Coastal Engineering 72 (2013) 56–68

Contents lists available at SciVerse ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

High resolution downscaled ocean waves (DOW) reanalysis in coastal areas

Paula Camus, Fernando J. Mendez *, Raul Medina, Antonio Tomas, Cristina Izaguirre

Environmental Hydraulics Institute IH Cantabria, Universidad de Cantabria, Spain

ARTICLE INFO

Article history:
Received 30 January 2012
Received in revised form 14 September 2012
Accepted 17 September 2012
Available online 7 November 2012

Keywords:
Dynamical downscaling
Hybrid downscaling
Reanalysis database
Statistical downscaling
Wave climate
Wave propagation

ABSTRACT

Large-scale wave reanalysis databases (0.1°–1° spatial resolution) provide valuable information for wave climate research and ocean applications which require long-term time series (>20 years) of hourly sea state parameters. However, coastal studies need a more detailed spatial resolution (50–500 m) including wave transformation processes in shallow waters. This specific problem, called downscaling, is usually solved applying a dynamical approach by means of numerical wave propagation models requiring a high computational time effort. Besides, the use of atmospheric reanalysis and wave generation and propagation numerical models introduce some uncertainties and errors that must be dealt with. In this work, we present a global framework to downscale wave reanalysis to coastal areas, taking into account the correction of open sea significant wave height (directional calibration) and drastically reducing the CPU time effort (about 1000×) by using a hybrid methodology which combines numerical models (dynamical downscaling) and mathematical tools (statistical downscaling). The spatial wave variability along the boundaries of the propagation domain and the simultaneous wind fields are taking into account in the numerical propagations to performance similarly to the dynamical downscaling approach. The principal component analysis is applied to the model forcings to reduce the data dimension simplifying the selection of a subset of numerical simulations and the definition of the wave transfer function which incorporates the dependency of the wave spatial variability and the non-uniform wind forcings. The methodology has been tested in a case study on the northern coast of Spain and validated using shallow water buoys, confirming a good reproduction of the hourly time series structure and the different statistical parameters.

1. Introduction

Wave retrospective analysis or *reanalysis* databases have become a powerful source of information for wave climate research and ocean applications over the last decades. These databases have good spatial coverage and provide continuous time-series of offshore wave parameters, over significant periods of time (more than 40 year-long), allowing the description of wave climate in locations where instrumental data is unavailable. However, i) they are not quantitatively perfect, ii) waves are poorly described at shallow water areas because the spatial resolution is not sufficiently detailed and iii) wave transformations due to the interaction with the bathymetry are not usually modeled.

The first problem related to the inaccuracy of the reanalysis wave data is corrected by means of calibration methods using instrumental observations (Mínguez et al., 2011a). The two other ones require modeling of the transformation processes and the increase of the spatial resolution (Camus et al., 2011b). This process, known as downscaling, is extremely important for design purposes in coastal engineering or for the evaluation of coastal wave energy resources.

Three different downscaling methods have been proposed: i) a dynamical approach consisting of nesting a wave propagation model for coastal areas which simulates wave transformation processes (refraction, bottom friction, shoaling, diffraction, breaking) from deep water to shallow water (Rusu et al., 2008); ii) a statistical approach establishing an empirical relationship between open ocean significant wave heights and a nearshore significant wave height in shallow water (e.g. using artificial neural networks, Browne et al., 2007; Kalra et al., 2005); iii) a hybrid approach which combines dynamical downscaling (numerical models) and statistical downscaling (an interpolation scheme, neural networks) in order to reduce the computational effort.

Dynamical downscaling is the most accurate approach providing a long time series with high spatial and temporal resolutions which allow a better statistical characterization of wave climate and extreme wave analysis. Statistical relations can be an effective method for nearshore height estimation with a little computational effort and an easy implementation. In applications where many simplifications must be adopted in numerical calculations due to the open ocean forcing and the bathymetry information available, the statistical methods can be more accurate than the numerical models (Browne et al., 2007). However, the main drawback is the requirement of coastal wave records at the location of interest to define the statistical model. Regarding hybrid methodologies, the most common

* Corresponding author at: Environmental Hydraulics Institute, IH Cantabria, Universidad de Cantabria, Parque Científico y Tecnológico de Cantabria, C/ Isabel Torres, 15, 39011, Santander, Spain. Tel.: +34 942 201616; fax: +34 942 266361.
E-mail address: mendezf@unican.es (F.J. Mendez).

0378-3839/$ – see front matter
http://dx.doi.org/10.1016/j.coastaleng.2012.09.002

ones consist of developing a transfer function for the transformation of offshore wave conditions to nearshore locations through the numerical propagation of a number of sea state conditions which characterize deep water wave climate (dynamical downscaling), see for instance Groeneweg et al. (2007) and Stansby et al. (2007). The representative cases are defined by means of several combinations of offshore wave and/or wind conditions at a specific location, without considering the spatial variability of these forcings. In order to correctly define the transfer function, a large number of sea states need to be simulated numerically, especially if the number of offshore wave parameters increases and therefore, the number of parameter combinations (Chini et al., 2010). Breivik et al. (2009) defined a linear downscaling based on 1-year hourly dynamical simulations, nested to the outputs of a third-generation wave model and forced by high resolution winds. However, the coastal wave height is estimated by means of a simpler linear relation with the height at a coarse-resolution open-ocean reanalysis grid, including the wave direction dependency via the definition of four regression models corresponding to four different directional sectors.

In addition to these hybrid methods, more sophisticated methodologies have been developed to obtain high resolution nearshore wave statistics. Galiskova and Weisse (2006) proposed three different statistical models based on linear regression, canonical correlation analysis and analogs to define a relation between instantaneous medium-scale wave fields from a hindcast database and higher resolution wave data in shallow water obtained dynamically. The empirical relations established are used to reconstruct certain percentiles of the significant wave height. Another statistical–dynamical approach, developed by Herman et al. (2009), uses a combination of a numerical model, principal component analysis and a neural-network method. This methodology reconstructs the spatial wave fields in shallow water as a function of the wave conditions, wind conditions and the sea level at a certain location because the forcings are highly uniform in the study area. These two methodologies require propagating several years of dynamical downscaling to generate the statistical model and its validation (Galiskova and Weisse, 2006; Herman et al., 2009).

The hybrid methodology proposed in Camus et al. (2011b), hereafter CMM, consists of the selection of a small number of representative wave conditions at deep water using the Maximum Dissimilarity Algorithm (MDA, see the analysis of selection algorithms of multivariate sea states presented in Camus et al., 2011a), the propagation of the selected cases using any state-of-the-art wave propagation model and the reconstruction of the wave time series at shallow water by means of the interpolation algorithm based on the radial basis functions (RBFs). The computational time required is significantly less than the other hybrid methodologies proposed because MDA covers the whole diversity of the offshore conditions with a reduced number of cases. Moreover, the RBFs allow establishing the statistical relation as a function of more offshore parameters.

The aim of this paper is to develop a methodology to generate hourly coastal wave time series trying to emulate the characteristics of the coastal wave reanalysis databases obtained by means of dynamical downscaling but reducing the computational time. In the application that will be shown, the computational time effort is reduced to three orders of magnitude (1000×) compared to the classical non-stationary simulations of a coastal wave reanalysis. The wave climate in deep water is transferred nearshore following the basis of the hybrid CMM methodology. However, in this previous paper, the offshore wave and wind conditions were defined at one location in deep water, assuming uniform forcings. In the present work, the dynamical propagations are nested to the outputs of a global/regional wave model. Therefore, the spatial wave variability along the boundaries of the propagation domain is taken into account and also the simultaneous wind fields in order to consider local wave generation. The CMM methodology needs to be improved and extended to deal with higher dimensional data. Although the MDA and RBF techniques are capable of dealing with multivariate data, the data dimension is reduced applying the principal component analysis (PCA), eliminating the information redundancy and facilitating the selection and the interpolation processes.

The generation of a coastal wave reanalysis database (downscaled ocean waves, DOW) by means of the proposed methodology requires the use of the wind and offshore wave reanalysis databases as forcings in order to obtain high temporal coverage. In this work, the long-term global NCEP/NCAR surface winds (Kalnay et al., 1996) and wave reanalysis GOW (Reguero et al., 2011) are used. Wave reanalysis models are a simplification of reality and they are also forced by discrete fields consisting of surface winds at different times. The wave generation outputs are calibrated to correct the differences when comparing with instrumental data (Mínguez et al., 2011a). Therefore, a global framework is proposed, which includes the previous calibration of the wave reanalysis data in deep water, in order to present a methodology with a wider application.

