Available online at www.sciencedirect.com
ScienceDirect
Continental Shelf Research 28 (2008) 391–398
CONTINENTAL SHELF RESEARCH
www.elsevier.com/locate/csr

A method for spatial calibration of wave hindcast data bases

Antonio Tomas[a,b], Fernando J. Mendez[a], Inigo J. Losada[a,*]

[a]*Ocean & Coastal Research Group, Instituto de Hidráulica Ambiental "IH Cantabria", Universidad de Cantabria, Spain*
[b]*Organismo Público Puertos del Estado, Ministerio de Fomento, Spain*

Received 21 February 2007; received in revised form 12 August 2007; accepted 19 September 2007
Available online 5 October 2007

Abstract

In the last years, new tools have been developed to describe wave climate. In particular, wave hindcast models allow a very detailed description in time and space. However, the data bases generated from these models require calibration with instrumental observations. In this work, a methodology for spatial calibration of wave hindcast data bases is developed. It is based on an Empirical Orthogonal Function decomposition and a non-linear transformation of the spatial–time modes. The method is applied to monthly long-term distribution of significant wave height in the Western Mediterranean. The calibration (transformation of the modes) is carried out using existing buoys in the area. After calibration, the validation with satellite data shows that this methodology is useful to better define wave climate from hindcast data bases.

Keywords: Calibration; Empirical Orthogonal Function; Hindcast; Monthly long-term distribution; Significant wave height; Spatial correlation; Wave climate; Wind wave model

1. Introduction

Recently, in situations where a long-term buoy record is not available, new wind wave data bases generated numerically appear to be a good source to obtain the wave climate at a particular site. However, from a quantitative point of view, some differences are found when comparing them with instrumental devices (see f.i. Caires and Sterl, 2005; Cavaleri and Sclavo, 2006). Moreover, near the coast and when the orography is complex, the inaccuracy of the results is evident due to a bad description of the wind field (Cavaleri and Bertotti, 2004). These wind wave numerical data bases usually express their results in terms of the significant wave height (H_s), the mean period and the mean direction in every hourly (or 3-hourly) sea state.

Conscious of the limitations of the quantitative validity of these hindcast data bases, several algorithms have been developed to correct the values of H_s, which is one of the key parameters to characterize wave climate. In the state-of-the-art, we can find parametric corrections (Cavaleri and Sclavo, 2006) as well as non-parametric approaches (Caires and Sterl, 2005). One of the disadvantages of the aforementioned techniques is that the calibration is applied on a point-to-point basis, disregarding the spatial correlation between adjacent nodes. In some particular locations, usually close to some archipelago, spatial wave climate is highly variable and the calibration using buoys or satellite data is not possible. This occurs, for instance in the Balearic Islands (Fig. 1), where the nodes from hindcast of the southwest face of Mallorca island or the eastern coast of Ibiza cannot be calibrated using point-to-point methods. In those cases, a spatial spreading of the calibration information is necessary and it might be good to consider the spatial correlation between neighbour nodes.

Many times, the purpose of the data base is to spatially characterize wave climate in terms of the (annual or monthly) long-term distribution of H_s. One possibility is to spatially analyse several H_s statistical parameters such as the annual mean, the 90th, 95th or 99th percentiles. Another option is to assume a given distribution (f.i., Weibull for minima, Log-Normal, Gamma) estimating the

*Corresponding author. Ocean & Coastal Research Group, Instituto de Hidráulica Ambiental "IH Cantabria", Universidad de Cantabria, Avda. de los Castros s/n, 3900 Santander, Spain. Tel.: +34 942201810; fax: +34 942201860.
E-mail address: losadai@unican.es (I.J. Losada).

