A Global Three-Dimensional Multivariate Statistical Interpolation Scheme

A. C. LORENC[1]

European Centre for Medium Range Weather Forecasts, Reading, Berkshire, England

(Manuscript received 3 September 1980, in final form 5 December 1980)

ABSTRACT

A three-dimensional statistical interpolation method, multivariate in geopotential height, thickness and wind, is described. The method has been implemented in the ECMWF operational global data-assimilation scheme, used for routine forecasting and for producing FGGE level III-b analyses. It is distinguished by the large number of data used simultaneously, enabling full exploitation of the potential of the three-dimensional multivariate technique, and by the incorporation of a statistical quality control check on each datum using the analysis itself. Some simple examples illustrating the properties of the method are presented, and a detailed study is made of the effect of various analysis parameters on one practical example.

1. Introduction

The analysis method described in this paper has been developed at the European Centre for Medium Range Weather Forecasts (ECMWF) as part of a data-assimilation scheme designed to provide initial states for the Centre's operational forecast model, and to produce analyses from the observations made during the First GARP Global Experiment (FGGE). The scheme must produce global analyses in numerical form, efficiently, with minimal human intervention, on a large fast vector-processing computer. The mass and wind analysis is a vital part of the assimilation. It has to combine a relatively accurate forecast "first-guess" provided by the data assimilation so far, with observations of many different types of accuracies and distributions, to detect and reject erroneous data, and form an analysis which exhibits the approximate balance between mass and wind fields observed in nature. An earlier version of the analysis was described by Lorenc *et al.* (1977). This paper is intended as a description of the analysis method; subsequent papers are planned documenting its behavior as part of an operational data-assimilation scheme.

A technique for combining first-guess and observed quantities in a way consistent with the estimated accuracy of each is optimum interpolation (Gandin, 1963). This method has been extended to the multivariate analysis of geopotential height and wind data by Rutherford (1973) and Schlatter (1975). Rutherford (1976) describes a split vertical and horizontal scheme capable of using height, temperature and wind data without fully taking into account their three-dimensional structure. Bergman (1979) describes a scheme using temperature and wind data three dimensionally. The method presented here is a fully three-dimensional version of optimum interpolation, multivariate in wind, height and thickness. It is thus capable of using, to good effect, widely differing types and distributions of data to produce a consistent three-dimensional height and wind analysis. For instance, the vertical and horizontal height gradient information provided by a sequence of satellite temperature soundings can be combined with wind data or reference level height data at any level.

Quality control of the data is an important function of an automatic analysis scheme. The statistical techniques of optimum interpolation lead naturally to a method of detecting data which are unlikely to be correct. Such a method is an integral part of the ECMWF scheme.

Since several similar versions of the interpolation equations are used, their derivation is grouped together in Section 3, after a description of the less mathematical aspects in Section 2. The prediction error covariance model, whose properties largely determine the local behavior of the analysis method, is described in Section 4 along with the assumed observational error covariances. Some simplified examples illustrating the potential of the three-dimensional multivariate method are presented in Section 5 and results obtained with real data in Section 6. In Section 7 computational aspects of the scheme are discussed and a summary is given in Section 8.

[1] Current affiliation: U.K. Meteorological Office, London Road, Bracknell, England.

0027-0644/81/040701-21$09.25
© 1981 American Meteorological Society

2. Description

a. Coordinates

The basic vertical coordinate used in the analysis is pressure. This simplifies the specification of consistent prediction error covariances and the processing of upper air data, most of which are reported at standard pressure levels. In order to reduce the effect of the coordinate transformations necessary when the analysis is part of a data assimilation based on a sigma-coordinate model, the analysis increments are interpolated from pressure to sigma coordinates before being added to the forecast first guess. Thus the smoothing effect of the transformations is minimized and in data voids the forecast is completely unchanged. At present the 15 standard pressure levels from 1000 to 10 mb are used. The analysis variables are geopotential height and wind components at these levels, and geopotential thickness between them. Since the forecast model usually used to provide a first guess has realistic boundary-layer parameterizations and a high vertical resolution near the surface, fields of surface pressure, 10 m wind and 2 m temperature are maintained in parallel with the pressure coordinate representation, and used in the processing of surface observations.

The program organization, data selection, and the bulk of the computation are independent of the horizontal analysis grid, which can thus be chosen to suit the forecast model used for data assimilation. At present a regular 1.875° latitude-longitude grid is used.

The analysis requires fields of the estimated prediction error, and produces estimates of the analysis error. The former is required to be smooth since its derivatives are ignored, while the latter is expensive to compute, so a much coarser 6° quasi-homogeneous horizontal grid is used for these, with grid points at the center of each of the analysis volumes described in Section 2c.

b. Processing of observed data

The analysis equations are derived assuming that all data are processed to give observed minus first-guess deviations for wind components at, geopotential height of, and the thickness between, the analysis levels. To obtain these, the first guess is interpolated bilinearly in the horizontal and if necessary observations are interpolated or extrapolated in the vertical to the nearest analysis level. Surface observations are processed using the first-guess surface fields. Sea level pressure deviations are converted to height deviations and assigned to the analysis level nearest the observed pressure, and 10 m wind deviations are assigned to the pressure level nearest the forecast model's surface. Apart from this use as the first guess of a surface wind from a forecast model including surface friction, no other correction is made to surface winds for frictional effects.

c. Data selection and organization

A novel feature of the ECMWF analysis scheme is the organization of the data selection and solution of the statistical interpolation equations. The scheme was designed for a vector processing computer especially suitable for the efficient solution of large linear systems of equations. On the other hand, the logical operations required for selecting only the best data in order to keep the systems small do not exploit the computer's full speed. Indeed, the design of algorithms to decide which data are best in a three-dimensional multivariate analysis is extremely difficult, as illustrated in Section 5b. So instead of carefully selecting a few data, typically 8–15 values, as in other analysis schemes (e.g., Rutherford, 1976; Schlatter, 1975; Bergman, 1979), the ECMWF scheme uses most nearby data, usually between 100 and 200 values. This enables the full potential of the analysis method to be exploited, since within such a large number of data it is possible to include height, wind and thickness data for several layers of the atmosphere. It is neither necessary nor practicable to set up and solve the large systems of equations this entails for each analysis grid point and variable. Instead this is done for analysis volumes about 660 km square and, in data-rich areas, a third of the atmosphere deep. In data-sparse areas the full depth of the atmosphere is done in one volume. The same selection of data and solution of the equations is used to check all the data within a volume and to evaluate the analysis for height and wind in all analysis grid points in it. In fact, in order to avoid discontinuities at analysis volume boundaries (particularly important for derived quantities such as divergence), the analysis is also evaluated for surrounding grid points as far as the center of neighboring analyses volumes. Several analyzed values are thus obtained for each point which are then averaged with weights varying linearly from 1 at the volume center to zero at a neighboring volume center.

The data selection is done in two stages: a pre-analysis data organization to eliminate clearly erroneous or redundant data, and the selection of data for each analysis volume. In the pre-analysis observations are sorted into boxes of 6° latitude by ~660 km longitude (identical in the horizontal to the analysis volumes). All data are transformed to deviations from the first guess, normalized by the estimated first-guess error. Redundant data are eliminated by first searching for pairs of observations within each box closer than ~150 km. ("Observation" is used for the collection of data at one horizontal location, "datum" for a single observed

value.) Each such pair is compared level by level, variable by variable, and if they agree they are combined into a single "super-observation" which is thereafter treated as an ordinary observation of somewhat greater accuracy.

The selection of data for each analysis volume is based on the principle that most nearby data remaining after the pre-analysis will be used. There are three steps: selection of data boxes, selection from these of observations, and selection from these of data. The criteria used are as follows:

(i) All boxes within or neighboring the analysis volume are selected. If these contain insufficient data then the neighbors of the box containing least data also are selected, and so on.

(ii) All observations within the analysis volume and its neighboring data boxes are selected. Beyond that only observations containing data for levels and variables with insufficient data so far are chosen.

(iii) All data within the analysis volume are selected. Beyond that data are selected in order of distance, up to a maximum for each level and variable and an overall maximum (currently 191).

d. Quality control of data

Checks are made at three stages during the analysis process. First, before the analysis starts, reports are checked for correct code formats, internal consistency and climatological reasonableness. Second, in the pre-analysis, a comparison is made with the forecast field, and within each pair of data considered for combination into a super-observation. These checks are designed to discard any clearly erroneous data. Data which appear unlikely to be correct, but which possibly could be, are retained for rechecking, but flagged so that they are not used to check other data.

The final check is made using the modified statistical interpolation equations derived in Section 3c. For each analysis volume all data within the volume and all selected data yet to be finally checked are compared with an interpolated value obtained not using the datum being checked or any flagged data. If the absolute value of this deviation is more than four times the estimated interpolation error then the datum is considered to have failed. If more than one fail, then the worst is flagged and the rest of the failures are rechecked not using it. Flags on data within the analysis volume are permanently recorded; those on data in volumes yet to be finally checked are only temporary. At this stage, to save computation, the grid-point analysis for points within the analysis volume can be made. Alternatively, a second scan can be made to produce the grid-point analysis using the final flags for all data.

3. Statistical interpolation equations

Three different modifications of the well-known optimum interpolation technique are used in the analysis: to form super-observations, to check data, and to produce grid-point values while excluding rejected data. Because of the assumptions made when modeling the prediction and observation errors and correlations, and in the linear least-squares theory, the interpolation is not truly optimum and the term statistical interpolation is preferred. It is convenient to group the derivations of the three sets of equations together in this section, after first repeating the derivation of the basic statistical interpolation equations.

a. Notation and basic method

The statistical techniques used are independent of the actual variables observed or interpolated, so I use here a notation which does not explicitly differentiate between them, allowing subscripts to range as appropriate over all observed or analyzed values whatever their positions, level or variable type. Thus B_i is any observed datum selected for the analysis and A_k any analyzed value within the analysis volume.

For all observed or analyzed values, the existence of predicted (first-guess) values P_i, P_k and true values T_i, T_k is assumed, the last being the value we wish to estimate in the analysis. Note that T is not necessarily the actual true value, since we do not wish to analyze atmospheric features below a certain scale.

Deviations from this "true" value are denoted by lower case letters

$$a = A - T, \tag{1a}$$

$$b = B - T, \tag{1b}$$

$$p = P - T. \tag{1c}$$

All analyzed, observed or predicted values have associated error estimates E defined by

$$E^a = \langle a^2 \rangle^{1/2}, \tag{1d}$$

$$E^0 = \langle b^2 \rangle^{1/2}, \tag{1e}$$

$$E^p = \langle p^2 \rangle^{1/2}, \tag{1f}$$

where angle brackets indicate an average over a large ensemble of similar realizations. It is convenient to derive equations in dimensionless form, and to have symbols for deviations from the prediction, so I define

$$\alpha = a/E^a, \tag{1g}$$

$$\beta = b/E^0, \tag{1h}$$

$$\pi = p/E^p, \tag{1i}$$

$$q = (B - P)/E^p, \tag{1j}$$

$$r = (A - P)/E^p, \tag{1k}$$

$$\epsilon^0 = E^0/E^p, \tag{1l}$$

$$\epsilon^a = E^a/E^p. \tag{1m}$$

All the above take subscripts i (or j) ranging over all observed values, or k ranging over all analyzed values, whatever their position level or variable.

The basis of the statistical interpolation method is that the analyzed deviation from the prediction is given by a linear combination of N observed deviations

$$r_k = \sum_{i=1}^{N} w_{ik} q_i, \tag{2}$$

with the weights (w) determined so as to minimize the estimated analysis error E_k^a.

