Stoch Environ Res Risk Assess
DOI 10.1007/s00477-011-0540-2

ORIGINAL PAPER

Wave height characteristics in the north Atlantic ocean: a new approach based on statistical and geometrical techniques

George Galanis · Peter C. Chu · George Kallos · Yu-Heng Kuo · C. T. J. Dodson

Abstract The main characteristics of the significant wave height in an area of increased interest, the north Atlantic ocean, are studied based on satellite records and corresponding simulations obtained from the numerical wave prediction model WAM. The two data sets are analyzed by means of a variety of statistical measures mainly focusing on the distributions that they form. Moreover, new techniques for the estimation and minimization of the discrepancies between the observed and modeled values are proposed based on ideas and methodologies from a relatively new branch of mathematics, information geometry. The results obtained prove that the modeled values overestimate the corresponding observations through the whole study period. On the other hand, 2-parameter Weibull distributions fit well the data in the study. However, one cannot use the same probability density function for describing the whole study area since the corresponding scale and shape parameters deviate significantly for points belonging to different regions. This variation should be taken into account in optimization or assimilation procedures, which is possible by means of information geometry techniques.

Keywords Numerical wave prediction models · Distribution of significant wave height · Radar altimetry · Information geometry · Fisher information metric

G. Galanis (✉)
Section of Mathematics, Hellenic Naval Academy, Xatzikyriakion, 18539 Piraeus, Greece
e-mail: ggalanis@mg.uoa.gr

G. Galanis · G. Kallos
Division of Environmental Physics-Meteorology, Atmospheric Modeling and Weather Forecasting Group, University of Athens, School of Physics, University Campus, Bldg. PHYS-V, 15784 Athens, Greece

P. C. Chu · Y.-H. Kuo
Department of Oceanography, Graduate School of Engineering & Applied Science, Naval Postgraduate School, Monterey, CA 93943, USA

C. T. J. Dodson
School of Mathematics, Manchester University, Manchester M13 9PL, UK

1 Introduction

In a demanding scientific and operational environment, the validity of high quality sea state information is constantly increasing. This is in direct correspondence with the significant number of applications that are affected: climate change, transportation, marine pollution, wave energy production and ship safety can be listed among them.

One of the most credible approaches towards accurate sea state forecasting products is the use of numerical wave prediction systems in combination with atmospheric models (see, e.g., WAMDIG 1988; Lionello et al. 1992; Komen et al. 1994; Chu and Cheng 2008). Such systems have been proved successful for the simulation of the general sea state conditions on global or intermediate scale. However, when focusing on local characteristics usually systematic errors appear (see Janssen et al. 1987; Chu et al. 2004; Chu and Cheng 2007; Makarynskyy 2004, 2005; Greenslade and Young 2005; Galanis et al. 2006, 2009; Emmanouil et al. 2007). This is a multi-parametric problem in which several different issues are involved: The strong dependence of wave models on the corresponding wind input, the inability to capture sub-scale phenomena, the parametrization of certain wave properties especially in areas with complicated coastal formation where overshadowing and inaccurate refraction wave features emerge, as well as the lack of

Published online: 20 November 2011

a dense observation network which, as in the case of atmospheric parameters over land, could help on the systematic correction of initial conditions. The latter increases the added value of satellite records for ocean wave parameters.

Within this framework, there are two main ways that the research community followed over the last few years in order to minimize the effects of the above mentioned difficulties: Assimilating available observations in order to improve the initial conditions (Janssen et al. 1987; Breivik and Reistad 1994; Lionello et al. 1992, 1995; Abdalla et al. 2005; Emmanouil et al. 2007) and optimization of the direct model outputs by using statistical techniques like artificial neural networks (Makarynskyy 2004, 2005), MOS methods, Kalman filters, etc. (Kalman 1960; Kalman and Bucy 1961; Rao et al. 1997; Galanis and Anadranistakis 2002; Kalnay 2002; Galanis et al. 2006, 2009).

In both cases the main idea is the minimization of a "cost-function" that governs the evolution of the error. Similar approaches are also adopted in purely statistical models used for the estimation of wave height (see, for example, Vanem 2011; Vanem et al. 2011). At this point a critical simplification is usually made: The "distance" between observed and modeled values or distributions is measured by means of classical Euclidean geometry tools—using, for example, least square methods. This is, however, not always correct. Recent advances, in particular the rapid development of information geometry, suggest that the distributions are elements of more complicated structures, non Euclidean in general. More precisely, distributions of the same type form a manifold, which is the generalization of a Euclidean space and in which the underlying geometry may differ significantly from the classical one (see Amari 1985; Amari and Nagaoka 2000; Arwini and Dodson 2007, 2008). The exact knowledge of the framework in which the data sets or distributions under consideration are classified may give more accurate criteria and procedures for the optimization of the final results.

The purpose of the present work is twofold: At first, the sea state characteristics in the north Atlantic ocean are analyzed by means of a variety of statistical indices. Special attention is given to the probability distribution function of the significant wave height (the average height of the highest one-third waves in a wave spectrum). In a second step, the derived statistical information is utilized for the estimation of possible biases in numerical wave predictions based on novel techniques provided in the framework of information geometry.

For the above purposes simulated wave data obtained from the state-of-the art numerical WAve prediction Model (WAM) (Komen et al. 1994; WAMDIG 1988; Jansen 2000, Bidlot and Janssen 2003) and corresponding records from all the available satellites covering the study area (Radar Altimetry Tutorial project, Rosmorduc et al. 2009) are employed. The distributions that the two data sets form are recovered based on different statistical tests, and intercomparisons are attempted.

An application of the proposed methodology is outlined by focusing on a restricted area (northwestern coastline of France and Spain) avoiding lumping data from different wave climate regions. Alternative scenarios for the estimation of model biases are discussed. The results and ideas presented in this work could be exploited for designing and using new methods for the optimization of the initial conditions and the final outputs of numerical wave prediction systems since they could support more sophisticated ways of realizing the corresponding cost functions taking into account the geometric properties (scale and shape parameters for example) of the space that the data under study form, and avoiding simplifications that the classical pattern (least square methods) impose.

The presented material is organized as follows: In Sect. 2 the wave model, the data sets and the methodology used are described. The statistical results obtained for the observations and the corresponding modeled values are analyzed in Sect. 3. In particular, Sect. 3.1 focuses on the optimum choice of distributions that fit to the data in the study, while in Sect. 3.2 a detailed study of the results obtained in a restricted area (northwestern coastline of Spain and France) is presented based on descriptive statistics and distribution fitting. In Sect. 4 a new approach dealing with the problem of distance estimation between observations and modeled values is proposed by using techniques of information geometry. Section 4.1 is devoted to the introduction of some general notions and results while in Sect. 4.2 a direct application to the wave data in the study is attempted. Finally, the main conclusions of this work are summarized in Sect. 5.

2 Models, data sets and methodology

2.1 The wave model

The model used for wave simulation is WAM Cycle 4—ECMWF version (Jansen 2000; Bidlot and Janssen 2003). This is a third generation wave model which solves the wave transport equation explicitly without any assumptions on the shape of the wave spectrum (WAMDIG 1988; Komen et al. 1994). The model was operated by our group (Atmospheric Modeling and Weather Forecasting Group, University of Athens, http://www.mg.uoa.gr) in an operational/forecasting mode (that is using forecasted wind forcing and not reanalysis data) for a period of 12 months (year 2008) covering the north Atlantic ocean (Latitude 0°N–80°N, Longitude 100°W–30°E, Fig. 1). The wave spectrum was discretized to 30 frequencies (range 0.0417–0.54764 Hz logarithmically spaced) and 24 directions (equally spaced).

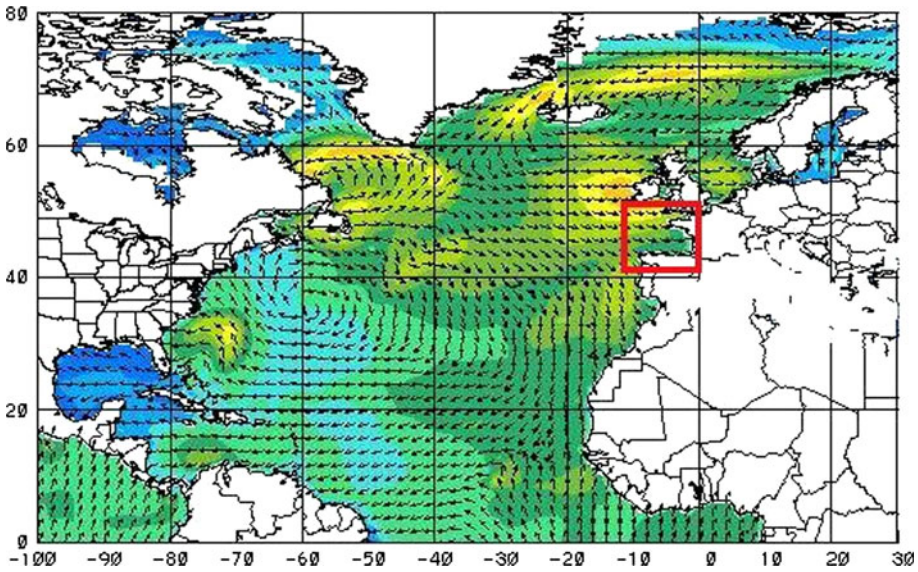

Fig. 1 The study area. The *red rectangle* denotes the borders of the restricted region. (Color figure online)

The horizontal resolution used was 0.5 × 0.5° and the propagation time step 300 s. WAM, ran on a deep water mode with no refraction, driven by 6-hourly wind input (10 m above sea level winds speed and direction) obtained by NCEP/GFS global model with horizontal grid resolution 0.5 × 0.5°. It should be noted that no assimilation procedure was employed since the available satellite data are used in our study as independent observations against which the modeled values are evaluated.

2.2 The satellite data

The observation data used in this study are obtained from the ESA-CNES joint project Radar Altimetry Tutorial (Rosmorduc et al. 2009). These data contain near-real time gridded observations for significant wave height obtained by merging all available relevant satellite records from official data centers: ERS-1 and ERS-2 (ESA), Topex/Poseidon (NASA/CNES), Geosat Follow-On (US Navy), Jason-1 (CNES/NASA), Envisat (ESA). The system is running daily in an operational mode. Each run is based on the available satellite data of the previous 2 days from which a merged map is generated. The produced interpolated outputs cover the whole area of study (0°N–80°N, 100°W–30°E) at a resolution of 1.0 × 1.0°. Data are cross-calibrated and quality controlled using Jason-1 as reference mission. The results are improved in case of additional mission availability. The period covered is again the whole year 2008.

2.3 Statistical approaches—methodology

Both observations and wave modeled data are studied by two statistical points of view: The first is based on descriptive statistical analysis methods where conventional indices are employed in order to capture the basic aspects of the data evolution spatially and temporally. The second approach is based on the study of the probability density function that fits to the available data. This is a complementary approach being able to provide additional information for the shape and scale of the data in the study including possible impact of extreme values. In this way, a complete view of the main characteristics of observational and simulated significant wave height values is obtained.

More precisely, the following statistical measures are used:

- *Mean value* of available data:

$$Mean = \mu = \frac{1}{N}\sum_{i=1}^{N} SWH(i)$$

Here *SWH* denotes the recorded (observed) or simulated significant wave height value and *N* the size of the sample.

- *Standard deviation*: $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(SWH(i)-\mu)^2}$
- *Coefficient of variation*:

$$c_v = \frac{\sigma}{\mu},$$

a normalized measure of the dispersion.

- *Skewness*:

$$g_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(SWH(i)-\mu)^3}{\sigma^3}$$

a measure of the asymmetry of the probability distribution.

- *Kurtosis*:

$$g_2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(SWH(i)-\mu)^4}{\sigma^4} - 3$$

that gives a measure of the "peakedness" of the probability distribution.

Additionally, the basic percentiles (P_5, P_{10}, $P_{25} = Q_1$, P_{50} = median, $P_{75} = Q_3$, P_{90} and P_{95}) are used.