0278-4343/$ - see front matter
doi:10.1016/j.csr.2007.09.009

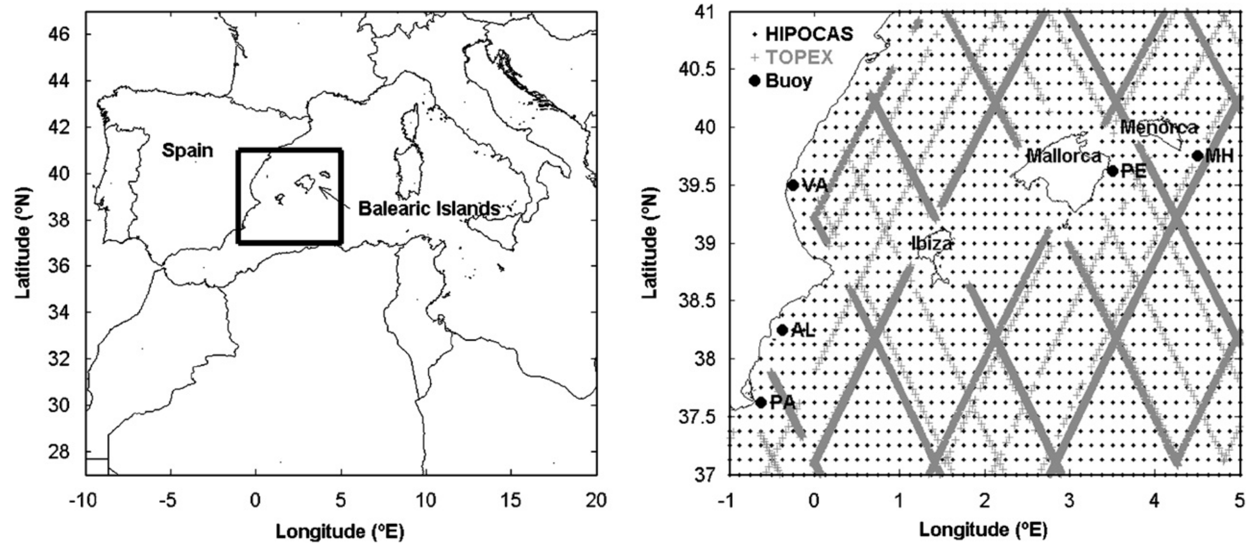

Fig. 1. Location of the study area and the data bases used in the analysis.

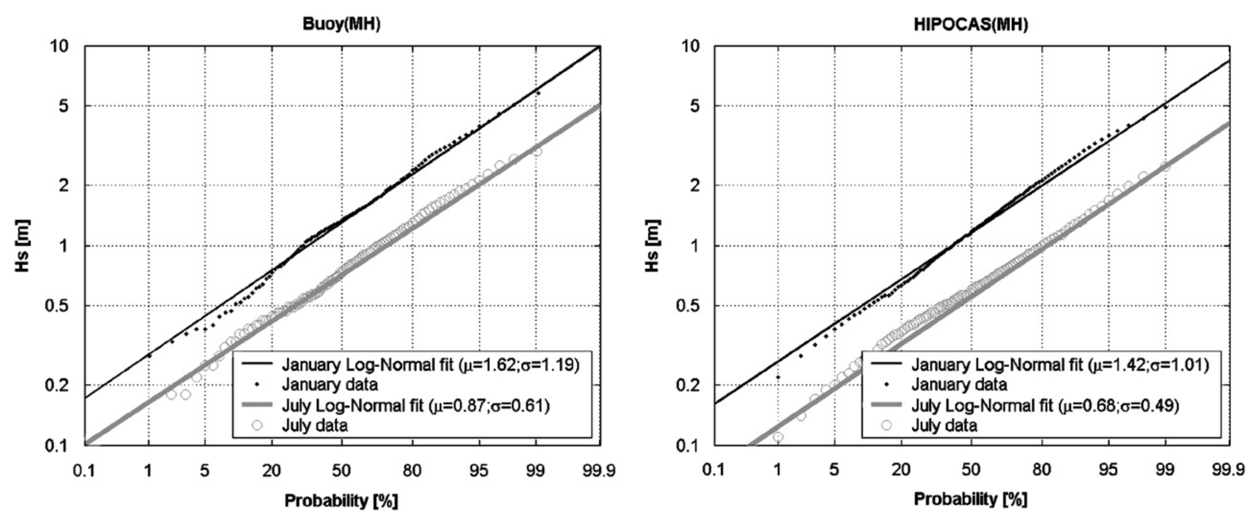

Fig. 2. Example of the Log-Normal distribution fit in Mahon position (buoy and hindcast data) for January and July.

parameters of the distribution for every specific point. The latter procedure is more flexible, since it allows the definition of whatever percentile you are interested in. In this work, we have assumed the monthly long-term distribution of H_s to follow a Log-Normal distribution (Massel, 1996; Holthuijsen, 2007). Several goodness-of-fit tests were applied to monthly data confirming that our hypothesis was adequate. The probability density function for the Log-Normal distribution can be expressed as

$$f(H) = \frac{1}{H\sqrt{2\pi}\sigma^*}\exp\left[-\frac{1}{2}\left(\frac{\log(H)-\mu^*}{\sigma^*}\right)^2\right] \quad (1)$$

where $0<H<\infty$, μ^* and σ^* are the location and scale parameters, respectively. It is usual to work with the mean, $\mu = \exp(\mu^* + \sigma^{*2}/2)$, and the standard deviation, $\sigma^2 = \exp(2\mu^* + 2\sigma^{*2}) - \exp(2\mu^* + \sigma^{*2})$, of this random variable. Therefore, the spatial monthly long-term distribution can be characterized using $\mu(x,\tau)$ and $\sigma(x,\tau)$, where x stands for every particular location and $\tau = \{1, \ldots, 12\}$ for every month (for instance, for January, $\tau = 1$). In Fig. 2, the Log-Normal distribution fit to the buoy data (Mahon buoy, see details in Fig. 1 and Table 1) and the hindcast data in this position for January ($\tau = 1$) and July ($\tau = 7$) are shown.