Substituting (1) in (2) gives

$$\alpha_k \epsilon_k^a = \pi_k + \sum_{i=1}^{N} w_{ik}(\beta_i \epsilon_i^0 - \pi_i). \tag{3}$$

Squaring (4) and taking the ensemble average gives

$$(\epsilon_k^a)^2 = 1 + 2\sum_{i=1}^{N} w_{ik}(\langle \pi_k \beta_i \rangle \epsilon_i^0 - \langle \pi_k \pi_i \rangle) + \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ik}(\langle \pi_i \pi_j \rangle + \epsilon_i^0 \langle \beta_i \beta_j \rangle \epsilon_j^0 - \epsilon_i^0 \langle \beta_i \pi_j \rangle - \langle \pi_i \beta_j \rangle \epsilon_j^0) w_{jk}. \tag{4}$$

These summations are conveniently manipulated using a vector and matrix notation:

$$\mathbf{w}_k = [w_{ik}], \tag{5a}$$

$$\mathbf{h}_k = [\langle \pi_k \pi_i \rangle - \langle \pi_k \beta_i \rangle \epsilon_i^0], \tag{5b}$$

$$\mathbf{q} = [q_i], \tag{5c}$$

$$\mathbf{M} = [\langle \pi_i \pi_j \rangle + \epsilon_i^0 \langle \beta_i \beta_j \rangle \epsilon_j^0 - \epsilon_i^0 \langle \beta_i \pi_j \rangle - \langle \pi_i \beta_j \rangle \epsilon_j^0]. \tag{5d}$$

Eqs. (2) and (4) then become

$$r_k = \mathbf{w}_k^{\mathrm{T}} \mathbf{q}, \tag{6}$$

$$(\epsilon_k^a)^2 = 1 - 2\mathbf{w}_k^{\mathrm{T}} \mathbf{h}_k + \mathbf{w}_k^{\mathrm{T}} \mathbf{M} \mathbf{w}_k. \tag{7}$$

One can now proceed to the derivation of the equation for the "optimum" weights, which minimize E_k^a. Since the ensemble average (angle braces) is assumed to be over a large number of similar realizations with the same estimated errors E, this is equivalent to minimizing the normalized error variance given by (4) or (7). By equating $\partial(\epsilon_k^a)^2/\partial w_{ik}$ to zero for $i = 1, N$ we get a set of linear equations for the weights which give

$$\mathbf{w}_k = \mathbf{M}^{-1} \mathbf{h}_k. \tag{8}$$

The analyzed value and estimated error corresponding to these weights are

$$r_k = \mathbf{h}_k^{\mathrm{T}} \mathbf{M}^{-1} \mathbf{q}, \tag{9}$$

$$(\epsilon_k^a)^2 = 1 - \mathbf{h}_k^{\mathrm{T}} \mathbf{M}^{-1} \mathbf{h}_k. \tag{10}$$

Since $\mathbf{M}^{-1}$ and $\mathbf{q}$ are independent of the point being analyzed it is convenient to evaluate their product once only, to give a vector of analysis coefficients $\mathbf{c}$. Thus for the grid-point analysis the weights $\mathbf{w}_k$ are not explicitly calculated; instead (9) becomes

$$\mathbf{c} = \mathbf{M}^{-1} \mathbf{q}, \tag{11}$$

$$r_k = \mathbf{c}^{\mathrm{T}} \mathbf{h}_k. \tag{12}$$

It is usual to call terms such as $\langle \pi_i \pi_j \rangle$ error correlations and terms such as $\langle p_i p_j \rangle$ covariances, although this is only true if biases such as $\langle p_i \rangle$ are zero. This is not strictly necessary for the above derivation, but if biases are non-zero Eq. (2) is not the best interpolation equation. I shall assume the biases to be zero. It is also usual to neglect the correlations between prediction error and observation error $\langle \pi_i \beta_j \rangle$ (e.g., Bergman, 1979). If an observation type is to be used for which these terms are known to be non-zero then their inclusion is straightforward. However, it seems likely that knowledge of such correlations will be limited to the local value $\langle \pi_i \beta_i \rangle$, and in this case a simple preprocessing technique similar to that of Section 3b can remove the correlation. In the current program all correlations $\langle \pi_i \beta_j \rangle$ are assumed negligible.

b. Super-observation formation

If the standard optimum interpolation technique were used to create the super-observation then the interpolated value would contain information from, and hence be correlated with, the predicted value P_k. This is inconvenient since for ordinary observations it is assumed that $\langle \pi_i \beta_j \rangle = 0$. Thus the interpolation equations used for super-observation formation are modified by imposing a constraint that no information is taken from the local predicted value. The constraint can be expressed as

$$\langle \alpha_k \pi_k \rangle = 0. \tag{13}$$

If we multiply (3) by π_k, take the ensemble average, and use matrix notation, we obtain

$$\mathbf{w}_k^{\mathrm{T}} \mathbf{h}_k = 1. \tag{14}$$

The interpolation error is still given by (7), i.e.,

$$(\epsilon_k^a)^2 = 1 - 2\mathbf{w}_k^{\mathrm{T}} \mathbf{h}_k + \mathbf{w}_k^{\mathrm{T}} \mathbf{M} \mathbf{w}_k. \tag{15}$$

Minimizing this with the constraint (14) gives the following equation for the constrained weights:

$$\mathbf{w}_k = (1 + \lambda)\mathbf{M}^{-1} \mathbf{h}_k. \tag{16}$$

Comparing this with (8) we see that the constraint forces a renormalization of the interpolation weights, with λ defined by substitution back in (14). The

error in the interpolated super-observation is given by substituting (14) and (16) in (15):

$$(\epsilon_k^a)^2 = \lambda. \tag{17}$$

In fact, in our present program super-observation formation is done univariately, level by level, for close pairs of observations only. So for the super-observation formation, but not for the data checking which precedes it, it is sufficiently accurate to assume $\langle \pi_i \pi_j \rangle = 1$. The interpolation then simplifies to a weighted averaging.

c. *Observation check*

The final check on each datum is to compare q_k with an interpolated value r_k using the data selected for the analysis volume. Hence it is appropriate when deriving the equations for this interpolation to minimize the expected variance of the difference between these, rather than the deviation from the true value. Thus instead of (7) we minimize

$$\langle (r_k - q_k)^2 \rangle = (\epsilon_k^0)^2 + 1 - 2\mathbf{w}_k^{\mathrm{T}}\mathbf{m}_k + \mathbf{w}_k^{\mathrm{T}}\mathbf{M}\mathbf{w}_k. \tag{18}$$

If the datum being checked is also used for the interpolation then $\mathbf{m}_k$ is a column of $\mathbf{M}$ and minimizing (18) leads to the trivial result

$$\mathbf{w}_k = \mathbf{d}_k, \tag{19}$$

where $\mathbf{d}_k$ is defined as a vector whose kth element is one and other elements are zero. Since we are trying to interpolate a value for a datum including its observational error, the best value is naturally the datum itself.

What we must do is minimize (18) subject to constraints that certain data (datum k and perhaps other data already rejected) are given zero weight.

If we let l_m ($m = 1$ to n) be a list of these data, the constraints can be written

$$\mathbf{d}_l^{\mathrm{T}}\mathbf{w}_k = 0 \quad (\text{for } m = 1, n). \tag{20}$$

Minimizing (18) subject to these constraints gives

$$\mathbf{w}_k = \mathbf{d}_k + \sum_{m=1}^{n} \lambda_m \mathbf{M}^{-1}\mathbf{d}_l. \tag{21}$$

We now write $\boldsymbol{\lambda}$ for the vector (dimension n) of multipliers λ_m, and $\mathbf{D}$ for the N by n matrix whose mth column is $\mathbf{d}_l$. Eqs. (20) and (21) become

$$\mathbf{D}^{\mathrm{T}}\mathbf{w}_k = 0, \tag{22}$$

$$\mathbf{w}_k = \mathbf{d}_k + \mathbf{M}^{-1}\mathbf{D}\boldsymbol{\lambda}. \tag{23}$$

The multipliers $\boldsymbol{\lambda}$ are given by multiplying (23) by $\mathbf{D}^{\mathrm{T}}$, and using (22) yields

$$\boldsymbol{\lambda} = -(\mathbf{D}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{d}_k. \tag{24}$$

Substituting (22) and (23) in (18) gives

$$\langle (r_k - q_k)^2 \rangle = (\epsilon_k^0)^2 + 1 - \mathbf{w}_k^{\mathrm{T}}\mathbf{m}_k. \tag{25}$$

It should be noted that this estimate of the interpolation error is arrived at assuming that the method, and all the estimated errors and correlations used, are perfect. In practice, however, an allowance for errors in these (currently somewhat arbitrarily set to 0.1) is added. A datum is thus considered to have failed the check if

$$(r_k - q_k)^2 > T^2[\langle (r_k - q_k)^2 \rangle + 0.1]. \tag{26}$$

The tolerance T is currently assigned the value 4.

d. *Grid-point analysis*

In order to be able to use interpolation equation (12), while giving zero weight to data which have been included in $\mathbf{M}$ but subsequently rejected using (26), we need to minimize (7) subject to constraints like those in (22).

Manipulations like those of Section 3c give

$$\mathbf{w}_k = \mathbf{M}^{-1}\mathbf{h}_k - \mathbf{M}^{-1}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{h}_k. \tag{27}$$

Substituting (27) in (6) gives the equivalent of (11)

$$\mathbf{c} = \mathbf{M}^{-1}\mathbf{q} - \mathbf{M}^{-1}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{q}. \tag{28}$$

4. Error statistics

a. *Description*

The specification of reasonable estimates for the prediction error E_i^p and prediction error correlations $\langle \pi_i \pi_j \rangle$ for all the horizontal positions, levels and variables covered by i and j is vital to the statistical analysis method. Not only are they necessary in order that statistical methods used should give reasonably "optimum" weights, but also the covariance functions determine the relationship between the analysis increments for different analysis grid points, levels and variables within each analysis volume. This can be seen from (12), which can be rewritten

$$A_k - P_k = \sum_{i=1}^{N} c_i E_i^p \langle \pi_i \pi_k \rangle. \tag{29}$$

Since within each analysis volume the list of data selected (denoted by subscript i) is fixed, while the values to be analyzed (subscript k) range over all locations and variables, this can be looked on as a linear combination of functions specifying the prediction error covariance with each datum position and variable. Hence any linear relationship obeyed by the functions is also obeyed by the analysis increments.

Three such relationships are valid for the prediction error covariance model described in Section 4b: (i) between geopotential heights at any two levels and the thickness between them; (ii) between stream-function and wind components, implying non-divergence; and (iii) between geopotential height and

streamfunction, approximately equivalent to the geostrophic relationship.

Relationships (ii) and (iii), while being useful fairly accurate descriptions of the local behavior of prediction errors, are not precisely true on larger scales. This is not in practice troublesome, since (29) only holds within each set of selected data. Larger scale deviations from the relationships indicated by observations, for instance, divergences between the wind observations in one set of data and those selected for the neighboring analysis volume, are drawn to by the analysis. The weighted averaging described in Section 2c of analyzed values from neighboring sets of data ensures that such deviations from the relationships occur smoothly, rather than at analysis volume boundaries. The height streamfunction relationship is relaxed in the tropics.

The observation error covariances have less direct impact on the structure of the analysis increments, but are rather the means of describing the properties of each observing system, so they can be used to best advantage. Thus, specifying positive horizontal correlations for satellite temperature-sounding observation errors causes relatively greater weight to be given to temperature gradient information between soundings (Seaman, 1977), and specifying negative correlations for errors at different levels in the same sounding causes greater weight to be given to the observed thickness of the combined deeper layer. Changes in the assumed vertical correlations can have a similar effect to a factor of 2 change in the error variance (Larsen *et al.*, 1977). For the majority of observations the error is assumed to be uncorrelated, and it is the ratios of the estimated observation error variance of those of other observations and the prediction error variance which determine the weight given to each observation.

b. Prediction error covariance model

Since in this section physical relationships between different variable types are considered, it is necessary to modify the notation to indicate the specific variable, horizontal position and level. Thus the single suffix is replaced by three: ϕ, t, ψ, u or v indicating geopotential height, thickness, streamfunction or wind components; i or j indicate horizontal position; and k or l indicate level. The omission of one of these suffices indicates that the value is independent of that characteristic.

1) HEIGHT-HEIGHT COVARIANCES

Following Rutherford (1976) I assume that these can be expressed as a product of errors and separable vertical and horizontal correlations

$$\langle p_{\phi ik} p_{\phi jl} \rangle = E^p_{\phi ik} E^p_{\phi jl} V_{\phi kl} F_{\phi ij}. \tag{30}$$

Since the analysis is made at, and data pre-interpolated to, a predefined set of pressure levels, $V_{\phi kl}$ can be specified as a matrix of vertical correlations between these levels.