Apart from the above descriptive statistical approach, the data in the study have been analyzed by a distributional point of view. More precisely, the optimum probability density functions (pdfs) that fit the observational and modeled significant wave height series are revealed. A variety of pdfs have been tested (Logistic, Normal, Gamma, Log-Gamma, Log-Logistic, Lognormal, Weibull, Generalized Logistic) at several levels of statistical significance by utilizing different fitting tests (Kolmogorov–Smirnov, Anderson–Darling as well as P–P and Q–Q plots) as well as statistical tools: Matlab (http://www.mathworks.com/products/matlab/) and EasyFit (http://www.mathwave.com/). The results reconfirm previous studies (Nordenstrøm 1973; Thornton and Guza 1983; Ferreira and Soares 1999, 2000; Prevosto et al. 2000; Muraleedharan et al. 2007; Gonzalez-Marco et al. 2008) proposing the Weibull distribution as a very good choice for fitting significant wave height data (see for example Fig. 2). However, the scale and shape parameters obtained vary spatially and temporarily (Sect. 3.1).

Apart from the above-mentioned "classical" statistical approaches, one of the main novelties proposed in this work is the utilization of non conventional statistical techniques obtained from a relatively new branch of Mathematics, the information geometry. This approach, discussed in detail in Sect. 4, allows the accurate description of the space to which the results under study belong and, based on the corresponding geometric properties, the better estimation of possible biasses. In this way, one avoids a classical simplification adopted in conventional statistics: the calculation of distances based on Euclidean measures.

3 Results and statistics

3.1 Probability density Function fitting

The data obtained for the significant wave height in the north Atlantic ocean, as simulated by the wave model (Sect. 2.1) and recorded by the Radar Altimetry Tool (Sect. 2.2), are studied here focusing on the distributions that they form. The use of all the statistical fitting tests mentioned earlier verified that, in most of the cases, the two-parameter Weibull distribution:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{(\alpha-1)} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, \quad \alpha, \beta > 0,$$

where α is the shape and β the scale parameter, fits well to the wave data at a statistical significance level of 0.05 or higher. An example is presented in Fig. 2. However, different parameters are obtained for the pdfs of satellite records and WAM values. On the other hand, a non-trivial spatial variability is revealed.

It should be noticed that the 3-parameter Weibull distribution fits also to the data in the study but with trivial differences from the 2-parameter case. Since an additional parameter would result in far more technical calculations in the proposed information geometry methodology without providing essential improvement of the obtained techniques, the 2-parameter Weibull has been adopted.

The data sets were partitioned into 3-monthly intervals (December–February, March–May, June–August and

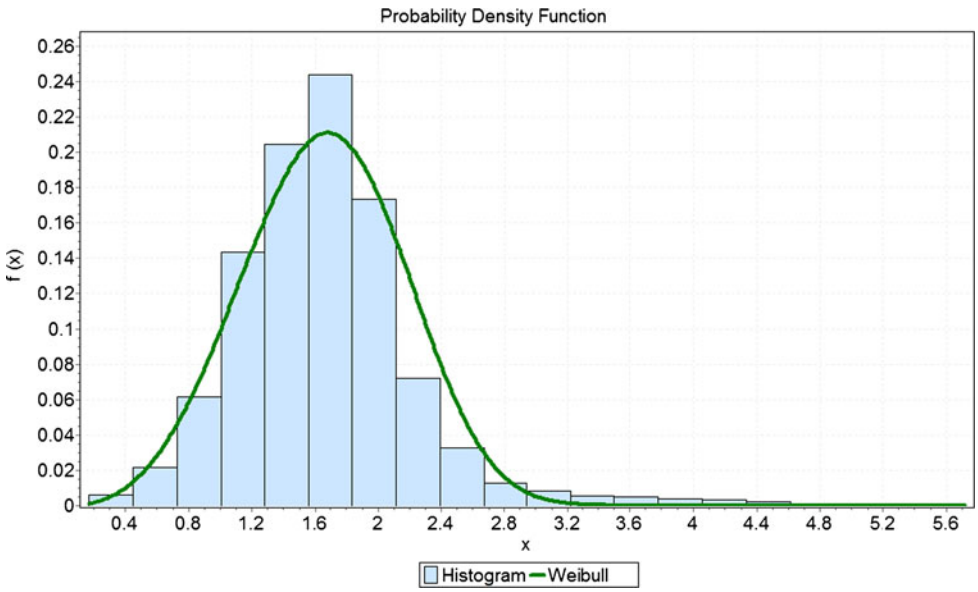

Fig. 2 Fitting of the 2-parameter Weibull distribution to the WAM modeled significant wave height data for May 2008

September–November) in order to have a clearer view of the seasonal variability of the sea state. In Figs. 3, 4, 5, and 6 the shape parameter of the obtained Weibull distribution fitted to the satellite data is plotted over the whole area of interest while Figs. 7, 8, 9, and 10 contain the corresponding values for the WAM outputs. It is worth underlining here that in both cases the values estimated are clearly increasing towards offshore areas. In particular, the maximum values emerged at the region southeast of Greenland and south of Iceland reaching values of 6.5 during the winter period (Figs. 3, 7). For the rest of the period, the same area keeps the maximum estimated values which, however, are significantly decreased. It is also noticeable that the estimated shape parameters for WAM outputs are elevated compared to those of satellite records in a relatively mild but systematic way.

The Weibull scale parameter values are presented in Figs. 11, 12, 13, and 14 for satellite records and Figs. 15, 16, 17, and 18 for their WAM counterparts. The wave model in this case seems to yield, in general, underestimated values. On the other hand, the increased values at the southern part of the domain, especially during summer months, can be partially attributed to the non uniform distribution of wave heights in this area.

It is important to underline at this point that the significant spatial variation of both shape and scale parameters, revealed in all the above cases, indicates that considering uniform ways of studying or correcting wave heights over

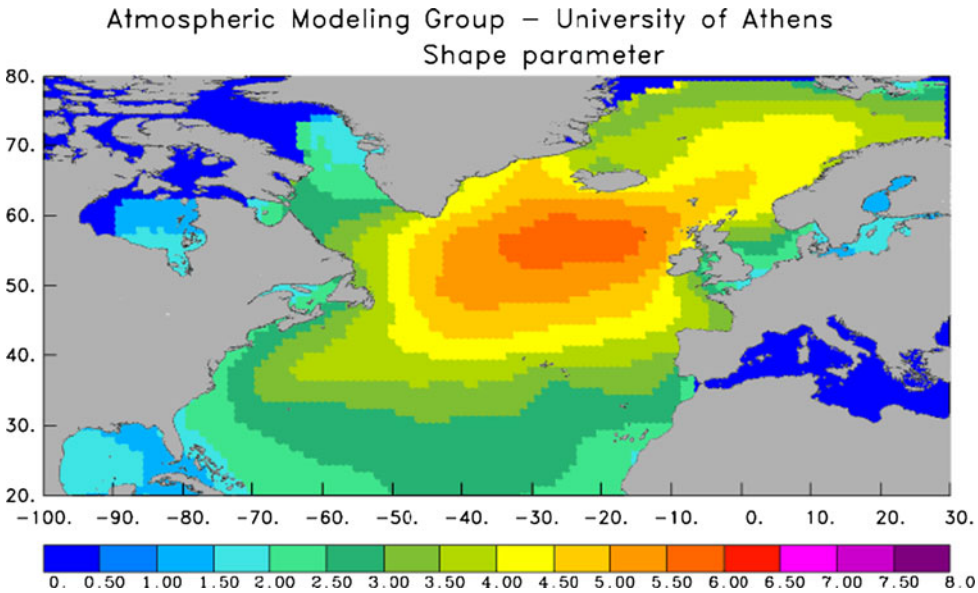

Fig. 3 The shape parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months December–February

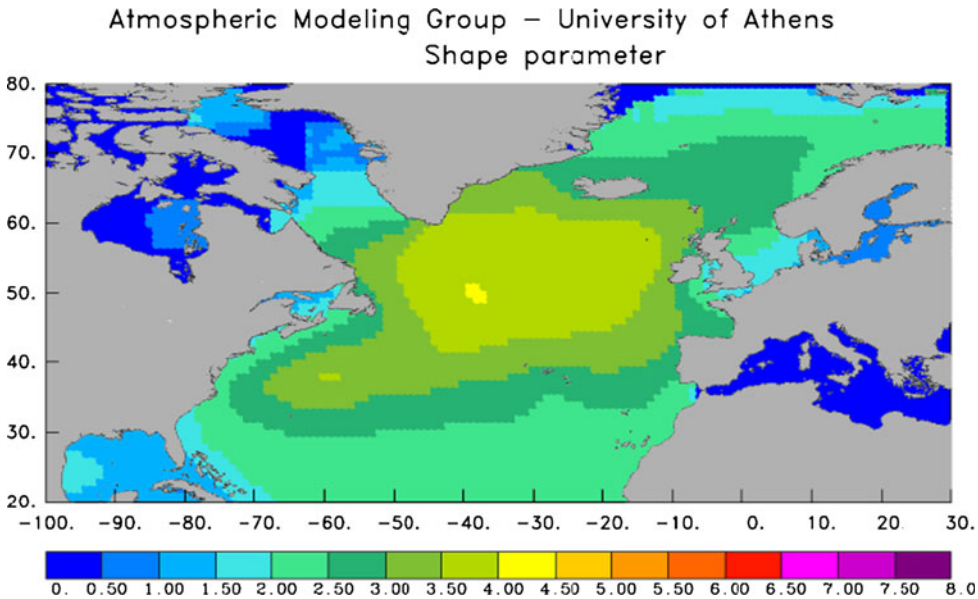

Fig. 4 The shape parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months March–May

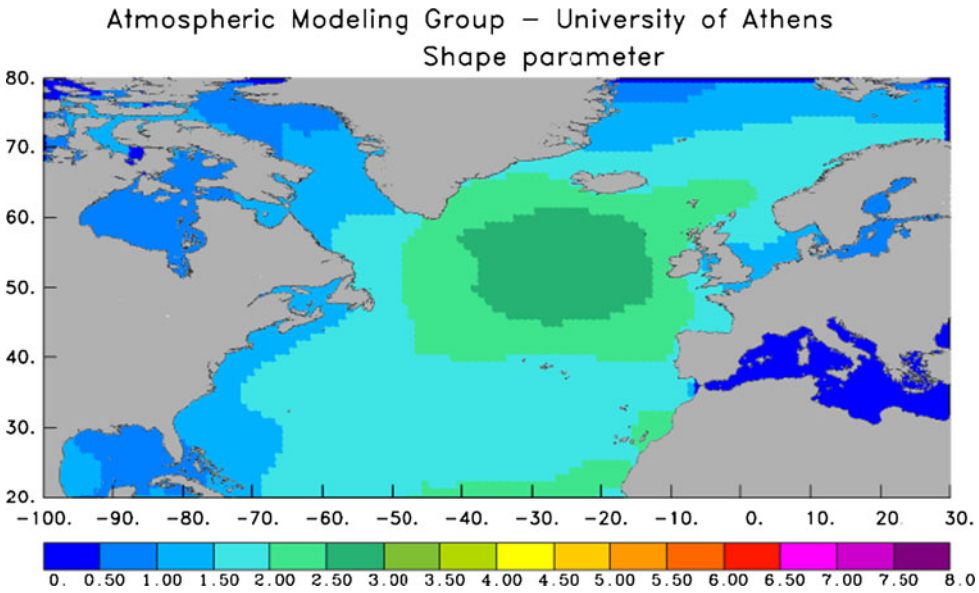

Fig. 5 The shape parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months June–August

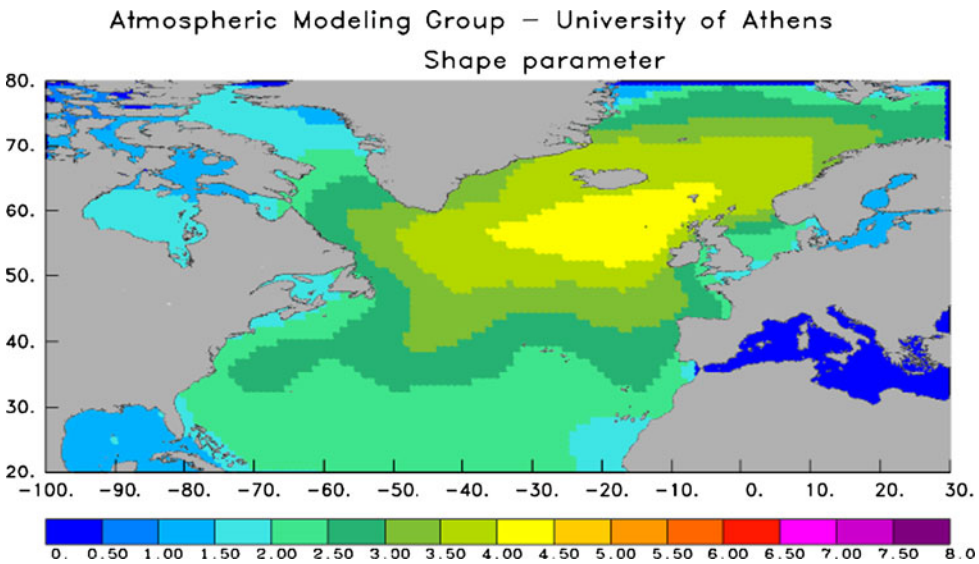

Fig. 6 The shape parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months September–November

the whole Atlantic ocean is an assumption of increased risk.