In this work, we use a hindcast data base generated by Puertos del Estado with the WAM model (Hasselman et al., 1998) in the Western Mediterranean (Ratsimandresy et al., 2007) in the frame of the HIPOCAS European project (Guedes Soares et al., 2002). It consists of a 44-year (1958–2001) hourly time series of H_s, mean period and mean direction. Spatial resolution is 0.125°. The area of analysis (see Fig. 1) covers part of the Spanish Mediterranean including the Balearic Sea and the coast of Valencia. For this particular area, there are several buoys from the Puertos del Estado network: four intermediate water depth buoys (Valencia, Alicante, Cabo de Palos, Cap de Pera)

Table 1
Description of the buoys used in the analysis

	Period of measurement	Longitude	Latitude	Water depth (m)
Valencia (VA)	1985–2001	00°17.0′W	39°28.0′N	20
Alicante (AL)	1985–2001	00°25.0′W	38°15.0′N	50
Cabo de Palos (PA)	1985–2001	00°38.3′W	37°39.2′N	67
Cap de Pera (PE)	1989–2001	03°29.1′E	39°39.0′N	48
Mahón (MH)	1993–2001	04°25.9′E	39°43.7′N	300

and one deep water buoy in Mahon (see details in Fig. 1 and Table 1). As this wave hindcast provides information in deep water (it was carried out switching off the refraction process), we have developed a procedure to back-propagate the buoy time series to deep water, thus obtaining five deep water buoys in the domain area. Fig. 1 also shows another complementary data base from TOPEX/POSEIDON satellite data (period 1995–2001) which will be later used in the validation procedure.

Comparing the mean, $\mu(x,\tau)$, and the standard deviation, $\sigma(x,\tau)$, of the monthly long-term distribution of the hindcast with the buoys (see Fig. 5: solid lines represent data from hindcast and dots from the buoys), one can see important discrepancies in the magnitude; the differences depending on the specific location and also on the variability within the year.

Therefore, the objective of this work is to spatially calibrate the monthly long-term distribution of significant wave height obtained from hindcast data. To achieve this goal, we spatially and temporally transform the wave hindcast trying to improve the fit at the buoys locations. An important condition of this "transformation" is that it must be able of extrapolating the information in space, not loosing the correlation between adjacent nodes. We have applied the methodology to the window selected in Fig. 1. Buoy data from the Spanish network are used for the calibration while satellite data from TOPEX/POSEIDON mission are used for the validation procedure.

2. Methodology

2.1. Summary of the approach

Once the objective has been presented in the previous section, the question to address is how to spatially and temporally transform the fields $\mu(x,\tau)$ and $\sigma(x,\tau)$ in order to fit in the positions of the buoys while still preserving the spatial correlation between adjacent nodes. Fig. 3 graphically shows the aim of the calibration, which consists of a transformation of a spatial–temporal field to fit into the locations of the instrumental data. One of the possibilities is to split up every spatial–temporal field into a number of modes using Empirical Orthogonal Function (EOF) decomposition (Section 2.2). The spatial modes and their temporal amplitudes (which explain a given percentage of the variance, say, 99%) are afterwards transformed using a non-linear parameterization (Section 2.3). The parameters are obtained by minimizing an objective function trying to reduce the errors at the locations of the buoys (Section 2.4). A detailed explanation of the steps carried out in the calibration applied to the study area follows.

2.2. Empirical Orthogonal Function decomposition

The EOFs technique is used to split up the spatial monthly long-term distribution of H_s in the study area, thus decomposing it into a series of q orthogonal functions (Baldacci et al., 2001). For example, for a generic variable $z(x,\tau)$ (which could be $\mu(x,\tau)$ or $\sigma(x,\tau)$), the following decomposition is obtained:

$$z(x,\tau) = z_M(x) + f_1(x)g_1(\tau) + f_2(x)g_2(\tau) + \cdots + f_q(x)g_q(\tau) \quad (2)$$

where $z_M(x)$ is the time-averaged spatial mode, f_i is the ith spatial mode and g_i is the ith time amplitude. For the study area, the total number of modes is $q = 1342$, which is the number of nodes considered in the domain (see dots in Fig. 1).