Again, following Rutherford I further assume that $F_{\phi ij}$ is homogeneous and isotropic. For simplicity I use a simple Gaussian function of distance r_{ij}, despite the fact that this is not the most appropriate form for climatological correlations (Julian and Thiebaux, 1975), *viz.*,

$$F_{\phi ij} = \exp[-\tfrac{1}{2}(r_{ij}/s)^2], \tag{31}$$

where s is the horizontal correlation scale parameter. More appropriate functional forms could be used when data on the prediction error correlation structure (or equivalently the error power spectrum) become available for a short-range forecast.

2) THICKNESS COVARIANCES

The thickness variable used in the analysis is defined as the geopotential height above the next lower level

$$P_{tik} = R_{\phi ik} - R_{\phi ik-1}. \tag{32}$$

This is sufficient to define all thickness covariances (and hence estimated errors and correlations) in terms of the corresponding geopotential height covariances

$$\langle p_{tik} p_{\text{any}} \rangle = \langle p_{\phi ik} p_{\text{any}} \rangle - \langle p_{\phi ik-1} p_{\text{any}} \rangle, \tag{33}$$

3) STREAMFUNCTION-STREAMFUNCTION COVARIANCES

Although the streamfunction is never directly used in the analysis it is convenient to define streamfunction prediction error covariances, for later use in deriving consistent covariances for wind components. I assume a separable model similar to that for geopotential height, i.e.,

$$\langle p_{\psi ik} p_{\psi jl} \rangle = E^p_{\psi ik} E^p_{\psi jl} V_{\psi kl} F_{\psi ij}, \tag{34a}$$

$$F_{\psi ij} = \exp[-\tfrac{1}{2}(r_{ij}/s)^2]. \tag{34b}$$

4) STREAMFUNCTION-WIND COVARIANCES

For a completely flexible specification of covariances for the horizontal velocity field I should now define velocity potential prediction error covariances. However, since these are in general an order of magnitude smaller than those for streamfunction, I shall follow Schlatter *et al.* (1976) and assume velocity potential prediction errors to be negligible. Covariances involving wind components can then be derived by differentiating the streamfunction covariance function (34). Like Buell (1972) I choose coordinates such that u is the wind component in the direction $i \rightarrow j$, and v is perpendicular to it.

Ignoring horizontal derivatives of E^p_ψ, and assuming $F_{\psi ij}$ to be isotropic, yields the simple wind pre-

diction error covariance model

$$\langle p_{vik}p_{\psi jl}\rangle = -E^p_{\psi ik}E^p_{\psi jl}V_{\psi kl}\frac{\partial}{\partial r_{ij}}F_{\psi ij}, \tag{35a}$$

$$\langle p_{\psi ik}p_{vjl}\rangle = E^p_{\psi ik}E^p_{\psi jl}V_{\psi kl}\frac{\partial}{\partial r_{ij}}F_{\psi ij}, \tag{35b}$$

$$\langle p_{uik}p_{\psi jl}\rangle = 0, \tag{35c}$$

$$\langle p_{\psi ik}p_{ujl}\rangle = 0, \tag{35d}$$

$$\langle p_{vik}p_{vjl}\rangle = -E^p_{\psi ik}E^p_{\psi jl}V_{\psi kl}\left(\frac{\partial}{\partial r_{ij}}\right)^2 F_{\psi ij}, \tag{35e}$$

$$\langle p_{uik}p_{ujl}\rangle = -E^p_{\psi ik}E^p_{\psi jl}V_{\psi kl}\frac{1}{r_{ij}}\frac{\partial}{\partial r_{ij}}F_{\psi ij}, \tag{35f}$$

$$\langle p_{uik}p_{vjl}\rangle = 0, \tag{35g}$$

$$\langle p_{vik}p_{ujl}\rangle = 0. \tag{35h}$$

Using earth-centered Cartesian coordinates it is straightforward to calculate the components of vector $\mathbf{r}_{ij}$ in the local latitude-longitude coordinates at points i and j, and hence the cosine and sine of the local bearings. These are then used to transform the covariances for the wind components in (35) into covariances for northward and eastward components used in the analysis program. This two-stage process for computing the covariances avoids the need to derive (35) in latitude-longitude coordinates. Expressed in any globally valid coordinate system the equations appear much more complicated, without the zero terms apparent in (35).

5) Height-streamfunction covariances

It only remains to couple the height and thickness covariance model of (30)–(33) with the streamfunction and wind model of (34)–(35). I do this by defining a local geopotential height-streamfunction prediction error correlation μ, and defining all covariances between geopotential height and wind in terms of this and the corresponding streamfunction wind covariance:

$$\langle p_{\phi ik}p_{\psi ik}\rangle = \mu, \tag{36}$$

$$\langle p_{\phi ik}p_{ujl}\rangle = \mu\langle p_{\psi ik}p_{ujl}\rangle. \tag{37}$$

By an appropriate choice of values this can be made equivalent to the geostrophic covariance models used by Schlatter (1975) and Rutherford (1976). Thus, if one were to set $\mu = 1$ and

$$E^p_{\psi ik} = (1/f)E^p_{\phi ik}, \tag{38a}$$

$$V_{\psi kl} = V_{\phi kl}, \tag{38b}$$

$$F_{\psi ij} = F_{\phi ij}, \tag{38c}$$

[and noting that in the derivation of (35) horizontal derivatives of $E^p_{\psi ik}$ and hence now of $1/f$ have been ignored], the model would link geopotential height and wind covariances geostrophically. On the other hand, if we set $\mu = 0$, then the geopotential height and wind analyses are completely uncoupled and the relationships (38) unnecessary. In practice μ is a function of latitude.

6) Mathematical constraints

Not all functions possible within the model just described are physically possible covariance functions. A mathematical property of valid functions is that all possible covariance matrices should be positive definite. Necessary conditions for this are that matrices $V_\phi V_\psi$ and functions $F_\phi F_\psi$ individually should be positive definite. This is the case for the Gaussian function used (Julian and Thiebaux, 1975), and such a condition is easy to impose on V_ϕ and V_ψ since they are pre-tabulated. Constraints on the combined geopotential height and wind covariance function depend on the absolute value of μ. For $|\mu| = 1$ then Eqs. (38b) and (38c) are necessary; however, as $|\mu|$ decreases these conditions can be relaxed. I have chosen to retain (38b) outside of the tropics and (38c) everywhere. In the tropics the deviation from (38b) imposes an upper limit on $|\mu|$.

It also is strictly necessary that the separability assumptions made in Eqs. (30), (34) and (37) are adhered to. However, in practice V_ϕ and V_ψ do vary with horizontal position, and the scale parameter s in F_ϕ and F_ψ varies with level and position. Also, μ clearly needs to be a function of latitude. Most non-positive definite matrices which might be caused by this can be avoided by assuming V and μ to be locally constant, that is by using one set of values for all computations associated with one analysis volume. However, vertical variations of s still occasionally lead to non-positive definite matrices. When this occurs the test for ill-conditioning described in Section 7 avoids problems in the analysis.

c. Observation error covariance model

1) Variances

The observation error E^0 is defined by (1e) as the expected deviation from the value we wish to analyze. Hence it contains contributions from variations on scales smaller than those we wish to analyze in both time and space, as well as the purely observational error. Furthermore, allowance must be made for errors introduced when transforming from the parameter observed to that used in the analysis. The quantity which actually enters the analysis equations is the ratio ϵ^0 (1i). Since the prediction error varies more in both time and space than the observation error they are specified independently.

2) CORRELATIONS

There are two sources of observation error correlation: (i) the real small-scale motions which are treated as error, and (ii) the characteristics of the instrument or data processing method used within one type of observation. The former is only significant for very close observations and is only considered when forming super-observations. There its effect is to increase the estimated observation error of the super-observation. The latter needs to be determined and specified for each observing system. The method of modeling the correlations is in principle the same as for the prediction errors. In practice super-observation formation, and the observing systems believed to have significant correlations (radiosonde height observations and satellite temperature soundings), are univariate. Moreover, the variation in time and space is less, and the accuracy with which they are known is poor. Thus considerable simplification is possible.

d. Determination of covariance model parameters

This is still an area of active research, so I shall not present actual values other than those given with the examples of Sections 5 and 6, but instead discuss how estimates may be obtained. The analysis method provides several opportunities for checking the consistency of the covariances used, since every value has associated with it an estimated variance. By accumulating actual and assumed variances the latter can be checked. The estimated variance is almost always a function of several covariance model parameters. By assuming some of these to be correct new estimates can be obtained for others. Thus

$$(\hat{E}^p)^2 = \overline{(B - P)^2} - \overline{(E^0)^2}, \tag{39}$$

where a caret indicates a new estimate and an overbar an average value accumulated over many analyses for a suitable division of season, horizontal position, level, variable and observation type. Eq. (39) follows from Eq. (1) and the assumption that $\langle \pi\beta \rangle$ is negligible, and depends on the error with which E^0 is known being small compared with E^p. Values given by (39) can be biased toward data-rich areas, and corrupted by very erroneous observations. This can be avoided by using the analysis instead of the observations for verification. In this case $\langle a_k p_k \rangle$ cannot be neglected, and one has to assume that the correlations used in the analysis are reasonable. It is easy to show that

$$\langle a_k p_k \rangle = (E_k^a)^2 \tag{40}$$

and hence that

$$(\hat{E}^p)^2 = \overline{(A - P)^2} + \overline{(E^a)^2}. \tag{41}$$

This new estimate is equal to the assumed prediction error in data voids and equivalent to (39) near isolated observations.

For inaccurate observations, for which E^0 is large compared with the error with which E^p can be estimated, instead of (39) one can get a new estimate for E^0. However, in general, E^p is insufficiently well known, and one must use analyzed values for verification when accumulating observation error estimates. This can be done while checking the data by accumulating terms from (26). Since the normalization factor E^p varies in time and space it is convenient to change this to E^0 by dividing by ϵ^0:

$$(\hat{E}^0/E^0)^2 = [\overline{(r - q)^2 - \langle (r - q)^2 \rangle}]/\overline{(\epsilon^0)^2} + 1. \tag{42}$$

Alternatively, observations can be compared with the final analysis made using them. In this case $\langle ab \rangle$ cannot be ignored and we have

$$(\hat{E}^0)^2 = \overline{(B - A)^2} + \overline{(E^a)^2} \tag{43}$$

as the equivalent of (41).

Horizontal prediction error correlations are also fairly easily accumulated (Rutherford 1972) for areas where there is a good coverage of observations with little observational error correlation. Alternatively, independent estimates of geopotential height and wind prediction errors, and the geostrophic relationship (38), determine the second derivative of the correlation function F at zero distance, and hence (for the simple one-parameter Gaussian function used) s. This latter method, however, is sensitive to the shape of the function assumed for F, and should perhaps be regarded as a test of its appropriateness. Vertical prediction error correlations are more difficult to accumulate, since most suitable observations have significantly correlated observation errors. Special studies such as that of Hollett (1975) are one approach. Alternatively, independent estimates of geopotential height and thickness prediction errors provide some off-diagonal elements of V, and some modeling assumptions may be used to fill in the rest.

Observation error correlations are even more difficult to separate from the uncertainties in the prediction or analysis error correlations, and their estimation is best done in separate studies (e.g., Schlatter and Branstator, 1979) with due regard to the correlations to be expected from the specific observation technique.

The height-streamfunction correlation must approach 1.0 and -1.0 in Northern and Southern Hemisphere extratropics, since the geostrophic relationship is approximately valid, and be zero at the equator. The behavior in the tropics may be determined from theoretical or practical studies of the accuracy of the geostrophic relationship (e.g., Bergman, 1979) or the validity of (38). The differences from ± 1 outside of this region will be highly dependent on the details of the data assimilation scheme which affect the gravity wave noise and ageostrophy of the first-guess.

5. Some simple examples

In this Section various properties of the multivariate analysis method in a simplified two-level analysis are illustrated, using the geostrophic prediction error covariance model (38). In each of the examples I take $\mu = 1$, $s = 500$ km, $E^p_{\phi\,1000\text{ mb}} = 18$ m, $E^p_{\phi\,500\text{ mb}} = 21$ m, $\langle \pi_{\phi\,1000\text{ mb}} \pi_{\phi\,500\text{ mb}} \rangle = 0.237$ and f appropriate for 60°N.

a. *Use of vertical and horizontal gradient information*

The steps taken in Section 4 to ensure prediction error covariances consistent with the physical relationship between various variables ensure that the analysis equations, although derived using statistical techniques, precisely reflect these relationships. This is best illustrated by a simple example using error-free data.

