Time series modeling of significant wave height in multiple scales, combining various sources of data

Ch. N. Stefanakos,[1,2] G. A. Athanassoulis,[1] and S. F. Barstow[3]

Received 22 April 2005; revised 12 June 2006; accepted 23 June 2006; published 3 October 2006.

[1] In the present paper, a composite stochastic model is formulated and validated, resolving the state-by-state, seasonal and interannual variabilities of H_S. The model is a combination of two cyclostationary random processes modeling the variability of mean monthly values and mean monthly standard deviations, respectively, and of a stationary random process modeling the residual, state-by-state, variability. In this way, the time series of significant wave height is given the structure of a multiple-scale composite stochastic process. The present model is a generalization of the nonstationary stochastic modeling introduced by the authors in previous works.

Citation: Stefanakos, Ch. N., G. A. Athanassoulis, and S. F. Barstow (2006), Time series modeling of significant wave height in multiple scales, combining various sources of data, *J. Geophys. Res.*, *111*, C10001, doi:10.1029/2005JC003020.

1. Introduction

[2] It is well known that long-term time series of significant wave height exhibit a number of features, namely random variability, serial correlation, seasonal periodicity and, possibly, a long-term climatic trend, evolving in different time scales. The first two authors have established a nonstationary modeling in a series of works [*Athanassoulis and Stefanakos*, 1995, 1998; *Stefanakos*, 1999; *Stefanakos and Athanassoulis*, 2001, 2003; *Stefanakos and Belibassakis*, 2005], according to which a many-year long time series of significant wave height is modeled as a cyclostationary stochastic process with yearly periodically varying mean value and standard deviation. A multiyear long-term trend can also be included in the model, if the data show that such a trend is present [see, e.g., *Athanassoulis and Stefanakos*, 1995; *WASA Group*, 1998; *Carter*, 1999; *Marshall et al.*, 2001] (and references cited therein). In the present work, we shall disregard this question, since the data we have at our disposal are not long enough to resolve this feature.

[3] Last decades, measurements from satellite altimeters have made available a large amount of wave data with a worldwide coverage. These data sets have been used for various sea wave applications, such as extreme value calculations [*Charriez et al.*, 1992; *Barstow and Krogstad*, 1993; *Cooper and Forristall*, 1997; *Panchang et al.*, 1999], wave climate studies [*Tournadre and Ezraty*, 1990; *Tournadre*, 1993; *Carter et al.*, 1995], etc.

[4] Further exploitation of satellite data can be accommodated by combining them with other sources of data (e.g., buoy measurements) to provide wave information for climatological and/or operational purposes. The present authors have developed a methodology for integrating wave data from different sources, combining especially long-term satellite altimeter data with a restricted amount of buoy measurements [*Athanassoulis et al.*, 2003]. This methodology exploits the nonstationary modeling of the H_S time series [*Athanassoulis and Stefanakos*, 1995], permitting to distinguish between the state-by-state (hourly) scale and the seasonal scale and, thus, to associate the hourly scale with the buoy measurements, and the seasonal scale with the satellite measurements. In this way, different features can be estimated by means of different type of data.

[5] The purpose of the present work is twofold. First, to further assess (using a number of long-term data sets from the Atlantic and the Pacific Ocean) the existing evidence that seasonal patterns of wave climate can be estimated equally well by means of either buoy or satellite data. Secondly, to stochastically model these seasonal patterns, i.e., the time series of monthly mean values and monthly mean standard deviations. These time series are given the structure of cyclostationary processes, independent from the stationary one used to model the hourly variability. In this way, a multiple-scale composite stochastic process is obtained providing an enhanced modeling of long-term time series of significant wave height. This model can be exploited for blending (integrating) available data from different sources, estimating different-scale features by using data from different sources. For example, already available satellite data can be integrated with a short (thus affordable and feasible) period of in situ buoy measurements, to obtain an artifact of a long-term measured time series. This kind of data is highly desirable for a number of important applications such as, e.g., coastal morphodynamics (sediment transport, beach erosion), coastal work planning, direct numerical simulation of nonlinear long-term responses of offshore structures, prediction of long-term extremes, down-time analysis etc.

[1]School of Naval Architecture and Marine Engineering, National Technical University of Athens, Zografos, Athens, Greece.

[2]Also at Department of Shipbuilding, Technological Educational Institute of Athens, Aegaleo, Athens, Greece.

[3]Fugro OCEANOR, Trondheim, Norway.

0148-0227/06/2005JC003020

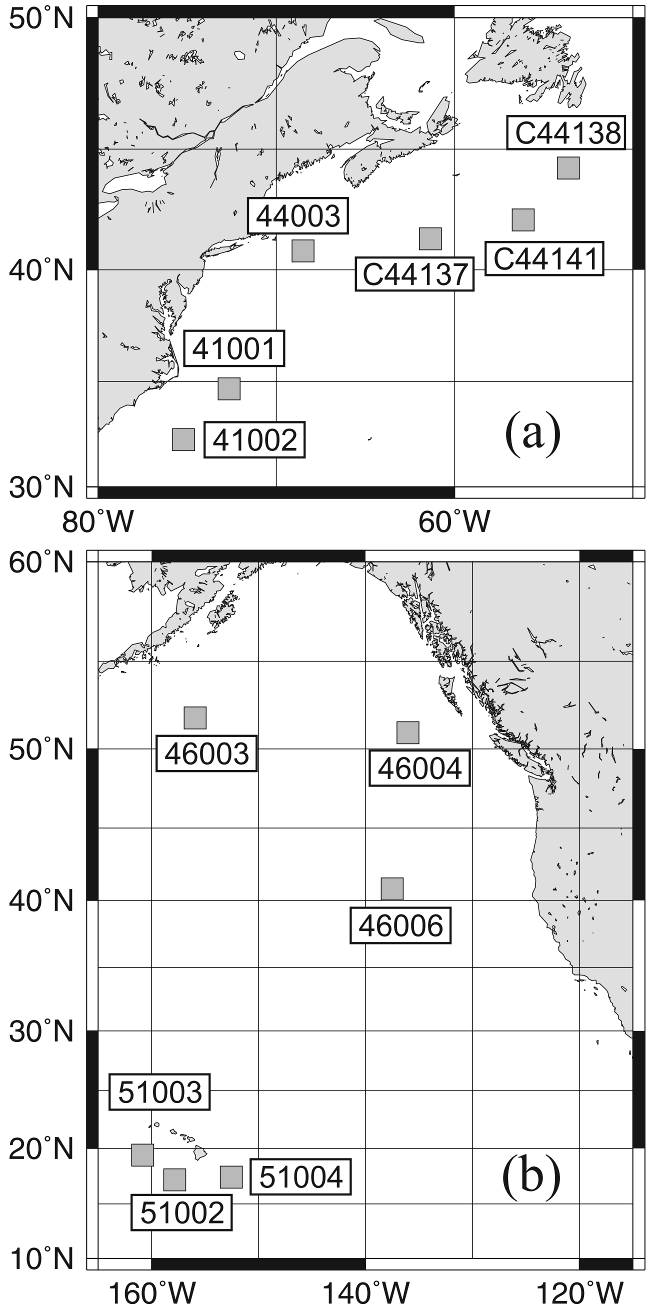

Figure 1. Examined sites in the (a) Atlantic Ocean and (b) Pacific Ocean.

[6] The structure of this paper is as follows. First, the stochastic modeling of *Athanassoulis and Stefanakos* [1995] is briefly presented and reformulated in accordance with the needs of the present work. Then, various statistics are introduced, defined by means of either time series of buoy measurements or space-time series of satellite data, associated with a given site. Systematic comparisons of the various statistics based on the two data sources are then presented, revealing which ones are interchangeable in the description of the seasonal patterns. After this assessment, the two time series of mean monthly values and mean monthly standard deviations are modeled as cyclostationary processes. Because of the established interchangeability of buoy and satellite statistics in the seasonal scale, the parameters of the above models can be estimated either from buoy measurements or from satellite data. A composite stochastic model is then obtained, integrating the three separate models, and giving a stochastic description of both hourly variability and second periodicity. Finally, a general discussion and some conclusions concerning the extent of applicability and the necessary precautions in using the present approach are presented.

