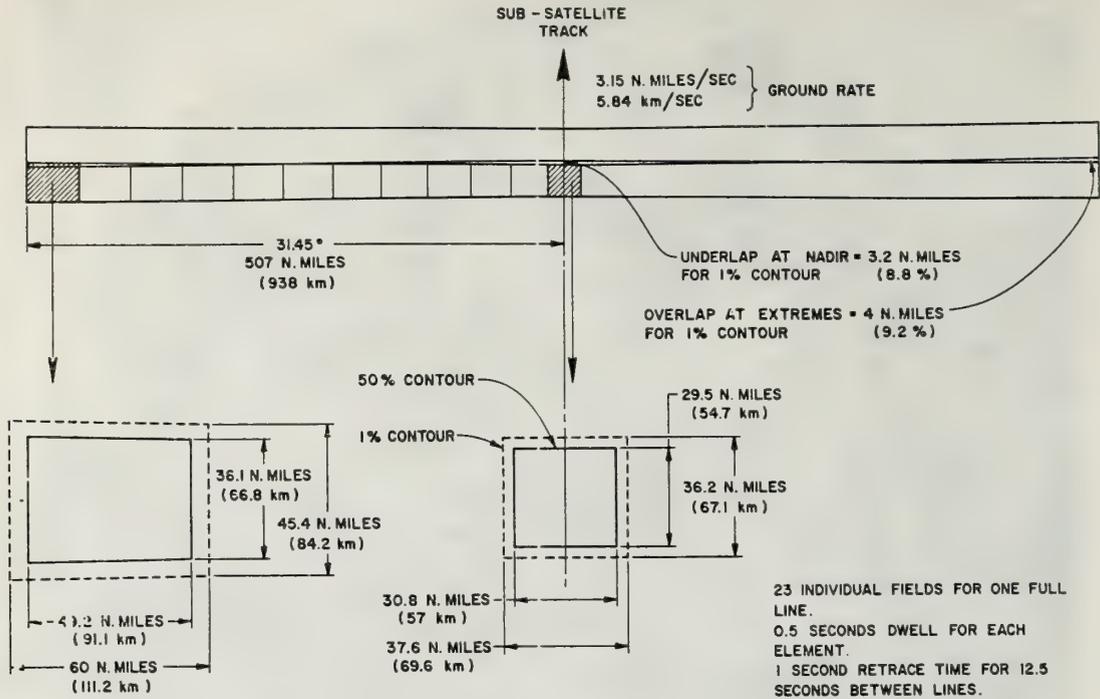


Figure 26.21 VTPR ground element and scan pattern.

day or night portions of each orbit. Daytime visible channel data may also be recorded. TIROS M and early ITOS recorder capacity limitations will prevent full day and night recorded data recovery when CDA contacts are missed, but AVCS cloud picture coverage requirements can be fully met.

A special feature of the ITOS SR is its broadcast for 94 sec of the 102 sec between APT transmissions. In this way about 75 lines of IR data can be provided with each APT frame. The area covered by IR is near the center of the next APT picture.

The FPR (Flat Plate Radiometer) and the SPM (Solar Proton Monitor) are considered to be secondary sensors. Only one of each is carried on the ITOS. Although their data are to be used operationally, their failure will not be deemed a mission failure.

The FPR measures energy in broad spectral bands with hemispheric angular acceptance. The sum of energy emitted by the earth and solar energy reflected by the earth is measured by black sensors sensitive between 0.3 and 30 μ . White sensors measure earth-emitted energy between 4 and 30 μ only. These measurements are used to determine the heat balance of the earth.

Two pairs of sensors have been used on the ITOS FPR. One black-white pair will operate as radiative equilibrium detectors in the same manner as the heat balance radiometers used on TOS. Data from this pair are expected to be closely comparable to data observed on the TOS flights. The two sets of data are expected to be sufficiently similar in calibration that they may be considered together for climatological purposes.

The second black-white sensor pair is of a new thermal feedback design. The energy required to maintain a constant temperature is measured. Except for calibrated losses to the mounts, energy (not its fourth root, temperature) is measured directly. As an added feature, this thermal feedback sensor pair rotates to view its housing for in-flight calibration of IR response.

The sensor measurements and auxiliary calibration data are recorded on a digital tape recorder for relay through CDA stations to the National Environmental Satellite Service for processing and study.

The SPM measures the flux of energetic particles (protons, electrons, etc.) in several energy ranges. The primary function of the SPM is to detect the arrival of energetic solar protons in the vicinity of the earth. The SPM measures the energetic particles flux in several ranges—protons in the 10, 30 and 60 MeV

range and electrons in the 100-750 keV range. The flux toward the earth along the vertical is measured with one set of detectors. A second set measures the flux along the orbit normal, from the side of the spacecraft away from the sun.

The fourth member of the ITOS series, ITOS-D due to be launched in July 1972, and the three or four following spacecraft will have several significant advances over the earlier members. In addition to the scanning radiometers, this spacecraft will carry a Very High Resolution Radiometer (VHRR) which is operationally very similar to the Scanning Radiometer but has a much higher resolution. The field of view in both the visible and infrared channels will be 0.6 mrad, which is an order of magnitude higher resolution than the SR. The amount of data collected by the VHRR exceeds the on-board storage capabilities of ITOS so that it will not be used to provide the full earth pictures as are the SR data. The VHRR will be broadcast in real time as APT and will be capable of being stored for limited areas of the earth for special events.

The visible channel, with a silicon detector, will respond to radiation of wavelengths between 0.55 and 0.75 μm at a signal-to-noise ratio of at least 4:1 when viewing a field whose albedo is 0.5%.

The IR channel detects radiation of wavelengths between 10.5 and 12.5 μm . The detector will be radiatively cooled to a temperature of 90°K, yielding an NE Δ T of 30°K for a source temperature of 185°K, and NE Δ T of 0.6°K for a source temperature of 300°K.

The output of this instrument will be used to determine the distribution of cloud cover, and to provide a map of cloud top temperatures and surface temperatures (in the cloudless areas). A dynamic range of 160:1 will be preserved in the transmission of the output data from each channel. This will permit making a visual display of the data with an eight-level gray scale.

It is important to note that on the ITOS D spacecraft, the vidicon systems have been completely replaced by the two scanning radiometers, the SR and the VHRR. The vidicons suffer a number of problems which have made it virtually impossible to quantitatively interpret their data. Among these are: (1) shading or variation of response across the sensitive area and with time; (2) susceptibility to magnetic or electrostatic interference; and (3) drift in black level or compression of dynamic range. The scanning radiometers, with a single detector, are more stable, are capable of onboard calibration, and have a much wider dynamic range.

Another important advance with ITOS D will be the initiation of temperature soundings on an operational basis. Based on the successful SIRS, IRIS and SCR experiments on Nimbus (27.2.2) the VTPR (Vertical Temperature Profile Radiometer) will measure the radiation emitted by the CO₂ and H₂O vapor in the atmosphere in 8 spectral bands. Temperature profiles will be computed from these radiances as described in 27.2.2 and in Chapter 16 of this text. An advanced feature of the VTPR is its scan pattern. Whereas SIRS, IRIS and SCR stared at the nadir below Nimbus (SIRS-B did have limited coverage to both sides of nadir), VTPR will scan a continuous swath nearly 2000 km wide with an instantaneous field of view of 67 km square at nadir. The scan pattern is illustrated in (F26.21). In addition to the obvious increase in geographical coverage, this scan will provide a means of correcting for scattered clouds in the field of view by a statistical comparing of many contiguous instantaneous fields (Smith, 1968). A single temperature profile will be produced from the processing of between 80 and 100 instantaneous looks.

GOES (Geostationary Operation Environmental Satellite)

The spin-scan cameras on the ATS 1 and 3 satellites have been so fruitful that an operational Geostationary Satellite system is under development. There will be four satellites in the full system, with two currently planned. They are called SMS (Synchronous Meteorological Satellite) and are scheduled for launch in 1973 and 1974. The principal sensor aboard is the VISSR (Visible Infrared Spin Scan Radiometer), a large dual channel radiometer with a primary aperture of 16" diameter. The radiometer looks out the side of the spacecraft which spins on its axis at 100 rpm. On each rotation the VISSR scans one line across the earth and then steps down one line before the next rotation. The radiometer has a visual channel at 0.54 to 0.7 μm and an infrared channel at 10.5 to 12.6 μm . The infrared field of view is 0.2 X 0.2 mrad (8 km spot size on the earth) and the visible field of view of 0.025 mrad achieved by a linear array of 8 detectors. This is possible because of the much greater sensitivity of visible detectors and the fact that resolving power for the diffraction-limited optical system is 20 times smaller in the visible region than in the infrared.

A cut away drawing of the VISSR is shown in (F26.22). The spacecraft orientation and the scan pattern are illustrated in (F26.23) and the VISSR specifications are listed in (T26.5). The sensitivity needed in the infrared channel of this instrument requires the use of a HgCdTe detector which must be cooled to < 100 °K. For an operational spacecraft, mechanical refrigerators or stored, solid cryogenes are impractical and a radiation cooler is used. In this device a black target patch is restricted to viewing cold space by a highly-reflecting dual cone assembly pointing out the top of the spacecraft. The patch, to which the detector is mounted, cools by radiating to space which is essentially at 0 °K. The temperature achievable is dependent upon the residual heat leaks in the mechanical support for the patch and the reflectance of the cones. Temperatures down to 80 °K have been achieved in vacuum chamber tests in the laboratory.

26.3 Trends in Research and Development

26.3.1 Passive Sensors.

Passive sensors are, and will continue to be in the near future, the most important for spacecraft application. Active sensors are required for many desirable measurements—particularly where a range determination is necessary—but the power requirements of active systems exceed the capabilities of most practical satellites. This is especially true of the operational satellites. Active current research efforts are aimed at determining the concentration of atmospheric constituents, both gaseous and particulate, and in some cases the vertical distribution of concentration. Limb scanning and absorption spectroscopy are primary techniques being employed and the microwave region is receiving increased attention.

Limb Scanning

Limb scanning from a satellite involves observing the horizon of the earth either in emission or transmission and deducing temperature or constituent concentration profiles from the observations. The technique is essentially limited to the stratosphere of the earth because cloud interference below the tropopause becomes too severe in most instances. Microwave observations might be considered to reduce the cloud interference, but achieving acceptable vertical resolution with a microwave radiometer would require an enormous and impractical antenna diameter.

A good example of current activity in this area is the Nimbus-F Limb Radiance Inversion Experiment, for which Professor John Gille of the Florida State University, is the Principal Investigator. This experiment is aimed at determining temperature profiles from observations of emission in the 15 μm CO_2 band and O_3 and H_2O concentration profiles from the measurement of radiances in absorption bands of these gases. First the temperature must be determined from CO_2 radiance because this gas is present in a uniform, stable concentration of 320 parts/million and hence emissivities can be calculated. Using 2 channels in the CO_2 band (14.0 to 16.3 and 14.0 to 16.7 μm), the experiment is predicted to be capable of an accuracy of ± 3 K from 15 to 40 Km and ± 6 K from 25 to 60 km with a 2 km vertical resolution. Then, using the temperatures so derived, a 9.4 to 10.0 μm channel will be used to determine O_3 density to an accuracy of 10^{-3} cm NTP/km over the height range 15 to 50 km with a vertical resolution of 2 km. H_2O vapor will be measured with a 20-40 μm channel to ± 1 part/million from 15-35 km and ± 2 ppm from 35-42 km with a 5 km vertical resolution. It must be remembered that these figures, taken from the technical proposal for this instrument, are predictions of performance.

Another limb scanning technique is to measure absorption using as a source the sun, the moon, or an emitter carried aboard a second satellite. This technique is capable of much greater precision than observing limb emission, particularly if the sun is used as a source, but the coverage of the earth's atmosphere is severely restricted. Observing from a polar orbiting satellite, the sun rises and sets only near the poles, and equatorial coverage could not be possible. For solar occultation measurements an inclined orbit like that of Apollo (26.1.2) is the best. Theoretically any gas having a clear absorption band unobstructed by bands from other constituents could be measured by absorption spectroscopy on the limb. The technique is being evaluated for potential application to the measurement of H_2O and pollutants such as CO , N_2O and NO_2 in the stratosphere. The principal problems to be investigated are the effects of refraction (including the strong variation of refractive index through an absorption band) and the ability to predict or relate laboratory values of absorption strengths for stratospheric conditions. There have been some experimental programs whose results indicate great promise for the technique. The British

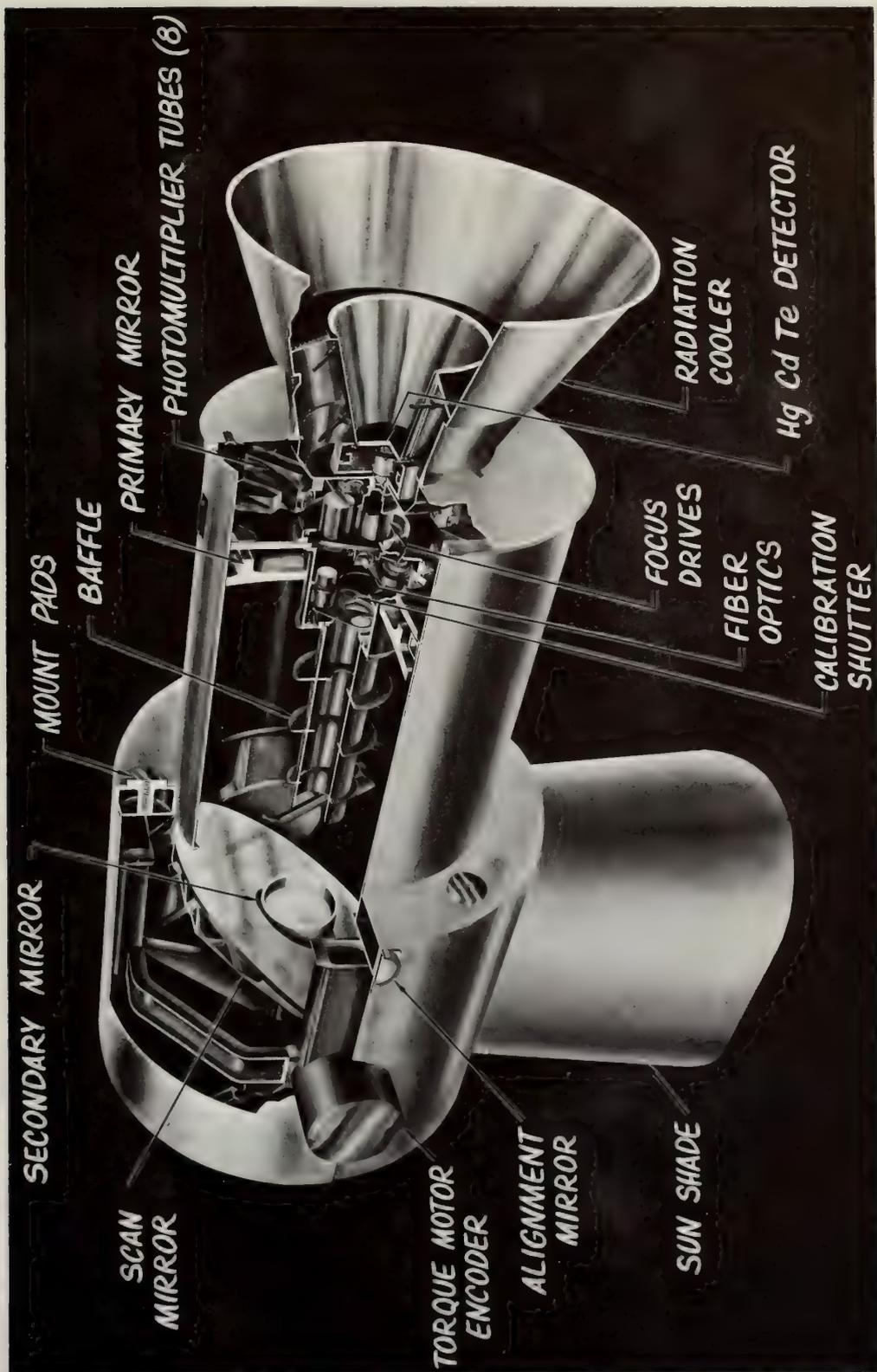


Figure 26.22 Cut-away drawing of VISSR instrument (courtesy of Santa Barbara Research Center).

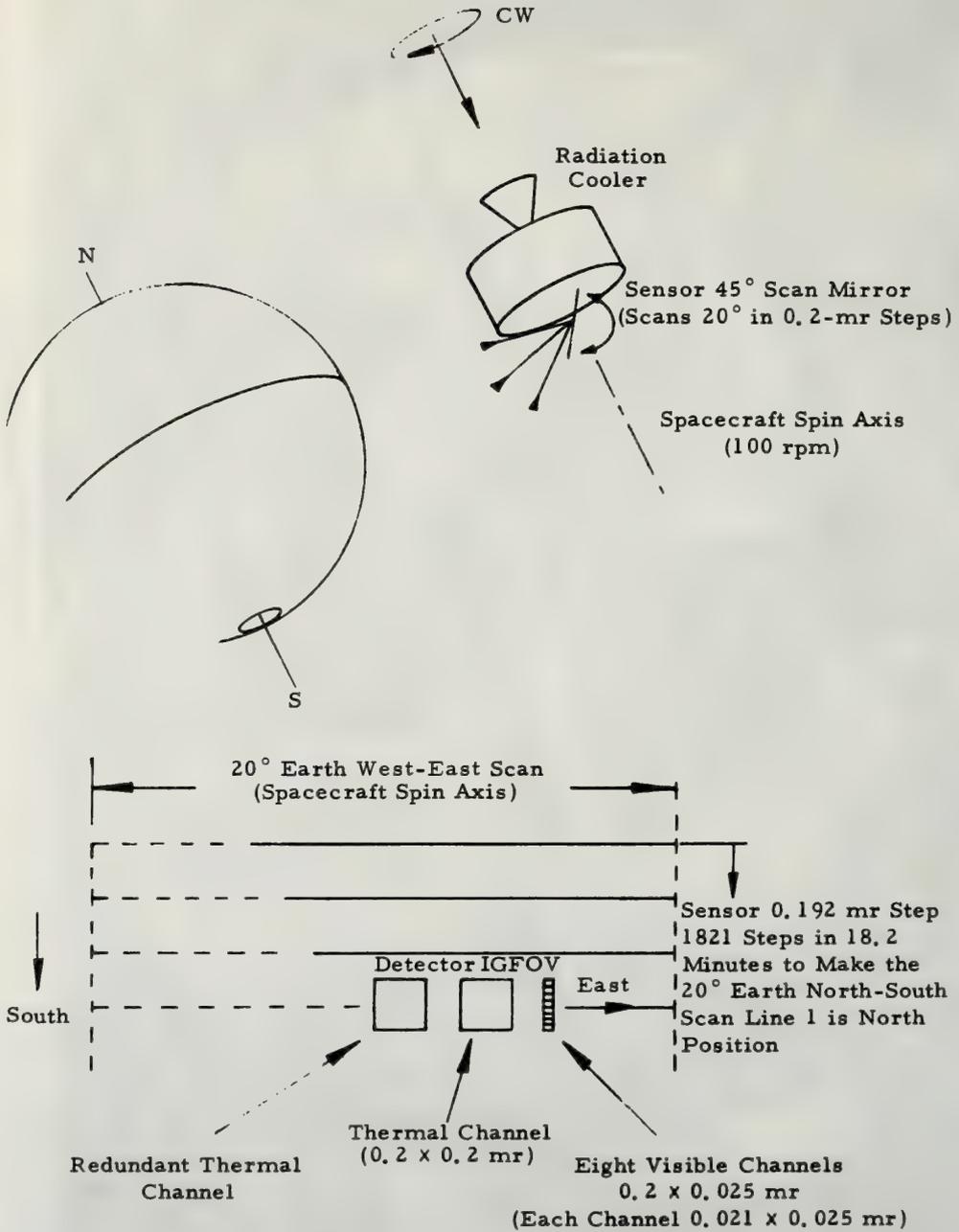


Figure 26.23 SMS/GOES spacecraft orientation and VISSR scan pattern (courtesy of Santa Barbara Research Center).

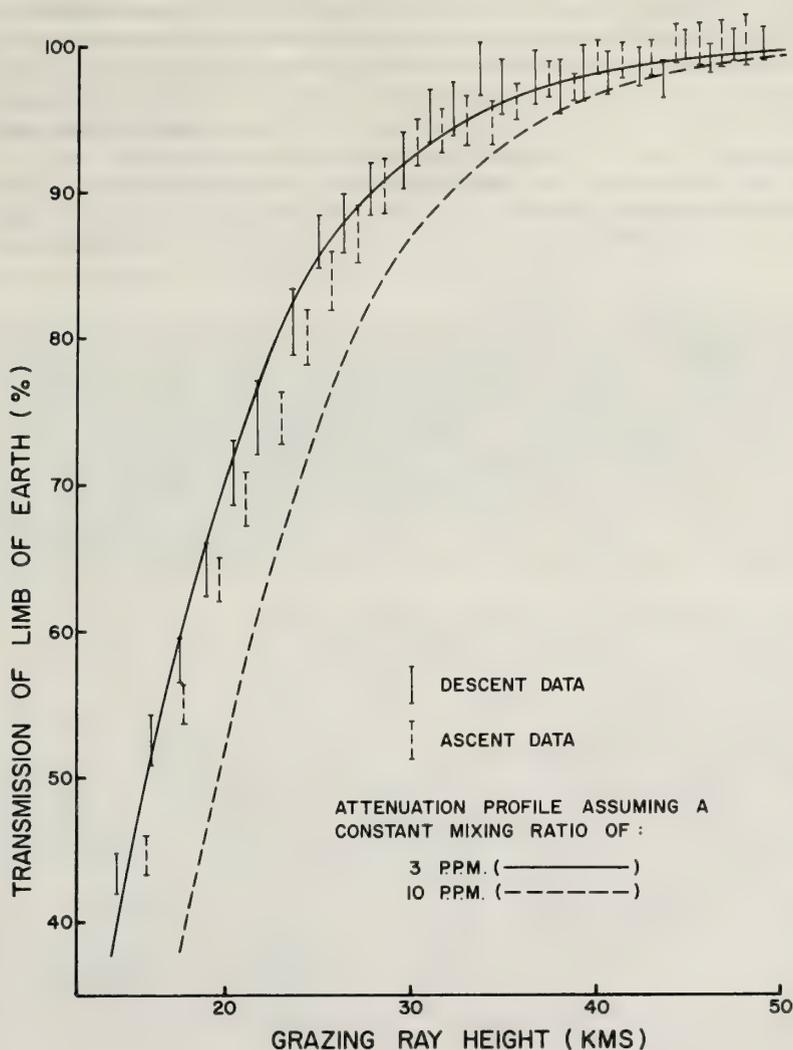


Figure 26.24 Stratospheric H_2O measurement from solar occultation measurement.

Meteorological Office (Dr. Miller—private communication of unpublished data) has flown a spectrometer on a Skylark sounding rocket at the Woomera Range, Australia, and observed the sun through the earth's limb. Water vapor concentration was calculated from the measured absorption at 3830 cm^{-1} with a 15 cm^{-1} bandwidth. The second channel was situated in a region of no absorption and served as the normalizing signal. The measured points from a flight during the summer of 1970 are shown plotted in (F26.24), along with computed contours for 3 and 10 parts/million of water vapor. The indicated 3 ppm average is in good agreement with the 2 to 4 ppm results of balloon sampling observations. With refined instruments and techniques, it is estimated that a precision of $\pm 0.1\text{ ppm}$ could be achieved in such a measurement of stratospheric water vapor.

Absorption Spectroscopy

Absorption spectroscopy is a powerful quantitative analytical technique. Its application from a satellite depends upon the existence of a suitable source of radiation. In the preceding section an example was given using the sun as a source. Looking down from a spacecraft, earth-bound discrete sources could be

used for limited coverage or the radiation emitted and reflected by the earth could be used as a source. Barringer, 1966, has applied this technique to the measurement of SO_2 from a balloon. A correlation spectrometer looks at the earth and responds to the presence and quantity of absorber in the intervening atmosphere. The absorption bands of SO_2 in the ultraviolet region were used. Working in the thermal infrared would not be possible because a good temperature contrast between the earth (source) and atmosphere (sample) is required and in the natural environment the earth can be either warmer or colder than the air, has a small difference in either event, and can fluctuate rapidly spatially.

Barringer uses a separate optical channel to observe fluctuations in the source intensity as the instrument scans along over the terrain below it. There are, of course, limitations to the technique—it measures total burden of absorber and cannot discern vertical distribution—but it appears from limited tests to be an effective and promising tool. Balloon measurements around Chicago agreed with ground measurements to $\pm 20\%$ over much of the flight.

An interesting application of absorption spectroscopy to other purposes—specifically the measurement of cloud top altitudes—is found in the CAR (Cloud Top Altitude Radiometer) described by Saiedy, Hilleary and Morgan (1965). The instrument observes, from a spacecraft, the brightness of a cloud top in three wavelength regions, two at 0.763 and $0.7600 \mu\text{m}$ in the absorption region of oxygen and one just outside the band at $0.765 \mu\text{m}$. The ratio of in-band to out-of-band intensities (two in-band frequencies are employed to extend the available dynamic range) is related to the total amount of oxygen traversed by the radiation in travelling from the sun down through the atmosphere to the cloud and back through the atmosphere to the spacecraft. The total oxygen traversed is, in turn, a function of the altitude of the cloud. The method was tested from the Gemini V spacecraft using a hand-held photographic spectrometer to measure selected cloud

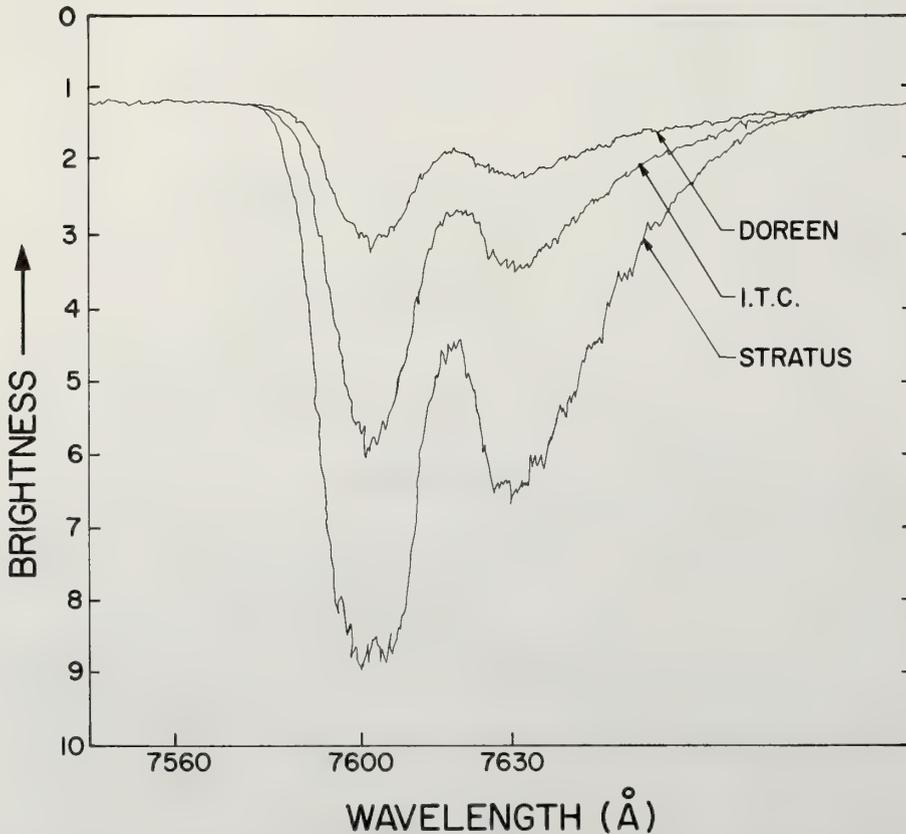


Figure 26.25 CAR evaluation, Gemini V flight.

formations. Three spectra, above low-lying stratus, intermediate inter-tropical convergence clouds and the high altitude top of Hurricane Doreen, are shown in (F26.25). The conversion of absorptance to height requires a knowledge of the incidence and observations angles and, for a sharp, well-defined cloud top can yield an altitude accurate to 100 meters. Unfortunately if the cloud top is ill defined or the cloud optically thin, a significant amount of the radiation received comes either from below the cloud or after multiple reflections within the cloud and both components make the cloud look lower than it actually is, necessitating the derivation of a correction factor. For thin high cirrus, the most troublesome cloud, this correction factor can exceed the altitude of the cloud and so it must be determined with high precision. The concept is being expanded to include infrared channels which will yield data on the cloud characteristic which will be used to calculate the correction factors. When proven, it will be considered for spaceflight testing.

Microwave Sensors

As mentioned earlier, the microwave region of the spectrum has a very desirable advantage over the visible and infrared regions: at the longer microwave wavelengths, scattering by aerosol and cloud particles is orders of magnitude reduced and in many cases is negligible. It has been demonstrated that passive microwave and radar imaging systems can function satisfactorily through very dense cloud cover. Of course, the effects of aerosol and clouds, though small, are still not zero and increase with particle size to the point where raindrops do have a significant effect – otherwise a weather radar could not function. An important part of current microwave research for satellite application is involved with investigating the quantitative atmospheric effects and their bearing on accuracy in atmospheric temperature and constituent sounding with microwave radiometers. The experiments aboard Nimbus E and F, (26.2.1) illustrate the trend of the present efforts.

A very desirable use of microwave radiometers would be to provide a measure of surface temperature through cloud cover. Such a measure, aside from the direct value to oceanographers, etc., is needed for the measurement of temperature profiles by inversion techniques. Unfortunately liquid water has a very high reflectivity – between 50 and 60% – throughout the useful microwave region and this causes serious problems. Radiation received from open water has two primary components: (1) that emitted by the water at an emissivity of about 40% and (2) the sky and atmosphere reflected by the water at about 60% reflectivity. The indicated temperature of the sea – whose actual temperature may be 290 °K – will range from perhaps 140 °K on a clear day when the apparent temperature of the sky is low, around 100 °K, to above 200 °K for an overcast day when the sea mirrors the warmer clouds. The apparent temperature is also influenced by sea state because a rough sea mirrors more of the warmer horizon than a calm sea and the emissivity goes up slightly due to foam and trapping. Corrections can be applied for these effects, but they are so large that the resultant accuracy is not very promising. Over land, dry soil has a high emissivity and yields a good radiometric temperature, but the presence of even small amounts of moisture reduces the apparent temperature by increasing the reflectivity.

The synergistic benefits of combining infrared and microwave radiometers is under consideration. Infrared systems give a true radiometric temperature for most situations because the emissivity is high, >90%, for the majority of land features and water as well. However, infrared systems are seriously affected by clouds and aerosol. Microwave systems are fairly free of cloud and aerosol interference, but sensitive to the nature of the target. Combined systems should provide results unattainable with either alone.

26.3.2 Active Sensors

Laser Probing

Active systems are in general the only means of supplying a measure of range and hence the only technique for providing constituent profiles of the atmosphere from a spacecraft. There is an exception in the case of deducing temperature profiles from radiance measurements in the CO₂ band. Here temperature is related to pressure, assuming the CO₂ to exhibit an exponential concentration profile and this in turn relates to altitude when a pressure reference – the barometric surface pressure for example – is known. This is, therefore, a passive technique with a range capability.

Lasers are enormously powerful probing tools and are being exploited with great success in remote sensing applications. Chapter 23 deals with such systems immersed in the atmosphere. Unfortunately when removed from the target (atmosphere) to satellite orbit distances – 800 to 200 km for most polar and inclined orbits – the sensitivity of Resonance, Raman and even stimulated emission backscatter systems is below practical levels. Resonance scattering may be an observable phenomenon above 100 km (at higher

pressures reradiation is quenched by collisions between molecules) but there is no particular need or application for such measurements in this region.

One area of potential promise for LIDAR (Light Detection and Ranging) from spacecraft is in the measurement of particle size distribution and concentration. There is great interest in this information for clouds and aerosols but no satisfactory means of collecting such data on a large scale. By using lasers covering a wide range of wavelengths — an ultraviolet laser would be appropriately scattered by aerosol particles but stopped by a dense cloud, whereas a laser at 100 micrometers would respond appropriately to a cloud while being essentially unaffected by the aerosol particles — particle size and concentration may be measurable under some conditions from spacecraft. Power requirements and safety regulations will always be serious considerations for any such system, however.

Radar

Radar has limited application from spacecraft both because of the power requirements and the size of antennas needed to achieve acceptable angular resolution. One application of promise, however, is the radar scatterometer for the measure of sea state. As a radar scans out in angle from the nadir, the return echo falls off very rapidly to essentially zero for a calm, specularly reflecting ocean. As the surface becomes more rough, the fall-off becomes less pronounced (i.e. the back-scatter increases for rough water) and the rate of fall-off with angle is relatable to sea state. The technique has been demonstrated from aircraft and proposed for spacecraft test but has not been flown to date.

Separated Source and Receiver

Several systems employing multiple satellites have been proposed for studying the atmosphere but none have been implemented. In general they involve a source of radiation on one satellite (or on several satellites in an array) and a receiver on a following satellite whos orbit is chosed so that the source satellite is viewed through the atmosphere. It is a powerful method but the expense of multiple satellites and the complexity of maintaining the orientation of the satellites reduce the practicality of the method significantly.

26.4 Compilation of Electronic and Mechanical Imaging Devices for Spacecraft Applications

Tables 26.6 and 26.7 summarize the characteristics and capabilities of the large variety of imaging devices which have been and will be carried on planned space missions. They are taken from NASA SP-8078 "Spaceborne Electronic Imaging Systems" dated June 1971.

Table 26.6 Image Tube Cameras

	Spacecraft	First launch date	Orbit	Application	Sensor	Horizontal resolution (TV lines)	Frame time, sec	Video bandwidth, kHz
TIROS	TIROX I-IX	4-1-60	550 km (300 nmi) inclined	Visible cloud cover	1.27-cm (½-in.) vidicon EM defl and foc	400	2.0	62.5
AVCS	Nimbus	8-28-64	1,110 km (600 nmi) sun-synchronous (noon)	Visible cloud cover	2.54-cm (1-in.) vidicon EM defl and foc	800	6.5	60.0
	TOS/ESSA		1,390 km (750 nmi) sun-synchronous (3 pm)					
	ITOS		1,460 km (790 nmi) sun-synchronous (3 pm)					
	Applications Technology Satellite (ATS)		11,120 km (6000 nmi) equatorial					
APT	TIROS	12-21-63	740 km (400 nmi) inclined	Visible cloud cover	2.54-cm (1-in.) storage vidicon EM defl and foc	700	200	1.6
	Nimbus	8-28-64	1,110 km (600 nmi) sun-synchronous (noon)					
	TOS/ESSA	2-28-66	1,390 km (750 nmi) sun-synchronous (3 pm)					
	ITOS	1-23-70	1,460 km (790 nmi) sun-synchronous (3 pm)					
Ranger	Ranger 3-5	-62	Translunar to impact	Lunar terrain	2.54-cm (1-in.) vid-ES	200	10	2.0
	Ranger 6-9	-65			2.54-cm (1-in.) vid-EM	700	2.56	200.0
Photo di-electric tape	Nimbus	Not flown	1,110 km (600 nmi) sun-synchronous (noon)	Visible cloud cover	35 mm D/C Tape	236.2/cm (600/inch) @ 50% response	3.25	650
OAO	Orbiting Astronomical Observatory	4-8-66	800 km (430 nmi) inclined	Telescope control	2.54-cm (1-in.) vidicon ES defl and foc	350	1.0	60.0
Apollo	Apollo 7, 8	-68	Translunar and lunar	Black and white real-time	2.54-cm (1-in.) hybrid vidicon	220	0.1	500.0
	Apollo 9, 11	-69		Black and white real-time	SEC hybrid	220	0.1	500.0
	Apollo 10, 11, 12, 13	-69		Color real-time	SEC-EM	360	0.033	4500 ^a

^a Limited to 1800 by communications link.

Table 26.6 – (continued)

	Spacecraft	Dynamic range		SNR (peak signal rms noise), dB	Optics	Field of view		Exposure, ms	Weight		Power, W	Contractor	Reference	
		lumen/m ² -s [cd/m ²]	ft-cd-s [f-L]			rad	deg		kg	lb				
TIROS	TIROX I-IX	0.22–10.76	0.02–1.0	40	5 mm f/1.5	1.832	105	1.5	4.717	10.4	11.6	RCA	2	
AVCS	Nimbus	0.04–4.31	0.004–0.4	32	17 mm f/4	0.873	50	40.0	8.256	18.2	21.0	RCA	3	
	TOS/ESSA			32	5.7 mm f/1.8	1.885	108	1.5	8.528	18.8	16.0			
	ITOS			32	5.7 mm f/1.8	1.885	108	1.5		18.8	16.0			
APT	Applications Technology Satellite (ATS)	0.11–7.53	0.01–0.7	40	200 mm f/4	0.052	3	40.0	10.297	22.7	31.0		4	
	TIROS			32	5.7 mm f/1.8	1.885	108	1.5	10.070	22.2	15.0	RCA		
	Nimbus								40.0					
	TOS/ESSA								1.5					
Ranger	ITOS							1.5						
	Ranger 3-5	0.11–3.23	0.01–0.3	36	1016 mm f/5.6	0.009	0.5	20.0	2.722	6.0	5.2	RCA	5	
	Ranger 6-9	0.04–7.32	0.004–0.68		76.0 mm f/2	0.209	12	5.0	7.893	17.4	30.7			
Photo di-electric tape	Nimbus	0.03–2.15	0.003–0.2	40	125 mm f/3.8	Panoramic		35.0	34.020	75	20 Write 25 Read	RCA	6	
OAO	Orbiting Astronomical Observatory	7th – 2nd magnitude stars		43	76 mm f/0.87	0.209	12	No shutter	9.526	21.0	9.0	RCA	7	
Apollo	Apollo 7, 8	0.03–322.92	0.003–30	43	8 mm f/2.0	1.396	80	No shutter	2.041	4.5	6.0	RCA	8	
	Apollo 9, 11	0.003–13,557.60	0.0003–1260	40	25 mm f/4	0.611	35		3.289	7.25	6.5	West-ing-house	9	
	Apollo 10, 11, 12, 13	0.323–4,304.00	0.03–400	43	25–150 mm (zoom)	0.122–0.733	7–42		5.897	13.0	16.0	West-ing-house	10	

Table 26.6 – (continued)

	Spacecraft	First launch date	Orbit	Application	Sensor	Horizontal resolution (TV lines)	Frame time, sec	Video bandwidth, kHz
RAE	Radio Astronomy Explorer (RAE)			Boom tip observation	1.27-cm (½-in.) vidicon EM defl and foc	—	13.11	2.5
RBV	Earth Resources Technology Satellite (ERTS)	(estimated) 1972	930 km (500 nmi) sun-synchronous (10 am)	Earth surface observation	5.08-cm (2-in.) RBV	4500	3.5	3000
Image Orthicon	ATS 4	-1968 (did not achieve orbit)	35,760 km (19,300 nmi) geostationary	Day-night cloud cover map	5.08-cm (2-in.) image orthicon	800	6.75	60
Image dissector	Nimbus	4-19-69	1,110 km (600 nmi) sun-synchronous (noon)	Visible cloud cover	2.54-cm (1-in.) image dissector S-11 surface	800	200	1.6
	ATS	11-5-67	35,760 km (19,300 nmi) geostationary			1300	13.3 min	28.0
NRL	Aerobee Rocket	11-4-69	Suborbital	X-UV solar flare observation	SEC tube EM defl and foc	450	0.033	4500
Surveyor	Surveyor	5-30-66	Lunar landing	Lunar terrain photography	2.54-cm (1-in.) hybrid vidicon	600 200	1.2 20.6	220 1.2
Mariner	Mariner 4	11-28-64	Mars flyby	Mars observation and survey	2.54-cm (1-in.) vidicon ES	200	24	6.94
	Mariner 6, 7	2-24-69			2.54-cm (1-in.) vidicon EM	945	42.24	9.45
Uvicon	Orbiting Astronomical Observatory	12-7-63	800 km (430 nmi) inclined	UV star maps	0.0000011-0.0000032 nm (1100-3200 Å) Uvicons (4) Electrostatic	250	10.5	62
ATS	ATS 2, 4, 5	4-5-67	11,110 km (6,000 nmi) and 35,760 km (19,300 nmi)	Gravity gradient boom observation	2.54-cm (1-in.) vidicon	600	0.033	3500
DODGE	DODGE	7-1-67	35,760 km (19,300 nmi)	Attitude measurement; boom observation	2.54-cm (1-in.) vidicon	500	200	0.65
Redstone	MR-2	1-31-61	Boost phase	Booster stability and separation data	2.54-cm (1-in.) vidicon EM defl and foc	350	1 30	4.5
Saturn I	SA-5, 6, 7	1964	Boost phase	S-IV separation and ignition ^b	2.54-cm (1-in.) vidicon EM defl and foc	450	1/30	5.6
Saturn I	SA-8, 9	1965	~ 130 nmi	Pegasus deployment	2.54-cm (1-in.) vidicon EM defl and foc	450	1 30	5.6
Saturn IB	AS-202	8-25-66	Suborbit	Panel deployment	1.27-cm (½-in.) vidicon EM defl and foc	400	1 30	5.0
Saturn IB	AS-203	7-5-66	~ 120 nmi	Liquid hydrogen in orbit	2.54-cm (1-in.) vidicon EM defl and foc	550	1 30	7.0
FUTURE								
Skylab	Apollo Telescope Mount	Early 1973	235 nmi	Solar observation	Two 2.54-cm (1-in.) vidicons EM defl and foc	650	1 30	8.5
					Two SEC vidicons EM defl and foc	550	1 30	7.0

^bAlso, for SA-6, engine gimbaling.

Table 26.6 — (continued)

	Spacecraft	Dynamic range		SNR (peak signal rms noise), dB	Optics	Field of view		Exposure, ms	Weight		Power, W	Con- tractor	Re- ference
		lumen/m ² -s [cd m ⁻²]	ft-cd-s [f-L]			rad	deg		kg	lb			
RAE	Radio Astronomy Explorer (RAE)	[685.2518–27,408 cd/m ² (no shutter)]	[200–8000 fL (no shutter)]	—	5.5 mm f/1.8	1.047	60	No shutter	2.404	5.3	5.3	RCA	11
RBV	Earth Resources Technology Satellite (ERTS)	0.038–3.766	0.0035–0.35	35	126 mm f/2.8	0.279	16	12	21.773	48	50	RCA	12
Image Orthicon	(ATS) 4	[0.0003426–34260.00 cd/m ²]	[10 ⁻⁴ –10 ⁴ fL]	30	16.25 mm f/6.4	0.074	4.25	30–6630.0	30.845	68.0	50.0	Hazel-tine	12
Image dis- sector	Nimbus	[68.520–34,260.00 cd/m ²]	[20–10,000 fL]	40	5.7 mm f/3	1.571	90	No shutter	5.670	12.5	12.0	ITT	13
	ATS	[1,713.00–34,260.00 cd/m ²]	[500–10,000 fL]	—	49 mm f/2	0.255	14.6	—	9.072	20.0	20.0	—	—
NRL	Aerobee Rocket	—	—	38	80 cm f/9.3	0.031	1.8	30–2000	3.629	8.0	8.0	West- ing- house	14
Surveyor	Surveyor	[0.027–8,907.60 cd/m ²]	[0.008–2600 fL]	36	25–100 mm f/4 (zoom)	0.105–0.436	6–25	150	8.301	18.3	10.2 (without heater)	Hughes	15
Mariner	Mariner 4	0.054–6.456	0.005–0.6	40	30.5 cm f/8	0.018	1.05	200	5.126	11.3	8.0	JPL	16
	Mariner 6, 7	0.022–4.304	0.002–0.4	45	50 mm f/5.6 500 mm f/2.35	0.192 and 0.019	11 and 1.1	90–180 6–12	21.727	47.9	29.0	EOS	17
Uvicon	Orbiting Astronomical Observatory	107,600.00:1	10 ⁴ :1	16	610 mm f/2	0.035	2	No shutter	—	—	25.0	EMR	18
ATS	ATS 2, 4, 5	—	—	35	10 mm f/	1.396	80	No shutter	3.175	7	10	Lear Siegler	19
DODGE	DODGE	[685.200–34,260.00 cd/m ²]	[200–10 ⁴ fL]	25	6.5 mm f/2.5 18 mm f/2.5	1.396 0.524	80 30	1,000	8.618	19.0	9.	APL	20
Redstone	MR-2	10.7–10.7 × 10 ³	1.0 to 10 ⁴	32	6.2 mm f/1.5	—	30 × 45	33	2.73	6.0	18	Lockheed	—
Saturn I	SA-5, 6, 7	5.4–10.7 × 10 ⁴	0.5 to 10 ⁴	32	12.5 mm f/2.5	—	17 × 23.5	33	5.0	11.0	14	Lear Siegler	—
Saturn I	SA-8, 9	5.4–10.7 × 10 ⁴	0.5 to 10 ⁴	32	5.7 mm f/1.8	—	80 × 96	33	5.0	11.0	14	Lear Siegler	—
Saturn IB	AS-202	5.4–10.7 × 10 ⁴	0.5 to 10 ⁴	28	8.0 mm f/1.8	—	47 × 60	33	4.09	9.0	15	GEC	—
Saturn IB	AS-203	5.4–10.7 × 10 ⁴	0.5 to 10 ⁴	40	12.5 mm f/2.5	—	17 × 23.5	33	4.09	9.0	15	MSFC	—
FUTURE	Skylab	Two vidicons Two SEC vidicons	—	43 36	Telescope Telescope	—	—	33 33	19.2 22.7	—	18 18	MSFC MSFC	—

* General Electrodynamics Corporation.

Table 26.7 Mechanical Scanner

Type	Spacecraft	First launch date	Orbit	Application	Scan	
					Type and scan angle	Instantaneous FOV, (mrad) (spectral band designation)
High-Resolution Infrared Radiometer (HRIR)	Nimbus I, II, III	8-28-64	1110 km (600 nmi) polar sun-synchronous (noon)	Day-night cloud maps and cloud temperature	Cross-course rotating mirror (45 r/min), 2.059 rad (118 deg)	7.5
Medium-Resolution IR Radiometer (MRIR)	Nimbus II, III	5-15-66 (II)	1110 km (600 nmi) polar sun-synchronous (noon)	Albedo, water vapor and CO ₂ distribution, day and night cloud maps, and cloud temperature.	Cross-course rotating mirror (8 r/min), 2.059 rad (118 deg)	43(A-E) ^a
Spin-Scan Cloud Camera (SSCC)	ATS I	12-6-66	Geostationary	Daytime cloud cover maps	Spacecraft spin (100 r/min) and latitude step of telescope, 0.314 rad (18 deg)	0.1
Multi-color Spin-Scan Cloud Camera (MSSCC)	ATS III	11-5-67	Geostationary	Three-color daytime cloud cover maps	Spacecraft spin (100 r/min) and latitude step of mirror 0.314 rad (18 deg)	0.1(A-C)
High-Resolution Scanning Radiometer (HRSR)	ITOS	1-23-70	1460 km (790 nmi) polar sun-synchronous (3 p.m.)	Day-night cloud maps and cloud temperature	Cross-course rotating mirror (48 r/min), 1.920 rad (110 deg)	2.7(A) 5.6(B)
Temperature-Humidity IR Radiometer (THIR)	Nimbus IV	4-8-70	1110 km (600 nmi) polar sun-synchronous (noon)	Water vapor distribution, day-night cloud maps, and cloud temperature	Cross-course rotating mirror (48 r/min), 2.059 rad (118 deg)	7(B) 21(A)
Multi-spectral Scanner (MSS)	ERTS A, B	First quarter, 1972	930 km (500 nmi) polar sun-synchronous (10 a.m.)	Earth resources survey	Cross-course oscillating mirror (15 Hz), 0.202 rad (11.6 deg)	0.077(A-D) 0.2(E)
Imaging Photopolarimeter	Pioneer F, G	First quarter, 1972	Jupiter flyby at 203,830.00 km (110,000 nmi)	Photometry and polarization of zodiacal light, asteroids, and Jupiter. Two-color mapping of Jupiter	Spacecraft spin (4.8 r/min) and cone angle step of telescope, 0.506 rad (29 deg)	0.5(A, B)
Visible-IR Spin-Scan Radiometer (VISSR)	Synchronous Meteorological Satellite (SMS)	Second quarter, 1972	Geostationary	High-resolution day and night cloud maps and temperature	Spacecraft spin (100 r/min) and latitude step of mirror, 0.314 rad (18 deg)	0.025(A) 0.2(B)
Very-High-Resolution Radiometer (VHRR)	ITOS-D	First quarter, 1972	1,460 km (790 nmi) polar sun-synchronous (3 p.m.)	High-resolution day and night cloud maps, and temperature	Cross-scan rotating mirror (400 r/min), 2.007 rad (115 deg)	0.6
Facsimile	Ranger			Lunar surface photo	Nodding mirror, camera rotation	1.5
	Explorer			Antenna position monitor	Rotating, nodding mirror 6.282 rad (360 deg)	2.3
Surface Composition Mapping Radiometer (SCMR)	NIMBUS E	First quarter, 1972	1,110 km (600 nmi) polar sun-synchronous (noon)	High-resolution maps of terrestrial mineral characteristics	Cross-scan rotating mirror (600 r/min), 1.571 rad (90 deg)	0.6
Very-High Resolution Radiometer (VHRR)	ATS-F	Fourth quarter, 1972	Geostationary	High-resolution day and night cloud maps	Raster scan of servo-stepped, gimbaled flat mirror	0.3(A) 0.15(B)

^a Letters in parentheses refer to respective channels of multispectral scanners.