3.2 Focusing on a restricted area

In this section, the attention is focused on a restricted area of increased interest due to several activities raised recently concerning mainly wave energy applications: the northwest coastline of France and Spain (inner rectangle in Fig. 1). Indeed, several European and national projects require the exact knowledge of the local wave climate as well as the accurate sea state prediction in order to estimate the available energy potential.

The sea wave characteristics are studied here by two different points of view: Descriptive statistical measures, giving the main information for the data in the study, as well as distribution fitting in order to categorize them in a more uniform way, appropriate for the new techniques proposed in this work.

In Table 1 the main descriptive statistical indices, as described in Sect. 2.3, are presented in monthly intervals for the available satellite data. The time period covered is again the year 2008 and the sample size exceeds 2 million values. The corresponding results for the whole time period as well as divided in "Summer" (April–September) and "Winter" months (October–March) can be found in Table 2. The first conclusions are rather expected: The range of the observations as well as their mean value and variability are higher during winter. Furthermore, the increased kurtosis during March and May reveals that a

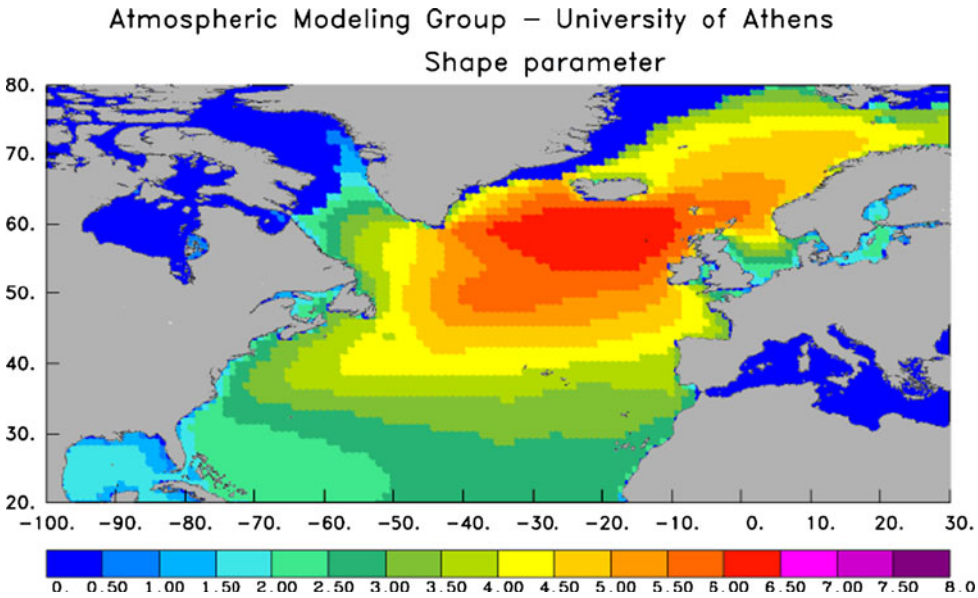

Fig. 7 The shape parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months December–February

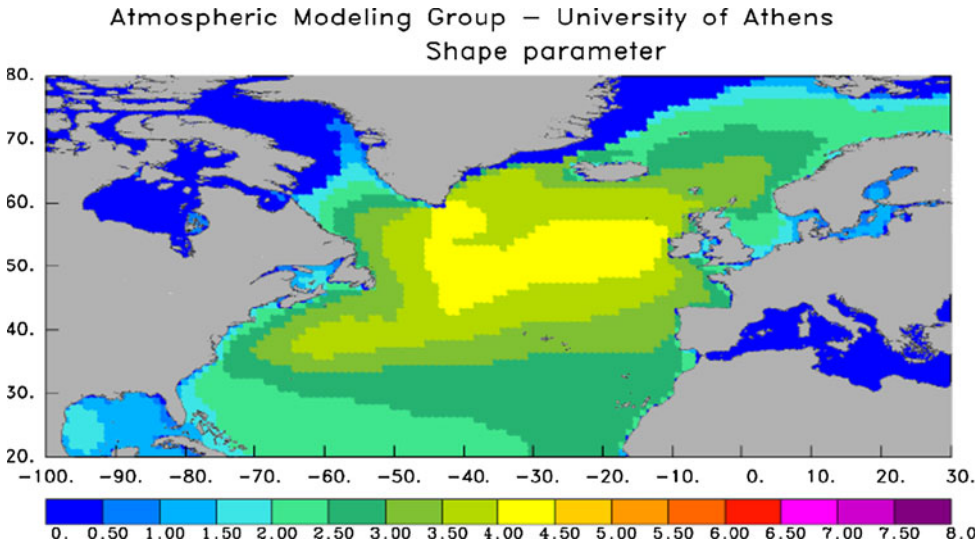

Fig. 8 The shape parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months March–May

significant part of the variability is related to non frequent outliers. The percentiles of the satellite records are presented in Tables 3 and 4.

The corresponding statistics for WAM outputs are presented in Tables 5, 6, 7, and 8. The basic descriptive statistical measures can be found in Tables 5 and 6 while the corresponding percentiles are presented in Tables 7 and 8. The same results are graphically represented in Figs. 19, 20, 21, and 22.

Interesting conclusions can be stated here for the accuracy of the numerical wave model WAM in an open sea area:

- WAM slightly, but constantly, overestimates wave heights through the whole study period (Fig. 19). The time independence of this divergence is worth mentioning.
- The variability of both observations and modeled values is increased during winter, something expected due to the unstable weather conditions. What needs to be mentioned is the consistently, again, higher values of the standard deviation of WAM (Fig. 20).
- Significant discrepancies exist between the ranges of the wave height results in the two sets (WAM simulations and satellite observations). This can be, at least partly, attributed to the fact that the observation data set is obtained by merging different satellite measurements, a procedure that always includes some smoothness of the final results due to interpolation. On the other hand, the well known difficulties of WAM on successfully simulating the swell decay (WISE Group 2007) contribute also to this problem.

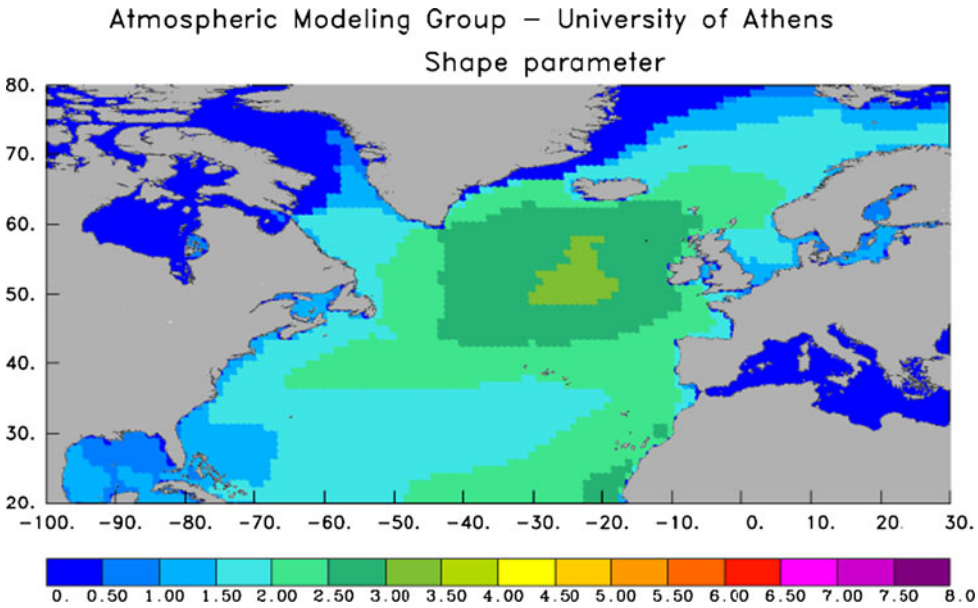

Fig. 9 The shape parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months June–August

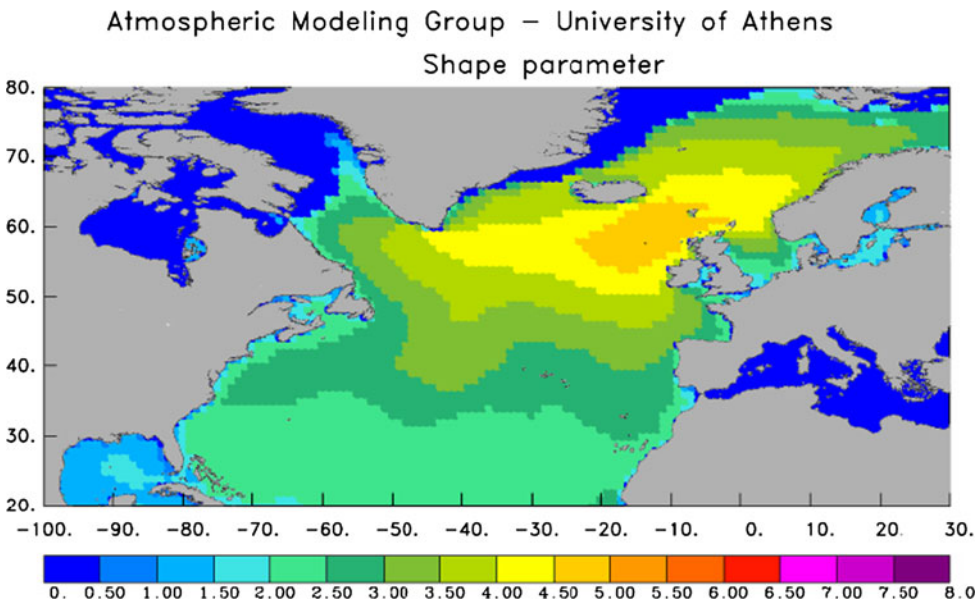

Fig. 10 The shape parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months September–November

- The relatively higher values of the corresponding percentiles as well as the monotonic increased distances between them (Tables 3, 4, 7, 8) confirm the overestimation of the data by WAM simulations and the non negligible influence of extreme values to their distribution. Although the purpose of this work is not to concentrate on problems of the wind/wave models that may lead to such deviations, it should be noted that the latter are closely related to the wind input used (atmospheric models discrepancies). On the other hand, the inclusion of current in wave forecasting is still lacking in WAM, while problems with the accurate simulation of the swell waves and especially their decay, as already mentioned earlier, also contribute to these discrepancies.