[7] The data used in the present work come from two sources. First, long-term time series of significant wave height from buoy measurements; see Figure 1 and Table 1. Buoy data are either freely available in the Internet through the site of the National Data Buoy Center, National Oceanic and Atmospheric Administration (NDBC/NOAA) or have kindly been provided by J. Gagnon, Marine Environmental Data Service (MEDS). Second, long-term space-time series of significant wave height from various altimeters around the buoy locations. Satellite data come from the archive of Fugro OCEANOR and has been extracted using its advanced software tool WWA. Note that the NDBC/NOAA buoys are some of the several buoys used in deriving the calibration procedure applied to satellite altimeter data sets [see, e.g., *Krogstad and Barstow*, 1999]. In the first phase, altimeter data used have (at least partial) overlapping in time with buoy measurements; namely Geosat and Topex altimeters (Geosat: 11.1986–12.1989 and Topex: 9.1992–12.1998). In a second phase, data from all available altimeter have been processed and compared with buoy data. In Table 2, a complete list of satellite data is given.

2. Modeling and Analysis of α-Hourly Long-Term Time Series

[8] The terminology "α-hourly" time series is used in order to denote any time series of measurements with time step $\Delta\tau = \alpha$ hours. Usually, spectra or spectral parameters are recorded (or calculated) every 1, 3, 6 or 12 hours, thus $\alpha = 1, 3, 6, 12$. However, any value $1 \leq \alpha \leq 12$ is possible.

[9] Let us denote by $X(\tau_i)$, $i = 1, 2, \ldots, I$, the α-hourly many-year long time series of significant wave height $H_S(\tau)$ or an appropriate transform of it. Usually, the shifted logarithms of $H_S(\tau)$ are considered, i.e., $X(\tau) = [\log H_S(\tau) + c]$, where c is a small positive constant between 0.2 m and 1 m. The constant c is introduced in order to avoid zeros and minimize the skewness of the probability distribution of $X(\tau)$. The log-transformed data are often approximately Gaussian, which greatly facilitates the analysis and the simulation procedure. According to the modeling introduced by *Athanassoulis and Stefanakos* [1995], such a time series $X(\tau)$ admits to the following decomposition:

$$X(\tau) = \overline{X}_{tr}(\tau) + \mu(\tau) + \sigma(\tau)W(\tau), \tag{1}$$

where $\overline{X}_{tr}(\tau)$ is any possible long-term (climatic) trend, $\mu(\tau)$ and $\sigma(\tau)$ are deterministic periodic functions with period of one year, and $W(\tau)$ is a zero-mean, stationary, stochastic process. The functions $\mu(\tau)$ and $\sigma(\tau)$ are called seasonal mean value and seasonal standard deviation, respectively, and are used to describe the exhibited seasonal patterns. In the sequel, we shall consider that $\overline{X}_{tr}(\tau) = \overline{X} = const$ and this

Table 1. Metadata Information (Identification Code, Location, Depth and Period) of Buoy Measurements Used

Identification Code	Source	Lat, °N	Long, °W	Depth, m	Period
		Atlantic Ocean			
41001	NDBC/NOAA	34.68	72.64	4389	1978–1998
41002	NDBC/NOAA	32.28	75.20	3786	1978–1998
44003	NDBC/NOAA	40.80	68.50	N/A	1979–1984
C44137	MEDS	41.32	61.35	4500	1988–1997
C44138	MEDS	44.25	53.62	1500	1988–1997
C44141	MEDS	42.12	56.13	4500	1990–1997
		Pacific Ocean			
46003	NDBC/NOAA	51.85	155.92	4709	1978–1998
46004	NDBC/NOAA	51.00	136.00	N/A	1978–1988
46006	NDBC/NOAA	40.84	137.49	4023	1977–1998
51002	NDBC/NOAA	17.19	157.83	5002	1992–1998
51003	NDBC/NOAA	19.40	160.81	4883	1992–1998
51004	NDBC/NOAA	17.44	152.52	5304	1992–1998

constant will be incorporated into $\mu(\tau)$. Let it be noted that, the whole methodology presented herein can be equally well applied to the case where a climatic trend $\overline{X}_{tr}(\tau)$ is present, if the data (from the same or other sources) permits us to identify such a trend.

[10] Thus, in the present work, decomposition (1) will be rewritten as

$$X(\tau) = \mu(\tau) + \sigma(\tau)W(\tau). \quad (2)$$

[11] The principal aim of the present work is to examine if and how it is possible: (1) to assess the methodology of obtaining reasonable estimates of $\mu(\tau)$ and $\sigma(\tau)$ by means of satellite data, and (2) to model $\mu(\tau)$ and $\sigma(\tau)$ as two cyclostationary processes, with similar structure as process modeled by equation (2), and then embed them in model (2).

[12] The time series $X(\tau)$ is usually reindexed, in order to properly treat variability at different time scales, by using the double Buys-Ballot index (j, τ_k), where j is the year index and τ_k ranges within the annual time [*Athanassoulis and Stefanakos*, 1995]. In the present work, a triple index of similar philosophy, introduced by *Athanassoulis et al.* [2003] and denoted by (j, m, τ_k), will be used. The first component j is again the year index. The second component m is a month index, ranging through the set of integers {1, 2,..., M = 12}. The third component τ_k represents the monthly time, the index k ranging through the set of integers {1, 2,..., K_m}, where K_m is the number of α-hourly observations within the m-th month. Clearly, the meaning of the symbol τ_k in the triple index (j, m, τ_k) used herewith, is different from the meaning of the same symbol in the double index (j, τ_k), used in previous studies [*Athanassoulis and Stefanakos*, 1995; *Stefanakos and Athanassoulis*, 2001].

[13] According to the new, three-index notation, the time series $X(\tau_i)$ is reindexed as follows:

$$\{X(j,m,\tau_k), \quad j = 1,2,\ldots,J, \quad m = 1,2,\ldots,M, \quad k = 1,2,\ldots,K_m\}. \quad (3)$$

[14] The three indices (j, m, τ_k) represent three different time scales, making it possible to explicitly define statistics with respect to each one of them, separately. In the following sections, use will be made of the subscripts 1, 2, 3 to denote various statistics (mean value and standard deviation) with respect to the corresponding (first, second, third) index. In order to clarify the structure of this notation, we present a number of examples, some of which will also be used in the sequel:

$$M_1(m,\tau_k) = \frac{1}{J}\sum_{j=1}^{J} X(j,m,\tau_k), \quad (4a)$$

$$S_1(m,\tau_k) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}[X(j,m,\tau_k) - M_1(m,\tau_k)]^2}, \quad (4b)$$

$$M_3(j,m) = \frac{1}{K_m}\sum_{k=1}^{K_m} X(j,m,\tau_k), \quad (4c)$$

$$S_3(j,m) = \sqrt{\frac{1}{K_m}\sum_{k=1}^{K_m}[X(j,m,\tau_k) - M_3(j,m)]^2}. \quad (4d)$$

Table 2. Metadata Information (Altimeter Name, Duration and Period) of Satellite Measurements Used

Altimeter	Duration, years	Period
	First Phase	
Geosat	3	1986.11.09–1989.09.26
Topex	6.5	1992.09.26–1998.12.29
	Second Phase	
Topex-all	10	1992.09.26–2002.08.09
Geosat-Follow On	6	2000.01.08–2005.12.07
Jason	3.5	2002.01.15–2005.09.12
Topex2	3	2002.09.26–2005.09.10
Envisat	3	2002.10.04–2005.09.15

[15] Similarly, by successively taking mean values with respect to two indices, we can also define two-index statistics. For example:

$$M_{13}(m) = \frac{1}{K_m} \sum_{k=1}^{K_m} \frac{1}{J} \sum_{j=1}^{J} X(j, m, \tau_k) = M_{31}(m). \qquad (4e)$$

[16] In the next sections, we formulate appropriate estimates of seasonal patterns $\mu(\tau)$ and $\sigma(\tau)$, based on the above presented statistics.

3. Modeling and Analysis of Monthly Mean Values

3.1. Buoy Measurements

[17] It is a straightforward matter to define the time series of monthly mean values (MMV) of $X(\tau_i)$. In fact, equation (4c) defines this time series by averaging α-hourly observations over each month. In Figure 2, the MMV time series, obtained from the α-hourly time series mentioned in Table 1, is presented as solid circles. Averaging $M_3(j, m)$ over all the examined years, we obtain the overall MMV (per month):

$$\tilde{M}_3(m) = \frac{1}{J} \sum_{j=1}^{J} M_3(j, m) = M_{31}(m) = M_{13}(m). \qquad (5a)$$

[18] The time series of monthly standard deviations (MSD) of $X(\tau_i)$ is defined by means of the equation (4d). See also Figure 2, where it is shown as a sequence of open circles. Averaging $S_3(j, m)$ over all the examined years, we obtain the overall MSD (per month):

$$\tilde{S}_3(m) = \frac{1}{J} \sum_{j=1}^{J} S_3(j, m). \qquad (5b)$$

[19] It should be noted that $\tilde{S}_3(m)$ is not the standard deviation of the time series $M_3(j, m)$. The selection of $\tilde{S}_3(m)$ as the representative quantity for the variability of MMV $M_3(j, m)$ about the overall MMV $\tilde{M}_3(m)$ has been dictated by the data analysis. Indeed, after extensive numerical experimentation it was found that it is exactly this quantity, i.e. $\tilde{S}_3(m)$, that can be related with (estimated by) an appropriately defined quantity obtained from satellite altimeter measurements; see next section. Further, in a subsequent section, an attempt will be made for the stochastic modeling of monthly series $M_3(j, m)$ and $S_3(j, m)$ of equations (4c)–(4d).