The proposed global framework and the case study for the application of the methodology are presented in Section 2. The deep water wave reanalysis database is described in Section 3. The steps involved in the proposed methodology: calibration, selection, propagation, and time series reconstruction are described in Sections 4, 5, 6, and 7 respectively. The selection and reconstruction processes are described in more detail because most innovative adaptations of the methodology to generate coastal wave reanalysis database are implemented. The validation of the methodology is detailed in Section 8. Finally, some conclusions are given in Section 9.

2. Global framework

The development of the DOW database implies several steps, which are summarized in Fig. 1. The steps of the proposed global framework are: a) analysis of the reanalysis databases available in the study area b) calibration of the reanalysis databases in deep water with instrumental data; c) selection of a limited number of cases which are the most representative of wave and wind hourly conditions in deep water; d) propagation of the selected cases using a wave propagation model; e) reconstruction of the time series of sea state parameters at shallow water; f) validation of the coastal wave data with instrumental data; and g) characterization of wave climate by means of a statistical technique.

The proposed methodology is applied to the northern coast of Spain (Fig. 2). The GOW Iberia grid with a resolution of 0.5°×0.5°, the GOW Cantabrico grid with a resolution of 0.1°×0.1° and the wind NCEP/NCAR database with a spatial resolution of 1.9° are shown in Fig. 2. The instrumental data located in the study area are: Bilbao moored buoys located in deep water at a depth = 600 m and near the coast at a depth = 53 m (belonging to Puertos del Estado), Pasaia acoustic doppler current profiler at a depth = 25 m (belonging to EUSKALMET) and Virgen Mar (depth = 32 m) and Santoña (depth = 28 m) moored buoys (belonging to Vigia Network from the Government of Cantabria). Although there a lot of information available, these data are spatially scarce and discontinuous in time, as can be seen in the time series of the GOW reanalysis gridpoint (marked with a circle) and the time series of the Bilbao buoys (lower panel of Fig. 2) during the year 2006.

The spatial domain has to be defined before applying the proposed methodology. This domain is nested to the outputs of the wave generation model, the GOW reanalysis database in our case study. Stationary wave simulations are assumed in order to consider the subset propagations as independent, which is a requirement of the proposed methodology. Therefore, the domain has to be small enough so that the wave propagation across this area occurs at a faster rate than the change in offshore forcing at the domain boundary. These restrictions are obviously inaccurate for global or basin-scale models but are reasonable for a smaller domain (Rogers

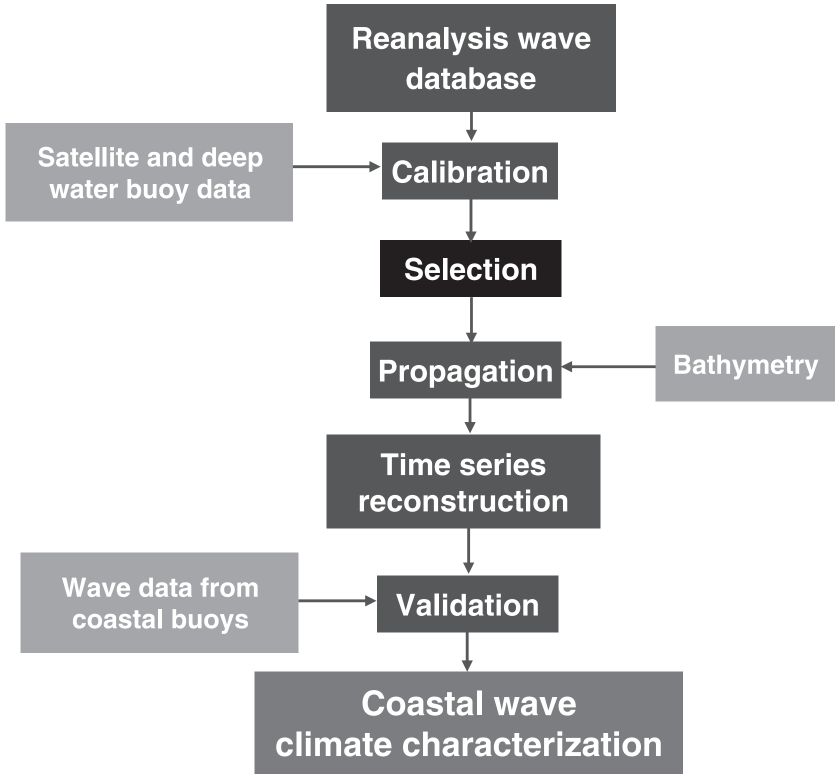

Fig. 1. Global framework to obtain coastal wave databases.

et al., 2007). Besides, if the stationary assumption is incorrect, a simple phase-shift in a time series will be noticed when comparing the instrumental data with numerical propagations.

The dimensions of the downscaling grid are 4.3°×0.8° with a resolution of 0.01° in longitude and 0.008° in latitude. The dimension of the computational grid along the main propagation direction (NW) is around 50 km, considered within the limits of the stationary simulation hypothesis. The bathymetry of the dynamical downscaling grids is defined by means of the global bathymetry "General Bathymetric Chart of the Oceans" (GEBCO), with a spatial resolution of 1′ from a combination of sounding waves and satellite data, available at the British Oceanographic Data Centre (BDOC), and the Spanish coastal charts, providing a detailed representation of the shallow water areas.

3. Deep water wave reanalysis databases

The global ocean wave (GOW) reanalysis is used to define the wave climate in deep water. This database is a large (from 1948 onwards) and up-to-date wave dataset with a global coverage and an hourly resolution (Reguero et al., 2011). The simulations at a global scale are computed on a grid with a spatial resolution of 1.5° in longitude and 1° in latitude using the model WAVEWATCH III (Tolman, 2002) forced with 6-hourly wind fields from the NCEP/NCAR reanalysis project (Kalnay et al., 1996). Bathymetry data used for the simulation comes from the ETOPO dataset (NOAA, 2006). A post-process using altimetry data has been applied consisting of a) the identification of possible outliers due to tropical cyclones (Mínguez et al., 2011b), not correctly simulated because of insufficient resolution in the wind forcing and b) a directional calibration procedure (Mínguez et al., 2011a) obtaining more accurate significant wave heights, especially remarkable for large values of wave heights. The GOW database presents a similar performance than other existing global analyses with the advantage of longer time records. Moreover, the global, wave simulations have been nested to a regional grid (the GOW Iberia grid with a resolution of 0.5°×0.5°) and to a local grid (the GOW Cantabrico grid with a resolution of 0.1°×0.1°), see Fig. 2. The output parameters are: the significant wave height (H_s), the peak period (T_p), the mean wave directions (θ_m) and the directional energy spectra in the boundaries of the DOW grid (see the stars in Fig. 4).

4. Calibration

Wave reanalysis databases allow a detailed description of the wave climate, since they provide long continuous time series records with a good spatial coverage. However, reanalysis models present inaccuracy mainly due to a bad description of wind fields, insufficient forcing and spatial and temporal model resolutions. A parametric calibration method depending on the mean wave direction is applied to correct significant wave heights with instrumental data from a satellite (Mínguez et al., 2011a). The model can be mathematically expressed as:

$$H_s^C = a^R(\theta)\left[H_s^R\right]^{b^R(\theta)} \tag{1}$$

where H_s^R is the reanalysis significant wave height, H_s^C is the calibrated or corrected significant wave height and $a^R(\theta)$ and $b^R(\theta)$ are the parameters dependent on the mean wave direction θ from reanalysis. For more details about the methodology and its hypothesis see Mínguez et al. (2011a).