Fig. 4 shows, for the parameter $\mu(x,\tau)$, the time-averaged spatial mode, $\mu_M(x)$, and the first three spatial modes ($f_1(x)$, $f_2(x)$, $f_3(x)$) with their respective time amplitudes ($g_1(\tau)$, $g_2(\tau)$, $g_3(\tau)$). One can see how $\mu_M(x)$ is higher in open waters, indicating that the intensity of the wave climate is stronger in offshore areas. Although the decomposition in EOFs is a statistical technique without physical meaning, the first modes usually explain some kind of climatic behaviour. For instance, the modulation of the first time amplitude $g_1(\tau)$ indicates seasonality (waves are higher in winter than in summer). The second mode can also help to detect intense sea states in the North of the Balearic Islands in October and November and near Cabo de Palos in May and June.

These patterns can also be observed in the EOF decompositions for $\sigma(x,\tau)$ (not shown) where $\sigma_M(x)$ is more variable in open waters, which also represents the winter–summer fluctuations in the wave climate variability.

For the initial q modes, we selected the first p modes that explain at least 99% of the variability of the variable $z(x,\tau) - z_M(x)$. For our study area, we obtained $p = 4$ modes (for both $\mu(x,\tau)$ and $\sigma(x,\tau)$). Consequently, $z'(x,\tau)$

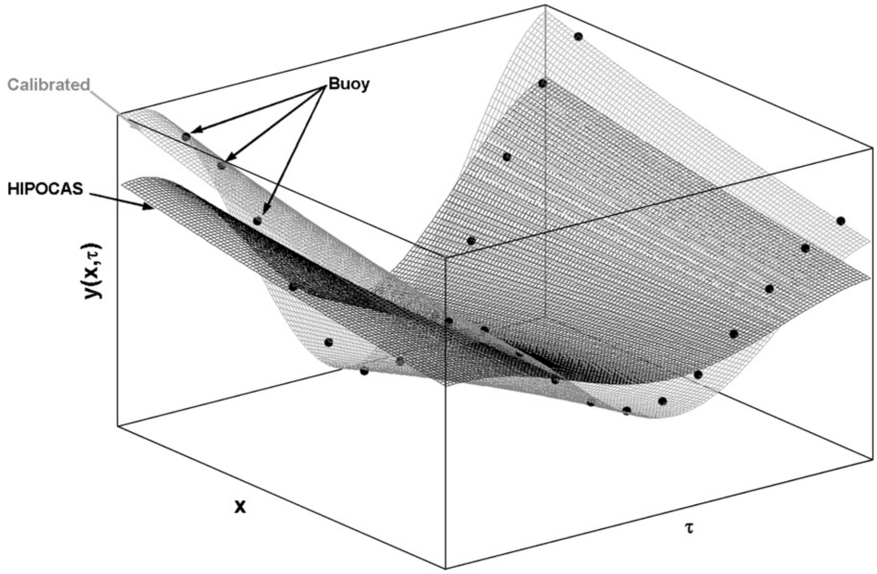

Fig. 3. Example of the transformation of a spatial–temporal field to fit in the locations of instrumental data available (dots).