3.2. Satellite Measurements

[20] Let us now turn our attention to satellite altimeter measurements of H_S, obtained along specific (satellite dependent) ground tracks. Clearly, successive satellite observations are not referred to the same point in the sea. Thus, satellite wave data do not have the structure of a time series. If, however, we assume that the wave field is spatially homogeneous for an area S_A, surrounding a specific site of interest A, then we can associate to this site all satellite observations within the area S_A [*Tournadre and Ezraty*, 1990; *Panchang et al.*, 1999]. Of course, the extent and the shape of the area S_A are satellite dependent (sufficient data), and site dependent (local meteorological conditions).

[21] Then, the set of the observations (population) can be given the structure of a three-index data set [*Athanassoulis et al.*, 2003]:

$$\{X^{sat}(j, m, \chi_\ell), \quad j = 1, 2, \ldots, J, \quad m = 1, 2, \ldots, M, \quad \ell = 1, 2, \ldots, L_m\}, \qquad (6)$$

where again j is the year index, m is the month index, and χ_ℓ is just a monthly counter, i.e., an index counting the number of observations within the area S_A, during the month m of the year j. Clearly, for given values of j and m, the individual values $X(j, m, \tau_k)$, $k = 1, 2, \ldots, K_m$, and $X^{sat}(j, m, \chi_\ell)$, $\ell = 1, 2, \ldots, L_m$, are not directly comparable.

[22] Despite the structural differences between the data sets $X(j, m, \tau_k)$ and $X^{sat}(j, m, \chi_\ell)$, it can be expected that appropriate statistics of $X(j, m, \tau_k)$ can be approximated by analogous statistics of $X^{sat}(j, m, \chi_\ell)$, provided that the sea area S_A has been chosen appropriately. This expectation is based on the following assumptions concerning the time-space field of significant wave height $H_S(\tau, \vec{r})$, where τ is time and $\vec{r}$ is the horizontal position of the measurement point: (1) Observations $X(j, m, \tau_k)$ and $X^{sat}(j, m, \chi_\ell)$ are considered as two different samples from the same field $H_S(\tau, \vec{r})$. (2) $H_S(\tau, \vec{r} = const)$ is (approximately) stationary within each month. (Of course, in finer scales, short-duration energetic events (e.g., frontal passages) may occur that do not comply with the stationarity assumption. These events, which should be modeled by using different (finer scale) stochastic processes, will be not considered here.) (3) $H_S(\tau = const, \vec{r})$ is (approximately) homogeneous within the area S_A. (4) A dispersion relation holds for the wave field $H_S(\tau, \vec{r})$ [*Tournadre*, 1993, section 5.3].

[23] Some results concerning the correspondence of temporal and spatial scales of $H_S(\tau, \vec{r})$ have been presented by *Monaldo* [1988, 1990], *Tournadre* [1993], and *Krogstad and Barstow* [1999].

[24] The triple-index notation greatly facilitates the definition of various statistics on $X^{sat}(j, m, \chi_\ell)$, and the comparison with analogous statistics on $X(j, m, \tau_k)$. We present below some definitions of monthly mean values (MMV) and monthly standard deviations (MSD) related with $X^{sat}(j, m, \chi_\ell)$:

$$M_3^{sat}(j, m) = \frac{1}{L_m} \sum_{\ell=1}^{L_m} X(j, m, \chi_\ell), \qquad (7a)$$

$$S_3^{sat}(j, m) = \sqrt{\frac{1}{L_m} \sum_{\ell=1}^{L_m} [X(j, m, \chi_\ell) - M_3(j, m)]^2}, \qquad (7b)$$

$$\tilde{M}_3^{sat}(m) = \frac{1}{J} \sum_{j=1}^{J} M_3^{sat}(j, m), \qquad (8a)$$

$$\tilde{S}_3^{sat}(m) = \frac{1}{J} \sum_{j=1}^{J} S_3^{sat}(j, m). \qquad (8b)$$

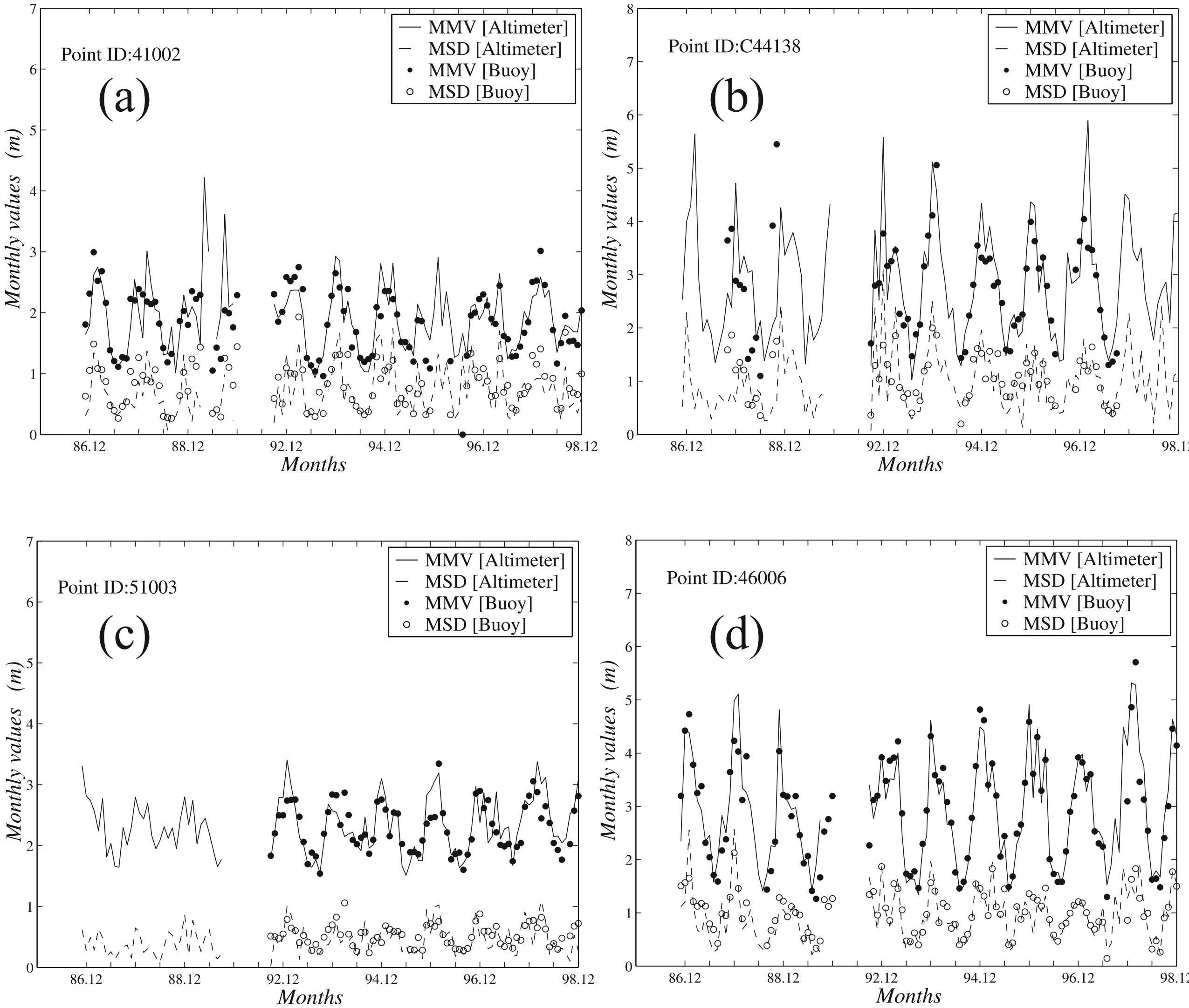

Figure 2. Time series of monthly mean values (satellite: solid line, buoy: solid circles), and monthly standard deviations (satellite: dashed line, buoy: open circles). (a) Point 41002, (b) Point C44138, (c) Point 51003, and (d) Point 46006.