It is worth noticing at this point that when wind sea and swell components are considered, a spectral partitioning adopted will affect the accuracy of wind sea and swell statistics. The Hanson and Phillips formulation (developed by the Applied Physics Department of Johns Hopkins University, 2001) for labeling wind sea and swell is commonly applied. The main drawback of this approach is related to fully developed wind seas with a small wind decay but still in the same direction of the wave field, as shown by Quentin (2002), and later by Loffredo et al. (2009); if the new condition cannot satisfy the formulation adopted by Hanson and Phillips, the old wind sea will be treated as swell and the new wind sea set to zero. Further, as documented in Loffredo et al. (2009),

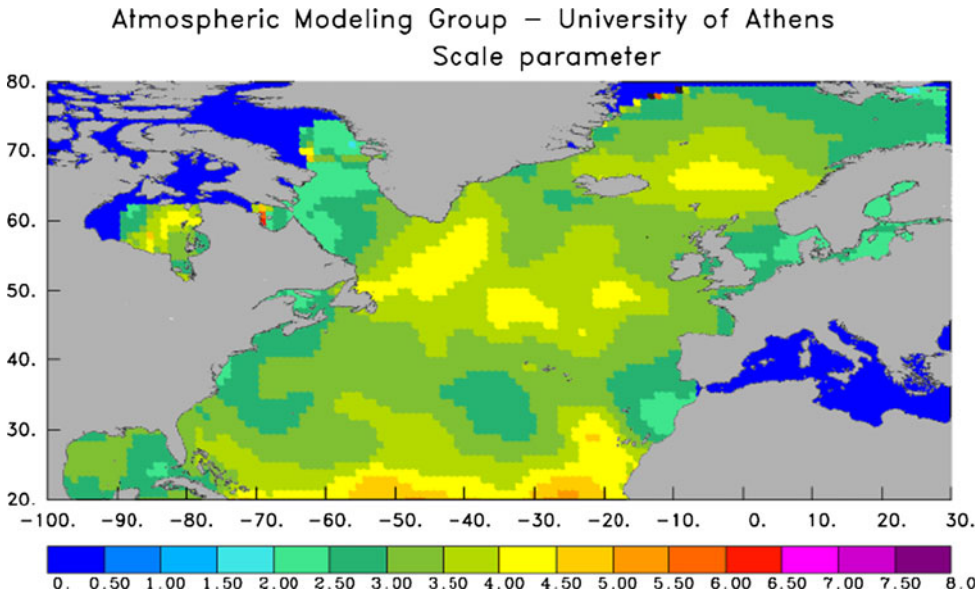

Fig. 11 The scale parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months December–February

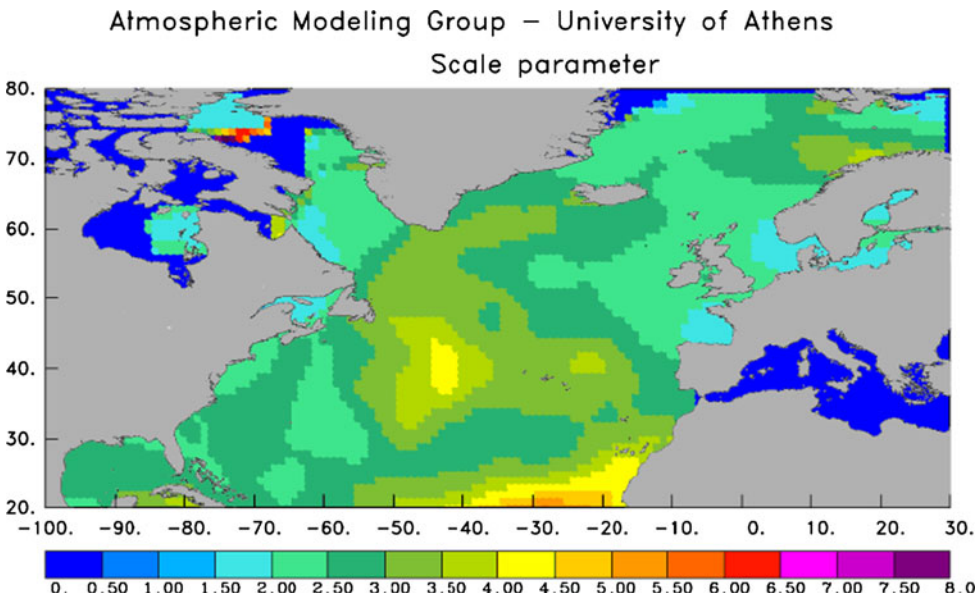

Fig. 12 The scale parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months March–May

the Hanson and Phillips formulation for labeling wind sea and swell may increase the number of wind seas as compared to other commonly used approaches for partitioning of wind sea and swell.

- Skewness is increased in WAM outputs compared to the observations (Fig. 21). This higher positive asymmetry indicates that a non-negligible portion of the modeled significant wave height is concentrated to relatively smaller values something that is less obvious in the corresponding observations.
- Elevated kurtosis for WAM outputs can be attributed to the increased influence of extreme values. This situation is more obvious during March and the summer months (Fig. 22).

Studying now the same data from a distribution fitting point of view, following the methodology discussed in Sect. 3.1, the following points may be emphasized:

- The 2-parameter Weibull distribution seems to fit well to the data in the study both for WAM and observed values.
- The shape parameter (α) both for the recorded and simulated values of SWH seems to deviate from the case of Rayleigh distribution (Tables 9, 10, 11, 12; Fig. 23) where $\alpha = 2$. The latter was the pdf proposed in previous works (e.g., Muraleedharan et al. 2007) indicating that the use of the general 2-parameter Weibull probability density function is more appropriate.

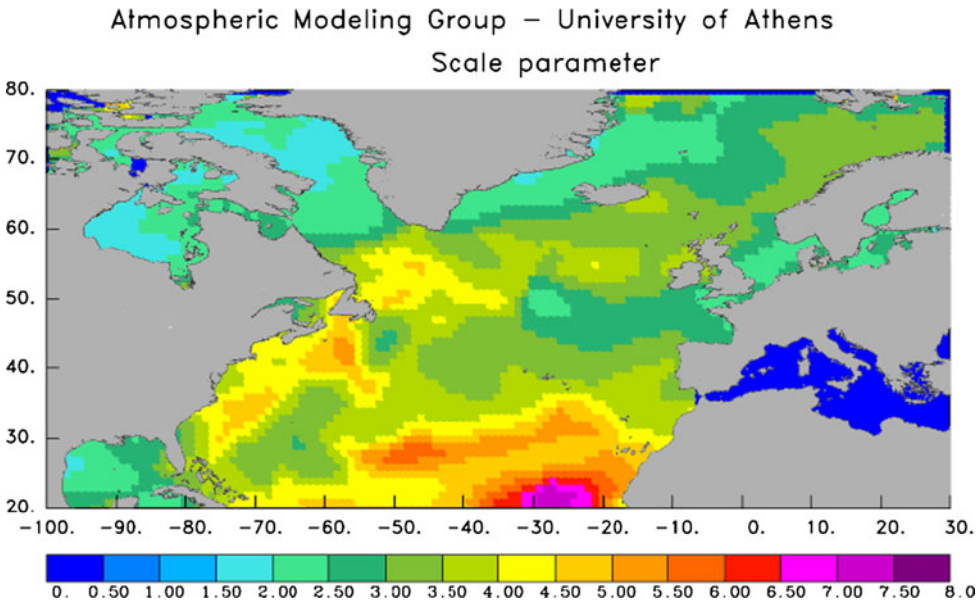

Fig. 13 The scale parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months June–August

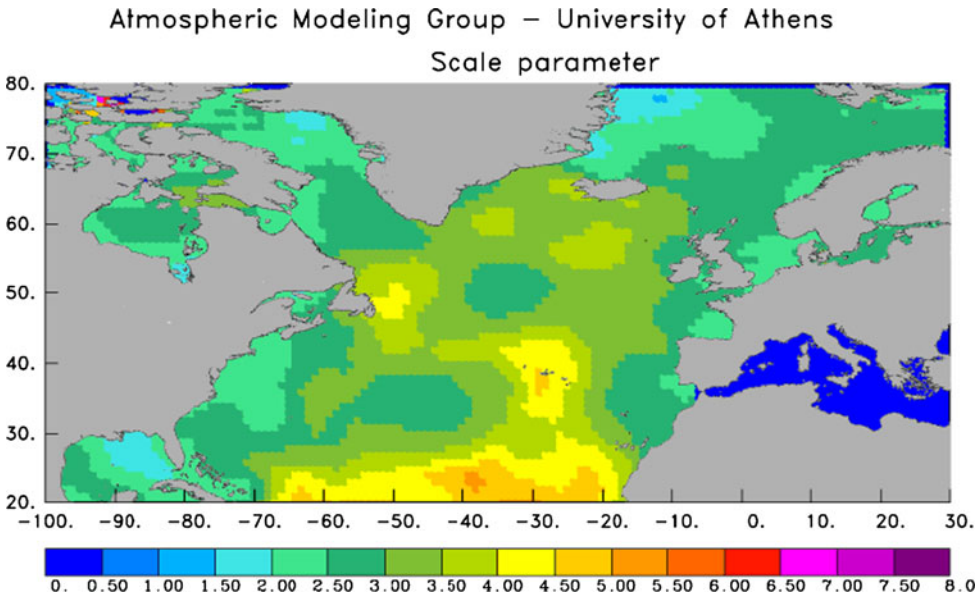

Fig. 14 The scale parameter of the Weibull distributions that fit to the significant wave height satellite data over the north Atlantic ocean for the months September–November

- The increased values of the scale parameter (β) for WAM (Fig. 24) reconfirms the overestimation of modeled values as already noticed based on the descriptive statistical measures. Moreover, the values of β for both cases follow the pattern of the mean values being reduced during summer months.
- The discrepancies between the parameters of the Weibull distributions obtained for satellite records and modeled wave height values are not major. Therefore, the techniques described in Sect. 4.2.1 for estimating the distance between WAM outputs and the corresponding observations can be exploited.

4 Estimation of the distance between observations and simulated values using information geometrical techniques

In the previous sections special attention was given on the main statistical characteristics as well as the distributions formed by WAM values and the corresponding satellite records for the area of the north Atlantic ocean. The obtained results reveal non negligible differences between the two data sets that should be taken into consideration in order to optimize the accuracy of the wave model. Some new ideas towards this direction based on information

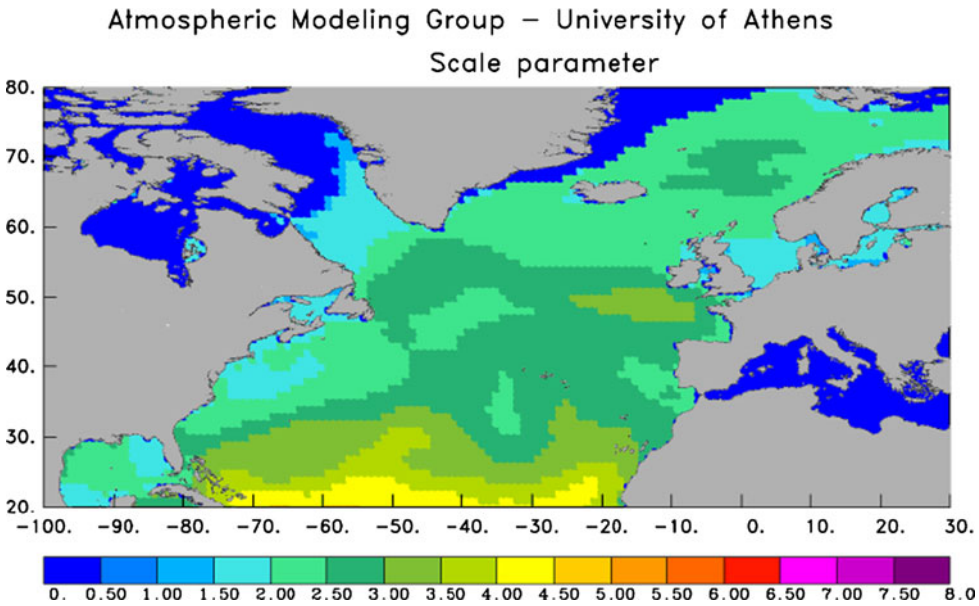

Fig. 15 The scale parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months December–February

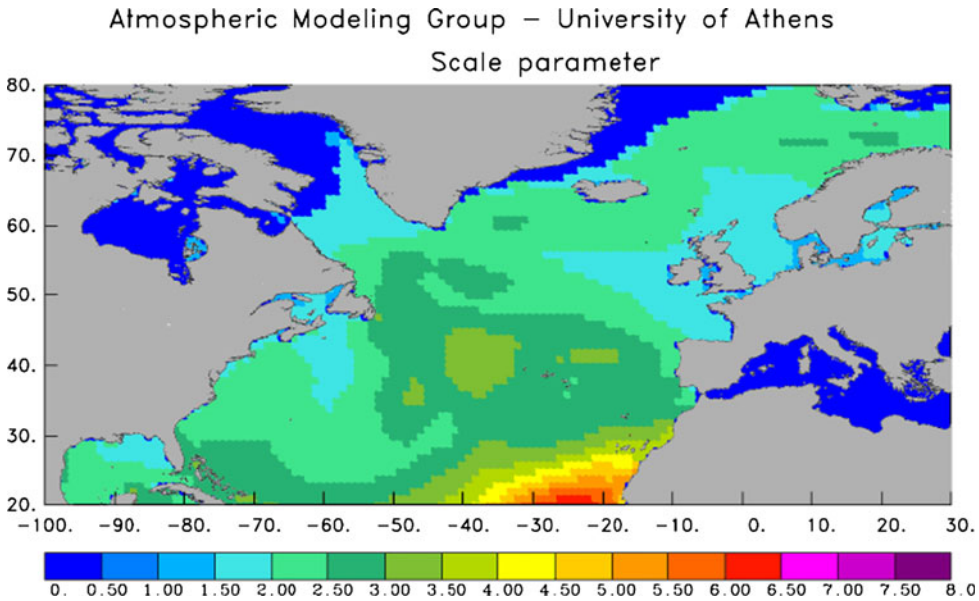

Fig. 16 The scale parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months March–May

geometry (IG) techniques are discussed in the present work. More precisely, having already defined the best-fitting distributions to the data in the study, a detailed description of the space that they form is attempted, the corresponding geometric entities are investigated and new techniques are proposed for the accurate estimation of the distance between observations and modeled values.