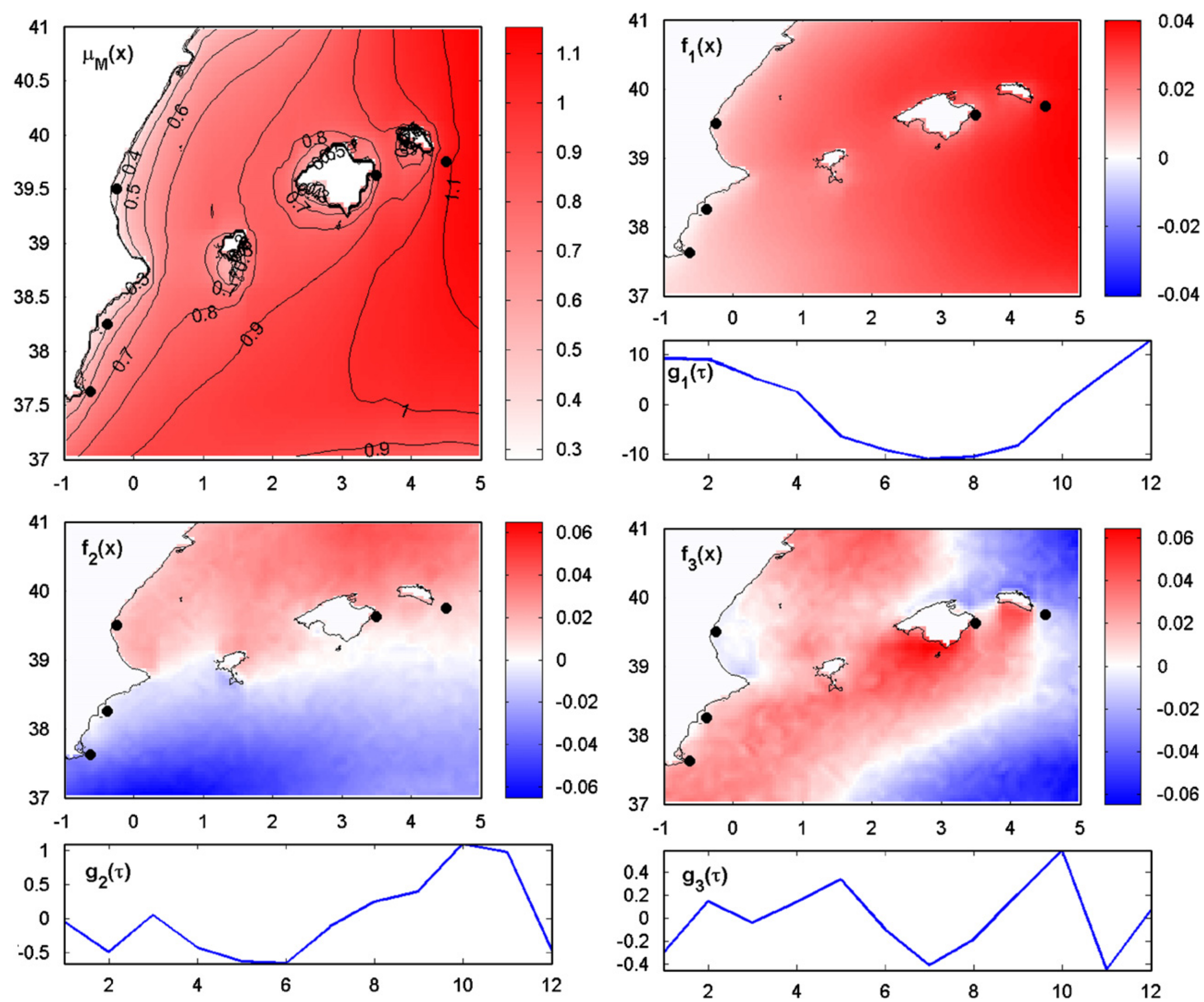

Fig. 4. Example of the time-averaged spatial mode for $\mu(x,\tau)$ and the first three spatial modes along with their respective time amplitudes.

is considered to be a close approximation of $z(x,\tau)$:

$$z'(x,\tau) = z_M(x) + f_1(x)g_1(\tau) + f_2(x)g_2(\tau) + f_3(x)g_3(\tau) + f_4(x)g_4(\tau) \quad (3)$$

Thus, the monthly long-term distributions are now, $\mu'(x,\tau)$ and $\sigma'(x,\tau)$, split up in one time-average mode and four spatial–temporal modes that explain, at least, 99% of the variance. Note that the purpose of maintaining the spatial correlation is achieved using this EOF decomposition.

2.3. Time–space transformation. Vector parameter

We propose a possible transformed (calibrated) spatial and temporal field $z_C(x,\tau)$ using a combination of linear and potential parameterizations as

$$z_C(x,\tau) = b_0 z_M(x)^{a_0} + b_1 f_1(x) g_1(\tau)\left|g_1(\tau)\right|^{a_1-1} + b_2 f_2(x)g_2(\tau) + b_3 f_3(x) g_3(\tau) + b_4 f_4(x) g_4(\tau) \quad (4)$$

where | | means absolute value. The parameters b_0, b_1, b_2, b_3 and b_4 are the linear coefficients that increase (if they are larger than 1) or decrease (if they are smaller than 1) each of the terms of $z_C(x,\tau)$. To spatially transform $z_C(x,\tau)$, the non-linear coefficient a_0 is introduced into the time-averaged spatial field $z_M(x)$. For the transformation in time, a non-linear coefficient a_1 is introduced into the time amplitude term of the first mode $g_1(\tau)$ (the first mode explains more than 95% of data variability). We tried different parameterizations not obtaining a significance improvement in the fitting. The set of parameters can be packed into a vector parameter $\theta = \{a_0, a_1, b_0, b_1, b_2, b_3, b_4\}$. All the coefficients were negative and positive bounded (between 0.3 and 2.5) to avoid a too large transformation.

2.4. Minimization of error

The objective is to determine the vector parameter estimates of $\hat{\theta}$ that minimizes the error between the calibrated field $z_C(x,\tau;\hat{\theta})$ and the correct values at the positions of the buoys, $z_B(x_i,\tau)$, where x_i refers to the position of every buoy. This results in an optimization problem in which the objective function to be minimized $J(\theta)$ must be defined. This function can be defined as

$$J(\theta) = \sum_{i=1}^{n_b}\sum_{j=1}^{12}\left[\frac{z_B(x_i,\tau_j) - z_C(x_i,\tau_j;\theta)}{\bar{z}_B(x_i)}\right]^2 w(x_i,\tau_j) \quad (5)$$

where n_b is the number of buoys, $\bar{z}_B(x_i)$ is the time-averaged value, $\bar{z}_B(x_i) = \sum_{j=1}^{12} z_B(x_i,\tau_j)/12$, and $w(x_i,\tau_j)$ is a weighting term that assigns the importance of every buoy and every month in the objective function. In this work, we have assumed a constant value $w(x_i,\tau_j) = 1$.