[25] Definitions (7a)–(7b) and (8a)–(8b) correspond to (4c)–(4d) and (5a)–(5b), respectively.

[26] Clearly, $M_3^{sat}(j, m)$ and $S_3^{sat}(j, m)$ are monthly time series generated by spatial/time averaging over the area S_A. In Figure 2, these time series, calculated from Geosat (1986–1989) and Topex (1992–1997) altimeter data are shown with solid line for $M_3^{sat}(j, m)$ and with dashed line for $S_3^{sat}(j, m)$. For comparison purposes, $M_3(j, m)$ and $S_3(j, m)$, based on buoy measurements are plotted as solid and open circles, respectively. As can be seen from Figure 2, the agreement between satellite monthly values and buoy monthly values is very satisfactory.

[27] It should also be stressed that, concerning the overlapping between the period of in situ measurements and the period of the satellite measurements (Geosat: 11.1986–12.1989 and Topex: 9.1992–12.1998), there are four cases: (1) buoy period completely overlaps with satellite (buoys: 41001, 41002, 46003, 46006), (2) buoy period partially overlaps with satellite (buoys: C44137, C44138), (3) buoy period overlaps only with Geosat (buoys: 44003, 46004), and (4) buoy period overlaps only with Topex (buoys: C44141, 51002, 51003, 51004), see also Table 1.

[28] From the analysis above, it seems reasonable to consider $M_3^{sat}(j, m)$ and $S_3^{sat}(j, m)$ as substitutes for $M_3(j, m)$ and $S_3(j, m)$. However for our study the weaker assumption that the monthly time series $M_3^{sat}(j, m)$ and $S_3^{sat}(j, m)$ are statistically equivalent with the monthly time series $M_3(j, m)$ and $S_3(j, m)$, is sufficient.

4. Mean Seasonal Patterns

[29] The mean values (over a number of years) $\tilde{M}_3(m)$ and $\tilde{S}_3(m)$ are defined as the mean seasonal patterns (at a given site in the sea). Based on the assumption of statistical periodicity of the wave climate, we expect that $\tilde{M}_3^{sat}(m) \approx \tilde{M}_3(m)$ and $\tilde{S}_3^{sat}(m) \approx \tilde{S}_3(m)$. The couples $(\tilde{M}_3(m), \tilde{M}_3^{sat}(m))$ and $(\tilde{S}_3(m), \tilde{S}_3^{sat}(m))$ are shown in Figure 3 by a solid (buoy) and a dashed (altimeter) line. The agreement between $\tilde{M}_3(m)$ and $\tilde{M}_3^{sat}(m)$ is, in general, very good, except for some (localized in time) discrepancies that might be (partly)

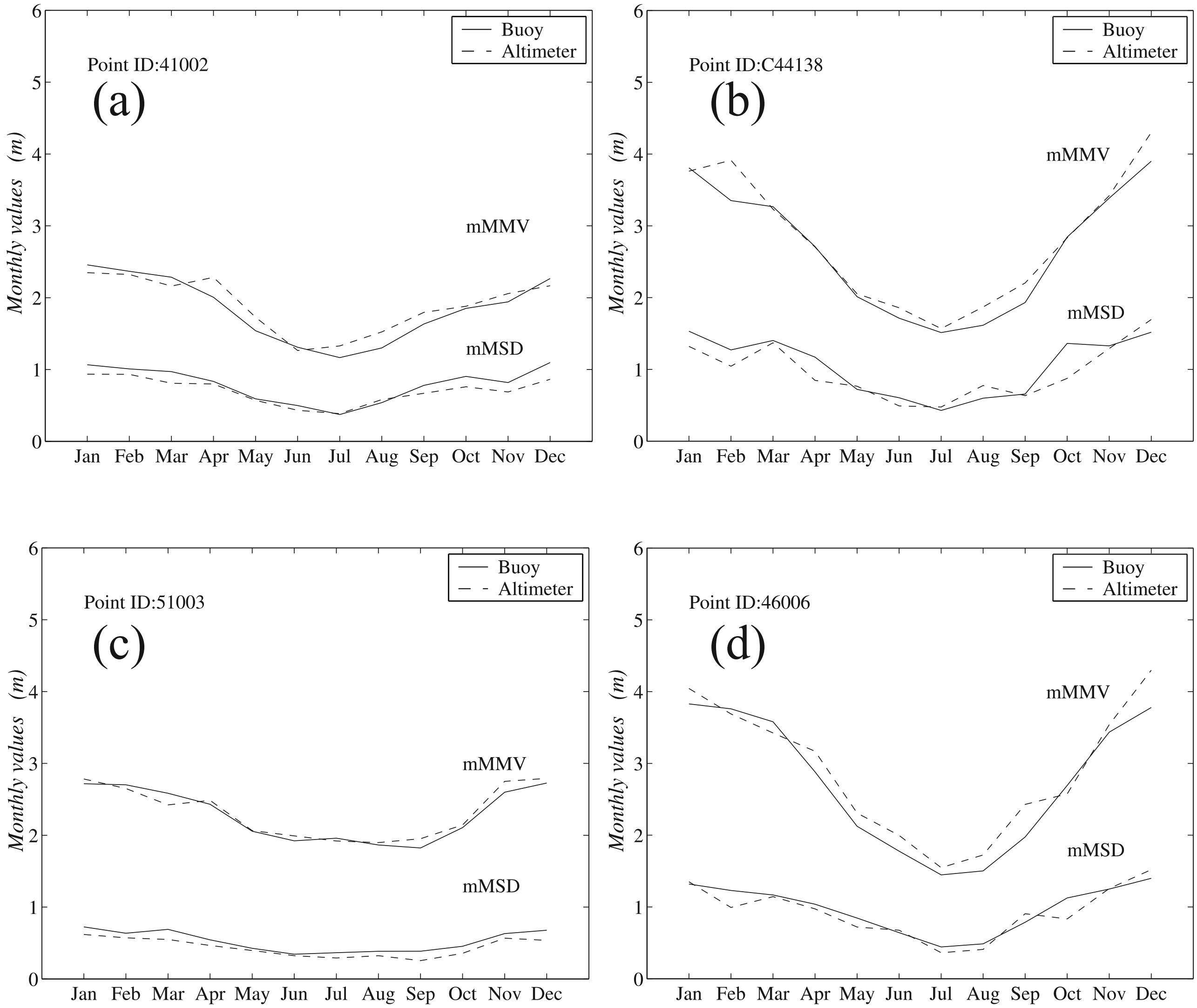

Figure 3. Seasonal patterns of significant wave height. Comparison of overall monthly mean values $\tilde{M}_3(m)$ ($\tilde{M}_3^{sat}(m)$) and overall monthly standard deviations $\tilde{S}_3(m)$ ($\tilde{S}_3^{sat}(m)$) from buoy (solid line) and satellite (dashed line) measurements. (a) Point 41002, (b) Point C44138, (c) Point 51003, and (d) Point 46006. (Coincident data.)

explained by studying finer-scale meteorological phenomena. The agreement between $\tilde{S}_3(m)$ and $\tilde{S}_3^{sat}(m)$ is also very good.

[30] The method of obtaining quantities $\tilde{M}_3^{sat}(m)$ and $\tilde{S}_3^{sat}(m)$ is further validated by calculating these quantities based on satellite data not coinciding in time with buoy measurements; see Table 2. The results of this calculation are shown in Figure 4 along with the results presented in Figure 3 for comparison purposes.

[31] In general, the pattern exhibited is impressively good taking into account the fact that the examined data do not coincide in time with buoy measurements. Some discrepancies that are present could be partly attributed to local (in space and time) meteorological phenomena. For example, in point 41002 the major discrepancies are exhibited in the month of September, which is the peak period for hurricane activity in the area; see Table 3. This fact in combination with the small number of satellite measurements in some data sets (Jason: 40, Envisat: 17, Topex2: 14) magnifies the influence of hurricane measurements in the calculation of mean monthly value of September, resulting in the present discrepancies.