4.1 Basic information geometric concepts

In order to make this work as self-contained as possible, a short presentation of the main notions and terminology of information geometric techniques needed here follows. More details and results can be found in Amari 1985; Amari and Nagaoka 2000; Arwini and Dodson 2007, 2008.

Information geometry is a relatively new branch of mathematics in which the main idea is to apply methods and techniques of non-Euclidean geometry to probability theory and stochastic processes. In particular, information geometry realizes a smoothly parametrized family of probability distributions as a manifold on which geometrical entities such as Riemannian metrics, distances, curvature and affine connections can be introduced. To be more precise, a family of probability distributions

$$S = \{p_\xi = p(x;\xi) | \xi = [\xi_1, \xi_2, \ldots, \xi_n] \in \Xi\} \quad (1)$$

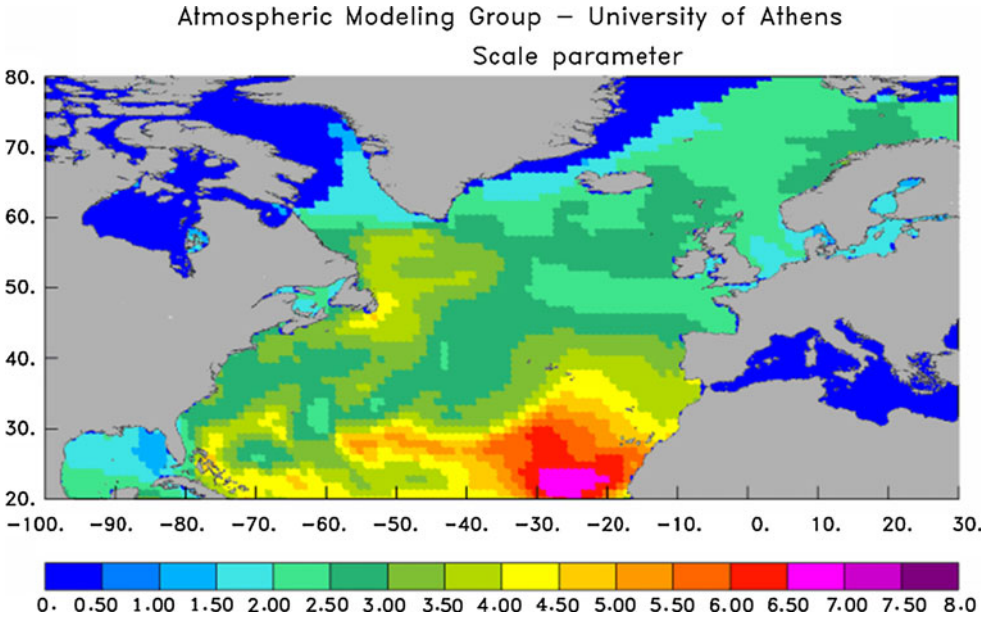

Fig. 17 The scale parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months June–August

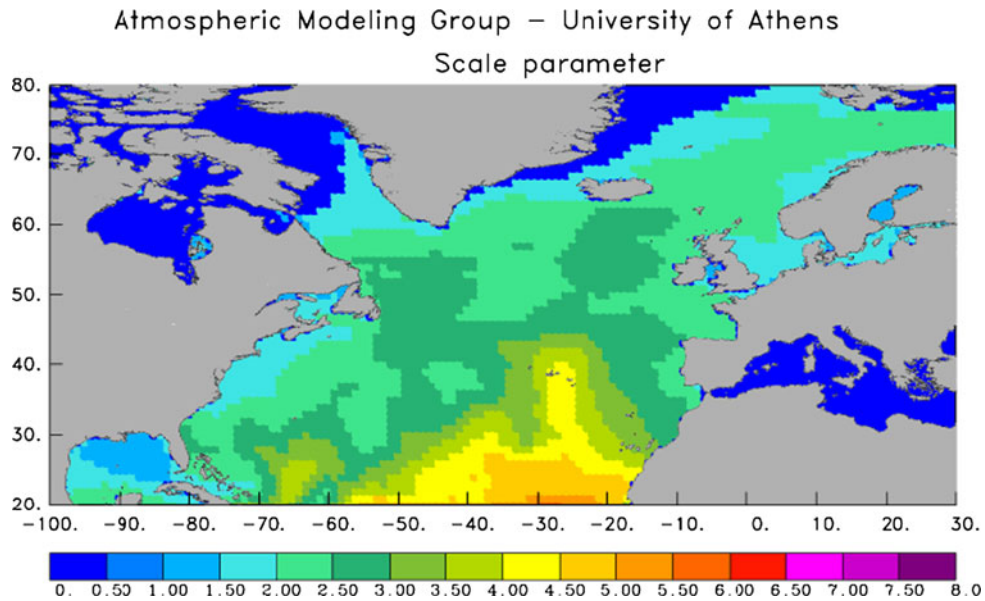

Fig. 18 The scale parameter of the Weibull distributions that fit to the WAM modeled significant wave height over the north Atlantic ocean for the months September–November

Table 1 The main statistical parameters for satellite data in the restricted area per month

Statistical parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Range	6.25	5.75	8.23	4.36	3.02	3.33	3.03	4.47	4.69	4.40	6.21	6.72
Mean	3.66	2.70	3.49	2.33	1.46	1.50	1.70	2.07	2.07	2.56	2.73	3.22
Std. deviation	1.16	1.06	1.42	0.79	0.53	0.50	0.61	0.78	0.92	0.79	1.25	1.15
Coef. of variation	0.32	0.39	0.41	0.34	0.37	0.33	0.36	0.38	0.45	0.31	0.46	0.36
Skewness	0.24	0.49	1.14	0.44	1.15	0.83	0.84	1.06	0.82	0.55	0.83	0.75
Kurtosis	−0.31	−0.57	1.46	−0.30	1.47	0.78	0.04	0.70	0.28	−0.13	−0.01	0.46

where each element may be parametrized using the n real valued variables $[\xi_1, \xi_2, \ldots, \xi_n]$ in an open subset Ξ of $\mathbb{R}^n$ while the mapping $\xi \rightarrow p_\xi$ is injective and smooth, is called a n-dimensional statistical manifold. The geometrical entities in a statistical manifold are dependent on the *Fisher information matrix* which at a point ξ is a $n \times n$ matrix

Table 2 The main statistical parameters for satellite data in the restricted area summarized for the whole study period, the summer and winter months

Statistical parameter	Overall	Summer	Winter
Range	5.04	3.82	6.26
Mean	2.46	1.86	3.06
Std. deviation	0.91	0.69	1.14
Coef. of variation	0.37	0.37	0.37
Skewness	0.76	0.86	0.66
Kurtosis	0.32	0.49	0.15

$$G(\xi) = [g_{ij}(\xi)], \tag{2}$$

defined by

$$g_{ij}(\xi) = E_{x|\xi}\left[\partial_i \ell(x;\xi)\partial_j \ell(x;\xi)\right] = \int \partial_i \ell(x;\xi)\partial_j \ell(x;\xi)p(x;\xi)dx, \quad i,j = 1,2,\ldots,n. \tag{3}$$

Here ∂_i stands for the partial derivative with respect to the i-th factor, ℓ is the log-likelihood function:

$$\ell(x;\xi) = \ell_\xi(x) = \log[p(x;\xi)] \tag{4}$$

and

$$E_{x|\xi}[f] = \int f(x)p(x;\xi)dx \tag{5}$$

denotes the expectation with respect to the distribution p.

The matrix $G(\xi)$ is always symmetric and positive semi-definite (Amari and Nagaoka 2000). If, in addition, $G(\xi)$ is positive definite, then a Riemannian metric (see Spivak 1965, 1979; Dodson and Poston 1991) can be defined on the statistical manifold corresponding to the inner product induced by the Fisher information matrix on the natural basis of the coordinate system $[\xi_i]$:

$$g_{ij} = \langle \partial_i | \partial_j \rangle. \tag{6}$$

This Riemannian metric is called the *Fisher metric* or the *information metric*. The corresponding geometric properties of this framework are characterized by the so-called Christoffel symbols $\left(\Gamma^i_{jk}\right)$ defined by the relations:

Table 4 Percentiles for satellite data in the restricted area for the whole study period, the summer and winter months

Percentile	Overall	Summer	Winter
P_5	0.62	0.44	0.81
P_{10}	1.24	0.96	1.52
$P_{25} = Q1$	1.43	1.11	1.75
P_{50} (median)	1.77	1.35	2.20
$P_{75} = Q3$	2.30	1.73	2.88
P_{90}	3.00	2.21	3.80
P_{95}	3.75	2.89	4.62

$$\Gamma_{jk,h}(\xi) = E_\xi\left[\left(\partial_j\partial_k\ell_\xi + \frac{1}{2}\partial_j\ell_\xi\partial_k\ell_\xi\right)(\partial_h\ell_\xi)\right], \quad i,\ j\ ,h = 1,2,\ldots,n, \tag{7}$$

$$\Gamma_{jk,h} = \sum_{i=1}^{2} g_{hi}\Gamma^i_{jk}(h = 1,2). \tag{8}$$

The minimum distance between two elements f_1 and f_2 of a statistical manifold S is defined by the corresponding *geodesic* ω which is the minimum length curve that connects them. Such a curve

$$\omega = (\omega_i) : \mathbb{R} \to S \tag{9}$$

satisfies the following system of 2nd order differential equations:

$$\omega_i''(t) + \sum_{j,k=1}^{n} \Gamma^i_{jk}(t)\omega_j'(t)\omega_k'(t) = 0, \quad i = 1,2,\ldots,n. \tag{10}$$

under the conditions $\omega(0) = f_1, \omega(1) = f_2$.

It worth noticing that information geometric techniques have been, directly or not, tested on different applications. Iguzquiza and Chica-Olmo (2008), for example, utilized the Fisher information matrix for geostatistical simulations for restricted samples. On the other hand, Cai et al. (2002) applied information theoretic analysis on self-clustering of amino acids along protein chains. Resconi (2009) is also based on non-Euclidean geometric tools for a risk analysis study. However, to the author's knowledge, the current work is the first try to apply such tools on meteorology/oceanography.

Table 3 Percentiles for satellite data in the restricted area per month

Percentile	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
P_5	1.89	1.28	1.67	1.21	0.80	0.82	0.92	1.15	0.84	1.49	1.17	1.64
P_{10}	2.13	1.47	1.98	1.43	0.89	0.95	1.06	1.29	1.02	1.63	1.37	1.92
$P_{25} = Q1$	2.74	1.86	2.52	1.76	1.08	1.14	1.27	1.50	1.34	1.94	1.74	2.41
P_{50} (median)	3.71	2.54	3.12	2.22	1.35	1.42	1.55	1.89	1.92	2.48	2.39	3.02
$P_{75} = Q3$	4.46	3.49	4.24	2.82	1.69	1.79	1.98	2.41	2.55	3.08	3.58	3.95
P_{90}	5.08	4.23	5.33	3.52	2.21	2.19	2.71	3.34	3.38	3.63	4.61	4.83
P_{95}	5.56	4.63	6.37	3.83	2.51	2.38	2.97	3.70	3.92	4.03	5.07	5.37

Table 5 The main statistical parameters for WAM outputs in the restricted area per month

Statistical parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Range	11.28	8.69	18.27	7.09	5.55	6.35	9.11	8.59	7.64	7.47	9.26	11.06
Mean	4.06	3.13	3.99	2.54	1.66	1.74	2.00	2.11	2.28	2.74	2.92	3.57
Std. deviation	1.50	1.24	1.99	1.07	0.56	0.63	0.85	1.05	1.09	1.10	1.47	1.53
Coef. of variation	0.37	0.40	0.50	0.42	0.34	0.36	0.43	0.50	0.48	0.40	0.50	0.43
Skewness	0.82	0.79	1.92	0.75	1.24	1.14	1.77	1.96	1.30	0.66	1.19	1.11
Kurtosis	1.07	0.50	6.61	0.68	4.17	2.57	5.95	5.21	2.38	0.44	1.52	1.90

Table 6 The main statistical parameters for WAM outputs in the restricted area summarized for the whole study period, the summer and winter months

Statistical parameter	Overall	Summer	Winter
Range	9.20	7.39	11.01
Mean	2.73	2.06	3.40
Std. deviation	1.17	0.88	1.47
Coef. of variation	0.43	0.42	0.43
Skewness	1.22	1.36	1.08
Kurtosis	2.75	3.49	2.01

Table 8 Percentiles for WAM outputs in the restricted area for the whole study period, the summer and winter months

Percentile	Overall	Summer	Winter
P_5	1.23	0.98	1.48
P_{10}	1.48	1.16	1.80
$P_{25} = Q1$	1.92	1.46	2.37
P_{50} (median)	2.52	1.88	3.15
$P_{75} = Q3$	3.31	2.48	4.15
P_{90}	4.24	3.17	5.32
P_{95}	4.95	3.74	6.16

4.2 Application to WAM outputs and satellite data

The significant wave height data obtained in the present study, both from satellite records and WAM model, have been proved in Sect. 3.1 to follow 2-parameter Weibull distributions. The corresponding parameters however seem to differ between the two data sets and to fluctuate in time and space.