Table 26.7 - (continued)

Type	Spacecraft	Optics aperture and focal ratio (including relay and immersion optics)	Detector type (spectral band)	Information bandwidth	System dynamic range blackbody temperature K; Earth albedo, %; Effective bright scene radiance range
High-Resolution Infrared Radiometer (HRIR)	Nimbus I, II, III	10.16 cm (4 in.): f/0.95 Cassegrain	PbSe photoconductor	300 Hz	200-340 K (night) $35 \times 10^{-3} \text{ W cm}^{-2} \text{ sr}^{-1}$ (day)
Medium-Resolution IR Radiometer (MRIR)	Nimbus II, III	4.37 cm (1.72 in.): f/0.27 (B, D, E) 4.37 cm (1.72 in.): f/0.9 (A, C) Cassegrain	Thermistor bolometer	4 Hz (A-E)	0-80% (A) 0-270 K (B), 0-330 K (C) 0-270 K (D), 0-290 K (E)
Spin-Scan Cloud Camera (SSCC)	ATS I	12.70 cm (5 in.): f/2 folded paraboloid	S-11 PMT	160 kHz	$\geq 1000:1$
Multi-color Spin-Scan Cloud Camera (MSSCC)	ATS III	12.70 cm (5 in.): f/3 Wynne-Rosin	S-11 PMT (A, B) S-20 PMT (C)	160 kHz (A, B, C)	$\geq 1000:1$ (A, B, C)
High-Resolution Scanning Radiometer (HRSR)	ITOS	12.70 cm (5 in.): f/3.4 (A)	Silicon photodiode (A)	910 Hz (A)	0-80% (A)
		12.70 cm (5 in.): f/0.25 (B) Cassegrain	Thermistor bolometer (B)	455 Hz (B)	0-330 K (B)
Temperature-Humidity IR Radiometer (THIR)	Nimbus IV	12.70 cm (5 in.): f/0.24 (A), 12.70 cm (5 in.): f/0.26 (B) Cassegrain	Thermistor bolometer (A, B)	115 Hz (A) 345 Hz (B)	0-270 K (A) 0-330 K (B)
Multi-spectral Scanner (MSS)	ERTS A, B	22.86 cm (9 in.): f/3.6 (A-D)	S-20 PMT (A, B) S-25 PMT (C)	35.5 kHz (A-D)	27×10^{-4} (A), 22×10^{-4} (B) 17×10^{-4} (C), 28×10^{-4} (D)
		22.86 cm (9 in.): f/2 (E) Ritchey-Chretien	Si PD (D) HgCdTe (E)	13.7 kHz (E)	$W \text{ cm}^{-2} \text{ sr}^{-1}$ 0-310 K (E)
Imaging Photopolarimeter	Pioneer F, G	2.54 cm (1 in.): f/3 Maksutov	S-20 Channel multiplier (A, B)	500 Hz (A, B)	8×10^{-5} (A), 4×10^{-5} (B) $W \text{ cm}^{-2} \text{ sr}^{-1}$
Visible-IR Spin-Scan Radiometer (VISSR)	Synchronous Meteorological Satellite (SMS)	40.64 cm (16 in.): f/6.3 (A) 40.64 cm (16 in.): f/1.3 (B) Ritchey-Chretien	S-20 PMT (A) HgCdTe photoconductor (B)	210 kHz (A) 26 kHz (B)	0-80% (A) 0-320 K (B)
Very-High-Resolution Radiometer (VHRR)	ITOS-D	12.70 cm (5 in.): f/0.89 Dall-Kirkham	Silicon photodiode (A) HgCdTe photoconductor (B)	35 kHz	0.5-80% (A) 180-315 K (B)
Facsimile	Rangor	0.36 cm (0.14 in.): f/2.4	Silicon	100 Hz	$8.0-2500 \mu\text{W cm}^{-2} \text{ sr}^{-1}$
	Explorer	0.86 cm (0.34 in.): f/	Photodiode	2.5 kHz	$8.5 \text{ to } 2800 \mu\text{W cm}^{-2} \text{ sr}^{-1}$
Surface Composition Mapping Radiometer (SCMR)	NIMBUS E	20.32 cm (8 in.): f/0.92	HgCdTe (A, B)	50 kHz	257-330 K (A, B)
	ATS-F	20.32 cm (8 in.): f/1.7	HgCdTe (A) Silicon photodiode (B)	1200 Hz	185-335 K (A) 1-100% (B)

Table 26.7 — (continued)

Type	Spacecraft	NE Δ T at scene temperature K or SNR at $\left\{ \begin{array}{l} \text{scene} \\ \text{albedo, \%} \\ \text{irradiance,} \\ \text{W cm}^{-2} \end{array} \right.$	Scanner size		Total weight		Maximum power, W	Contractor	Reference
			cm	in.	kg	lb			
High-Resolution Infrared Radiometer (HRIR)	Nimbus I, II, III	0.22 K at 270 K (B) 0.23 K at 330 K (C) 0.20 K at 270 K (D) 0.26 K at 290 K (E)	25.40 \times 40.64 \times 22.86	10 \times 16 \times 9	8.62	19.0	4	ITT	21
Medium-Resolution IR Radiometer (MRIR)	Nimbus II, III		16.51 \times 16.51 \times 33.02	6.5 \times 6.5 \times 13	6.58	14.5	7.5	SBRC	22
Spin-Scan Cloud Camera (SSCC)	ATS I	>30 at 10 ⁻¹¹ W cm ⁻²	25.40 \times 27.94 \times 43.18	10 \times 11 \times 17	9.07	20	21	SBRC	23
Multi-color Spin-Scan Cloud Camera (MSSCC)	ATS III	>30 at 10 ⁻¹¹ W cm ⁻²	30.48 \times 27.94 \times 43.18	12 \times 11 \times 17	10.66	23.5	23	SBRC	24
High-Resolution Scanning Radiometer (HRSR)	ITOS	19 at 0.5% } (A) 3000 at 80% } 1.4 K at 185 K } (B) 0.3 K at 300 K }	16.26 \times 40.39 \times 21.34	6.4 \times 15.9 \times 8.4	8.30	18.3	6.5	SBRC	25
Temperature-Humidity IR Radiometer (THIR)	Nimbus IV	4 K at 185 K } (A) 0.2 K at 300 K } 1.5 K at 185 K } (B) 0.3 K at 300 K }	17.78 \times 19.05 \times 39.62	7 \times 7.5 \times 15.6	9.03	19.9	7.5	SBRC	26
Multi-spectral Scanner (MSS)	ERTS A, B	100 at 1.6 \times 10 ⁻¹¹ W cm ⁻² (A) 73 at 1.3 \times 10 ⁻¹¹ W cm ⁻² (B) 44 at 1.0 \times 10 ⁻¹¹ W cm ⁻² (C) 73 at 1.7 \times 10 ⁻¹¹ W cm ⁻² (D) 1.2 K at 310 K (E)	38.10 \times 38.10 \times 91.44	15 \times 15 \times 36	52.16	115	25	SBRC	27
Imaging Photopolarimeter	Pioneer F, G	25 at 2.2 \times 10 ⁻¹¹ W cm ⁻² (A) 25 at 1.1 \times 10 ⁻¹¹ W cm ⁻² (B)	17.78 \times 38.10 \times 15.24	7 \times 15 \times 6	4.08	9	4	SBRC	28
Visible-IR Spin-Scan Radiometer (VISSR)	Synchronous Meteorological Satellite (SMS)	3 at 0.5% (A) 1.7 K at 200 K } (B) 0.4 K at 300 K }	50.80 \times 50.80 \times 147.32	20 \times 20 \times 58	60.10	132.5	23	SBRC	29
Very-High-Resolution Radiometer (VHRR)	ITOS-D	20 at 0.5% (A) 1.5 K at 185 K } (B) 0.5 K at 300 K }	20.32 \times 20.32 \times 48.26	8 \times 8 \times 19	9.07	20	5.0	RCA	30
Facsimile	Ranger	3 at 5.6 \times 10 ⁻¹¹ W cm ⁻² (at limiting bandwidth)	3.12 dia \times 24.99	1.23 dia \times 9.84	1.18	2.6	16	Philco-Ford	31
	Explorer	3 at 14.1 \times 10 ⁻¹¹ W cm ⁻² (at limiting bandwidth)	3.81 dia \times 16.51	1.5 dia \times 6.5	0.50	1.1	3		32
Surface Composition Mapping Radiometer (SCMR)	NIMBUS E	1.0 K at 280 K	22.86 \times 42.16 \times 54.86	9 \times 16.6 \times 21.6	19.05	42	15	ITT	
Very-High Resolution Radiometer (VHRR)	ATS-F	1.0 K at 200 K (A) 30 at 1% (B)	67.31 \times 53.34 \times 38.10	26.5 \times 21 \times 15	27.22	60	35	ITT	

26.5 References

- Barnett, J. J., Cross, M. J., Harwood, R. S., Houghton, J. T., et al. (1971), "The First Year of the Selective Chopper Radiometer on NIMBUS IV," *Quart. J. Roy. Met. Soc.*, 98, 1972, p. 17.
- Barringer, A. R. and Shock, J. P. (1966), "Progress in the Remote Sensing of Vapours for Air Pollution, Geologic and Oceanographic Application," *Proceedings of the 4th Symposium on Remote Sensing of the Environment*, University of Michigan.
- Hanel, R. A. and Conrath, B. J. (1970), "Thermal Emission Spectra of the Earth and Atmosphere from the NIMBUS 4 Michelson Interferometer Experiment." *Nature*, 328, (5267), 1970, p. 143.
- Kaplan, L. D. (1959), "On the Inference of Atmospheric Structure from Remote Radiation Measurements," *J.O.S.A.* 49; p. 1004.
- King, J. I. F., 1958: "The Radiative Heat Transfer of Planet Earth," *Scientific Uses of Earth Satellites*, Second Revised Edition, edited by J. A. Allen, University of Michigan Press, Ann Arbor, 316 pp.
- Saiedy, F., Hilleary, D. T., and Morgan, W. A. (1965), "Cloud Top Altitude Measurements from Satellites," *Applied Optics* 4, p. 495.
- Smith, W. L. (1968), "An Improved Method for Deducing Tropospheric Temperature and Moisture Profiles from Satellite Radiation Measurements," *Monthly Weather Review*, 96 (6) pp. 387-396.
- Wark, D. Q. (1970), "SIRS: An Experiment to Measure the Free Air Temperature from a Satellite," *Applied Optics* 9; p. 1861, Dec.

List of Symbols

S/N	Signal-to-noise ratio
NE Δ T	Noise equivalent temperature difference
Δ T	Least discernible temperature contrast
τ	The transmittance of the atmosphere (dimensionless)
d τ /dx	Derivation of transmittance

Chapter 27 REACTION RATES OF ATMOSPHERIC CONSTITUENTS

Eldon E. Ferguson
Aeronomy Laboratory
Environmental Research Laboratory
National Oceanic and Atmospheric Administration

A concise introduction to chemical reaction rate theory is presented. Several categories of reactions are described and illustrated by example. Specific tropospheric reaction schemes relevant to the problem of the photochemical oxidation of nitric oxide and the oxidation of hydrocarbons to produce aldehydes and peroxyacylnitrates (eye irritants) are described.

27.0 Introduction

This chapter is an introduction to atmospheric chemistry and is directed primarily toward persons concerned with atmospheric measurements and properties whose specialization is not chemical kinetics or atmospheric photochemistry. The chapter attempts to accomplish two tasks, in a necessarily brief or even superficial manner. The first of these is to give an introduction to reaction rate theory (kinetic theory). Clearly anyone desiring very detailed knowledge in this area will go to one of the numerous standard reference works in the field. The second goal is to give a cursory view of some of the neutral reactions and reaction schemes that are of concern in the troposphere. In this case, one desiring more detailed information will wish to consult current reviews and research papers in this rapidly progressing field.

Chemical kinetics, like much of modern science, is a field in which the general principles are believed to be well known but only for the very simplest examples is a rigorous theoretical analysis tractable. The sequence of events following a collision between two molecules, which would include reaction as well as elastic scattering, rotational and vibrational excitation, and other phenomena, is in principle contained in the solution of the time dependent Schrödinger equation describing the collision. For any reaction of interest in the atmosphere, this constitutes a many body problem (all the electrons) on a formidable scale. Rigorous quantum mechanical kinetic theory is currently very much at the $H + H_2 \rightarrow H_2 + H$ level and indeed comparison of theory and experiment (facilitated by the use of isotopes) on this reaction is very much in the forefront of present advanced chemical kinetics research.

It is clear that most of us, with somewhat pressing practical chemical problems to contend with, cannot wait for theoretical solutions to these problems. On the other hand the very mass of data necessarily involved in reaction kinetics almost demands some theoretical systematization. If, for each reaction of interest to us, we had to have a separate table (or graph) of measured rate constants as a function of temperature, the matter would be unwieldy, to say the least. We adopt then, in common with most other disciplines in science, simplified theoretical models which allow us to condense and simplify experimental data, to extrapolate data beyond the range of measurements, and to make some general predictions about unmeasured reactions. Fortunately, a great deal of kinetic data is described reasonably well by extremely simple models. We shall first describe these models and then discuss several specific reaction systems of tropospheric interest.

27.1 Reaction Rate Theory

27.1.1 The Arrhenius Equation

We consider first exothermic binary reactions between two neutral molecules to produce chemically distinct products. This is the class of reactions of most interest in the troposphere and indeed the most studied class of reactions in kinetics. It has been known for a long time that the dependence of reaction rate constant on temperature for most reactions is given rather well by a simple, two-parameter equation, the so-called Arrhenius equation

$$k = A \exp(-E/RT) \quad (27:1)$$

in which A is called the pre-exponential factor and E is called the activation energy; A is usually assumed to be temperature independent. Although it can be shown theoretically that A must depend on temperature, the exponential term so dominates the temperature dependence that A can satisfactorily be considered constant. A more general equation sometimes used is

$$k = A'T^n \exp(-E'/RT) \quad (27:2)$$

however almost no kinetic data of sufficient precision exists to support a three-parameter equation and unless there is a theoretical basis for independently picking one of the parameters, it is most satisfactory to stay with the Arrhenius expression. The drastic temperature dependence of (27:1) follows not from any subtlety of the chemical reaction mechanism but from a very simple model in which an energy barrier to reaction exists (such as the breaking of a chemical bond while new bonds are being formed) so that the exponential term is simply the Boltzmann factor describing the distribution of reactant energies. Reaction is normally limited to the high energy reactants in the tail of the Boltzmann distribution.

The definition of the rate constant k follows from the basic kinetic equation, the law of mass action. If we assume a *simple* reaction



then

$$-\frac{d[A]}{dt} = -\frac{d[B]}{dt} = \frac{d[C]}{dt} = \frac{d[D]}{dt} = k[A][B]. \quad (27:4)$$

Brackets denote concentrations, i.e. [A] may be in molecules cm^{-3} , moles liter $^{-1}$, parts per million, or some other convenient (or conventional) concentration unit. *Simple*, as applied to reactions, has the technical connotation that the mechanism is suggested by the reaction equation, in this case one molecule of A strikes one molecule of B, then one molecule of C and one molecule of D emerge from the collision. Reactions whose overall effect is as simple as (27:3) may involve more complicated kinetics and not obey the law of mass action (27:4) which applies to simple *steps* of reactions. Such reactions are called *Complex*. The complexity may arise because A first dissociates into free radicals before reacting with B, from catalysis, as a result of reactions on surfaces, etc. A well known example is the combustion of hydrogen and oxygen mixtures, in which H, O, and OH radicals all play a role.

Some experiments in kinetics measure cross sections rather than rate constants (e.g., beam experiments) and the relation between the two is

$$k(T) = \left[\frac{8}{\pi\mu(kT)^3} \right]^{1/2} \int_0^{\infty} \exp(-E/RT) Q(E, T) E dE \quad (27:5)$$

for Boltzmann distributions of states. The symbol μ denotes reduced mass of the reactant pair,

$$\mu = \frac{M_1 M_2}{M_1 + M_2}$$

and R is the gas constant. This is the fundamental link between the macroscopic and microscopic kinetics of binary reactions. The rate constant k is a quantity which measures macroscopic conversion rates, Q describes the interaction at the molecular level. It is clear that rate constants can be readily deduced from cross sections but the converse is not ordinarily so.

With regard to units, aeronomers invariably express concentrations in particles cm^{-3} and hence k for binary reactions in $\text{cm}^{-3} \text{sec}^{-1} \text{particle}^{-1}$, except that they almost invariably do not bother to write particle

(or molecule) as part of the units. Traditional kineticists often measure concentrations in moles liter⁻¹ and hence k in liter mole⁻¹ sec⁻¹. Pollution chemists on the other hand often measure concentration in parts (by number) per million or billion (ppm or ppb) and hence k might be in units such as pphm⁻¹ hr⁻¹.

27.1.2 Activation Energy

The quantity E in (27:1) is called the activation energy. The general definition of activation energy is

$$-R \frac{d \ln k}{d(1/T)} \equiv E \quad (27:6)$$

which is valid whatever the functional form of $k(T)$. It can be shown that E is the difference between the average energy of those molecules which do react and the average energy of *all* the reactant molecules. It is simplest to imagine E as an energy barrier (F27.1). The difference between the reverse and forward activation energies is the heat of reaction $E_r - E_f = \Delta E_T$. The activation energies (and heat of reaction) depend on the temperature strictly speaking, although this refinement is not encompassed in the Arrhenius Equation (27:1).

27.1.3 Collision Theory of Bimolecular Reactions

If one assumes a functional form for $Q(E,T)$ in (27:5), the integral can be evaluated to yield a *collision theory rate* constant. One conventional assumption is that reaction will occur with a hard sphere cross section σ_R at all relative energies $E \geq E_0$ along the line of centers of the reactants. This leads to the *line-of-centers rate* constant

$$k(T) = \left[\frac{8\pi kT}{\mu} \right]^{1/2} \sigma_R \exp(-E_0/RT) \quad (27:7)$$

The model behind (27:7) is that reaction is controlled only by a translational threshold. An extended collision model allows for the possibility that other degrees of freedom, particularly vibrational, contribute to reaction. Assuming s classical oscillators to be involved leads to the *activation-in-many-degrees-of-freedom* rate constant,

$$k(T) = \left[\frac{8\pi kT}{\mu} \right]^{1/2} \sigma \Gamma(s)^{-1} \left[\frac{E_0}{RT} \right]^{s-1} \exp(-E_0/RT) \quad (27:8)$$

in which Γ is the gamma function. This can lead to much larger rate constants than (27:7). This simple statistical model in which all internal degrees of freedom are assumed to be equivalent is clearly naive. It has been found experimentally and theoretically, that in some cases, vibrational energy is very much more effective than translation, in promoting reaction. Again, determination of reaction rate data depends on laboratory experimentation, simple models such as (27:7) and (27:8) are not adequate bases to make rate constant estimates.

27.1.4 Activated Complex Theory

A very powerful theory for certain kinds of reactions, particularly for those involving more complex molecules, was developed by Eyring and others in the thirties and is variously known as *activated complex theory*, *transition state theory*, or the *absolute reaction rate theory*. The assumption is basically that the reactants are in a pseudo-equilibrium with activated complexes, represented by the transition state of (F27:1). This effectively reduces the kinetic problem to a thermo-dynamic problem. This model obviously has limitations but has proven extremely valuable and, in the hands of experienced kineticists, it provides a

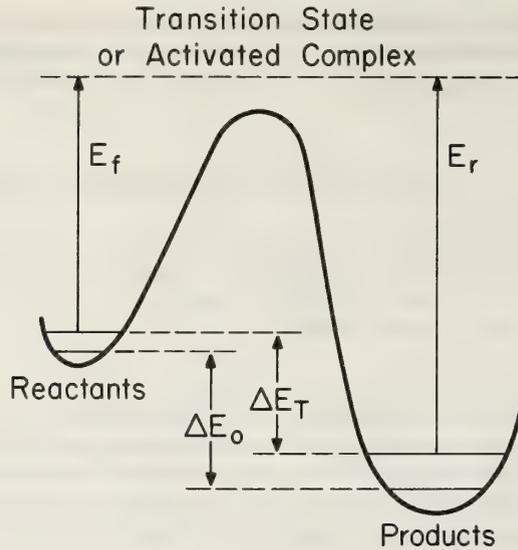


Figure 27.1 *Transition state or activated complex.*

good deal of predictability. The *activated complex* is not susceptible to measurement so that its thermodynamic properties need to be judiciously estimated. The Activated Complex Theory rate constant becomes

$$k = K \nu K \quad (27:9)$$

where K is a transmission coefficient, often put equal to unity, ν is the frequency with which the activated complex crosses the transition state to products, and K is the equilibrium constant between reactants and the transition state. It is practical to factor the reaction coordinate out of the equilibrium constant. Doing this and setting $K = 1$ leads to

$$K = \frac{kT}{h} K^\ddagger \quad (27:10)$$

where K^\ddagger is the equilibrium constant without the reaction coordinate included. From thermodynamics, $K^\ddagger = e^{-\Delta G^\ddagger/RT}$ and $\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$ so that one can write

$$k = \frac{kT}{h} \exp(\Delta S^\ddagger/R) \exp(-\Delta H_T^\ddagger/RT) \quad (27:11)$$

Alternatively, one can consider the problem from the point of view of statistical mechanics and obtain

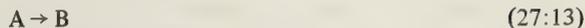
$$k(T) = K \frac{kT}{h} \frac{\Omega^\ddagger}{\Omega_A \Omega_B} \exp(-E/RT) \quad (27:12)$$

where the Ω 's are the partition functions for the activated complex (minus the reaction coordinate) and the reactants A and B. It appears from (27:11) that the pre-exponential factor is associated with the entropy change required to form the activated complex. This allows many generalizations and predictions about pre-exponential factors to be made.

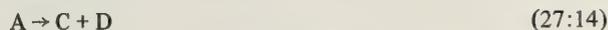
27.2 Categories of Reactions

27.2.1 Unimolecular Reactions

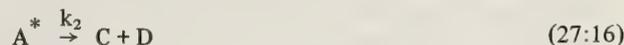
One is sometimes concerned about reactions of 1st order, e.g.,



or



where (27:13) may be an isomerization for example, the more common (27:14) is simply dissociation. The classification of (27:14) as 1st order, or unimolecular, follows from the kinetic expression $d[A]/dt = -k[A]$. At the molecular level, dissociations are at least binary processes:



in which collisions are necessary to energize A to a level at which it can dissociate. At high pressures, (i.e., where $k_{-1} [M] \gg k_2$) collisions maintain an equilibrium concentration of A^* , which is then independent of $[M]$, and the reaction appears to be 1st order. At low pressure ($k_{-1} [M] < k_2$) all unimolecular processes become bimolecular. The Arrhenius Equation (27:1) gives a useful description of unimolecular processes and Activated Complex Theory (ACT) is applicable. It can be predicted from ACT that unimolecular decomposition reactions will have large pre-exponential factors. Consider, for example, the decomposition of ethane, by way of an activated complex



It seems obvious that the activated complex $C_2H_6^\ddagger$, will have an increased C—C bond length. This will lower the vibrational and rotational energy levels of $C_2H_6^\ddagger$ as compared to C_2H_6 and thus there is an entropy increase in going from C_2H_6 to $C_2H_6^\ddagger$. From (27:11) this clearly leads to increase in the pre-exponential factor over the “normal” $\frac{kT}{h}$ value ($= 6.3 \times 10^{12} \text{ sec}^{-1}$ at 300°K).

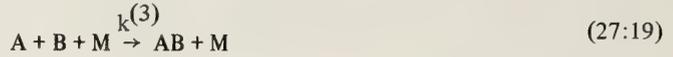
It is generally true that the activation energy for dissociation is less than the energy of the bond being broken! For example, the bond dissociation energy of O_3 is 24.2 kcal/mole and the observed activation energy is 22.7 kcal/mole.

27.2.2 Three-Body Association Reactions

The converse of a dissociation reaction is a recombination reaction



Recombination reactions play an important role in atmospheric chemistry at all altitudes up to the F-region of the ionosphere. Just as in the case of unimolecular dissociation, recombination is actually a process whose order varies with pressure. At high enough pressure, the process appears to be bimolecular (27:18), just the inverse of unimolecular decomposition (27:14) at high pressure. However, for the pressure regions of most common application, recombination is in its low pressure regime and follows termolecular kinetics,

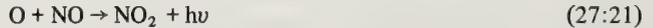


The role of M is obviously to carry away sufficient energy to leave A and B bound. In most cases of interest to us (27:19) is not very chemically specific for M, i.e., we may expect factors of 3 or 4 variation in rate constant as M varies from O₂ to N₂ but not order or magnitude variations. There are exceptions to this in cases where M is chemically reactive with A or B.

There are exceptions, principally radiative recombination,



in which the emitted photon plays the role of third body. The rate constant for (27:20) is never very large however. This can easily be seen to be the case as follows. When A and B collide to form a transitory AB*, then if a stabilizing event does not soon occur, A and B (which may of course be either atoms or molecules) will fly apart. The lifetime of AB* is roughly that of a vibrational period, $\tau \sim 10^{-13}$ sec. The lifetime of an allowed electronic transition, on the other hand is $\sim 10^{-8}$ sec, so we may roughly estimate the probability of radiative stabilization in a favorable case, to be $P \sim 10^{-13}/10^{-8} \sim 10^{-5}$, i.e. improbable. One prominent atmospheric example of (27:20) is



which gives rise to the well known atmospheric airglow. The rate constant for (27:21) is low as expected. When the recombination does not lead to an electronically excited state which has an allowed transition to a lower state, then radiative recombination becomes even less likely. Allowed radiative vibrational lifetimes are very long, the order of milliseconds or greater and the radiative lifetimes of vibrationally excited homonuclear diatomics are essentially infinite. One can estimate an order of magnitude rate constant for $k^{(3)}$ (27:19) by assuming a very simple energy transfer model



followed by



from which in the low pressure limit,

$$k^{(3)} = k_1 k_2 / k_{-1} \quad \text{or} \quad k_1 k_2 \tau \quad (27:24)$$

since the unimolecular rate constant k_{-1} is simply the reciprocal lifetime of AB^* . Taking for k_1 and k_2 simple gas kinetic collision rate constants and for τ a typical vibrational lifetime leads to

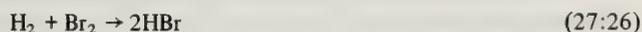
$$k^{(3)} \approx 10^{-10} \text{ cm}^3/\text{sec} \times 10^{-10} \text{ cm}^3/\text{sec} \times 10^{-13} \text{ sec} \approx 10^{-33} \text{ cm}^6/\text{sec}. \quad (27:25)$$

Kinetics and thermodynamics can be related by equilibrium considerations. Any reaction must, obviously, proceed at the same forward and reverse rates at equilibrium. The ratio of forward to reverse rate constants $k_f/k_r = K(T)$ is the equilibrium constant for the reaction and is completely determined by thermodynamic (or statistical mechanical) considerations, essentially the energy levels of the molecules involved and the temperature. The equilibrium constant can be calculated from such data (or may have been measured) and consequently knowledge of either k_f or k_r may allow the calculation of the other one. Of course, in the case of the inverse processes of dissociation and recombination, it is usually the case that the processes are of concern in different temperature regimes, for example, collisional dissociation is measured in shock tubes or re-entry wakes, etc., and recombination is measured in afterglows or in cool atmospheres. Nevertheless, utilization of (27:25) has been most valuable in kinetics.

Since, as discussed above (F27:1), the activation energy for recombination is related to that for dissociation by $E_r = E_d - \Delta E_T$, where ΔE_T is the endothermicity of the dissociation, *and* since it is usually found experimentally that $|E_d| < |\Delta E_T|$ it follows that the activation energy for recombination is usually negative. This means that the recombination rate constant increases with decreasing temperature. This is readily comprehensible on the basis of a simple picture. The recombination process is an attempt to join two separate species, and it follows that they will more readily adhere the lower their relative kinetic energy. From a slightly more sophisticated point of view, the lifetime of the transient complex $\tau(AB^*)$ increases with decreasing energy content, a problem which is solved mathematically for grossly oversimplified models. Typically, three-body association reaction rate constants behave something like $k^{(3)} \sim k_0(T/T_0)^{-1}$.

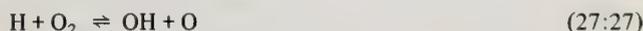
27.2.3 Chain Reactions and Free Radical Mechanisms

It was mentioned above that reactions may be complex, i.e., the reaction process may be much more involved than is suggested by the stoichiometric equation. This is often due to free radical, chain propagating steps. An often cited example is the reaction



The rate of disappearance of H_2 (or Br_2) is not given by $d[H_2]/dt = k[H_2][Br_2]$, but rather by $k[H_2][Br_2]^{1/2} [1 + k'[HBr]/Br_2]$. The similar reaction of H_2 with I_2 on the other hand, does follow the simple rate law, at least at low temperature, so that the matter is not an obvious one as to whether a reaction will be simple or complex.

The combustion of H_2 is similarly complex and involves free radical steps such as



and



The overall oxidation of H_2 is controlled by the reactions of the very reactive free radicals (species with unsatisfied valences) O, H, and OH. A reaction such as (27:28) is called a chain reaction, one free radical (OH) reacting with a reactant molecule (H_2) to produce a product molecule (H_2O) and another free radical (H) which further reacts (27:27) to propagate the chain. Reaction (27:27) produces *two* free radicals and is called a *branched chain* reaction. It is easy to see that branched chain reactions can lead to explosions in a manner similar to nuclear fission explosions. Unbranched chain reactions can also lead to explosion as well, if the system is unable to dissipate the heat released by reaction so that the temperature rises. This leads, in general, to more rapid reaction according to (27:1) and this can be a run-away situation, giving rise to thermal explosions.

Free radical mechanisms are of importance in hydrocarbon oxidation in atmospheric pollution situations. Usually free radical reactions do not have activation energies and are often very fast, i.e. $k \sim 10^{10} \text{ cm}^3/\text{sec}$. This is essentially a consequence of their having unsatisfied valences which can form new bonds without rupturing old bonds.

27.2.4 Ion-Molecule Reactions

Binary exothermic ion-neutral reactions have a somewhat simpler behavior than neutral reactions. Usually they do not have activation energies and have very little energy dependence. This is understandable on the basis that an ion and neutral are strongly attracted together by long range charge-induced dipole attractive forces (the potential is $U = e^2\alpha/2r^4$, where α is the neutral polarizability) so that at the time of impact their relative velocity is sufficient to overcome any activation barrier. Since ions are normally free radicals the barriers would usually be small anyway. If the neutral has a substantial dipole moment, such as H_2O does, then the long range forces are even stronger. Ion-neutral reactions often have large rate constants equal to the collision rate constant. For non-polar neutrals, this leads to

$$k_L = 2\pi e \sqrt{\alpha/\mu} \quad (27:29)$$

where μ is the reduced mass of the reactants. This is often called the Langevin rate constant, sometimes the Gioumousis and Stevenson rate constant and is typically a number $\sim 10^9 \text{ cm}^3/\text{sec}$ for small atmospheric species. This corresponds to a collision cross section for typical thermal velocities $v \sim 10^5 \text{ cm/sec}$ of $\sigma = k/v \approx 10^{-14} \text{ cm}^2 = 100 \text{ \AA}^2$. Note that k_L does not depend on the reactant energy.

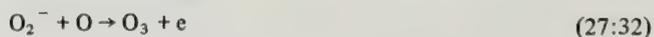
Ion-neutral reactions may be analogous to neutral reactions in that old bonds are broken and new bonds are formed, e.g.



or they may simply involve charge-transfer (an electron jump), e.g.



There is no essential difference between positive and negative ion reactions in these two categories. There is a kind of binary negative ion reaction which has no counterpart in positive ion reactions, associative-detachment; e.g.



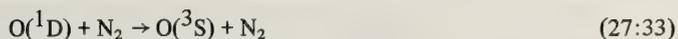
When exothermic, such reactions are almost invariably fast, $k \sim 10^{10} \text{ cm}^3/\text{sec}$. Three-body ion association

reactions, like their neutral counterparts, have negative temperature dependences, i.e., increasing rate constants with decreasing temperature. Because of the strong ion-neutral forces, association is much more general for ions than for neutrals. Any ion in the troposphere, for example, can be expected to be hydrated to a considerable degree, i.e., in the form $A^{\pm}(H_2O)_n$. The role of such hydration in atmospheric nucleation e.g. noctilucent cloud formation and aerosol formation is at present somewhat open.

27.2.5 Excited State Reactions, Quenching, Selection Rules

The reaction of an electronically excited species should be considered as a separate, independent, and unrelated process rather than as a simple modification of the reaction of the ground state species. In other words there is no necessary relationship between either the pre-exponential factors or activation energies in the two cases, indeed they will almost certainly differ. Vibrational excitation on the other hand may or may not lead to drastic changes in rate constant, little such data is available. Reactions are known in which the rate constant can change orders of magnitude with a small change of vibrational quantum number. Rotational relaxation is so rapid that it is almost never out of equilibrium with translation and the question of selective excitation does not arise.

A new phenomenon does arise in the case of excited states however, namely quenching. Quenching reactions may involve the conversion of electronic energy into translational or vibrational, e.g.



in which case it is not known where the energy does go although it is a matter of greatest concern in the E-region of the ionosphere. Alternatively, they may involve chemical reaction, e.g.,



a reaction which is apparently the long missing NO formation process in the earth's ionosphere. Electronic excitation transfer may also occur, e.g.

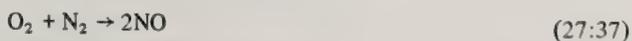


may be of great importance in urban atmospheres, the NO_2^* being produced by visible light absorption and the O_2^* (singlet oxygen) being important in the oxidation of hydrocarbons.

All of the theoretical discussion outlined above has implicitly assumed adiabatic reaction conditions, i.e., that the reactants and products were connected by a continuous potential surface. The curve of (F27.1) may be taken as a path over the lowest barrier height on such a multidimensional surface. In some cases, the ground states of products and reactants are not so connected and this can present a substantial hindrance to reaction. For example the dissociation



involves a change in electronic configuration, specifically the singlet reactant state goes to a triplet product state. This reaction therefore violates what is known as the spin conservation rule and is anomalously slow. It also follows, of course, that the reverse reaction, the association of ground state O atoms and CO molecules to produce CO_2 is anomalously slow, a matter of some significance in the atmospheres of Mars and Venus. Recently, very useful orbital symmetry selection rules have been developed, mostly for organic reactions. For example, the endothermic reaction



violates orbital symmetry conservation and is hence likely to be slow even when sufficient energy is supplied.

27.3 Specific Tropospheric Reactions

27.3.1 An Overview

It is not entirely appropriate to classify atmospheric reactions according to altitude regimes, i.e., tropospheric, stratospheric, or mesospheric, because vertical transport of critical species between these layers cannot be ignored. For example, the transport of ozone from the stratosphere to the troposphere is important. Nevertheless, we discuss some of the reactions primarily of tropospheric concern, keeping this limitation in mind.

The photochemistry of the natural atmosphere is exceedingly complex, indeed the photochemistry of pure oxygen systems in the laboratory is already of such complexity that widespread research continues in attempts to resolve unanswered questions. Johnston (1968) has given a very detailed analysis of the kinetics of an oxygen system with a comprehensive and critical evaluation of available data up to the time of his review. When to the "natural" atmosphere (which includes oxygen, nitrogen, oxides of nitrogen, water, hydrocarbons and sulfur compounds) one adds the effluents of civilization, the photochemistry becomes extremely complex indeed. Since certain aspects of this photochemistry manifest themselves in increasingly unpleasant ways, the scientific effort engaged in understanding (and solving?) problems of atmospheric photochemistry has become substantial. In the broad perspective it must be said that our present level of understanding of atmospheric photochemistry is very limited.

One of the complicating features of atmospheric chemistry is the very small concentrations of critical reactants that are involved. For example, the most powerful oxidizer, atomic oxygen, occurs only on the order of a few parts in 10^{14} in a typical polluted atmosphere. The problems associated with measuring such concentrations are formidable. Additionally, many reactions involve unstable species (free radicals) which are both difficult to detect in the atmosphere and difficult to study in the laboratory. Thus many critical reaction rate constants remain unmeasured. Reactions are not easy to carry out in the laboratory at the minuscule concentrations appropriate to the atmosphere (because of wall reactions, etc.) and this presents a serious problem since the reaction kinetics can alter between very low and relatively high concentrations.

Table (T27:1) is a representative collection of typical tropospheric compounds. The photochemical problem might be posed: What is the relationship between these species in the presence of sunlight? A photochemical pollution problem might be posed: If certain mixtures of hydrocarbons and nitrogen oxides are added to this photochemical scheme, what will then happen?

In addition to gas phase reactions, tropospheric chemistry is influenced greatly by reactions on particle surfaces, in both natural and polluted atmospheres. The eventual washing out of certain atmospheric species, e.g., sulphates, plays an important role in their chemical balance. The bible in this field remains the classic work of Leighton (1961) even though it must be substantially augmented as a consequence of a decade's progress. A comprehensive compilation of quantitative reaction rate data has been prepared by Schofield (1967).

27.3.2 Nitrogen Oxides

The most plentiful nitrogen compound in the atmosphere is the relatively inert nitrous oxide, N_2O which is produced largely by biological action in the soil. An excellent discussion of the sources and abundances of gaseous atmospheric pollutants has been prepared by Robinson and Robbins (1968). McElroy and McConnell (1971) have recently proposed that N_2O provides a natural source of NO, via the reaction

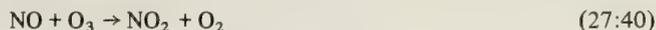


The source of $O(^1D)$ is the photolysis of ozone,



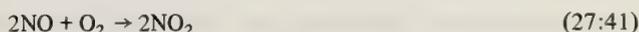
which produces $O(^1D)$ at the rate of $2.4 \times 10^{-5} [O_3] \text{ sec}^{-1}$ for an overhead sun (Levy, 1971a).

One of the more pressing problems in current pollution photochemistry is how the NO produced by internal combustion engines in urban atmospheres is oxidized to nitrogen dioxide, NO₂. NO₂ is a major visible manifestation of photochemical smog, giving rise to the brownish pall over cities like Los Angeles, Denver, and Boulder. In rural areas the rapid oxidation of NO by O₃



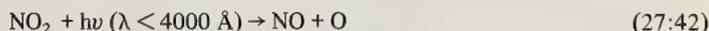
leads to worldwide ratio of NO₂/NO of about 2:1 (Berry and Lehman, 1971). The ozone arises from downward diffusion from the stratosphere. In urban areas where the morning NO concentration may approach 0.1 ppm or more (compared to ~ 0.2 – 4 ppb for rural areas) the ozone concentration is quite insufficient to account for the observed rapid, essentially complete conversion of NO to NO₂.

With sufficient NO concentrations, the three-body reaction



will occur, however at the low atmospheric NO concentrations (27:41) is negligible.

The essence of the NO → NO₂ conversion problem is that the major atmospheric oxidation source is

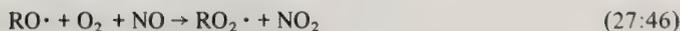


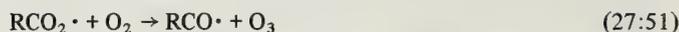
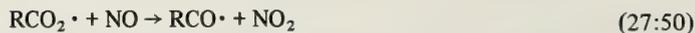
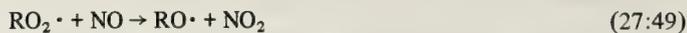
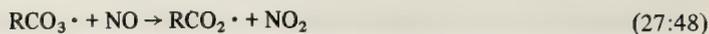
where the photolysis proceeds with unit efficiency for $\lambda < 3600 \text{ \AA}$ and to a lesser extent for $\lambda < 4000 \text{ \AA}$. It is obviously not quite cricket to ask the NO₂ to provide the source for its own production, i.e., the oxidant to convert NO to NO₂!

Following Berry and Lehman (1971) we shall give several NO oxidation schemes which have been proposed. None of these has yet been shown to be sufficient or to be the actual atmospheric process.

27.3.2.1 The Atomic Oxygen Mechanism

The mechanism of NO oxidation by atomic oxygen is indicated by the following series of reactions where R• indicates a hydrocarbon radical.





This radical reaction chain repeatedly converts NO to NO₂ while regenerating radicals which can continue the chain. Production of O₃ in (27:51) is equivalent to NO→NO₂ conversion because of (27:40). The chain terminates by radical loss due to radical-radical recombination, radical scavenging by NO₂ and less important processes. There does not seem to be enough atomic oxygen present to account for NO → NO₂ conversion by this scheme, however.

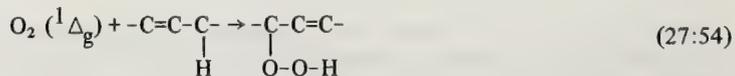
27.3.2.2 Singlet Oxygen Mechanisms

Singlet oxygen (i.e. metastable excited oxygen) can be produced by excitation transfer from NO₂,

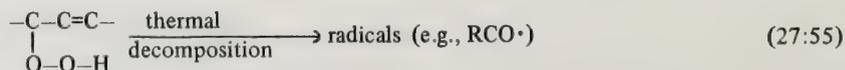


This process has recently been studied by Frankiewicz and Berry (1971) and Jones and Bayes (1971). Singlet oxygen may also be produced by excitation transfer from other species, e.g., SO₂* or excited states of hydrocarbons. Also the photolysis of ozone (27:39) will lead to some O₂(¹Δ_g).

Once it is produced, O₂(¹Δ_g) (which is only slowly quenched by N₂ and O₂, in contrast to the rapid quenching of O(¹D)) may then react with olefins to give hydroperoxides:



The hydroperoxides are thermally decomposed to free radicals

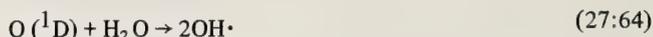


These radicals then undergo chain reactions, e.g., (27:47), etc. in the same way as the atomic oxygen generated radicals.

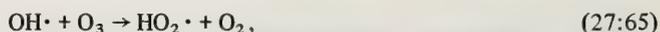
Reactions (27:62) and (27:63) are important scavenging reactions for NO_2 , which along with inorganic nitrate formation accounts for most of the NO_2 removal.

The lowest member of the peroxyacylnitrate family, peroxyacetylnitrate (PAN) is a strong eye irritant and is considered to be a major culprit in smog irritation. Another peroxybenzoylnitrate (PBN) has recently been found to be a much stronger eye irritant and may also be important even though present in less concentration. In PBN, R is the benzene radical.

Levy (1971a,b) has recently developed detailed reaction schemes for the normal (non-polluted) troposphere, commencing with $\text{O}(^1\text{D})$ production by ozone photolysis (24:39) and leading to a substantial production of formaldehyde by way of



for which the rate constant is very uncertain at present. This is critical since (27:64) must compete with very fast quenching of $\text{O}(^1\text{D})$ by both N_2 and O_2 . The OH's oxidize CO (27:58), they destroy ozone



and they destroy methane



The $\text{CH}_3\cdot$ formed in (27:66) rapidly attaches to the plentiful O_2 to form a methylperoxyl radical



This radical oxidizes NO to give a methoxy radical



which in turn reacts with O_2 to give formaldehyde and a hydroperoxyl radical



The $\text{HO}_2\cdot$ can react with NO to regenerate OH and continue the chain



Levy computes that the large formaldehyde concentration of (T27:1), 1.5×10^{10} molecules/cm³ at noon, will arise from this chain.

27.4 Conclusions

These very brief and necessarily superficial lectures should convey the message that atmospheric photochemistry is an extremely complex subject in which only the barest dent has been made. Certainly an essential key in further progress in this field will depend on the development of extremely sophisticated

atmospheric measuring techniques to monitor the trace constituents occurring in the various proposed reaction schemes. The vital role of atmospheric photochemistry in our increasingly urbanized, populated, and mechanized planet behooves us to make progress in this field.

Table 27.1
Some Representative Values of Atmospheric Constituent Concentrations

Species	Concentration (molecules/cm ³)	
Air = M	2.5 (19)	
O ₂	5 (18)	
H ₂ O	5 (17)	50% relative humidity
N ₂ O	6 (14)	at 30°C
CH ₄	3.8 (13)	
CO	3 (12)	
O ₃	1.2 (12)	
NH ₃	1.5 (11)	
NO + NO ₂	~3 (10)	Rural area
	~3 (12)	Morning air of large city
HNO ₃	1 (10)	
HNO ₂	2 (8)	
NO ₃	1 (7)	
N ₂ O ₅	2 (6)	
H ₂ C=O	1.5 (10)	
HO ₂	5 (8)	
OH	5 (6)	
CH ₃ O ₂	1 (4)	
O(³ P)	~3 (5)	Polluted air (Leighton 1961)
O(¹ D)	~2 (-2)	

These values are largely from Levy (1971), and are only for the purpose of orientation. They are in many cases highly variable, often quite uncertain, and the list is not a complete one.

27.5 References

- Berry, R. S. and P. A. Lehman, (1971): Aerochemistry of Air Pollution, *Ann. Reviews of Physical Chemistry* 22 47.
- Frankiewicz, T. and Berry, R. S., (1972): Singlet O₂ Production from Photoexcited NO₂, *Environ. Sci. Technol.* 6, 365.
- Johnston, H. S., (1968): "Gas Phase Reaction Kinetics of Neutral Oxygen Species" NSRDS-NBS 20.
- Jones, I. T. N. and K. D. Bayes, (1971): Energy Transfer from Electronically Excited NO₂, *Chem. Phys. Letters* 11, 63.
- Leighton, P. A., 1961: *Photochemistry of Air Pollution*. Academic Press, New York.
- Levy, H., (1971a): Normal Atmosphere: Large Radical and Formaldehyde Concentrations Predicted, *Science* 173, 141.
- Levy, H., (1971b): Photochemistry of the Lower Troposphere, *Planet Space Sci.*, in press.

McElroy, M. B. and J. C. McConnell, (1971), Nitrous Oxide: A Natural Source of Stratospheric NO, *J. Atm. Sci.*, in press.

Schofield, K., (1971): An Evaluation of Kinetic Rate Data for Reactions of Neutrals of Atmospheric Interest, *Planet Space Sci.* 15, 643.

List of Symbols

A	pre-exponential factor
E	activation energy
μ	reduced mass of reaction pair
R	gas constant
Q	describes the interaction at the molecular level
A,B,C,D	reactants
[A]	concentration, molecules cm^{-3} , moles liter $^{-1}$, or parts per million
T	temperature, Kelvin
E_r	reverse activation energies
E_f	forward activation energies
σ_R	hard sphere cross section
Γ	gamma function
k	rate constant
κ	transmission coefficient
ν	frequency of crossing transition state, electromagnetic frequency
K	equilibrium constant between reactants and the transition state
K^\ddagger	equilibrium constant without the reactant coordinate included
Ω	partition function for the activated complex
K	in the combination $K \frac{T}{h}$, is Boltmann's constant, ($K \frac{T}{h} = 6.3 \times 10^{12} \text{ sec}^{-1}$ at 300 °K)
τ	lifetime of a quantum state
k_{-1}	unimolecular rate constant
k_1, k_2	gas kinetic collision rate constants
ΔE_T	endothermicity of dissociation
α	neutral polarizability

Chapter 28 DATA GATHERING AND PROCESSING ASPECTS OF REMOTE TROPOSPHERIC SENSING SYSTEMS

S. W. Maley

University of Colorado

Remote tropospheric sensing systems are becoming of increasing importance because of the urgent need for environmental preservation. Effective tropospheric sensing requires elaborate data gathering and processing systems. Communication systems suitable for data gathering networks are discussed with particular attention given to the system complexity required for an adequate sensing system. Data storage and processing are also discussed in general terms; and recommendations are made concerning the form of information carrying signals that are the most desirable from the point of view of ease of transmission, storage, and processing of information. Finally, it is observed that remote tropospheric sensing will require radio spectrum allocations, probably eventually, in such quantity as to require discontinuance of some current spectrum uses.

28.0 Introduction

Remote sensing systems for monitoring the troposphere have become a topic of great interest recently because of increasing efforts being made to control atmospheric pollution. This trend is likely to increase in the foreseeable future.

The overall problem of remote sensing of the troposphere may be considered to consist of several parts, sensor design, storage and processing of sensor output signals at sensor locations, transmission of data from sensors to data centers, storage and processing of information at data centers, transmission of information between data centers and dissemination of data produced by the remote sensing system.

Sensor design is a topic broad in scope and under widespread investigation at the present time. There are various types of sensors for monitoring the various characteristics of the troposphere. Some of these are discussed in other lectures of this series. This discussion, however, will be limited to consideration of selected aspects of storage, processing, and transmission of information in remote sensing systems. The sensor will simply be regarded as an information source characterized by the rate at which it generates information.

28.1 Data gathering systems

Tropospheric remote sensing systems may be of various sizes depending upon their location and the specific objectives of the system. The smallest systems to be considered here consist of a set of sensors and a data gathering system to transmit information from the sensors to a central data storage and processing facility at which the data is analyzed and the results of the analysis are made available. Such a system may be designed for a city or part of a city or perhaps a single industrial complex. On the other hand it could be designed for an area larger than a city. Such a system will be called a local system.

A number of local systems can be combined into a regional system by providing data communication facilities among the data centers of the various local systems. One of these centers can also serve as a regional data center.

A regional tropospheric sensing system may include a group of cities, or perhaps a state, or a section of a country.

A group of regional tropospheric sensing systems can be connected together with data communication facilities into a larger tropospheric sensing system. This may be of such size as to include an entire country or even a continent or an ocean. Again, in this case, one of the regional data centers can become the data center for the entire national, continental, or oceanic tropospheric sensing system.

Finally these national, continental, or oceanic tropospheric sensing systems can be organized into a world wide system with facilities for transmission of information between data centers all over the world.

Nothing has been said yet about the positioning of the sensors. In the simplest system for monitoring the troposphere over land, near the surface of the earth, the sensors can be mounted on ground based structures and the communication links can be wires or radio links. If it is necessary to have sensors above the surface of the earth, towers could be used for small heights. For large heights balloons,

aircraft, rockets, or satellites could be used. In this case radio communication links must be used. The need for a communication link could be avoided in some cases by recording data at the sensor and transporting the recorded data to the data center. This has been done in the case of sensors mounted in aircraft or rockets.

Sensing systems over oceans generally must make greater use of radio communication links than those over land. Those close to land could transmit information directly from the sensors to a data center on land. Sensing systems remote from land, present a more difficult problem. One solution is the use of a ship to serve as a data gathering center. Information could then be processed on the ship or it could be transmitted using a radio link to a data center on land. This transmission could be by way of a communication satellite. Alternatively, a satellite could be used as a data gathering center. The information would then be transmitted from the satellite to a data center probably on land.

The spacing of the sensors in a sensing network is dependent upon the accuracy desired in the spacial distribution of the particular characteristic of the atmosphere that is being measured. Optimal spacing can be determined from the rates of spacial variation in the quantities being measured and from sampling theory (see App. A). The complexity of the data gathering network as a function of sensor spacing will be treated in the following paragraphs.

The basic building block of tropospheric remote sensing systems is the local system. It will be treated in greater detail here because larger systems simply consist of local systems connected together using conventional data communication systems and data processing systems.

Consider a local tropospheric remote sensing system designed to monitor the troposphere over an area perhaps the size of a city. The system will consist of a number say, N , of sensors at selected locations which will monitor one or more characteristics of the troposphere and will supply output signals which may be stored temporarily and may receive preliminary processing at the monitor site. However, these signals must then be transmitted to a data center.

The sensor may be of a type that generates a continuous signal characterizing a particular property of the troposphere (temperature for example). Furthermore this continuous signal may be continuously transmitted to the data center or, alternatively, it may simply be sampled periodically with sample values being sent to the data center (For example a continuous record of temperature is probably not necessary; a reading every several minutes is probably sufficient). On the other hand the sensor may be of a cyclic type that goes through a series of operations and supplies an output consisting of a sequence of numbers at the end of the cycle (for example a spectrometer). In still another case the sensor may make several different types of measurements simultaneously.

It is apparent that the communication system may have to continuously transmit a single data signal or it may have to periodically transmit a single signal or it may have to periodically transmit a sequence of numbers or signals. The communication system must be adaptable to any of these possibilities. Furthermore to provide flexibility in the use of the tropospheric sensing system it is desirable, in many cases, to make provisions for control of the mode by which information is transmitted from the sensor to the data center by signals from the data center to the sensor location. To do this, the communication system must be capable of transmitting information in both directions (most communication systems are readily adaptable to two-way communication; so this is a minor additional requirement). It is also necessary to have circuitry at the sensor which decodes signals from the data center and controls the mode of data transmission accordingly. The sensor in this case functions as a transponder. This mode control could allow for such things as a change in the sampling rate, for disabling some of the functions of the sensor, for a change in the sensitivity of the measurement process, etc.