5. Stochastic Modeling of the Seasonal Patterns

[32] In the present section, the monthly time series $M_3(j, m)$ and $M_3^{sat}(j, m)$, and $S_3(j, m)$ and $S_3^{sat}(j, m)$ are given the structure of cyclostationary processes with yearly periodic mean value and standard deviation. That is, we assume that

$$\mu(\tau) = \mu_1(\tau) + \sigma_1(\tau)W_1(\tau), \text{ and} \tag{9a}$$

$$\sigma(\tau) = \mu_2(\tau) + \sigma_2(\tau)W_2(\tau), \tag{9b}$$

where $\mu_i(\tau)$ and $\sigma_i(\tau)$ are deterministic periodic functions, and $W_1(\tau)$ and $W_2(\tau)$ model the residual series. Note that

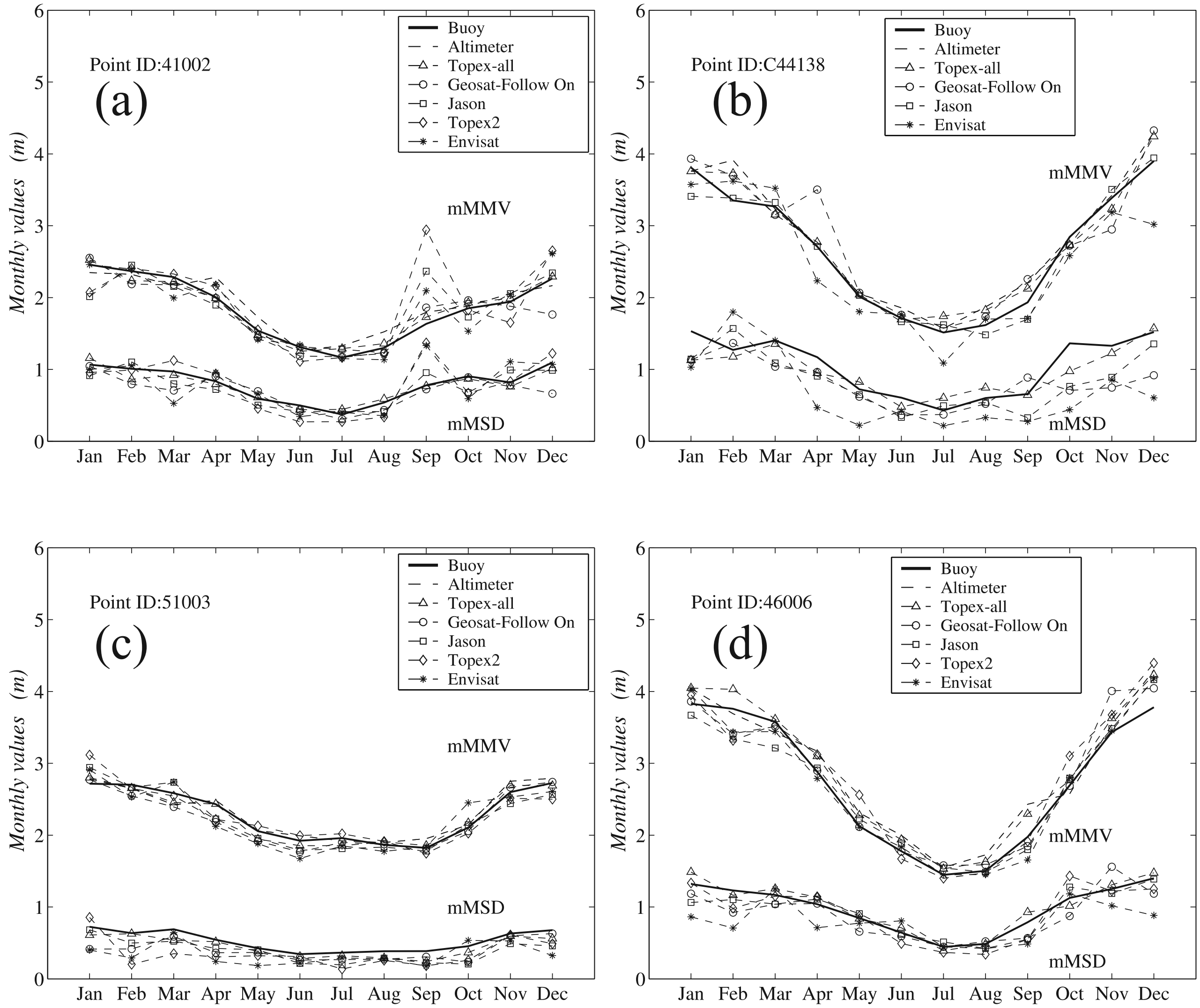

Figure 4. Seasonal patterns of significant wave height. Comparison of overall monthly mean values $\tilde{M}_3(m)$ ($\tilde{M}_3^{sat}(m)$) and overall monthly standard deviations $\tilde{S}_3(m)$ ($\tilde{S}_3^{sat}(m)$) from buoy (solid line) and satellite (dashed line) measurements. (a) Point 41002, (b) Point C44138, (c) Point 51003, and (d) Point 46006. (Noncoincident data.)

mean values $\mu_1(\tau)$ and $\mu_2(\tau)$ are estimated by means of $\tilde{M}_3(m)$ (or $\tilde{M}_3^{sat}(m)$), and $\tilde{S}_3(m)$ (or $\tilde{S}_3^{sat}(m)$), respectively. Working similarly, $\sigma_1(\tau)$ and $\sigma_2(\tau)$ are estimated as follows:

$$sdM_3(m) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left[M_3(j,m) - \tilde{M}_3(m)\right]^2}, \tag{10a}$$

$$sdS_3(m) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left[S_3(j,m) - \tilde{S}_3(m)\right]^2}, \tag{10b}$$

if buoy data are considered, or

$$sdM_3^{sat}(m) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left[M_3^{sat}(j,m) - \tilde{M}_3^{sat}(m)\right]^2}, \tag{11a}$$

$$sdS_3^{sat}(m) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left[S_3^{sat}(j,m) - \tilde{S}_3^{sat}(m)\right]^2}, \tag{11b}$$

if satellite data are used.

Table 3. Hurricanes in September Near Point 41002

Altimeter	Date, yyyymmdd	H_S, m	Hurricane
Geosat	19890907	5.01	Gabrielle
Topex	19960905	6.08	Fran
Topex-all	19960905	6.30	Fran
	19990915	5.14	Floyd
Geosat-Follow On	20050911	5.91	Ophelia
Jason	20020909	5.03	Gustav
	20050914	5.40	Ophelia
Topex2	20030917	10.02	Isabel
	20050910	5.63	Ophelia
Envisat	20030917	6.25	Isabel
	20040920	5.77	Jeanne