We now consider three perfect observations which together contain enough information to specify the 500 mb geopotential height at the analysis point: a wind observation 250 km away at 500 mb, a 1000–500 mb thickness observation 500 km away in the same direction, and a 1000 mb height observation of the same location as the thickness. Columns 1–4 in Table 1 show the estimated analysis error and the interpolation weights for all combinations of these observations. Using none of the observations the estimated analysis error equals the assumed prediction error (used singly or in pairs the observations only reduce this slightly), while all three together reduce the estimated analysis error to only 1.9 m. This small residual error is a measure of the error with which one wind observation specifies the average height gradient over 500 km; by using more wind data the error can be decreased toward zero.

In practice, of course, observations are not perfect while the first-guess prediction is usually good, often having similar accuracy to that of the observations.

TABLE 1. Observation errors and prediction errors for 1000 mb height, 1000–500 mb thickness, and 500 mb wind; and analysis error and weights when using various combinations to analyse a 500 mb height.

	Perfect data				Typical observation errors		
	ϕ	t	v		ϕ	t	v
E^o =	0.0 m	0.0 m	0.0 m s^{-1}	E^o =	7.0 m	29.8 m	3.90 m s^{-1}
E^p =	18.0 m	24.2 m	3.26 m s^{-1}	E^p =	18.0 m	24.2 m	3.26 m s^{-1}
E^a (m)	w_ϕ	w_t	w_v	E^a (m)	w_ϕ	w_t	w_v
21.0	—	—	—	21.0	—	—	—
20.8	0.143	—	—	20.8	0.125	—	—
19.1	—	0.419	—	20.3	—	0.166	—
18.9	—	—	0.441	20.1	—	—	0.182
14.4	—	0.611	0.628	19.1	—	0.191	0.206
18.4	0.192	—	0.461	19.9	0.142	—	0.188
16.7	0.520	0.699	—	19.7	0.225	0.215	—
1.9	0.853	1.147	0.880	18.3	0.262	0.250	0.224

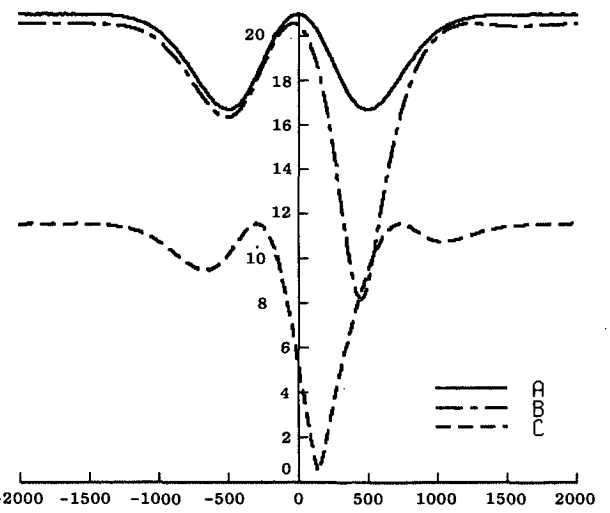

FIG. 1. Estimated analysis error (m) of 500 mb geopotential height as a function of wind datum position for (A) wind only (B) wind plus a height datum at 900 km, and (C) wind plus a height datum at 300 km.

In this case each observation has much less impact, as illustrated in columns 5–8 of Table 1. (This result should not be used directly to judge the impact of particular observing systems because in practice more data are used to analyze each grid point, and because the prediction is only accurate because of the impact of observations in previous analyses.)

b. *The effect of gradient data on data selection*

The last example shows how the analysis method is capable of using vertical and horizontal gradient information from winds and temperature soundings. The usefulness of such information depends greatly on the presence of some absolute observations to act as a reference. Conversely, a distant absolute observation has its usefulness enhanced by suitable gradient information. Data selection algorithms which evaluate each datum independently (e.g., Bergman, 1979) cannot take this effect into account, and may therefore not select useful data. As a further example, Fig. 1 shows the 500 mb height analysis error as a function of the position of perfect 500 mb observations of wind and height. The analysis point and the observations are assumed to be on a straight line. Three curves giving the 500 mb height error as a function of wind observation position are shown: 1) for a wind observation alone, 2) with a height observation at 900 km distance, and 3) with a height observation at 300 km. It can be seen from curve A that a wind observation at the analysis point is useless, the analysis error (21 m) being equal to the prediction error. With a nearby height observation however, such a wind observation is useful. One can see from curve C that it reduces the error from 11.5 m (the error with no wind observation

TABLE 2. Estimated analysis error (m s^{-1}) of a 500 mb wind component for various distributions of error-free 1000–500 mb thickness data (t) and 1000 mb wind data (v).

	Analysis error	Distance of data from analysis point (km)				
		−500	−250	0	250	500
a	3.26					
b	2.96					t
c	2.51	t				t
d	2.37		t		t	
e	0.38		t	v	t	
f	3.17			v		

equals the asymptotic value with the wind very distant) to 5.5 m. A comparison of curves A and B, which almost coincide at most points, shows that a height observation at 900 km distance is almost useless unless there are wind data specifying the intermediate gradient.

When analyzing wind using height data there is a corresponding effect—the effectiveness of a datum is enhanced by the presence of another positioned to enable a gradient to be obtained. With thickness data the effect is further complicated since even an adequate horizontal distribution specifies only the wind shear, and a reference level wind is necessary to get full effect from the data. This is illustrated in Table 2, which shows the estimated analysis error for a 500 mb wind component for various distributions of error-free data. For the prediction error covariances I have assumed the 1000–500 mb thickness which correlates best with the analysis variable is 500 km away, with a correlation of 0.42. It can be seen from row b that such a datum is only marginally useful. With another such observation in the other direction (row c) the gradient is better determined. When calculating gradients the optimum distance is not usually the same as for the data individually; row d shows the effect of data at half the distance. Even this pair determines only the wind shear, and hence is only statistically related to the analyzed wind. The addition of a reference level wind (row e) thus has a dramatic effect, even though the wind alone (row f) is not useful.

These examples illustrate the difficulties of devising simple data selection algorithms to select a few best data for the analysis of each value, a problem avoided by the large order systems and large analysis volumes of the ECMWF scheme.

c. *Sensitivity to prediction error covariance model*

In Section 6 the effect of varying the prediction error covariance model parameters on a practical analysis example is described. However, in such a complex system the individual effect of each parameter is difficult to isolate. For instance when varying the horizontal correlation scale (s) one also has to vary the ratio of height and wind errors according to (38a) and often the largest effects are due to different data being rejected in the automatic data checking, rather than directly due to the covariance model. Moreover, because the accuracy of the first-guess forecast is usually high, the changes made during the analysis are normally small and the effect of varying a parameter on these changes is difficult to display. Thus I present here a simple one-dimensional example representing a 1000 km wavelength wave in the 1000 mb height field, just resolved by nine observations with a 500 km spacing and deviations of alternately +20 m and −20 m from the first guess which is 40 m everywhere. Fig. 2 shows analyses of this wave for three different values of the prediction error scale parameter s, with prediction and observation errors of 18 and 7 m, respectively. Because of the symmetrical situation only half of each analysis is shown. It is clear that, for these error values, analyses made with s greater than about half the wavelength do not resolve the wave, instead averaging the observed deviations to give an analysis almost equal to the 40 m first guess. If the ratio of observation error to prediction error is decreased the fit of the analyses to the observations improves. For example, Fig. 3 shows the analyses with the assumed prediction error doubled. In the extreme case of observation errors assumed to be zero (not shown) all the analyses fit the data exactly, but the overshooting in the data void already apparent in Fig. 3 reaches extreme values of 96 m for s = 500 km and 311 m for s = 700 km. The effects of the assumed error correlations and variances on the analyzed fields are thus interrelated.

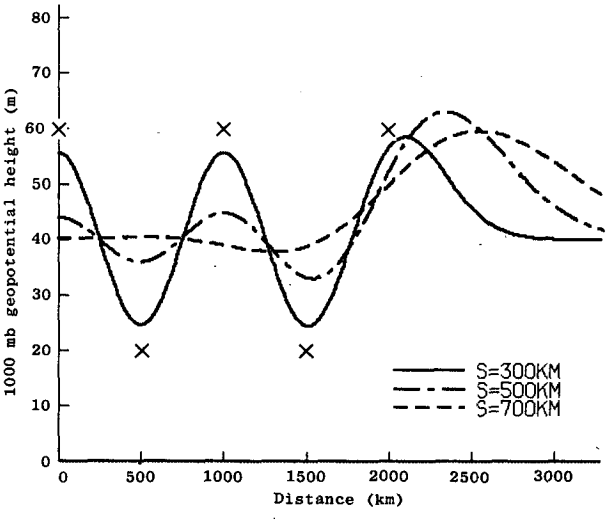

FIG. 2. Analyses of a 1000 km wavelength wave from nine observations marked X for various horizontal prediction error correlation scales (s). Only half of the symmetric situation is shown.

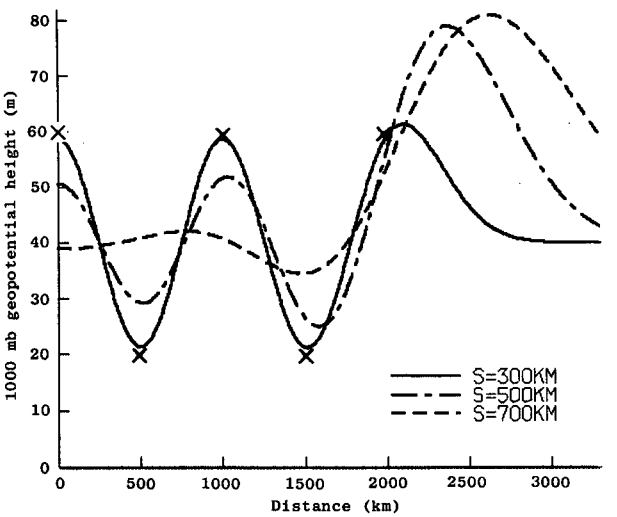

FIG. 3. As in Fig. 2 with estimated prediction error doubled.

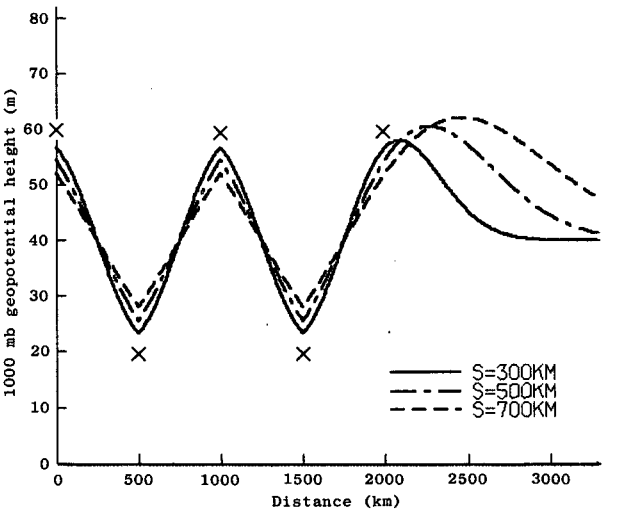

FIG. 4. As in Fig. 2 selecting only the two nearest observations to each point.

Inappropriate values can lead to unjustifiable extremes in the analyses when extrapolating into a data void from a region with observations. The interrelation of appropriate error correlation and variance estimates can be explained since in Section 3 I defined "error" as the deviation from the state which we wish to analyze. Thus, if the prediction model used for data assimilation cannot resolve waves as short as 1000 km we might define the "truth" as the actual state with all waves shorter than 1000 km removed. This would have the effect of increasing the appropriate observation error estimates to include these scales, decreasing the prediction error estimates by the same amount, and increasing the horizontal correlation scale (s) of the prediction errors.