This correction is applied to each boundary node in the DOW grid. For every location, the pairs of data for the calibration are obtained choosing all the satellite data in a radius of 1.5° (see example in Fig. 3). For this particular case, the $a^R(\theta)$ and $b^R(\theta)$ coefficients are displayed in the middle panel of Fig. 3. Finally, some QQ-plots, scatter-plots, cumulative distributions and roses of the instrumental data and the reanalysis data before and after the calibration are

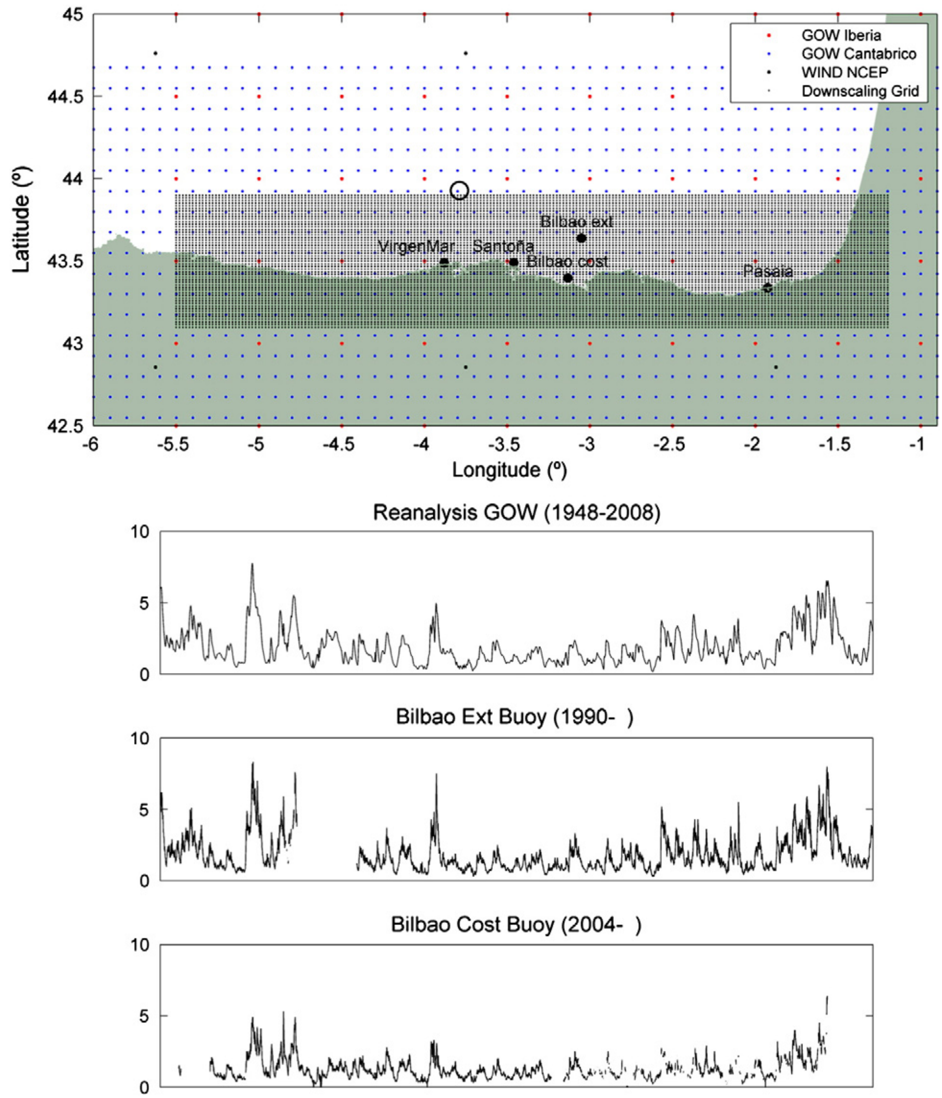

Fig. 2. Reanalysis databases and instrumental data available in the area of interest and the downscaling grid proposed to obtain coastal waves.

shown in Fig. 3. As seen, the method corrects the discrepancies between the reanalysis and the satellite data.

5. Selection

The aim of the selection process is to extract a subset of wave situations representative of available ocean conditions from the reanalysis database. The maximum-dissimilarity algorithm (MDA) has been proved to identify a subset of sea states comprising the most dissimilar wave conditions in a database (Camus et al., 2011a), even the extreme conditions, which is very suitable for the time series reconstruction using an interpolation technique. This algorithm was first described by Kennard and Stone (1969) and is widely applied in molecules in a high throughput screening in drug discovery (Snarey et al., 1997; Willet, 1996). The subset is initialized by transferring one vector from the data sample D_1. The rest of the $M-1$ elements are selected iteratively, calculating the dissimilarity between each remaining data in the database and the elements of the subset, and then transferring the most dissimilar one to the subset. The process finishes when the algorithm reaches M iterations. The algorithm is described in detail in Camus et al. (2011a), including the more efficient version of this algorithm (Polinsky et al., 1996).

This step of the methodology can be subdivided in several stages: a) set wind grid points and wave grid points which define forcings of the numerical propagations. Standardize these calibrated data after the wave and wind directions have been transformed to the x and y components. b) Apply the principal component analysis to the standardized forcings. Select the number of principal components (i.e. the variables in the new reduced space) that produces an acceptable root-mean-square error reconstruction. c) Select a representative number of offshore conditions using MDA in the reduced space and identify these select cases in the original space. An explanatory sketch of the stages of the selection step is shown in Fig. 4 and is next explained in detail.

The wave reanalysis nodes along the boundaries of the propagation domain and the simultaneous wind fields define the wave and wind forcing conditions of the wave simulations in shallow water, taking into account the wave spatial variability and the local wind wave generation. In the case study the GOW Cantabrico nodes, with a spatial resolution of 0.1°, are used to define the boundaries of the DOW grid (called G01). The NCEP/NCAR wind databases are used to define the wind fields, which are also the forcings used in the generation of the GOW reanalysis database. Fig. 5 shows the dynamical downscaling grid nested to the GOW database and the NCEP/NCAR wind nodes (defined by those closest to the study area). The hourly sea state parameters in deep water, which are going to be used in the selection process and in the time series

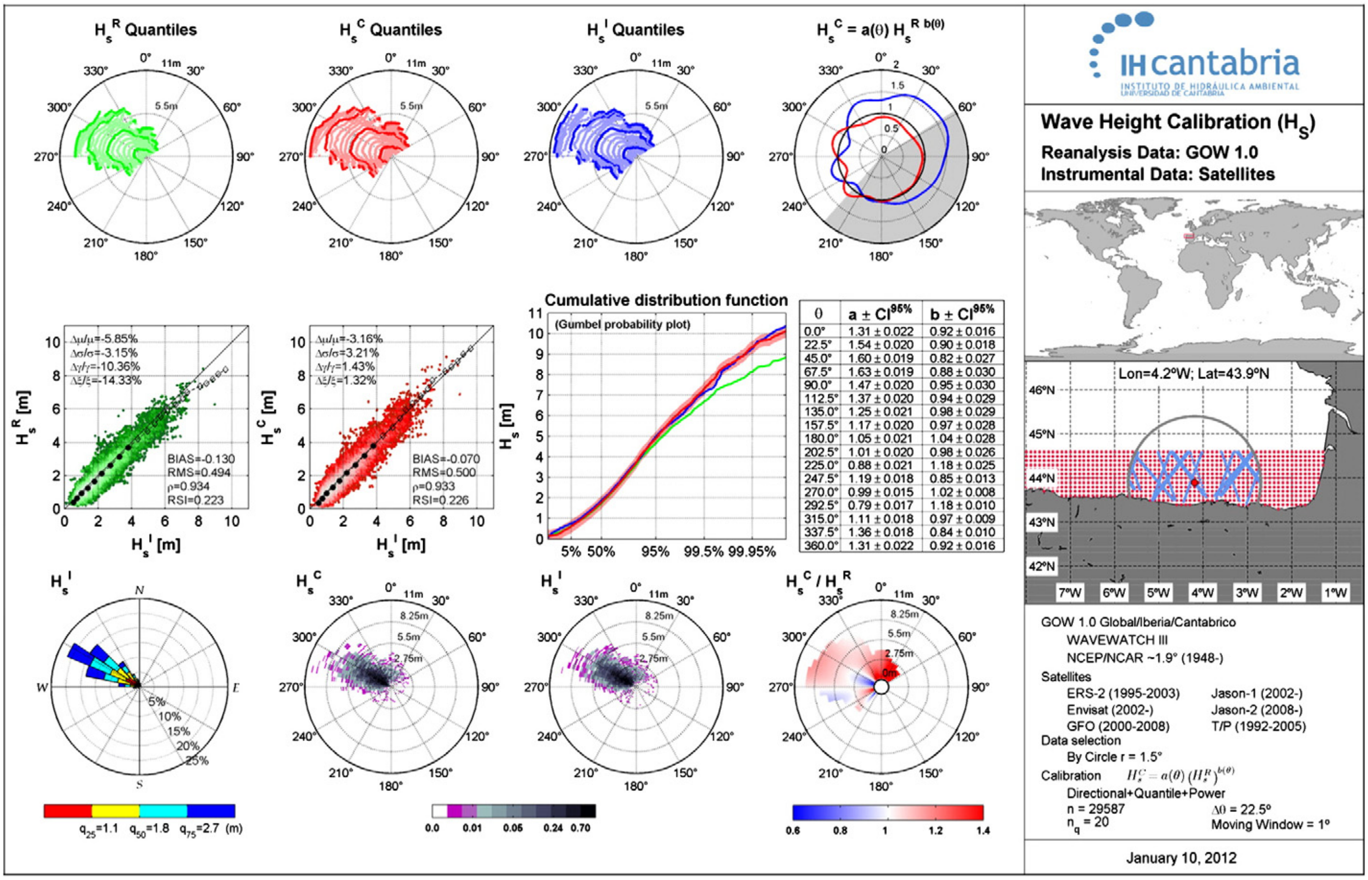

Fig. 3. Significant wave height GOW calibration with satellite data (Mínguez et al., 2011a). Instrumental data in blue, reanalysis data in green and calibrated data in red.