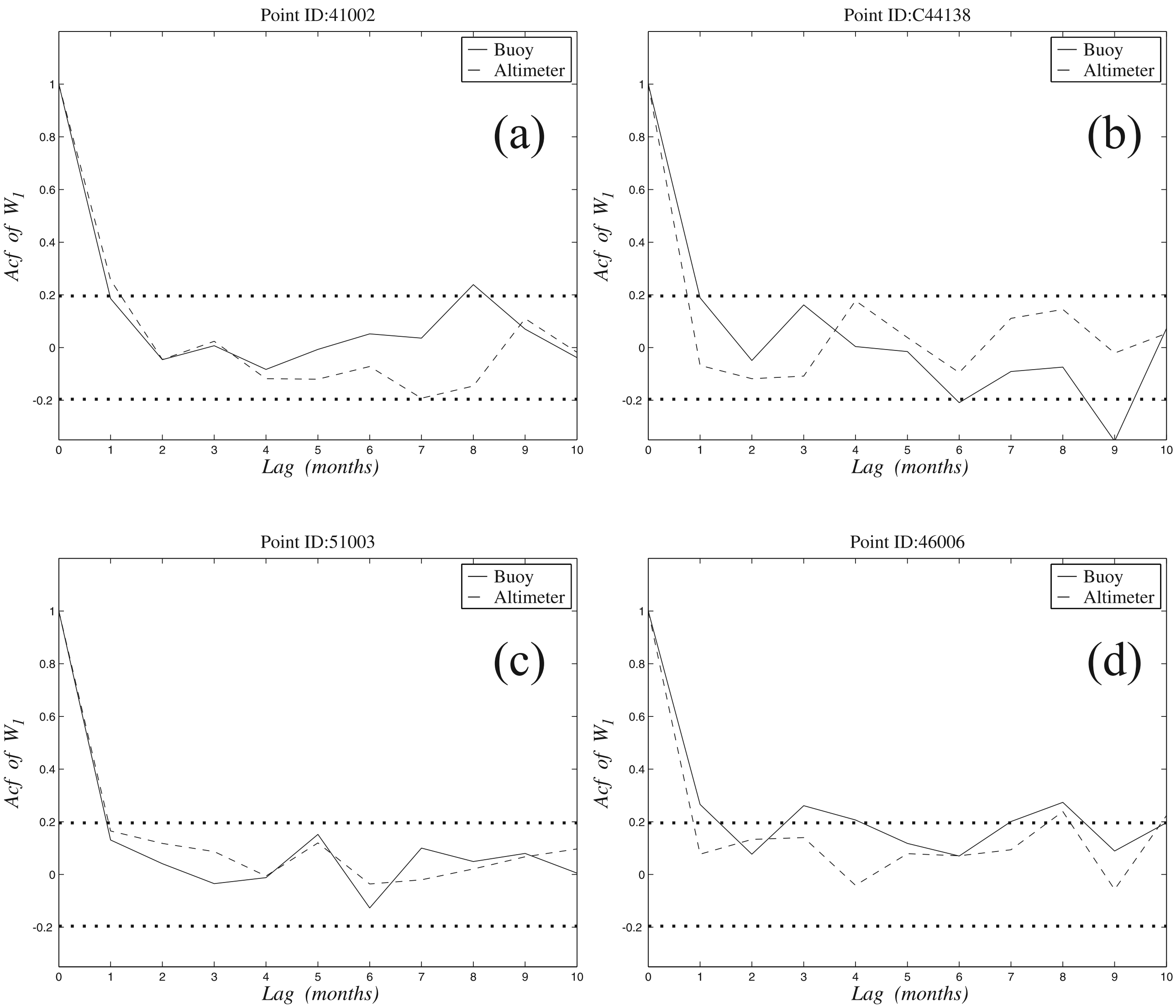

Figure 5. Autocorrelation coefficient function of $W_1(\tau)$ for the points 41002, C44138, 51003, and 46006. Solid line: buoy data, dashed line: altimeter data.

[33] It should be noted here that the processes appearing in equations (9a)–(9b) are being evolved on monthly time scales, while the stationary process $W(\tau)$, appearing in equations (1)–(2), evolves on an α-hourly time scale.

[34] Thus, the processes $\mu(\tau)$, $\sigma(\tau)$, $W_1(\tau)$ and $W_2(\tau)$ can be considered independent from $W(\tau)$.

[35] In order to check the validity of the models (9), the autocorrelation coefficient function (acf) of the residual series $W_1(\tau)$ and $W_2(\tau)$ have been computed and examined. In Figure 5, the acf's of $W_1(\tau)$ for four sites are shown. One can observe that acf's become approximately zero from the first lag. The same holds true for the acf's of $W_2(\tau)$. That is, residual series $W_1(\tau)$ and $W_2(\tau)$ can be considered uncorrelated.

[36] In order to study the nearness of the probability distribution of the residual values coming from the buoy and the satellite measurements, we have plotted the quantiles of the empirical distribution based on the buoy data against the corresponding quantiles based on the satellite data; see Figure 6. This is a version of the well-known QQ plot [*Wilk and Gnanadesikan*, 1968]. The results are very near to the straight line, which means that the data from the two sources can be considered to belong to the same population.

[37] Another application of the QQ plot is shown in Figure 7. In this figure, the quantiles of the empirical distributions coming from buoy and satellite data for the series $W_1(\tau)$ are compared with the quantiles of the standardized normal distribution $N(0,1)$. Again, results are very near to the straight line, i.e. we can say that the data from both sources belong to the normal distribution with zero mean and standard deviation unity. The same result holds true for $W_2(\tau)$. Results for all points mentioned in Table 1 are available from the authors upon request.

6. Multiple-Scale Stochastic Model

[38] By combining equations (2) and (9a)–(9b), we obtain the following composite stochastic process:

$$X\left(t;\beta_\mu,\beta_\sigma,\gamma\right) = X_\mu\left(t;\beta_\mu\right) + X_\sigma\left(t;\beta_\sigma\right)\, W(t;\gamma), \qquad (12)$$

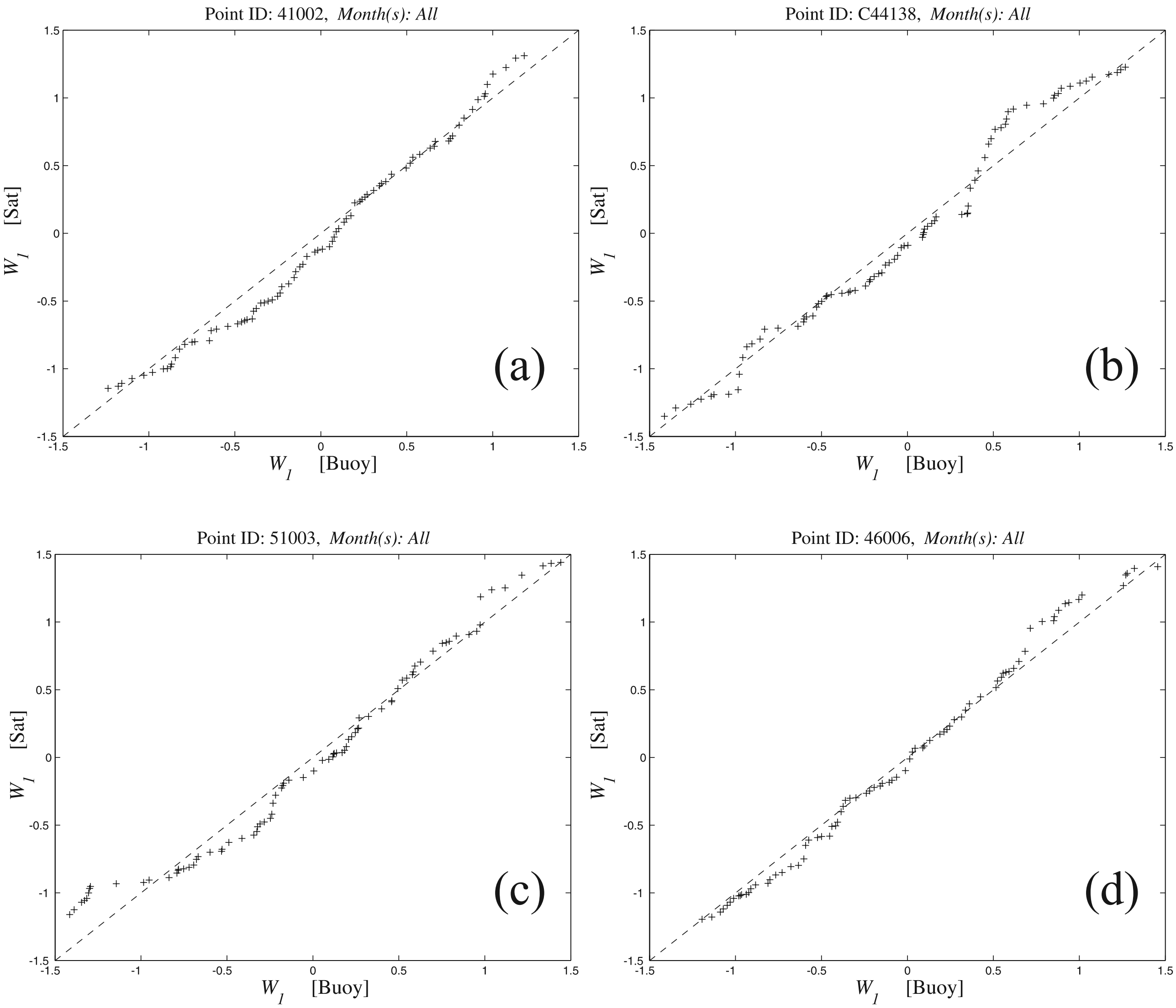

Figure 6. QQ plot of the quantiles of the empirical distributions of $W_1(\tau)$ for the points 41002, C44138, 51003, and 46006.

where $X_\mu(t;\ \beta_\mu)$ and $X_\sigma(t;\ \beta_\sigma)$ are cyclostationary random processes of the form

$$X_\mu(t;\beta_\mu) = \mu_1(t) + \sigma_1(t)W_1(t;\beta_\mu), \quad (13)$$

$$X_\sigma(t;\beta_\sigma) = \mu_2(t) + \sigma_2(t)W_2(t;\beta_\sigma), \quad (14)$$

modeling the variability of mean monthly values and mean monthly standard deviations, respectively, and $W(t;\ \gamma)$ is a stationary random process modeling the residual, state-by-state, variability (β_μ, β_σ, γ are appropriate stochastic arguments).

[39] In this way, the time series of significant wave height is given the structure of a multiple-scale compound stochastic process. Model (12) is a generalization of the nonstationary stochastic model introduced by *Athanassoulis and Stefanakos* [1995].

[40] The parameters of both processes $X_\mu(t;\ \beta_\mu)$ and $X_\sigma(t;\ \beta_\sigma)$ are obtained equally well by means of either buoy or satellite measurements, whereas the parameters of process $W(t;\ \gamma)$ are estimated using only buoy measurements.