In this section different scenarios will be discussed, based on information geometric techniques, concerning the optimum way of estimating the distance between the two data sets. The obtained results can be exploited in assimilation or optimization procedures for better defining the involving cost functions targeting at the improvement of the final modeled products.

Following the formalism presented in Sect. 4.1, the family of the two parameter Weibull distributions can be considered as a 2-dimensional statistical manifold with $\xi = [\alpha, \beta]$, $\Xi = \{[\alpha, \beta];\ \alpha \text{ and } \beta > 0\}$ and

$$\mathrm{p}(\mathrm{x};\xi) = \frac{\alpha}{\beta}\left(\frac{\mathrm{x}}{\beta}\right)^{\alpha-1} e^{-\left(\frac{\mathrm{x}}{\beta}\right)^{\alpha}} \tag{11}$$

The log-likelihood function becomes:

$$\begin{aligned}\ell(x;\xi)) &= \log[p(x;\xi)] \\ &= \log\alpha - \log\beta + (\alpha-1)(\log x - \log\beta) - \left(\frac{x}{\beta}\right)^{\alpha}\end{aligned} \tag{12}$$

while the Fisher information matrix (Amari 1985; Amari and Nagaoka 2000) takes the form:

$$G(\alpha,\beta) = \begin{bmatrix} \alpha^2\beta^2 & \beta(1-\gamma) \\ \beta(1-\gamma) & \frac{6(\gamma-1)^2+\pi^2}{6\alpha^2} \end{bmatrix} \tag{13}$$

Here $\gamma = \lim_{n\to+\infty}\left(\sum_{k=1}^{n} 1/k - \ln n\right) \cong 0.577215$ is the Euler Gamma. The Christoffel symbols of the 0-connection

Table 7 Percentiles for WAM outputs in the restricted area per month

Percentile	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
P_5	1.97	1.45	1.64	1.04	0.87	0.93	1.04	1.05	0.92	1.14	1.16	1.49
P_{10}	2.31	1.75	2.02	1.30	1.03	1.08	1.20	1.19	1.13	1.46	1.37	1.91
$P_{25} = Q1$	2.94	2.25	2.68	1.79	1.30	1.29	1.44	1.44	1.51	1.96	1.84	2.57
P_{50} (median)	3.89	2.90	3.57	2.38	1.62	1.63	1.78	1.80	2.08	2.59	2.59	3.35
$P_{75} = Q3$	4.93	3.81	4.88	3.20	1.93	2.07	2.38	2.48	2.79	3.41	3.64	4.25
P_{90}	6.01	4.92	6.25	3.96	2.27	2.56	3.14	3.38	3.69	4.23	4.98	5.55
P_{95}	6.76	5.55	7.36	4.54	2.57	2.91	3.65	4.34	4.40	4.75	5.88	6.67

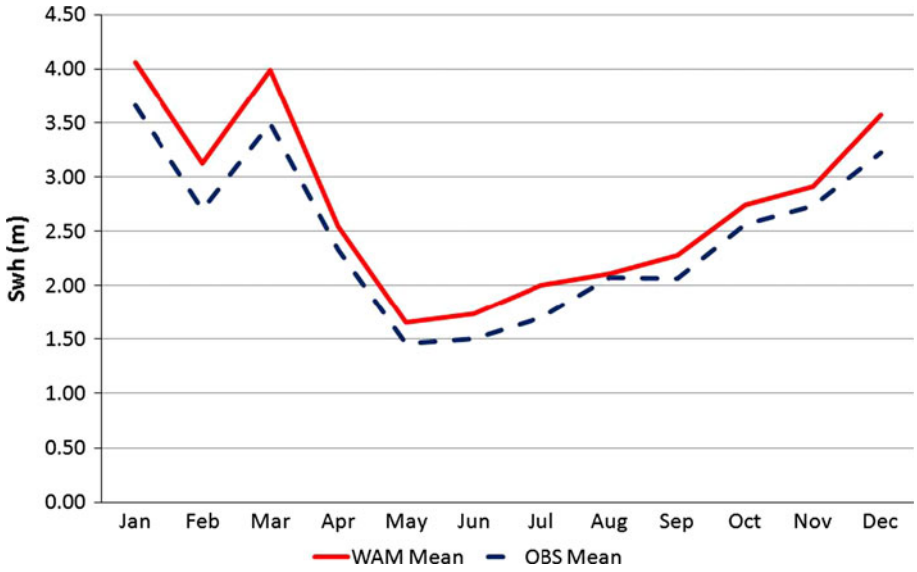

Fig. 19 The evolution of mean value for WAM modeled and satellite recorded significant wave height in the restricted region through the whole study period

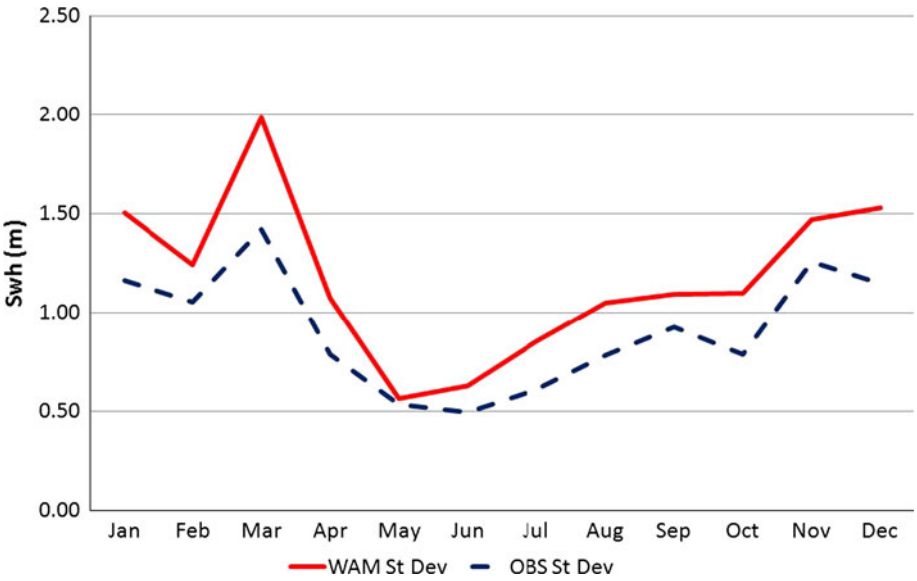

Fig. 20 The evolution of standard deviation for WAM modeled and satellite recorded significant wave height in the restricted region through the whole study period

(see Amari and Nagaoka 2000; Arwini and Dodson 2007, 2008) in this case are:

$$\Gamma^1_{11} = \frac{6\left(\gamma\alpha - \alpha - \frac{\pi^2}{6}\right)}{\pi^2\beta} \quad \Gamma^2_{11} = \frac{-\alpha^3}{\pi^2\beta^2}$$
$$\Gamma^1_{21} = \Gamma^1_{12} = \frac{6\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha}$$
$$\Gamma^2_{21} = \Gamma^2_{12} = \frac{6\alpha(1-\gamma)}{\pi^2\beta} \qquad (14)$$
$$\Gamma^1_{22} = -\frac{6(1-\gamma)\beta\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha^3}$$
$$\Gamma^2_{22} = -\frac{6\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha}$$

The main-general question that is raised is:

> With the Weibull parameters α and β known, which is the optimum way of estimating the distance between observations and WAM outputs?

Two scenarios are proposed.

4.2.1 Working for points in the same neighborhood

A first approach supported by the information geometric techniques can be based on the projection of the distributions, which fit the data sets, to the same tangent space. Then, their distance is calculated based on the corresponding inner product. For example, the Weibull distribution followed by the satellite data obtained in the

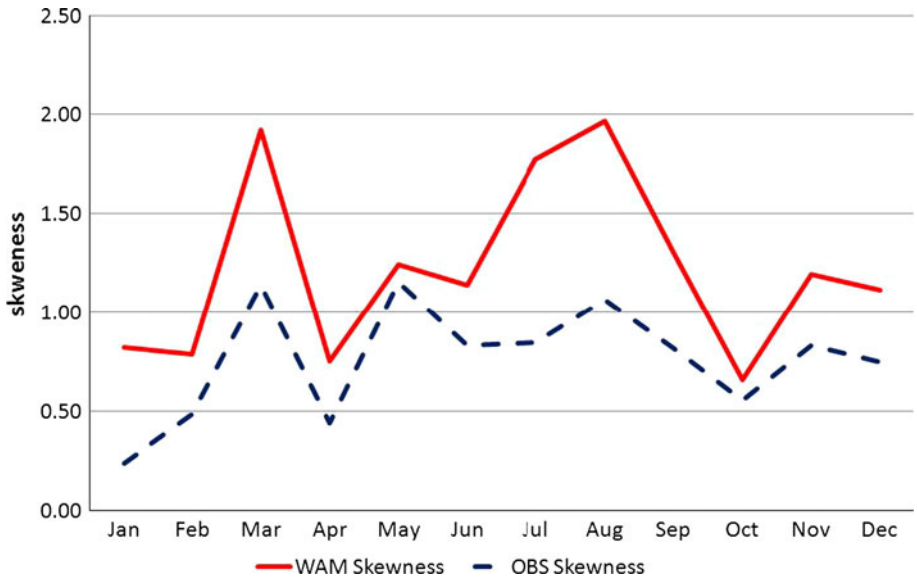

Fig. 21 The evolution of skewness for WAM modeled and satellite recorded significant wave height in the restricted region through the whole study period

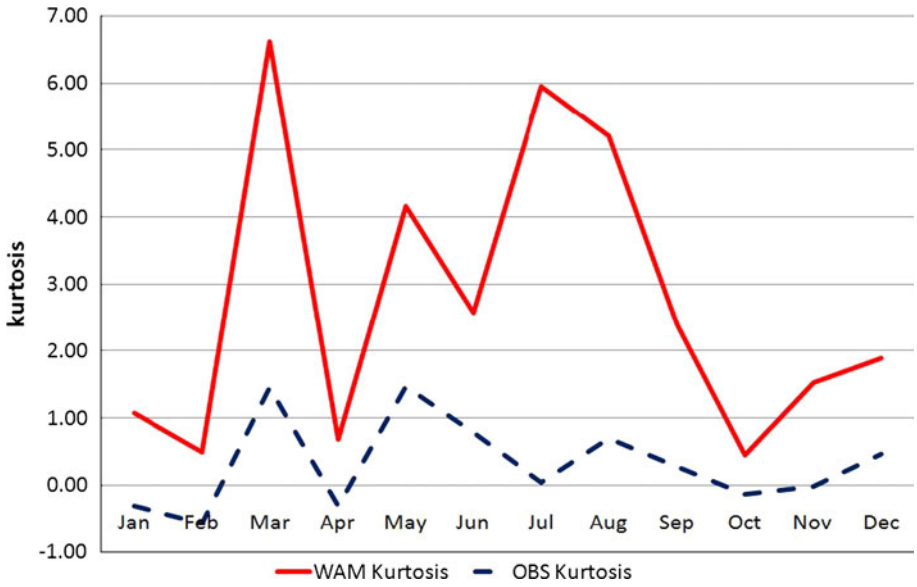

Fig. 22 The evolution of kurtosis for WAM modeled and satellite recorded significant wave height in the restricted region through the whole study period