A diagram of a local tropospheric sensing system is shown in (F28.1). The area to be monitored will be denoted A . Area A is shown to be divided into N cells (49 in this case). Each cell is monitored by a single sensor. The area of a cell is $a = A/N$; and in the design of the system it is assumed that a single sensor can adequately monitor area a . In (F28.1) the boundaries between individual cells are sketched; it must be realized, however, that there can be no clear cut boundary line between cells. Sensors in most cases are more sensitive to areas closer to the sensor than to more remote areas. Thus the sensor output is effectively a weighted average over the approximate area sketched as its cell in (F28.1).

Each sensor must be connected to the data center by a communication link. The amount of information transmission capability required depends upon the geometric configuration of the transmission facility and upon the amount of information that must be transmitted by each sensor to the data center. It is customary to measure the rate of information transmission in bits per second. One bit of information can

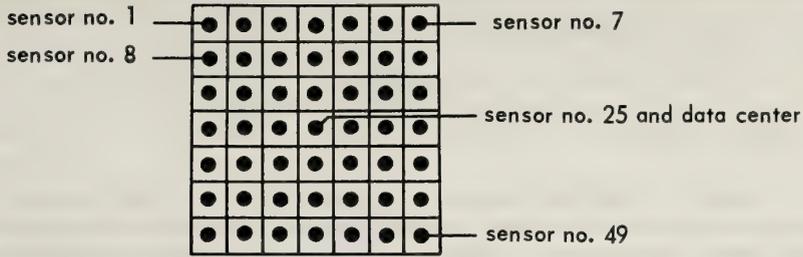


Figure 28.1 A local tropospheric monitoring system.

be represented by one binary digit if the information is perfectly encoded. Under these conditions the rate of transmission of information will be equal to the rate of binary digits representing that information. On the other hand with nonperfect encoding the binary digit rate is greater than the information rate. The extra digits will be redundant. Sometimes redundant digits are purposely added to data for error control purposes. Sometimes they exist in the output of sensors, and it is simply not worthwhile to recode the information so as to eliminate them. In any case the communication system must be designed to handle all of the binary digits (redundant or not).

In the present discussion let the output of the N sensors have data rates denoted by D_i , $i = 1, 2, \dots, N$ binary digits per second, average. The units of measurement for these data rates will hereafter be referred to as bits per sec. even though the encoding may not be perfect. That is these information rate figures are inflated by inclusion of any redundant digits.

It should be noted that the data rates are specified to be average. If the sensor output data rate is not uniform then buffer memories may be used to allow for transmission of the data without having excess capacity in the transmission system. If buffers are not available then the data rate specified for a sensor is the highest data rate that may occur.

The transmission capacity of communication links will be measured in km-bits per second which is the capacity required to transmit one bit per second a distance of one kilometer. The total transmission capacity† required for the system shown in (F28.1) is

$$T = \sum_{i=1}^N D_i d_i \tag{28:1}$$

where d_i is the distance along the transmission path from sensor i to the data center. In order to minimize the cost of the communication system, the data center should be placed so as to minimize T .

To gain insight into the amount transmission capacity needed assume all sensors have the same output data rate, D . The minimal value of T results from the use of straight line transmission paths from each of the sensors to the data center and from locating the data center in the position which will minimize T . This results in (see Appendix B)

$$T_{\min} = \frac{2}{3} N \left\{ \frac{A}{\pi} \right\}^{1/2} D = \frac{2}{3} N^{3/2} \left\{ \frac{a}{\pi} \right\}^{1/2} D ; \tag{28:2}$$

it is assumed in this analysis that the number, N , of sensors is not too small although the result is within a few percent down to only six or so sensors.

†The term channel capacity is avoided here since it pertains specifically to the rate at which data is transmitted without any regard for the distance it is transmitted. Transmission capacity is felt to be a more meaningful quantity in this case because it can be related to the cost of the communication system.

This system however involves a total length of the data transmission system (transmission line, radio link, or whatever type of communication system is used) that is

$$L = \frac{2}{3} N \left\{ \frac{A}{\pi} \right\}^{1/2} = \frac{2}{3} N^{2/3} \left\{ \frac{a}{\pi} \right\}^{1/2} \quad (28:3)$$

This may be excessive because with many data transmission systems it is less expensive to use a smaller total length of transmission equipment and a greater data (binary digit) rate over that shorter distance. This can be done using a system such as shown in (F28.2). It is shown in Appendix C that for this configuration of transmission facilities the transmission capacity, T, is, assuming as before that the average data rate for each sensor is D,

$$T = \frac{NA^{1/2}}{2} D = \frac{N^{3/2} a^{1/2}}{2} D ; \quad (28:4)$$

and the total length of the transmission system is

$$L = (NA)^{1/2} = Na^{1/2} \quad (28:5)$$

Two data transmission systems have been presented. The question arises as to which should be used. There is no clear cut answer to this question. The answer depends upon the type of information transmission system used. If the cost of the information transmission system is a linear function of the transmission capacity in km-bits/second then the configuration shown in (F28.1) is preferable because it minimizes the transmission capacity. However for many communication systems the length of the system has the primary influence on the cost. The channel capacity, that is the number of bits per second it will handle, has only a secondary influence. This is true, for example, in the case of a telephone system utilizing open wire or cable transmission facilities. The cost of construction of wire lines or cables (on poles or in underground trenches) will be substantial, but additional channel capacity can be achieved relatively inexpensively by simply adding additional wire pairs or adding multiplexing equipment. The open wire or cable transmission facility is perhaps the most clear cut example, but the same may be true for radio information transmission systems also.

If such a communication system is to be used then the configuration of the type shown in (F28.2) may be preferable. In (F28.2) the transmission facility (telephone lines for example) is long enough to link all of the sensors together. It is not intended to imply here that the transmission facility routing in (F28.2) gives the absolute minimum length because this depends on the shape of area A and other factors not considered in detail here. However it does give approximately the minimum length; and any improvement in routing would result in expressions T and L that depend upon a and N in nearly the same way.

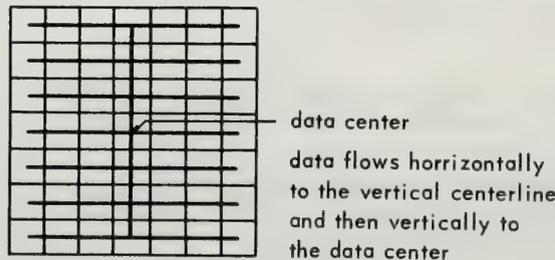


Figure 28.2 Sketch of a local data gathering network designed so as to avoid excessive transmission system length.

Comparison in the routing in (F28.1) and (F28.2) shows that the transmission capacity in the two cases depends upon N , a , and D in the same way but the constants in the expressions are such that, for the routing of (F28.2) T is less (approximately 25%) than for the case of the routing of (F28.1). However the length of the transmission facility in the two cases depends upon N in a different manner. In the case of the routing of (F28.1) the length is proportional to $N^{3/2}$ while the routing in (F28.2) is such that the length is a linear function of N . This may make the routing of (F28.2) preferable depending upon the communication system costs.

It has been suggested that large tropospheric sensing systems probably will be organized into several levels: the local level, the regional level, the national level, etc. because of legislative, judicial and other structures that make use of the data output of the sensing system. The question arises whether there is an optimum size for local remote sensing systems based upon the cost of the data gathering system. This can be studied by reference to the system shown in (F28.3). The entire figure represents a regional system made up of a number of local systems. Assume that the local systems are all the same size and consist of N sensors each producing data at the rate of D bits per second average. Let the total number of sensors in the entire regional system be N_1 then the number of local systems in the regional system is N_1/N . The transmission capacity required for a single local system (assuming communication links in the configuration shown in (F28.2)) is

$$\frac{1}{2} N^{3/2} a^{1/2} D. \tag{28:6}$$

Since there are N_1/N local systems, the total transmission capacity within all local systems is

$$(N_1/2)N^{1/2}a^{1/2} . \tag{28:7}$$

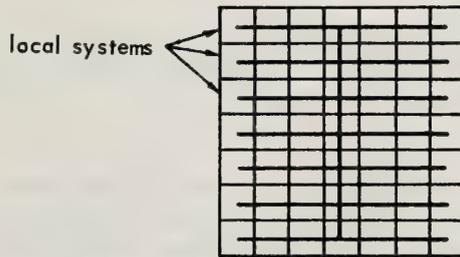


Figure 28.3 A regional sensing system consisting of N_1/N local sensing systems.

The transmission link configuration for the regional system will be assumed to be of the same type as the configuration of the local systems. Since there are N_1/N data sources (local systems) of the size $(Na)^{1/2}$ on each side each producing data at an average rate ND , the regional communication links (connecting the local data centers to the regional data center) requires a transmission capacity

$$\frac{1}{2} (N_1/N)^{3/2} (Na)^{1/2} ND = \frac{1}{2} N_1^{3/2} a^{1/2} D \tag{28:8}$$

The total transmission capacity, T_t , required in the regional system is

$$T_t = \frac{N_1}{2} N^{1/2} a^{1/2} D + \frac{1}{2} N_1^{3/2} a^{1/2} D = \frac{N_1 a^{1/2} D}{2} (N^{1/2} + N_1^{1/2}) \tag{28:9}$$

It is seen from this expression that the total transmission capacity is minimum for $N = 1$ that is for the case in which there is no local system but in which each sensor is connected directly to the regional center. This line of reasoning suggests that there should be only one level for the world wide sensing system. It may be asked then why several levels were suggested. The answer is although a world wide system with each sensor connected to a single data center would require the least transmission capacity it would not be practical because of several factors. The most important factor is the fact that the data obtained from sensing systems is needed on the various levels suggested before (local, regional, national, etc.). Another important reason is concerned with the manner with which sensing systems will be built up and expanded. They will be built first for local areas where the need is the greatest and only later will these local sensing systems be organized into regional and larger systems.

Another possibility that should be mentioned is that of leasing a communication facility. This often may be the preferred procedure. In that case the above cost analysis is still useful because leasing costs should be proportional to the cost expressions given above.

Communication systems for the transmission of information in local tropospheric sensing systems have been discussed in some detail. These communication systems are somewhat unique because the communication needs are somewhat unusual. The needs are for a low rate of transmission continuously or at regular intervals from a large number of sensors. Communication between local data centers and between local data centers and regional centers, on the other hand, is a simpler problem because such data communication is more conventional and conventional data commercially available communication channels can be used. For this reason nothing more will be said here about data communication systems between data centers.

One further aspect of information transmission systems concerns radio transmission. If the system uses radio information transmission systems then it is necessary to have frequency assignments as necessary for the transmitters. Since some systems will require radio communication links and since the frequency spectrum is overcrowded, consideration should be given early to obtaining the necessary frequency assignments. The assignments needed for tropospheric sensing throughout the world are likely to be great, so great in fact that it may be necessary to use assignments presently used for other purposes. It may take much time to obtain the necessary assignments for world wide tropospheric remote sensing; so it would be wise to initiate the efforts soon.

Another need for frequency assignments is for use by sensors which use transmission of electromagnetic radiation in the performance of the monitoring activity. Frequency assignments must be obtained for these also.

28.2 Data analysis

The output of a tropospheric sensor is a random signal owing to the randomness of the characteristics of the troposphere. In other words the output of a sensor is a sample function of a random process. The objective of tropospheric sensing is the characterization of this random process which in turn is a partial characterization of the troposphere. Random processes and the determination of their characteristics is discussed in Appendix D.

The output of a tropospheric sensor may require little or no processing before transmission to the data center or it may require substantial processing. Actually the system designer has considerable choice in the matter because the processing can be done at the sensor or at the data center. However for some types of sensors and for locations at which there are several sensors some data processing is almost essential to put the signal into proper form for transmission to the data center.

Before signal conditioning and processing are discussed further something should be said about types of signals. In general, signals are of two types, analog and digital. Most sensors produce analog output signals. Such signals are easily handled by communication systems. It may seem logical then to simply transmit analog output signals to the data center. However further thought on the subject indicates this may not be wise. It may be necessary to store the signals at the data center indefinitely and it may be necessary to be able to recall the signals quickly in a form suitable for processing on a digital computer. It will also often be necessary to store signals temporarily at sensor locations. These requirements are more easily met by digital storage. Digital storage of information has additional advantages concerned with error control during transmission of information, adaptability of stored data to various processing systems and procedures, ease with which monitoring procedures can be modified, the simplicity and wide spread

availability of digital storage devices, and the simplicity of standardization of data formats from different types of sensors. For these reasons it is strongly recommended that sensors be designed to have digital outputs. If sensors characteristically have analog outputs then analog to digital converters could be added to the circuits to change the outputs to digital form by a sampling process. Sampling theory is discussed in Appendix A.

28.3 References

Bendat, J. S., A. G. Piersol, 1966: *Measurement and Analysis of Random Data*, Wiley, New York.

Carlson, A. B., 1968: *Communication Systems*, McGraw-Hill, New York.

Hancock, J. C., 1961: *An Introduction to the Principles of Communication Theory*, McGraw-Hill, New York.

Stein, S., J. J. Jones, 1967: *Modern Communication Principles*, McGraw-Hill, New York.

Appendix A – Sampling Theory

An analog signal such as $f(t)$ shown in (F28A.1) has a finite information rate provided it has no power above a certain upper limiting frequency, that is if it is band limited. If this is true then all of the information the signal possesses can be determined from the values of the amplitude of the signal at a regularly spaced set of points on the time scale. The spacing of the points must be such that there are at least two points (average) during each cycle for the upper limiting frequency at which power is present in the signal. The determination of the amplitudes at these points is called sampling and the individual numbers representing these amplitudes are called samples. The sampled version of the signal of (F28A.1) is shown in (F28A.2). The pulses representing the samples need have only infinitesimal width; in practical sampling, of course, small but finite width is used. The sampling theorem implies that the information in (F28A.2) is sufficient to exactly recover the entire signal of (F28A.1). In fact all that is required is to pass the signal of (F28A.2) through a low pass filter and the output will be proportional to the signal of (F28A.1). Sampling is useful because a sampled signal has much unused space on the time scale between samples. It is possible to use this space for other signals.

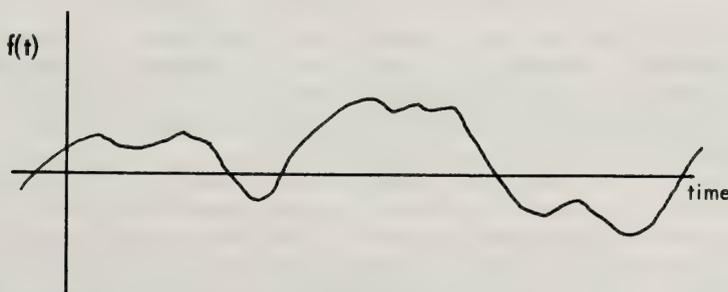


Figure 28A.1 *An analog signal.*

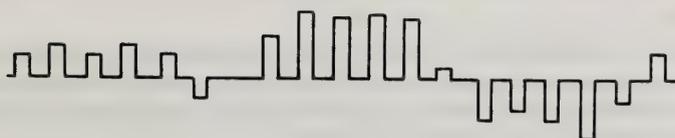


Figure 28A.2 *Sampled version of the signal of (F28A.1).*

In the case of a narrow band signal another version of the sampling theorem is useful. If the signal has power only in the band of frequencies from $f_m - B$ to f_m then the sampling rate must be $2f_m/k$ where k is the largest integer not exceeding f_m/B . The number of samples then is between $2B$ and $4B$ approaching $2B$ for narrowband signals having a bandwidth small compared with the center frequency of the band (Hancock, 1961). Actually a sampling rate of $2B$ is always adequate if the signal is first shifted down to the band from 0 to B .

The sampling of a signal may be thought of as the multiplication of the signal by a sampling function. A sampling function is shown in (F28A.3). The pulse width and height are related so that the product is unity, but this is not a restriction since a multiplicative constant will be allowed which permits any relationship between the height and duration. When a signal, $x(t)$ is multiplied by a sampling function the result is as shown in (F28A.4). This may be called a shaped top sampled signal.

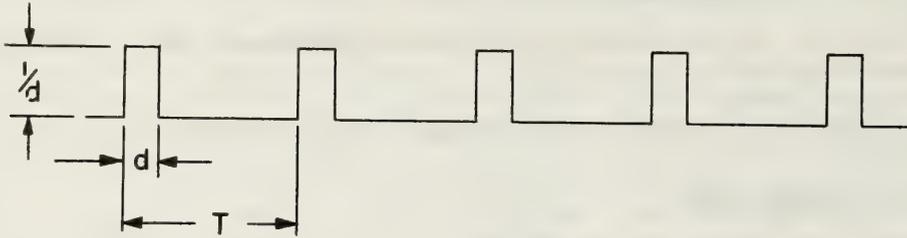


Figure 28A.3 Sampling functions.

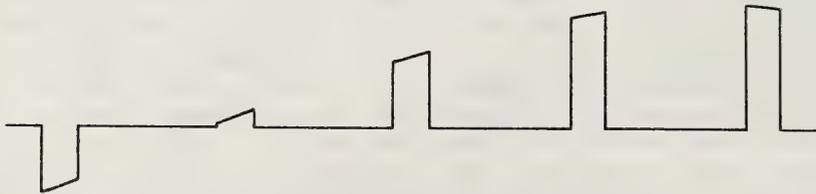


Figure 28A.4 Shaped top sampled signal.

If this sampled signal is transmitted over a communication system that preserves the shape of the tops of the pulses then the signal $x(t)$ can be recovered to any desired degree of accuracy from the sampled signal provided that there is no power in the signal above the frequency f_m (Stein and Jones, 1969). Preservation of the shape of the tops of the pulses requires an analog communication system. It is often inconvenient to have to preserve the shapes of the tops of the pulses. The importance of such shape preservation can be diminished by reducing the pulse duration.

In the limit as $d \rightarrow 0$ this produces a sampling function consisting of a series of impulses and the sampled signal consists of a sequence of impulses of strength, at each sampling point, equal to the signal amplitude at that point. This mathematical idealization is called ideal sampling. Impulses cannot be handled by communication systems, but the sequence of impulses can be represented by a sequence of flat topped pulses each pulse having a height proportional to the strength of the impulse it represents. Such a sequence of flat topped pulses can be handled more simply than shaped topped pulses by a communication system. The fact that the pulse is of finite width rather than being an impulse causes some distortion in the signal recovery circuitry called aperture effect; however, this distortion can be completely eliminated by the use of a suitably designed equalization filter (Carlson, 1968).

The sequence of flat topped pulses can be handled by analog or digital communication systems. If a digital communication system is used the pulse height will be quantized. This amounts to changing each pulse height to the nearest one of the quantized levels. These quantized pulses can then be handled by some sort of pulse modulation scheme, PCM, for example

The quantization process, since it alters the pulse heights, distorts the signal. This distortion is called quantization noise; the extent to which quantization noise degrades the signal is discussed below.

Suppose a signal ranging in amplitude from $-V_m/2$ to $V_m/2$ and is quantized into n steps of magnitude, s , as shown in (F28A.5). The input to the quantizer is V_i and the output is V_o . The operation of the quantizer may be described as follows. If

$$(j - 1)\frac{s}{2} < V_i < (j + 1)\frac{s}{2} \tag{A:1}$$

then $V_o = js$ where j is an integer

$$-\frac{n-1}{2} \leq j \leq \frac{n+1}{2} \tag{A:2}$$

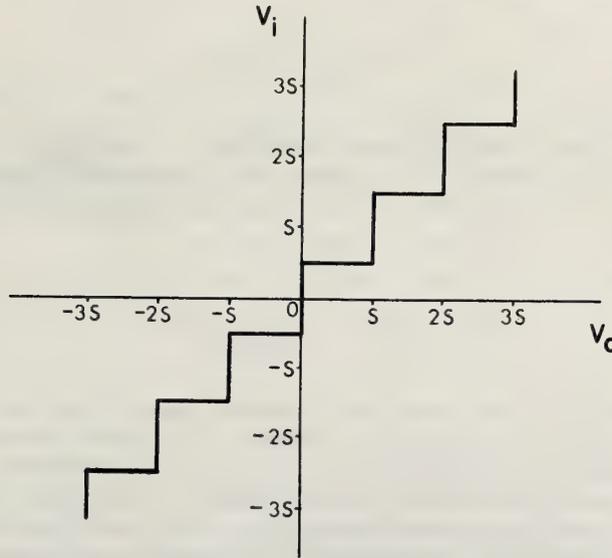


Figure 28A.5 Characteristic of a seven level quantizer.

and it is assumed that n is an odd integer. It is apparent that V_i and V_o can differ at most by the magnitude $s/2$. This difference $V_q = V_i - V_o$ is called quantization noise and it effects the signal in the same way as additive random noise (Stein and Jones, 1967); but it is not like gaussian noise because it is peak limited. If the statistics of the input voltage are such that the quantization noise amplitude (V_q) is uniformly distributed in the range

$$-\frac{s}{2} < V_q < \frac{s}{2} \tag{A:3}$$

then it has a mean square of

$$\bar{V}_q^2 = \frac{s^2}{12} \tag{A:4}$$

The effect of quantization noise on the signal can be expressed in terms of a signal to noise ratio (SNR). In order to determine the signal to noise ratio some assumptions must be made concerning the amplitude distribution of V_i . If V_i is uniform in the interval

$$-\frac{V_m}{2} < V_i < \frac{V_m}{2} \tag{A:5}$$

then the mean square of V_i is approximately

$$\bar{V}_i^2 = \frac{V_m^2}{12} \quad (\text{A:6})$$

but $V_m = ns$, so

$$\bar{V}_i^2 = \frac{n^2 s^2}{12} \quad (\text{A:7})$$

and the quantization signal to noise ratio is

$$\text{SNR} = \frac{n^2 s^2}{12} \left(\frac{s^2}{12} \right)^{-1} = n^2 \quad (\text{A:8})$$

If the amplitude distribution of the input signal is not uniform the result can vary somewhat from n^2 ; but this gives a reasonably good estimate of the degrading effect of quantization. Another characterization of quantization noise is in terms of the RMS (root mean square) error amplitude as a proportion of the RMS signal amplitude, this is

$$(\bar{V}_q^2)^{1/2} / (\bar{V}_i^2)^{1/2} = \left(\frac{1}{\text{SNR}} \right)^{1/2} = \frac{1}{n} \quad (\text{A:9})$$

In the discussion thus far it was assumed that rectangular pulses are used in the sampling process. Rectangular pulses, of course, are mathematical idealizations that can only be approximated in practice. However, rectangular pulses are not necessary. Pulses of any shape can be used regardless of the particular sampling process being used. The details are well known and will not be repeated here (Carlson, 1968).

Recovery of a signal from a sampled representation of that signal is not possible unless the signal originally sampled was band limited. More specifically if the signal $x(t)$ has no power at frequencies greater than f_m and if it is sampled at $2/f_m$ samples per second (that is two samples per cycle at the frequency f_m) then if $x(t)$ has any power at frequencies greater than f_c then the accuracy of recovery is limited. The character of these limitations will now be discussed. Suppose $x(t)$ has a power spectrum as sketched in (F28A.6). The sampling process reproduces the power spectrum at intervals of $2f_m$ on the frequency axis. The reproductions may be exact or modified depending upon the sampling process. In any case $x(t)$ can be recovered to any desired degree of accuracy from the sampled signal by passing it through a low pass filter with cut-off frequencies f_m and also an equalization filter if needed for the particular sampling process in use.

If $x(t)$ has power at frequencies above f_m as shown in (F28A.7a), then the reproductions of the spectrum of $x(t)$ overlap (F28A.7b). It is apparent that the signal emerging from the recovery circuit will have a spectrum such as that shown in (F28A.8). The recovered signal has lost its frequencies above f_m thereby distorting the signal, but those lost frequencies have effectively been shifted to lower frequencies (aliased) and added back in. This adding back in distorts the signal further.

To minimize distortion, the signal $x(t)$ should be passed through a low pass filter with cut off frequency f_m before being sampled. This filtering distorts the signal, but subsequent sampling then recovering the signal does not introduce any further distortion. The end result is only half as much as distortion as without presampling filtering.

Sampling Theory has been discussed in terms of functions of time, but other variables, distance for example, may be used in place of time. Furthermore the extension to more than one variable is straightforward.

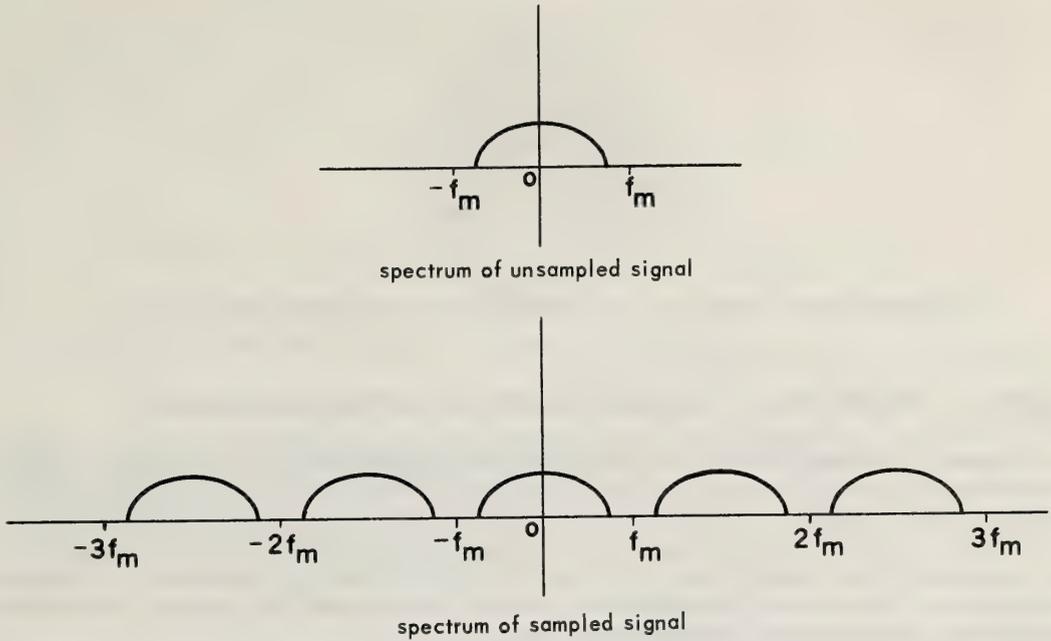


Figure 28A.6 Power spectrum of an unsampled signal with no power above frequency f_m and of the same signal after sampling at $2f_m$ samples per second.

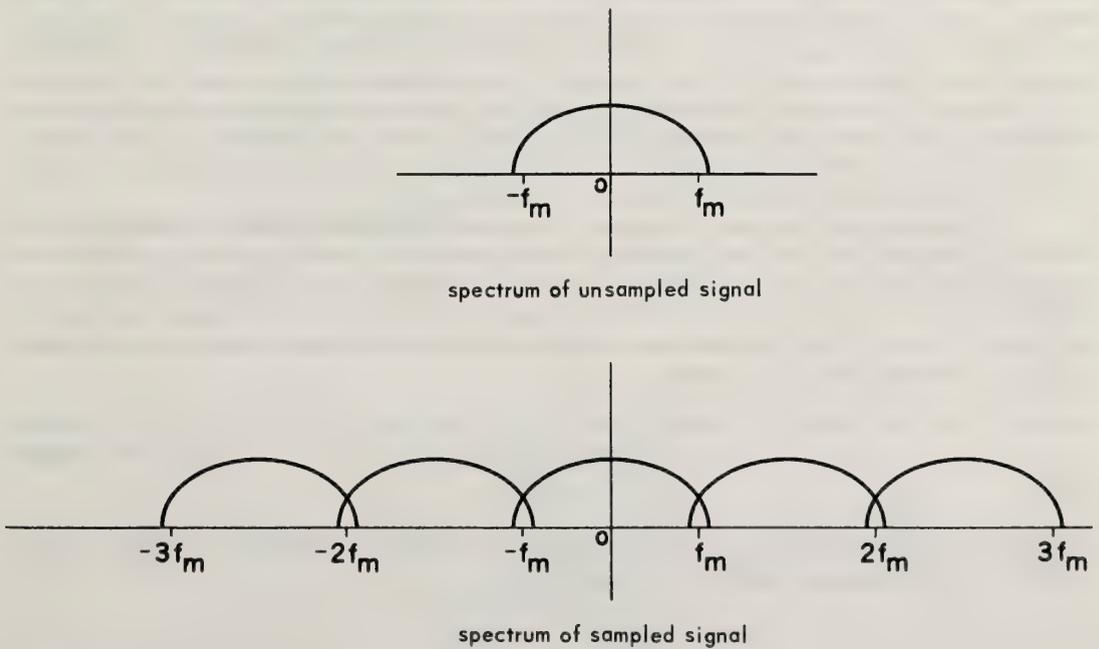


Figure 28A.7 Power spectrum of an unsampled signal with finite power above frequency f_m and of the same signal after sampling at $2f_m$ samples per second.

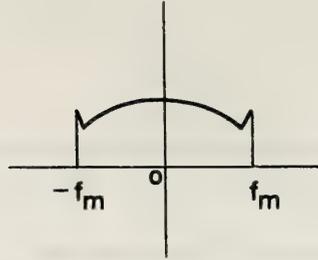


Figure 28A.8 Power spectrum of the signal recovered from the sampled signal of (F28A.7).

Appendix B – Minimum Values of Transmission Capacities for Local Data Gathering Systems

Assume a monitoring network which monitors an area A , on the surface of the earth with N sensors. Assume that each sensor is connected to a data center by a communication link. It is desired to find the minimum possible transmission capacity required for all communication links. The transmission capacity is defined to be the number of bits per second that can be handled by the link multiplied by its length. Thus the dimensions will be kilometers-bits per second. This minimum will be found assuming all sensors transmit data at the same rate, D bits per second. It will be further assumed that all sensors monitor equal areas; thus the area monitored by a single sensor is $A/N = a$.

The problem is to find the geometric configuration of N nonoverlapping areas and the position of the data center such that the sum of the distances from the centroids of the N areas to the data center is minimum. It is easily argued that the configuration must be symmetric about the data center. If there are no restrictions on the shapes of the areas they should be chosen as concentric annuluses. This shape, however, is not a practical solution to the problem since each sensor should be expected to monitor an area consisting of all points within a certain distance. This implies the area monitored by a single sensor is a circle. Circles, however, cannot be fitted together in a nonoverlapping fashion to make up a larger area; so for purposes of this analysis, which is to find a lower limit for the transmission capacity, the small areas monitored by single sensors will be permitted to vary from a circular shape sufficiently to allow the areas to fit together to completely fill the larger area A . A deformation from the circle to a sector of an annulus (or a circle in the limiting case near the data center) will be permitted.

It is apparent, then, that the configuration for minimum transmission capacity is approximately† sketched in (F28A.9). The total area monitored is A ; each sector of an annulus has area a , and there are N such areas. The data center is at the center of the circle. It is assumed that N is sufficiently large that the number of sectors in each annulus, except perhaps several near the center, is much greater than one. For large values of N and for the outer annuluses the small areas will be approximately square and the dimension of a side of the square will be $a^{1/2}$ where

$$a^{1/2} = (A/N)^{1/2} \quad . \quad (B:1)$$

This, then, is the radial dimension of each annulus and the number of annuluses is approximately the radius of a circle having area A divided by $a^{1/2}$ or

$$(A/\pi)^{1/2}/(A/N)^{1/2} = (N/\pi)^{1/2} \quad . \quad (B:2)$$

†It is recognized that this analysis lacks precision, but it is felt that the results will be sufficiently accurate for the present objectives.

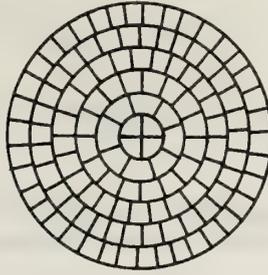


Figure 28A.9 An arrangement of cells which minimizes the transmission capacity of a sensing system.

The radius of the i 'th annulus from the data center is $(A/N)^{1/2}$; the circumference is $2\pi i(A/N)^{1/2}$. The number of segments in the i -th annulus from the data center is $2\pi i(A/N)^{1/2}/(A/N)^{1/2} = 2\pi i$. The total transmission capacity is

$$\begin{aligned}
 T_{\min} &= \sum_{i=1}^{(N/\pi)^{1/2}} i(A/N)^{1/2} 2\pi i D \\
 &= 2\pi(A/N)^{1/2} \frac{D}{6} (N/\pi)^{1/2} [(N/\pi)^{1/2} + 1] [2(N/\pi)^{1/2} + 1] \\
 &\cong (2/3) (A/N)^{1/2} \pi(N/\pi)^{3/2} D = (2/3) N(A/\pi)^{1/2} D \quad .
 \end{aligned}
 \tag{B:3}$$

In terms of $a = A/N$, the area covered by a single sensor, this is

$$T_{\min} \cong (2/3) N^{3/2} (a/\pi)^{1/2} D \quad .
 \tag{B:4}$$

The total length of the transmission system is

$$L = T_{\min}/D = (2/3) N^{3/2} (a/\pi)^{1/2} \quad .
 \tag{B:5}$$

It is of interest to note that for cases in which the small areas that make up the area A are concentric annuluses, as in (F29A.10), the transmission capacity can be found as follows. There are N concentric annuluses each of area $a = A/N$. The radius of the i -th annulus is r_i where

$$r_i = i^{1/2} (A/\pi N)^{1/2} \quad .
 \tag{B:6}$$

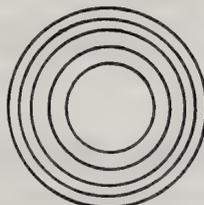


Figure 28A.10 An arrangement of annular areas which minimizes the transmission capacity of a sensing system.

The transmission capacity is approximately

$$\begin{aligned} \sum_{i=1}^N r_i D &= (A/\pi N)^{1/2} \sum_{i=1}^N i^{1/2} D \approx (A/\pi N)^{1/2} D \int_0^N x^{1/2} dx \\ &= (A/\pi N)^{1/2} (2/3)N^{3/2} D = (2/3)N(A/\pi)^{1/2} D = (2/3)N^{3/2}(a/\pi)^{1/2} D \end{aligned} \tag{B:7}$$

In this derivation a summation was approximated by an integral; this is justified if N is sufficiently large. The result is the same as was obtained for the configuration shown in (F28A.9). The total length of the transmission system is also the same.

Appendix C – Minimum Length Data Gathering Systems

In Appendix B a communication system, of minimum transmission capacity, for a sensing system was studied. If it is desired to minimize the length rather than the transmission capacity a different routing of the communication links is required; for example consider the routing of (F28A.11). Assume the output data rate is D bits per second from each sensor. Let the data center be at the center of the square area of (F28A.11), and let the transmission link run in the serpentine pattern shown. At each point the transmission link carries data from all the sensors beyond that point (from the data center). Each sensor is at the center of a square of width $(A/N)^{1/2}$. The length of the transmission path would not change at all if all of the small squares were placed side by side in a single row. In this case the length of the row of squares would be $A/(A/N)^{1/2}$. Then subtracting half the widths of the two end squares the length of the transmission link; it is

$$A/(A/N)^{1/2} - (A/N)^{1/2} = (NA)^{1/2} - (A/N)^{1/2} \tag{C:1}$$

or approximately $(NA)^{1/2}$. In terms of $a = A/N$ this is $Na^{1/2}$. This is considerably shorter than the total length of the system considered in Appendix B; but the total transmission capacity, T, is greater. The transmission capacity will be found in the following paragraph.

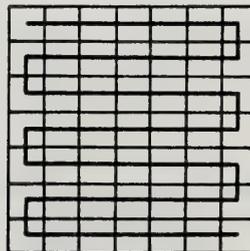


Figure 28A.11 A sensing system having a serpentine transmission link.

On each side of the data center there are $N/2$ sensors with spacing $(A/N)^{1/2}$. Progressing along the transmission link toward the data center the data rate increases by D as each sensor is traversed. The total transmission capacity is approximately

$$\begin{aligned}
 T &= 2 \sum_{i=0}^{N/2} \left[\frac{(NA)^{1/2}}{2} - \frac{(NA)^{1/2}}{2N} - i(A/N)^{1/2} \right] D \\
 &= 2 \left[\frac{(NA)^{1/2}}{2} - \frac{(NA)^{1/2}}{2N} \right] \left[\frac{N}{2} + 1 \right] D - (A/N)^{1/2} \left[\frac{N}{4} \right] \left[\frac{N}{2} + 1 \right] D \\
 &= 2 \left[\frac{N^{3/2} A^{1/2}}{4} + \frac{(NA)^{1/2}}{2} - \frac{(NA)^{1/2}}{4} - \frac{(NA)^{1/2}}{2N} - \right. \\
 &\quad \left. \frac{N^{3/2} A^{1/2}}{8} - \frac{N^{1/2} A^{1/2}}{4} \right] D \\
 &\cong 2 \left[\frac{N^{3/2} A^{1/2}}{4} - \frac{N^{3/2} A^{1/2}}{8} \right] D = \frac{N^{3/2} A^{1/2}}{4} D
 \end{aligned} \tag{C:2}$$

In these expressions it is assumed that N is sufficiently large that only terms involving the highest power in N need be retained. In terms of $a = A/N$, the area monitored by a single sensor, this is

$$T = \frac{N^2 a^{1/2}}{4} \tag{C:3}$$

Suppose that the data center is at one end of the serpentine pattern of the transmission link rather than at the center then the transmission capacity, T_1 , is approximately

$$\begin{aligned}
 T_1 &= \sum_{i=0}^{N-1} \left[(NA)^{1/2} - \frac{(NA)^{1/2}}{N} - i(A/N)^{1/2} \right] D \\
 &= \left[(NA)^{1/2} N - \frac{(NA)^{1/2}}{N} N - (A/N)^{1/2} \left[\frac{N-1}{2} \right] N \right] D \\
 &= \left[N^{3/2} A^{1/2} - N^{1/2} A^{1/2} - \frac{N^{3/2} A^{1/2}}{2} + \frac{N^{1/2} A^{1/2}}{2} \right] D \\
 &\cong \frac{N^{3/2} A^{1/2}}{2} D = \frac{N^2 a^{1/2}}{2} D
 \end{aligned} \tag{C:4}$$

where, again, it is assumed that N is sufficiently large that only terms of the highest power in N need be retained. It is seen that the transmission capacity doubles if the data center is placed at one end of the serpentine shaped transmission link rather than at the center.

It is observed that for the configuration of minimum transmission capacity discussed in Appendix B the capacity is proportional to $N^{3/2}$, $a^{1/2}$ and to D and the length is proportional to $N^{3/2}$ and to $a^{1/2}$. In this appendix using the serpentine pattern shown in Fig. 1 the total transmission capacity is proportional to N^2 , $a^{1/2}$ and to D ; and the total length is proportional to N and to $a^{1/2}$. The behavior of the transmission capacity as a function of N is superior to that of the configuration of Appendix B; but the behavior of the length is superior for the serpentine configuration shown in (F28A.11) of this appendix. It would be desirable to find a configuration with the advantages of both. Such a configuration is possible; it is shown in (F28.2.14). This configuration is such that the transmission link from any sensor to the data center (at the center of the square) runs horizontally (except for those sensors on the vertical center line of the array) to the vertical center line and then it runs vertically to the data center. Obviously the total length of this transmission system is the same as for the serpentine configuration in (F28A.11); this is

$$(NA)^{1/2} - (A/N)^{1/2} \cong (NA)^{1/2} \quad (C:5)$$

which in terms of a is $Na^{1/2}$. The transmission capacity will now be found. Each of the horizontal branches to the left and to the right of the vertical center line of the array has the same transmission capacity. There are $N^{1/2}/2$ sensors in each of these branches and the branch has length

$$A^{1/2}/2 - (A/N)^{1/2}/2 \quad (C:6)$$

The analysis of the transmission capacity varies somewhat depending on whether $N^{1/2}$ is an integer and upon whether the number of sensors in a single row is even or odd. An approximate analysis will be presented which produces results of good accuracy for all cases if N is large. The transmission capacity of one horizontal branch on one side of the vertical center line is

$$\begin{aligned} & \sum_{i=0}^{N^{1/2}/2} (A^{1/2}/2 - (A/N)^{1/2}/2 - i(A/N)^{1/2})D \\ &= \{[A^{1/2}/2 - (A/N)^{1/2}/2] (N^{1/2}/2 + 1) - (A/N)^{1/2} (N^{1/2}/4) (N^{1/2} + 1)\}D \quad (C:7) \\ &= [N^{1/2}A^{1/2}/4 - A^{1/2}/4 + A^{1/2}/2 - N^{1/2}A^{1/2}/2 - N^{1/2}A^{1/2}/8 \\ &\quad - A^{1/2}/4]D \\ &\cong N^{1/2}A^{1/2}D/8 = Na^{1/2}D/8 \end{aligned}$$

where it is assumed again that N is large enough that all terms except those of highest power in N may be neglected. There are $2(N)^{1/2}$ of these branches; so the total transmission capacity in the horizontal direction is

$$2N^{1/2}Na^{1/2}D/8 = N^{3/2}a^{1/2}D/4 \quad (C:8)$$

In the vertical direction the transmission link increases its data rate by $N^{1/2}D$ as it crosses each row of sensors progressing toward the data center. Therefore the data capacity of the vertically oriented transmission links is

$$\begin{aligned} & 2 \sum_{i=0}^{N^{1/2}/2} (A^{1/2}/2 - (A/N)^{1/2}/2 - i(A/N)^{1/2})N^{1/2}D \quad (C:9) \\ & \cong NA^{1/2}D/4 = N^{3/2}a^{1/2}D/4 \end{aligned}$$

where again only the highest powers in N are retained. The total transmission capacity of the system of (F28 A.2) is

$$T \cong NA^{1/2}D/4 + NA^{1/2}D/4 = NA^{1/2}D/2 \quad (C:10)$$

or

$$T \cong N^{3/2} a^{1/2} D/2 \quad (C:11)$$

It is now seen that for this configuration the transmission capacity varies with $N^{3/2}$ instead of N^2 as for the configuration of (F28A.11). Thus the configuration of (F28.2) retains the short length of the serpentine configuration of (F28A.11) and yet it has transmission capacity that varies with N in the same way as the configuration analyzed in Appendix B.

Appendix D – Description of Random Processes

A random process is an ensemble of random functions. The ensemble will be denoted by $\{x(t)\}$ and individual sample functions will be denoted $x_1(t)$, $x_2(t)$, $x_3(t)$, etc. The independent variable, t , is time, but it could just as well be a space coordinate. Alternatively, there may be multiple independent variables involving time and space and possibly other variables.

A random process can be partially described by a variety of characteristics. The simplest is the mean,

$$\mu_{x(t_1)} = \langle x(t_1) \rangle, \quad (D:1)$$

at a specified time, t_1 . This is simply the expected value at time, t_1 , obtained by averaging over the ensemble (the angular brackets will denote an ensemble average). Another characteristic is the auto correlation function $R_x(t_1, t_1 + 2) = \langle f(t_1)f(t_1 + 2) \rangle$. Other characteristics include higher moments,

$$\langle [x(t_1)]^n \rangle, \quad (D:2)$$

at various values of time and joint moments,

$$\langle x(t_1) x(t_2) \rangle, \quad (D:3)$$

at two or more values of time. Other partial descriptions include probability density functions such as $p[x(t_1)]$ for the distribution (over the ensemble) of x at time t_1 . Joint density functions may also be written for the amplitude at a set of values of time for example

$$p [x(t_1), x(t_2), x(t_3)] . \quad (D:4)$$

It is apparent that there are an unlimited number of partial descriptions of a general random process. A complete description is not possible from a practical point of view.

Random processes are broadly classified as stationary or nonstationary. The process is said to be strongly stationary (or stationary in the strict sense) if all of the moments and joint moments are independent of time. A somewhat weaker classification is that of weak stationarity (or stationarity in a wide sense); this categorizes random processes for which the mean and the autocorrelation function (but not all higher moments or joint moments) are independent of time.

Stationarity is a characteristic of ensembles. The same concept applied to an individual random function is called self-stationarity. A function is said to be self-stationary if characteristics (mean, variance,

etc.) calculated in different time intervals do not differ more than would be expected due to statistical fluctuations. Self-stationarity, like stationarity, can be strong or weak.

If a random process is stationary (in either of the senses mentioned above), it may be ergodic or non-ergodic. It is ergodic if the mean and auto correlation function can be determined by averaging over time rather than over the ensemble; that is if

$$\langle x(t) \rangle = \overline{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt, \quad (D:5)$$

$$\langle x(t) x(t+\tau) \rangle = \overline{x(t) x(t+\tau)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) x(t+\tau) dt \quad (D:6)$$

where the bar denotes a time average.

A widely used partial description of random processes is the power spectral density function $G_x(f)$. It is a function of frequency, f , and at any frequency is proportional to the mean square (power) of the output of a narrow band filter having a sample function of the random process as an input. For a stationary random process, the power spectral density function is the Fourier transform of the auto correlation function.

Often a partial description of the relationship between two random processes is of interest. One such relation is the cross-correlation function

$$R_{xy}(t_1, t_2) = \langle x(t_1) y(t_2) \rangle \quad (D:7)$$

where $\{x(t)\}$ and $\{y(t)\}$ are the two processes in question. For stationary processes the cross-spectral density function $G_{xy}(f)$ may be defined as the Fourier transform of the cross-correlation function.

Sometimes it is convenient to use the co-variance and cross-co-variance functions rather than correlation and cross correlation functions. The co-variance function for stationary processes is

$$C_x(\tau) = [x(t) - \overline{x(t)}] [x(t+\tau) - \overline{x(t)}] = x(t) x(t+\tau) - [\overline{x(t)}]^2 \quad (D:8)$$

and the cross co-variance function is

$$C_{xy}(\tau) = [x(t) - \overline{x(t)}] [y(t+\tau) - \overline{y(t)}] = x(t) y(t) - [\overline{x(t)}] [\overline{y(t)}] \quad (D:9)$$

The cross co-variance function is sometimes normalized to give a correlation coefficient function which is given by

$$\rho_{xy}(\tau) = \frac{C_{xy}(\tau)}{[C_x(0) C_y(0)]^{1/2}} \quad (D:10)$$

Another partial description of the relationship between two random processes is the coherence function defined by Bendat and Piersol (1966)

$$H_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_x(f) G_y(f)} \quad (D:11)$$

In the processing of data (sample functions of a stationary random process) to obtain means, variances, co-variances, power spectral density functions, etc. The accuracy of the result, that is, the closeness of the result to the correct result for the random process in question depends upon the amount of data that enters into the analysis. For high accuracy the length of the data record (measured in time) should be long. A shorter record will require less computation, but it will also result in a larger expected error.

The length of record needed for a specified accuracy is dependent upon the type of calculation being made. This general problem is treated in detail in the literature (Bendat and Piersol, 1966). Space and time limitations preclude any attempt to even summarize the problem here, but some rough guidelines will be given.

Assume that the data is band limited with a bandwidth B. Then for a uniform power spectrum within B, the length, T_r , of data required to estimate the mean within a normalized standard error, ϵ (ϵ is the ratio of the standard deviation of the estimate of the parameter under computation to the true value of that parameter) is

$$T_r = \frac{1}{2\epsilon^2 B} \left(\frac{\sigma}{\mu}\right)^2 \quad (D:12)$$

where μ and σ are the true mean and standard deviation.

An estimate of the second moment of a random process requires

$$T_r = \frac{1}{\epsilon^2 B} \quad (D:13)$$

In order to find the probability density function for the amplitude of a random process, let the range of values of the random variable be divided into sub-intervals. The proportion of time that the amplitude is in each sub-interval is determined to give a step function approximation to the probability density function. The accuracy of the result may be determined by comparison with probabilities calculated from the true distribution assuming the same sub-intervals. When this is done it may be concluded that the length of data needed is approximately

$$T_r = \frac{1}{2\epsilon^2 BA} \quad (D:14)$$

where A is the area under the density function curve for the sub-interval for which the density function is being calculated. A^{-1} will be the number of points on the curve.

An estimate of the value of the auto correlation function $R(\tau)$ for argument τ_0 requires data of length

$$T_r = \frac{1}{2\epsilon^2 B} \left\{ 1 + \left[\frac{R(0)}{R(a)} \right]^2 \right\} \quad (D:15)$$

An estimate of the power spectral density function may be made by dividing the range of frequency for which power is present into sub-intervals. The power in each sub-interval is calculated to give a step function approximation to the spectral density function. The accuracy of this result can be evaluated by comparison with results obtained from the true spectral density function for the same choice of sub-intervals. When this is done, it is found that the length of data required for accuracy, ϵ , is approximately

$$T_r = \frac{N}{\epsilon^2 B} \quad (D:16)$$

where N is the number of sub-intervals and also the number of points on the curve.

It should be remembered that the results of this section are approximate and are specifically for the case in which the spectral density is uniform across the bandwidth B . Furthermore the above results are less accurate and probably should not be used for larger values of ϵ than about 0.2. For more accurate results or for other conditions the reader is referred to other sources (Bendat, and Piersol, 1966). If it desired to state an approximate criterion for the length of data required, that is, an order of magnitude estimate which might be committed to memory for quick approximate calculations it could be: the length of data required is $1/(2B\epsilon^2)$ for an estimate of the mean or of a single point on the probability density function curve. The length of data required for an estimate of the second moment or of a single point on the autocorrelation function curve or of a single point on the spectral density curve is $1/(B\epsilon^2)$. For estimates of curves the record lengths should be several times as long.

List of Symbols

A	Area to be monitored by a local sensing system	V_q	Magnitude of quantization error
N	Number of sensors in a local sensing system	SNR	Signal to noise ratio
a	Area to be monitored by a single sensor	μ	Mean of a random variable
D	Data rate in binary digits per sec.	σ	Standard deviation of a random variable
d_i	Distance from sensor i to the data center	R_x	Autocorrelation function for the random process $\{x(t)\}$
T	Transmission capacity in km-bits per sec.	C_x	Co-variance function for the random process $\{x(t)\}$
N_1	Number of local sensing systems in a regional system	C_{xy}	Cross co-variance function for random processes $\{x(t)\}$ and $\{y(t)\}$
f_m	Maximum frequency at which power is present in a signal	ρ_{xy}	Correlation coefficient function for random processes $\{x(t)\}$ and $\{y(t)\}$
k	Largest integer not exceeding f_m/F	H_{xy}	Coherence function for random processes $\{x(t)\}$ and $\{y(t)\}$
d	Duration of a sampling pulse	T_r	Data record length
PCM	Pulse code modulation	ϵ	Normalized standard error
V_m	Maximum amplitude of a signal	B	Signal bandwidth
V_i	Amplitude of input signal	G_x	Power spectral density of random process $\{x(t)\}$
V_o	Amplitude of output signal		
n	Number of quantization steps		
s	Magnitude of quantization steps		

Chapter 29: CHANNEL MODELING AND SYSTEM OPTIMIZATION FOR REMOTE SENSING: A CANONICAL APPROACH†

David Middleton*

Physicist and Applied Mathematician
127 E 91 St., New York, N.Y. 10028

A general program for channel modeling, data acquisition and validation, and system optimization and evaluation is described. Attention for the most part is directed to the basic, active single-link system consisting of transmitter (signal source), channel (with or without secondary sources-e.g., scatterers), and receiver, which here is coupled to the medium of propagation by an M-element array of sensors (multi-dimensional sensing). For passive reception a similar receiver is used, to respond to ambient sources localized or distributed in the medium. A short summary of the main elements of statistical communication theory is first presented, followed by construction of channel models, including various typical propagation mechanisms for electromagnetic, acoustic, and mechanical radiation, for both scattered and ambient radiation fields. This leads to a quantitative description of received waveforms, with various statistical formulations for the random processes generated or arising in the system, particularly in the channel itself. The statistic of principal interest is the covariance of the process in question, which is illustrated by specific relations for scattered and ambient cases, including experimental results for scattering from a water-atmosphere interface. A short account of some of the statistical techniques required for the validation and processing of statistical data is then given. Some examples of system optimization – principally for reception – are next presented, and a short review of the main ideas behind these recent approaches to problems of data acquisition and remote sensing completes the discussion.