The sensitivity of the analysis to the values specified for the error correlations and variances is enhanced by the large number of data included in the system for each analysis. Fig. 4 shows the same analyses as Fig. 2, but with only the two nearest data to each point used. The effect of varying the prediction error correlation scale (s) is much less dramatic. This can be explained theoretically in terms of the spectral window resolved by the selected observation (Julian and Thiebaux, 1975). The difference between Figs. 2 and 4 is an exaggerated analogy of the difference between the analysis scheme described in this paper and earlier schemes (e.g., Rutherford, 1973; Schlatter, 1975; Bergman, 1979) which use fewer data at once. Despite the large number of data used in each analysis volume in the ECMWF scheme, because of the multivariate three-dimensional distribution of the data and its varying quality the number of accurate data for any one variable in one dimension is never as great as in Fig. 2. However, it does illustrate the fact that the advantage of needing less care in selecting observations noted in Section 5b is balanced by the need for more care in specifying error covariances.

In practice, the largest impact of a change in the error covariances used on the analysis is indirect, via the automatic quality control of data. Fig. 5 shows the analysis comparable with Fig. 2 which result if the algorithm described in Sections 2d and 3c are used. For s = 500 km three data (at 0 km and $\pm$500 km) failed during the first check scan; after the worst failure (the central datum) was rejected, the others passed on the second scan. For s = 700 km five data (at 0 km $\pm$ 500 km and $\pm$1000 km) failed during the first scan. After the data at $\pm$1000 km were rejected the others passed. In general, the behavior

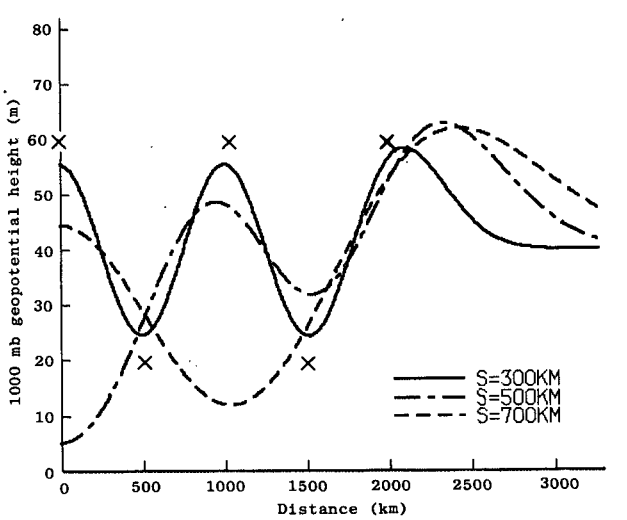

FIG. 5. As in Fig. 2 with automatic data checking.

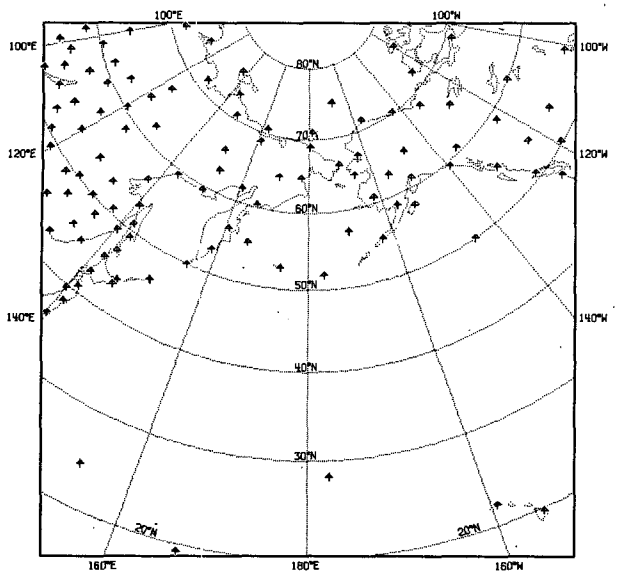

FIG. 6. Radiosonde observation positions 0900–1500 GMT 19 January 1979 in FGGE dataset.

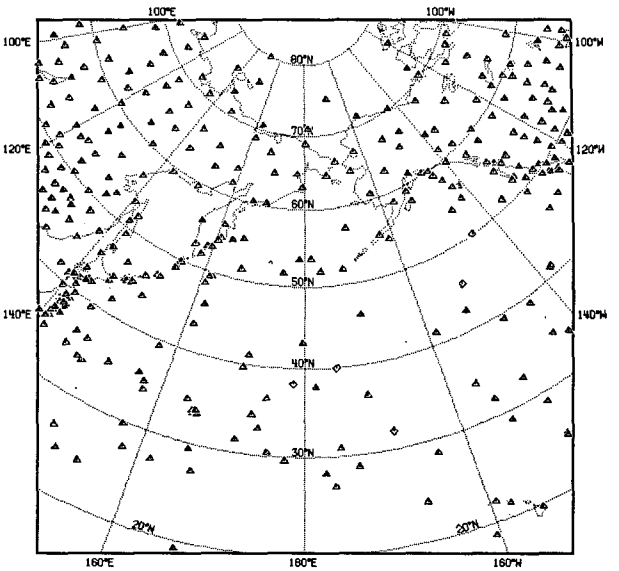

FIG. 8. As in Fig. 6 for surface observations.

of the automatic quality control is very satisfactory, agreeing well with careful subjective judgement and needing no manual intervention. However, when, as in these examples, the available data barely resolve a feature with amplitude large compared to the assumed prediction error, the checking is critical and changes in the assumed covariances can alter which data are rejected, with large effect on the analysis. If the prediction error is increased as in the example of Fig. 3, or if the data density is increased by adding observations of 40 m midway between those originally used, or if the density is effectively decreased as in the example of Fig. 4, then none of the data is rejected. When the data checking is implemented, inappropriate error covariances usually cause good data to be rejected, rather than affecting directly the analysis. Thus the overshooting in the data void noted in Fig. 3 and the extreme overshooting described with zero observation errors does not occur. Instead sufficient data are rejected for the analyses to differ from the first guess by only a few times the assumed prediction error.

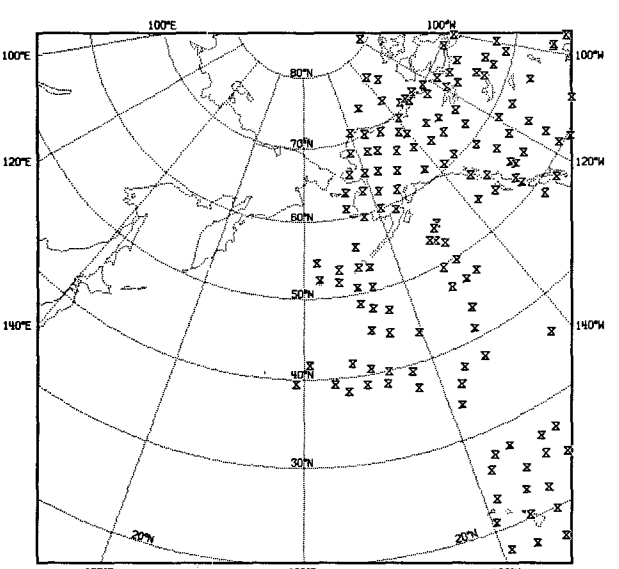

FIG. 7. As in Fig. 6 for satellite temperature soundings.

6. An example analysis

In this section an analysis for 1200 GMT 19 January 1979 made as part of a global 6 h intermittent data-assimilation of FGGE data is described. The horizontal prediction error correlation scale (s) was assumed to be 400 km at 1000 mb, increasing to 1000 km at 10 mb, and the height-streamfunction prediction error correlation (μ) was assumed to be 0.85 north of 30°N. Assumed prediction errors were a function of level, latitude, and of the estimated analysis errors in the previous analysis of the data assimilation cycle.

In order to illustrate the behavior of the analysis method when given a poor first guess, the analysis was repeated using climatology instead of a forecast. Prediction errors appropriate for the atmospheric variability were specified, s was assumed to be 1000 km at all levels and μ was 0.95 north of 30°N.

An active region in the North Pacific was chosen for close study. Figs. 6–9 show the observation positions for the period 0900–1500 GMT and Fig. 10 shows the estimated geopotential height prediction and analysis errors at the center of the region. The 6 h forecast fields of 300 mb geopotential height and

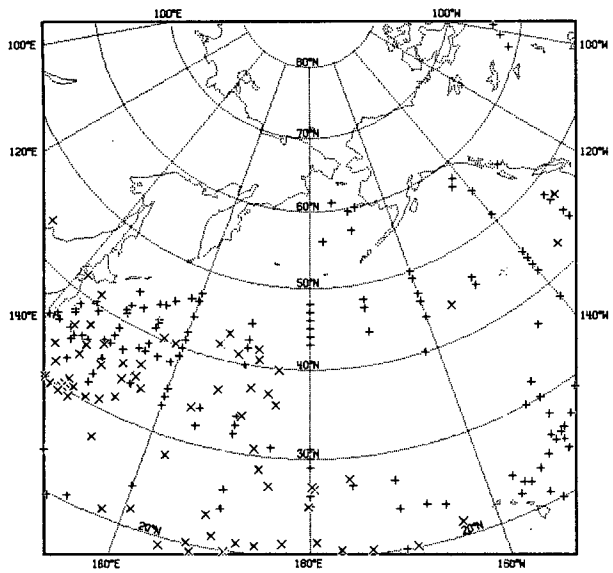

FIG. 9. As in Fig. 6 for aircraft reports and cloud motion winds.

wind speed are shown in Fig. 11, the analysis made from this first guess in Fig. 12, and an independent analysis from a climatological guess in Fig. 13.

By comparing Figs. 11 and 13 one can see that both the forecast and the observations must specify the fields rather well in most areas since differences are limited to details. This is in accordance with Fig. 10, the estimated root-mean-square 300 mb height errors being 42 and 25 m respectively, less than half a contour interval on the plotted charts. The statistical interpolation method merges the two sources of information accordingly; at 40°N, 180°E there is less reliable upper air information than at 50°N (see Fig. 6) and the estimated errors are about equal, so the analysis which has both sources of information available (Fig. 12) takes an average value. The slightly enhanced ridge in this region, associated with slightly more southerly and northerly flows to the west and east and the deeper surface low to be discussed later, is the main difference between the forecast and analyzed fields (Figs. 11 and 12) in the region being studied. The main difference between the standard analysis and that from climatology (Figs. 12 and 13) is the failure of the latter to resolve the sharp trough and wind shear at 160°W, falling as it does in a data-sparse area.

Fig. 14 shows the 6 h forecast sea level pressure and wind field used as first guess and the surface observations. A calibration 10 m s^{-1} arrow for the

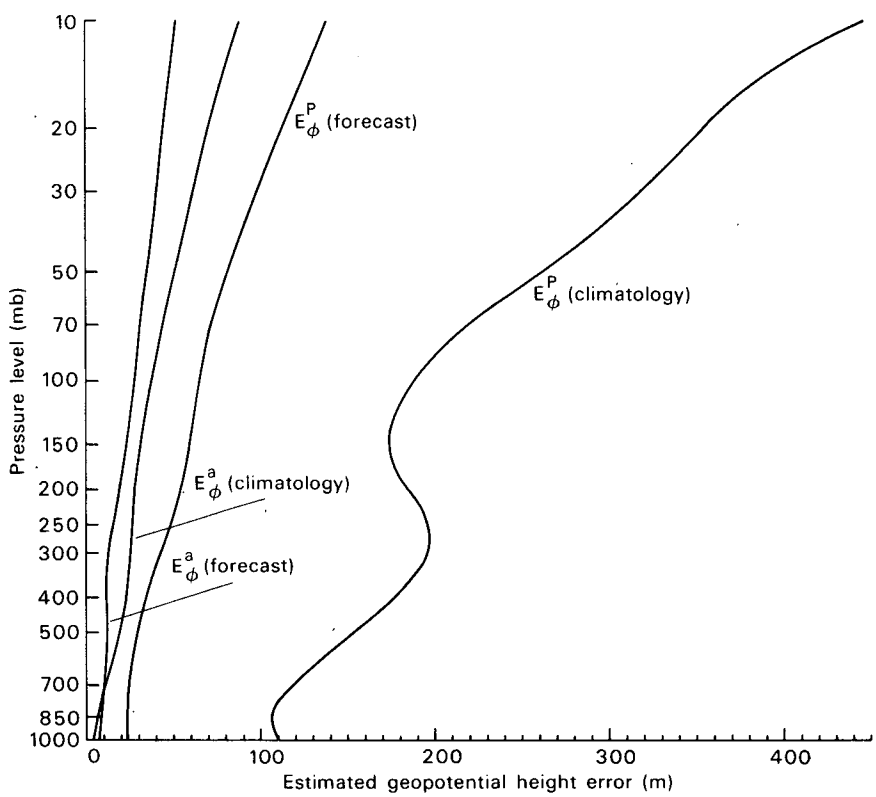

FIG. 10. Estimated geopotential height prediction error (E^p_ϕ) and analysis error (E^a_ϕ) for 50°N, 180°W, 1200 GMT 19 January 1979, for both climatological and forecast first-guesses.