[41] Thus, if, for example, a restricted amount of buoy data (say, one year) is available, and several years of satellite altimeter measurements, model (12) can be used to derive a many-year long time series by simulation, which combines all the basic statistical structure of the wave data. The whole methodology can be considered as an efficient way of blending (integrating) already available satellite data with a short (thus affordable and feasible) period of in situ measurements, to obtain an artifact of a long-term measured time series.

[42] If we compare equation (12) with equation (2), we see that the new feature introduced in the new modeling is an additional sampling variability of purely random nature. These random variabilities permit us to represent more accurately (realistically) the slowly varying (monthly) mean values. Also, the compound model (12)–(14) is expected to be more appropriate for simulation purposes.

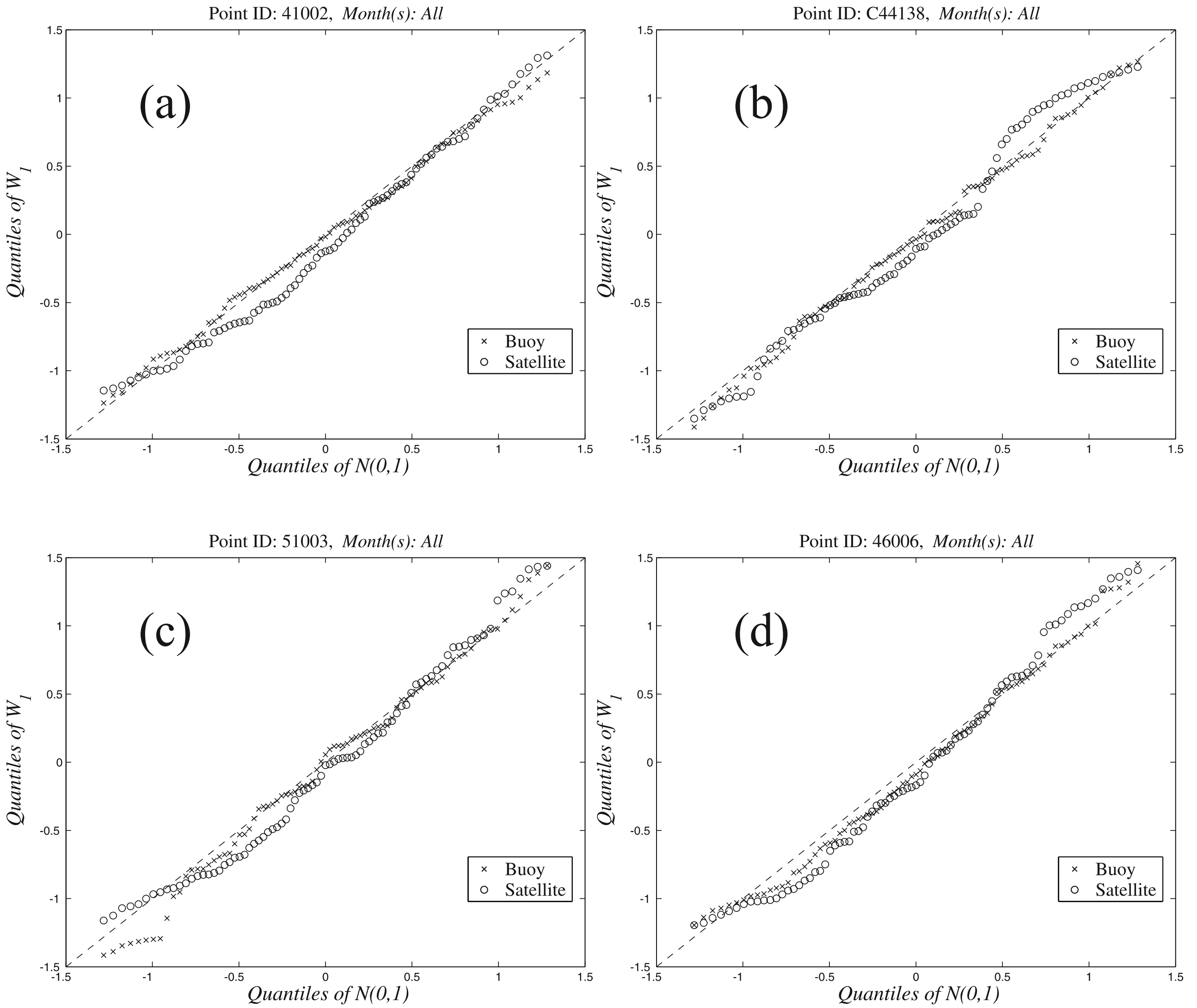

Figure 7. QQ plot of the quantiles of the empirical distributions of $W_1(\tau)$ against the quantiles of the standardized normal $N(0, 1)$ for the points 41002, C44138, 51003, and 46006.

[43] In order to further demonstrate the advantages of stochastic modeling (12), an example for the extreme-value prediction of the wave height will be given. A new enhanced method for the calculation of return periods of significant wave height, called MENU method, will be applied, exploiting the nonstationary modeling (12). For a description of MENU method; see *Stefanakos* [1999], *Stefanakos and Athanassoulis* [2006], and *Stefanakos and Monbet* [2006].

[44] For the application of MENU method, two components are required: (1) estimations of the seasonal mean value and seasonal standard deviation and their derivatives, and (2) estimations of the parameters of the joint probability density function of $W(t)$ and $W(t + \tau)$.

[45] First, estimates of 1 are produced using all available satellite data mentioned in Table 2, regardless if they coincide (or not) in time with buoy data. In Figure 8, the return periods of significant wave height for the point 46006 are shown using all the above mentioned estimates. If results based on buoy data are considered as the more reliable, then one can observe how close are the results

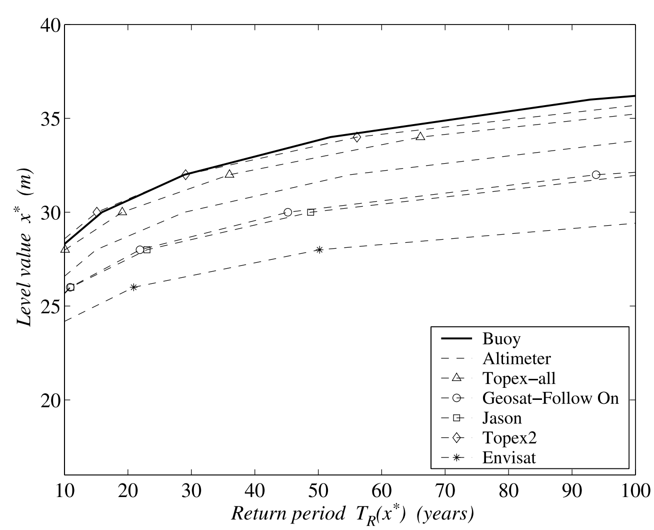

Figure 8. Return periods of significant wave height for the point 46006.

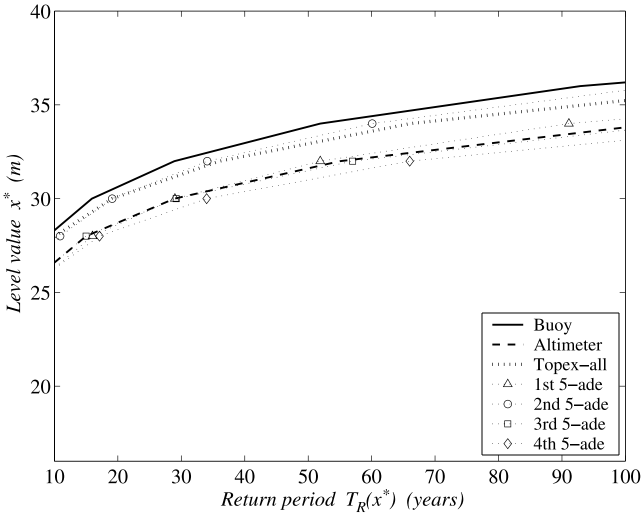

Figure 9. Return periods of significant wave height for the point 46006 using shorter periods.

based on the various satellite altimeters. For example, concerning the value of wave height corresponding to return period 50 years (the so-called 50-year wave height), all values range from 33.84 m (Buoy) to 27.99 m (Envisat). That is, the span between the minimum and the maximum value is only 17% of the maximum value. If, further, the Envisat estimate is excluded, the minimum value is 30.05 m (Jason) and the corresponding span is only 11.2% of the maximum value.

[46] Second, keeping estimates of 1 from only one altimeter (Topex-all), the estimates of 2 are calculated using (1) the whole amount of data and (2) shorter five-year long data sets. The estimated return periods are depicted in Figure 9, along with the results based on the buoy and the initial altimeter data. One can observe that shorter time series (of, say, five years) can equally well be used for extreme-value calculations. Moreover, the span between the minimum and the maximum value is less than in the previous case; namely 6.7% of the maximum value.