reconstruction process, are: the significant wave height (H_s), the mean wave period (T_m) and the mean wave direction (θ_m) of every five nodes (0.5°) at the computational boundaries and the time series of the wind parameters (W_{10x}, W_{10y}) of the nodes at the upper boundary of the wind grid. These wave and wind grids are marked in Fig. 5. The bathymetry of the downscaling area is also shown in Fig. 4. Each hourly situation is defined by the wave and wind fields around the area of interest: $X_i^* = \left\{H_{s,1}, T_{m,1}, \theta_{m,1}, \dots, H_{s,n1}, T_{m,n1}, \theta_{m,n1}, W_{10x,1}, W_{10y,1}, \dots, W_{10x,n\,2}, W_{10y,n2}\right\}_i \quad i = 1, \dots, N$, where $n1$ $(=9)$ is the number of wave data locations, $n2$ $(=4)$ is the number of wind data locations and N $(=534.000)$ is the total amount of hourly situations (Fig. 6).

The hourly situations are highly correlated among different grid points of a given variable and among different variables. The high dimensionality of spatial fields can be reduced using the principal component analysis (PCA), to extract as much correlations as possible from the spatial fields while maintaining the diversity of the climate situations. The previous dimensionality reduction simplifies the selection process and time series reconstruction. The wave and wind directions are transformed to x and y components and all the variables are then standardized (with a zero mean and a standard deviation of one) for each grid point, to avoid problems due to different scales. After these transformations, the dimensionless data are defined as:

$$X_i = \left\{H_1, T_1, \theta_{x,1}, \theta_{y,1}, \dots, H_{n1}, T_{n1}, \theta_{x,n1}, \theta_{y,n1}, \dots, W_{x,1}, W_{y,1}, \dots, W_{x,n2}, W_{y,n2}\right\} \times (i = 1, \dots, N).$$

The PCA reduces the dimension of the data by means of a projection in a lower dimensional space preserving the maximum variance of the sample data. Given the spatial–temporal variable $X_i(x,t_i)$, where x are the spatial standardized variables of dimension $4n1+2n2$ and t_i is time, PCA is applied to obtain a new d-dimensional space. The eigenvectors (empirical orthogonal function, EOFs) of the data covariance matrix define the vectors of the new space. The idea of PCA is to find the minimum d linearly EOFs, so that the transformed components of the original data (principal components, PC) explain the maximum variance necessary for the problem at hand. The original data can be expressed as a linear combination of EOFs and PCs:

$$X(x, t_i) = EOF_1(x)\cdot PC_1(t_i) + EOF_2(x)\cdot PC_2(t_i) + \dots + EOF_d(x)\cdot PC_d(t_i). \quad (2)$$

Once PCA is applied, the data are defined by the principal components:

$$X_i^{EOF} = \{PC_1, PC_2, \dots, PC_d\}_i; i = 1, \dots, N. \quad (3)$$

Each PC accounts for a fraction of the variability in the data in a decrease order. Original data can be expressed in terms of a number of PCs (d), in which each one explains a certain variance, but supposing a reconstruction error when the transformed d-dimensional vectors (d PCs) are projected back to the original space. The explained variance is higher and the reconstruction error is lower when the number of PCs considered is larger. The criterion applied in this work for selecting an appropriate number of PCs is based on the reconstruction root-mean-square-error (RMSE) of the offshore wave and wind conditions for an increasing number of PCs (meaning the same that an increasing fraction of variance explained). The reconstruction RMSE of the original data variables have been computed

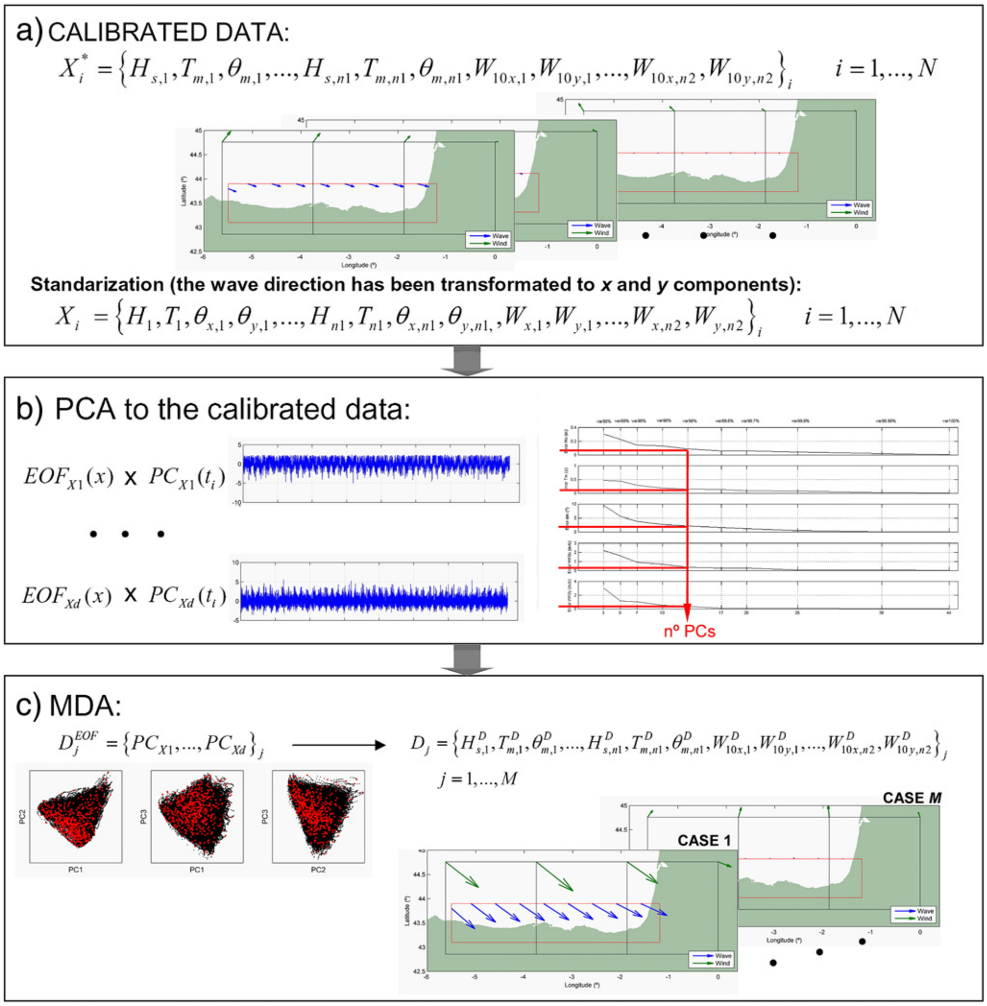

Fig. 4. Sketch of the stages of the selection step of the proposed methodology.

for an increasing number of PCs (ranging from 3 to 44) or explained variance (ranging from 70% to 100%). Fig. 7 shows the results, where the reconstruction error of each of the five variables (H_s, T_m, θ_m, W_{10x}, W_{10y}) is computed separately. For example, considering a number of PCs equal to 13, the errors are 0.1 m, 0.2 s, 2.5° and smaller to 1 m/s in the reconstruction of H_s, T_m, θ_m, W_{10x} and W_{10y},