29.0 Introduction

In this chapter we shall address ourselves in a very general way to the basic problems associated with the title of this book and considered in detail in the preceding chapters. Our approach is canonical, in that we wish to describe, necessarily rather compactly for reasons of space, general approaches to channel modeling, data fields, certain questions in the handling of statistical data, and optimization of transmitting and receiving systems to be used for remote sensing.

The problems of remote sensing are often very different in character from problems of communication, although the forms of mathematical methodology may be similar. For example, an important problem in communication is the detection of a generally deterministic signal in the presence of random noise or other, deterministic signals. In remote sensing of geophysical parameters the problem is often that of detecting a random signal (statistical physical variable) in the presence of a random noise (a fluctuating transmission medium). In such cases conventional engineering practices, often so close to optimum, are not always sufficient for optimization of remote data gathering systems. This chapter endeavors to lay the ground work for analysis of the more subtle problems to be encountered in querying nature. It is clearly a forerunner of the future. The application of these methods to remote sensing is only at its beginning. The examples to be quoted later are largely from fields other than tropospheric remote sensing, often from problems of oceanic acoustic sounding, on which by far the greatest effort at channel modeling and optimization has so far been expended. However, these examples embody many of the same problems which must be faced in tropospheric studies, and may accordingly serve as pertinent illustrations of the kinds of concepts and methods which will be needed in the fully developed approaches of the future.

In channel modeling, scattering by inhomogeneities (both of comparatively large sizes and of molecular order) in the medium and at bounding interfaces is a major problem and source of much information about the channel, as Chapters 8, 10-15, 18, 21-25, and 26 have indicated. In addition, turbulence, when actively

†Preparation of a portion of this Chapter was supported by the Office of Naval Research under Contract N00014-70-C-0198.

* Professor, Electrical Engineering, Univ. of Rhode Island, Kingston, R.I. 02881

sensed, is very closely related to scattering phenomena (see Chapters 1-4, 6, 7, 13, 14, 20, and 25). In both instances, absorption mechanisms may also be significant. Data acquisition, of course, is the central endeavour, as discussed in most of the earlier chapters.

Very important for the interpretation of results are appropriate techniques of statistical data analysis (SDA). Here the validation of ensemble data, namely, the determination that the data may be said, with near unity probability, to belong to an ensemble, is a critical requirement, before one can proceed to obtain reliable statistical estimates. We shall touch on this briefly, also.

Finally, various classes of systems (e.g., radio, radar, optical, acoustical, etc.) are needed for our fundamental task of “interrogating” the medium in order to learn its properties and predict its behavior (see Chapters 12, 15, 18-24, 26, and in particular, Chapter 27). It is, therefore, both natural and necessary to consider in this regard such basic questions as system design and evaluation of performance, which have become standard in the more familiar civilian and military communication usage. This includes such topics as system optimization, signal design, and methods of coupling to the medium under study.

Because our subject involves elements of many physical disciplines and a correspondingly diverse arsenal of mathematical techniques and concepts, the reader should not be discouraged by his real or apparent lack of familiarity with portions of the material of this chapter. Those with electrical engineering backgrounds, for example, should encounter little difficulty with the communication theoretical features of our approach, but may have some trouble with the physical aspects of the channel modeling. On the other hand, those whose training and interests are primarily in particular areas, and directed toward the experimental investigation of those areas, may find the physical statistical communication concepts and techniques difficult at first, especially since we are not able to present any of these subjects in any detail here. Rather, the reader should feel quite satisfied if he has achieved some sense of overview and a feeling for the possibilities, as yet largely unexplored in remote sensing activities, of the joint approaches to data acquisition and analysis, and optimal and near-optimal system design for these purposes.

29.0.1 What The General Problem Is

Fundamentally, we are concerned here with the general problem of:

- (1) the *acquisition of data* by one or more distinct and distributed sensors, where these data represent the relatively long-range interrogation, i.e., remote sensing, of the medium; and
- (2) the *processing* of such data, in order to measure in some optimal, or near optimal fashion, one or more desired observables, which in turn will enable us to deduce the physical properties of the phenomena under study, whether it be the medium itself or some event taking place in the medium. Both *detection* (“is a desired signal observed in the presence of undesired or background noise?”), and *extraction* (“how much, or how large, etc., is the received signal, or the desired quantity measured by the received signal?”) are the generic operations whereby the information sought for is ultimately obtained. “Signal” here, of course, may be the “background” itself: very often we may wish to measure the properties of the medium alone. In other instances we may wish to study a more localized phenomenon (received as a “signal”), which is embedded in the background field. For the latter we can have both the detection and extraction modes of reception.

29.0.2 The General Approach

Thus, in the broad sense we have a problem in communications: we are asking questions of nature and seeking a reply. This may be done *actively* by transmitting various types of signals and receiving signals modified in some fashion by the medium or the phenomenon under observation. Or it may be carried out *passively*, where the phenomenon itself produces “signals”, which we seek to acquire in our receiving complex of sensors and data processors.

For this general purpose, the basic approach described here requires:

- (1) *A plan for data sampling*, i.e., where to locate the sensors, and what kind of temporal sampling to use, i.e., discrete (digitalized) or continuous sampling in time;
- (2) *Data processing procedures*, whereby optimal or near-optimal data processing is to be carried out, to extract maximum usable information from the acquired data. This includes the space-time sampling approaches of (1) above.
- (3) *Statistical Data Analysis (SDA)*, whereby one validates the ensemble of experimentally obtained data, so that the needed statistical measures, e.g., moments, distributions, etc., can be reliably estimated. This

is very important, in as much as there is almost always a significant random component in the phenomenon under investigation: either inherently, or because of unavoidable medium or system noise, or all three.

(4) *Physical models* of the channel and/or fields which are being sensed. These are vital to the successful processing of the data and maximal elimination of unwanted noise and signals. They are also critical elements of interest in their own right, and often the goal of the investigation itself. This is particularly true in the present applications to remote sensing of the troposphere.

The extent to which we can successfully combine the factors 1-4 determines both the quantity and quality of the results of our remote sensing effort. In subsequent sections we shall provide a concise, occasionally detailed development of these ideas.

29.0.3 Previous Work

We shall base our present treatment mainly on some recent work of Middleton (1970 and 1969). For instance, Middleton (1970) contains a detailed outline, with some specific examples, of much of (2) and (4) above, while Middleton (1960, 1965) provides an extensive technical background to (2). Application of various statistical techniques to the validation of physical data (3), in particular where reasonably good models are available, is described in Middleton (1969). For space-time sampling plans and their rationale for optimum data processing, we refer to the analyses of Petersen and Middleton (1963 and 1965) and to Chapter 22. Other, more specialized references are cited therein and later in the text. The reader is also encouraged to examine the preceding chapters, particularly for specific and detailed applications.

29.0.4 Organization of Chapter 29

We begin in (29.1) with a concise, partial introduction to Statistical Communication Theory, applied to the situations where specifically there may be multiple $M(\geq 1)$ sensors in the receiver complex. However, a single transmitted signal is assumed in both the active and passive cases: only one signal is used to interrogate the medium or (localized) phenomenon under study in the former, and but one signal is emitted from the latter. (Extension to multiple sources is readily made but will not be given here.) An operator representation of the communication processes involved is also included, and the important topics of signal detection, extraction (i.e., measurement), and their optimization are briefly described.

Section 29.2 is devoted to modeling the medium, and a variety of radiating sources, which may emit deterministic or random signals, which may in turn produce scattered and/or ambient fields, respectively. We illustrate this with a prototypical scalar and vector example.

Section 29.3 considers reception and the characteristic received waveforms, with attention to preformed and adapted beams, for receivers with multiple sensors immersed in the medium, or in the signal and noise (background) fields described in 29.2.

Section 29.4 illustrates the preceding discussion with a number of important physical examples involving acoustical, mechanical and electromagnetic phenomena, both man-made and natural. This includes fields and the received waveforms.

Section 29.5 then extends these essentially classical results to a general statistical formulation, in terms of the associated Langevin equations and their probabilistic solutions.

Section 29.6 offers several specific theoretical and experimental examples of the general statistical approach outlined in 29.5 .

Section 29.7 gives a short, illustrated account of some SDA methods for problems of the above type.

Section 29.8 gives some examples of optimized processing systems, which are illustrative of the general methods of 29.1 above. A very brief, mainly referential account of space-time sampling procedures is given in 29.9 . Section 29.10 concludes the chapter with a short summary of the principal implications of our canonical approach for the remote sensing problem.

Finally, because of the diversity of examples, the variety of techniques, and the possibly unfamiliar concepts and approaches touched upon here, the following reading plan may prove helpful: Start with Section 29.1 , the beginnings of 29.2 , 29.3 , 29.5 , all of 29.9 and 29.10 ; then, return to Sections 29.4 and 29.5 , and finally, to the remaining sections. Hardy souls may wish to plunge in and omit nothing at a first reading. In any case, the reader is encouraged to sample, as long as he does not lose sight of the fact that this is a canonical and interdisciplinary approach.

29.1 A Concise Survey of Some Elements of Statistical Communication Theory (SCT)

We begin with an operator description of the generalized link between source and receiver and the communication channel in which they are embedded. This is followed by a short account of signal detection, extraction, and system optimization, from the viewpoint of SCT, and involving a multidimensional formulation.

29.1.1 The Single-Link System

Whenever it is meaningful to consider the transfer, or “mapping” of “information” from one point in space-time to another, we encounter in the broadest sense a basic problem of statistical communication theory (SCT), independent of the particular physical mechanisms involved. If, in addition to information transfer or mapping, decision-making in the face of uncertainty is required, such as signal detection or extraction (measurement), the concepts and methods of SCT provide the appropriate apparatus for handling problems of this type. Our general purpose here is to present a concise account of developments in that area of SCT devoted to the detection and extraction (measurement) of signals which have joint spatio-temporal characteristics.

During the last decade and particularly in the last few years, the roles of the propagation process, associated system geometry, and the manner of coupling to the medium in question, have been generally recognized as critical elements in the design and evaluation of optimum and suboptimum systems. Thus, an effective theory of the communication process, here remote sensing activity including detection and extraction (measurement) of signals, now requires that these elements be explicitly incorporated. Quantitative attention to such a program is required if the fullest realization of system capabilities in a real environment is to be achieved. The relevant physics of the channel, principally the medium of propagation, but including an (often) localized phenomenon of interest, must be quantitatively included in the model of the communication process (See references 1-45, 53, 69, 71, and 72, in Middleton, 1970). Indeed, the channel itself is usually the subject of prime interest. Although we are concerned mainly in this book with atmospheric channels, our approach here is in no way limited to them, but applies canonically to other media as well, for example, the oceans, seismological domains, space, and so on. For this reason our examples will not be confined to the atmosphere alone, but will draw on a somewhat broader spectrum of geophysical interest.

The systems we consider here are for the most part “single-link,” active systems, which may be compactly represented in operator form by (Middleton, 1960, 1965)

$$\{v\} = \mathbf{T}_R^{(N)} \mathbf{T}_M^{(N)} \mathbf{T}_T \{u\} = \mathbf{T}_{R_0}^{(N)} (\mathbf{T}_{AR} \mathbf{T}_M^{(N)} \mathbf{T}_{AT}) \mathbf{T}_{T_0} \{u\} \quad (29:1a)$$

The $\{u\}$ are a set of messages to be transmitted while the $\{v\}$ constitute an ensemble of received messages, or decisions, which are consequent upon the set $\{u\}$. The \mathbf{T}_T , $\mathbf{T}_M^{(N)}$, and $\mathbf{T}_R^{(N)}$ are general operators that describe the processes of transmission, propagation through the medium in question, and reception, respectively. The superscript (N) indicates that noise is injected in some fashion at the designated stage of the system.

The operator notation, here employed as a compact means of representing more detailed mathematical calculations, means that the ensemble, or set of possible messages or signals, either mechanical, electromagnetic, or electronic quantities, are changed in form by the medium, or by antenna or transducers, or by mechanical or electronic operations. Thus, if we wish to signify the effect of a receiver on the messages received, we write the operator \mathbf{T}_R to the left of $\{u\}$. Successive operations are applied in orders to the left of preceding operations. Clearly, the operations in general are not commutative: their order of application is important.

As the second relation of (29:1) shows, $\mathbf{T}_R^{(N)}$ and \mathbf{T}_T can be decomposed into an ordered sequence of two other, more primary operations: $\mathbf{T}_T = \mathbf{T}_{AT} \mathbf{T}_{T_0}$ and $\mathbf{T}_R^{(N)} = \mathbf{T}_{R_0}^{(N)} \mathbf{T}_{AR}$. Here \mathbf{T}_{T_0} and $\mathbf{T}_{R_0}^{(N)}$ represent time processing only, which occurs in the “encoding” of the message set $\{u\}$ and in the “decoding” that results in the received output ensemble $\{v\}$. However, \mathbf{T}_{AT} and \mathbf{T}_{AR} , which represent the transmitting and receiving apertures, respectively, embody both temporal and spatial processing, which take place when these apertures are coupled to the medium. These, together with the medium operator $\mathbf{T}_M^{(N)}$, enable us to define a “coupled-medium” operator $\mathbf{T}_{RMT}^{(N)} \equiv \mathbf{T}_{AR} \mathbf{T}_M^{(N)} \mathbf{T}_{AT}$, which contains all the spatio-temporal operations

inherent in our single-link system. Thus, (29:1) can be equivalently written

$$\{v\} = T_{R_0}^{(N)} T_{RMT}^{(N)} T_{T_0} \{u\}, \tag{29:1b}$$

where we have separated the purely time-processing functions from the space-time operations, which are determined by the geometry and physics of the particular medium intervening between source and receiver.

For passive operation, where there is only a receiver awaiting signals, S_I , generated by a localized source in the medium, or by the medium itself, i.e., an ambient "signal" field, we have, instead of (29:1a, 29:1b)

$$\{v\} = T_R^{(N)} T_M^{(N)} \{S_I\} = T_{R_0}^{(N)} (T_{AR} T_M^{(N)}) \{S_I\}, \tag{29:2}$$

where now $T_{AR} T_M^{(N)} \equiv T_{RM}^{(N)}$ is the coupled-medium operator in this case. Figure (29.1) illustrates these operator models of the single-link systems described above.

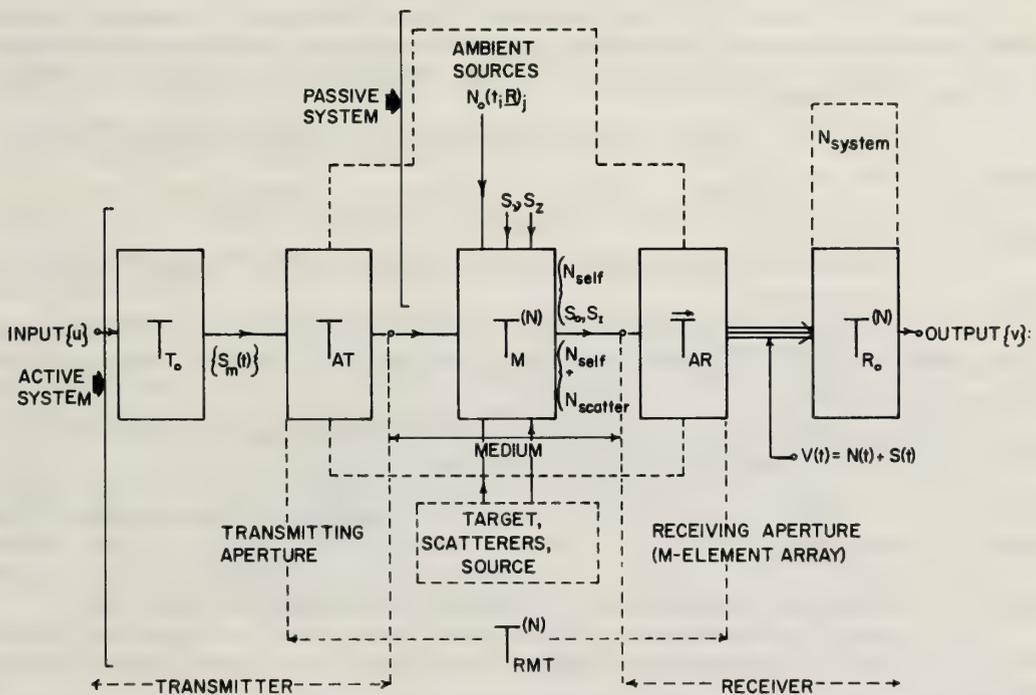


Figure 29.1 Multidimensional detection and extraction of signals in random media.

For example, consider an acoustic sounder whose mission is to determine the wave and turbulent structure of the medium at a given point. The $T_T \{u\}$ represent a set of sounding signals applied to the transmitting aperture, or acoustic radiator, which in turn "broadcasts" a typical signal in the form of a space-time differential pressure field into the atmosphere. This field is "directed" (by the aperture T_{AT}) toward the regions we wish to study. By using a pulsed signal we can resolve in time, and therefore in space, the various regions of medium-signal interaction $T_M^{(N)} T_{AT} T_{T_0} \{u\}$, (just as in the familiar radar cases), by studying the return as it is mapped (by the receiving aperture T_{AR}) into a purely time wave, entering the receiver's processor $T_{R_0}^{(N)}$. This is ultimately displayed (at the output of $T_{R_0}^{(N)}$) as, say, a voltage vs. time (or range). With a passive system, we have only the receiver, with its "beam" (T_{AR}) directed toward the spatial regions of interest, awaiting the locally generated signal from which we are finally to deduce the desired structural properties.

For signal detection and extraction we are concerned initially[†] with $\mathbf{T}_{R_0}^{(N)}$ (i.e., with all receiver elements following the aperture \mathbf{T}_{AR}). In order to obtain optimum or near-optimum receivers for these purposes, and to evaluate their performance (and, in fact, the performance of any specified receiver, $\mathbf{T}_{R_0}^{(N)}$), we must know the physical and statistical structure of their inputs. This, in turn, means that we must obtain corresponding information about the signal and noise fields in which the receiver as a whole is immersed. As noted in Middleton (1970), with “passive” operation, where the receiver has no control over signal sources, $\mathbf{T}_M^{(N)} \{S_I\}$ embodies the effects of all (external) signal (and noise) sources and is, of course, a space-time phenomenon. In the “active” mode of operation, (Middleton, 1970), the complete link represents the functional stages of the system. The criterion of optimality used here is the standard one of Bayes, or minimum average cost (Middleton, 1970). In what follows we shall outline the key steps needed for the successful solution of these multidimensional detection and extraction (measurement) problems.

29.1.2 Past and Current Studies: Our Program for SCT Here:

We remark that most earlier work has concentrated on one or more aspects of these spatio-temporal problems, with varying degrees of physical realism, including idealizations of the geometry, the channel, and the manner of coupling to it. Significant exceptions to this in various respects may be cited in a hierarchy of increasing scope and complexity:

(1) For work devoted primarily to propagation models (the “uncoupled” medium $\mathbf{T}_M^{(N)}$), we have Middleton, (1970), References 4-14.

(2) For attention centered mainly on coupling to the medium (array processing \mathbf{T}_{AT} and \mathbf{T}_{AR}) see Middleton, (1970), References 15-18.

(3) Papers emphasizing models of the medium and coupling to it, e.g., $\mathbf{T}_{AR}\mathbf{T}_M^{(N)}$, are listed in Middleton, (1970), References 19-21.

(4) Work concentrating on spatial processing (arrays) and the subsequent time processing (detection and extraction), e.g., $\mathbf{T}_{R_0}^{(N)}\mathbf{T}_{AR}$ and sometimes $\mathbf{T}_{R_0}^{(N)}\mathbf{T}_{AR}(\mathbf{T}_M^{(N)})$ are cited in Middleton, (1970), References 22-40, 71, and 72.

(5) Finally, combinations of all features of (1) – (4) in the general program symbolized by (29:1a), (29:1b), and (29:2), are contained in Middleton (1970), References 41-45.

New features of this treatment lie not in the specific results presented – because of limitations of space these can only be occasional, illustrative highlights – nor in the documentation, which is necessarily selective and finite. They reside, instead, in

(1) our attempt to give a unified picture of the approach here to the measurement of fields and particularly to detection and measurement of multidimensional signals in general noise fields, and

(2) to emphasize the interdisciplinary character of the methods used, and the fact that statistical communication theory is a *canonical theory*, which is independent of any particular physical domain of application.

Accordingly, to achieve an operational outline of (1) and (2), we have organized the subsections below as follows:

In 29.1.3 we extend our operator formalism to models of the channel and signal sources in terms of both active and passive situations, where in general the coupling to the medium employs multiple sensors distributed in some fashion in the medium. Section 29.1.4 summarizes various basic elements of the SCT approach.

29.1.3 Received Signal and Noise Processes and Multidimensional Formulation

For receivers with multisensor inputs we write

$$\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_m, \dots, \mathbf{S}_M] , \quad (29:3)$$

[†] The structures of \mathbf{T}_{AT} and \mathbf{T}_{AR} are fixed for the time being. We point out in [Middleton, 1971] the importance of also optimizing \mathbf{T}_{AT} and \mathbf{T}_{AR} in system design.

a collection (or vector) of signal vectors, representing the desired signal outputs from the $m = 1, \dots, M$ sensors into the processing portion of the receiver $\mathbf{T}_{R_0}^{(N)}$. Here

$$\mathbf{S}_m = [S_m(t_1), \dots, S_m(t_n)] = [S_m(t_k)] = [S_{mk}] \quad (29:4)$$

is the vector of the m^{th} received signal, $S_m(t)$, sampled at times t_1, \dots, t_n , ($k = 1, \dots, n$). The indexing on (29:4) accordingly denotes a discrete sampling in space-time: spatial sampling by the M sensors and temporal sampling (preceding $\mathbf{T}_{R_0}^{(N)}$) over some observation period ($t_0, t_0 + T$). Similarly, the actual data \mathbf{V} which may or may not contain the desired signals, is represented by the vector

$$\mathbf{V} = [V_1, \dots, V_m, \dots, V_M] , \quad (29:5)$$

with

$$V_m = [V_m(t_1), \dots, V_m(t_n)] = [V_{mk}] , \quad (29:6)$$

cf.(29:4). Thus, $V_{mk} [\equiv V_m(t_k)]$ is the k^{th} time-sample of the data output of the m^{th} sensor, into the receiver processor $\mathbf{T}_{R_0}^{(N)}$, cf. (F29.1). [We remark that sampling in time is taken at discrete instants, basically as a mathematical device which enables us to construct the needed statistical models from the usual statistical methods, wherein one deals with a finite, denumerable set of random variables, here, say, the V_{m1}, \dots, V_{mn} when considered over the ensemble of possible V_m 's. Of course, continuous sampling in time is common: one has $S(t), t_1 < t < t_2$, for all t in that interval, say. However, now we must deal with probability *functionals*, instead of the simpler probability functions (densities and distributions), and the analytical treatment becomes much more involved.]

Next, remember that \mathbf{S} and \mathbf{V} both represent inputs to the last stage $\mathbf{T}_{R_0}^{(N)}$ of the system in (F29.1). In order to determine the various statistical distributions governing \mathbf{S} and \mathbf{V} , required for application of SCT, and which contain implicitly the physics of the phenomenon in question, we must relate \mathbf{S} and \mathbf{V} to their original sources and determine their underlying causal and statistical mechanisms. Accordingly, we consider the general case of the M -channel receiver ($M \geq 1$), where the receiving aperture coupling to the medium is a set of M transducers, or more generally, M apertures, arbitrarily distributed in space, and each output from which is separately applied to the time processor $\mathbf{T}_{R_0}^{(N)}$. Thus, $\vec{\mathbf{T}}_{AR}$ can be expressed as an M -component operator, arbitrarily indexed according to the individual elements or apertures comprising the now M -element array.

In the *passive mode of operation* the noise inputs to $\mathbf{T}_{R_0}^{(N)}$ (F29.1) can be represented by the vector

$$\mathbf{N}(t) = \vec{\mathbf{T}}_{AR} \{N_{self}(t; \mathbf{R})\} + N(t)_{system} . \quad (29:7)$$

Here $N(t)_{system}$ refers to system noise produced in the M -aperture channels in \mathbf{T}_{AR} , which notationally we continue to regard as effectively noise-free[†]. In (29:7) $N_{self}(t; \mathbf{R})$ is an ambient, or self-generated space-time noise field, continuous in time and space and specified at every point $\mathbf{P}(\mathbf{R})$, $\mathbf{R} = \hat{i}_x x + \hat{i}_y y + \hat{i}_z z$ (outside the regions occupied by the transmitter and receiver), cf. (F29.2). Each channel ($m = 1, \dots, M$) is likewise assumed to have a system noise contribution $N_{sys}(t)_m$. If $N_o(t; \mathbf{R})$ is a local, ambient noise source at the spatial point $\mathbf{P}(\mathbf{R})$, we have also $N_{self}(t; \mathbf{R}) = \mathbf{T}_M^{(N)} \{N_o(t; \mathbf{R})\}$. Similarly, for a desired (external) signal S_D that enters \mathbf{T}_{R_0} , we can write

$$\mathbf{S}(t) = \vec{\mathbf{T}}_{AR} \{N_{self} \oplus S_D\}; = 0 \text{ if } S_D = 0 . \quad (29:8)$$

[†] Additional system noise may arise in the processor $\mathbf{T}_{R_0}^{(N)}$, but this is usually ignorable vis-à-vis N_{system} . Accordingly, we shall consider \mathbf{T}_{R_0} here and subsequently to be noise-free, and drop the superscript (N).

In general, $V(t) = S(t) \otimes N(t)$, where \otimes indicates “in combination with” and is not necessarily additive. However, in almost all cases of interest the medium is additive with respect to what it is meaningful to define as signal and noise (undesired signals), so that \otimes becomes (+) and from (29:8) $S(t) = T_{AR} \{S_D\} \neq 0$. We have also $V(t) = S(t) + N(t)$, or $N(t)$ alone, if the desired signal is absent. (In many cases here, the “desired signal is the “noise” field itself, and we are asked to explore its properties.)

The situation is more complex *in the active mode*. Here we may have signal-generated noise fields (scatter), multipath [separation of the injected signal, S_{in} , (which is the output of the modulator or “encoder” $T_{T_0} \{u\}$ – see F29.1) into a number of resolvable signal components], and scatter from one or more localized “targets” or reflecting, secondary sources (including possible multipath effects here, also) in addition to the ambient noise fields and system noise previously described. For the usually additive media encountered in practice we can, accordingly, decompose the active (i.e., signal-dependent) portion of $T_M^{(N)}$ into three components:

$$T_{M\text{-active}}^{(N)} \{ \} = T_{M\text{-scatter}}^{(N)} \{ \} + T_{M\text{-multipath}}^{(N)} \{ \} + T_{M\text{-targets}}^{(N)} \{ \}, \quad (S_{in} \neq 0). \quad (29:9)$$

When $T_{AT} T_{T_0} \{S_{in}\} = S_{in}(t; \mathbf{R})$ is the exciting signal field propagated in the medium by the driving signal $S_{in}(t)$ at the transmitter, the first component of $T_M^{(N)}$ acting on $S_{in}(t; \mathbf{R})$ may generate a random noise field – the scatter field, of which clutter in radar and reverberation in underwater sound are familiar examples. Acoustic scattering by inhomogeneities in the atmosphere, electromagnetic scattering by molecules, by turbulent layers, and other inhomogeneities in the troposphere, etc., provide the principal examples here. The second and third components may be regarded as producing the desired (deterministic) signal inputs to the receiver. Thus, for the M-channel receiver the (vector) noise and signal processes into the receiving processor T_{R_0} become, in the active mode,

$$N(t) = \vec{T}_{AR} \{N_{self}\} + N(t)_{sys} + \vec{T}_{RMT}^{(N)} T_{T_0} \{S_{in}\}_{scatter}, \quad (29:10a)$$

$$S(t) = \vec{T}_{RMT} T_{T_0} \{S_{in}\}_{target} + \vec{T}_{RMT} T_{T_0} \{S_{in}\}_{multipath}, \quad (29:10b)$$

where for practical purposes the additivity of the medium implies its linearity. In fact, the coupled-medium operators, $T_{RMT}^{(N)}$, are also linear†, since the apertures T_{AR} and T_{AT} are linear operators as well. This is a fortunate physical circumstance in most instances, since it vastly reduces the difficulty of obtaining solutions to specific problems. Again, we distinguish the situations of signal $V(t) = S(t) + N(t)$ and no signal, $N(t)$. More complicated signal and noise processes can be constructed along similar lines. Note that $T_{RMT}^{(N)}$ need not be completely deterministic, but can introduce various random effects (fading, path delays, etc.), which are now indicated formally by random parameters $\underline{\theta}$ in $S(t; \underline{\theta})$.

In fact, what is “signal” to one observer may be a noise background to another: In passive observation it is often the input field itself which is of interest, when there are no localized, embedded sources which might be considered “signals”. In the active mode, we probe the medium, and the ambient background is usually an interference or noise effect.

Finally, once the central problems of constructing the statistical – physical models of N_{self} , $N(t)_{system}$, $T_{RMT}^{(N)} \{S_{in}\}_{scatter}$, etc., have been resolved, we can proceed (formally, at least) to the derivations of the needed probability density functions (pdf’s). (This is done in most instances by calculating the characteristic functions of these pdf’s from the governing pdf’s of the input signal and noise processes, and then inverting them, cf. 29.5.)

† They may be time varying, with memory, see (29:34c).

29.1.4 SCT: Signal Detection and Extraction

At this stage of the formulation we must call attention to the *economic* factors which control the decision-making processes of measurement and detection. These exist because there is always a cost associated with decision-making in the face of uncertainty, above any cost of experimentation itself. This cost is that associated with making errors in the possible decisions: inexact measurements are scientifically more “expensive” than exact ones. (Of course, strictly speaking, no measurement is exact; some are just more precise than others: there is always an inherent “noise” that corrupts the observation process, whether due to a background interference, or inherent in the statistical nature of the phenomenon itself that is being studied.) The effect of all this is to put a cost upon the measurement or observation process. This, in turn permits us to raise the question of minimizing such costs, by seeking better ways of measuring the phenomenon, i.e., better ways of processing the descriptive data thereof, and so the notion of system optimality may naturally be introduced (Middleton, 1960; Part IV). We shall touch on this here briefly.

The scope and power of the canonical approach is perhaps best appreciated by proceeding from a general framework to more particular forms, and ultimately, to results of specific applications. Accordingly, let us begin by postulating for the moment that the various couplings (i.e., apertures T_{AT} and T_{AR}) to the medium are specified, as well as the possible noise mechanisms and input signals (at least statistically). For optimum and suboptimum receivers ($T_{R_0}^{(N)*}$ and $T_{R_0}^{(N)}$) and their performance as detectors and extractors of signals, we introduce the following general expression for the average cost (or risk) (see Middleton, 1960, Chapter 18; 1965, Chapter 1, for the detailed development leading to 29:11):

$$R(\sigma, \delta) = \int_{\Omega} \sigma(S) dS \int_{\Gamma} F_L(V|S) dV \int_{\Delta} \delta(\gamma|V) C(S, \gamma) d\gamma. \tag{29:11}$$

In the usual way, cf.(F29.1), we have
 $\sigma(S)$ = (a priori) probability density function of received signals S ;

$F_L(V|S)$ = (a priori) probability density function of the received data V , given S , following the receiver’s aperture T_{AR} , as input to the time-processor $T_{R_0}^{(N)}$, vide (F29.1), for discrete sampling† on the time interval $(t_0, t_0 + T)$. Here $L = Mn$, the order of F_L and σ , and is the number of components in S and V .

For the additive case the conditional density in (30:11) is $F_L(V|S) = W_L(V - S)_N$, where $W_L(\cdot)_N$ is the joint L th order p.d.f. of the L -component vector $N = [N_m(t_i)]$, ($i = 1, \dots, n$; $m = 1, \dots, M$), for the now time-sampled vector process N , as in (29:7).

$\delta(\gamma|V)$ = the decision rule for the set of possible decisions $\gamma = [\gamma_1, \dots, \gamma_D]$ consequent upon the original input message set $\{u\}$, and based on the received data V , above. In “passive” operation the input message set is now the set of all external signal (and) noise sources available to the receiver $T_{R_0}^{(N)}$. The decision rule is the functioning element of the receiver; in fact, it is the receiver $T_{R_0}^{(N)}$, including both the space-time processing of the input fields and the decision-making processing of the input fields and the decision-making consequent upon them.

$C(S, \gamma)$ = a preassigned cost function, which to each class of signals (S) and each decision γ_i ($i = 1, \dots, D$) assigns appropriate costs. (For the usual average risk and Bayes systems discussed specifically here, $C(S, \gamma)$ is independent of the decision rule δ .) See Middleton (1960), 18.1-18.4, and Eqs.(1.7) (1.8), Middleton (1965).

Here Ω , Γ , and Δ are, respectively, the abstract signal-, data-, and decision-spaces (Middleton, 1960, Fig. 18.5).

† With continuous sampling on $(t_0, t_0 + T)$, $n \rightarrow \infty$ suitably ($T < \infty$), so that functions of vectors (and matrices) are (formally) replaced by appropriate functionals. Thus, $\sigma(S) \rightarrow \sigma[S(t)]$, $F_L(V|S) \rightarrow F_T[V(t)|S(t)]$, etc., become the associated functionals.

For estimation or measurement of the signal – its waveform \mathbf{S} or parameters $\underline{\theta}$ – when it is known surely to be present in some noise background, the decision rule† now becomes a probability density and takes the form

$$\delta(\underline{\gamma}|\mathbf{V}) = \delta(\underline{\gamma} - \underline{\gamma}_\sigma(\mathbf{V})) \tag{29:16}$$

where the right-hand member is a delta function. Thus, the estimator $\underline{\gamma}_\sigma(\mathbf{S}$ or $\underline{\theta}|\mathbf{V}) = \mathbf{T}_{R_0}\{\mathbf{V}\}$ is time processing performed on the input data \mathbf{V} from the M-channel coupling aperture $\vec{\mathbf{T}}_{AR}$, and as such $\mathbf{T}_{R_0}\{\mathbf{V}\}$ here embodies the structure of the estimator $\underline{\gamma}_\sigma = [\gamma_{\sigma q}]$, [$q = 1, \dots, Q (= n)$ for estimation of waveform; $q = 1 \dots, K$ for estimation of K (time-independent) parameters]. With specified receiver $\mathbf{T}_R^{(N)}$, (29:16) in (29:11) gives directly the average cost

$$R(\sigma, \delta) = \int_{\Omega} \sigma(\mathbf{S}) \, d\mathbf{S} \int_{\Gamma} F_L(\mathbf{V}|\mathbf{S}) C(\mathbf{S}, \underline{\gamma}_\sigma) \, d\mathbf{V}, \tag{29:17}$$

with \mathbf{S} replaced by $\underline{\theta}$ in σ and $C, \mathbf{S} = \mathbf{S}(\underline{\theta})$ in F_L for parameter estimation. As in detection, the average cost will depend on our choice of cost function C . A common selection is the *quadratic cost function* (QCF), $C(\mathbf{S}, \underline{\gamma}_\sigma) = C_0(\mathbf{S} - \underline{\gamma}_\sigma) \cdot (\mathbf{S} - \underline{\gamma}_\sigma) = C_0|\mathbf{S} - \underline{\gamma}_\sigma|^2$ where C_0 is a (positive) scale factor. Other useful cost functions are discussed in Middleton (1960), sections 21.2, 3, and Middleton (1965), section 3.2.

29.1.5 Optimization

If the structure of the receiver is not specified a priori, we can find the optimum (\equiv “op”) receiver in standard fashion by minimizing the average risk $R(\sigma, \delta)$ with respect to the decision rule (\equiv receiver structure), viz.,

$$\min_{\delta \rightarrow \delta^*} R(\sigma, \delta) = R^*(\sigma, \delta^*), \tag{29:18a}$$

with

$$\text{op}\{\mathbf{v}\} = \underset{\mathbf{T}_{R_0} \rightarrow \mathbf{T}_{R_0}^*}{\text{op}} (\mathbf{T}_{R_0} \vec{\mathbf{T}}_{RMT} \mathbf{T}_{T_0}\{u\}), \tag{29:18b}$$

for optimum performance, where R^* is called the *Bayes cost (or risk)* and δ^* is now the embodiment of the optimum M-channel receiver $\mathbf{T}_{R_0}\{\mathbf{V}\}$.

In our binary example of signal detection above it is found that the optimum decision rules $\delta^*(\gamma_0|\mathbf{V})$ and $\delta^*(\gamma_1|\mathbf{V})$, and the optimum decision process, are given by (29:12) and (29:13) with $\mathbf{T}_{R_0}\{\mathbf{V}\}$ replaced by $\mathbf{T}_{R_0}^*\{\mathbf{V}\}$, the (logarithm of a) *generalized likelihood ratio* Λ_L . Thus, the decision rule depends specifically on a generalized likelihood ratio Λ_L

$$\Lambda_L = \frac{p \int_{\Omega} \sigma(\mathbf{S}) F_L(\mathbf{V}|\mathbf{S}) \, d\mathbf{S}}{q F_L(\mathbf{V}|\mathbf{O})} \tag{29:18c}$$

where, again, $p =$ probability that a signal is present, $q = (1 - p) =$ probability that only noise occurs. Like $\vec{\mathbf{T}}_{R_0}(\mathbf{V})$, Λ_L is now compared with the cost ratio $\kappa = (C_{01} - C_{00}) / (C_{10} - C_{11})$; (usually it is convenient to compare $\log \Lambda_L$ with $\log \kappa$) with the algorithm, cf. (29:13):

$$(i) \text{ Decide noise alone (i.e., set } \delta^*(\gamma_0|\mathbf{V}) = 1; \delta^*(\gamma_1|\mathbf{V}) = 0) \text{ if } \log \Lambda < \log \kappa \tag{29:18d}$$

$$(ii) \text{ Decide signal present (i.e., set } \delta^*(\gamma_0|\mathbf{V}) = 0; \delta^*(\gamma_1|\mathbf{V}) = 1) \text{ if } \log \Lambda \geq \log \kappa$$

Clearly, the decision rules $\delta^*(\underline{\gamma}|\mathbf{V}) = \delta^*(\gamma_0|\mathbf{V})$, $\delta^*(\gamma_1|\mathbf{V})$ are the embodiment of the time processing in the now optimum receiver. The corresponding Bayes risk R^* and error probabilities $\beta_1^{(0)*}$ and $\beta_0^{(1)*}$ are determined from (29:14) and (29:15) with $\delta^*(\gamma_0|\mathbf{V})$ and $\delta^*(\gamma_1|\mathbf{V})$ therein. (See Middleton, 1960, section 19.1-19.3 for details.)

† Like the decision rules in detection, it is also nonrandomized, (see Middleton, 1960, p. 961).

In signal extraction the results also depend on our choice of cost function. Two of the most useful are the QCF, already noted above, and the *simple cost function* (SCF), described by $C(\mathbf{S}, \gamma_{\sigma}) = C_0 \{ \mathbf{A} - \delta(\gamma_{\sigma} - \mathbf{S}) \}$. With the former, the optimum receiver or Bayes estimator is represented by the well-known conditional expectation of \mathbf{S} , given \mathbf{V} , while with the SCF, the optimum receiver is an *unconditional maximum likelihood estimator* (UMLE)†.

Still further optimizations are possible: over \mathbf{T}_{T_0} and over \mathbf{T}_{AT} , \mathbf{T}_{AR} , the coupling to the medium, subject to appropriate constraints. These additional optimizations are important, because they represent (when realized) the full exploitation of the system designer's options vis-a-vis the medium, signals in the medium, etc., which he can use in principle, at least, to enhance his observations and further improve his questioning of nature (Middleton, 1971) and 29.8.3 .

To complete the necessary link between the canonical formulations outlined above and the desired results to be obtained in specific applications, we need now to provide the physical models of the signal and noise processes in question. This is done in 29.2 following.

29.2 Approaches to the Channel Model: Sources and Fields

We turn now to the main problem of modeling the medium, the coupling to it, and various sources therein. Here our task is threefold:

- (1) To obtain the waveforms of signals and noise propagated in the medium and observed at the receiver \mathbf{T}_{R_0} , as in (29:1).
- (2) To derive the explicit structures of the coupling and medium operators \mathbf{T}_{AT} , \mathbf{T}_{AR} , and \mathbf{T}_M , as in (29:1).
- (3) To determine the statistics of the received signal and noise processes, as noted in (29.1.3).

Since we must deal with signals of finite bandwidth (and duration) and with sources and media which are jointly random and variable in both space and time, a combination of wave theory, ray theory, and statistical modeling is needed (Middleton, 1967 and 1972). The local character of the propagation is "microscopically" described by appropriate differential equations with stochastic forcing terms—the Langevin equations (Middleton, 1960, Chapter 10). The general geometries are "macroscopically" determined by the ray paths that delimit the flow of energy in the medium and specify the domains of illumination and scatter.

In contrast with the usual, or "classical" treatment here, (see, for example, Chapters 8 and 11; also, Middleton, 1967, Section 1; and 1972a, Sections 12, 20), we avoid the intractable problems presented by the complex local boundary conditions that occur at interfaces and inhomogeneities of the medium by introducing their phenomenological equivalent: spatially dependent linear and time-varying stochastic filters, which represent the boundary effects and the distributed random properties of the medium (Middleton, 1967 and 1972a). In this way a realistic yet manageable theory can be constructed, which includes in addition to the realistic geometries and general signal waveforms, the key feature of a generalized statistical description, usually absent from the earlier classical models and essential to the newer goals of system analysis and design.

It should be emphasized that this representation of the medium in terms of an appropriate (linear) stochastic filter still embodies, as it must, the desired observables: *the statistics* of such a filter, e.g., its mean, fluctuations, variance, spectrum, correlation functions, distributions, *are*, in fact, all that we can observe of what is basically a random space-time process, viz., the field whose properties we are investigating. We can, for example (and do, in the sections below), postulate that the inhomogeneities of the medium may act like an ensemble of point sources, distributed in space and with temporally varying characteristics, which may also depend on location as well as time.

These "point" sources may embody the scattering properties under illumination from some active source, or they may be used to describe inherent radiative properties, representing the generation of an ambient field. What is important here is the fact that this is a workable model, whose "scientific truth" extends down to the order of the wavelengths (λ) of the radiation involved: macroscopically, we observe "smooth" or continuum properties of the medium in time and space; microscopically, the "point" model may or may not be realistic at very small dimensions ($\sim \lambda$).

† For a general discussion and details, see Middleton, 1960, section 21.2, 3, and Middleton, 1965, section 3.2.

We remark that the classical linear filtering and prediction theories of Wiener and Kolmogoroff are also included in the present formalism (see Middleton, 1960, section 21, 2, 21.4.)

In any case, we can use this approach to obtain a statistically predictive description of localized and widely distributed phenomena, in space and time. This is observationally an entirely adequate physical model on a continuum basis. It is, also, on a discrete basis, as well, where we are actually dealing with point sources or “point” mechanisms (to the limits of our system’s resolution), e.g., molecules in EM scattering situations, occluded gas bubbles in the ocean (in volume or near-surface scatter of acoustic waves), for example. This stochastic approach has already achieved quantitative verification in recent work on reverberation in underwater sound (Plemons, Shooter, Middleton, 1971; Plemons, 1971). See also the discussion in Section 24 of Middleton (1972a).

29.2.1 Signal Fields (Middleton, 1967 and 1972a,b)

Our first task is to obtain the desired waveform of the propagation. We shall consider two cases: the generation of a prototypical scalar and *vector* field, as two examples† with broad application, when a given signal, S_{in} , is injected into the medium.

Thus, we start canonically with the idealized but common case of a homogeneous nondissipative isotropic medium in which a scalar field $p(t, \mathbf{R})$ of some sort can be propagated. The source of this field is a locally distributed signal generator that is assumed motionless in the medium††. The quantity $\mathbf{R} = \hat{i}_x x + \hat{i}_y y + \hat{i}_z z$ is a vector from the origin of our coördinate system, locating the point $P(\mathbf{R})$ in space where we wish to describe the field, cf. (F29.2). An appropriate *dynamical equation* for many situations is the well-known wave equation (in rectangular coördinates)

$$\nabla \cdot \nabla p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = -G_T(t, \xi), \text{ or } 0, \tag{29:19}$$

respectively, in the region V_T containing sources, or where there are no sources ($G_T = 0$). The quantity G_T is for the moment a nonrandom source function defined by

$$G_T(t, \xi) = \int_{-\infty}^{\infty} a_T(\xi, t - \tau) S_{in}(\tau, \xi) d\tau. \tag{29:20}$$

Here α_T is the *aperture weighting* associated with the radiating element. The $\nabla \cdot \nabla \equiv \nabla^2$ is the (vector) Laplacian; c is the velocity of propagation; $\xi = \hat{i}_x \xi_x + \hat{i}_y \xi_y + \hat{i}_z \xi_z$ is the vector from the origin 0 to a radiating element $d\xi$ in V_T , cf. (F29.2), and $\partial p^2 / \partial t^2 = d^2 p / dt^2 \equiv p$, since the source producing G_T is motionless vis-à-vis the medium; $S_{in}(t, \xi)$ is a driving signal applied at ξ in the transmitting aperture.

The aperture weighting α_T represents the (linear) filtering action of the radiating element $d\xi$, of the transducer (which is physically distributed over V_T), and which converts the driving signal S_{in} applied at $P(\xi)$ into a local field. Thus, G_T is a “forcing” or driving term, expressed as a differential field (vide the left member of (29:19)). We note that α_T is time-dependent (i.e., has memory), and correspondingly, that its Fourier transform

$$A_T(\xi, f) = F_t \{ a_T \} = \int_{-\infty}^{\infty} a_T(\tau, \xi) e^{-i\omega\tau} d\tau, \quad (\omega = 2\pi f), \tag{29:20a}$$

is frequency dependent††† – components with different frequencies driving such apertures produce different weightings. With each aperture weighting a *beam pattern* A_T is associated, defined as the spatial Fourier transform of A_T

$$A_T(\nu, f) \equiv F_{\xi}^{-1} \{ A_T \} = \int_{V_T} A_T(\xi, f) e^{2\pi i \nu \cdot \xi} d\xi, \tag{29:21}$$

where $\nu = \nu(i_T) = \hat{i}_x \nu_x + \hat{i}_y \nu_y + \hat{i}_z \nu_z = \mathbf{k} / 2\pi = \hat{i}_x \lambda_x^{-1} + \hat{i}_y \lambda_y^{-1} + \hat{i}_z \lambda_z^{-1}$ is the (vector) spatial frequency for the direction $i_T = \mathbf{R} / R$ of the point $P(\mathbf{R})$ with $R = |\mathbf{R}|$ the magnitude of \mathbf{R} ; \mathbf{k} is the corresponding (vector) wave number, and the λ_x , etc., are wavelengths (of the propagated field) in the medium. Physically, the square

† We must usually distinguish between vector and scalar cases, since the former may interact quite differently than the latter with discontinuities in the medium (interfaces, scatterers, etc.). For example, we cannot talk of polarization or anisotropic interactions in the scalar cases (29.6.1.1).

†† Extension to the important cases of moving sources is readily made (Middleton, 1970, Section 4.4), for example.

††† To determine the structure of α_T , or $A_T(\xi, f)$, is a nontrivial boundary-value problem. We shall assume throughout, however, that A_T (or α_T) has been obtained (often in practice by experimental means, Middleton, 1970, Reference 51).

magnitude $|A_T|^2$ of the beam pattern indicates the relative allocation of the energy of the emission in wave number and frequency, cf. sec. 29.3.1 .

The solution of (29:19) *outside* V_T , i.e., where $G_T = 0$, is known to be (Stratton, 1941)

$$p(t, \mathbf{R}) = \int_{V_T} \frac{d\xi}{4\pi r} \int_{-\infty}^{\infty} A_T(\xi, f) S_{in}(f, \xi) e^{i\omega(t-r/c)} df, \tag{29:22}$$

where $S_{in}(f, \xi)$ is $F_t \{S_{in}(t, \xi)\}$ and $r = |\mathbf{r}| = |\mathbf{R} - \xi|$. (We assume negligible coupling of boundaries, interfaces, scatterers, etc., with the source function G_T – an entirely practical situation in most cases.)

Next, in a very large number of applications we have a *far-field* or Fraunhofer condition (expressed by the fact that $2L_{max}^2/\lambda_{min} \ll R$, where L_{max} is the largest physical dimension of the radiating aperture and λ_{min} is the shortest wavelength of the driving signal S_{in}), so that we can now express (29:22) compactly in terms of the beam pattern as

$$p(t, \mathbf{R}) \doteq \frac{1}{4\pi R} \int_{-\infty}^{\infty} A'_T(\hat{\mathbf{r}}_T f/c | S_{in}) e^{i\omega(t-R/c)} df, \tag{29:23a}$$

with $\hat{\mathbf{r}}_T = \hat{\mathbf{r}}_T f/c$, or in terms of the aperture, by

$$p(t, \mathbf{R}) \doteq \frac{1}{4\pi R} \int_{V_T} d\xi \int_{-\infty}^{\infty} A_T(\xi, f) S_{in}(f, \xi) \cdot e^{i\omega(t-R/c + \xi \cdot \hat{\mathbf{r}}_T/c)} df, \tag{29:23b}$$

where $A'_T = F_{\xi} \{A_T \equiv A_T(\xi, f) S_{in}(f, \xi)\}$ is a *generalized beam pattern*, corresponding to the generalized aperture weighting A'_T . Specifically, $\hat{\mathbf{r}}_T = \hat{\mathbf{i}}_x \cos \phi \sin \theta + \hat{\mathbf{i}}_y \sin \phi \sin \theta + \hat{\mathbf{i}}_z \cos \theta$ vide. . . (F29.2).