Table 9 Weibull parameters for satellite data in the restricted area per month

Weibull parameters	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
α	3.70	3.06	3.16	3.48	3.53	3.81	3.53	3.43	2.74	4.01	2.69	3.49
β	4.05	3.00	3.89	2.59	1.61	1.66	1.88	2.30	2.30	2.82	3.05	3.57

restricted area of Northwestern European coastline (Sect. 3.2) during August 2008 has shape parameter $\alpha = 3.43$ and scale $\beta = 2.30$ m (see Tables 9, 11). The corresponding values for WAM modeled significant wave height are $\alpha = 2.82$ and $\beta = 2.35$ m. Therefore, the observed and modeled data can be considered as elements $u_0 = W(3.43, 2.30)$, $u_1 = W(2.82, 2.35)$ of the statistical manifold S of all Weibull distributions being projected to the same tangent space. The latter can be chosen to be the tangent space $T_{u_o}S$ of u_0 where the inner product, and hence the distances, is defined by the Fisher information matrix at u_0:

Table 10 Weibull parameters for satellite data in the restricted area for the whole study period, the summer and winter months

Weibull parameters	Summer	Winter	Overall
α	3.39	3.42	3.35
β	2.73	2.06	3.40

$$G = \begin{bmatrix} (3.43)^2(2.30)^2 & 2.30(1-\gamma) \\ 2.30(1-\gamma) & \frac{6(\gamma-1)^2+\pi^2}{6(3.43)^2} \end{bmatrix} = \begin{bmatrix} 62.23 & 0.97 \\ 0.97 & 0.16 \end{bmatrix}, \tag{15}$$

The correct distance between u_0 and u_1 would be in this case:

$$d(u_o, u_1) = \sqrt{(u_o - u_1)^T G (u_o - u_1)} \tag{16}$$

which should replace the classical $\sqrt{(u_o - u_1)^T (u_o - u_1)}$ used by least square methods in assimilation or other optimizations procedures.

In a similar way one may also estimate the distance between any elements of the same tangent space. The novelty compared to the classical least square methods *is the use of the Fisher information matrix instead of the identity*, incorporating in this way the geometrical structure of the manifold of distributions.

The present approach simplifies the estimation of the distance since there is no need of solving complicated systems of differential equations as those corresponding to geodesics (relation 10). However, an approximation error should be expected.

4.2.2 Using geodesics

The full exploitation of the information geometric framework proceeds by the use of geodesic curves $\omega = (\omega_1, \omega_2) : \mathbb{R} \to S$ for the estimation of the distances on a statistical manifold S. This results to a system of second order differential equations (Eq. 10). By substituting the values of the Christoffel Γ^i_{jk} (Spivak 1965, 1979; Dodson and Poston 1991) obtained for the Weibull statistical manifold (Eq. 14), the system becomes:

Table 12 Weibull parameters for WAM outputs in the restricted area for the whole study period, the summer and winter months

Weibull parameters	Summer	Winter	Overall
α	3.01	3.10	2.92
β	3.03	2.29	3.78

$$\begin{aligned}
&\omega_1''(t) + \frac{6\left(\gamma\alpha - \alpha - \frac{\pi^2}{6}\right)}{\pi^2\beta}\left(\omega_1'(t)\right)^2 + \frac{12\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha} \\
&\omega_1'(t)\omega_2'(t) - \frac{6(1-\gamma)\beta\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha^3}\left(\omega_2'(t)\right)^2 = 0, \\
&\omega_2''(t) - \frac{\alpha^3}{\pi^2\beta^2}\left(\omega_1'(t)\right)^2 + \frac{12\alpha(1-\gamma)}{\pi^2\beta}\omega_1'(t)\omega_2'(t) \\
&- \frac{6\left(\gamma^2 - 2\gamma + \frac{\pi^2}{6} + 1\right)}{\pi^2\alpha}\left(\omega_2'(t)\right)^2 = 0,
\end{aligned} \tag{17}$$

In most of the cases, this cannot be solved analytically and the use of approximation methods is necessary.

A relevant example is presented here. The Weibull distribution that fits to the satellite data obtained in the restricted area of Northwestern European coastline during August 2008 are used again. Therefore, the probability density function of the satellite records has shape parameter $\alpha = 3.43$ and scale $\beta = 2.30$ m, while for the relevant WAM outputs $\alpha = 2.82$ and $\beta = 2.35$ m. The minimum length curve that gives the distance between the two distributions is a two dimensional curve $\omega = (\omega_1, \omega_2)$ that can be obtained as the solution of the differential system:

$$\omega_1'' - 0.82\left(\omega_1'\right)^2 + 0.65\omega_1'\omega_2' - 0.02\left(\omega_2'\right)^2 = 0$$

$$\omega_2'' - 0.77\left(\omega_1'\right)^2 + 0.77\omega_1'\omega_2' - 0.32\left(\omega_2'\right)^2 = 0$$

under the conditions

$$\omega_1(0) = 3.43, \quad \omega_2(0) = 2.30, \quad \omega_1(1) = 2.82, \quad \omega_2(1) = 2.35$$

By numerically solving this nonlinear system, one reaches the solution presented in Fig. 25. The graphical representations of the geodesic are far from being linear which should be the case if the classical (linear regression) statistical approach has been adopted. In the same figure, the

Table 11 Weibull parameters for WAM outputs in the restricted area per month

Weibull parameters	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
α	3.34	3.08	2.70	2.78	3.64	3.52	3.17	2.82	2.67	2.97	2.53	2.88
β	4.50	3.48	4.43	2.84	1.84	1.92	2.22	2.35	2.54	3.06	3.25	3.98

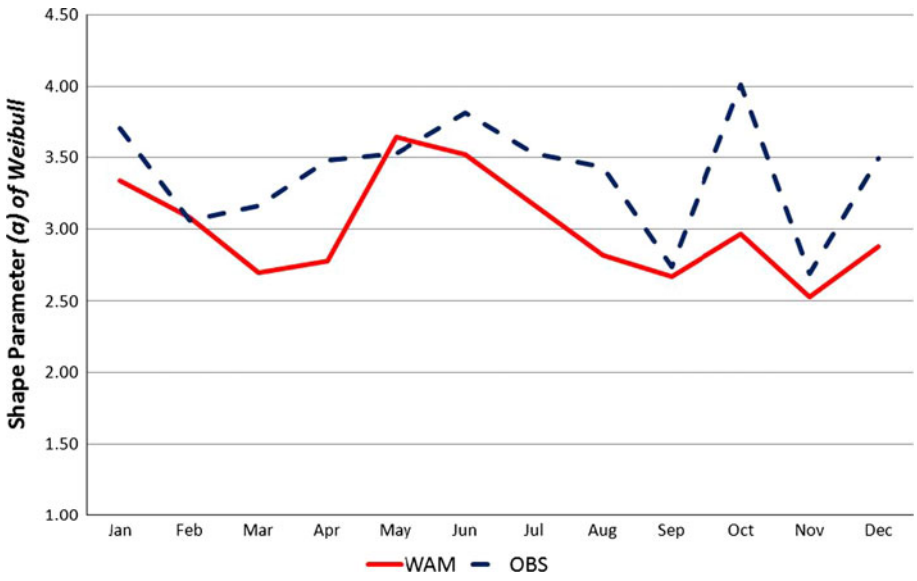

Fig. 23 The shape parameter α of the Weibull distributions that fit to WAM modeled and satellite recorded significant wave height in the restricted region through all months of 2008

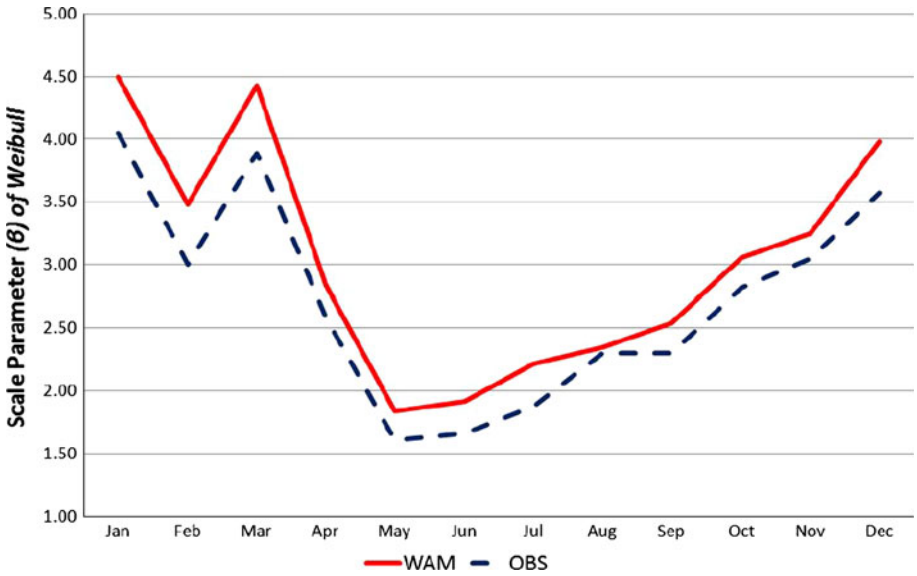

Fig. 24 The scale parameter β (in meters) of the Weibull distributions that fit to WAM modeled and satellite recorded significant wave height in the restricted region through all months of 2008

spray of other geodesics emanating from the same initial point (3.43, 2.30) is also presented.

An attempt to visualize further the above approach is made in Fig. 26 a and b where the statistical manifolds formed by the satellite records and WAM outputs (monthly values) are presented as elements of the non-Euclidean space that the totality of Weibull distributions define.

5 Conclusions

The results of the numerical wave prediction model WAM for an area of increased interest (the north Atlantic ocean) concerning the significant wave height over a period of 1 year were evaluated against corresponding satellite measurements. Special attention was given to the probability distribution functions formed. The outcomes were utilized in order to discuss novel statistical procedures for the quantification of the bias, based on a relatively new branch of mathematics, information geometry, which has not been exploited so far in atmospheric sciences and oceanography. The most important conclusions made follow:

- Similar but not identical two-parameter Weibull distributions seem to fit to the observational and modeled significant wave height values. In particular, the shape parameter values both for satellite records and WAM outputs increase as moving to offshore areas. The maximum values emerge at the sea area southern of

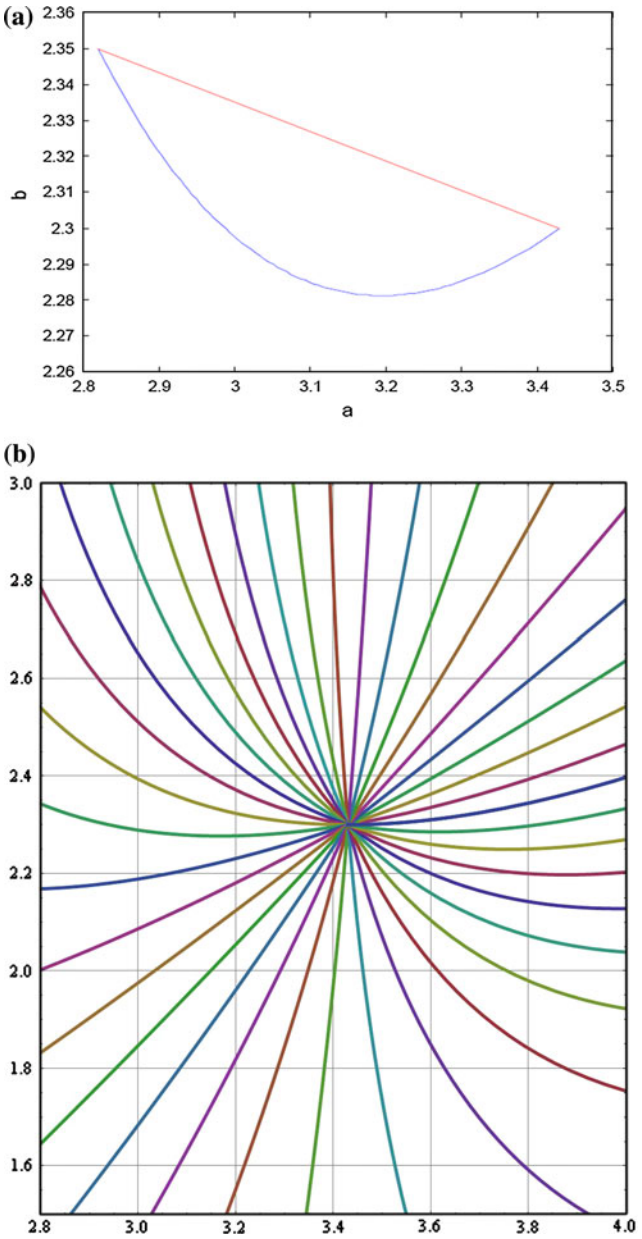

Fig. 25 **a** The graphical representation of the geodesic (*curved line*) that gives the minimum length curve connecting the satellite observations with WAM outputs for August 2008. The *straight line* corresponds to the Euclidean (classical) geodesic. **b** The graphical representation of a numerical solution spray of geodesics emanating from (3.43,2.30) including the one to (2.82, 2.35) that gives the minimum length curve connecting the satellite observations with WAM outputs for August 2008. (Color figure online)

Iceland. On the other hand, increased scale parameters for both observations and model outputs in the western coast of central Africa can be attributed to non uniform distribution of the sea state in this area.