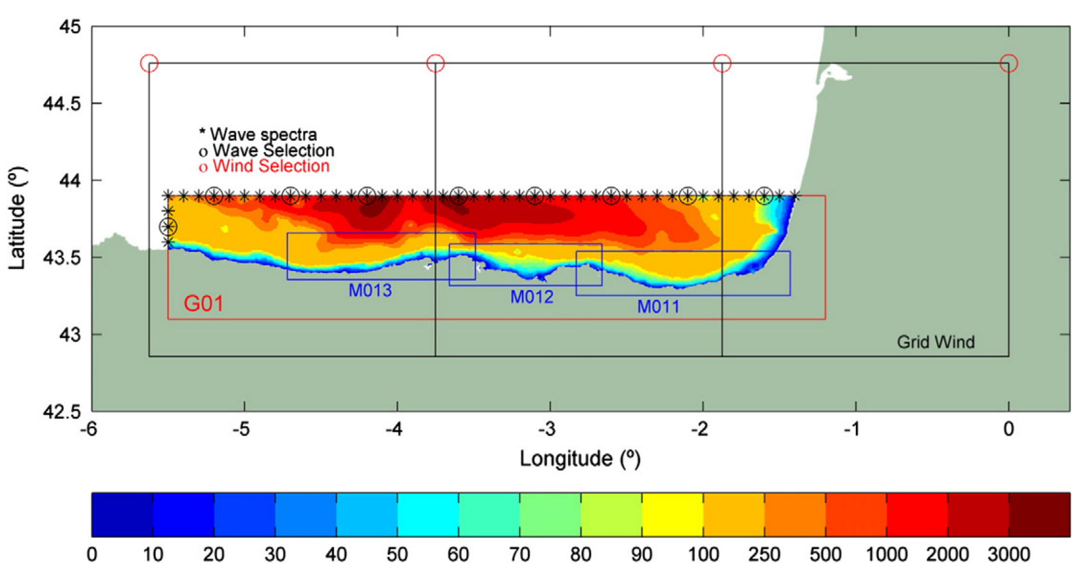

Fig. 5. Downscaling grids: G01 (0.01º × 0.008º) and M011, M012, M013 (0.005º × 0.004º). NCEP/NCAR wind grid (~1.875°) associated to wave propagations. The directional spectra at the boundaries of the computational (each 0.1°), the wave data and wind data considered in the definition of the computational conditions in the selection process. Bathymetry of the downscaled area (depth in m).

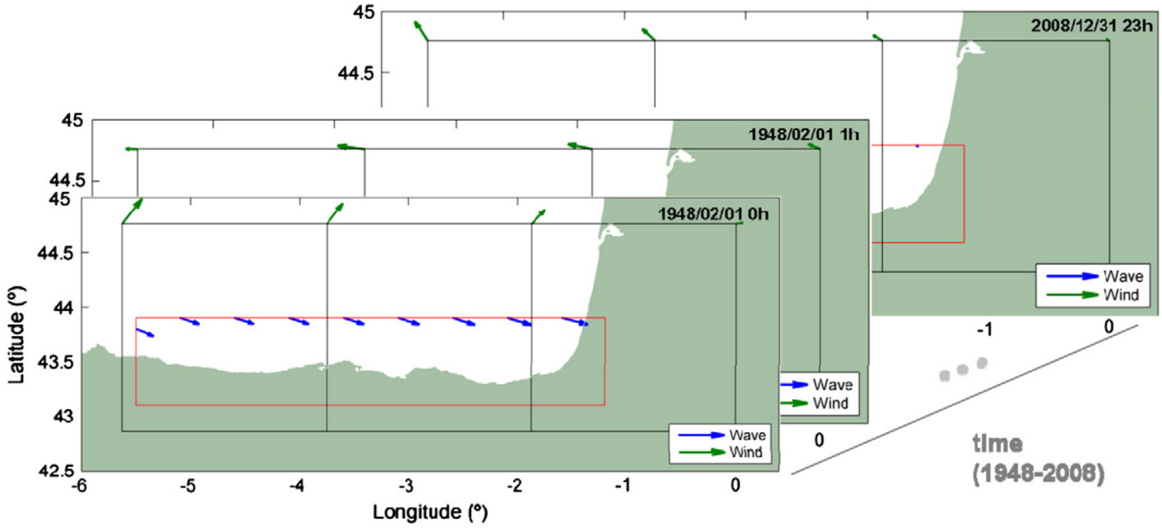

Fig. 6. Time series of the computational conditions (wave data along the grid boundaries of the wave computational model and wind data from the NCEP/NCAR database).

respectively. In this work, a number $d=13$ of PCs that explained 99.0% of the variance for the original data have been considered. Therefore, the dimension of the hourly wave and wind conditions is reduced from 35 to 13 with no significant information loss.

The next step consists of selecting a representative subset of size P using MDA $X_j^{EOF}=\{PC_1,PC_2,\ldots,PC_d\}_j; j=1,\ldots,P$. The first element selected is the one with the largest significant wave height, identified in the original space. Fig. 8 shows the subset of size $M=500$ elements selected in the EOF space $D_j^{\mathrm{EOF}}=\left\{PC_1^D,\ldots,PC_d^D\right\}_j \quad j=1,\ldots,M$, where we can see how the selected cases are fairly distributed in the data space. This subset selected by MDA is not projected back to the original space. The selected elements are identified in the original time series of

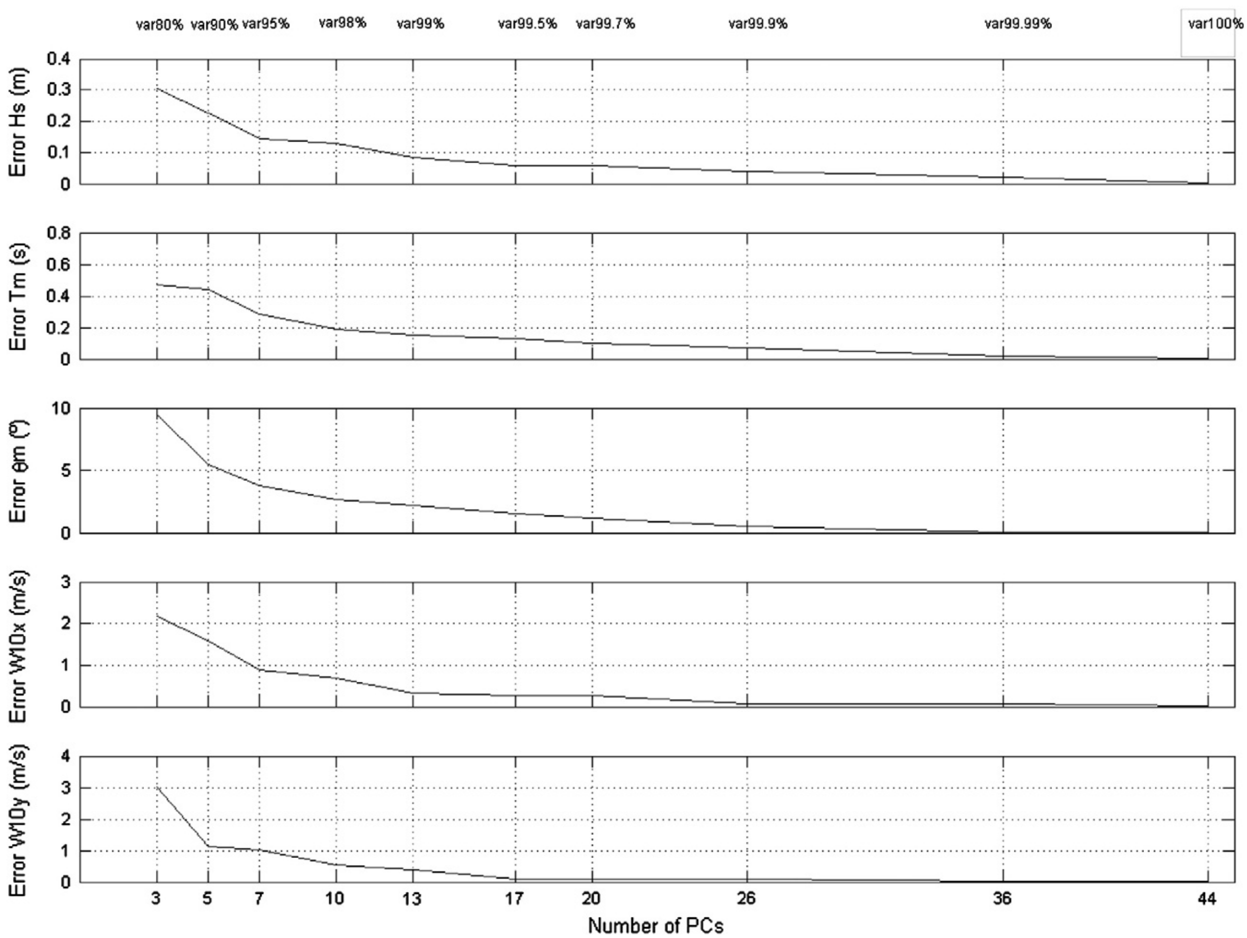

Fig. 7. Root mean square error of the variables that define the grid boundaries as a function of the explained variance.