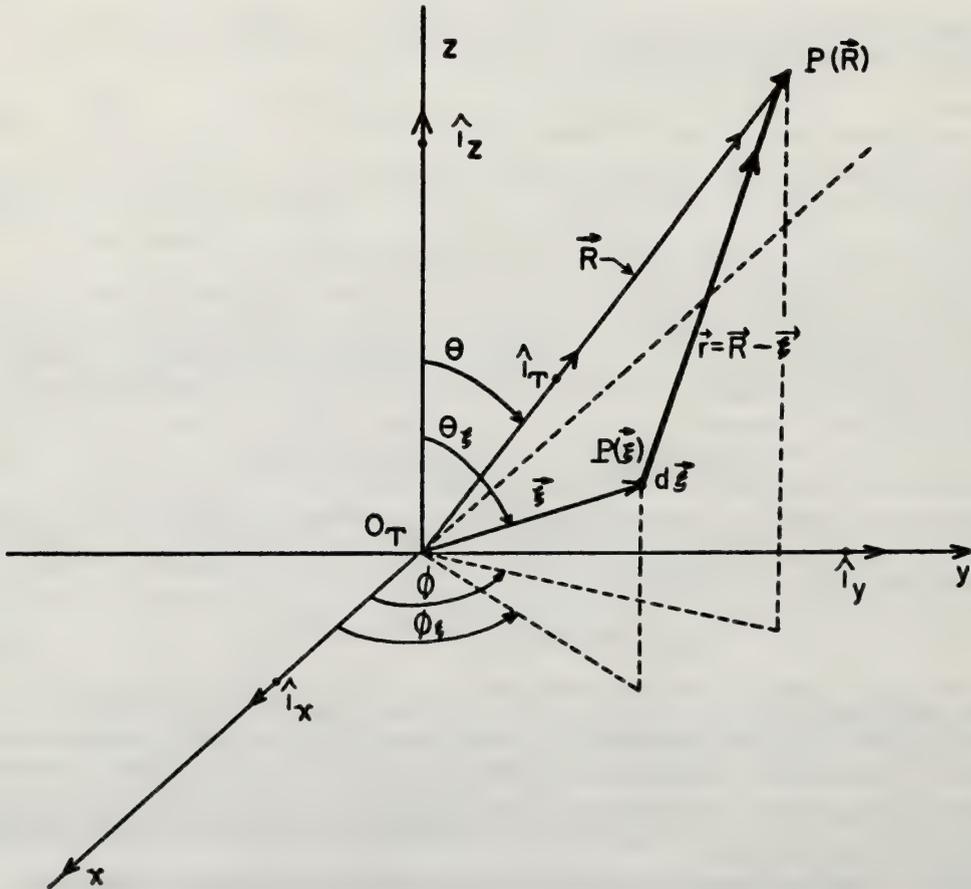


Figure 29.2 Geometry of a signal source.

As our second prototypical example, we have the case of propagation of a *vector* field, $\mathbf{L}(t, \mathbf{R})$, also in a homogeneous, isotropic, and non-dissipative medium, which obeys a vector analogue of the scalar wave equation (29:19), viz.,

$$\nabla \cdot \nabla \mathbf{L} - \frac{1}{c^2} \frac{\partial^2 \mathbf{L}}{\partial t^2} = -\mathbf{G}_T(t, \xi), \text{ or } 0. \quad (29:24)$$

Solutions, $\mathbf{L}(t, \mathbf{R})$, of (29:24) in the regions V_T outside the driving sources are the vector analogs (Stratton, 1941) of (29:22) and (29:23), with $A_T(\xi, f)$ replaced now by the *vector* weighting $A_T(\xi, f) \mathbf{L}$, where in the usual way for the far-field cases (see 29:23) $|\mathbf{R} - \xi|$ is replaced by R and $R/c - \hat{\mathbf{i}}_T \cdot \xi/c$, respectively, in the amplitude and phase. The various sources and aperture weightings are related by

$$\mathbf{G}_T(t, \xi) = \int_{-\infty}^{\infty} A_T(\xi, f) \mathbf{L} S_{in}(f, \xi) e^{i\omega t} df. \quad (29:25)$$

The vector weighting A_T has the corresponding *vector beam pattern* $A_T(\nu, f) \equiv F_{\xi}^{-1} \{A_T\}$, cf. (29:21). The far-field solutions of (29:24) are thus found to be

$$\mathbf{L}(t, \mathbf{R}) \doteq \frac{1}{4\pi R} \int_{V_T} d\xi \int_{-\infty}^{\infty} A_T(\xi, f) S_{in}(f, \xi) e^{i\omega(t - R/c + \xi \cdot \hat{\mathbf{i}}_T/c)} df, \quad (29:26a)$$

cf. (29:23b) above. In terms of a *vector* generalized beam pattern,

$$A'_T \equiv F_{\xi} \{A'_T \equiv A_T(\xi, f) S_{in}(f, \xi)\}, \quad (29:26b)$$

(29:26a) is alternatively

$$\mathbf{L}(t, \mathbf{R}) \doteq \frac{1}{4\pi R} \int_{-\infty}^{\infty} A'_T(\hat{\mathbf{i}}_T f/c, f) S_{in} e^{i\omega(t - R/c)} df, \quad (29:26c)$$

again with $\nu \rightarrow \nu_T = \hat{\mathbf{i}}_T f/c$ for the spatial frequencies.

29.2.2 Noise Fields: Ambient and Scattered Fields

As we remarked earlier (see 29:7 and 29:10a), there are two principal types of random field that appear in the channel: an ambient “self-noise” produced locally throughout the medium and a scatter “noise” arising from local inhomogeneities. The former is independent of any man-made signal source, while the latter is signal-dependent. Both these fields are of critical interest to us in the remote sensing of the troposphere, as they often *are* the “signals” which we wish to observe here. The former is generated in the medium and appears in passive reception, while the latter is produced when we “interrogate” the medium. The signal fields described in 29.2.1, are really special cases of the former, which it is convenient to designate separately when we wish to emphasize the communication aspects of the theory. Here now we have control of the type of signal at the transmitter and wish to direct our attention to it after it has passed through the medium in question.

Thus, a versatile model of self-noise for the cases discussed in 29.2.1 may be constructed by postulating a set $\{j\}$ of distinct radiating point sources, distributed in a volume or on a surface, or both. A disturbance emitted from the j th source at $P(\mathbf{R}_j)$ obeys a vector dynamical equation[†] similar to (29:19), (29:24)

$$\begin{aligned} \nabla^2 \mathbf{L}_{self-j} - \frac{1}{c^2} \frac{\partial^2 \mathbf{L}_{self-j}}{\partial t^2} &= -\mathbf{G}_{self}(t, \mathbf{r} | \mathbf{R}_j) \\ &= -\delta(\mathbf{r} - 0) \mathbf{n}_{self}(t, \mathbf{R}_j)_j, \end{aligned} \quad (29:27)$$

and \mathbf{L}_{self-j} can be \mathbf{L}_j , or $\mathbf{p}(t, \mathbf{R}_j)_j$. The vector field \mathbf{L} is described at all points $P(\mathbf{R}'_j)$ by an equation like (29:24), where the vectors $\mathbf{r}, \mathbf{R}'_j$ are drawn from the point source at \mathbf{R}_j as the new origin. Specifically, we can write

$$\mathbf{L}_{self}(t, \mathbf{R}'_j) = \mathbf{n}_{self}(t - R'_j/c) / 4\pi R'_j \quad (29:27a)$$

[†] Scalar fields and scalar dynamical equations are obtained at once, formally, by replacing vectors by scalars in what follows.

It is instructive to discuss the structure of n_{self-j} in more detail. We write

$$n_{self}(t, \mathbf{R}_j)_j = \int_{-\infty}^{\infty} H'_{self}(\tau, t | \mathbf{R}_j)_j L_{0j}(t - \tau, \mathbf{R}_j)_j d\tau \tag{29:28}$$

where H'_{self} (with dimensions length/time), is a (scalar) stochastic (realizable) linear time-varying filter, when considered over the ensemble of sources. L_{0j} represents the j th inherent source mechanism. The statistical properties of n_{self} depend, of course, on the detailed assumptions concerning H'_{self} and L_{0j} (Middleton, 1967 and 1972a), (Karp, et al., 1968). The total self-noise stochastic field available to the receiver's aperture T_{AR} is $N_{self}(t, \mathbf{R}') = \sum_j L_{self}(t, \mathbf{R}'_j)_j$. The set of dynamical equations (29:27) is called the *Langevin equation* of the process N_{self} (Middleton, 1960, Chapter 10; also, Section 29.5 following.)

The model for signal-generated noise or "scatter" (clutter, reverberation, etc.) is formally similar to that of self-noise fields. However, each scatterer now modifies the field incident upon it and reradiates it in variable amounts in all directions. In the case of scalar sources and media the scattered field L_{scat} is given by (29:29) with L_{self} replaced by L_{scat} and G_{self} by G_{scat} . In (29:28), n_{self} , H'_{self} , and L_0 , respectively, are replaced by n_{scat} , H'_{scat} , and L_{inc} , which is now the incident, signal-generated scalar field $p(t, \mathbf{R}_j)$, etc., as in (29:19).

In the case of vector fields there is a critical modification of our scalar relation for the scattered field. Equation (29:27) again formally applies ($L_{self} \rightarrow L_{scat}$, etc.), but now H'_{self} is replaced by a dyadic (or second-rank tensor) \hat{H}'_{scat} , so that n_{scat} is specifically

$$n_{scat}(t, \mathbf{R}_j)_j = \int_{-\infty}^{\infty} \hat{H}'_{scat}(\tau, t | \mathbf{R}_j)_j \cdot L(t - \tau | \mathbf{R}_j)_j d\tau . \tag{29:29}$$

The scattered field at $P(\mathbf{R}')$ is $N_{scat}(t, \mathbf{R}') = \sum_j L_{scat-j}(t, \mathbf{R}'_j)_j$. The dyadic \hat{H}'_{scat} is a nine-component tensor equal to $\hat{i}_x \hat{i}_x (\hat{H}')_{xx} + \hat{i}_x \hat{i}_y (\hat{H}')_{xy} + \dots + \hat{i}_z \hat{i}_z (\hat{H}')_{zz}$. It represents a mechanism whereby incident radiation falling upon the scatterer is reradiated in different amounts, in different directions, and with consequently different polarizations. Ionized layers and optical scattering by small particles in the atmosphere offer some examples of such tensor media (see F29.3).

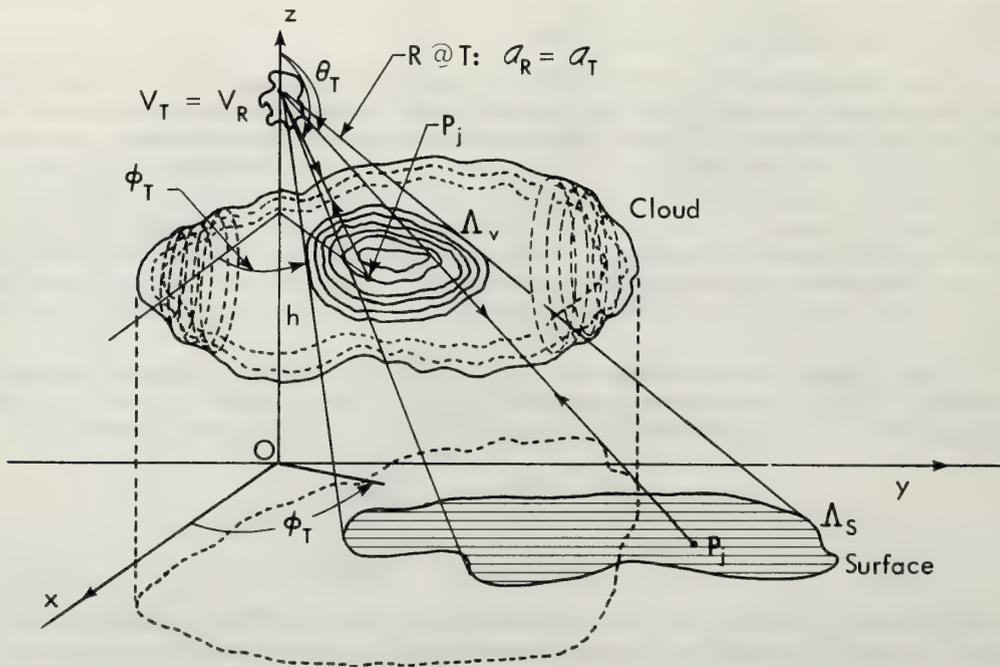


Figure 29.3 Monostatic system with volume and surface scatter.

29.3 Reception

To complete the link between transmitter and receiver we need now to couple to the medium by means of the receiving aperture \vec{T}_{AR} , (see 29:7). Thus, recalling that our receiver couples to the medium by means of M distinct sensors, distributed over some region, V_R , we note that the input to each component T_{AR-m} of the M -channel aperture \vec{T}_{AR} is either a vector or scalar process, depending on the underlying physics of the situation.

29.3.1 Preformed and Adapted Beams

The question of whether or not one has a *preformed beam* ($M = 1$), an *adapted beam* ($M > 1$), or uses *adaptive beam forming*, depends on the type of signal drive at the transmitter in the active cases, and in reception, on how the elements of the receiving aperture are combined. As examples of different types of drive, which produce different kinds of distributions in the medium, first consider the important special (approximate) case where the array consists of M point-elements, located at $\underline{\xi} = \underline{\xi}_m, m = 1, \dots, M$, at each of which different input signals $S_m(t, \underline{\xi})$ are applied. Then, $S_{in}(t; \underline{\xi}) = \sum_{m=1}^M S_m(t, \underline{\xi}) \delta(\underline{\xi} - \underline{\xi}_m)$, while if each element has the same driving signal, this reduces to $S_{in}(t; \underline{\xi}) = S(t) \sum_m \delta(\underline{\xi} - \underline{\xi}_m)$. Similar extensions are readily made for various combinations of point, line, surface, and volume elements.

We can easily include *the cases of adaptive beam forming*, (Middleton and Groginsky, 1965, and Middleton, 1966). Adaptive beam forming may be described (for point elements) if now we set $S_{in}(t; \underline{\xi}) = \sum_m S(t - \tau_m, \underline{\xi}) \delta(\underline{\xi} - \underline{\xi}_m)$ and let τ_m be selected delays, chosen in some appropriate fashion for both transmission and reception, but usually in the latter, *after* processing the received data, when for example, one wishes to optimize a possible signal detection or bearing estimate. (Middleton and Groginsky, 1965, and Middleton, 1966). A choice of the τ_m independent of the particular data at hand produces, of course, a *preformed beam*, in the usual way. Steering and scanning by preformed beams are readily included: steering by preselection of the τ_m , scanning by changing the τ_m (slowly) in time, so that $\tau_m = \tau_m(t)$. Similar remarks apply directly for distributed array elements.

Now, if the receiver uses a preformed beam, then, in effect, $M = 1$: all geometric elements of the array, or aperture, are regarded as forming a single aperture, and there is only a single input to the processor T_{R_0} , cf. (29:1), (29:1a).

For the M -channel or multidimensional receiver ($M > 1$) considered here, however, \vec{T}_{AR} represents an *adapted beam*. The individual outputs of each component aperture $\{T_{AR-m}\}$ are combined in some preset fashion in the subsequent (M -component) processor \vec{T}_{R_0} to form a beam, in effect, as a result of the processing embodied in $\vec{T}_{R_0}\{V\}$ itself. Receiver optimization ($\vec{T}_{R_0} \rightarrow \vec{T}_{R_0-opt}$) automatically determines the manner in which the M -channel input data are to be combined (Middleton and Groginsky, 1965, and references in Middleton, 1970). These "post-processing" beams are "adapted" to the input data, and the process of forming such a beam is called *adaptive beam forming*. Clearly, such a dynamic procedure offers further possibilities for extending receiver performance, since the constraint of a preformed beam pattern, fixed for all inputs, is now removed.

In fact, the extension to time-apertures (filters), apertures constructed of discrete spatial arrays, is a very important aspect of the *geometrical* descriptions of sampling plans. This includes, mostly for future study, the design of optimum apertures, both with regard to sensor design, and deployment, and data processing, e.g., "synthetic apertures." See 29.8.3, 29.9. (For some exploration of these ideas, see Petersen, 1963, and Petersen and Middleton, 1973.)

29.3.2 The Received Waveform, U_j

The elements of the noise and signal input to the processor \vec{T}_{R_0} (see Stratton, 1941, and Middleton, 1966a) may now be specified by considering the various vector (or scalar) radiation fields impinging upon an element $(d\underline{\eta}_0)_m$ at $(\underline{\eta}_0, \underline{\eta}_0 + d\underline{\eta}_0)_m$ in each receiving aperture in the M -channel receiver. As in (29:21) (29:27), etc., we can define a (vector) receiving beam pattern[†] by $A_R(\underline{\nu}_{Rm}, f)_m = F_{\eta_0}^{-1} \{A_T(\underline{\eta}_0, f)_m\}$, where, like A_T

[†] Here $\underline{\nu}_{Rm}$ is the (vector) spectral frequency $\underline{\nu}_{Rm}(\hat{r}_R)$, where \hat{r}_R is a unit vector normal to the wavefront impinging upon the receiving aperture; specifically, $\hat{r}_R = \mathbf{R}_R |\mathbf{R}_R|^{-1}$.

and g_T , we have $A_{R-m} = F_t \{g_R(\eta_0, t)\}$ relating the time and frequency forms of receiver aperture weightings in the m^{th} channel ($m = 1, \dots, M$). For a typical j^{th} self-noise source, or for a scatterer, we find that the inputs to the processor are now†

$$U(t)_j = [U_m(t)_j] = \left[\int_{V_{Rm}} d\eta_{0m} \int_{-\infty}^{\infty} g_R(\eta_{0m} + \mathbf{r}_m; t - \tau)_m \cdot L(\tau, \mathbf{R}_{Rj} - \eta_{0m} - \mathbf{r}_m)_j d\tau \right], \quad (29:30)$$

where $L = L_{\text{self}}$ or L_{scat} ; \mathbf{R}_{Rj} is the vector drawn from the point source, or scatterer, at $P(\mathbf{R}_j)$ to the origin O_R of the coordinate system of the entire receiving aperture; \mathbf{r}_m is the vector from O_R to the center of the m^{th} component aperture and V_{Rm} is its volume, vide (F29.4).

Summing over j for the contributions of the individual noise sources that radiate into the M beams A_{Rm} , ($m = 1, \dots, M$), gives the total noise processes $X_{\text{self}}(t)$ and $X_{\text{scat}}(t)$ that constitute the input to the processor \vec{T}_{R_0} . Thus, in the passive case (Stratton, 1941), the total noise input is $N(t) = X_{\text{self}}(t) + N(t)_{\text{system}}$, while in the active case (29:10a) it becomes $N(t) = X_{\text{self}}(t) + N(t)_{\text{system}} + X_{\text{scat}}(t)$.

Particular signals are handled in precisely the same way. Thus, for (point-source) signals we replace L in (29:30) by $S_D(\tau, \mathbf{R}_D - \eta)$ and for localized inhomogeneities (or “targets,” in the more familiar radar and sonar applications) by $S_{\text{target}}(\tau, \mathbf{R}_R - \eta | S_{\text{in}}, \mathbf{R})$, respectively, to obtain the possible signal inputs to the processor. Extensions to distributed ambient or scatter signal sources raises no conceptual difficulties. We must replace the present “point aperture” by a distributed one, as we have already done for the transmitting and receiving apertures of our communications link (Middleton, 1970).

29.3.3 The Medium and Coupling Operators

The noise and signal inputs above may be compactly described in terms of the channel and system operators:

$$X_{\text{self}} = \sum_j U_{\text{self}-j} = \sum_j \vec{T}_{ARj} \{L_{\text{self}-j}\} \equiv \vec{T}_{AR} \{N_{\text{self}}\}, \quad (29:31)$$

symbolically:

$$X_{\text{scat}} = \sum_j U_{\text{scat}-j} = \sum_j \vec{T}_{ARj} \cdot \hat{T}_{Mj}^{(N)} \cdot T_{ATj} T_{T_0} \{S_{\text{in}}\} = \sum_j \vec{T}_{RMTj} T_{T_0} \{S_{\text{in}}\}. \quad (29:32)$$

Similarly, for localized signal sources we have

$$S_{(\text{passive})} = T_{AR} \{S_D\}, \quad (29:33a)$$

$$S_{\text{“target”}} = (\vec{T}_{AR} \cdot \hat{T}_{M-\text{tar}} \cdot T_{AT}) T_{T_0} \{S_{\text{in}}\} = \vec{T}_{RMT-\text{tar}} T_{T_0} \{S_{\text{in}}\}, \quad (29:33b)$$

where again “target” refers to scattered signals from some local inhomogeneity in the medium, and

$$\vec{T}_{RMT-j}^{(N)} = (\vec{T}_{AR} \cdot \hat{T}_M^{(N)} \cdot T_{AT})_j, \text{ etc.}, \quad (29:33c)$$

cf. 29.1.3 above.

Single path and multipath signals are handled in the same way. Note that for vector fields \hat{T}_M is a dyadic or 2nd rank tensor, (cf. Page, 1935) and that T_{AT} is a 3- vector. Observing that these operations

† For scalar field inputs we have formally $a_{R \rightarrow g_R}, L \rightarrow L$, etc.

received waveform structure, U_j , in these cases, mainly for the narrow-band signals that customarily occur in practice. These examples are of general geophysical interest, and for the moment are not specialized to acoustic and electromagnetic propagation in the atmosphere, although this is the area of principal concern in this book. We broaden the domain of application here in order to emphasize the canonical nature of our statistical-physical approach in conjunction with the methods of statistical communication theory.

29.4.1 Some Dynamical Equations and Fields

As our first example, we consider the propagation of incremental[†] (scalar) pressure waves in the ocean (and atmosphere) for general, typical, physical conditions. When the medium can be approximated by a homogeneous, nondissipative model our prototypical results (29:19)-(29:23b) may be directly applied: the dynamical equation is (29:19), and the field at some point $P(\mathbf{R})$, external to the source, is $L(t, \mathbf{R}) = p(t, \mathbf{R})$, (29:22), (29:23).

However, when appreciable absorption and a nonvanishing velocity gradient ($\nabla c \neq 0$) occur, a frequent situation in the ocean, this simple model must be suitably modified. From Middleton, 1967 and 1972a, Part III, we have the extension of (29:19) and (29:23). The dynamical equation is now [Middleton, 1972a, (Eq. 15.7 in Part III)] for *small* gradients:

$$\left(1 + \frac{\alpha}{c^2(z)} \frac{\partial}{\partial t}\right) \nabla^2 p - \frac{1}{c^2(z)} \frac{\partial^2 p}{\partial t^2} = -G_T(\xi, t), \text{ or } 0, \quad (29:35)$$

where α is an absorption parameter (≥ 0). The (unscattered) incremental (far-field) at $P(\mathbf{R})$ is found to be [for frequencies below 0(40 kHz)],

$$p(t, \mathbf{R}) \nabla_{c \neq 0} \doteq \frac{1}{4\pi\ell_0(\mathbf{R})} \int_{-\infty}^{\infty} \beta(\ell_0|\mathbf{R}) A'_T(-\mathbf{a}_1 f, f|S_{in}) Y_{path}(f|\mathbf{R}) e^{2\pi i[t - \tau_{\ell_0}(\mathbf{R})]} df, \quad (29:36)$$

in which ℓ_0 and τ_{ℓ_0} are respectively the path length and path delay to $P(\mathbf{R})$, and \mathbf{a}_1 ($\neq \hat{\mathbf{r}}_T f/c$) is a new vector spatial frequency, which has the direction at O_T of the ray initiated at O_T and traveling along the ray path to $P(\mathbf{R})$. The quantity $\beta(\ell_0|\mathbf{R})$ is a ray focus-defocusing factor (unity for $c = 0$) that accounts for the concentration or dilution of the energy in the beam, in the course of its propagation. The quantity Y_{path} represents the effects of absorption along the path in the medium, and is given by (Eq. 15.9 in Middleton, 1972a, III),

$$Y_{path}(f|\mathbf{R}) \doteq e^{-\alpha\omega^2\tau_{\ell_0}^3/2\ell_0^2 + ik\pi/2}, \quad \omega = 2\pi f, \quad (29:37)$$

where k ($=0, 1, 2, \dots$) indicates the number of turning points (i.e., points at which the ray reverses its direction) that the ray may experience in the medium in the course of propagation to $P(\mathbf{R})$.

Figure (29.5) shows four important modes of propagation in a deep ocean, when ($\nabla c \neq 0$), for various locations of transmitting and receiving apertures (cf., Section 17 of Middleton, 1972a). These are *convergence zone* (CVZ), *SOFAR* channel, *bottom bounce* (BB), and *surface ducting* (SD), with their attendant geometries. Idealized linear profiles [$c(z) = c_{00} - bz$] are assumed.

Our second physical example involves a (scalar) model of man-made noise or interference in the urban radio environment. From the viewpoint of the receiver this is an ambient noise mechanism obeying a dynamical equation of the type (29:19), for which the (far-)field at $P(\mathbf{R}_j)$, from a typical (j^{th}) source, providing say, ignition noise from an automobile engine, is (Middleton, 1972b)

$$L_j(t, \mathbf{R}_j) \doteq \frac{1}{4\pi R_j} \int_{-\infty}^{\infty} A'_I(\hat{\mathbf{r}}_{R_j} f/c, f|S_I) e^{i\omega(t - R_j/c)} df, \quad (29:38)$$

where $S_I = F_t\{S_I(t; \mathbf{R}_j)\}$ is the amplitude spectrum of the interfering source S_{Ij} , and A'_I is the generalized beam-pattern, cf. (29:23b).

Dynamical equations of the form (29:19) also occur when (small amplitude) propagation in an elastic medium is considered. Here, however, one now generally has a vector field $\mathbf{u}(t, \mathbf{R})$ for the displacement of the

[†] By incremental pressure here is meant the (fluctuation) in pressure about the static, or mean pressure in the medium.

medium from equilibrium, instead of the simpler, scalar fields $p(t, \mathbf{R})$. The dynamical equation is (Page, 1935)

$$c_p^2 \nabla \cdot \nabla \mathbf{u} - c_s^2 \nabla \times \nabla \times \mathbf{u} - \frac{\partial^2 \mathbf{u}}{\partial t^2} = -\mathbf{F}_T(t, \xi), \text{ or } = 0, \quad (29:39)$$

respectively, in or outside the region V_T containing the driving source \mathbf{F}_T (force per unit volume). Equation (30:39) can be resolved into two wave equations representing purely compressional (“longitudinal”) or P-waves and purely shear (“transverse”) or S-waves if we set $\mathbf{u} = \mathbf{u}_p + \mathbf{u}_s$, where the vector fields \mathbf{u}_p and \mathbf{u}_s , respectively, obey the auxilliary conditions $\nabla \times \mathbf{u}_p = 0$ (\mathbf{u}_p is irrotational) and $\nabla \cdot \mathbf{u}_s = 0$ (\mathbf{u}_s is solenoidal). Then \mathbf{u}_p and \mathbf{u}_s obey

$$\begin{aligned} \nabla^2 \mathbf{u}_p - \frac{1}{c_p^2} \frac{\partial^2 \mathbf{u}_p}{\partial t^2} &= -\mathbf{F}_T(t, \xi)/c_p^2 \\ \nabla^2 \mathbf{u}_s - \frac{1}{c_s^2} \frac{\partial^2 \mathbf{u}_s}{\partial t^2} &= -\mathbf{F}_T(t, \xi)/c_s^2, \end{aligned} \quad (29:40)$$

(with $\mathbf{F}_T = 0$ outside V_T), where c_p and c_s ($c_p > c_s$) are the (phase velocities of the P- and S-waves in the medium. The joint occurrence of both P- and S-modes are generally typical of seismic phenomena and elastic media. In gases and liquids, which do not support a transverse stress, only P-waves occur, and one has a scalar field† for the resulting *incremental acoustic pressure* $p(t; \mathbf{R})$, obeying (29:19) (when the fluid is incompressible), as noted earlier.

While acoustic phenomena are of central interest for propagation in the earth, the atmosphere, and underwater (seismology, sound, sonar, etc.) electromagnetic phenomena (radio, radar, optics, etc.) play an even broader role from the viewpoint of communication theory and almost as broad a one in the remote sensing situations of primary concern to us here. The various fields here also obey vector wave equations of the form (29:24), compactly represented by the relations (see Chapter 8)

$$\nabla^2 \underline{\Pi}^{(*)} - \mu_0 \epsilon_0 \frac{\partial^2 \underline{\Pi}^{(*)}}{\partial t^2} = \left(-\frac{1}{\epsilon_0} \mathbf{P}, \text{ or } -\mathbf{M} \right) = -\mathbf{G}_T^{(*)}(t, \xi), \mu_0 \epsilon_0 = c^{-2}, \quad (29:41)$$

where $\underline{\Pi}$, $\underline{\Pi}^{(*)}$ are electric and magnetic hertzian potentials (in volt-meters and ampères per cubed meter). Here \mathbf{P} (coulombs per square meter) is the electric dipole moment per unit volume of the free charge distribution produced in the aperture by the driving voltage S_{in} and \mathbf{M} (ampère-turns per meter) is the magnetic moment per unit volume generated by an appropriate exciting current S_{in} ; c is the phase velocity of propagation. The electric (\mathbf{E}) and magnetic (\mathbf{H}) fields are obtained from

$$\mathbf{E} = \nabla \nabla \cdot \underline{\Pi} - \frac{1}{c^2} \frac{\partial^2 \underline{\Pi}}{\partial t^2} \text{ and } \mathbf{H} = \epsilon_0 \nabla \times \frac{\partial \underline{\Pi}}{\partial t}. \quad (29:42)$$

and the \mathbf{E}^* and \mathbf{H}^* associated with \mathbf{M} are given by (29:42) when $\mathbf{E} \rightarrow \mathbf{H}^*$, $\mathbf{H} \rightarrow \mathbf{E}^*$, $\underline{\Pi} \rightarrow \underline{\Pi}^*$, and $\epsilon_0 \rightarrow \mu_0$ explicitly therein. Solutions of these dynamical equations (29:40)-(29:42) have the forms (29:26 a, c) above, suitably specialized to the physics of the problem.

Other propagation equations of interest (particularly of meteorological interest) are the Rossby, barotropic, baroclinic equations, which are significant in various contexts. Here, in fact, solutions to the common wave equation are deliberately “filtered out” of many current models of atmospheric dynamics. Lack of space here prevents further discussion of this important topic (Petersen and Middleton, 1973 est.).

29.4.2 Some Received Waveforms, \mathbf{U}_j

We combine the results of 29.3.2 now with the field solutions of 29.4.1 above to obtain explicitly typical waveforms \mathbf{U}_j , or U_j , at the input to the receiver’s processor $\mathbf{T}_{R_0}^{(N)}$, for both bistatic ($R \neq T$) and monostatic operation.

In the active bistatic scattering case, where arbitrary input (scalar) signals are used and there is no

† Derived from \mathbf{u}_p through the relations $\mathbf{u}_p = \nabla \phi = \mathbf{v}$; $p(t, \mathbf{R}) = -p\phi$ and $(\nabla \rho = 0)$, where ρ is the medium density.

velocity gradient ($\nabla c = 0$ and negligible absorption), it can be shown that (Middleton, 1967, Eq. 4.36) the desired waveform is

$$\mathbf{U}_j = [\mathbf{U}_{mj}] = \left[\left[(4\pi)^2 \mathbf{R}_T \mathbf{R}_R \right]^{-1} \int_{-\infty}^{\infty} [A_R(\hat{\mathbf{i}}_R f/c, f)]_m e^{i\omega t} df \cdot \int_{-\infty}^{\infty} Y_M(f', f - f' | \mathbf{R}_T) A'_T(\hat{\mathbf{i}}_T f'/c, f' | S_{in}) df' \right]_j \quad (29:43)$$

Here $Y_M(f, \nu | \mathbf{R}_j) \equiv \mathbf{F}_T \mathbf{F}_\tau H_M(\tau, t | \mathbf{R}_j)$ is the *bifrequency function* associated with the (time-varying) response functions, H_{Mj} , of the j^{th} scatterer; [see the remarks preceding and following (29:28)]. For a point scatterer of cross-section $\gamma_0(\mathbf{R}_j)$, in relative motion vis-à-vis one or more of the following: either or both platforms (T, R), or the medium, we get for Y_M in (29:43) specifically

$$Y_M(f', f - f' | \mathbf{R}_T)_j = \left\{ \mu \gamma_0 e^{-i\omega' \mu T_0} \delta(f - \mu f') \right\}_j, \quad (29:44)$$

where $\mu \equiv 1 +$ a sum of the pertinent Doppler velocities; T_0 is the roundtrip delay between T and R.

For the situation discussed in (29.4.1), when $\nabla c \neq 0$ and there is random scattering off, say, the ocean bottom, the result (29:43) is now modified for monostatic operation (R@T) (Middleton, 1972a, Eq. 17.17b):

$$[\mathbf{U}(t_m)]_{\nabla c \neq 0} = \left\{ (4\pi \ell_0)^{-2} \int_{-\infty}^{\infty} [A_R(a_1 f_2, f_2)]_m \beta(\ell_{0s} | \mathbf{R}) \beta(\ell_{s0} | \mathbf{R}) Y_{\text{path}}(f_2 | \ell_0) e^{i\omega_2 t} df_2 \cdot \int_{-\infty}^{\infty} Y_M(f_1, f_2 - f_1 | \ell_0) \cdot A'_T(a_1 f_1, f_1 | S_{in}) Y_{\text{path}}(f_1 | \ell_0) df_1 \right\}_j, \quad (29:45)$$

with Y_{path} given by (29:37), when absorption may be important.

When *narrow-band input signals* are used, (29:43) reduces to the much simpler, but important result, for $\nabla c = 0$:

$$\mathbf{U}_{j - \text{n.b.}} = \left[\left[\text{Re} \left\{ \frac{\mu \gamma_0 e^{i\omega_0 \mu (t - T_0) + \hat{\mathbf{i}}_R \cdot \mathbf{r}_m / c}}{(4\pi)^2 \mathbf{R}_T \mathbf{R}_R} \hat{S}_0(\mu [t - T_0]) \cdot A_R(\mathcal{L}_R - \mathcal{L}_0 R, f_0)_m A_T(\mathcal{L}_T - \mathcal{L}_0 T, f_{0m}) \right\} \right]_j \right], \quad (29:46)$$

under the more restricted but still useful conditions (29:44), and uniform signal drive on A_T , cf. Section 29.3.1. Here $f_0 (= \omega_0 / 2\pi)$ is the carrier frequency of S_{in} , and \hat{S}_0 is the complex signal envelope. When $\nabla c \neq 0$ (and there is absorption) we find that (29:46) becomes

$$\mathbf{U}(t)_j |_{\nabla c \neq 0} = \left[\left[\text{Re} \left\{ \frac{\mu Q_{RT} e^{i\omega_0 \mu (t - T_0) + \hat{\mathbf{i}}_R \cdot \mathbf{r}_m / c}}{(4\pi \ell_0)^2} Y_{\text{path}}(\mu \omega_0 | \ell_0) \beta(\ell_{0s} | \mathbf{R}) \beta(\ell_{s0} | \mathbf{R}) \cdot Y_{\text{path}}(\omega_0 | \ell_0) \hat{S}_0(\mu [t - T_0]) \right\} \right]_j \right], \quad (29:47)$$

where now we write

$$\cdot \left(Q_{RTm}^{(M)} \right)_j \equiv \left\{ A_R(a_1 f_0, f_0)_m \gamma_0 A_T(-a_1 f_0, f_0)_m \right\}_j \quad (29:48)$$

as the *coupled-system cross-section* associated with the j^{th} scatterer; it has the dimensions of an area.

As our final example, we consider the narrow-band vector cases, and we employ the usual point scatterer model. We then have the vector analogue of (29:46), viz:

$$\mathbf{U}_{\text{scat}}(t)_j = [\mathbf{U}_{\text{scat}}(t)_{mj}] \tag{29:49a}$$

$$= \left[\text{Re} \left\{ \frac{\mu e^{i\omega_0 \mu(t - T_0 + \hat{\mathbf{r}}_R \cdot \mathbf{r}_m/c)}}{(4\pi)^2 R_R R_T} \hat{S}_0(\mu[t - T_{0j}]) Q_{RT}^{(m)} \right\} \right]_j,$$

where

$$Q_{RT-j}^{(m)}(\mathbf{R}, \mathbf{R}_R | f_0)_j \equiv A_R(\hat{\mathbf{r}}_R f_0/c - \mathcal{L}_{0R}, f_0)_m \cdot \hat{\gamma}_0 \cdot A_T(\hat{\mathbf{r}}_T f_0/c - \mathcal{L}_{0T}, f_0)_j \tag{29:49b}$$

is a generalized version of the *coupled-system cross-section* (CS), (29:48), of the *j*th scatterer. (For scalar fields (29:49b) reduces directly to the simpler form $Q^{(m)} = \gamma_0 A_R^{(m)} A_T$.) The effects of platform and scatterer motion are once more included through $\mu_j = 1 + \epsilon_j$, where ϵ_j is the sum of the Doppler components along the line joining the *j*th scatterer to the (origin of the) transmitting and receiving apertures (see Middleton, 1967 and 1972a, Figs. 4.4, 16.1). Again $\omega_0 = 2\pi f_0$ is the angular carrier frequency of the transmitter and \hat{S}_0 is the complex envelope of the driving signal $S_{in}(t, \xi) = \text{Re}\{\hat{S}_0(t) \exp[i\omega_0 t - 2\pi i \mathcal{L}_{0T} \cdot \xi]\}$, where now \mathcal{L}_{0T} is a steering delay, employed to change the direction of the transmitter's beam. Finally, as an example of an ambient waveform, from, say, a narrow band interfering source in the medium, we have the analogue of (29:46), viz:

$$\mathbf{U}_{j-\text{n.b.}} = \left[\text{Re} e^{i\omega_0 \mu(t - T_0 + \hat{\mathbf{r}}_R \cdot \mathbf{r}_m/c)} (4\pi R_R)^{-1} \hat{S}_0(t - T_0) \cdot A_R(-\hat{\mathbf{r}}_R f_0, f_0)_m A_I(\hat{\mathbf{r}}_R f_0, f_0) \right]. \tag{29:49c}$$

This concludes our outline of channel modeling 29.2-29.4. We are now able to specify the various signal and noise components of the received waveforms U_j , \mathbf{U}_j which constitute the received process

$$\mathbf{X}(t) = \sum_j \mathbf{U}_j(t), \text{ or } \mathbf{X}(t) = \sum_j \mathbf{U}_j(t), \tag{29:50}$$

respectively for preformed or adapted beams (see 29.3.1). These signal and “noise” components of U_j , namely, $(U_j)_{\text{sig}}$, $(U_j)_{\text{scat}}$, $(U_j)_{\text{self}}$, etc., are combined according to (29:31)-(29:33); their detailed structure is illustrated by the results of this section, for many important physical situations.

29.5 Statistical Formulations

In order to carry out the program outlined at the beginning of 29.2 for multidimensional detection and extraction of signals in random media as well as, and including, the measurement of medium properties themselves, we must next determine the hierarchy ($L = 1, 2, \dots$) of probability densities $W_L(\mathbf{N})_N$, $W_L(\mathbf{V} - \mathbf{S})_N$, etc., which describe the various “noise” fields (29.2.2), which may or may not be accompanied by signals, such as those used here to interrogate or probe the medium itself, as well as to establish the possible presence of localized inhomogeneities or secondary sources.

Two principal approaches are available. In the first or *postulational approach*, we begin by assuming some reasonable statistics for the basic random parameters and random mechanisms in the physical models for the received waveforms $\mathbf{X}_{\text{scat}} (= \sum \mathbf{U}_{j-\text{scat}})$ and $\mathbf{X}_{\text{self}} (= \sum \mathbf{U}_{j-\text{self}})$. (Note that we must have here the explicit forms U_j , cf. (29:50), and 3.4.2.) Next, we calculate $W_L(\mathbf{X})$, etc., by the usual technique of obtaining the appropriate *L*th-order characteristic function, defined as the statistical average $F_L(i\underline{\xi}; \mathbf{X}) \equiv \exp(i\underline{\xi} \cdot \mathbf{X})$, followed by Fourier inversion $W_L(\mathbf{X}) = F_{\underline{\xi}}^{-1} F_L(i\underline{\xi}; \mathbf{X})$. (We remark that the transform relation between characteristic function and probability-density are unique: knowing the one we have the other, in principle. The utility of the characteristic function lies in its greater simplicity for analytic manipulations, and, particularly, as a device for the explicit calculation of probability models, as the following material in 29.5.2 shows.)

In the second, or *Markovian approach*, we start with the premise that \mathbf{X} belongs to, or is a projection of, a simple (first-order) Markoff process† (Middleton, 1960, Section 1.4; also, 1.3-1.5).

29.5.1 Langevin Equations

We next construct the Langevin equations of the process: this is the set or ensemble of the appropriate dynamical equations of $\mathbf{X}(=\Sigma U_j)$, for both ambient and scattered sources, where the set is constructed from all configurations of the $j = 1, \dots, J$ sources or scatterers in the domain, Λ , of “illuminated” and/or viewed sources. We begin with the Langevin equation for the fields L_{self-j} , or L_{scat-j} , etc., (29:27), which is

$$\left\{ \left(\nabla^2 L - \frac{1}{c^2} \frac{\partial^2 L}{\partial t^2} \right) \right\} = \{ -\mathbf{G}_j \}, \text{ or } \{ 0 \}, \tag{29:51}$$

where $\{ \}$ denotes the ensemble. In order to obtain the desired Langevin equation for the process $\mathbf{X}(=\Sigma U_j)$, we must next relate U_j to the deterministic solution L_j , which is done in (29:30) formally. From this we obtain the appropriate Langevin equation for \mathbf{X} , when possible. Once this has been done we then solve the Langevin equation for the set $\mathbf{X} = \{ \mathbf{X}_j \}$. “Solution” here means obtaining the desired probability distributions of \mathbf{X} from an integrodifferential equation (the Boltzmann or Smoluchowski [Chapman-Kolmogoroff] equations), in which the Langevin equation is suitably embedded (Middleton, 1960, Chapter 10, Sec. 10.3).

The first approach, however, effectively bypasses the Langevin equation and its solution, and is direct and comparatively simple to carry out. It has the conceptual (and sometimes practical) disadvantage that we are unable to determine whether or not we have a complete description (Middleton, 1960, Chapter 10.3) of the random process \mathbf{X} , etc., when W_L is obtained in this way. Usually, however, this W_L is sufficient for these detection and measurement situations. Since the principal disadvantage of the Markovian approach in most instances is the very considerable technical difficulty of obtaining explicit solutions from it, the postlational, or direct, approach is recommended.

29.5.2 The Poisson Process

A very important model for many physical scatter and self-noise mechanisms postulates a Poisson process $\mathbf{X}(=\Sigma U_j)$ in space (Middleton, 1967, Sec. 1, 1972), (and, therefore, in time, since $ct=R$ for these random sources (when $\nabla c \neq 0$, cf. 29:35, 29:36, 29:45, and 29:47). This is particularly suited to situations where primary scattering or emission provide the major contribution to the generated field: negligible interaction exists between individual scatterers or sources, which may then be treated as statistically independent. Examples are ocean reverberation, radar “clutter”, automobile ignition noise, and some special cases involving thin clouds and atmospheres. The L th-order characteristic function for M -channel receivers here is found to be, by a direct extension of (2.9 of Middleton, 1967)

$$F_L(i\xi)_{\mathbf{X}} = \exp \int_{\Lambda} \rho(\lambda) \left\langle \exp i \left\{ \sum_{m=1}^M \sum_{\ell=1}^n \xi_{\ell m} U_m(t_{\ell}, \lambda, \underline{\theta}) \right\} - 1 \right\rangle_{\underline{\theta}} d\lambda \tag{29:52}$$

where $\lambda = (t_j = R_j/c, \theta_j, \phi_j)$ are the coordinates of the j th scatterer or ambient sources, $\underline{\theta}$ are a set of random parameters associated with the j th “event” (the corresponding emission or scattering of a field from the j th source). As before, $\mathbf{X} = \sum_j U_j$, (cf 29.3.3 and 29.5.0 above), is the sum of the basic waveforms U_j produced by the sensors or scatterers.

† A first order Markoff process is one about which all (statistical) information is contained in the second-order probability density $W_2(\mathbf{X}_1, t_1; \mathbf{X}_2, t_2)$; higher-order densities $W_n(\mathbf{X}_1, t_1; \dots; \mathbf{X}_n, t_n)$ are all reducible to product combinations of first- and second-order densities W_1, W_2 of the process variable \mathbf{X}_k at times $t_k, k = 1, \dots, n$. For example, it can be shown that $W_n(\mathbf{X}_1, t_1; \dots; \mathbf{X}_n, t_n) = W_1(\mathbf{X}_1, t_1) \prod_{k=2}^n W_2(\mathbf{X}_{k-1}, t_{k-1} | \mathbf{X}_k, t_k), n \geq 2$, where $W_2(1)$ is the conditional p.d., of \mathbf{X}_k at time t_k , given \mathbf{X}_{k-1} at time $t_{k-1}; (t_{k-1} < t_k \leq t_n)$. (See Sections 1.4-3 of Middleton, 1960.)

This $\rho(\underline{\lambda})$ is the (nonnegative) *density* of this M-dimensional Poisson process, which “counts” the number of events arising in the common region Λ of illumination and observation. Thus, ρ depends on the specific geometry of the receiver vis-à-vis the various sources. For example, when these are distributed on a surface, with physical density σ_S , we may follow the analysis of Section 3.1 of Middleton, [1967], and obtain, for monostatic operation (R@T) in the active mode (i.e., the sources are scatterers):

$$\rho_S(\underline{\lambda}) = \sigma_S \left(\frac{\lambda}{2}, \theta_T = \cos^{-1}(h_T/c\lambda), \phi_T \right) \lambda c^2 / 4, \quad c\lambda \geq h_T, \quad (29:53)$$

where h_T is the height of the receiver (O_R) above the surface, etc. For such sources distributed in a volume, with physical density σ_V , we get similarly

$$\rho_V(\underline{\lambda}) = \sigma_V \left(\frac{\lambda}{2}, \theta_T, \phi_T \right) \lambda^2 c^3 \sin \theta_T / 8. \quad (29:54)$$

The differential elements $d\lambda$ are $d\lambda d\phi_T$, $d\lambda d\theta_T d\phi_T$, respectively, in (29:52). On the other hand, for reception in an ambient field, where now the (point) sources in the medium are the primary emitters, and it is this emission field which constitute the medium properties under study, (29:53) and (29:54) become

$$\rho_S(\underline{\lambda}) = \sigma_S(\lambda, \theta_R = \cos^{-1}(h_R/c\lambda), \phi_R) \lambda c^2, \quad (29:55a)$$

$$\rho_V(\underline{\lambda}) = \sigma_V(\lambda, \theta_R, \phi_R) \lambda^2 c^3 \sin \theta_R, \quad (29:55b)$$

with the differential elements $d\lambda$, formally the same as before.

Of special interest are the cases $L = 1, 2$. From (29:52) we have then

$$F_1(i\underline{\xi}_1, t_1)_X = \exp \int_{\Lambda} \rho(\underline{\lambda}) \left\langle \exp \left(i \sum_{m=1}^M \xi_{1m} U_m(t_1; \underline{\lambda}, \underline{\theta}) \right) - 1 \right\rangle_{\underline{\theta}} d\lambda, \quad (29:56)$$

with $\underline{\xi}_1 =$ the vector $[\xi_{1m}]$, and

$$F_2(i\underline{\xi}_1, t_1; i\underline{\xi}_2, t_2)_X = \exp \int_{\Lambda} \rho(\underline{\lambda}) \left\langle \exp \sum_{m=1}^M [i\xi_{1m} U_m(t_1; \underline{\lambda}, \underline{\theta}) + i\xi_{2m} U_m(t_2; \underline{\lambda}, \underline{\theta})] - 1 \right\rangle_{\underline{\theta}} d\lambda, \quad (29:57)$$

where the U_m are the typical received waveforms after the m^{th} sensor or m^{th} channel into \mathbf{T}_{R_0} , cf., 29.4.2 above. Of particular importance, also, are the (vector) mean and the (dyadic) covariance functions:

$$K_X(t_1, t_2)_{km} = \left[- \frac{\partial F_2}{\partial \xi_{1k} \partial \xi_{2m}} - \left(-i \frac{\partial F_2}{\partial \xi_{1k}} \right) \left(-i \frac{\partial F_1}{\partial \xi_{2m}} \right) \right]_{\xi_1 = \xi_2 = 0} \quad (29:58)$$

$$\langle \mathbf{X}(t_1) \rangle = -i \frac{\partial F_1}{\partial \xi_1} \Big|_{\xi_1=0} = \int_{\Lambda} \rho(\underline{\lambda}) \langle U(t_1; \underline{\lambda}, \underline{\theta}) \rangle_{\underline{\theta}} d\lambda = \langle \mathbf{X}_m(t_1) \rangle, \quad (29:58a)$$

(Note that for a completely deterministic source or target location, we have $\rho = \delta(\underline{\lambda} - \underline{\lambda}_0)$, and $\langle \mathbf{X}(t_1) \rangle = \langle \mathbf{X}(t_1; \underline{\lambda}_0, \underline{\varrho}) \rangle$ which is the correct result of \mathbf{X}_1).

Since

we have the covariance *matrix*

$$\mathbf{K}_X(t_1, t_2) = \int_{\Lambda} \rho(\underline{\lambda}) \langle \mathbf{U}(t_1; \underline{\lambda}, \underline{\varrho}) \mathbf{U}(t_2; \underline{\lambda}, \underline{\varrho}) \rangle_{\underline{\varrho}} d\underline{\lambda} = [\mathbf{K}_{km}(t_1, t_2)] , \quad (29:58b)$$

with elements

$$\mathbf{K}_{km}(t_1, t_2) = \int_{\Lambda} \rho(\underline{\lambda}) \langle U_k(t_1; \underline{\lambda}, \underline{\varrho}) U_m(t_2; \underline{\lambda}, \underline{\varrho}) \rangle_{\underline{\varrho}} d\underline{\lambda} , k, m = 1, \dots, M . \quad (29:58c)$$

29.5.3 The Gauss Process

In many cases the density ρ is sufficiently large that \mathbf{X} is an M -dimensional gauss process, now completely determined by (29:58b,c). The general n^{th} -order d.d. of $\mathbf{X}(t)$ is given by

$$W_L(X_1, t_1; \dots; X_n, t_n | \Lambda) = [\exp \left(-\frac{1}{2} \tilde{\mathbf{x}} \mathbf{K}_X^{-1} \mathbf{x} \right)] / (2\pi)^{n/2} (\det \mathbf{K}_X)^{1/2} \quad (29:59a)$$

with the c.f.

$$F_n(i\xi_1, t_1; \dots; i\xi_n, t_n | \Lambda) = \exp \left(-\frac{1}{2} \tilde{\xi} \mathbf{K}_X \xi + i\xi \cdot \langle \mathbf{X} \rangle \right) . \quad (29:59b)$$

Here, $\mathbf{K}_X = [x(t_k)x(t_l)] = \tilde{\mathbf{K}}_X$, ($k, l = 1, \dots, L$), is the symmetrical) covariance matrix of the *fluctuation* $x_m = [x_{mk}] = [\mathbf{X} - \langle \mathbf{X} \rangle]_m$, and $\mathbf{x} = [x_m]$, etc. Specifically, $(\mathbf{K}_X)_{k1}$ is obtained from (29:58b, c). For a variety of important special cases, see Section 10.3 of Middleton, (1967).

29.6 Some Statistical Results: Covariance of Received Scattered and Ambient Fields

Let us develop the exposition of the preceding sections further with some specific examples, both theoretical and experimental. Because of such technical advances as the highspeed, high memory computer, and reliable, rapid methods of data acquisition and recording (e.g., magnetic tapes), etc., it has become possible in the last few years to study and test the validity and practicality of theoretical models like those outlined in (29.4,5) above.