- The estimated shape parameters for WAM outputs outmatch those of satellite records in a mild but systematic way while the scale analogous values for the wave model outputs, concerning the whole area of study, are slightly underestimated indicating that the satellite records form stretched out distributions.
- WAM seems slightly but consistently to overestimate the significant wave height through the whole study period. The same holds also for the variability of the simulated values as expressed by the standard deviation that constantly outmatch that of observations.
- Non negligible differences exist between the ranges of SWH values for WAM outputs and observations. This can be attributed to WAM problems with swell decay as well as to the way of calculation (merging) of satellite records.
- An increased part of the distribution of modeled values, compared to the corresponding observations, is concentrated at relatively smaller values. This positive asymmetry is highlighted by the increased values of skewness.
- The variability of WAM outputs is more dependent on extreme values than satellite observations as the increased kurtosis indicates, especially during the summer months.
- The parameters of the probability density functions that fit the modeled and observational data appear to have significant spatial variation. As a result, the use of the same cost function in optimization systems for the whole domain of the study is a serious simplification. In this respect information geometry techniques provide possible ways out.
- Two different scenarios for the estimation of distances between the data sets in the study are discussed taking into account that the Weibull distributions form a 2-dimensional non-Euclidean space, in particular a Riemannian manifold, avoiding simplifications that classical statistics adopt (use of Euclidean distances):
  - The first approach utilizes the tangent spaces at the points of interest avoiding solving the complicated differential systems that arise within the information geometric framework. An approximation error is expected in this case.
  - In the second scenario the proposed geometric methodology is fully exploited and the distances are obtained based on the geodesic curves of the statistical manifold that the data in the study form.
- In both cases the obtained results deviate from those resulted in the classical case.
- An example/application of the proposed techniques to the northwestern coastline of France and Spain is discussed clarifying the alternative way for the estimation of distances between observations and modeled values.

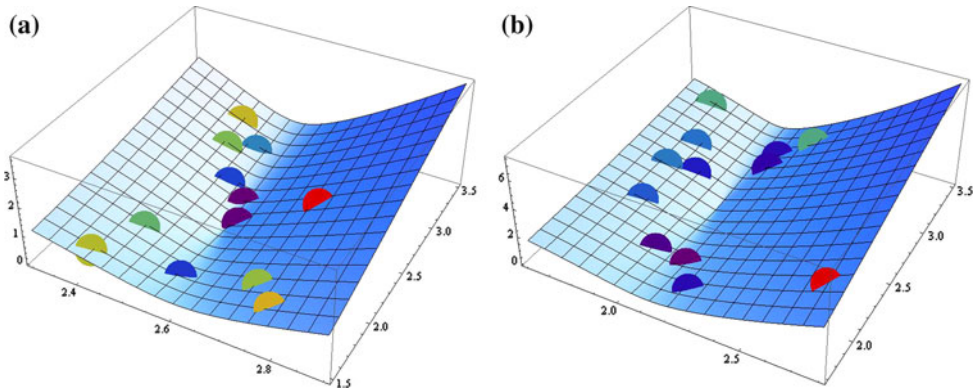

Fig. 26 The statistical manifolds formed by the monthly values of the satellite records (**a**) and WAM outputs (**b**) as elements of the non-Euclidean space of all Weibull distributions. A classical "BlueGreenYellow" color palette has been used depending on their approximate divergence from annual averages. (Color figure online)

Acknowledgments This work was partially supported by the MARINA project (7th Framework Programme, Grant agreement number: 241402, http://www.marina-platform.info/), and the E-wave project (funded by the Research Promotion Foundation of Cyprus, http://www.oceanography.ucy.ac.cy/ewave/). The anonymous reviewers are also acknowledged for their constructive suggestions that essentially contributed to the final form of this work.

References

Abdalla S, Bidlot J, Janssen P (2005) Assimilation of ERS and ENVISAT wave data at ECMWF. In: ENVISAT & ERS symposium, Salzburg, 6–10 Sep 2004 (ESA SP-572, Apr 2005)

Amari S-I (1985) Differential geometrical methods in statistics. Springer lecture notes in statistics, vol 28. Springer-Verlag, Berlin

Amari S-I, Nagaoka H (2000) Methods of information geometry. American Mathematical Society, Oxford University Press, Oxford

Arwini K, Dodson CTJ (2007) Alpha-geometry of the Weibull manifold. In: Second basic science conference, Tripoli

Arwini K, Dodson CTJ (2008) Information geometry: near randomness and near independence. Lecture notes in mathematics, vol 1953. Springer-Verlag, Berlin

Bidlot J, Janssen P (2003) Unresolved bathymetry, neutral winds and new stress tables in WAM. ECMWF Research Department Memo R60.9/JB/0400

Breivik LA, Reistad M (1994) Assimilation of ERS-1 altimeter wave heights in an operational numerical wave model. Weather Forecast 9(3):440–450

Cai Y, Dodson CTJ, Doig A, Wolkenhauer O (2002) Information-theoretic analysis of protein sequences shows that amino acids self-cluster. J Theor Biol 218(4):409–418

Chu PC, Cheng KF (2007) Effect of wave boundary layer on the sea-to-air dimethylsulfide transfer velocity during typhoon passage. J Mar Syst 66:122–129

Chu PC, Cheng KF (2008) South China Sea wave characteristics during Typhoon Muifa passage in winter 2004. J Oceanogr 64:1–21

Chu PC, Qi Y, Chen YC, Shi P, Mao QW (2004) South China Sea wave characteristics. Part-1: validation of wavewatch-III using TOPEX/Poseidon data. J Atmos Ocean Technol 21(11): 1718–1733

Dodson CTJ, Poston T (1991) Tensor geometry graduate texts in mathematics, vol 120, 2nd edn. Springer-Verlag, Berlin

Emmanouil G, Galanis G, Kallos G, Breivik LA, Heilberg H, Reistad M (2007) Assimilation of radar altimeter data in numerical wave models: an impact study in two different wave climate regions. Ann Geophys 25(3):581–595

Ferreira JA, Soares CG (1999) Modelling distributions of significant wave height. Coast Eng 40:361–374

Ferreira JA, Soares CG (2000) Modelling the long-term distribution of significant wave height with the Beta and Gamma models. Ocean Eng 26:713–725

Galanis G, Anadranistakis M (2002) A one dimensional Kalman filter for the correction of near surface temperature forecasts. Meteorol Appl 9:437–441

Galanis G, Louka P, Katsafados P, Kallos G, Pytharoulis I (2006) Applications of Kalman filters based on non-linear functions to numerical weather predictions. Ann Geophys 24:2451–2460

Galanis G, Emmanouil G, Kallos G, Chu PC (2009) A new methodology for the extension of the impact in sea wave assimilation systems. Ocean Dyn 59(3):523–535

Gonzalez-Marco D, Bolanos-Sanchez R, Alsina JM, Sanchez-Arcilla A (2008) Implications of nearshore processes on the significant wave height probability distribution. J Hydraul Res 46(2, Suppl. SI):303–313

Greenslade D, Young I (2005) The impact of inhomogenous background errors on a global wave data assimilation system. J Atmos Ocean Sci 10(2):61–93

Iguzquiza E, Chica-Olmo M (2008) Geostatistical simulation when the number of experimental data is small: an alternative paradigm. Stoch Environ Res Risk Assess 22:325–337

Jansen PAEM (2000) ECMWF wave modeling and satellite altimeter wave data. In: Halpern D (ed) Satellites, oceanography and society. Elsevier, New York, pp 35–36

Janssen PAEM, Lionello P, Reistad M, Hollingsworth A (1987) A study of the feasibility of using sea and wind information from the ERS-1 satellite, part 2: use of scatterometer and altimeter data in wave modelling and assimilation. ECMWF report to ESA, Reading

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME D 82:35–45

Kalman RE, Bucy RS (1961) New results in linear filtering and prediction problems. Trans ASME D 83:95–108

Kalnay E (2002) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge

Komen G, Cavaleri L, Donelan M, Hasselmann K, Hasselmann S, Janssen PAEM (1994) Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge

Lionello P, Günther H, Janssen PAEM (1992) Assimilation of altimeter data in a global third generation wave model. J Geophys Res 97(C9):14453–14474

Lionello P, Günther H, Hansen B (1995) A sequential assimilation scheme applied to global wave analysis and prediction. J Mar Syst 6:87–107

Loffredo L, Monbaliu J, Bitner-Gregersen E, Toffoli A (2009) The role of spectral multimodality in wave climate design. In: Wave Hindcasting Workshop, Halifax

Makarynskyy O (2004) Improving wave predictions with artificial neural networks. Ocean Eng 31(5–6):709–724

Makarynskyy O (2005) Neural pattern recognition and prediction for wind wave data assimilation. Pac Oceanogr 3(2):76–85

Muraleedharan G, Rao AD, Kurup PG, Unnikrishnan N, Mourani S (2007) Modified Weibull distribution for maximum and significant wave height simulation and prediction. Coast Eng 54: 630–638

Nordenstrøm N (1973) A method to predict long-term distributions of waves and wave-induced motions and loads on ships and other floating structures. Der Norske Veritas, Publication No. 81

Prevosto M, Krogstad HE, Robin A (2000) Probability distributions for maximum wave and crest heights. Coast Eng 40:329–360

Quentin C (2002) Etude de la surface océanique, de sa signature radar et de ses interactions avec le flux turbulent de quantite de mouvement dans le cadre de l'expérience FETCH (in French). PhD thesis, Universite de Paris

Rao ST, Zurbenko IG, Neagu R, Porter PS, Ku JY, Henry RF (1997) Space and time scales in ambient ozone data. Bull Am Meteor Soc 78(10):2153–2166

Resconi G (2009) Geometry of risk analysis (morphogenetic system). Stoch Environ Res Risk Assess 23:425–432

Rosmorduc V, Benveniste J, Lauret O, Maheu C, Milagro M, Picot N (2009) In: Radar altimetry tutorial. Benveniste J, Picot N (eds) http://www.altimetry.info

Spivak M (1965) Calculus on manifolds. W.A. Benjamin, New York

Spivak M (1979) A Comprehensive introduction to differential geometry, vol 1–5, 2nd edn. Publish or Perish, Wilmington

Thornton EB, Guza RT (1983) Transformation of wave height distribution. J Geophys Res 88(C10):5925–5938

Vanem E (2011) Long-term time-dependent stochastic modelling of extreme waves. Stoch Environ Res Risk Assess 25:185–209

Vanem E, Huseby A, Natvig B (2011) A Bayesian hierarchical spatio-temporal model for significant wave height in the North Atlantic. Stoch Environ Res Risk Assess. doi:10.1007/s00477-011-0522-4

WAMDIG, The WAM-Development and Implementation Group: Hasselmann S, Hasselmann K, Bauer E, Bertotti L, Cardone CV, Ewing JA, Greenwood JA, Guillaume A, Janssen PAEM, Komen GJ, Lionello P, Reistad M, Zambresky L (1988) The WAM model—a third generation ocean wave prediction model. J Phys Oceanogr 18(12):1775–1810

WISE Group (2007) Wave modelling—the state of the art. Prog Oceanogr 75:603–674