7. Conclusions

[47] In the present work, we have established a multiscale compound stochastic process for modeling the time series of significant wave height H_S; see equation (12). In this model, different kinds of measurements from different sources, each one resolving a different time scale, can be used for the estimation of the parameters of the model.

[48] In the present work we have used buoy measurements for the state-to-state correlation structure, and satellite altimeter measurements for describing the mean seasonal pattern and the seasonal variability. The key point is that we can use the seasonal pattern as obtained from appropriately defined satellite monthly values, in order to deseasonalise the buoy measurements. Then, by analysing the deseasonalised buoy measurements (which can be assumed to be a stationary stochastic process [see, e.g., *Athanassoulis and Stefanakos*, 1995; *Athanassoulis et al.*, 2003], we can estimate the correlation structure associated with the state-to-state scale by calculating the corresponding autocorrelation (or spectral density) function.

[49] In applying the model (12) (having determined its parameters using data from different sources), we should bear in mind that, in principle, it contains exactly those characters that have been resolved in the stage of the analysis procedure. For example, intermediate scale phenomena (e.g., energetic frontal passages) not complying with the constitutive assumptions of our model are not included in it. There are, however, various benefits in using a carefully estimated model like (12), instead of a unique measured sample. For example, the model is free from gaps (missing values), it permits the performance of sensitivity studies either by obtaining a population of realisations (by using various independent identically distributed (iid) samples of the generating random sequence) or by varying the parameters of the model and, also, it gives us the ability to treat more complex problems by combining the present model with other ones.

[50] Among various possible generalisations of the model (12) and its applications, the following seems to be the most interesting one. The generalisation towards the inclusion of other phenomena evolving in different time scales. For example, finer scale phenomena may be modeled by pulse-like processes [see, e.g., *Lopatoukhin et al.*, 2000, 2001], while longer-scale phenomena might be included by introducing additional (longer) periods in the cyclostationary model (12).

[51] **Acknowledgments.** The authors wish to acknowledge the National Data Buoy Center, National Oceanic and Atmospheric Administration (NDBC/NOAA), for freely providing its data through Internet, and Fugro OCEANOR for providing the satellite data. Also, J. Gagnon, Marine Environmental Data Service (MEDS), Fisheries and Oceans, Canada, is gratefully acknowledged for kindly providing the measured data sets from buoys C44137, C44138 and C44141. Figure 1 was produced by means of public domain software package GMT, created by P. Wessel, School of Ocean and Earth Science and Technology, University of Hawaii, and W. H. F. Smith, Laboratory for Satellite Altimetry, Oceanic Research and Applications Division, NOAA. The work of the first author has been co-sponsored by Nyfex & Company LLC.

References

Athanassoulis, G. A., and Ch. N. Stefanakos (1995), A nonstationary stochastic model for long-term time series of significant wave height, *J. Geophys. Res.*, *100*(C8), 16,149–16,162.

Athanassoulis, G. A., and Ch. N. Stefanakos (1998), Missing-value completion of nonstationary time series of wave data, paper presented at 17th International Conference on Offshore Mechanics and Arctic Engineering, Lisbon, Portugal.

Athanassoulis, G. A., Ch. N. Stefanakos, and S. F. Barstow (2003), A methodology for integrating wave data from different sources permitting a multiscale description of the wave climate variability, in *Advances in the Applications of Marine Climatology - The Dynamic Part of the WMO Guide to the Applications of Marine Climatology*, *WMO/TD-1081*, pp. 187–198, World Meteorol. Org., Geneva.

Barstow, S. F., and H. E. Krogstad (1993), Analysis of extreme waves and recent storms in the Norwegian Sea, paper presented at Climatic Trends and Future Offshore Design and Operation Criteria Workshop, Reykjavik, Iceland.

Carter, D. J. T. (1999), Variability and trends in the wave climate of the North Atlantic: A Review, paper presented at 9th International Offshore and Polar Engineering Conference, ISOPE'1999, Brest, France.

Carter, D. J. T., P. G. Challenor, and P. D. Cotton (1995), Surface wave statistics from satellite altimeters, in *Proceedings of RINA International Conference on Seakeeping and Weather*, *Pap. 10*, R. Inst. of Nav. Architects, London, UK.

Charriez, P., M. Olagnon, and J. Tournadre (1992), Confidence intervals associated with satellite measurement of wave and wind environment, paper presented at 11th International Conference on Offshore Mechanics and Arctic Engineering, OMAE'1992, Calgary, Canada.

Cooper, C. K., and G. Z. Forristall (1997), The use of satellite altimeter data to estimate the extreme wave climate, *J. Atmos. Oceanic Technol.*, *14*(2), 254–266.

Krogstad, H. E., and S. F. Barstow (1999), Satellite wave measurements for coastal engineering applications, *Coastal Eng.*, *37*(3–4), 283–307.

Lopatoukhin, L. J., I. V. Lavrenov, V. A. Rozhkov, A. V. Boukhanovsky, and A. B. Degtyarev (2000), Wind and wave climate of oil fields in the Barents, Pechora, and Kara Seas (in Russian, English abstract), paper presented at 6th International Conference on Ships and Marine Structures in Cold Regions, ICETECH, St. Petersburg, Russia.

Lopatoukhin, L. J., V. A. Rozhkov, V. E. Ryabinin, V. Swail, A. V. Boukhanovsky, and A. B. Degtyarev (2001), Estimation of extreme wind wave heights, *WMO/TD-1041*, 75 pp., World Meteorol. Org., Geneva.

Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravanan, and M. Visbeck (2001), North Atlantic climate variability: Phenomena, impacts and mechanisms, *Int. J. Climatol.*, *21*, 1863–1898.

Monaldo, F. (1988), Expected differences between buoy and radar altimeter estimates of wind speed and significant wave height and their implications on buoy-altimeter comparisons, *J. Geophys. Res.*, *93*(C3), 2285–2302.

Monaldo, F. (1990), Corrected spectra of wind speed and significant wave height, *J. Geophys. Res.*, *95*(C3), 3399–3402.

Panchang, V., L. Zhao, and Z. Demirbilek (1999), Estimation of extreme wave heights using GEOSAT measurements, *Ocean Eng.*, *26*(3), 205–225.

Stefanakos, Ch. N. (1999), Nonstationary stochastic modelling of time series with applications to environmental data, Ph.D. thesis, Dep. of Nav. Architecture and Mar. Eng., Natl. Tech. Univ. of Athens, Athens, Greece.

Stefanakos, Ch. N., and G. A. Athanassoulis (2001), A unified methodology for the analysis, completion and simulation of nonstationary time series with missing-values, with application to wave data, *Appl. Ocean Res.*, *23*(4), 207–220.

Stefanakos, Ch. N., and G. A. Athanassoulis (2003), Bivariate stochastic simulation based on nonstationary time series modelling, paper presented at 13th International Offshore and Polar Engineering Conference, ISOPE'2003, Honolulu, Hawaii.

Stefanakos, Ch. N., and G. A. Athanassoulis (2006), Extreme value predictions based on nonstationary time series of wave data, *Environmetrics*, *17*(1), 25–46.

Stefanakos, Ch. N., and K. A. Belibassakis (2005), Nonstationary stochastic modelling of multivariate long-term wind and wave data, paper presented at 24th International Conference on Offshore Mechanics and Arctic Engineering, OMAE'2005, Halkidiki, Greece.

Stefanakos, Ch. N., and V. Monbet (2006), Estimation of wave height return periods using a nonstationary time series modelling, paper presented at 25th International Conference on Offshore Mechanics and Arctic Engineering, OMAE'2006, Hamburg, Germany.

Tournadre, J. (1993), Time and space scales of significant wave heights, *J. Geophys. Res.*, *98*(C3), 4727–4738.

Tournadre, J., and R. Ezraty (1990), Local climatology of wind and sea state by means of satellite radar altimeter measurements, *J. Geophys. Res.*, *95*(C10), 18,255–18,268.

WASA Group (1998), Changing waves and storms in the northeastern Atlantic, *Bull. Am. Meteorol. Soc.*, *79*(5), 741–760.

Wilk, M. B., and R. Gnanadesikan (1968), Probability plotting methods for the analysis of data, *Biometrika*, *55*(1), 1–17.

G. A. Athanassoulis and Ch. N. Stefanakos, School of Naval Architecture and Marine Engineering, National Technical University of Athens, 9 Heroon Polytechniou str., 15773 Zografos, Athens, Greece. (chstef@naval.ntua.gr.)

S. F. Barstow, Fugro OCEANOR, N-7462 Trondheim, Norway.