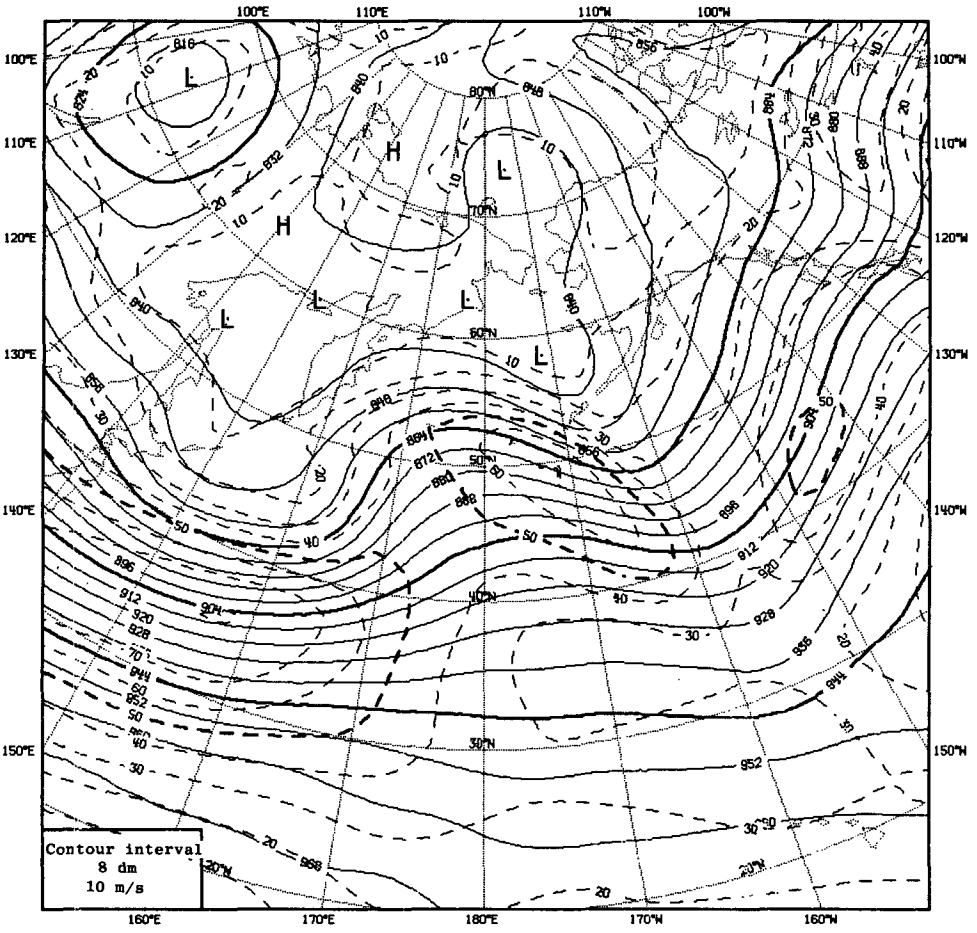

FIG. 11. 300 mb geopotential height and wind speed for 1200 GMT 19 January 1979 from the 6 h forecast used as first guess for the analysis in the data-assimilation cycle.

wind field is shown in the lower left corner. The rapidly developing warm sector depression has not deepened enough during the forecast and has moved too far north, as indicated by the observations near its center. The warm front running east from the center, which is active and clearly defined in the forecast model, is also too far north, as shown by the wind observations at 43°N, 180°E and 38°N, 178°W. There is an almost stationary depression near 60°N, 175°W which is too shallow in the forecast and an active rapidly moving trough to its southeast which is a little too weak. In the analysis (Fig. 15) these errors have been largely corrected, and the analyzed values fit the observations closely, except for some of the topographically affected data in the north. Since the prediction error covariances used were not representative of the error structures that occur when a front is mispositioned, the analysis has not done what a human analyst might have done and moved the frontal trough south, but rather it has drawn a more extended region of curvature. The analysis has moved the 1000–500 mb thickness pattern (not shown) ~150 km southward.

A series of experiments to determine the impact of various analysis scheme parameters on this example showed such subtle changes that inclusion of separate figures is not justified. Removal of the height-streamfunction coupling by setting $\mu = 0$ caused the main low to be 1 mb shallower and the trough in the surface pressures near 155°W to be less pronounced than in Fig. 15, while observations of winds which do not obey the geostrophic relationship, such as those near 52°N, 158°E and 64°N, 179°E, were fitted more closely. Doubling of the estimated prediction errors (E^p) similarly caused these winds to be fitted more closely, as well as those at 46°N, 170°W and 43°N, 180°E. Changes in analyzed pressures were very small except for an intensification of the ridge due to the erroneous acceptance of an observation in the group near 52°N, 175°E. An analysis using no wind data had similar pressure fields, and an analysis using no geopotential data had wind fields

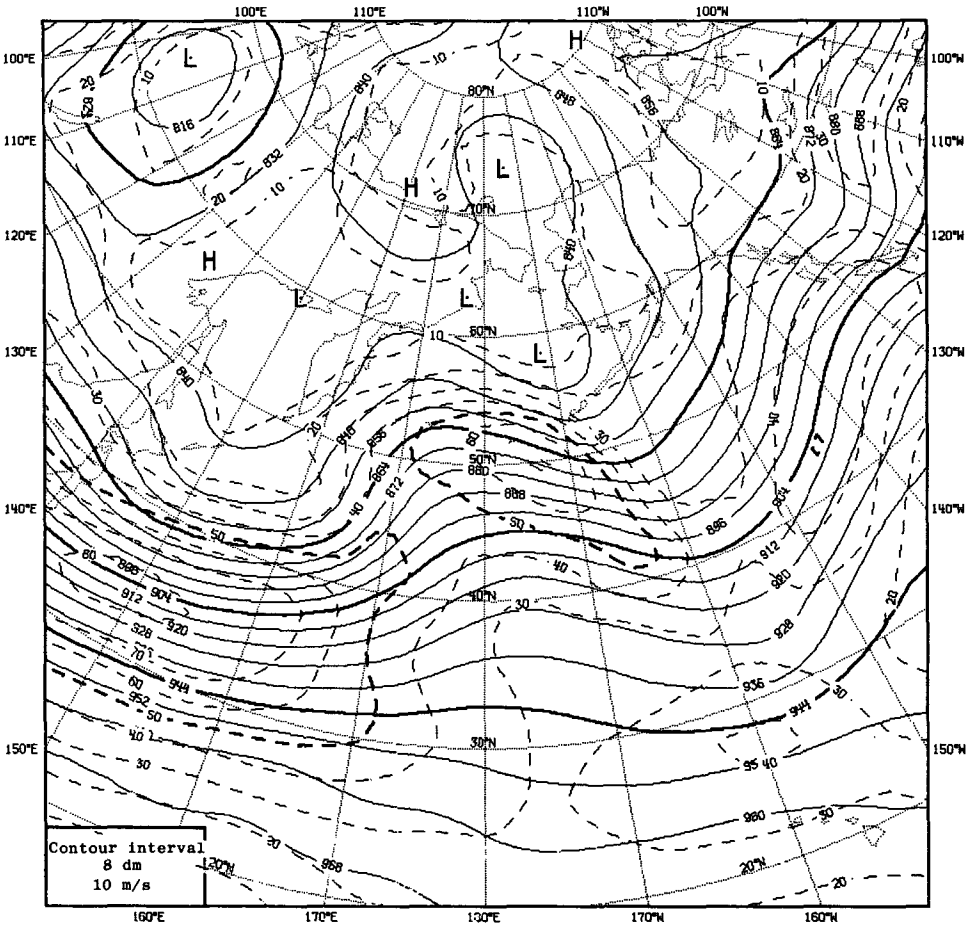

FIG. 12. As in Fig. 11 for the analysis in the data-assimilation cycle.

similar to those mentioned above obtained with $\mu = 0$. The analysis using geopotential data only did not show details in the wind field such as the strong winds west and southwest of the main low and the intensification (compared to the forecast) of the trough to the east; however, the main features of the wind analysis were correct and fitted the wind observations better than the forecast. On the other hand, the analysis using only wind data did not alter the large-scale pressure pattern of the first guess, changes being limited to details such as the trough at 45°N associated with the warm front, the positioning of the low at 58°N and the intensification of the trough to the east. An analysis made with the horizontal prediction error correlation scale (s) doubled, and the wind prediction errors halved to maintain consistency with (35) and (38), had remarkably few differences from that shown in Fig. 15. The central 955 mb contour of the depression enclosed a much smaller area with the lowest pressure in the same position, and the trough at 45°N was less marked. The trough in the east was a little deeper and more rounded, with a separate low indicated near 51°N, 165°W as in the forecast (Fig. 14). The 1000 and 995 mb contours near the coast in the north were a little straighter.

The analysis for the same region from a climatological first guess is shown in Fig. 16. Note that although the first guess was precisely geostrophic the analysis does show some cross-isobar flow in regions where data indicate it. This is because μ is not precisely 1.0, and because the overlapping of analysis volumes with a varying weighted averaging (described in Section 2c) allows gradients in the mass field increments not balanced by wind increments. This analysis was more sensitive to changes in the prediction error correlation model parameters than the analysis of Fig. 15. With $\mu = 0$ the trough in the pressure field near 165°W, already much weaker in Fig. 16 than Fig. 15, was not apparant at all. The main low was deeper, with central pressure 950 mb, and the low to the north had its lowest pressure at 60°N, 180°E. With s reduced to 700 km, and the wind prediction errors increased accordingly, all the

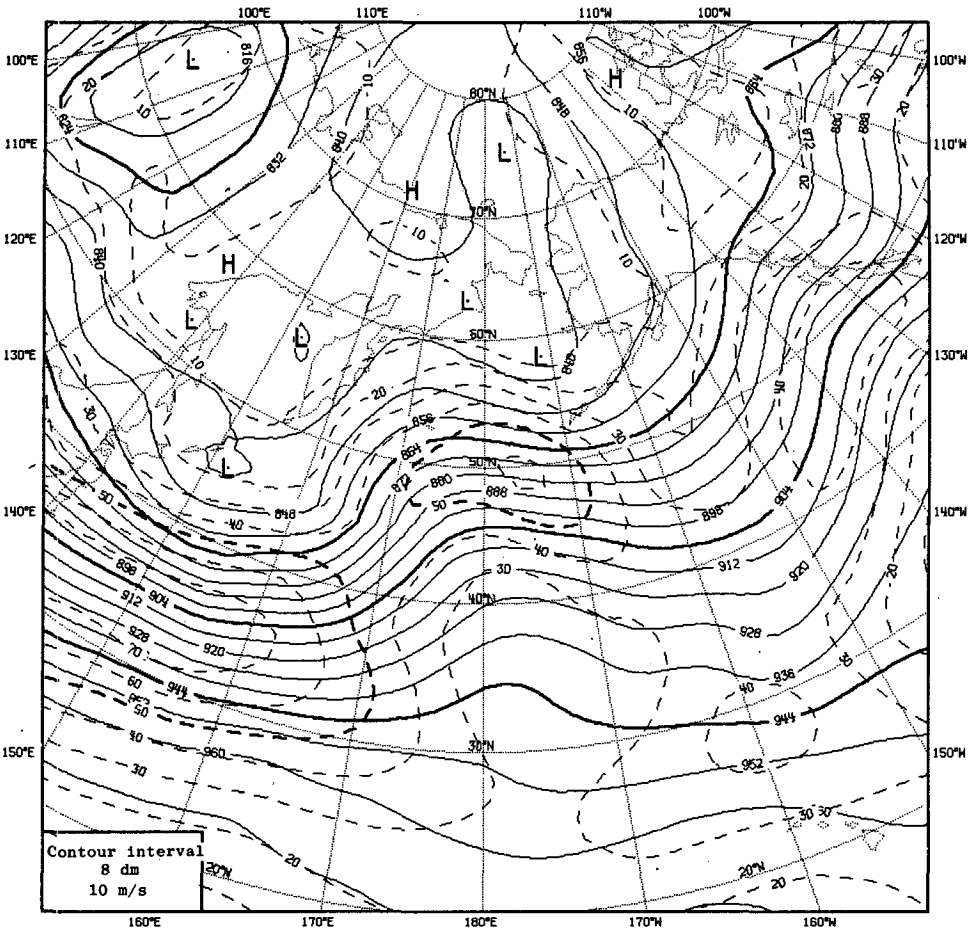

FIG. 13. As in Fig. 11 for an analysis made using a climatological first guess.

winds were fitted more closely. The main low was 1 mb less deep and a little farther west and the wind field had a distinct closed circulation at 47°N, 162°W.