We begin (see 29.6.1) with some explicit analytical results for the covariance† of some general geophysical processes, which may include tropospheric situations in particular cases. Our particular examples, however, are chosen mainly from oceanic studies, where our new methods have had their initial applications, and for which specific data are just now becoming available. Our examples here are, (i), a received scatter process, such as ocean reverberation and; (ii), a received ambient scalar acoustical field, such as the background noise in the sea produced by some biological organisms, for instance. An extension to the vector case (EM scattering) follows, and we conclude (see 29.6.2) with some experimental results for reverberation from a lake surface. (In this latter instance preformed beams were used, so that $M = 1$.)

29.6.1 Received Scattered and Ambient Fields

For illustration we restrict ourselves here to monostatic operations, small Doppler, (point)-scatterers (or sources) with time-invariant cross-sections, $\gamma_0(\mathbf{R})$, cf. (29:44). We also assume for the active mode narrow-band input signals to the medium, or in the passive cases involving ambient fields, those sources which emit narrow-band radiation. (This is not a conceptual restriction, as theory can handle the broad-band cases equally well; the resulting expressions are somewhat more complex, however, cf. (29:43), (29:45) vs. (29:46), (29:47) for typical waveforms.)

† Although $\langle \mathbf{X} \rangle \neq 0$ can occur (cf. the examples, 29.8 following), we usually need to consider only \mathbf{K}_X , either because of the way the ensemble of (received) waves is generated, or because of inherent mechanisms in the medium, scatterer, etc., which preclude a nonvanishing mean.

29.6.1.1 Scattered Fields

Combining (29:46) with (29:53) in (29:58 b, c) we obtain for received scatter from a surface illuminated by a narrow-band scalar field in the monostatic case ($\mathbf{R}@\mathbf{T}$), ($\nabla c = 0$), the covariance matrix

$$K_X(t_1, t_2) = (1/2)\text{Re}\{B_X^* e^{i\omega_0 \tau}\} = R_X(t_1, t_2) \cos \omega_0 \tau + \Lambda_X(t_1, t_2) \sin \omega_0 \tau, \quad (29:60)$$

with the (*) denoting the complex conjugate, and $\tau = t_2 - t_1$, where†

$$\begin{aligned} B_X^*(t_1, t_2) &= \tilde{B}_X(t_2, t_1) = [B_X^*(t_1, t_2)]_{km} \\ &= \left[\frac{1}{4^3 \pi^4 c^2} \int_{\Lambda_S} \frac{\sigma_s(\lambda)}{\lambda^3} \left\langle Q_{RT}^{(k)}(\lambda) Q_{RT}^{(m)}(\lambda) \right\rangle_{\gamma_0} \right. \\ &\quad \cdot \left\langle \mu^2 e^{i\omega_0 \epsilon \tau + i\omega_0 \mu \hat{a}_T \cdot (\mathbf{r}_m - \mathbf{r}_k)} \hat{S}_0(\mu[t_1 - T_0(\lambda)]^*) \right. \\ &\quad \left. \left. \cdot \hat{S}_0(\mu[t_2 - T_0(\lambda)]\phi) \right\rangle_{\mu, \epsilon = \mu - 1} d\lambda \right] \end{aligned} \quad (29:61)$$

is an $M \times M$ matrix, and $B/2 = R_X + i\Lambda_X$. Again, σ_S is the density of scatterers on the surface, $\lambda = (\lambda, \phi)$, $d\lambda = d\lambda d\phi$ for surfaces (λ measured in seconds, e.g., $c\lambda = R$), Λ_S is the domain of illuminated and observed scatterers, and $\hat{a}_T = \hat{\mathbf{i}}_T/c$ is a delay per unit length, (see 29:23b). Note that because of the dispersive nature of the medium, i.e., $\mu \neq 1$, there is an interelement coupling in the adapted beam, which will degrade beam forming vis-à-vis the nondispersive cases. Here we have an M -element, adapted beam, and each X_m in $[X_m] = \mathbf{X}$ is the resulting input, from the m^{th} aperture element (at \mathbf{r}_m from O_R , cf., F29.5), to the processor T_{R_0} . The coupled system cross-section, $Q_{RT}^{(m)}$, is given by (29:48), with $\mathbf{a}_1 \rightarrow -\hat{\mathbf{i}}_T/c = -\hat{\mathbf{i}}_T$, since $\nabla c = 0$ here.

In the vector cases, where, for example, the (narrow-band) driving signal produces an *electromagnetic field*, our result (29:60), (29:61) also, applies, provided we replace $Q_{RT}^{(m)}$ by the quadratic form (29:49b). We must distinguish, also, the mode of excitation; i.e., whether or not we have an electric or magnetic drive. We do this by an index α ($\alpha = E$, or M), so that we write $Q_{RT}^{(m)} \rightarrow (\alpha)Q_{RT}^{(m)} = \mathbf{A}_{RM}^{(\alpha)} \cdot \hat{\gamma}(\alpha) \cdot \mathbf{A}(\alpha)$; σ_S by $\sigma_S^{(\alpha)}$, and similarly for $B_X^{(\alpha)}(t_1, t_2)$, (29:61). In the general case, of course, we shall have some linear combination of the two. The resultant E-H field then obeys the dynamical equations (29:41), (29:42), as do the scattered fields. For a full development, see Middleton (1971).

In these vector field cases it should be pointed out that $Q_{RT}^{(\alpha)}$ is not generally symmetric, since $\gamma_{oj}^{(\alpha)} \neq \gamma_{oj}^{(\alpha)}$, inasmuch as the absorptive-radiative characteristics of the scatterers may not be symmetric. For this reason, reciprocity fails; interchanging receiver and transmitter does *not* yield the same waveform, $U_j(\alpha)$, at the receiver. See comments following (29:29). On the hand, when symmetry is a reasonable assumption e.g., $\hat{\gamma}_{oj}^{(\alpha)} = \hat{\gamma}_{oj}^{(\alpha)}$, or $[\hat{\gamma}_{kl}^{(\alpha)} = \hat{\gamma}_{lk}^{(\alpha)}]$, reciprocity does occur, and the system cross-section $Q_{RT}^{(\alpha)}$ is correspondingly simplified. In the most elementary situations, the scatterer introduces no mixing of the incident radiation—the component incident in the x -direction is emitted (albeit modified) in the x -direction without a portion being transferred to the y - or z -directions, etc., and there is no preferred direction of reradiation.

† The validity of (29:60) in these nonstationary cases, when a *complex* signal representation is used, depends on the manner in which the ensemble $X(t)$ of scattered waves is generated and observed. Thus, in more detail, we relate $X(t)$ to the observer by writing $X = X(t-\epsilon)$, where ϵ is a (random) epoch. By assigning various values and/or distributions to ϵ , we can generate different ensembles (see Middleton, 1960, Sections 1.3, 1.7). It is sufficient that ϵ be symmetrically distributed over an "RF" cycle ($1/f_0$). This is physically established here by the actual scatter process, independent of the observer's scheme of observation.

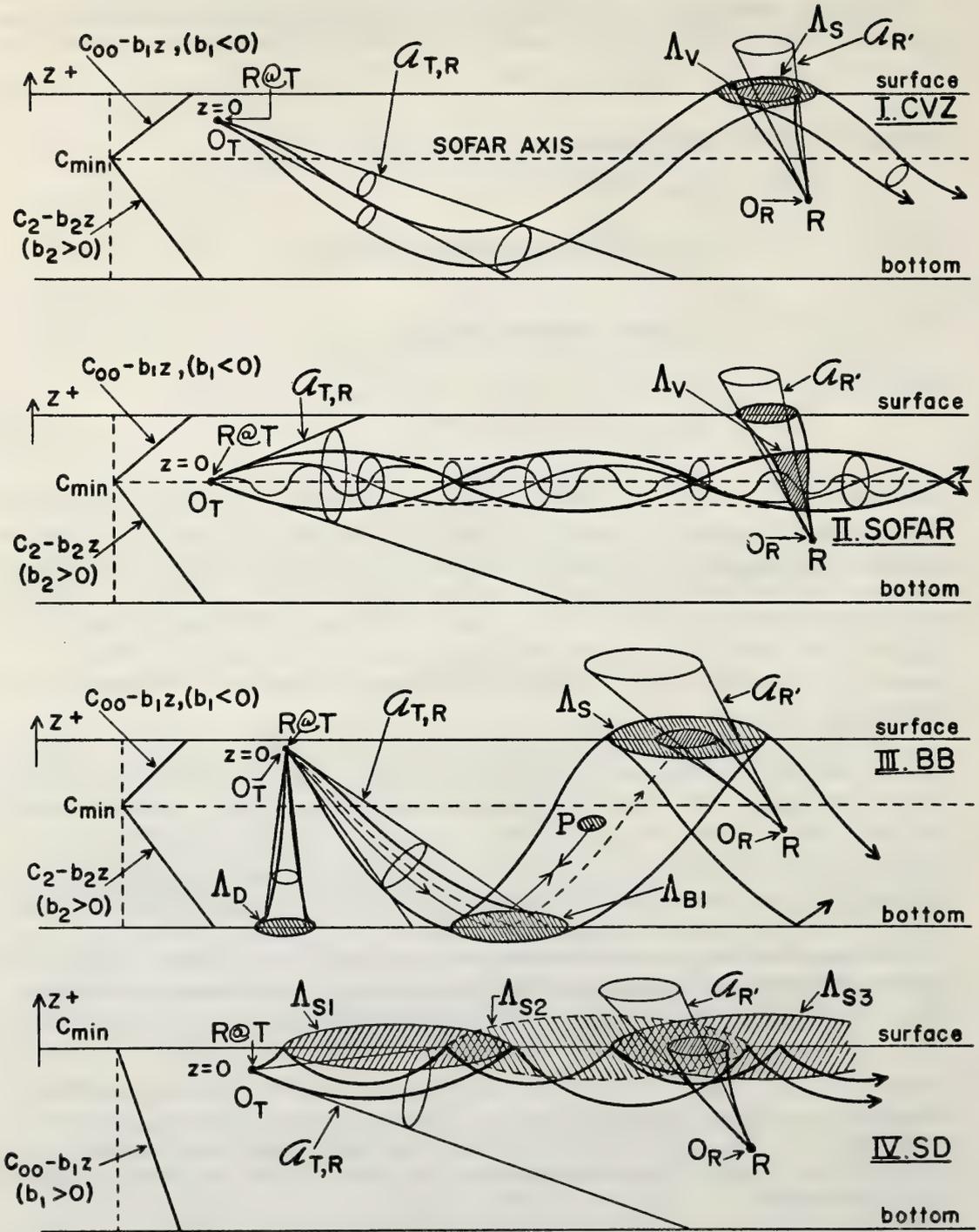


Figure 29.5 Four important modes of propagation in a deep ocean with a non-vanishing velocity gradient ($\Delta c \neq 0$).

29.6.1.2 Ambient Fields

Consider a set of (point-) sources distributed on a surface and radiating to a receiver at some point $P(\mathbf{R})$, not necessarily on the surface. Applying (29:53) and (29:49c) to (29:58b) in this instance, we get for a typical narrow-band received scalar ambient field, again the covariance matrix $\mathbf{K}_X(t_1, t_2)$, (29:60), but where \mathbf{B}_X , (29:61), now becomes explicitly

$$\begin{aligned} \mathbf{B}_X^{(*)}(t_1, t_2) &= [\mathbf{B}_X^{*}(t_1, t_2)_{km}] \\ &= \left[\frac{1}{(\pi c)^2} \int_{\Lambda_s} \frac{\sigma_I(\lambda)}{\lambda} \left\langle \left\langle q_{RI}^{(k)}(\lambda) \right\rangle^* q_{RI}^{(m)}(\lambda) \right\rangle_{(T)} \right. \\ &\quad \cdot \left\langle e^{[i\omega_0 \epsilon \tau + i\omega_0 \mu \hat{Q}_T \cdot (\mathbf{r}_m - \mathbf{r}_k)]} \hat{S}_0(\mu [t_1 - T_0(\lambda)]^* I) \right. \\ &\quad \left. \left. \cdot \hat{S}_0(\mu [t_2 - T_0(\lambda)] I) \right\rangle_{\mu, \epsilon = \mu - 1, I d\lambda} \right], \end{aligned} \tag{29:62}$$

where the subscript (T) in $\langle \rangle_{(T)}$ denotes the *average over* ambient source emission direction with the usual small Dopplers ($\mu \approx 1$), and where in these passive situations

$$q_{RI}^{(m)}(\lambda) \equiv A_R(-\hat{\mathbf{I}}_R f_0/c, f_0)_m A_I(\hat{\mathbf{I}}_R f_0/c, f_0) \tag{29:63}$$

is the analogue of the coupled-system cross-section (29:48) in the active scatter cases. The average $\langle \rangle_I$ is over possible random parameters in the signal source, I, where the (complex) envelope is \hat{S}_{0I} above.

For vector ambient fields, (29:62) still applies, but from (29:28) we see that (29:63) is typically transformed into the scalar product

$$q_{RI}^{(m)}(\lambda) \equiv A_R(-\hat{\mathbf{I}}_R f_0/c, f_0)_m \cdot A_I(\hat{\mathbf{I}}_R f_0/c, f_0), \tag{29:64}$$

where now A_R, A_I are *vector* beam patterns, cf. remarks following (29:25). In the electromagnetic cases we must again consider a combination of fields, depending on the mode of emission of the source, cf. 29.6.1.1 above.

29.6.2 Reverberation From A Lake Surface (Plemons, 1971)

As an example of the interaction between theory and experiment, involving a scatter channel of the types modeled above, we cite some results here of recent work (Plemons, 1971) on underwater acoustic scattering from a lake surface. In particular, we wish to determine the covariance (function) of the reverberation return in the active case, and the same quantity for ambient fields in the passive case. The conditions of operation were:

- (1) a horizontally directed beam, radiating into the volume at a depth $h = 8$ ft below the lake surface;
- (2) monostatic operation (R@T);
- (3) signal consisting of a set of incoherently generated, comparatively short pulses, "pinging" the medium;
- (4) fixed platform;
- (5) narrow-band signals, linear FM or "chirped" carrier; $f_0 = 110$ kHz, duration $T = 1.25$ msec; bandwidth 9.6 kHz;
- (6) volume reverberation was found to be negligible.

For this experimental configuration and with a preformed beam ($M=1$), the covariance matrix (29:60) can be

shown to reduce to the very simple, one-element form (for real signals)

$$K_X(t_1, t_2) = \int_{-\infty}^{\infty} G(\lambda) S_{in}(t_1 - \lambda) S_{in}^*(t_2 - \lambda) d\lambda, \quad (29:65)$$

where

$$G(\lambda) \equiv \frac{1}{4^3 \pi^4 c^2} [\langle \gamma_0^2(\lambda) \rangle \sigma_S(\lambda)] \left[\frac{1}{\lambda^3} \int_0^{2\pi} |A_R(\lambda, \phi) A_T(\lambda, \phi)|^2 d\phi \right], \quad (29:66)$$

for these sufficiently directional beam patterns; $\langle \gamma_0^2 \rangle \sigma_S \equiv J(\lambda)$, the backscattering coefficient. For the comparatively short-duration signals used, the further approximations apply:

$$K_X(t_1, t_2) \doteq K_X(\tau|t_1) = G(t_1) \int_{-\infty}^{\infty} S_{in}(\lambda) S_{in}^*(\lambda + \tau) d\lambda; \tau = t_2 - t_1. \quad (29:67)$$

[S_{in} , of course, is physically real.] In normalized form, (29:67) becomes

$$K_{norm}(\tau|t_1) \equiv K_X(\tau|t_1)/K_X(0|t_1) = \int_{-\infty}^{\infty} S_{in}(\lambda) S_{in}^*(\lambda + \tau) d\lambda / \int_{-\infty}^{\infty} |S_{in}(\lambda)|^2 d\lambda. \quad (29:68)$$

The denominator is (twice) the energy in the driving signal. The *mean intensity* of the received reverberation (after the aperture but before T_{R_0}) is

$$\langle X(t_1)^2 \rangle \equiv K_X(t_1, t_1) = G(t_1) \int_{-\infty}^{\infty} |S_{in}(\lambda)|^2 d\lambda. \quad (29:69)$$

Figure (29.6) shows the mean intensity $\langle X(t_1)^2 \rangle$ as a function of range ($\sim t_1$), on an arbitrary scale. Instead of one scattering surface, two were found, the second attributed to an abrupt discontinuity in the temperature profile at 28 feet below the lake surface. By applying the above to *each* scattering surface, a good theoretical agreement with experiment was obtained, as (F29.6) indicates. Figure (29.7) shows the experimental version, \bar{K}_O , of the envelope of $K_X(\tau|t_1)$, viz., $K_O(\tau|t_1) = \sqrt{R_O^2 + \Lambda_O^2} = |R_O - i\Lambda_O| (= \frac{1}{2} B_X^*|_1)$, cf. (29:60), (29:61), in both normalized and unnormalized forms. The peak in $(\bar{K}_O)_{non-norm}$, (F29.7), corresponds to the second maximum of $\langle X^2 \rangle$ in (F29.6), where the main vertical beam impinges on the deeper layer. (The first maximum of $\langle X^2 \rangle$, at about $t_1 = 26$ msec, corresponds overall to the main beam's intersection with the surface.) Finally, (F29.8) shows K_O as a function of τ , at two different ranges $t_1 = 54, 82$ msec. The symmetry $\bar{K}_O(\tau|t_1) \doteq \bar{K}_O(-\tau|t_1)$, as expected for these short signals, is good. (The departure from the dotted lines is caused by the frequency selectivity of the apertures here, and agrees well with the results when these measured responses are included; with no selectivity it would be the electronic signal waveform that would be received, and the data should coincide with the dashed curve.) The results of this study (Plemons, 1971) have shown that good agreement between theory and experiment can be obtained if proper care of, and attention to, the critical geometrical factors are exercised.

29.7 Statistical Data Analysis (SDA)

Essential prerequisites for the estimation of statistical quantities, such as the covariance functions and probability distributions of scatter and reverberation discussed in 29.5, and 29.6, are the construction and the validation of the data ensemble from which these quantities are to be estimated. Fundamentally, validation means that the observed data ensemble can be shown to be a subset of a statistically proper parent ensemble. This is achieved by showing that the members of the sample ensemble are *independent* and form a *homogeneous set*. The present tests for independence and homogeneity of reverberation data (i.e., here data for acoustic scattering into the medium (water), from a water-air interface) are briefly outlined below. They are applications of standard statistical techniques, which have recently been adapted (Middleton, 1969) to this particular class of problems. (For more details, see Middleton, 1969, and references therein.)

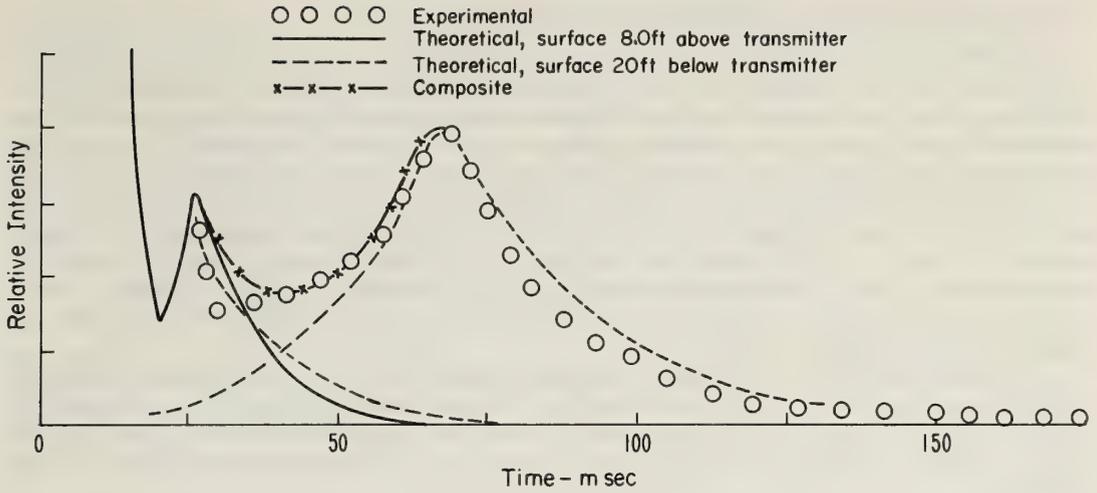


Figure 29.6 Theoretical and experimental reverberation intensities.
 Transmitter and receiver depth = 8.0 ft.
 Transmitted signal = 1.25 msec FM

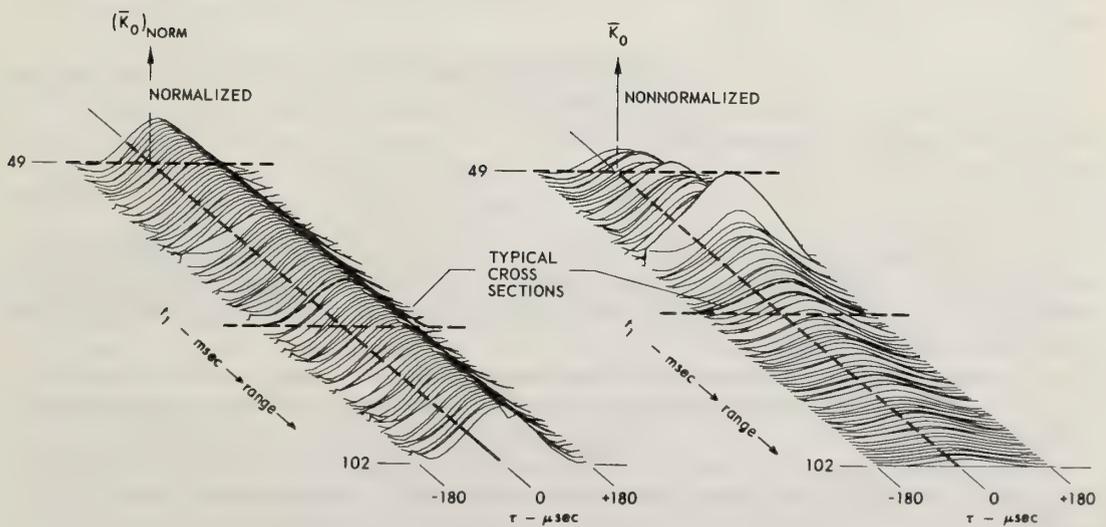


Figure 29.7 Two-dimensional envelope structure of the covariance of surface reverberation.
 Ensemble size = 150

29.7.1 A Test For Independence: The One-Sample Runs Test

Validation of the data begins by testing the data for independence, here the “null” hypothesis, H_0 , with the help of a one-sample runs test (Middleton, 1969, Sections III-E). The test for independence, or *randomness*, among the member elements of the sample ensemble is based on the number of runs in the data sample, where, in our case, a run is defined as a succession of similar events or attributes. To implement the runs test here, the positive and negative values of the high frequency reverberation amplitudes at given range (t_1) are chosen as the attributes to be tested. The number of runs in the data sample is a measure of the intersample randomness, or independence, of the reverberation returns. Thus, if a set of positive and negative numbers, say, representing the instantaneous reverberation amplitudes, produces very few runs because of a disproportionately large number of positive amplitudes, then it is concluded (*viz.*, the alternative hypothesis, H_1 , is accepted) that the data are not entirely random and that some deterministic mechanism, such as a nonzero mean value, is present.

Specifically we proceed as follows: In a given data sample let M_1 be the number of positive values (+’s) and M_2 be the number of (–’s) of the returns, $X(t_1)$. The total number of samples in our experimental ensemble is $M = M_1 + M_2$. The order in which these attributes occur in this data ensemble is observed and the number of runs, R , is counted. Two cases are usually distinguished, for which different distributions of R are obtained:

Small Samples [$M_1, M_2 < 20$] and *Large Samples* [$M_1, M_2 \geq 20$]. Since the data ensembles in our example here have 150 members (F 29.6-F 29.8), we have a large-samples situation, where the p.d. of R is asymptotically normal. The implementation of the runs test then consists of the following steps (*cf.*, Middleton, 1969, Section III-C):

1. In the 150-member ensemble, the number of +’s (M_1) and the number of –’s (M_2) and the number of runs (R) are determined.

2. The sample mean, $\mu_R = \bar{R}$, and variance, σ_R , are calculated from

$$\mu_R = \bar{R} = \frac{2M_1M_2}{M_1 + M_2} + 1, \quad (29:70a)$$

and

$$\sigma_R^2 = \frac{2M_1M_2(2M_1M_2 - M_1 - M_2)}{(M_1 + M_2)^2(M_1 + M_2 - 1)}. \quad (29:70b)$$

3. The test statistic Z_{sample} is computed from

$$Z_{\text{sample}} = \frac{R - \mu_R}{\sigma_R}. \quad (29:70c)$$

4. A *significance level* α (\equiv conditional probability of a Type 1 error, which is the probability of saying that the data are *not* homogeneous when they actually are, *cf.* 29:15) is next chosen (Middleton, 1969, Section III-C; Siegel, 1956), and from the appropriate Tables (Tables 4 and 5 of Middleton, 1969) a threshold Z_α is then determined.

5. Finally, a decision is made according to the following criteria:

(a) If $Z_{\text{sample}} \leq Z_\alpha$, then we conclude that the data ensemble is randomly generated: (H_0 is accepted).

(b) If $Z_{\text{sample}} > Z_\alpha$, then we conclude that the data ensemble is *not* randomly generated: (H_1 is accepted and H_0 rejected). This conclusion is correct, on the average, $(1 - \alpha)100\%$ of the time.†

29.7.2 A Test For Homogeneity: The Kolmogorov-Smirnov Test

Once the independence of the individual data samples has been established, the sample ensemble is next tested for homogeneity (H_0). This is done with the help of an appropriate Kolmogorov-Smirnov (K-S) test, which is based on a comparison of the cumulative probability distributions (here, of the amplitudes or

† In most applications, the Type II error probability, β , *cf.* (29:15) is not available, since the p.d.f. associated with the alternative hypothesis state (H_1) is not known.

magnitude of the envelope) of various subsets of the original data sample (Middleton, 1969, Section III-B). If the data sample does belong to a single parent population, then different subsets may be expected to have the same probability measures (within the confidence limits of the test). Thus, one may begin by dividing the original data sample into two subsets, and then compare their cumulative distributions. If the two distributions are the same, within a statistically reasonable difference (determined by some critical level, α), this is an acceptable (but, of course, not conclusive) indication that the data may be homogeneous. The K-S tests can be repeated for additional insight as to our conclusion, by subdividing the subsets and again comparing the cumulative distributions of the smaller subsets. Sample size, processing capability, and time are usually the factors which limit the extent to which we can make sample subdivisions and apply the K-S test. In the present example we stop at a single subdivision of the sample ensemble.

To test for homogeneity at the time t_1 , for example, we split the data sample into two subsets, each containing (for convenience only) an equal number of samples. Thus, two subsets each containing 75 samples are obtained from the 150-sample ensemble. The K-S test is then applied to compare the two cumulative distributions of these 75-sample subsets. The experimental test statistic is Z_{sample} , which for this test is defined as the maximum absolute difference between the two experimental (cumulative) distribution functions F_1 and F_2 :

$$Z_{\text{sample}} = \text{MAX}|F_1 - F_2| . \tag{29:71}$$

This statistic, Z_{sample} , is then compared to a pre-chosen threshold Z_α , and a decision is made according to

$Z_{\text{sample}} \leq Z_\alpha$: H_0 : Data accepted as homogeneous

or

$$\tag{29:72}$$

$Z_{\text{sample}} > Z_\alpha$: H_1 : Data rejected as inhomogeneous.

For a prechosen value of α , Z_α is obtained by reference to an appropriate set of tables (Middleton, 1969, Section III-B; see also Plemons, 1971).

29.7.3 Application: Validation Of Reverberation Data (FM Signals)

The data from which the statistical estimates shown in (F29.6-F29.8) were made proved to be independent *and* homogeneous, at all times (ranges), as (F29.9) shows. Here the time (range) interval is (58, 105) msec, which does not correspond to the entire region of analysis of the data. A total of 291 sample ensembles was selected from the (58, 105) msec intervals. The time interval between any two consecutive ensembles is (58, 105) msec/291 = 0.16 msec. The reverberation samples separated by 0.16 msec are statistically uncorrelated, which means that the 291 runs tests and K-S tests for homogeneity are independent tests. The number of failures of individual tests for independence (randomness of the data) is 15(= 5.2% of 291) and the number of failures of individual tests in regard to homogeneity (at $\alpha = 0.05$) is 17(= 5.8% of 291). These are to be compared to the expected 15 ($\approx \alpha 291$), or 5% failures (at the significance level $\alpha = 0.05$), and respective standard deviation $\sigma = [(291)(0.05)(0.95)]^{1/2} (\approx 4)$ about this mean. These data fall well within the bounds of one or two σ 's about the mean, so that we readily consider the reverberation as independently generated and homogeneous over the interval (58, 105) msec.

29.7.4 Tests Of The First Order Probability Distributions Of Reverberation And Its Envelope

In addition to the important second-order moments of the acoustic scatter, further statistical information, necessary in problems of signal processing, may be gained from the probability distributions of the process. Here we shall consider the first-order probability distributions that partially characterize these narrow-band reverberation processes and their envelopes. In many practical situations, theory tells us to expect gaussian distributions of amplitude, and Rayleigh distributions of the envelopes (Middleton, 1967, Section 10). However, a variety of mechanisms may intervene to modify these predictions: for example, a few strong scatterers imbedded in the much larger number of scatterers of approximately equal strength will produce a mixed impulsive-gauss process which can depart noticeably from a normal process. Active, biological sources in the medium can also produce similar modifications.

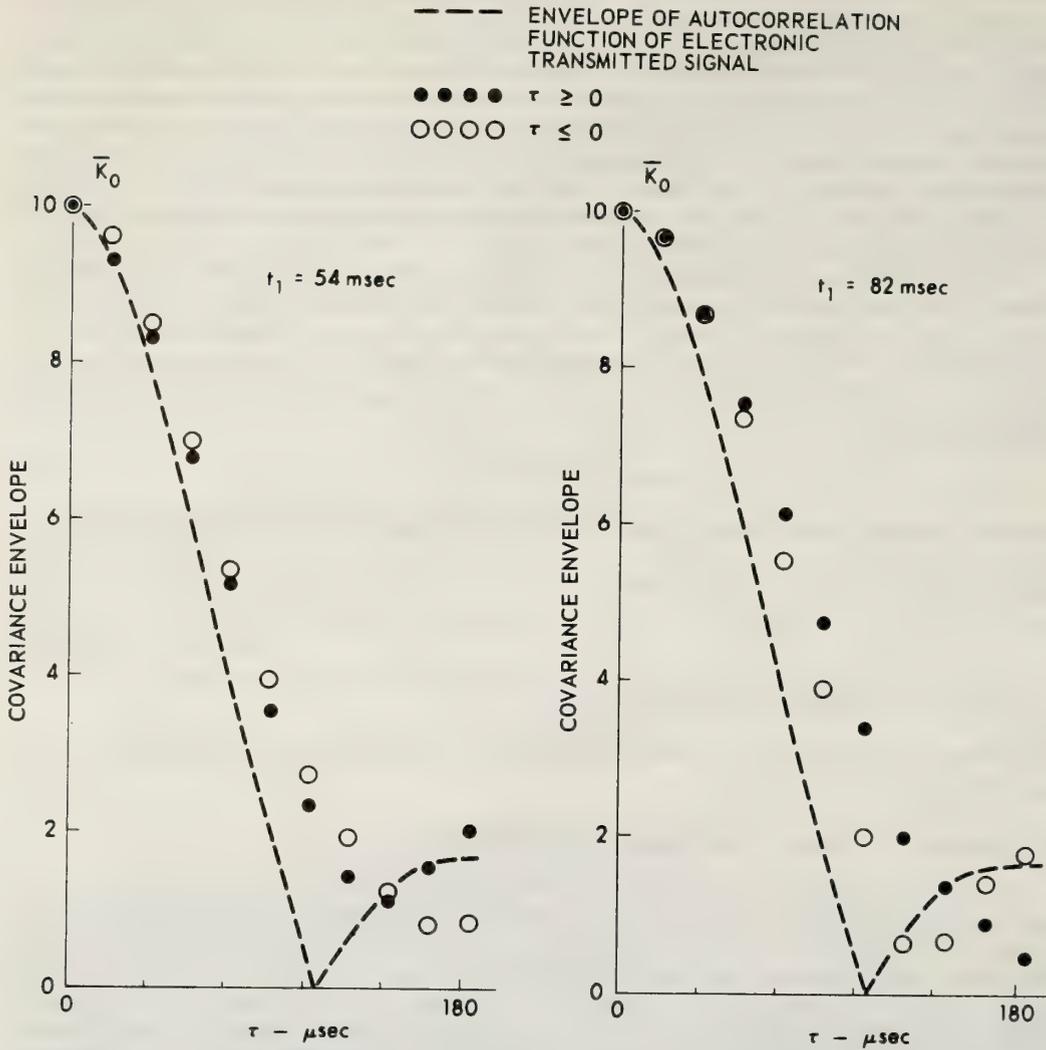
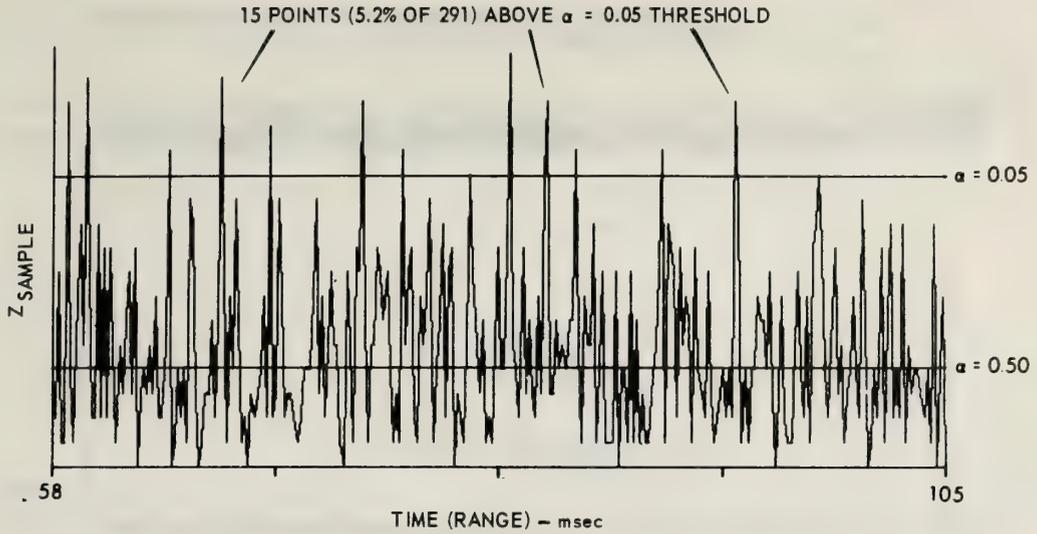


Figure 29.8 Envelope \bar{K}_0 of the non-normalized covariance of surface reverberation. Ensemble size = 150

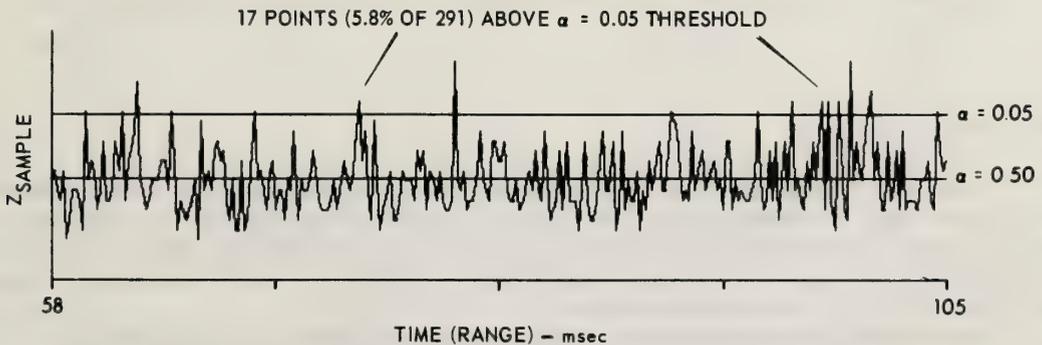
Accordingly, we should test our assumptions as to the gaussian and Rayleigh character of these scatter processes. This is done here with the aid of a Kolmogorov-Smirnov "Goodness-of-Fit" test (Middleton, 1969, Section IV-B). This compares the experimental cumulative distributions (of amplitude and envelope) with the expected, theoretical distributions. The test statistic is (Middleton, 1969, Eq. 4.10)

$$Z_{\text{sample}} = \max |F - S|_{t\varrho} \tag{29:73}$$

where $F(X(t\varrho))$ or $E(t\varrho)$ is the postulated cumulative distribution of the instantaneous amplitude or envelope, and $S(X(t\varrho))$ or $E(t\varrho)$ is the experimental distribution of these quantities. In the usual way the test itself compares Z_{sample} with a threshold Z_α . Thus, whenever Z_{sample} exceeds Z_α , the threshold value for the prechosen critical level α , our assumed distribution fails the test (H_1 accepted). We say then that the actual



(a) ONE SAMPLE RUNS TEST FOR SAMPLE INDEPENDENCE
(ENSEMBLE SIZE = 150)



(b) KOLMOGOROV-SMIRNOV TEST FOR HOMOGENEITY
(1st 75 MEMBERS OF EACH ENSEMBLE
COMPARED WITH 2nd 75 MEMBERS)

Figure 29.9 Tests for independence and homogeneity of FM reverberation.
The time interval (58,105) msec contains 291 independent cross sections (data ensembles).

and postulated distributions are (significantly) different. Otherwise, we conclude that they are the same, i.e., H_0 accepted, (with probability α that this conclusion is incorrect). Applying the above to the amplitude and envelope data, we see, as (F29.10) indicates, that this reverberation process is gaussian with a Rayleigh distribution of envelopes (at the $\alpha = 0.05$ significance level) for all ranges, as theory predicts.

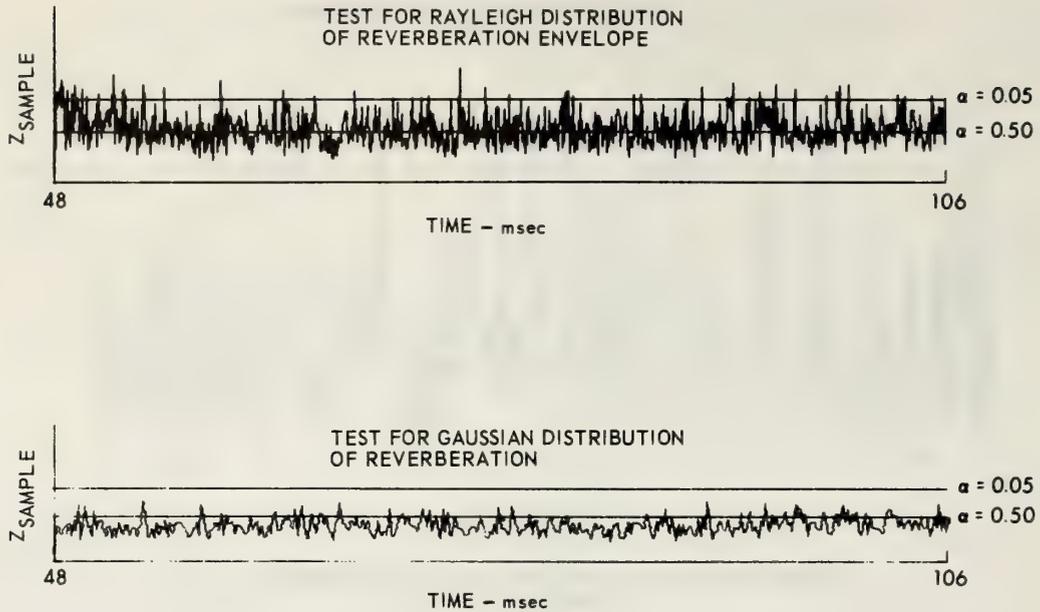


Figure 29.10 *Kolmogorov-Smirnov test for Rayleigh and Gaussian distribution of FM reverberation.*

29.8 Optimum Reception

We turn now to steps 4 and 5 in our program, outlined at the beginning of 29.1.2. This is the central task of incorporating the statistical-physical description of the medium and channel into the decision structure 29.1.4, which in turn permits us to find optimal, or near-optimal algorithms for processing the received data, and to evaluate and compare systems for so doing. We shall consider only optimum systems in the discussions below, 29.1.5, 29:18, where the main problems are to:

- (1) determine optimum processors, $T_{R_0}^*$, for signal detection or extraction (i.e., measurement);
- (2) determine the expected performance of such systems, as evaluated by a minimum average, or Bayes risk, R^* , (29:18).

In these examples, we are concerned with the multidimensional ($M \geq 1$) detection and extraction of signals in and from random media, in both the active and passive régimes 29.1.3. The framework for this has been described in Section 29.1. This includes as one major area of interest in remote sensing of the troposphere, the measurement of the “noise” or background field itself, which often is the desired “signal”, in our communication theory terminology. Methods of optimal sampling (measurement) of the field are referred to in Section 29.9. In fact, in our tropospheric measurements we are often sensing a given spatial location, observing the phenomena drifting through it with the wind. Sometimes these may be represented by “signal”, analogous to “targets”, e.g., reflecting discontinuities in a general background. Sometimes we may be observing self-radiating sources, and often, when probing the medium, we have to operate against our own self-generated “clutter” or reverberation. In many cases, too, we may wish to observe these phenomena, “targets,” fields, etc., continuously, over comparatively long times. To indicate and illustrate how systems may be generally designed for these purposes, and their performance evaluated, we cite below some examples from radar and sonar applications in analogous situations, which embody the SCT methods reviewed in this chapter.

We emphasize again the critical role of the channel physics and geometry, 29.2-29.4. Accordingly, let us illustrate the approach with some recent results, with an example, each, of signal detection and extraction when an M-channel (adapted beam) receiver is used.

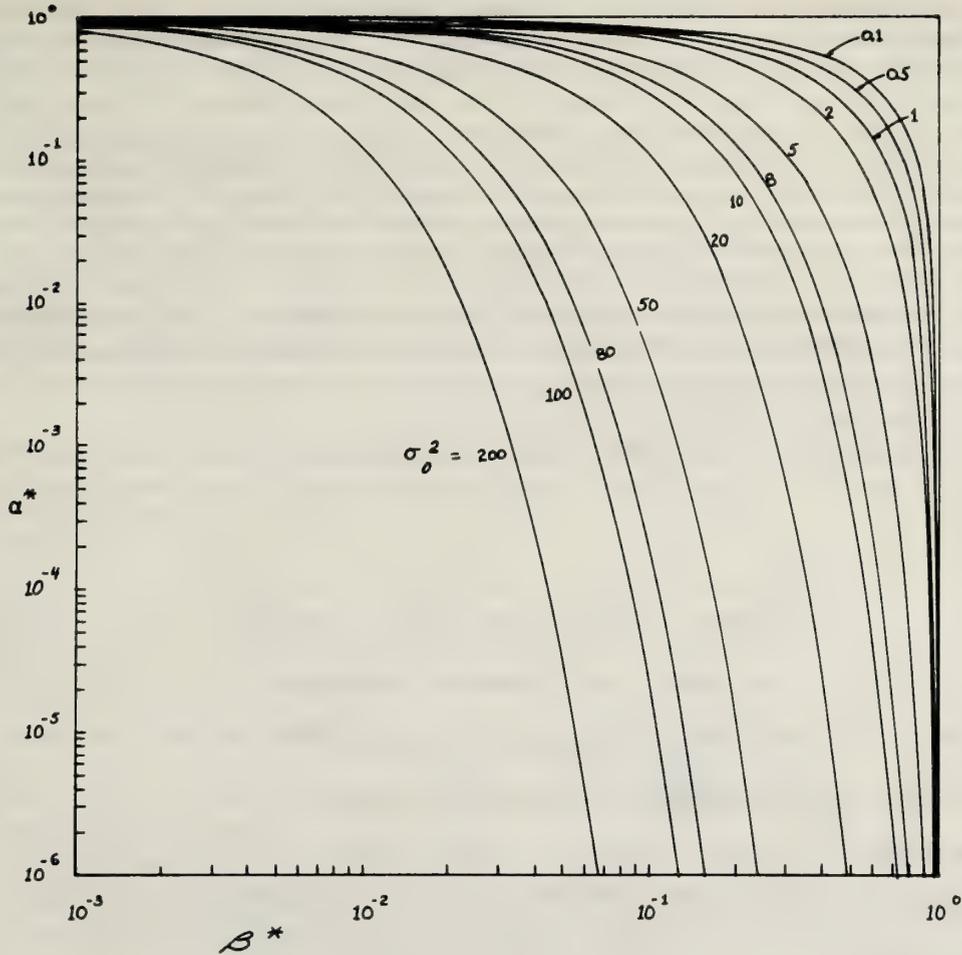


Figure 29.11 *Type I vs. Type II error probabilities; detection of a slowly fading target in reverberation or clutter.*

29.8.1 Example 1: Target Or Source Detection In A Scatter-Dominated Environment (Middleton and Viccione, 1970, 1973)

A problem of great interest in sonar and radar applications is the detection of a target in a reverberation or clutter-dominated environment. The analogous problem of detecting the presence of a source phenomenon, i.e., some locally distributed scatterer, by acoustical or electromagnetic "illumination" of it (and the medium), is also a frequent one in remote sensing studies of the environment, (cf., Chapters 8, 10-15, and 18, etc.). We summarize some recent results here, based on the canonical methods and models described in the preceding Sections.

Using an active system (see 29.1.3), we wish to detect during a period of T seconds, a comparatively distant (point) target in a sufficiently strong signal field that (acoustic) reverberation is the limiting interference. Narrow-band signals are used and detection is incoherent[†]. The (dimensionless) effective target cross-section, γ_T , is assumed to undergo (slow) Rayleigh fading (Middleton, 1965, Section 2.5). Our attention is confined to a single space-Doppler cell (target location and velocity are accordingly specified), and our task here is to detect the possible target in that cell. A high-density Poisson scatter process is assumed, so that the incremental acoustic pressure in the medium is a gaussian field (see end of 29.5.2), originating from a surface scatter mechanism, which dominates all volume effects in the chosen space-Doppler cell. Since the noise field here is signal-generated, noise and signal are comparable in intensity. The usual techniques for obtaining optimum (threshold) detector structures fail (Middleton, 1966a), and an "exact" treatment is required. Fortunately, this can be done for many useful cases. In particular, it is found that (Middleton and Viccione, 1972) optimum processors z^* , equivalent to but structurally simpler than $\mathbf{T}_{R_0}^* \{V\}$, are given by a functional $\psi_{OT}[V(t)]$ and expected performance is measured by a corresponding Bayes risk [see (29:14), (29:15), (29:18)], $R_{opt} = R^*(\sigma, \delta)$.

The decision process is a modified version of (29:12), with the threshold κ now replaced by an equivalent threshold K' , a function of κ (and known statistical parameters of the reverberation and the target) cf. (29:75).

We have specifically

$$H_0: \text{decide a target is not present if: } z^* = \Psi_{OT}[V(t)] \leq K' \quad (29:74a)$$

$$H_1: \text{decide a target is present if: } z^* = \Psi_{OT} > K', \quad (29:74b)$$

where the threshold K' here is determined from the original cost ratio κ , (Middleton, 1965), according to

$$K' = \frac{4\rho_0^2}{\langle \theta^2 \rangle} (1 + \sigma_{TR}^2) \log \left\{ \frac{\kappa}{\mu} (1 + \sigma_{TR}^2) \right\}, \quad \mu \equiv \rho/q, \quad (29:75)$$

with, specifically

$\theta \equiv a_0 \gamma_T :$	(normalized) illuminated target strength (amplitude);
$\rho_0 \equiv \bar{a}_0^2 \bar{\gamma}_R^2 \bar{\eta}_0 :$	total average scatter-to-(receiver) system noise power ratio;
$\sigma_{TR}^2 \equiv B_T^2 \bar{\gamma}_{TR}^2 / 2 :$	output "S/N"-cross-section ratio, of target to reverberation.
$\gamma_{TR} \equiv \gamma_T / \sqrt{\eta_0 \bar{\gamma}_R^2} :$	effective cross-section of point target divided by [average total mean-square scatter-cross-section of the reverberation] ^{1/2} ,
$\bar{\eta}_0 =$	average number of scatterers in domain of illumination Λ_S ,
$\gamma_T \gamma_R =$	(dimensionless) cross-sections or "reflectibilities" of point target and a typical scatterer.
$a_0 = A_0 / \sqrt{2\psi} :$	normalized (transmitted) signal amplitude; $\psi \equiv$ mean intensity of system noise in receiver, after the aperture; $A_0 =$ (peak) amplitude of transmitted signal.
$\sigma_0^2 = \rho_0 B_T + (\bar{a}_0 \bar{\eta}_0 \bar{\gamma}_R C_T)^2 :$	average output, total reverberation noise level, including a possible specular component ($\sim \bar{\gamma}_R$)
$\sigma^2 = \bar{a}_0^2 \bar{\gamma}_T^2 B_T^2 / 2\sigma_0^2 :$	(normalized) average output target (i.e., signal) to reverberation power ratio. (29:75a)

[†] This means that the RF epoch is unknown at the receiver, and is regarded as being uniformly distributed over a cycle of the RF carrier f_0 . [The RF epoch is a time, ϵ , relating the observer's "clock" time to the time-origin of the source, e.g., $S = \cos \omega_0(t - \epsilon)$, where $t = \epsilon$ is the origin of time as measured by the observer.]

The quantities B_T and C_T are *processing gains*, proportional to sample-size T , for the random and specular scatter fields, cf., (29:80); (29:81), below.

The Bayes error probabilities are found to be in this instance (Middleton and Viccione, 1970)

$$\begin{aligned} \alpha^* (= \beta_1^{(0)*}) &= \exp [-K'/2\sigma^2] ; \beta^* (= \beta_0^{(1)*}) = 1 - \exp [-K'/2\sigma^2(1 + \sigma_0^2)] \\ &= 1 - \exp^{(\log \alpha^*)/(1+\sigma_0^2)} \end{aligned} \quad (29:76)$$

Figure (29:11) shows α^* vs β^* for σ_0^2 as parameter. In monostatic operation (RT), a_0 is known at the receiver, so that $\bar{a}_0 = a_0$; $a_0^2 = a_0^2$, etc.; in bistatic operation ($R \neq T$) a_0 is not known, and accordingly, an appropriate average over signal level is assumed in θ , cf. (29:75a), here the Rayleigh pdf $w(\theta) = (2\theta/\bar{\theta}^2) \cdot \exp(-\theta^2/\bar{\theta}^2)$, $\theta \geq 0$. Finally, note that the Bayes average risk or cost here is given by (29:76) in (29:14), viz.,

$$R^* = (qC_{00} + pC_{11}) + \left(\frac{\kappa}{\mu}\right) \alpha^* + \beta^* p(C_{10} - C_{11}) . \quad (29:77)$$

Specifically, we can show that (Middleton and Viccione, 1970) the monotonic equivalent, z^* , of the likelihood ratio Λ_L , is

$$\Psi_{0T}[V(t)] = \left| \sqrt{\psi} \int_{-\infty}^{\infty} \left[V(t) - \langle X(t)_{\text{scat}} \rangle \right] \text{tr} Z_T(t) dt \right|^2 \quad (29:78)$$

where $Z_T [= 0; t \text{ outside } (0,T)]$ are a set of *Bayes matched filters* (Middleton, 1965, "Topics", Chapter 4) determined by the basic set of coupled integral equations

$$\int_{-\infty}^{\infty} K_X(t, u) Z_T(u) du = g_S(t)_T, \quad t \in T . \quad (29:79)$$

Note from (29:60), (29:61), typically, that the covariance is strongly conditioned by the geometry and waveforms of the overall system.