The minimization of Eq. (5) is carried out using a widely used optimization tool for highly dimensional non-linear problems, the shuffled complex evolution method SCE-UA (Duan et al., 1992), obtaining the parameter estimates $\hat{\theta}$. Table 2 shows the set of parameters obtained for the

Table 2
Set of parameter estimates for $\mu_C(x,\tau)$ and $\sigma_C(x,\tau)$

	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$
$\mu_C(x,\tau)$	0.49	1.47	1.25	0.41	2.40	2.49	1.15
$\sigma_C(x,\tau)$	0.72	1.03	1.09	1.13	0.68	0.35	1.01

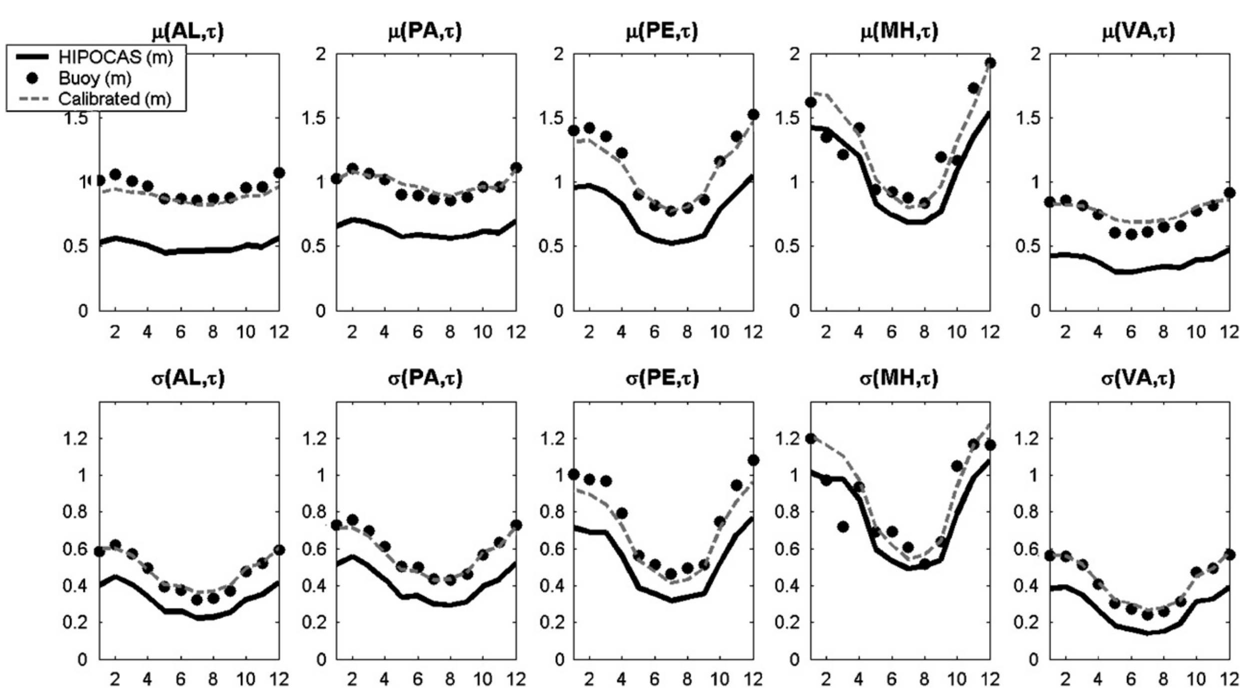

Fig. 5. Time evolution of μ and σ in the buoys located in the calibrated area.

calibration of $\mu_C(x,\tau)$ and $\sigma_C(x,\tau)$ for our particular study area.

3. Results

Fig. 5 represents the evolution over time (12 months) of the two parameters (μ and σ) for the five buoys located in the area. It is shown that in all the cases, the parameters calibrated for the numerical data base reproduce almost perfectly those determined by the obtained buoy data. One can observe how the goodness-of-fit is similar in all the buoys (not depending on the degree of activity of the wave climate on each buoy). The annual winter–summer fluctuation in the Log-Normal parameters of the distribution can also be noted, or what is the same, the evolution over time of the monthly long-term distribution (Ochi, 1998). This same fluctuation is always present in the time amplitude of the first mode ($g_1(\tau)$) of the EOF decomposition of both parameters.

Using the spatial calibration, the monthly long-term distribution can be defined for the entire study area. Fig. 6 shows the spatial distribution of the parameters $\mu_C(x,\tau)$ and $\sigma_C(x,\tau)$ in the area, not only for the mean time of the monthly long-term distribution but also for the monthly long-term distribution for January (month 1) and July (month 7). Note how the waves in January (winter) are larger and more variable than in July (summer). It is also remarkable how the transformation works are slightly modifying the shape of the wave fields.