Finally, in Figs. 17–19 I present the 700 mb vertical velocity fields corresponding to Figs. 14–16 (with the sea level pressure shown dashed to ease comparison), calculated directly from the winds using the continuity equation. It has already been noted that although the streamfunction based wind prediction error covariance model implies nondivergent wind increments within one analysis volume, the independent analyses of different volumes are not so constrained, and divergences on a scale which can be resolved by these can be analyzed. Thus the detailed vertical motion patterns from the forecast (Fig. 17) are mostly preserved in the analysis (Fig. 18). The changes made during the analysis, however, do seem meteorologically feasible: increased upward motion near the center of the low which was underdeveloped in the forecast, intensification of the cold front, and downward motion in the ridge near 180°E. The analysis from climatology (Fig. 19) shows clearly the low effective resolution. Yet, despite the zero first guess and the lack of any diagnostic relationship with the analyzed mass field the gross features associated with the main low are distinguishable.

7. Computational aspects

Care must be taken that the large matrices **M** used in the analysis are not ill-conditioned, otherwise ridiculous results can occur. It has been found in practice that the strictly positive definite covariance model, with parameters s and μ locally constant within each system and **V** and F positive definite, together with realistic observation errors, always gives acceptable analyses. This result, which holds even in the presence of close or collocated observations, is contrary to experience with some other optimum interpolation schemes (e.g., Bergman, 1979; Jones, 1976). It is perhaps due to the accurately positive definite correlation functions used, rather than approximate tabulated forms, and the

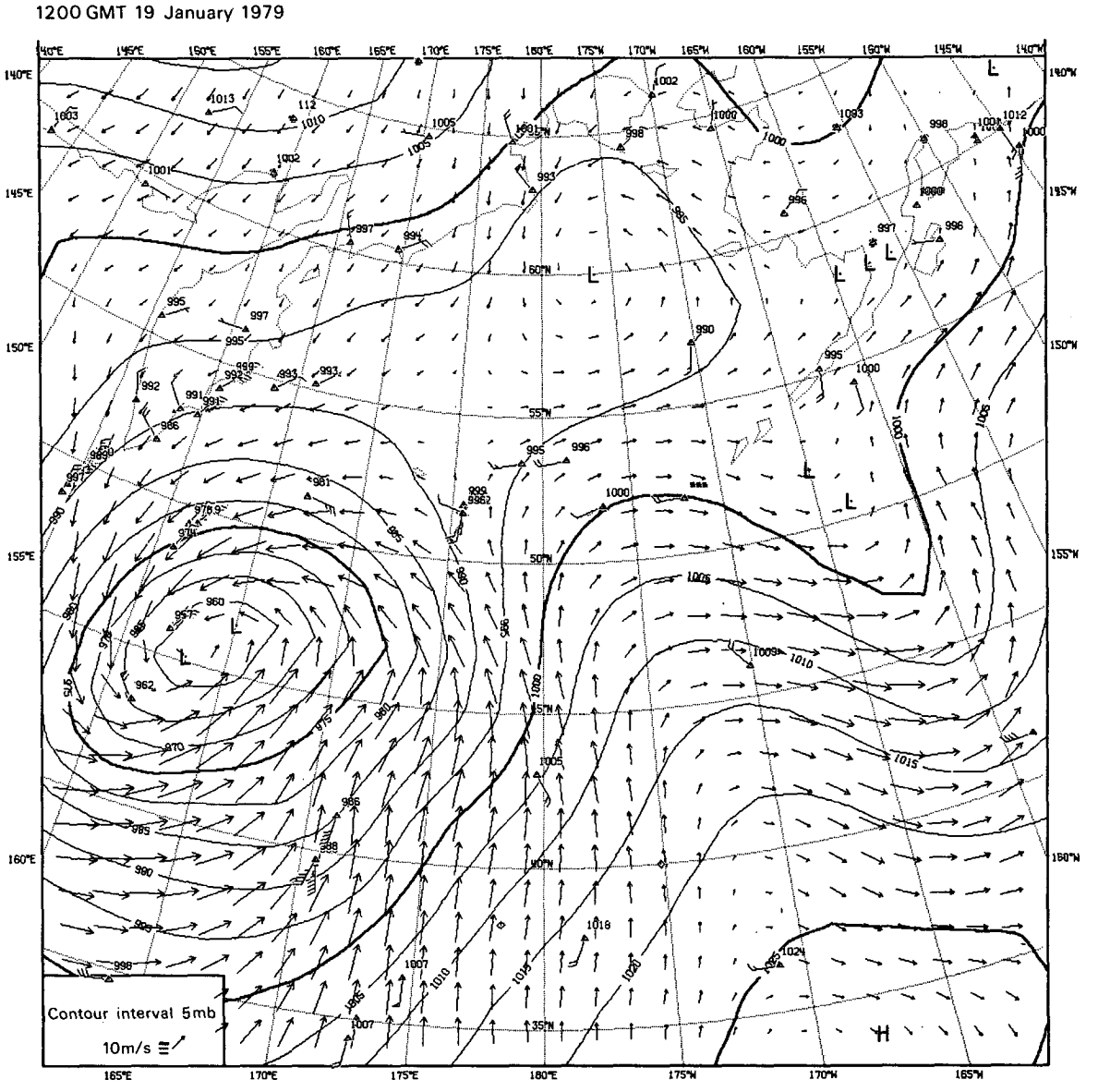

FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s^{-1}).

48-bit precision used for all computations. If the mathematical constraints on the covariance model are relaxed by allowing s to vary in the vertical, then ill-conditioning or even negative eigenvalues sometimes occur. When this is detected the simple expedient of increasing slightly the observation errors overcomes any problems.

If the basic analysis method of Section 3a were used to calculate grid-point values, then were would be no need to calculate an explicit matrix inverse $\mathbf{M}^{-1}$. Instead the analysis coefficients $\mathbf{c}$ could be found by solving (11) and each grid-point value evaluated by a scalar product (12), thus saving considerable computation. However, calculations of estimated analysis errors using (8) and (9) require the matrix inverse or a different solution for each error, as does the data checking of Section 3c and the exclusion of rejected data at Section 3d. So an explicit inverse is calculated using a root-free Cholesky decomposition (Wilkinson and Reinsch, 1971). Because this algorithm can be organized to take full advantage of the Cray 1 computer's vector processing hardware the actual expense of calculating the inverse is not excessive compared to that of other parts of the calculation. Approximate timings on the Cray 1 computer for a global analysis

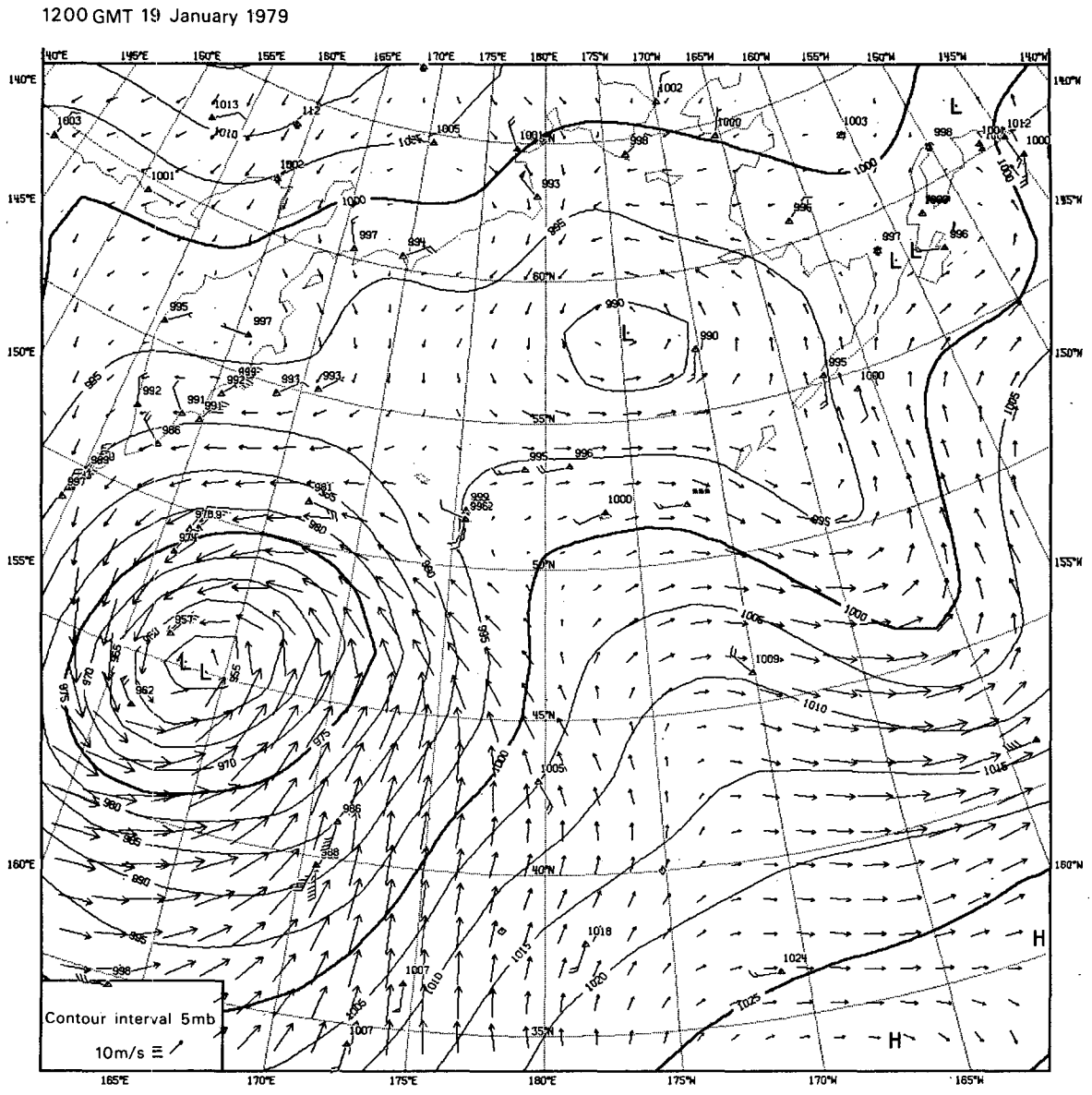

FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

from a FGGE observation set containing 8386 reports were as follows:

1) Preprocessing of observations, interpolation and subtraction of first-guess, "super observation" formation—16 s. (After this processing there remained 5772 observations containing 38747 data.)

2) Selection of data—40 s. (There were 2752 analysis volumes with a mean matrix order $\bar{N}$ of 105.)

3) Calculation of $\mathbf{M}$—51 s.

4) Calculation of $\mathbf{M}^{-1}$—55 s.

5) Data checking—20 s.

6) Calculation of analysis errors using (27) (9) and analysis coefficients using (28)—35 s.

7) Evaluation of grid point values using (12)—71 s.

8) Input, output, etc.—17 s.

The mean matrix order $\bar{N}$ is sensitive to parameters used in the super-observation formation and data selection. The time taken in 4) is approximately proportional to $\overline{N^3}$, in 3) and 5) to $\overline{N^2}$ and in 6) and 7) to $\bar{N}$.