For monostatic ($R@T$) operation (the M vector) $\langle X \rangle$, and ($M \times M$ matrix) K_X are given explicitly by the complex version of, say, (29:46), and by (29:49), specialized to a scalar field ($Q_{RT}^{(m)} = \gamma_{\text{o-target}} A_{R-m} A_T$). The processing gains B_T and C_T determining the Bayes risk (29:77) are

$$B_T = \frac{\psi}{2} \text{Re} \int_{-\infty}^{\infty} \tilde{g}_S(t|T_0, \hat{a}_S) Z_T^*(t) dt \quad (29:80)$$

$$C_T = \frac{\psi}{2} \text{Re} \int_{-\infty}^{\infty} \tilde{g}_{\langle X \rangle}(t) Z_T^*(t) dt ,$$

with (\sim) indicating the transposed vector or matrix, and $\alpha_S (= \hat{r}_R/c = -\hat{r}_T/c)$ the delay per unit distance of the impinging wavefront of the signal from the target. Here $g_S = [F_{Sm}(t) e^{i\omega_0 t - i\phi_{Sm}(t)}] = (\sqrt{2}/A_0) \times$ complex version of (29:47), or (29:49), say, now with scalar beam patterns, and $R_R = R$, $T_0 = 2R/c = 2\lambda$ for the present monostatic operation. Similarly, $g_{\langle X \rangle} = [F_{\langle X \rangle m}(t) e^{i\omega_0 t - i\phi_{\langle X \rangle m}(t)}] = (\sqrt{2}/A_0) \langle X \rangle$, where now

$$\langle X \rangle = \frac{1}{16\pi^2} \int_{\Lambda_S} d\lambda \frac{\sigma_S(\lambda)}{\lambda} \langle Q_{RT}^{(m)}(\lambda) \rangle_{\gamma} \cdot \langle \mu e^{i\omega_0 \mu(t - T_0(\lambda)) - \hat{a}_T \cdot \mathbf{r}_m} \hat{S}_0(\mu[t - T_0(\lambda)]) \rangle_{\mu} \quad (29:81)$$

is a complex quantity, which vanishes if the scattering surface is sufficiently rough, e.g., $\bar{\gamma}_R \rightarrow 0$. Here again, A_0 is the peak amplitude of the original driving signal S_{in} , and ψ throughout is the mean intensity of receiver system noise, used as a scale for the reverberation, cf. (29:75a). The quantities F_S , ϕ_S are respectively the envelope and phase of the narrow-band transmitted signal

$$S(t) = a_0 F_S(t) \cos [\omega_0 t - \phi_S(t)] . \quad (29:82)$$

In these monostatic cases $\langle \mathbf{X} \rangle$ and C_T always vanish, since coherent scatter from the surface occurs only in the forward, or specular direction, for the usual beam patterns and emitted signals. The results (29:78)-(29:80) are also directly applicable to the bistatic cases, with appropriate modifications of $\langle \mathbf{X} \rangle$, \mathbf{K}_X , and (29:61) to this situation (Middleton, 1967, and 1972).

Equation (29:78) can be interpreted in a variety of ways to represent different equivalent optimum structures having the general form of M Bayes matched filters, followed by identical multipliers and appropriate adders (Middleton and Groginsky, 1965; and Middleton, 1965, Chapter 4), to yield the test statistic ψ_{OT} . Figure (29.12) shows a typical structure for the optimum detector here. Finally, when a preformed beam is used in the receiver, we have only to set $M = 1$ in the above: \mathbf{V} , $\langle \mathbf{X} \rangle$, \mathbf{Z} , etc., become simple scalars. In all cases the explicit form of the (complex) matched filters \mathbf{Z}_T (i.e., $\hat{h}_x + ih_y$) is determined by the solution of (29:79), which can present major computational problems (Kagiyada, et al., I, II, 1968; Baggeror, 1970; Van Trees, 1968).

29.8.2 Example 2: Amplitude, Range, And Bearing Estimation: Passive Reception: (Kelley and Levin, 1964; and Middleton 1966):

Instead of an active system, in this example we employ a receiver only and assume that interference is a dominantly gaussian ambient noise field (see 29.3.2 and 29.5.3). Let us assume that unconditional maximum likelihood estimators are sought (29.1.4) for the unknown parameters θ [= amplitude, range, and bearing (θ, ϕ)] of a distant (point) source. The optimum M -channel receiver $\mathbf{T}_{R_c}^* \{ \mathbf{V} \}$ for this purpose is found to have the form $\chi[\underline{\theta} | \mathbf{V}(t)] = F(\Psi_{OT}, \mathbf{B}_T)$, where F depends on the a priori distribution postulated for the unknown parameters $\underline{\theta}$, and Ψ_{OT} . \mathbf{B}_T have the same structures as in our detection example (29:78)-(29:80), with \mathbf{X}_{scat} , etc., replaced by \mathbf{X}_{self} , (29:31) etc. Figure (29.12) again applies, with an additional computation representing $F(\Psi_{OT})$, the (joint) estimates of source amplitude ($\sim a_0$), range ($\sim T_0$), and bearing (θ, ϕ) . Without going into further detail, we see again how the channel geometry and physics specifically control structure and performance through Ψ_{OT} and \mathbf{B}_T .

29.8.3 Further Optimization And Future Problems: Waveforms and Apertures (29.1.5)

Two major current problems in detection and extraction theory are waveform selection ("cost coding", Middleton, 1960, Section 23.2) and the sampling, reconstruction, and extrapolation of data fields (Petersen and Middleton, 1962; and Petersen and Middleton, 1965). In the former, we seek a waveform \mathbf{S}_{in} or a "modulation" $\mathbf{T}_{T_0} \{ \mathbf{S}_{in} \}$, which, subject to various constraints (usually on bandwidth and average or peak power) will further optimize the link, represented by (29:1). In the latter, we wish to optimize data acquisition by choosing optimal array or aperture configurations, \mathbf{T}_{AR} and \mathbf{T}_{AT} for both transmission and reception, again with appropriate constraints (usually on size, number of elements, and channels). In each instance the accompanying optimum receiver is also required. Combination of both types of optimization is a third possibility (Middleton, 1971).

In any case, the first step toward solution demands results like those of the examples above for system structure and performance which include channel physics and geometry explicitly, and the measurement of channel properties, e.g., H'_{scat} and H'_{self} , etc. (Middleton, 1967 and 1972; Plemons, 1971; and Bello, 1969). An approach to the optimal array problem is referred to in 29.9 below, where we are concerned with the basic problem of data acquisition: in effect, sampling the received field by elements of a receiving array, in such a way as to obtain estimates of minimal error.

29.9 Space-Time Sampling Theory (Petersen and Middleton, 1963 and 1965; Petersen and Middleton, book in preparation, 1973 est.)

As the reader can see from the discussion in the previous sections, one central task is that of data sampling. This arises in all our general communications problems, where information transfer from point-to-point in space-time is the desideratum, be this a measurement task, i.e., interrogation of the medium

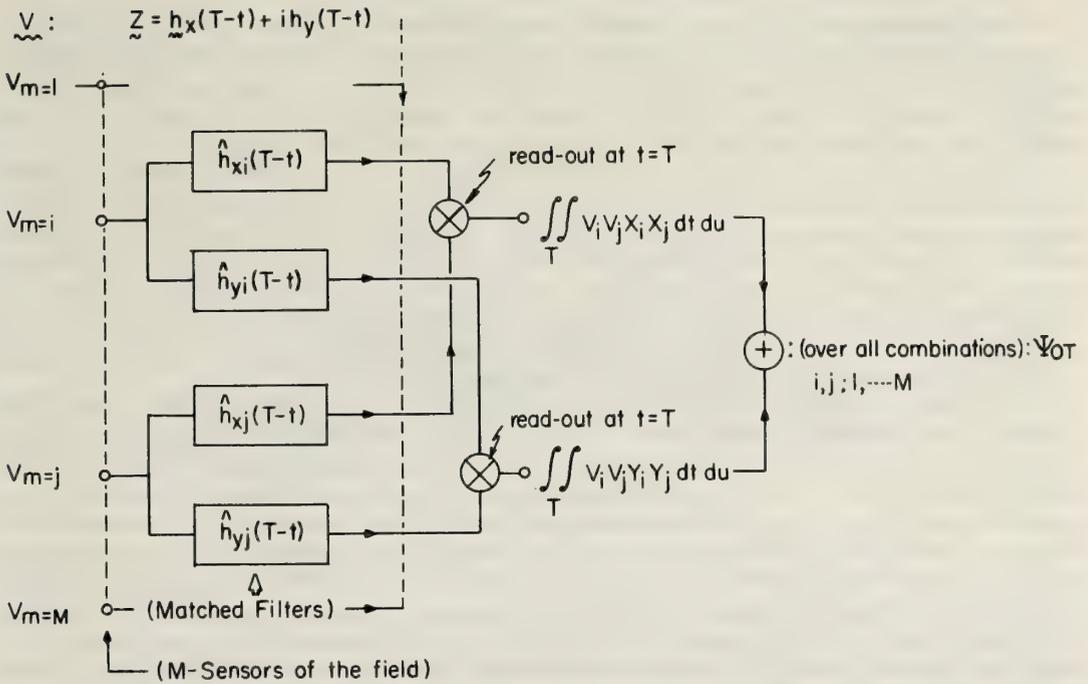


Figure 29.12 Scheme of optimum processor for target detection in reverberation using Bayes matched filters of the 2nd kind, type 3.

itself ("remote sensing"), or the transmission and reception of desired signals in undesired, i.e., "noise" backgrounds—self-generated or ambient. Where to place our sensors and how often to require their sensing of the medium are fundamental questions, which often have critical economic implications as well as scientific ones. Immediate applications, for example, are in weather prediction, oceanography, acoustics, seismic monitoring, and various other types of surveillance.

In this Section we shall give only a few summary remarks regarding the aims and results of a multidimensional sampling theory, as an introduction to the reader's own further pursuit of the literature (Petersen and Middleton, 1962, 1963 and 1965; Petersen and Middleton, book in preparation (1973 est.); and Petersen, 1963), and application[†] of these concepts.

Specifically, the Petersen-Middleton approach develops:

- (1) A theory of multidimensional sampling, using a delta-modulation technique, thereby sharpening the statement of the general sampling theorem and providing a clearer insight into the general sampling problem itself;
- (2) The general procedure for achieving a minimum sampling lattice for wave number limited stochastic processes;
- (3) Statements of critical properties of sampling lattices and reconstruction functions;
- (4) Sampling of isotropic functions;
- (5) Extensions to wave-number-limited stochastic processes and to optimum prefiltering and reconstruction of nonwave-number-limited processes; and
- (6) A discussion of the implications of the theory in measurement situations.

[†] For some meteorological and optical applications, see Petersen and Middleton, 1963.

29.9.1 Summary

The multidimensional sampling theorem (Petersen and Middleton, 1962) is found to be readily adaptable to the interpolation of homogeneous multidimensional stochastic processes. The general optimum post-sampling filter for nonwave-number-limited processes which yields a least mean-square error is incidentally derived, in the course of proving that wavenumber-limited processes can be reproduced with zero mean-square error from samples taken over a lattice with suitably small repetition vectors.

The problem of optimum prefiltering of multidimensional data, for sampling on a lattice of specified density, has been examined. It is found that the optimum operation (yielding minimum mean-square error *uniformly averaged* within a cell corresponding to the given sampling lattice consists of ideal bandlimiting to the cell of the wavenumber lattice enclosing the greatest spectral power. This relationship is the ultimate justification for the study of admittedly artificial “band-limited” processes. It permits the establishment of lower and upper bounds on the average mean-square error of *optimally reconstructed* sampled data. These are respectively, one and two times the total spectral intensity outside the most favorable wave-number cell within the constraint of repeatability corresponding to the given sampling lattice.

We remark on the key role played by the covariance, \mathbf{K} , of the random process which we are sampling in space and time. Very often we may know the form of this covariance function from the detailed physical model, (cf. 29.6 and 29.8.1, 2.3), or we may measure it (under such data controls as are indicated in 29.7). Here, of course, we need to include the spatial correlation as well, which can be obtained by suitable location of the M -sensors of our array or lattice sampling plan, along the lines suggested above. Knowledge of the Langevin equation (29.5.1), of the random process under study provides additional information, vis-à-vis an observed covariance function alone, and is equivalent to those cases above where we can specify the detailed physical model, and thus the explicit structure of \mathbf{K} . Furthermore, knowing the Langevin equations enables us to establish constraints on the estimates of covariance and spectral functions (Petersen and Truske, 1971). Finally, the Langevin equation, plus initial conditions and the (theoretical) covariance \mathbf{K} , give us the complete structure for a (linear) second-order sampling optimization, with minimization of mean-square-error (i.e., with QCF, cf. (29:17), et seq). These ideas have been developed in detail, in connection with multidimensional ($M \geq 1$) sensing or sampling of a data field. [Petersen and Middleton, (1973 est.), and Petersen 1963].

Finally, it is clear that the applications of multidimensional sampling theory are manifold. The above is merely an introductory remark on such central processing questions as the rôle of noise, limited data spaces, performance criteria other than mean-square and uniform averaging, vector processes, and the relationship of time and displacement correlation to the applicable dynamic equations. Moreover, specific *system* design applications, such as those embodied generally in the decision-theoretic formulation of communication theory, briefly outlined in 29.1, are needed in fields such as meteorology, oceanography, and optics – where constraints and tradeoffs involving filter realizability, lattice density in space and time, and sequential scanning, data transmission and processing, must be introduced.

29.10 Concluding Remarks

In this chapter we have presented what is in effect a very detailed outline of a canonical approach to the joint problem of channel modeling, data acquisition and validation, and system optimization. The key feature of such an approach is the basic method, which is not tied down to any one system or local philosophy of data handling. Rather, it seeks to provide a general framework wherein both the measurement problems of principal concern in this volume and the communication goals of signal detection and extraction may be recognized as aspects of a “global” communication theory, with techniques and methods independent of specific constraints and limited aims: in short, a canonical approach. The essential ingredients of this philosophy, are:

- (1) jointly related cost and evaluation procedures, e.g., decisions-theoretical ideas harnessed to the concepts of information transfer-retrieval;
- (2) incorporation of the controlling system and channel physics, which include the strongly influential geometrical factors (beam patterns, source and receiver locations, distributions of scatterers and sources, etc.);
- (3) statistical modeling of the random mechanisms in the system and channel, and space-time sampling planning for data acquisition; and
- (4) statistical evaluation of the acquired data, with emphasis on ensemble validation. A glance at 29.0.2 will show in summary how these notions have been developed here.

The principal task ahead of us is the detailed exploitation of the preceding approach for specific problems in different physical areas. Certainly, the statistical-physical anatomy of the channel in both studies of the medium and system performance needs much further development in many cases. Technically improved analytical results for the statistical description of the various nonnormal (i.e., nongaussian) noise and signal fields and waveforms are also required, for the analysis and prediction of system performance and the interpretation of measurements. Experimental techniques which recognize the ensemble nature of an adequate description of the phenomena in question when these phenomenon are probabilistic or exhibit, at least, a significant random component (e.g., ocean reverberation, for example cf. 29.6.2), must be practically (i.e., economically) integrated with appropriate computer processing.

Finally, a very important question of increasing practical concern is that of data acquisition or sampling, referred to in 29.9, where the space-time scanning of a data field is to be undertaken in some rational fashion, which includes economic as well as physical constraints. Spatial or space-time sampling, of course, is a form of data processing, and part of the decision-making sequence, and is to be included in the common framework of statistical communication theory, in the large. What has been presented here in Chapter 29, is a somewhat detailed, though incomplete, indication of how we may expect to proceed along these lines, with a maximum utilization of the inter- and multi-disciplinary techniques of contemporary Statistical Communication Theory.

29.11 References

- Baggeror, A. B., 1970: *State Variables and Communication Theory*, Res. Monograph, No. 61, MIT Press, Cambridge, Mass.
- Bello, P. A. (1969), Measurement of random time-variant linear channels, *IEEE Trans. Inform. Theory*, *IT-15*, 469-475, July.
- Kagiwada, I. H., R. Kalaba, and B. Vereeke (1968), The invariant embedding numerical method for Fredholm integral equations with degenerate kernels, *J. Approximation Theory*, *1*, 355-364.
- Kagiwada, I. H., R. Kalaba, and B. Vereeke (1968), Verification of the invariant embedding method for certain Fredholm integral equations, *J. Mathematical Analysis and Applications*, *23*, 540-550.
- Karp, S., R. M. Gagliardi, and I. S. Reed (1968), Radiation models using discrete radiator ensembles, *Proc. IEEE*, *56*, October, 1704-1711.
- Kelly, E. J., Jr. and M. J. Levin (1964), Signal parameter estimation for seismometer arrays, M.I.T. Lincoln Laboratory, Lexington, Mass.; Report 339, Jan. 8.
- Middleton, D., 1960: *Introduction to Statistical Communication Theory*. McGraw-Hill, New York. Part IV.
- Middleton, D., 1965: *Topics in Communication Theory*. McGraw-Hill, New York.
- Middleton, D. (1966a), Optimum and suboptimum bearing estimation for deterministic and random signals in a normal noise field, Submarine Signal Lab., Newport, R. I., Raytheon Research Report R452, October (also, references therein).
- Middleton, D. (1966b), Canonically optimum threshold detection, *IEEE Trans. Info. Theory*, *IT-12*, April, 230-243.
- Middleton, D. (1967), A statistical theory of reverberation and similar first-order scattered fields. Parts I, II; *IEEE Trans. Information Theory*, *IT-13*, No. 3, July, 372-414.
- Middleton, D. (1969), Acoustic Modeling, Simulation, and Analysis of Complex Underwater Targets, II. Statistical Evaluation of Experimental Data, June 26: Report ARL-TR-69-22 (Applied Research Laboratories, Univ. of Texas, Austin; Nav. Ships Sys. Comnd. Contract N00024-69C-1129).

- Middleton, D. (1970), Multidimensional detection and extraction of signals in random media, Proc. IEEE, 58, No. 5, May, 696-706.
- Middleton, D. (1971), Conceptual Approaches in the Design of Optimal Radar and Sonar Systems: II. Formulation of Statistical Scatter Models in the Electromagnetic Case., RAND Corp. Report R-707-PR, May.
- Middleton, D. (1972a), A statistical theory of reverberation and similar first-order scattered fields. Part III, IV; IEEE Trans. Information Theory, *IT-18*, No. 1, January, 35-90.
- Middleton, D. (1972b), Statistical-physical models of urban radio-noise environments: I. Foundations; IEEE Trans. on Electromagnetic Compatibility, *EMC-14*, May.
- Middleton, D., and H. L. Groginsky (1965), Detection of Random Acoustic Signals by Receivers with Distributed Elements: I. Optimum Receiver Structures for Normal Signal and Noise Fields, J. Acoust. Soc. Am., 38, November, 727-737.
- Middleton, D., and D. M. Viccione (1970, 1973), Target detection in a reverberation-dominated environment, in preparation for J. Acous. Soc. Amer., (vide. D. M. Viccione, "Detection of Underwater Signals in Reverberation Fields with Non-vanishing Velocity Gradients," Doctoral Dissertation, November 1970, Dept. of Electrical Engineering, Univ. of Rhode Island, Kingston, Rhode Island, 02881; also Raytheon Report, Submarine Signal Div., Portsmouth, Rhode Island, 1970.)
- Page, L., 1935: *Introduction to Theoretical Physics*. Van Nostrand, New York. Sections 47-52.
- Petersen, D. P., 1963: *Sampling of Space-Time Stochastic Processes with Application to Information and Decision Systems*, Dr. of Engineering Sciences Dissertation, Rensselaer Polytechnic Inst. Troy, New York.
- Petersen, D. P., and D. Middleton (1962), Sampling and reconstruction of wave-number limited functions in N-dimensional Euclidean spaces, *Information and Control*, 5, 279-323.
- Petersen, D. P., and D. Middleton (1963), On representative observations, *Tellus*, 15, 307-405.
- Petersen, D. P., and D. Middleton (1965), Linear interpolation, extrapolation, and prediction of random space-time fields with a limited domain of measurement, *IEEE Trans. Information Theory*, *IT-11*, January 18-30.
- Petersen, D. P., and Middleton, 1973 [est.]. *Sampling and Processing of Space-Time Fields*, book in preparation. MIT Press, Cambridge, Massachusetts.
- Petersen, D. P., and T. N. Truske, "A Study of Data Processing Techniques for Satellite Meteorological Observations," Final Report, Contract 1-35161 for National Environmental Satellite Service (NOAA), U. of New Mexico, Bureau of Engineering Research, Sept. 1971.
- Plemons, T. D., J. A. Shooter, and D. Middleton (1971), Underwater acoustic scattering from lake surfaces, I, II. Submitted November 1971 to JASA. See, also, T. D. Plemons, "Spectra, Covariance Functions, and Associated Statistics of Underwater Acoustic Scattering from Lake Surfaces," Doctoral Dissertation, May, Phys. Dept., Univ. of Texas, Austin, Texas.
- Siegel, S., 1956: *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- Stratton, J. A., 1941: *Electromagnetic Theory*. McGraw-Hill, New York. Chapter 8, Section 8.
- Van Trees, H. L., 1968, 1971: *Detection, Estimation, and Modulation Theory*, Parts I-II, John Wiley & Sons, New York.

List of Symbols

a_T, A_T	aperture weighting	L, L	fields
A_T	beam pattern	$\ell_0, \ell_0 s$	path length ($\nabla c \neq 0$)
A', A'	generalized beam patterns	$\lambda, \underline{\lambda}$	coordinates
a	absorption factor	μ	a Doppler factor
$\beta(\ell_0 \mathbf{R}), \beta(\ell_0 s \mathbf{R})$	ray focus-defocus factors	M_1, M_2	sample sizes
$\beta_1^{(0)}, \beta_0^{(1)}$	conditional error probs	$N(t), N, N_{\text{self}}, N_{\text{system}}$	noise waveforms
B_T	processing gain	ν	vector spatial frequency
$C(\mathbf{S}, \gamma)$	cost function	ω	angular frequency
C, C_{ij}	cost matrix; cost elements	ω_0	angular carrier frequency
c	velocity of propagation	W_L	pdf
C_T	processing gain	$p(t, \mathbf{R})$	pressure field
$\delta(\gamma \mathbf{V}), \delta(\gamma_0 \mathbf{V}), \delta(\gamma_1 \mathbf{V})$	decision rules	Q_{RT}	coupled cross section
$\nabla \cdot \nabla$	Laplacian	q_{RT}	coupled cross section
$\delta(\mathbf{x} - \mathbf{o})$	delta function	$R(\sigma, \delta)$	average cost or risk
∇	gradient operator	R^*	Bayes risk
$\nabla_{\mathbf{x}}$	curl operator	\mathbf{R}, \mathbf{R}	distance, distance vector
$\underline{\eta}, \underline{\eta}_0, \underline{\eta}_{0m}$	vector to receiving element $d\eta$	$\rho(\underline{\lambda})$	Poisson density
\mathbf{E}	electric field vector	\mathbf{r}	a vector
$F_L(\mathbf{V} \mathbf{S})$	a priori pdf of data	S_{in}	input signal
f	frequency	\mathbf{S}	signal vector
f_0	carrier frequency	SCT	statistical communication theory
F_L	characteristic function	SDA	statistical data analysis
G_T	source function	$\sigma(\mathbf{S})$	a priori pdf of signal
γ	decision vector	\hat{S}_0	complex envelope of n.b. signal
$\gamma_0, \hat{\gamma}_0$	scatter cross section	σ_S, σ_V	source densities
H', \underline{H}'	ambient source functions	σ_R	a variance
\mathbf{H}	magnetic field vector	$T_R, T_M, T_T,$ $T_{AR}, T_{AT}, T_{R_0}, T_{T_0},$ $T_{RMT}, \text{etc.}$	operators
$\hat{\mathbf{i}}_x, \hat{\mathbf{i}}_y, \hat{\mathbf{i}}_\rho, \hat{\mathbf{i}}_T, \hat{\mathbf{i}}_R$	unit vectors	τ_0	path delay ($\nabla c \neq 0$)
j	j th source or scatter	T_0	path delay
κ, K'	decision thresholds	t	time
$\underline{\xi}$	vector to radiating elements	$\{u\}$	transmitted message set
K_X, K_X, \bar{K}_0	covariance functions		

List of Symbols

U	basic received scattered, or ambient waveform
$\{v\}$	received message set
V	data vector
$\langle X \rangle$	mean value
$X(t), X_{\text{scat}}, X_{\text{self}}$	(total) received random process from ambient or scattered fields
Y_M	bifrequency function
F_{sample}	test statistic

CHAPTER 30 STATUS OF REMOTE SENSING OF THE TROPOSPHERE

C. Gordon Little

Wave Propagation Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration

This chapter attempts to summarize the status of remote sensing of the troposphere. It opens with a statement of the need for remote sensing and identifies the potential advantages inherent to such techniques. The main material of the chapter is in the form of a series of tables which identify the stage of development of each method (acoustic, radio or optical) for the measurement of the different meteorological parameters relevant to the troposphere. We differentiate between those capabilities which as yet are limited to the boundary layer, and those already applicable to the troposphere; and also between those techniques which are limited to clear air as opposed to those which can operate in the presence of fog or precipitation.

30.1 The Role of Remote Sensing in the Troposphere

As the size and complexity of our society increases, so the number and complexity of its interactions with our atmospheric environment grows. As a result, the economic costs to commerce, transport, communications and agriculture of natural disasters such as blizzards, ice storms, hurricanes, tornadoes, floods, droughts, etc., increase year by year. This growing sensitivity to weather conditions has, of course, led to a greatly expanded need for atmospheric information and forecasts of many different types, for many different purposes.

The attainable quality and scope of an environmental information or forecasting service is determined by the nature of the observational data set on which it is based — and especially by the density in time and space of the relevant observations. Thus, the present meteorological observing system of ground-released radiosondes and satellite measurements is well matched to synoptic scale weather patterns, i.e., to space scales of the order 1000 km and time scales (at one location) of the order one or two days. But recent analyses of national needs for improved weather services show that these needs lie primarily in the area of short-term *local* weather forecasts, e.g., forecasts for a local area 1–100 km in size, for time periods ranging from 0 to 6 hours. These improved short-term local weather forecasts inevitably will require a much denser array of meteorological data than is currently available from the present observational networks. The required increase in space-time density of relevant observations is huge — at least four orders of magnitude relative to the twice per day, 500 km horizontal spacing of the upper air radiosonde observing system. Cost estimates indicate that it is totally impracticable to achieve this desired increase in density of observations by mere expansion of the existing radiosonde network. Satellite remote sensing of meteorological conditions is, of course, already playing a very important role; however, because of the great need for improved resolution in height, time, and space, especially in the boundary layer of the atmosphere and under conditions of cloud, satellites can only provide a very incomplete answer. Obviously, some new breakthrough in weather observing is required; it is the thesis of the proponents of ground-based remote sensing that these techniques offer such a breakthrough. Moreover, they claim that this breakthrough will have great impact on the nature and conduct of the atmospheric research and services of the future.

The potential advantages of remote sensing are several and may be listed as follows:

- a. The observations are taken remotely, i.e., without requiring that in situ instrumentation be carried to the region of atmosphere to be measured.
- b. Ideally, remote sensing permits the measurement of the relevant parameters of the atmosphere in 1, 2, or 3 spatial dimensions, all as a function of time.
- c. Excellent space and time resolution is often obtainable.
- d. An increasingly broad range of parameters of the atmosphere may be sensed. While a standard radiosonde is limited to measurements of temperature, humidity, and wind, parameters such as spectrum of turbulence or momentum flux can be measured remotely, e.g., by Doppler radars.
- e. The measurement system does not modify the medium being measured. Under some situations this can be of serious concern, as in aircraft or meteorological tower measurements.

f. The remote sensing measurements typically provide a line, area or volume integration which helps to ensure that the observations are more representative of the medium than those of a single point sensor, which may be adversely affected by local or transient perturbations.

g. Remote sensing instrumentation is usually automatic and often can provide fully-processed data 24 hours per day with a minimum of manpower.

It should, of course, be noted that the above points are potential advantages, which are generally applicable to remote sensing as a whole, but in most cases have yet to be fully realized.

30.2 Categorization of the Status of Remote Sensing

In attempting to describe the status of remote sensing of the troposphere, it is necessary to describe the progress in terms of several different aspects of remote sensing. First, the range of parameters to be measured is large. Thus, a reasonably full knowledge of the condition of the atmosphere around the observing site requires information on the 3-dimensional distribution of six parameters, all as a function of time. These parameters are

Wind
Temperature
Humidity
Precipitation (hydrometeors)
Aerosol
Gaseous Pollutants

For each of these parameters different types of measurements may be required ranging from the relatively simple average value of the parameter along some line to the full 4-dimensional distribution of the parameter in time and space. These different types of measurements form a hierarchy of increasing information content (and usually of difficulty in acquisition) as follows:

line integral – the integrated value of the parameter along a line through the whole atmosphere

line average – the average value of the parameter along some line of known length

line profile – the distribution of the parameter along some line of known length

2-dimensional coverage – the distribution of the parameter over a plane

3-dimensional coverage – the distribution of the parameter in space around the instrument

structure constant – a measure of the intensity of small-scale fluctuations of the parameter in space and time

spectrum – the power spatial spectrum of variability of the parameter in space – i.e., how strong are the variations of the parameter as a function of the spatial size of the irregularity

flux – the rate at which mass, momentum or heat is being transported (usually in the vertical direction).

Remote sensing is based on measurements of the interaction of waves with the medium of study. In remote sensing of atmospheric conditions, it has been found important to make use of the unique advantages peculiar to acoustic, radio, and optical waves, and to recognize that passive remote sensing systems (using waves of natural origin) as well as active systems (using man-made waves) each have their own particular advantages.

In general, the development of a remote sensing concept can be seen to follow a logical sequence. In Step A, the concept is identified, and preliminary first-order estimates made of its feasibility. In Step B, the potential capabilities and limitations of the concept are analyzed theoretically in considerable detail. If the concept still appears attractive, the development of a research equipment for the experimental evaluation of these capabilities and limitations takes place in Step C. Assuming that this quantitative experimental evaluation of the concept is successful, the next stage (Step D) is to build a development model (as opposed to a research model) which is thought of as a prototype of an operational unit which is to be capable of being used in the field by research workers or technicians other than the original research group. If the concept continues to show promise, Step E involves working with industry to obtain commercially built units for field evaluation. Once this stage has been successfully completed, the final stage (Step F) requires that fully evaluated commercial units be routinely available for procurement.

The status of the remote sensing of the troposphere is now discussed in terms of these stages, A-F, of development of acoustic, radio and optical remote sensing techniques for the different types of measurement

of the six meteorological parameters listed above.

30.3 Acoustic Remote Sensing of the Troposphere

Table (30.1) summarizes the status of acoustic remote sensing of the boundary layer under clear air conditions, using letters to denote the different stages of development described in the previous section. The footnotes below (T30.1) (and succeeding tables) refer to the superscript numbers in the table, and indicate the nature of the equipment responsible for each specific measurement category. Where possible, a recent reference is given to the work on which the entry was based.

The status of acoustic remote sensing of the troposphere (as opposed to the boundary layer) is shown in (T30.2) for clear air conditions. As of this time, acoustic echo techniques are not being actively applied to tropospheric remote sensing. However, as discussed in Chapter 19, the prospects for acoustic echo sounding of the troposphere are good.

The status of acoustic echo sounding under precipitation or fog conditions is summarized for the boundary layer in (T30.3) and for the troposphere in (T30.4). Present experience with the operation of acoustic echo sounders under conditions of fog, rain or snow is very limited. Analysis by Little (1972a) indicates that hydrometeor echoes should be readily obtainable, (at least in the lowest few hundred meters) provided frequencies of the order several thousand Hz are used. It should therefore be possible to derive some information on the distribution, and fall velocity, of hydrometeors in the lower regions of the boundary layer, using such a system.

Because of the λ^{-4} dependence of the hydrometeor scattering cross section upon wavelength, the effects of the hydrometeor echoes can be greatly reduced by using low frequencies. Relatively little is known experimentally of any increase of acoustic absorption due to the presence of hydrometeors, though the theoretical work of Cole and Dobbins (1970, 1971) suggests that this will not be significant for rain or fog conditions. However, major uncertainties exist as to the increase in noise level during rainfall conditions, due to precipitation impacting the antenna. Present information suggests that cloud, fog and drizzle conditions will have little effect upon standard boundary-layer and tropospheric acoustic echo sounders though snow conditions may enhance the acoustic absorption, and rain could seriously enhance the ambient noise level on the receiving antenna. Parentheses are therefore used in (T30.3 and T30.4) to denote that acoustic echo-sounding systems will experience serious degradation or failure during rain.

30.4 Radio Remote Sensing of the Troposphere

In general, radio waves interact more weakly with the clear atmosphere than optical waves or acoustic waves. Nevertheless, both passive and active radio systems have been used. The passive systems rely on the absorption (and therefore emission) of radiowaves by water vapor at about 1.35 cm wavelength, and by oxygen at 5-6 mm wavelength. Such systems are capable of measuring integrated water vapor and profiles of temperature respectively. The active systems include line-of-sight and forward scatter systems, as well as radar systems. The backscattering radar cross section due to clear air is usually very low, and high sensitivity radars (either high power pulsed systems, or FM-CW systems) are required. Even these systems do not give continuous echoes from all volumes, but detect only the regions of most intense fluctuations in radio-frequency refractivity. Alternatively, artificial targets (chaff) may be introduced into the atmosphere; more rarely, the density of small insects may be sufficiently high that their echoes can be used as tracers of the air motion.

Tables (30.5) and (30.6) give the status of radio remote sensing of the clear atmosphere in the boundary layer and troposphere respectively, and indicate in parentheses those radio remote sensing capabilities which are dependent upon the presence of chaff. Also included in (T30.5) is a hybrid acoustic-radiowave technique, variously designated EMAC (electromagnetic-acoustic) or RASS (radio-acoustic sensing system), which may be used to derive temperature and/or wind profiles. In these studies, a powerful acoustic pulse is tracked by a radar system, which obtains echoes because of the spatial modulation of radio frequency refractive index created by the pressure oscillations of the acoustic wave. As implied in (T30.5 – T30.7) radio waves are essentially unaffected by aerosol and gaseous pollutants at the densities normally found in the atmosphere.

The abrupt change in radio frequency refractive index represented by a hydrometeor is enormously larger than the refractivity fluctuations present in clear air, and hence appreciable radio power is scattered,

especially for wavelengths comparable in size with the drop. (At much longer wavelengths, the scattering is weak and can be ignored.) For many years, microwave radar echoes from precipitation have been used to monitor the location and intensity of precipitation. In the past decade, major efforts have been made to use Doppler techniques to determine the velocities of the precipitation, and hence of the horizontal velocity field of the air. The status of this work is shown in (T30.7), which covers both boundary layer and troposphere studies of radio remote sensing under conditions of precipitation. This table also covers the use of passive microwave radiometric techniques to derive integrated liquid water content along the beam, using the 5-10 GHz radio waves emitted by the water drops.

30.5 Optical Remote Sensing of the Troposphere

The interaction of optical waves with the constituents of the atmosphere is extraordinarily rich and varied, and we may therefore expect that in the long run optical techniques will eventually become dominant in remote sensing of the clear atmosphere. While passive optical remote sensing systems will undoubtedly be used, the principal progress is likely to come with the development of the field of "Lidar Spectroscopy." This field combines the remote sensing advantages of radar with the chemical specificity of spectroscopy, and the coherent properties of lasers, to permit (in concept at least) the remote interrogation of small volumes of the atmosphere for specific constituents. The potential advantages of lidar spectroscopy, relative to radio or acoustic techniques, may be listed as:

- excellent angular resolution
- excellent range resolution
- excellent Doppler resolution
- excellent chemical specificity
- very fast informate rate
- broad range of parameters conceptually available.

This latter point is well illustrated in (T30.8), which indicates the status of optical remote sensing of the (clear) boundary layer. A very full range of remote sensing measurements is *conceptually* practicable though as of this time relatively few optical measurement techniques have been tested experimentally. Currently, the problem is primarily technological. Ideally, one would like to be able to transmit short (e.g., 10 nano-second) pulses, at very high power (e.g., 10^{19} photons/pulse), at high pulse repetition rates (e.g., 10^4 pulses per second) from highly directive telescopes (beam divergence $\leq 10^{-4}$ radian) using tunable laser systems capable of operating in the U-V, optical and near infrared. While these parameters can be met or exceeded separately, they cannot yet be attained simultaneously in the same system. Nevertheless, laser technology continues to make rapid progress. It should also be emphasized that much of the detailed spectroscopic information needed for remote sensing purposes is not available, and will have to be obtained if the field is to advance efficiently.

The possibility of optical Doppler studies of motions in the clear boundary layer is obviously important, and is at the heart of many of the velocity or flux measuring concepts listed in (T30.8). At the moment, the likelihood of optical Doppler measurements of clear air at distances of several kilometers seems remote, and hence (T30.9), which gives the status of optical remote sensing in the troposphere, is much less complete than for the corresponding boundary layer (T30.8).

30.6 Comparison of Status of Acoustic, Radio and Optical Remote Sensing of the Troposphere

Tables (30.10) through (30.13) compare the status of acoustic, radio and optical remote sensing of the troposphere. As before, these tables use the letters A, B, C, etc., to denote the stage of development of a measurement capability, with suffixes 1, 2, or 3 to denote whether the technique is acoustic, radio or optical. Thus, the entry $C^{1,2} B^3$ under Line Profile of longitudinal component of wind indicates that experimental tests (Stage C) of acoustic (1) and radio (2) methods of measuring the line profile of longitudinal velocity (i.e., component along the line) have been made, whereas an optical technique (3) has only reached Stage B, i.e., detailed theoretical analysis. As in previous table sets, (T30.10 and T30.11) deal respectively with the boundary layer and troposphere under clear air conditions; (T30.12 and T30.13) cover precipitation conditions. Tables (30.10) through (30.13) include the information contained in (T30.1) through (T30.9); the footnotes of these earlier tables may be used to identify the techniques under consideration and (where available) an appropriate reference.

Table (30.10) shows that a very full measurement capability (at least in terms of the eight types of measurement of the six parameters considered here) is being investigated for the clear boundary layer. The corresponding troposphere table, (T30.11), is much less completely filled, with coverage under precipitation conditions much less complete, though microwave Doppler radars provide excellent capabilities for boundary layer and troposphere wind fields and precipitation fields.

30.7 Estimates of Accuracy of Remote Sensing Measurements

The specification of accuracy of a remote sensing system involves very many qualifications (e.g., at what range; with what integration time; under what meteorological conditions; with what background noise level; etc., etc.). A further problem is that the accuracy of remote sensing instrumentation is often not well known, primarily because of the difficulty of obtaining data from an appropriate network of accurate, in situ, sensors. Also, there is the question of what do we mean by accuracy? For example, absolute accuracy, or relative accuracy, or precision? Despite these problems, (T30.14 – T30.17) have been prepared in order to give the reader some feeling for the accuracy believed available from the respective remote sensing systems when operating under designed operating conditions. It must be emphasized that the very limited entries in these tables are at best order of magnitude estimates. It is believed that the experimental accuracies actually attained with present systems typically are within a factor of three of these numbers, and that the relative accuracies of adjacent measurements made by the same system would be perhaps 3 to 10 times better. In most cases, the theoretical limits to the accuracy of the measurement system are much smaller than those presently attained, and we may expect significant improvement in measurement accuracy with time.

For more detailed information on the performance of remote sensing systems, the reader is referred to the appropriate chapters of this book.

30.8 Pattern Recognition in Remote Sensing

Up to this point, the chapter has been concerned with the status of various remote sensing techniques to make *quantitative* measurements, of various levels of information content and sophistication, of meteorologically significant parameters. It is important to recognize that much of the value of remote sensing techniques lies not in the ability to reproduce the kinds of quantitative data the meteorologist is accustomed to, but in the ability to provide in real time totally new types of data and data presentations. These data can often take the form of “maps” or 2-dimensional displays of the intensity of some parameter. Examples of such maps are the PPI plots of radar echo intensity, or the height vs. time facsimile recordings of C_T from monostatic acoustic echo sounders. Pattern recognition will play a major role in the use of such maps. Man has long been accustomed to using his eyes as a superb passive optical remote sensing system and it is important to recognize that the information is processed by pattern recognition and *not* by quantitative physical measurements. (The large grey structure over there is a thunderstorm. That transient line of enhanced brightness was a flash of lightning. In that direction it is raining already. Over there the sky is still clear, etc., etc.)

It is only in the last 25 years that local weather forecasters have been able to supplement the local weather information available from the human eye with an additional remote sensing capability – namely, weather radar. Again, pattern recognition (in this case while looking at “maps” of echo strength), is important to the interpretation of the data.

Under clear sky conditions, the human eye and the weather radar receive no information (other than the fact that there is no pattern) but already new remote sensing techniques permit the internal structure of the boundary layer to be monitored continuously under clear sky conditions. In recent years, we have learned how to use many different tracers, additional to hydrometeors, for remote sensing purposes. Specifically, small scale (of order $\lambda/2$) fluctuations in temperature and velocity can be used acoustically to monitor the internal structure and processes of the atmosphere; similarly, the small scale humidity fluctuations can be used as tracers of boundary layer structure by FM-CW radar techniques. At optical wavelengths, aerosol are readily detected by lidar techniques and may therefore also be used to monitor the boundary layer. The resultant patterns are already being recognized in terms of thermal plumes, inversions, internal gravity waves, breaking waves, etc. As the local weather forecaster gains skill in recognizing patterns in mesoscale and microscale maps of parameters such as velocity, temperature and humidity (and their small-scale structure parameters, C_v^2 ,

C_T^2 and C_e^2), so he should greatly enhance his ability to identify the existence and location of meteorological phenomena or processes (e.g., thermal plumes, radar hook echoes, and temperature inversions) and hence to observe and predict their development. Note that these maps do not have to have high absolute accuracy; the many orders of magnitude variability of parameters such as C_T , C_v , C_e , and radar and lidar reflectivities, mean that relative accuracies of 3 dB and absolute accuracies of 10 dB may be acceptable for pattern recognition purposes. The wide range of 2-dimensional or 3-dimensional maps potentially available to the meteorologist of the future will provide him with totally new abilities to watch for changing atmospheric conditions, and hence will radically affect both the nature and scope of future mesoscale research and services.

30.9 Tables

Table 30.1 Status of Acoustic Remote Sensing of the Boundary Layer Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Average Line Profile	C^1 C^3	C^2 C^3	C^8 A^9	A^9		
2-D Coverage 3-D Coverage	A^4		C^{10}			
Structure Constant	A^5					
Spectrum	A^6		A^{11}			
Flux	A^7					

¹ Spaced line-of-sight, cross-correlation techniques (Mandics, 1971)

² Sonic anemometer (Little, 1969)

³ Doppler echo sounding (Beran and Clifford, 1972)

⁴ Steerable or multibeam Doppler echo sounding (Hall, private communication)

⁵ Bistatic echo sounding (Little, 1969)

⁶ Doppler echo sounding (Little, 1969)

⁷ Doppler echo sounding (Little, 1972b)

⁸ Sonic anemometer/thermometer (Little, 1969)

⁹ Multi-frequency monostatic echo sounding (Gething and Jessen, 1971)

¹⁰ Monostatic echo sounding (Little, 1969)

¹¹ From C_T measurements (Wyngaard, Izumi and Collins, 1971)

Table 30.2 Status of Acoustic Remote Sensing of the Troposphere Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Integral Line Average Line Profile	A ¹	A ¹	A ⁴	A ⁴		
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux	A ² A ³ A ¹ A ¹					

¹ Doppler acoustic echo sounding (Beran and Clifford, 1972)

² Steerable or multi-beam Doppler echo sounding (Hall, private communication)

³ Bistatic echo sounding (Little, 1969)

⁴ Multi-frequency monostatic echo sounding (Gething and Jessen, 1971)

⁵ Monostatic echo sounding (Little, 1969)

Table 30.3 Status of Acoustic Remote Sensing of the Boundary Layer Under Precipitation Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Precipitation	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal					
Line Average Line Profile		(A ¹)			(B ²)		
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux	(A ¹) (A ¹) (A ¹)				(A ²)		

¹ Doppler acoustic echo sounding (Beran and Clifford, 1972)

² High frequency echo sounding to a few hundred meters range (Little, 1972a)

Table 30.4 Status of Acoustic Remote Sensing of the Troposphere Under Precipitation Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Integral Line Average Line Profile	(A ¹)	(A ¹)				
2-D Coverage 3-D Coverage Structure Constant	(A ¹)					
Spectrum Flux	(A ¹)	(A ¹)				

¹Doppler acoustic echo sounding (Beran and Clifford, 1972)

Table 30.5 Status of Radio Remote Sensing of the Boundary Layer Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Average Line Profile	C ¹ (C ²)	(C ³)	D ⁴ C ⁵	C ⁶		
2-D Coverage 3-D Coverage Structure Constant	(C ²) (B ²) (A ²)			C ⁷		
Spectrum Flux	(A ³)	(A ³)				

¹Spaced line-of-sight cross-correlation techniques (Lee and Waterman, 1968)

²Multi-station Doppler radar, using chaff (Lhermitte, 1969)

³Single station Doppler radar, using chaff (Lhermitte, 1969)

⁴Passive multi-frequency and/or multi-angle scan of O₂ absorption at ~ 5 mm (Snider, 1972)

⁵Hybrid acoustic-radio (RASS) echo sounding (Marshall, Peterson and Barnes, 1972)

⁶Line-of-sight time-of-flight measurements of integrated refractivity (Bean et al., 1969)

⁷From FM-CW radar echo power, assuming refractivity fluctuations are primarily caused by water vapor (Bean, 1971)

Table 30.6 Status of Radio Remote Sensing of the Troposphere Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Integral Line Average Line Profile	C ¹	(A ²)	D ⁴	C ⁵		
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux	(A ³) (A ³) (A ³) (A ²) (A ²)					

¹ Forward scatter Doppler measurements with RAKE system (Birkemeier et al., 1969)

² Single station Doppler radar, using chaff (Lhermitte, 1969)

³ Multi-station Doppler radar, using chaff (Lhermitte, 1969)

⁴ Passive O₂ radiometer measurements (to ~ 5 km height only) (Snider, 1972)

⁵ Passive 20 GHz radiometer measurements (Guiraud, Decker, and Westwater, 1972)

⁶ FM-CW radar (Richter, 1969)

Table 30.7 Status of Radio Remote Sensing of the Boundary Layer or Troposphere Under Precipitation Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Precipitation	Aerosol	Gaseous Pollu- tants
	Transverse	Longitudinal					
Line Integral Line Average Line Profile	C ²	C ¹ C ¹			C ⁴ C ⁵ C ⁵		
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux		C ² B ³ C ¹ C ¹			C ⁶ C ⁶ C ⁶ C ⁷		

¹ Single station Doppler radar (Lhermitte, 1969)

² Scanning single Doppler radar in a plane will give 2-D plot of *radial* velocity. Two-dimensional coverage of the *vector* wind requires two Doppler radars, preferably operating in the coplane mode (Lhermitte, 1969)

³ Three (or more) Doppler radar system (Lhermitte, 1969)

⁴ From echo power on calibrated radar, or passive microwave radiometry at 5-10 GHz (Atlas, 1964; Decker and Dutton, 1970)

⁵ Calibrated radar (Atlas, 1964)

⁶ Scanning calibrated radar (Atlas, 1964)

⁷ From Z-R relationships (Atlas, 1964)

Table 30.8 Status of Optical Remote Sensing of the Boundary Layer Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Average Line Profile	C ¹ B ²	A ³ B ²	C ⁵ C ⁶	C ⁸ C ⁹	C ¹¹	A ¹³ A ¹⁴
2-D Coverage	A ⁴		A ⁶	A ⁹	C ¹¹	A ¹⁴
3-D Coverage	A ⁴		A ⁶	A ⁹	A ¹¹	A ¹⁴
Structure	A ⁴					
Constant						
Spectrum	A ⁴		A ⁶	A ⁹	A ¹¹	A ¹⁴
Flux	A ⁴		A ⁷	A ¹⁰	A ¹²	A ¹⁵

¹ Laser beam scintillation transverse wind system (Lawrence, Ochs and Clifford, 1972)

² Spaced lidar beams, cell correlation method (Derr and Little, 1970)

³ Laser beam scintillation, longitudinal wind component system (Lawrence, private communication)

⁴ Optical Doppler (Owens, 1969)

⁵ Line-of-sight, time-of-flight measurements of integrated refractivity (Lawrence, 1969)

⁶ Lidar, using nitrogen Raman density measurements; or Stokes/anti-Stokes ratio (Strauch, Derr and Cupp, 1971; Salzman, Masica and Coney, 1971)

⁷ Eddy correlation method, using optical Doppler and N₂ Raman

⁸ Relative absorption of HCN laser lines (Derr, private communication)

⁹ Lidar, using H₂O Raman, or differential absorption in vicinity of water vapor line (Strauch, Derr and Cupp, 1972; Schotland, 1969)

¹⁰ Eddy correlation method, using optical Doppler and H₂O Raman

¹¹ On frequency lidar (Collis, 1970)

¹² Eddy correlation method, using optical Doppler and aerosol scatter

¹³ Infrared absorption (Hanst, 1970)

¹⁴ Raman, resonance Raman of fluorescent spectroscopy (Derr and Little, 1970)

¹⁵ Eddy correlation method, using optical Doppler and spectroscopic lidar

Table 30.9 Status of Optical Remote Sensing of the Troposphere Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Integral	A ¹ A ²	A ²	C ³	C ⁴	C ⁵	
Line Average			C ³		C ⁶	
Line Profile			A ³	A ⁴	C ⁶ A ⁶	
2-D Coverage						
3-D Coverage						
Structure						
Constant						
Spectrum						
Flux						

- ¹ Laser beam scintillations, using geostationary satellite-borne laser
- ² Spaced cell correlation method, using aerosol scatter (Derr and Little, 1970)
- ³ Lidar, using nitrogen Raman density measurements (Strauch, Derr and Cupp, 1971) or Stokes/anti-Stokes ratio (Salzman, Masica and Coney, 1971)
- ⁴ Lidar, using H₂O Raman density measurement (Strauch, Derr and Cupp, 1972)
- ⁵ Solar extinction measurements (Fegley, Blifford and Gillette, 1972)
- ⁶ On-frequency lidar (Collis, 1970)

Table 30.10 Comparison of Acoustic, Radio and Optical Remote Sensing Capabilities in the Boundary Layer Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Average	C ^{1,2,3} C ¹ (C ²)B ³	C ¹ , A ³ C ¹ (C ²)B ³	C ^{1,3} A ¹ D ² , C ³	C ^{2,3} A ¹ C ³	C ³	A ³ A ³
Line Profile						
2-D Coverage	A ^{1,3} (C ²)		A ³	A ³	C ³	A ³
3-D Coverage	A ³ (B ²)		A ³	A ³	A ³	A ³
Structure	A ^{1,3} (A ²)		C ¹	C ²		
Constant						
Spectrum	A ^{1,3} (A ²)		A ³	A ³	A ³	A ³
Flux	A ^{1,3} (A ²)		A ^{1,3}	A ³	A ³	A ³

- 1 = Acoustic
- 2 = Radio
- 3 = Optical
- A = Concept identified
- B = Concept analyzed theoretically
- C = Concept tested experimentally
- D = Development model tested
- E = Commercial model tested
- F = Commercial model available

Parenthesis used to denote chaff required as tracers for Doppler radars operating under clear air conditions.