4. Validation

The last part of the methodology aims to validate with satellite data from altimeters the results obtained in the

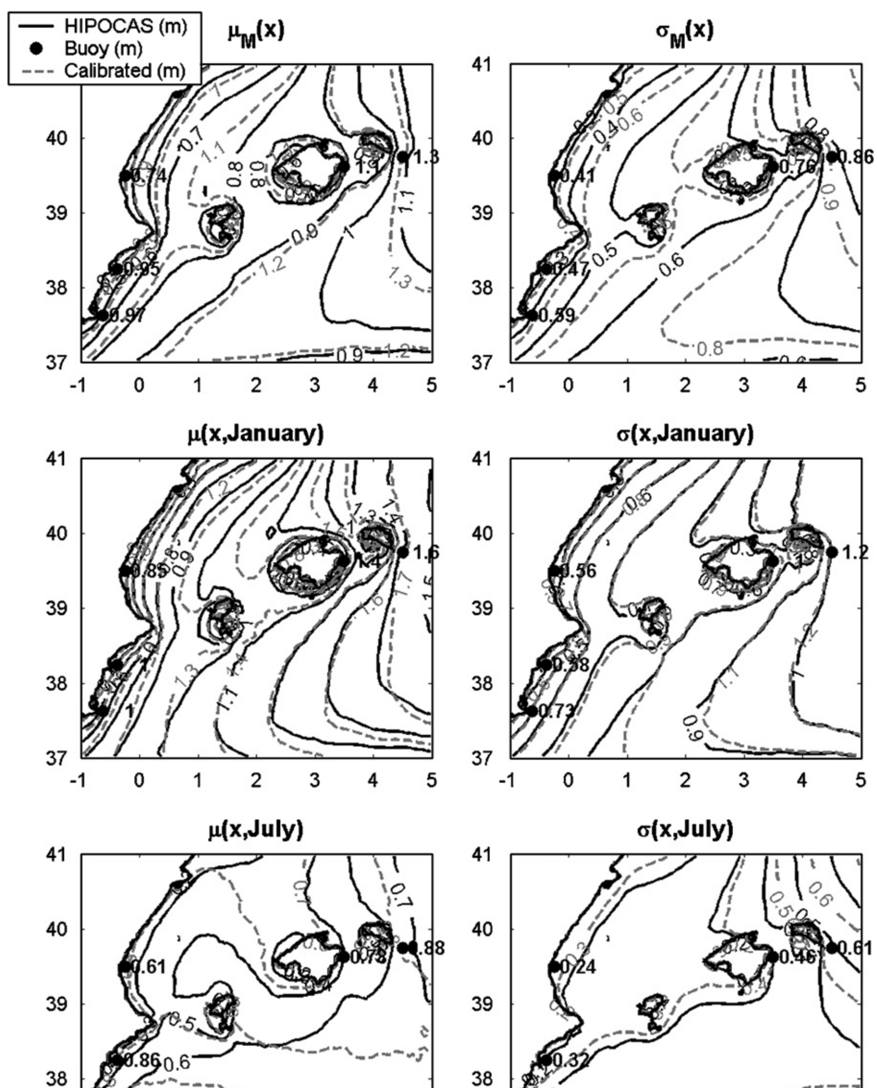

Fig. 6. Spatial evolution of $\mu(x,\tau)$ and $\sigma(x,\tau)$ in the calibrated area, for its time-averaged fields and for the months of January and July.

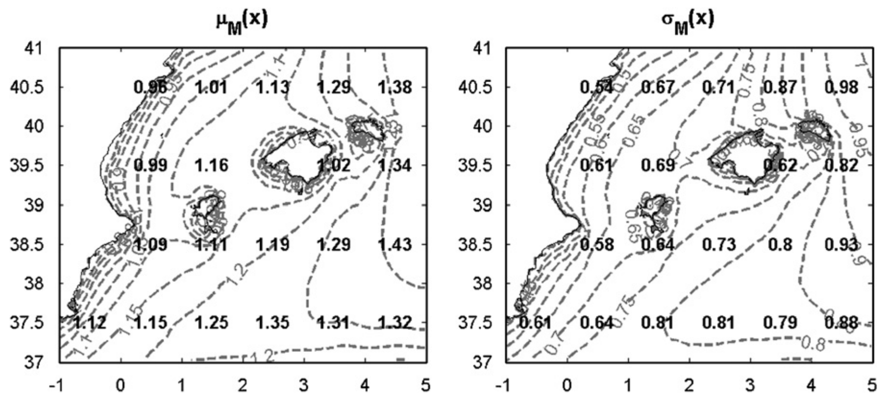

Fig. 7. Comparison between the calibrated wave fields $\mu_C(x,\tau)$ and $\sigma_C(x,\tau)$ and the $1^\circ \times 1^\circ$ spatial-averaged values from TOPEX satellite.

calibration using five buoys. Measurements of H_s in the study area between May 1995 and December 2001 from the TOPEX/POSEIDON mission are used. This data base was obtained from the *Physical Oceanography Distributed Active Archive Center* (PO.DAAC) of the *NASA Jet Propulsion Laboratory* (http://podaac.jpl.nasa.gov/). Following Krogstad and Barstow (1999) we have corrected this data base using the expression $H_s^* = 1.1H_s - 0.165$ (m).