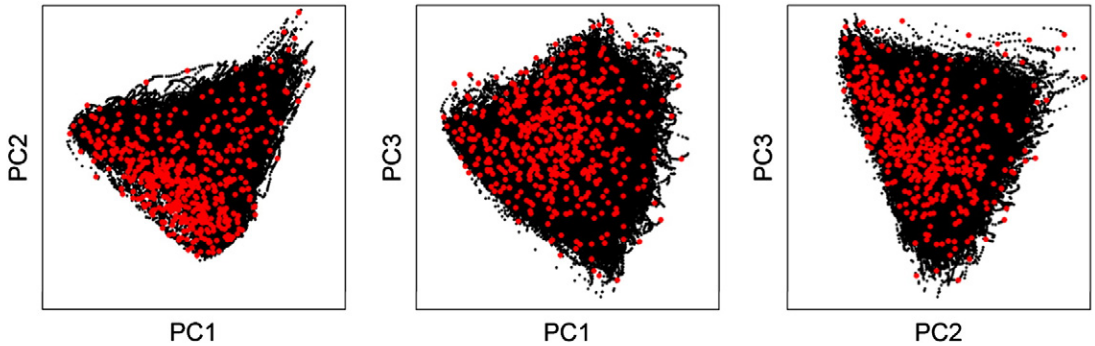

Fig. 8. $M = 500$ grid boundary selected by MDA in the *EOFs* space.

the wave conditions $D_j = \left\{H^D_{s,1}, T^D_{m,1}, \theta^D_{m,1}, \ldots, H^D_{s,n1}, T^D_{m,n1}, \theta^D_{m,n1}, W^D_{10x,1}, W^D_{10y,1}, \ldots, W^D_{10x,n2}, W^D_{10y,n2}\right\}_j$ $j = 1, \ldots, M$. Fig. 9 shows the $M = 500$ cases selected from the time series of the wave conditions in deep water, representative of the diversity of the deep water conditions.

6. Deep to shallow water wave transformation

The $M = 500$ selected cases by MDA, representative of the wave climate in deep water, are propagated to coastal areas using the numerical wave model SWAN (Booij et al., 1999). Note that the selection is applied to wave conditions defined by means of the sea state parameters H_s, T_m, θ_m and wind conditions defined by W_{10x}, W_{10y} at the grid nodes located in open water. However, the directional spectra at these grid nodes are available from the GOW database (calibrated in terms of H_s). Therefore, the spectrum and the corresponding wind fields are identified in the dates of the selected cases using MDA and define the wave boundaries of the computational grid and wind forcing. Besides the G01 computational grid with a resolution of 0.01°×0.008°, three other grids (M011, M012, M013) are considered with a resolution of 0.005°×0.004° (see Fig. 4).

Different sea state parameters at the nodes of the DOW grids are stored for each case (j): the propagated significant wave height ($H_{sp,j}$), the peak period ($T_{pp,j}$) and the mean wave direction ($\theta_{mp,j}$).

The subset of the $M = 500$ propagations in the downscaling domains defined a library (catalog) of cases formed by the $M = 500$ hourly sea state parameters $(H_{sp}, T_{pp}, \theta_{mp})_j$, corresponding to a certain sea state condition in deep water. An example of the catalog of the $M = 500$ situations for the wave parameter H_s is represented in Fig. 10.

7. Time series reconstruction

The reconstruction of the time series of wave parameters in any of the nodes of the DOW grids is carried out by an interpolation technique based on radial basis functions (RBF), a scheme which is very convenient for scattered and multivariate data (see details in Camus et al., 2011b).

For the implementation of the RBF interpolation technique in the sea state time series reconstruction, we have M d-dimensional points $D^{EOF}_j = \{PC^D_1, \ldots, PC^D_d\}_j; j = 1, \ldots, M$, corresponding to the M cases which are representative of the wave climate conditions in deep water. These are selected by the MDA algorithm in the EOF space (being d the number of PCs considered) and the associated real propagated parameters obtained by the numerical propagation at the shallow water location. These parameters are the propagated significant wave height $\{H^D_{sp,j}\}$, the propagated peak period $\{T^D_{pp,j}\}$ and the components x- and y- of the propagated mean direction $\{\theta x^D_{mp,j}, \theta y^D_{mp,j}\}$. The mean wave direction θ_{mp} is reconstructed after the interpolation of the components x- and y-. Therefore, the aim of the RBF application is the evaluation of the interpolation function for each sea state parameter.

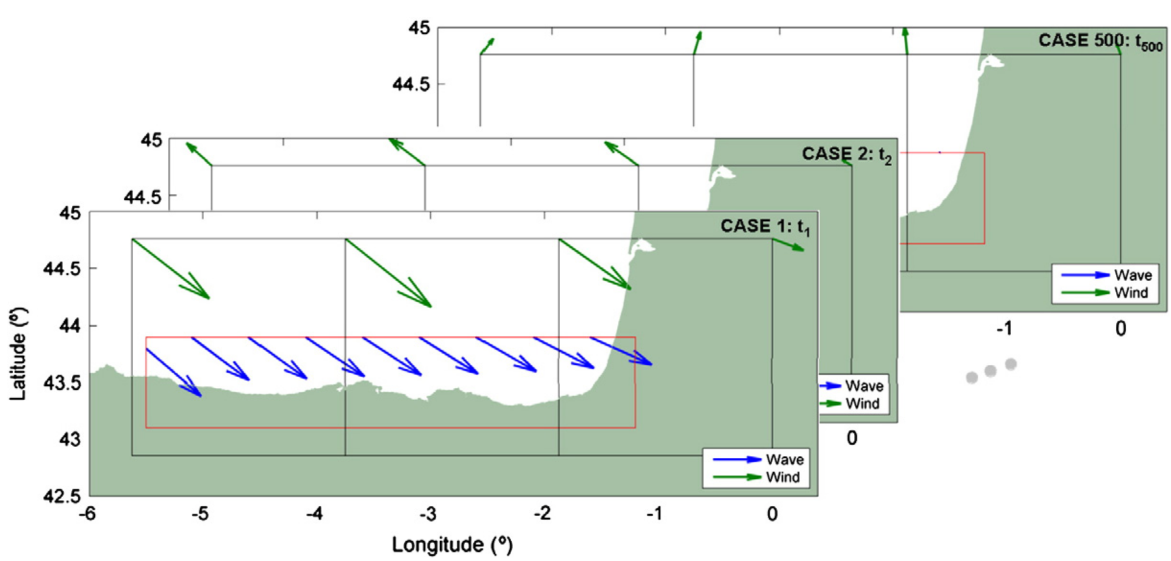

Fig. 9. $M = 500$ grid boundary selected by MDA.

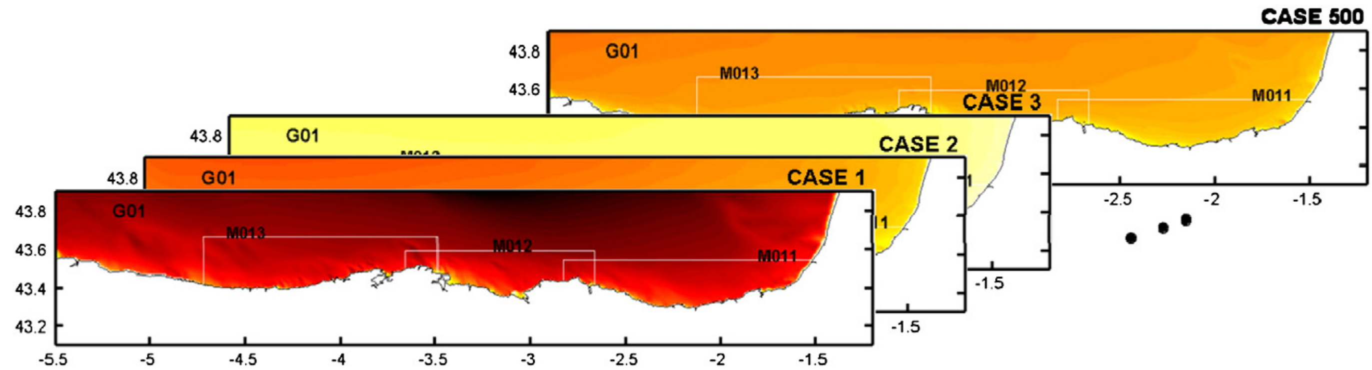

Fig. 10. Library with $M=500$ propagations.