Table 30.11 Comparison of Acoustic, Radio and Optical Remote Sensing Capabilities in the Troposphere Under Clear Air Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal				
Line Integral Line Average Line Profile	A ³ A ^{1,3} C ¹	A ^{1,3} (A ²)	C ³ A ¹ , D ² , C ^{2,3}	C ² C ³ C ³	C ³ C ³	
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux	A ¹ (A ²) (A ²) A ¹ (A ²) A ¹ (A ²) A ¹ (A ²)		A ³ A ¹	A ³ A ²	C ³ A ³	

1 = Acoustic

2 = Radio

3 = Optical

- A = Concept identified
- B = Concept analyzed theoretically
- C = Concept tested experimentally
- D = Development model tested
- E = Commercial model tested
- F = Commercial model available

Parenthesis used to denote chaff required as tracers for Doppler radars operating under clear air conditions.

Table 30.12 Comparison of Acoustic, Radio and Optical Remote Sensing of the Boundary Layer Under Precipitation Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Precipitation	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal					
Line Average Line Profile	C ²	C ² C ² (A ¹)			C ² C ² (B ¹)		
2-D Coverage 3-D Coverage Structure Constant Spectrum Flux	C ² (A ¹) B ²				C ² (A ¹) C ²		
	C ² (A ¹) C ² (A ¹)				C ² C ²		

1 = Acoustic

2 = Radio

3 = Optical

- A = Concept identified
- B = Concept analyzed theoretically
- C = Concept tested experimentally
- D = Development model tested
- E = Commercial model tested
- F = Commercial model available

Parenthesis used to denote chaff required as tracers for Doppler radars operating under clear air conditions.

Table 30.13 Comparison of Acoustic, Radio and Optical Remote Sensing of The Troposphere Under Precipitation Conditions

Parameter Type of Measurement	Wind		Temperature	Humidity	Precipitation	Aerosol	Gaseous Pollutants
	Transverse	Longitudinal					
Line Integral					C ²		
Line Average					C ²		
Line Profile	C ² (A ¹)	C ² (A ¹)			C ²		
2-D Coverage	C ² (A ¹)				C ²		
3-D Coverage	B ²				C ²		
Structure							
Constant							
Spectrum	C ² (A ¹)				C ²		
Flux	C ² (A ¹)				C ²		

1 = Acoustic

2 = Radio

3 = Optical

A = Concept identified

B = Concept analyzed theoretically

C = Concept tested experimentally

D = Development model tested

E = Commercial model tested

F = Commercial model available

Parenthesis used to denote chaff required as tracers for Doppler radars operating under clear air conditions.

Table 30.14 Estimates of Accuracy of Wind Velocity Measurements

Type of Measurement	Technique	Estimated Accuracy		
		Acoustic	Radio	Optical
Line average of transverse velocity	Line-of-sight scintillations	$\pm 10\% \pm 0.5 \text{ ms}^{-1}$	$\pm 10\% \pm 0.5 \text{ ms}^{-1}$	$\pm 5\% \pm 0.3 \text{ ms}^{-1}$
Line average of longitudinal velocity	Sonic Anemometer	$\pm 5\% \pm 0.3 \text{ ms}^{-1}$		
Line profile of transverse velocity	Multistation Doppler	$\pm 10\% \pm 0.5 \text{ ms}^{-1}$	$\pm 5\% \pm 0.3 \text{ ms}^{-1}$	
Line profile of longitudinal velocity	Single station Doppler	$\pm 5\% \pm 0.3 \text{ ms}^{-1}$	$\pm 5\% \pm 0.2 \text{ ms}^{-1}$	
2-D coverage of vector wind	Multistation Doppler		$\pm 5\% \pm 0.3 \text{ ms}^{-1}$	

Remote measurements of the spectrum of turbulence, and of the momentum flux are currently available only by microwave Doppler radar. The accuracies of these spectra and flux measurements will be determined by the accuracies (estimated above) of the original Doppler measurements from which they are calculated.

Status of Remote Sensing of the Troposphere

Table 30.15 Estimates of the Accuracy of Temperature Measurements

Type of Measurement	Technique	Estimated Accuracy		
		Acoustic	Radio	Optical
Line integral	Time-of-flight	$\pm 1^\circ\text{K}$		$\pm 0.2^\circ\text{K}$
Line profile	Multifrequency O ₂ radiometer N ₂ Raman lidar	$\pm 2^\circ\text{K}$		$\pm 1^\circ\text{K}$
Structure Constant	Monostatic acoustic sounder	$\pm 50\%$		

Table 30.16 Estimates of Accuracy of Water Vapor Content

Type of Measurement	Technique	Estimated Accuracy		
		Acoustic	Radio	Optical
Line integral	Microwave radiometry		$\pm 10\% \pm 1 \text{ gram/cm}^2$	
Line average	Time-of-flight HCN laser		$\pm 10\%$	0.1 mb
Line profile	H ₂ O Raman			0.1 mb
Structure Constant	FM-CW radar		Factor of 2	

Table 30.17 Estimates of Accuracy of Aerosol Reflectivity

Type of Measurement	Technique	Estimated Accuracy		
		Acoustic	Radio	Optical
Line profile	Lidar			Factor of 2*
2-D coverage	Scanning lidar			Factor of 2*

* Relative accuracy of adjacent measurements with the same system should be within $\pm 10\%$.

30.10 References

- Atlas, D. (1964), Advances in radar meteorology, *Adv. in Geophysics*, 10, 317-478.
- Bean, B. R. (1971), Comparisons of remote and in situ measurements of meteorological parameters and processes, *Proc. of NATO Adv. Study Inst.*, April 15-22.
- Bean, B. R., R. E. McGavin, C. B. Emmanuel and R. W. Krinks (1969), Radiophysical studies of evaporation at Lake Heffner, 1966 and 1967, *ESSA Tech. Rpt. ERL 115-WPL7*.
- Beran, D. W. and S. F. Clifford (1972), Acoustic Doppler measurements of the total wind vector, *A.M.S. Sec. Symp. on Meteor. Observ. and Instru.*, March 27-30, San Diego, CA.
- Birkemeier, W. P., P. F. Duvoisin, A. B. Fontaine and D. W. Thomson (1969), Indirect atmospheric measurements utilizing RAKE tropospheric scatter techniques – Part II Radiometric interpretation of RAKE channel-sounding observations, *Proc. I.E.E.E.*, 57, 4, 552-559.
- Cole, J. E. and P. A. Dobbins (1970), Propagation of sound through atmospheric fog, *J. Atmos. Sci.*, 27, 426-434.
- Cole, J. E. and P. A. Dobbins (1971), Measurements of the attenuation of sound by a warm air fog, *J. Atmos. Sci.*, 28, 202-209.
- Collis, R. T. H. (1970), Lidar, *Appl. Opt.*, 9, 8, 1782.
- Decker, M. T. and E. J. Dutton (1970), Radiometric observations of liquid water in thunderstorm cells, *J. Atmos. Sci.*, 27, 785-790.
- Derr, V. E. and C. G. Little (1970), A comparison of remote sensing of the clear atmosphere by optical, radio and acoustic radar techniques, *Appl. Opt.*, 9, 9, 1976-1992.
- Fegley, R. W., I. H. Blifford, Jr. and D. A. Gillette (1972), A comparison of measured atmospheric solar extinction with that predicted from direct aerosol data (in publication).
- Gething, J. T. and D. Jenssen (1971), Measurements of temperature and humidity by acoustic echo sounding, *Nature* 231, 198-200.
- Guiraud, F. O., M. T. Decker and E. R. Westwater (1972), Experimental investigation of the conceptions of electrical range by passive microwave radiometry, *NOAA Tech. Rpt. ERL 221-WPL19*.
- Hanst, P. L. (1970), Infrared spectroscopy and infrared lasers in air pollution research and monitoring, *App. Spectros.*, 24, 2, 161.
- Lawrence, R. S. (1969), Remote atmospheric probing by ground-to-ground line-of-sight optical methods, *Report of Remote Atmospheric Probing Panel, Committee on Atmospheric Sciences, Nat'l. Acad. of Sci. – Nat'l. Res. Coun.* 2, 91-110.
- Lawrence, R. S., G. R. Ochs and S. F. Clifford (1972), The use of scintillations to measure average wind across a light beam, *Appl. Opt.*, 11, 2, 239-243.
- Lee, R. W. and A. T. Waterman, Jr. (1968), Space correlations of 35 GHz transmissions over a 28-km path, *Radio Sci.*, 3, (New Series), 2, 135-139.
- Lhermitte, R. M. (1969), Atmospheric probing by Doppler radar, *Report of Remote Atmospheric Probing Panel, Committee on Atmospheric Sciences, Nat'l. Acad. of Sci. – Nat'l Res. Coun.* 2, 253-285.

- Little, C. G. (1969), Acoustic methods for the remote sensing of the lower atmosphere, *Proc. I.E.E.E.*, 57, No. 4, 571-578.
- Little, C. G. (1972a), On the detectability of fog, cloud, rain and snow by acoustic echo-sounding methods, *J. Atmos. Sci.* (in press).
- Little, C. G. (1972b), Prospects for acoustic echo sounding, Chapter 19.
- Mandics, P. A. (1971), Line-of-sight acoustic probing of atmospheric turbulence, Tech. Rpt. 4502-1, Stanford Elect. Labs., Stanford Univ., Stanford, CA.
- Marshall, J. M., A. M. Peterson and A. A. Barnes, Jr. (1972), Combined radar-acoustic sounding system, *Appl. Opt.*, 11, 108-112.
- Owens, J. C. (1969), Optical Doppler measurement of microscale wind velocity, *Proc. I.E.E.E.*, 57, 4, 530-536.
- Richter, J. H. (1969), High resolution tropospheric radar sounding, *Radio Sci.*, 4, No. 12, 1261-1268.
- Salzman, J. A., W. J. Masica and T. A. Coney (1971), Determination of gas temperatures from laser Raman scattering, NASA Tech. Note TN D-6336, Washington, D.C.
- Schotland, R. M. (1969), Some aspects of remote atmospheric sensing by laser radar, Report of Remote Atmospheric Probing Panel, Committee on Atmospheric Sci., Nat'l. Acad. of Sci. – Nat'l. Res. Coun., 2, 179-200.
- Snider, J. B. (1972), Ground-based sensing of temperature profiles from angular and multi-spectral emission measurements, *J. Appl. Meteor.* (in press).
- Strauch, R. G., V. E. Derr and R. E. Cupp (1971), Atmospheric temperature measurement using Raman lidar, *Appl. Opt.*, 10, 12, 2665-2669.
- Strauch, R. G., V. E. Derr and R. E. Cupp (1972), Atmospheric water vapor measurement using Raman lidar, *Remote Sensing of Environment* (in press).
- Wyngaard, J. E., Y. Izumi and S. A. Collins, Jr. (1971), Behaviour of the refractive-index-structure parameter near the ground, *J. Opt. Soc. Am.*, 61, 1646-1650.

SUBJECT INDEX

- Absorbing spheres, 8.6
- Absorption
 - acoustic-gravity waves, 21.1
 - coefficient, 9.4, 9.7, 10.2.1, 24.1.1
 - of acoustic energy, 18.1
 - of radiation, 9.1.2
- Acoustic Doppler measurements, 18.5
- Acoustic echo sounder design, 18.3
- Acoustic echo sounding, 18, 19, 20
 - at airports, 19.5.5
 - bistatic, 18, 19.2
 - Doppler sounding, 18, 19.3
 - in urban meteorology, 19.5.5
 - monostatic, 19.2
 - of gravity waves, 19.5.5
 - of hydrometeors, 19.2
 - of momentum flux, 19.5.1
 - of radiation inversions, 19.5.5
 - of temperature structure constant, 18, 19.1, 19.5.5
 - of thermal plumes, 18, 19.3, 19.5.5
 - of turbulence, 18, 19.5.1, 19.2
 - of velocity field, 19.5.1
 - of velocity structure constant, 19.1, 19.5.5
 - physics of scatter, 10, 18, 19.1
 - potential measurement capabilities, 19.2
 - refraction effects, 19.3, 19.4.2
 - with spaced sounders, 19.5
- Acoustic remote sensing status, 30.3
 - (under precipitation conditions)
- Acoustic waves, 3.3.3, 7
- Acoustics, atmospheric, 18
- Activation energy, 27
- Active systems, 30.1.1, 30.1.3
- Adapted beam, 29.3.1
- Adiabaticity, 5.1
- Adiabatic lapse rate, 1.5.2
- Adiabatic motions, 3.4
- Adjoint operator, 16.3.3
- Advantages of optical remote sensing, 30.5
- Advantages of remote sensing, 30.1
- Advection, mathematical concepts, 1.7.1
- Aerosols, 1.4
 - attenuation effects of, 24.1.2
 - characteristics of, 23.4
 - content, 23.4
 - indices of refraction of, 23.4
 - polydisperse size distributions of, 10.5.3
 - scattering by, 8.6
 - size distribution of, 23.4
 - spherical and non-spherical, 23.4
- Ageostrophic wind, 1.8.2
- Air density, 23.4
 - measurement (by lidar), 23.4
 - fluctuation vs. time, 23.4
- Air quality measurements, 24.1
- Air-to-water interface scattering, 24.2.2
- Aliasing, 28A.1
- Alternating tensor, 2.4
- Altimeter measurement of sea scatter, 12.5
- Ambient field, 29.6, 29.6.1
- Ambient noise field, 29.2.2
- Amplitude covariance function, 11.3.2
- Amplitude fluctuations of a plane wave, 11.3.1
- Amplitude spectra of lightning, 17.2.1
- Angular dependence of scattering, 10.3.3, 10.4.3
- Angular dispersion, 21.4
- Angular momentum, 9.1.1, 9.2.1, 9.3.2
- Antenna effects, 15.2.2.1
- Aperture, 29.1.1, 29.2.1
 - operator, 29.1.1
 - beam, 29.1.1
 - array, 29.1.1
 - weighting, 29.2.1
- APT (Automatic Picture Transmission), 26.2.2
- Arrays, pressure sensor, 21.4
- Arrhenius equation, 27
- Artificial Source 21.7
- Association reactions, 27
- Atmospheric
 - boundary layer, 3.4
 - composition, 1.1
 - electricity, 17.4
 - energy budget, global scale, 1.6.3
 - gravity waves, 3.3.3
 - motion, meso-scale, 2.11
 - motion scales, 1.0, 1.6.4, 2.1, 2.9, 2.17
 - tides, 3.3.2
 - transmission, 9.6
 - waves, 21.1
 - windows, 22.0
- Atmospherics, 17.2, 17.3, 21.6
- ATS (Application Technology Satellite), 26.1.1, 26.2.1
- Attenuation, acoustic, 18.1
- Attenuation of radiation, 9.6
- Aurora, infrasound from, 21.5
- Austausch coefficients. 1.8.2, 2.17, 2.18, 2.19, 6.2.1, 6.2.2
- Autocorrelation function, 28A.4
- Average, ensemble, 11.1.2, 1.7.2, 28A.4.4
- Average cost, risk, 29.1.4
- Average value, definition, 1.7.4, 4
- Backscatter cross-section, 13.1.3, 13.1.3.1
- Backscattered signal
 - distributed target, 13.1.2
 - point target, 13.1.1
- Balloon, 27
- Baroclinic instability, 2.22, 3.5.2
- Barotropic instability, 3.5.2
- Basis set, 16, 10.1.1
- Bayes matched filters, 29.8.1
- Bayes risk, 29.1.4
- Beam-limited radar altimeter, 12.5.3
- Beam pattern, 29.2.1
- Beam spreading, 25.1.3
- Bearing extraction, 29.8.2
- Billows, radar detection of, 14.5
- Birds, radar detection of, 14.3
- Bistatic sea scatter, 12.2.3
- Blackbody radiation, 22.1.2
- Boltzmann distribution law, 9.2.3
- Bomex, 22.1.4.3
- Boundary conditions, 8.1
 - sound-hard, 8.3
 - sound-soft, 8.3
- Boundary layer
 - atmospheric, 2.14, 2.17, 4, 6, 20
 - dynamics of, 20
 - equatorial, 6.1
 - planetary, turbulence in, 2.14
 - pressure fluctuations in, 21.3
 - stable, 6.1
- Boussinesq approximation, 2.6, 6.2.3, 21.2
- Bragg angle reflection, 11.1.3
 - ocean waves, 12.2.1
 - indirect, 12.4
- Brightness temperature, 22.2.1.1
- Brillouin scattering, 10.5.5
- Brunt frequency, 1.5.2, 21.2
- Buoyancy, 2.6

- Buoyancy oscillations, convective cell, 21.6
- Buoyant forces, 2.14
- Buoyant production of energy, 6.5.3
- Burst rate, 17.3.1, 17.3.3, 17.3.4

- Canonical theory, 29, 29.0, 29.1.2
- Capillary waves, 12.1.1
- Cartesian coordinate system, 2.4
- Chain reactions, 5.6
- Channel capacity, 28.2
- Channel model, 29.2, 29.4, 29.6, 29.8
- Chimonas instability, 7.4.4
- Chinook, 2.21
- Chlorophyll, 26.3
- Chromaticity
 - coordinates, 22.3.1.1
 - diagram, 22.3.1.1
- Clear air radar, 20
- Clear air turbulence, 2.4, 2.8ff, 2.13, 2.21, 14.6, 20
 - non-equilibrium patches of, 2.18
 - sensing of, 14.6, 20, 25.2.5
- Climatic change, 2.1
- Cloud backscatter, 24.2.4
- Cloud, motion of, 26.2.2
- Coefficient, eddy characteristics, 1.8.2, 2.24, 2.25, 2.29, 6.2.1, 6.2.2

- Coherence, 21.4
- Coherence function, 28A.4
- Coherent radar, 13.1.4
- Coherent scattering, 10.5.1
- Collision broadening, 9.4.3
- Collisions, mean time between at STP, 10.2
- Color perception, 22.3.1
- Color triangle, 22.3.1
- Committee on atmospheric science, 2.1
- Comparison of acoustic, radio and optical remote sensing, 30.6
- Composite rough surface model, 12.4
- “Concorde”, 2.21
- Conduction, molecular, 6.2.2
- Continuity equation, 1.6.1, 4, 6.2.3, 13.2.3.1
- Contrast, 26
- Convection, 6.4.4
 - acoustic detection of, 18.4.1
 - clear air convective cells, 14.4
 - convective field, 14.4
 - mesoscale convective patterns, 14.4
 - radar detection of clear air, 14.4
- Convective processes, 20
- Convective storm, 13.2.4
 - infrasound from, 21.6
- Coriolis term, 2.2
- Correlation functions, 4.2.2, 28A.4
- Correlation-slope method, wind measurement, 25.3.3
- Correlation spectrometer, 24.1.2
- Correlator, analog, 21.4
- Cosmic radiation, 22.2.6
- Cospectrum, 2.8, 2.17, 28A.4
- Cost, 29.1.4
- Coupled-system cross-section, 29.4.2
- Coupling operator, 29.3.3
- Covariance, see eddy correlation, 1.8.2
 - function, 11.1.2, 28A.4, 29.5.2, 29.6
 - function of scintillations, 25.1.2
 - two-dimensional, 11.3.2
- Creation – destruction interaction, 10.3.1
- Critical layer, 21.8
- Critical Richardson number, 6.6.1
- Cross-correlation
 - analysis, 21.4
 - function, 28A.4
- Cross-covariance function, 28A.4
- Cross-power spectral analysis, 21.4
- Cross-spectral density function, 28A.4
- Cross section
 - Mie scattering, 10.4.3, 23.4
 - quantum scattering, 10.2.2
 - Raman scattering, 23
 - scattering, 11.1.3
- Curvature of a laser beam, 25.5
- Cyclostrophic flow, 3.2.4

- Damping factor, 10.3.1
- Data gathering networks, 28.2
- Data record length, 28A.4
- Data sampling, 29.0.2, 29.7, 29.9
- Data validation, 29.0, 29.7
- Debye potentials, 10.4.1
- Decision rule, 29.1.4
- Deep-water waves, 12.1.1
- Density, spectral, 2.8, 2.13
- Detection, 29.0.1, 29.1.1, 29.1.4, 29.8.1
- Detector noise, 26.1.4
- Diabatic surface layer, 6.4
- Dielectric scattering centers, 10.0
- Dielectric spheres, 8.6
- Diffusion coefficient, 6.2.1
- Dimensional analysis, 2.2
- Dipole moment, 9.1.2
- Direction of arrival, 17.3.2
- Discharge current, 17.1, 17.2
- Dispersion
 - angular, 21.4
 - relationship, ocean gravity waves, 12.1.3
- Dissipation
 - acoustic-gravity wave, 21.8
 - gravity waves, 5.5
 - rate of, 2.8, 2.14
- Diurnal tide, 5.10
- Divergence factor, sphere, 8.6
- Dominant wavelength, 22.3.1.1
- Doppler
 - acoustic, 18.5, 19.3
 - laser radar, 23.4
 - line broadening, 9.4.3
 - shift, 13.1.4
 - shift, due to scattering, 10.5.4
 - spectrum, 13.1.5
 - wind measurements, acoustic, 18.5
- Drop-size distribution, 13.2.2
- Duct, 8.5
- Duration, wave, 12.1.1
- Dust devils, 6.5.4
- Dye laser, 24.1.2, 24.2.1
- Dynamical equation, 29.2.1, 29.4.1

- Earthquakes, infrasound from, 21.5
- Earth Resources Technology Satellite (ERTS), 26.1.1; 26.2.1
- Echo-sounding, acoustic, 18, 19
- Eddies, 2.1, 2.2, 2.4, 2.6
- Eddy
 - correlation, 6.2.2
 - heat transfer coefficient, 6.2.2
 - viscosity, 1.8.2, 6.2.1, 2.17ff, 2.19ff
- Eigenvalue and eigenvector, 10.1
- Eikonal equation, 8.4
- Einstein
 - coefficients, 9.1.2
 - summation convention, 2.4
- Ekman
 - instability, 6.3.3
 - layer, 2.17
 - spiral, 2.17, 3.2.5, 6.3.3

- Elastic scattering, 24.2.4
 Electric charge, 17.1
 Electromagnetic radiation, 8, 9, 11, 17.2
 Elliptical coordinates, bistatic sea scatter, 12.2.3
 Emission
 atmospheric, 9.5
 of radiation, 9.1.2
 probability (coefficient), 10.2.1
 Energy
 available potential, 2.31
 cascade of, 2.2, 4
 dissipation, 2.2
 equation, 6.2.4
 flux, wave-associated, 21.8
 kinetic, 2.7, 2.12ff, 2.19, 2.20, 2.21ff
 kinetic, dissipation of, 2.9
 kinetic, spectrum of, 2.1
 potential, 2.13, 2.22
 production, 6.5.3
 spectra, 4.2.2
 transfer between eddy scales, 2.9
 Ensemble
 average, 1.7.2, 11.1.2, 28A.4, 29.7
 validity, 29.7.3
 Enthalpy, 6.2.2
 Entropy, 2.7
 Environmental noise, 1.4, App. 1
 Equation
 energy, 2.7
 momentum, 2.7
 of motion, 2.5, 3.1, 6.2.1
 of state, 1.2, 2.3
 Ergodic processes, 1.7.2, 28A.4
 Eulerian coordinate system, 1.7.1
 Extinction
 coefficient, 23.1.1
 measurement of, 23.4
 Extraction, 29.0.1, 29.1.1, 29.1.4

 Fabry-Perot interferometer, 24.1.2, 24.2.1
 Fair weather field, 17.4
 Fall velocity, 13.2.1, 13.2.3, 13.2.4.1, 13.2.2
 Far field assumption, 11.1.3
 Fetch, 12.1.1
 Fickian diffusion, 2.19
 Filter
 effect of on spectra, 9.6.1
 high-pass, 2.4, 2.17, 2.19
 low-pass, 2.5, 2.8, 2.19
 Filtering, atmospheric, 21.6
 Fine structure, 9.4.5
 Flow, smooth, 6.3.1.1
 Fluid equations, 4.1
 Fluid shear, 1.8.2
 Fluorescent scattering, 10.2.2
 Fluorescence, 10.2.2
 lidar, 24.2.3
 Flux, 4, 6, 6.5.4
 profile relationships, 6.4.2
 Richardson number, 6.4.1
 Forces, frictional, 2.17
 Forecasting, long-range, 2.24
 Forel scale, 22.3.1
 Fourier
 Stieltjes integral, 11.3.1
 transform, discrete, 13.1.5.2
 Fredholm, integral equation of the first kind, 16.2
 Free convection, 6.4.4
 Free-space, Green's function, 8.3
 Frequency, Brunt-Väisälä, 21.2, 1.5.2
 Friction term, 2.2
 Friction velocity, 1.8.2

 Fully developed seas, 12.1.1
 Funnel cloud, 17.3

 Galactic radiation, 22.2.6
 Gap, spectral, 2.17
 GARP, 2.9, 2.21
 GATE, 2.21
 Gauss process, 29.5.3
 Gauss' theorem, 11.1.3
 General circulation, concepts, 1.6.2
 Generalized, adapted beams, 29.3.1
 Geometric acoustics, 21.6
 Geometrical cross section, 8.6
 optics, 8.4
 Geostationary Operational Environmental Satellite (GOES), 26.2.2
 Geostrophic wind, 3.2.2
 equation, 1.8.2
 Global monitoring, 22.1.1
 Globar, 24.1.1
 Golay cell, 24.1.1
 Görtler instabilities, 2.20
 Gradient flow, 3.2.3
 Gravity waves, 6.3.3, 12.1.1, 19.5.5
 critical layers, 7.4.1
 impedance relation, 7.2
 dissipation by viscosity—boundary layer, 7.2
 dispersion equation, 5.15
 ducting by boundary-layer temperature, wind structure, 7.2
 energy flux, 7.3
 momentum flux, 7.3
 turbulence interactions, 7.4.3
 wave interactions, 7.4.2
 Great Plains, 2.17, 2.19
 Green's function, 11.1.3
 Greybody, 22.1.2
 Groundwave propagation, 17.2.2
 Ground-wave sea backscatter, 12.2.2
 Gulf stream, 22.1.1

 Hadley cell, 1.6.2
 Hamiltonian, 10.1.2
 Harmonic oscillator representation, 10.1.2
 Heat
 conduction equation, 6.2.2
 equation, first law, 1.5.2
 equator, 1.6.2
 latent, 2.17, 2.19
 sensible, 2.17
 transfer coefficient, 6.2.2
 Height, dimensionless, 6.4.1
 Helmholtz wave equation, 8.2
 Heterosphere, 1.1
 Histogram technique, 22.1.4.3
 Homosphere, 1.1
 Hot tower, 2.13
 convection, 2.17
 Hough functions, 5.9, 5.10, 5.11
 HRIR, 22.1.1
 Hydrocarbon Oxidation, 27
 Hydrometeors, 19.2, 1.4, 8, 23
 Acoustic echo-sounding of, 13, 14, 19.2
 Hydrostatic approximation, 3.2.1, 6.2.1
 equation, 1.5.1, 23
 Hyperfine structure, 9.4.5

 Ice, 22.2.5
 Ideal fluid, 1.8.2
 Ill-conditioned matrices, 16.2.2
 Ill-posed problems, 16.2
 Improved TIROS Operational Satellite (ITOS), 22.1.1, 26.1.1, 26.2.2

- Impulses, 17.2, 17.3
- Incompressibility, 3.4, 6
- Inelastic aerosol scattering, 24.2.2
- Inertial instability, 3.5.2
- Infrared absorption, 24.1.1
 - detection, 22.1.1
 - lasers, tunable and multiline, 24.1.1
 - radiometers, 15.2.3
- Infrared Temperature Profile Radiometer (ITPR), 26.2.1
- Infrared Interferometer Spectrometer (IRIS), 26.1.1
- Infrasound, 21.2
- Inner scale of turbulence, 25.2
- Insects, radar detection of, 14.3, 20
- Instability
 - inertial, 3.5.2
 - in gravity waves, 5.5
 - Kelvin-Helmholtz, 14.5, 20
- Intensity, acoustic, 18.1
 - constants for scattering, 10.3.3
- Interaction, non-linear, 2.1ff, 2.9, 2.17, 2.22
- Interference filters, 24.1.1, 24.1.2
- Internal boundary, 6.7.1
- Internal gravity waves, 21.2
- Intrinsic frequency of gravity waves, 5.7
- Inversion, atmospheric, 6, 18, 19.5.5, 20
- Inversion techniques, mathematical, 16, 16.3
 - aerosol distribution, 23.4
- Inversion techniques mathematical, 16.3
- Ion-molecule reactions, 27
- Ionospherically-propagated sea backscatter, 12.2.2
- Ionospheric propagation, 17.2.2
- Ionosphere, infrasound in, 21.6
- Iisentropic process, 1.5.2
- Isobar, 1.8.2
- Isothermal, 1.5.2
- Iterative inversion techniques, 16.3.3
- ITOS, 22.1.1
- Jet stream, 2.13, 2.21
 - infrasound from, 21.5
 - low-level, 2.17, 2.19
- Kelvin-Helmholtz waves, 2.9, 2.14, 2.20
 - instability, 14.5, 20
- Kernel of integral equation, 16.1, 16.2
- Kinematic properties of mean wind, 13.2.3.1
 - viscosity, 5.5, 6.2.1
- Kinetic energy balance, 6.2.4
- Kirchoff-Fresnel diffraction formula, 8.3
- Kirchhoff's law, 22.1.2
- Kolmogorov-Smirnov test, 29.7.2
- Kolmogorov theory of turbulence, 4.2.4, 11, 25.1.1
- Kronecker delta, 2.4
- Lagrangian coordinate system, 1.7.1
- Laminar flow, 6.7.1
- Landweber iteration, 16.3.3
- Langevin equation, 29.2.2, 29.5.1
- Lapse rate, 6.7.3
- Laser line width, 10.2.2
- Laser radar
 - qualitative measurements, 23.4
 - calibration, 23.4
 - line broadening, 23.4
 - systems, 23.2.4
- Latent heat, 2.12
- Lee waves, 1.5.2, 7.5, 21.5
- Length, Obukhov, 6.4.1
- LIDAR (Light Detection and Ranging), 26.3.2
 - applications, 23, 24.2.1
- Lifetime, collision, 9.4.3
- Lightning, 17
 - discharge, 17.1
 - infrasound from, 21.6
 - stroke location, 17.3.2
- Linear atmosphere, 8.5
- Line-of-sight propagation, 17.2.2
 - laser, 25
- Linewidth, spectral, 9.4, 10.2.2
- Log amplitude variance, 11.3.2
- Logarithmic wind profile, 6.3.2
- LORAN A sea-scattered signals, 12.2.3
- Lorentz force equation, 8.1
- Markovian approach, 29.5
- Mass conservation, 6.2.3
- Maxwell's equations, 8.1, 10.4.1, 11.1.1
- Mean, departures from, 2.4
- Mean wind, 13.2.3.1
- Measurement techniques, 15.2
- Medium operator, 29.3.3
- Meridional cell, 1.6.2
- Meteors, infrasound from, 21.5
- Michelson interferometer, 24.1.1
- Microbarograph, 21.3
- Microwave atmospheric range correction, 15.3.2
 - radiometry, 15.2.2, 26.3.1
- Mie scattering, 8, 10, 10.4, 23.1.4
 - coefficients, 10.4.1
 - lidar, 23.4, 24.2.4
- Mirage, 25.5
- Mixing length, 6.3.1
- Model atmosphere, 22.1.3.1
- Moisture variables, 1.3.1
- Molecular scattering, 10, 11, 23, 23.4
- Molecules, diatomic, 9.2
 - polyatomic, 9.3
- Momentum, 2.19
 - angular, 2.22, 9
 - flux, 2.21, 6, 19.5.1
 - flux, wave-associated, 21.8
- Monostatic acoustic echo-sounding, 19.2
- Motion, convective scale, 2.17
 - scales of, 2.1, 2.3, 1.6.4
- Mountains, infrasound from, 21.5
- MRIR, 22.1.1
- Multiparameter scattering measurements, 24.2.4
- Multiple Doppler radar, 13.2.4.2
- Multiple scattering, 10.5.2, 23.4
- Nadir, 22.1.1
- Natural line breadth, 9, 10.2.2
- Negative viscosity, 2.21
- NEMS (NIMBUS E Microwave Spectrometer), 26.2.1
- Neutral boundary layer, 6.3
- NIMBUS spacecraft, 22.1.1, 26.1.1, 26.2.1
- Nitrogen density measurements, 23.3.1
- Noise field, 29.2.2, 29.4, 29.4.1, 29.6
 - in laser radars, 23.2.2, 23.2.3
 - pressure fluctuations, 21.3
 - process, 29.3.2
- Nondispersive infrared analyzers, 24.1.1
- Nonlinearities in gravity waves, 5.4
- NO oxidation, 27
- Norton attenuation factors, 12.2.1
- Nuclear explosions, 21.7
- Number operator, 10.1.2
- Ocean, 22.0
 - color, 22.3
 - temperature, 22.2.2
- Operator, quantum mechanical, 10.1
- Operators, "coupled-medium", 29.1.1
 - propagation, 29.1.3
 - transmission, 29.1.2
- Optical index of refraction, 4.2
 - propagation, 25

- remote sensing status, 30.5
 - advantages, 30.5
 - under precipitation conditions, 30.5
- scattering, 10, 23.4
- spectrum analyzer, 24.2.1
- Optimization, 29.1.5, 29.8, 29.8.3
- Optimization, reception, 29.8
- Orbits, 27.1.2
- Parametric oscillator, 24.1.2
- Passive microwave system, 22.2
- Passive systems, 29.1.1, 29.1.3
- Pattern recognition, 30.7
- Penetrative convection, 21.6
- Perturbation approach to sea scatter, 12.2.1
 - expansion, 11.1.3
 - motion, 2.7
 - operator, 10.1.2
- Phase, 13.1, 13.1.4
 - fluctuations of a plane wave, 11.3.1
 - fluctuations produced by turbulence, 25.1.3
 - structure function, 11.3.2
 - velocity of acoustic waves, 19.1
- Phillips waveheight spectrum (of ocean), 12.1.4
- Photic zone, 22.3.2
- Photochemistry of the troposphere, 26
- Photopic vision, 22.3.1.1
- Phytoplankton, 22.3
- Photon counting, 23.2.1, 23.2.2
- Planck's law, 22.1.2
- Planetary boundary layer, relationship to numerical forecasting, 1.5.3
- Planetary waves, 1.6.4
- Plumes, 6.5.4
 - thermal, 19.5.5
 - acoustic echo-sounding of, 18, 19.3
- Point matching calculations, 10.4.3
- Poisson process, 29.5.2
- Polarization, 22.2.1.2
 - of scattered laser light, 23.4
 - relations, 21.3
- Pollutant mapping, 24.2, 24.2.2
- Polynyas, 22.2.5
- Polytropic atmospheres, 23.4
- Postulational approach, 29.5
- Potentials, scalar and vector, 8.2
- Potential temperature, 1.5.2
- Power density spectrum, 13.1.5.1
- Poynting vector, 11.1.3
- Precipitation
 - acoustic echoes from, 19.5.4
- Prediction, numerical, 2.2
- Prefocused beam, 29.3.1
- Pressure
 - atmospheric, 21.3
 - broadened line width, 10.2.2
 - gradients, wave-associated, 21.4
 - sensor, 21.3
- Primary colors, 22.3.1
- Primitive equations (Navier-Stokes equations), 1.5.3, 1.8, 2.3, 4, 6
- Probability density function, 11.1.2
 - distribution function, 11.1.2
- Propagation
 - acoustic, 18.1
 - acoustic-gravity wave, 21.6
 - electromagnetic, 9
 - equation, 29.2.1
 - line-of-sight, 11.3.1
- Pulse-limited radar altimeter, 12.5.3
- Pure color, 22.3.1.1
- Pure ocean water, 22.3.2.1
- Quadratic cost function, 29.1.5
- Quantity, averages of, 2.4
- Quantization noise, 28A.1
- Quantum number, 9.1.1, 9.2., 9.3
- Quasi-specular scatter from sea, 12.4, 12.5.1
- Quenching of fluorescence, 10.2, 10.2.2
- Radar
 - clear air echoes, 14
 - Doppler, 13
 - detection, CAT, 14.6
 - insects, birds, 14.3
 - clear air convection, 14.4
 - equation, 13.1.3, 13.1.3.1
 - laser, 23
 - range equation, 12.2.1
 - reflectivity, 13.1.3
 - reflectivity, clear air scattering, 14.1, 14.2
 - reflectivity factor, 13.1.3.1
 - scattering, clear air, 14.1
 - wave patterns, clear air, 14.5
 - weather, 21.6
- Radial velocity, 13.2.1
- Radiation field, 17.2
 - inversion, 19.5.5
- Radiative absorption, 6.2.2
 - transfer, 15.1
 - equation, 16.1
- Radiometer, basic, 15.2.1
- Radio refractive index of air, 1A, 2.19
- Radio wave remote sensing status, 30.4
 - under precipitation conditions, 30.4
- Rayleigh-Jeans law, 22.2.1
- Raman
 - density measurements, 23.3.1
 - scatter lidar, 23, 24.2.2
 - scattering, 10, 10.3.2, 23.1.5, 23.3.1
 - cross sections, 24.2.2
- Random fields, 11.1.2
 - homogeneous, 11.1.2
 - isotropic, 11.1.2
- Random processes, 28A.4
 - a realization of, 11.1.2
- Rate of occurrence, 17.3
- Ray equation, 8.4
- Rayleigh number, 3.5.2
- Rayleigh scattering, 8.6, 10.3.2, 10.3.3, 23.1.3
- Ray's asymmetry parameter, 9.3.2
- Ray tracing, acoustic-gravity wave, 21.6
- Reaction rate theory, 27
- Received waveform, 29.4.2
- Reference spectra, 24.1.1
- Reflection, acoustic, 18.2
- Refraction
 - acoustic, 19.3, 19.4.2
 - laser beam, 25.5
- Refractive index
 - acoustic waves, 11.2.2
 - aircraft measurements, 14.2
 - EM waves, 11.1.2
 - radar detection of inhomogeneities, 14.1
 - relative, 10.4.1
 - turbulence, 25.1.1
- Refractivity
 - of air, 25.1.1
 - spectrum, 11.1.2
- Remote sensing, 2.14, 12.26, 30
 - advantages, 30.1
 - under precipitation conditions, 30
 - acoustic, 30.3
 - optical, 30.5
 - radio, 30.4
- Remote versus in-situ measurements, 20
- Resolution, 26.1.1
- Resonance factor, 10.2.2
- Resonance reradiation, 10.2.2

- lidar, 24.2.3
- Resonant absorption, 23.1.6
- Reynolds averaging, 1.7.4, 2.4, 2.6, 2.7
- Reynolds number, 1.8.2
 - stresses, 1.8.2, 3.3.4, 6
- Richardson number, 2.11, 2.21, 3.5.2, 6.4.1
- Richardson stability criterion, 7.1
- Risk, 29.1.4
- Rocky Mountains, 2.21
- Role of remote sensing in the troposphere, 30.1
- Rolls, 6.5.5
- Rosby number, 6.3.2
 - waves, 3.3.1
- Rotational Raman scattering, 23.4
- Roughness, surface, 6.3.1
- Runs, test (independence), 29.7.1

- Salinity, 22.2.3
- Sampling theory, 28A1
- Sargasso Sea, 22.3.1
- Satellite
 - altimeter, reflection from sea, 12.5.2
 - bistatic HF sea scatter experiment, 12.2.3
 - detectors, 22.1.1
- Saturation
 - of scintillation, 25.2
 - of spectral lines, 9.4.4
- Scale of atmospheric motion, 2.2, 1.6.4
- Scaling parameters, 6.5.3
- Scatter, acoustic, 19.1
- Scattered field, 29.2.2, 29.6, 29.6.1
- Scattering, 29.1, 29.2.2, 29.4, 29.6, 29.8
 - acoustic, 18.2
 - acoustic waves, 11.2.1
 - clear air, by radar, 14.1
 - cross section, 23.1
 - by aerosols,
 - as a function of wavelength, 23.4
 - as a function of scattering angle, 23.4
 - as a function of refractive index, 23.4
 - echo area, 8.6
 - of sea, average, 12.2.1
 - electromagnetic wave, 11.1
 - formulas, summary, 10.6.1
 - single step (two photon), 10.3
 - sound waves, 18.2
 - two step (single photon), 10.2
- Scintillations, 25
 - covariance of, 25.3.2
 - noise on long path spectra, 24.1.1
 - power spectrum of, 25.3.1
 - production by turbulence, 25.1.2
- Sea
 - foam, 22.2.4
 - roughness, 22.2.4
 - salts, 22.3.2.1
 - scatter, radar, MF/HF, 12.2
 - state, 12.1.1, 22.2.4
 - surface, description of, 12.1
 - temperature, 22.1.1
 - waves, infrasound from, 21.5
- Secchi disc, 22.3.1
- Secondary flow, 6.5.5
- Second-order sea scatter, 12.3
- Selection rules, 9.1.3
- Semidiurnal tide, 5.9
- Severe storm, 17.3
- Severe weather, infrasound from, 21.5
- Shear, 6.3.1
 - production, 6.2.4
 - of energy, 6.5.3
- Shearing stress, 1.8.2

- Signal
 - detection, 29.1.1, 29.1.4, 29.8.1
 - extraction, 29.1.1, 29.1.4
 - field, 29.2.1, 29.4, 29.4.1
 - vector, 29.1.3
- Signal-to-noise ratio, 19.4.1
- Similarity, 6.3.2
- Simple cost function, 29.1.5
- Simultaneous interaction, 10.3
- Single link system, 29.1.1
- Single scattering, 23.4
- SIRS (Satellite Infrared Radiation Spectrometer), 26.2.1
- Size distribution of aerosols, measurement of, 23.4
- Size parameter, Mie scattering, 10.4.1
- SKYLAB, 26.1.1
- Skywave sea backscatter, 12.2.2
- Smoothing methods for inversion, 16.3.1
- SMS (Synchronous Meteorological Satellite), 26.2.2
- Snell's law, 8.4
- Solar power, spacecraft, 26.1.1
- Solar source (for spectroscopy), 24.1.1, 24.1.2
- Sonic boom, pressure signature from, 21.5
- Sound, 21.1
- Source function, 17.2.1
- Source term, equation of motion, 6.2.1
- Space-time sampling, 29.9, 29.9.1, 29.10
- Specific humidity, 2.19
- Spectra
 - atmospheric constituents, 9.7
 - electronic, 9.1
 - rotational, 9.1
 - vibrational, 9.1
- Spectral density function, 28A4
 - interference, 24.1.1
 - radiance, 22.3.2
 - radiant emittance, 22.1.2
- Spectrum
 - analysis, 2.14, 2.23, 24.1.1
 - function, 2.8
 - Kolmogorov, 11.1.4
 - locus, 22.3.1.1
 - one-dimensional, 2.2, 2.19
 - pressure fluctuation, 21.5
 - three-dimensional spatial, 11.1.2
 - turbulent, 2.14
 - two-dimensional spatial, 11.3.1
- Specular-point theory, 12.5.1
- Spheres, absorbing, 8.6
- Spin
 - electron, 9.4.5
 - nuclear, 9.4.5
- Squall line, 2.8, 2.12
- Stability, 1.5.2, 6.7.2
 - parameter, 6.4.1
- Stable planetary boundary layer, wind structure, 7.1
- Stages in development of remote sensing system, 30.2
- State equation, 6.2.3
- State vector, 10.1
- Static stability, 3.5.1
 - equation of, 1.5.2
- Stationary phase, method of, 8.6
 - processes, 28A4
 - states, 10.1.2
- Statistical communication theory, 29.1, 29.1.4, 29.10
- Statistical data analysis, 29.0, 29.0.2, 29.7
- Statistical inversion techniques, 16.3.2
- Statistical stationarity, homogeneity, & isotropy, 4.2.2
- Statistical validity, 29.7.3
- Status of remote sensing, 30
 - acoustic, 30.3
 - advantages of, 30.5
 - optical, 30.5

- radio wave, 30.4
 - under precipitation conditions
 - acoustic, 30.3
 - optical, 30.5
 - radio, 30.4
- Stochastic process, 1.7.2, 29.5, 29.5.1, 29.5.2
- Strength of spectral lines, 9.4.3, 9.6
- Structure constant
 - refractive index, 11.1.4
 - temperature, 11.2.3
 - acoustic echo-sounding of, 19.1, 19.5.5
 - velocity, 11.2.3
 - acoustic echo-sounding of, 19.1, 19.5.5
- Structure function of turbulence, 4, 11.1.2
- Sub-grid scale, parameterization, 1.5.3
- Sub-grid term, 2.2
- Subsonic waves, 21.2
- Substantial derivative, 1.7.1, 4
- Surface layer, 6.3.1
 - roughness effects, 4.3.1
 - stress, 6.3.1
- Surface-wave sea backscatter, 12.2.2
- Swell, 12.1.1
- Tangent plane coordinate system, 1.8.1
- Taylor dissipation length, 4.2.4
- Taylor hypothesis, 1.7.3
- Temperature
 - atmospheric profiles of, 15.3.1, 23, 26.1.1
 - fluctuations in atmospheric turbulence, 25.1.1
 - gradient, measurement by beam curvature, 25.5
 - inversions, 20
 - acoustic detection, 18.4.2
 - lapse rate, definition, 1.5.2
 - microwave measurement of, 15.3.1, 15
 - optical measurement of, 25.4
 - profile measurement, 15.3.1, 23, 26.1.1
 - structure constant, 19.1, 19.5.5
 - acoustic echo-sounding of, 19.1, 19.5.5
 - effect on gravity waves, 5.7
 - function, 4.3.3
 - parameter, 18.2
 - of turbulence, 25.1.1
 - turbulence, 25.1.1
- Terrain effects, 6.7
- Thermal atmospheric boundary layer, 4.3.2
- Thermal plumes, 19.3, 19.5.5
- Thermal wind, 1.6.2
- Thermal wind equation, 3.2.2
- Threshold, 29.1.4
- Thunder, infrasound from, 21.6
- Thunderstorm, 17.1, 17.2, 17.3
 - electricity, 17.1
- Tides, 5.9, 5.10, 5.11
- Time series analysis, 11.1.2
- TIROS, 22.1.1, 26.2.2
- Tornado, 2.1, 17.3
- Trace velocity, 5.8, 21.4
- Transition probability, 9.1.2, 10.1.2
- Transmission
 - capacity, 28.2
 - of radiation, 9
- Transport equation, 8.4
- Trichromatic coefficients, 22.3.1.1
- Tristimulus values, 22.3.1.1
- Tropopause, 1.5.2
- Troposphere, 1.5.2, 6
- Turbidity
 - definition of, 23.4
 - measurement of, 23.4
 - index, 23.4
- Turbulence, 1.8.2, 2.20, 3.3.4, 4, 6, 19.1, 19.2, 19.5.5
 - aircraft measurements, 14.6
 - clear air, as revealed by radar, 14.6
 - convective storm, 21.6
 - detection of clear air, 14.6
 - effects on pressure sensors, 21.3
 - equilibrium state, 2.14
 - "fossil", 2.20
 - homogeneous, 2.8
 - inertial subrange of, 2.2
 - inner scale of, 11.1.4
 - intermittent, 2.14
 - Kolmogorov description, 4, 11, 25.1.1
 - locally homogeneous, 11.1.2
 - optical propagation through, 25
 - outer scale of, 11.1.4
 - remote sensing of, 13, 14, 17, 18, 20, 25.2
 - structure function, 4, 11.1.2, 11.1.4, 11.2.3, 19.1, 19.5.5, 25.1.1
- Turbulent energy
 - distribution, 6.5.3
 - equation, 6.2.4
- Tyndall scattering, 10.0
- Unconditional maximum likelihood estimator, 29.1.5
- Undulance, 2.20, 2.21ff
- Unimolecular reactions, 27
- Unstable boundary layer, 6.5
- Urban meteorology, 19.5.5
- Variance, 2.4
 - of Doppler spectrum, 13.2.3.2
- Vector potential, 6, 10.1.2
- Velocity
 - azimuth display, 13.2.3, 13.2.3.1
 - defect law, 6.3.2
 - group, 2.21, 2.24
 - phase, 19.1, 2.27, 2.34
 - structure constant, 19.1, 19.5.5
 - acoustic-echo sounding of, 19.1, 19.5.5
 - structure parameter, 18.2
- Vertical beam, 13.2.2, 13.2.4.1
 - distribution of water vapor from satellite infrared spectrometer measurements, 15.3.3
 - ozone profiles, determination of, 16.1, 16.4
 - temperature profiles, determination of, 16.1, 16.4, 23
 - water vapor profiles, determination of, 16.1, 16.4, 23
- VHHR (Very High Resolution Radiometer), 22.1.1, 26.2.2
- Virtual temperature, 1.3
- Viscosity, 1.8.2, 6.7.1
- Viscous terms, equation of motion, 6.2.1
- Visibility
 - distance, definition of, 23.4
 - measurement of, 23.4
- VISSR (Visible Infrared Spin Scan Radiometer), 26.1.1, 26.2.2
- Volume backscatter coefficient, 23.1.1
- Von Karman constant, 6.3.1
- Vortex, wing-tip, 2.17
- Water vapor density measurements, 23.3.1
- Wave
 - baroclinic, 2.11
 - cyclone-scale, 2.12
 - equation, 29.2, 29.3, 29.4
 - form, 29.4, 29.4.1
 - function, 2.5, 2.19, 2.21, 10.1
 - gravity, 2.7, 2.23ff, 2.28, 5
 - inertia, 2.19, 2.20ff
 - height (of ocean)
 - distribution, 12.1.2
 - rms, 12.1.1, 12.1.5
 - significant, 12.1.1
 - spectrum, 12.1.4

- identification, 21.4
- lee, 2.5, 2.19, 2.21
- number, 2.2, 2.7
- planetary, 2.13, 2.22
- slope (of ocean)
 - distribution, 12.1.2
 - rms, 12.1.5
- Waves
 - acoustic-gravity, 21.1
 - atmospheric, 3, 5, 6, 21.1
 - deep water, 12.1.1
 - gravity, 5, 21.2
 - instabilities, 5.4
 - nonlinear, 5.3
 - plane, 5.2
 - Kelvin-Helmholtz billows, 14.5
 - radar detection of, 14.5
 - Rossby, 2.19
 - subsonic, 21.2
- Weather, definition, 1.5.3
- Weather radar, 21.6
- Weighting functions, 16.2
 - for scintillations, 25.1.2, 25.2.4, 25.3.4
- Wind, 18, 23, 26.2.1
 - measurements, acoustic, 18.5
 - nocturnal jet, 18.5
 - noise, 21.3
 - remote measurement of, 25.3
 - structure, effect on gravity waves, 5.8
 - vertical, 18.5
 - waves, 12.1.1
- WKB method, 8.4
- Yellow substance, 22.3.2.2

2

PENN STATE UNIVERSITY LIBRARIES



A000070945492