8. Summary

A three-dimensional statistical interpolation scheme, multivariate in geopotential height, thickness and wind, has been described. The scheme is

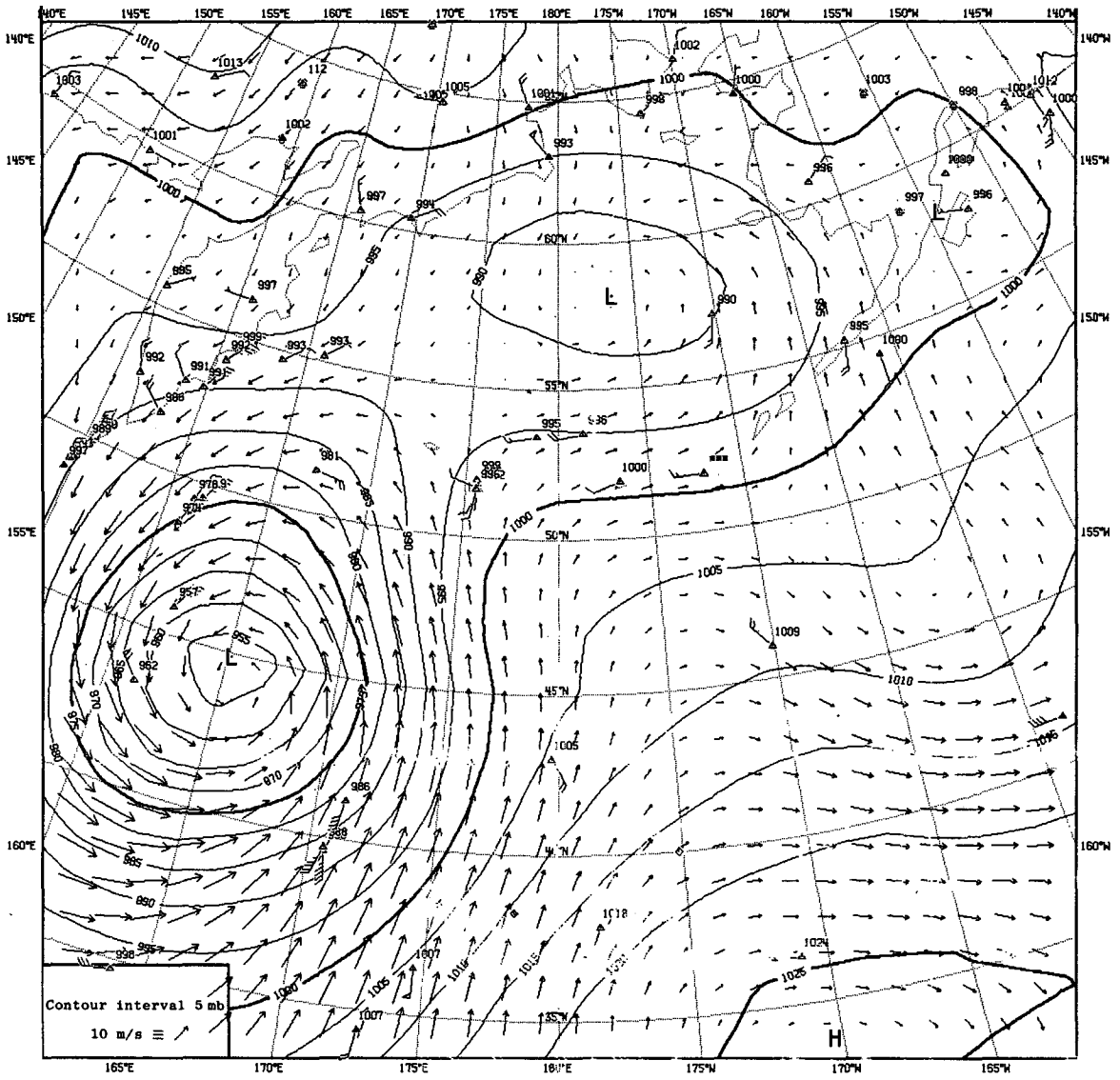

FIG. 16. As in Fig. 14 for an analysis made using a climatological first-guess.

novel in that the analysis equations are set up and solved for a volume of space rather than the individual grid points and variables, making feasible the use of a large number of data.

The multivariate relationships used relate deviations from the first guess of wind components via the streamfunction, vertical gradients of geopotential height and thickness, and (outside of the tropics) horizontal gradients of geopotential height and wind components. It is shown in examples that the method is capable of using such multivariate data consistently, no matter how they are distributed. However, the usefulness of gradient data (wind, thickness) when analyzing geopotential height is greatly enhanced by the simultaneous use of suitably located reference level data, and conversely the usefulness of geopotential height or thickness data when analyzed wind is enhanced if they are distributed suitably for calculating gradients. The large number of data used for each analysis reduces the data selection problem this entails. The use of these relationships does not necessarily imply that the analysis is geostrophic or nondivergent.

The scheme includes an automatic check of each datum against an analysis made not using it, and data which are unlikely to be correct are excluded from the final analysis. In principle, the analysis method is sensitive to the estimated error covariances used; more so because many data rather than just the nearest are used for each grid point. In most

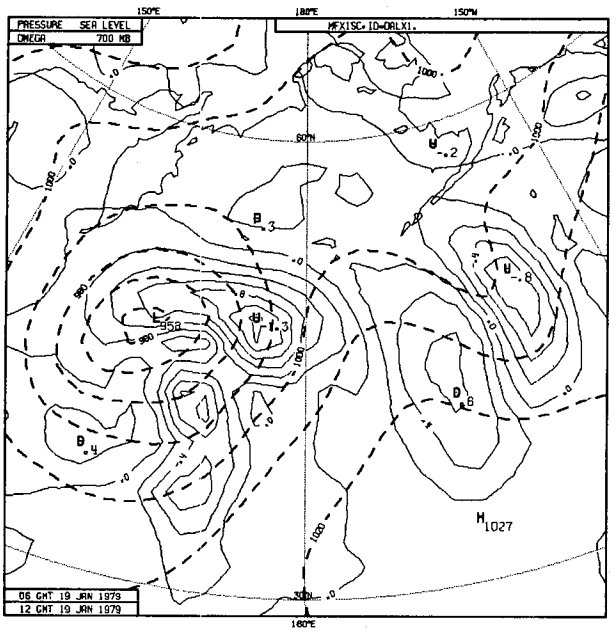

FIG. 17. 700 mb vertical velocity forecast (Pa s^{-1}) corresponding to Fig. 14, with sea level pressure shown dashed.

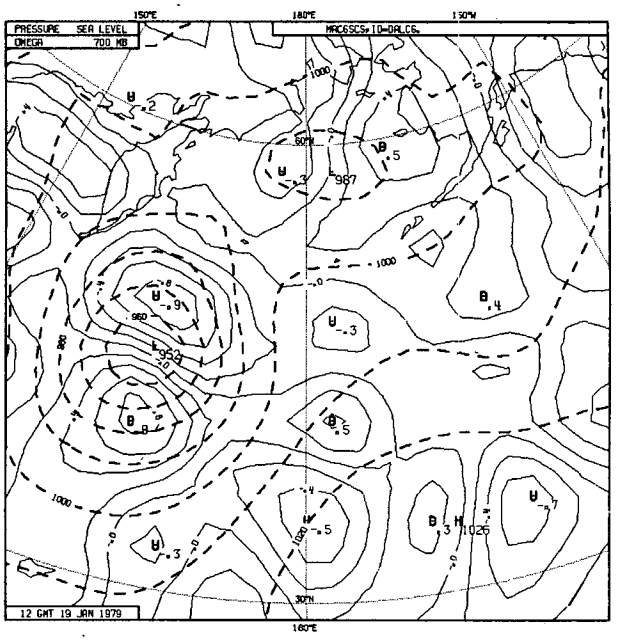

FIG. 19. As Fig. 17 for an analysis made using a climatological first guess.

practical cases, however, the direct effect of changing covariance model parameters is smaller than the indirect effect arising from different data being rejected by the automatic quality control check.

Acknowledgments. Dr. Ian Rutherford and Mr. Gorm Larsen contributed greatly to the initial design of the scheme. Mr. Chris Clarke, Mr. Chris Little and Dr. Stephano Tibaldi assisted with the programming. Discussion with these and other staff members and visiting scientists at ECMWF was invaluable, as was their close study of analyses made using the scheme. Dr. Tony Hollingsworth made several useful suggestions to clarify an earlier version of this paper.

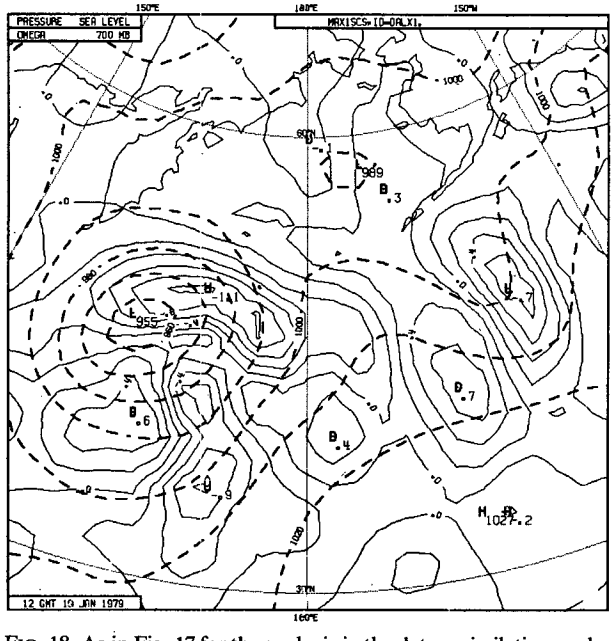

FIG. 18. As in Fig. 17 for the analysis in the data assimilation cycle.

REFERENCES

Bergman, K. H., 1979: Multivariate analysis of temperatures, and winds using optimum interpolation. *Mon. Wea. Rev.*, **107**, 1423–1444.

Buell, C. E., 1972: Correlation functions for wind and geopotential on isobaric surfaces. *J. Appl. Meteor.*, **11**, 51–59.

Gandin, L. S., 1963: *Objective analysis of meteorological fields.* [Translated from Russian by the Israeli Program for Scientific Translations, 1965.], 242 pp.

Hollett, S. R., 1975: Three-dimensional spatial correlations of PE forecast errors. MS thesis. Dept. of Meteorology, McGill University, 73 pp.

Jones, D. E., 1976: The UK Meteorological Office objective analysis scheme for GATE. *Proc. JOC Study Group Conference on Four-dimensional Data Assimilation*, GARP program on numerical experimentation, Rep. No. 11, 122–131.

Julian, P. R., and H. J. Thiebaux, 1975: On some properties of correlation functions used in optimal interpolation schemes. *Mon. Wea. Rev.*, **103**, 105–616.

Larsen, G., C. Little, A. Lorenc and I. Rutherford, 1977: Analysis error calculations for FGGE. ECMWF Research Dept., Int. Rep. No. 11, 63 pp.

Lorenc, A., I. Rutherford and G. Larsen, 1977: The ECMWF

analysis and data assimilation scheme—Analysis of mass and wind fields. ECMWF Tech Rep. No. 6, 47 pp.

Rutherford, I., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.

——, 1973: Experiments on the updating of PE forecasts with real wind and geopotential data. *Preprints Third Symposium on Probability and Statistics*, Boulder, Amer. Meteor. Soc., 198–201.

——, 1976: An operational 3-dimensional multivariate statistical objective analysis scheme. *Proc. JOC Study Group Conference on Four-Dimensional Data Assimilation*, Paris, The GARP programme on numerical experimentation, Rep. No. 11, WMO/JCSU, 98–111.

Seaman, R. S., 1977: Absolute and differential accuracy of analysis achievable with specified observational network characteristics. *Mon. Wea. Rev.*, **105**, 1211–1222.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103**, 246–257.

——, and G. W. Branstator, 1979: Estimation of errors in Nimbus 6 temperature profiles and their spatial correlation. *Mon. Wea. Rev.*, **107**, 1402–1413.

——, —— and L. G. Thiel, 1976: Testing a global multivariate statistical objective analysis scheme with observed data. *Mon. Wea. Rev.*, **104**, 705–783.

Wilkinson, J. H., and Reinsch, C., 1971: *Handbook for Automatic Computation*, Vol. II, *Linear Algebra*. Springer-Verlag, 441 pp.