In order to calculate the interpolation functions, the PCs which define each wave climate condition in deep water are normalized using a linear transformation which scales the values between 0 and 1. Each sea state in deep water is defined as $X_i^{EOF,\mathrm{norm}}=\{PC_1^{\mathrm{norm}},\ldots,PC_d^{\mathrm{norm}}\}_i$, while each selected case, where the real propagated parameters are available, is expressed as $D_j^{EOF,\mathrm{norm}}=\{PC_1^{D,\mathrm{norm}},\ldots,PC_d^{D,\mathrm{norm}}\}_j$.

The interpolation function is calculated by means of this expression:

$$\mathrm{RBF}\left(X^{EOF,\mathrm{norm}}\right)=p\left(X^{EOF,\mathrm{norm}}\right)+\sum_{j=1}^{M}a_j\Phi\left(\left\|X^{EOF,\mathrm{norm}}-D_j^{EOF,\mathrm{norm}}\right\|\right) \quad (4)$$

where $\Phi$is the radial basis function, being |||| the Euclidian norm; $p(x)$ is a monomial basis $\{p_0,p_1,\ldots,p_d\}$, formed by a number of monomials of degree 1 equal to the data dimension (d) and a monomial of degree 0, being $b=\{b_0,b_1,\ldots,b_d\}$ the coefficients of these monomials. Therefore, $p(X^{EOF,\mathrm{norm}})=b_0+b_1PC_1^{\mathrm{norm}}+b_2PC_2^{\mathrm{norm}}+\ldots+b_dPC_d^{\mathrm{norm}}$ and Φ is a gaussian function with a shape parameter c:

$$\Phi\left(\left\|X^{EOF,\mathrm{norm}}-D_j^{EOF,\mathrm{norm}}\right\|\right) = \exp\left(-\frac{\left\|X^{EOF,\mathrm{norm}}-D_j^{EOF,\mathrm{norm}}\right\|^2}{2c^2}\right). \quad (5)$$

The optimal shape parameter is estimated using Rippa (1999) algorithm. The coefficients $b_l=[b_0,b_1,b_2,\ldots,b_d]^T$ of the monomials and the coefficients $a_j=[a_1,\ldots,a_M]^T$ of the radial basis functions are obtained by enforcing the interpolation conditions:

$$\mathrm{RBF}\left(D_j^{EOF,\mathrm{norm}}\right)=f_j\left(D_j^{EOF,\mathrm{norm}}\right)=D_{p,j};\quad j=1,\ldots,M \quad (6)$$

where the real functions $D_{p,j}$ are defined by the propagated parameters $\left\{H_{sp}\right\}_j$, $\left\{T_{pp}\right\}_j$, $\left\{\theta_{xp}\right\}_j$ or $\left\{\theta_{yp}\right\}_j$, corresponding to the selected sea states in MDA algorithm D_j.

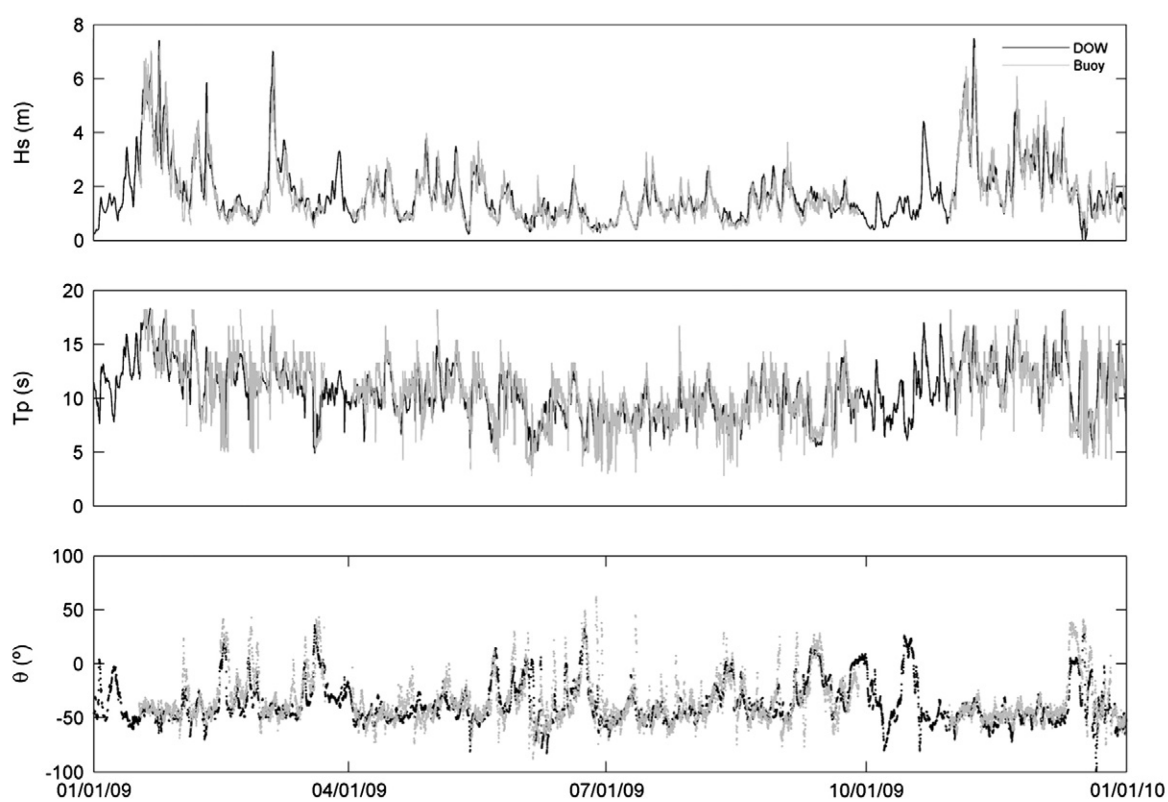

Fig. 11. Comparison between reconstructed and measured H_s, T_p and θ_m at Virgen Mar buoy 2009.