The satellite obtained H_s data base does not suffice to define the long-term distribution in every particular point, due to which the data is aggregated in a $1^\circ \times 1^\circ$ square grid. Fig. 7 represents the annual values of μ and σ from the TOPEX satellite for every $1^\circ \times 1^\circ$ square. It can be seen that the calibrated wave fields reproduce, very adequately, the values obtained from satellite data, confirming the ability of the methodology. The root-mean-square relative error of the hindcast (compared with the satellite data) has been reduced from 28% (before calibration) to 5% (after calibration) for the mean and from 20% to 8% for the standard deviation.

5. Conclusions

A spatial calibration methodology for wave hindcast has been developed. To do this, the space–time fields of the monthly long-term distribution of H_s are decomposed into modes, utilizing Empirical Orthogonal Functions, and transforming them to obtain the best match possible with buoy observations. The difference between the observed and transformed long-term distribution is minimized with the SCE-UA method.

The calibration method has been applied to the significant wave height of the HIPOCAS data base for an area in the Mediterranean Sea which covers Valencia and the Balearic Islands, using the buoy registered data base to carry out the calibration. Validation of the calibration has been based on satellite-registered data from TOPEX/POSEIDON, obtaining very satisfactory results. This methodology can be extended to any other data base of geophysical variables, for instance wind fields and storm surge fields, obtained from hindcast models. This is considered to be an important tool to be able to use these data bases from a quantitative point of view.

Acknowledgements

The authors would like to thank Puertos del Estado for providing the information, both from the HIPOCAS data base as well as that of the instrumental observation networks. F.J.M. and I.J.L. thank the Ministerio de Educación y Ciencia for its financial support within the "Ramón y Cajal" program as well as that of the GALERNA (ENE2004-08172) and TECLIMAR (CGL2005-05365/CLI) Projects.

References

Baldacci, A., Cosini, G., Grasso, R., Manzella, G., Allen, J.T., Cipollini, P., Guymer, T.H., Snaith, H.M., 2001. A study of the Alboran sea mesoscale system by means of empirical orthogonal function decomposition of satellite data. Journal of Marine Systems 29, 293–311.

Caires, S., Sterl, A., 2005. A new non-parametric method to correct model data: application to significant wave height from the ERA-40 hindcast. Journal of Atmospheric and Oceanic Technology 22, 443–459.

Cavaleri, L., Bertotti, L., 2004. Accuracy of the modelled wind and wave fields in enclosed seas. Tellus 56A, 167–175.

Cavaleri, L., Sclavo, M., 2006. The calibration of wind and wave model data in the Mediterranean Sea. Coastal Engineering 53, 613–627.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28 (4), 1015–1031.

Guedes Soares, C., Carretero, J.C., Weisse, R., Alvarez, E., 2002. A 40 years hindcast of wind, sea level and waves in European waters. In: Proceedings of OMAE 2002: 21st International Conference on Offshore Mechanics and Arctic Engineering. OMAE 2002-28604.

Hasselman, S., Hasselman, K., Janssen, P.A.E.M., Komen, G.J., Bertotti, L., Lionello, P., Guillaume, A., Cardone, V.C., Greenwood, J.A., Reistad, M., Zambresky, L., Ewing, J.A., 1998. The WAM model—a third generation ocean wave prediction model. Journal of Physical Oceanogaphy 18, 1775–1810.

Holthuijsen, L.H., 2007. Waves in Oceanic and Coastal Waters. Cambridge University Press.

Krogstad, H.E., Barstow, S.F., 1999. Satellite wave measurements for coastal engineering applications. Coastal Engineering 37, 283–307.

Massel, S.R., 1996. Ocean Surface Waves: Their Physics and Prediction. World Scientific.
Ochi, M.K., 1998. Ocean Waves: The Stochastic Approach. Cambridge University Press.
Ratsimandresy, A.W., Sotillo, M.G., Carretero, J.C., Alvarez, E., Hajji, H., 2007. A 44-year high-resolution ocean and atmospheric hindcast for the Mediterranean basin developed within the HIPOCAS Proyect. Coastal Engineering, in press.