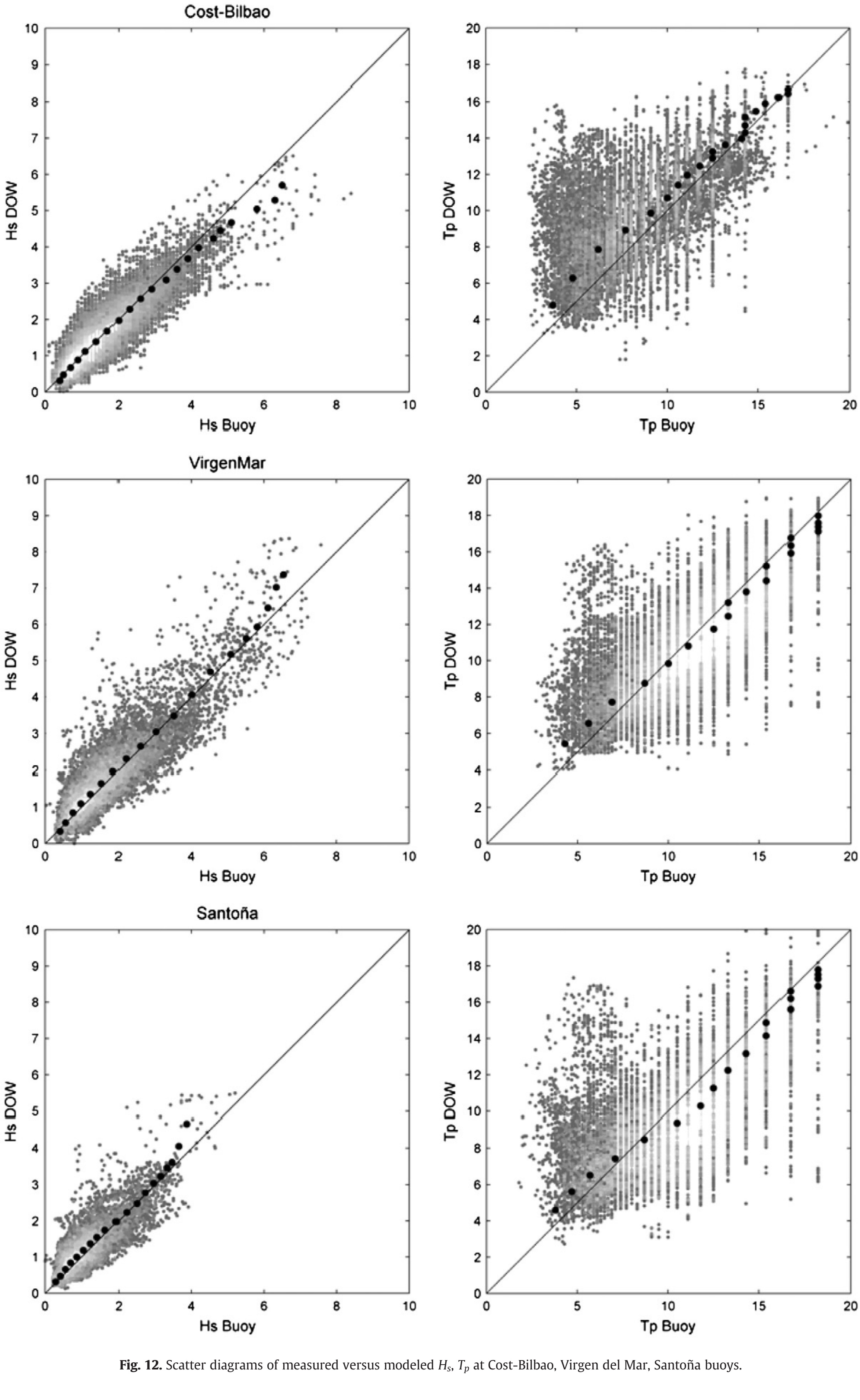

Fig. 12. Scatter diagrams of measured versus modeled H_s, T_p at Cost-Bilbao, Virgen del Mar, Santoña buoys.

Therefore, the time series $X^{EOF,\text{norm}} = \{PC_1^{\text{norm}}, \ldots, PC_d^{\text{norm}}\}$are transferred from deep water to the location of interest in shallow water by means of the RBF functions calculated for each propagated parameter. These functions are defined as:

$$H_{sp,i} = \text{RBF}_H\left(\left\{D_j^{EOF,\text{norm}}, H_{sp,j}(j=1,\ldots,M)\right\}, X_i^{EOF,\text{norm}}\right); i=1,\ldots,N \quad (7)$$

$$T_{pp,i} = \text{RBF}_{T_p}\left(\left\{D_j^{EOF,\text{norm}}, T_{pp,j}(j=1,\ldots,M)\right\}, X_i^{EOF,\text{norm}}\right); i=1,\ldots,N \quad (8)$$

$$\theta x_{mp,i} = \text{RBF}_{\theta_x}\left(\left\{D_j^{EOF,\text{norm}}, \theta x_{mp,j}(j=1,\ldots,M)\right\}, X_i^{EOF,\text{norm}}\right); i=1,\ldots,N \quad (9)$$

$$\theta y_{mp,i} = \text{RBF}_{\theta_y}\left(\left\{D_j^{EOF,\text{norm}}, \theta y_{mp,j}(j=1,\ldots,M)\right\}, X_i^{EOF,\text{norm}}\right); i=1,\ldots,N. \quad (10)$$

A general transfer function for a specific location can be defined as:

$$X_{p,i} = \text{RBF}\left(\left\{D_j^{EOF,\text{norm}}, D_{p,j}(j=1,\ldots,M)\right\}, X_i^{EOF,\text{norm}}\right); \quad i=1,\ldots,N. \quad (11)$$

The final result is the reconstructed time series at a specific location at shallow water:

$$X_{p,i} = \left\{H_{sp,i}, T_{pp,i}, \theta_{mp,i}\right\}; \quad i=1,\ldots,N. \quad (12)$$

8. Validation

8.1. Time series

The validation of the reconstructed time series was made with measured data in shallow (intermediate) water buoys at four different locations in the study area, as shown in Fig. 2. The time period for each validation was chosen according to the available measured hourly data (the comparison with the buoy measurements was made for the following periods during which data was available): Cost-Bilbao Buoy (10.02.2004–16.04.2009), Pasaia Buoy (01.01.2003–01.10.2009), Virgen del Mar Buoy (17.01.2009), and Santoña Buoy (17.01.2009). For the second buoy, only the significant wave height was available while for the other three directional data were available.

Some direct comparisons between the reconstructed time series using the proposed methodology and instrumental time series at Virgen Mar (year 2009) are provided in Fig. 11 for the significant wave height, the peak period and the mean direction. The discontinuities in the curves describing the buoy data reflect some gaps that were encountered in the measured data field. The comparison between the time series depicted in the figures shows an overall good agreement between the measured and predicted sea state parameters with an accurate reproduction of the time series structure, even that of the extreme events.

Fig. 12 shows scatter and quantile–quantile (20 equally distributed Gumbel quantiles) plots of the measured versus modeled H_s, T_p for the entire dataset of each buoy indicating the general good quality of the results obtained. Several diagnostic statistics are calculated to compare model performance with respect to instrumental data, such as the root mean square error (RMSE), the Pearson's correlation coefficient (ρ), the systematic deviation between two random variables (BIAS) and the residual scatter index (SI).

Table 1 provides the values of these diagnostic statistics for significant wave height and peak wave period, respectively, comparing the DOW reanalysis data versus buoy observations of the Cost-Bilbao, Pasaia, Virgen del Mar and Santoña buoys. The scatter indexes and the H_s correlation coefficients are below 0.33 and above 0.85 respectively. The biases related to H_s are practically zero. Correlations related to T_p are lower (0.74–0.80), being consistent with the results from global and regional reanalyses, but the results are quite adequate as well as the scatter index (<0.34).

Table 1
Correlation statistics for significant wave height and peak period between DOW and buoy observations.

Buoy name	H_s (m)				T_p (s)			
	RMSE (m)	ρ	BIAS (m)	SI	RMSE (m)	ρ	BIAS (m)	SI
Bilbao	0.34	0.92	0.00	0.25	2.17	0.77	−0.95	0.24
Pasaia	0.46	0.88	0.00	0.33	–	–	–	–
Virgen Mar	0.49	0.89	−0.09	0.31	2.22	2.22	−0.01	0.22
Santoña	0.36	0.85	−0.11	0.35	2.73	2.73	0.30	0.29

8.2. Statistical distributions of H_s and θ_m

In order to validate the statistical distribution of H_s and θ_m at buoy locations, Fig. 13 displays the empirical bivariate distribution of H_s versus θ_m for the directional buoys (Cost-Bilbao, Virgen Mar and Santoña). The distribution of buoy data is shown in the left panel while the distribution of the downscaled reanalysis data is in the right panel. It can be observed that the reconstruction of the time series reproduces the wave directional distribution. The wave climate at this region is influenced by waves from the NW–NE sectors with the most frequent and most energetic sea states coming from the NW in deep water. These most frequent and most energetic waves are from the NW direction at Virgen del Mar buoy, turning to NNW at Bilbao buoy and with a higher north component at Santoña buoy due to the orientation of the coast at these locations. The directional distribution of the reconstructed time series reproduces the wave transformation processes as can be observed in similarity to the buoy directional distribution. The main differences are found in the northeastern sea states, which are mainly wind seas, due to a deficient description of the wind fields. This shortcoming also affects the general differences in the wave direction to all sectors. Therefore, further research is needed to improve the performance of the wave parameters using dynamical downscaled winds as forcing in the generation models at regional and local scales.

9. Conclusions

A global framework to downscale wave reanalysis to coastal areas, which takes into account a correction of open sea significant wave height (directional calibration) and a hybrid methodology based on numerical models (dynamical downscaling) and mathematical tools (statistical downscaling), is presented. The hybrid method extends and improves the previous one developed by Camus et al. (2011b), which consists of a selection of a subset of sea states representative of the wave and wind conditions in deep water using MDA, the numerical propagation of these selected cases and the time series reconstruction using the RBF interpolation technique. In this case, the spatial variability of the wave boundaries of the propagation domain and the local wind wave generation, using the simultaneous wind fields, are taken into account, in a similar way as if the coastal wave databases were generated by means of dynamical downscaling. Therefore, the new methodology is adapted to high dimensional wave and wind data in deep water. Although the MDA and RBF methods are able to deal with highly dimensional data, a previous reduction of the dimension is applied using PCA in order to simplify the selection and reconstruction processes. Therefore, once the computational domain is defined, the wave boundaries and the wind fields are calibrated to correct the deficiencies with respect to

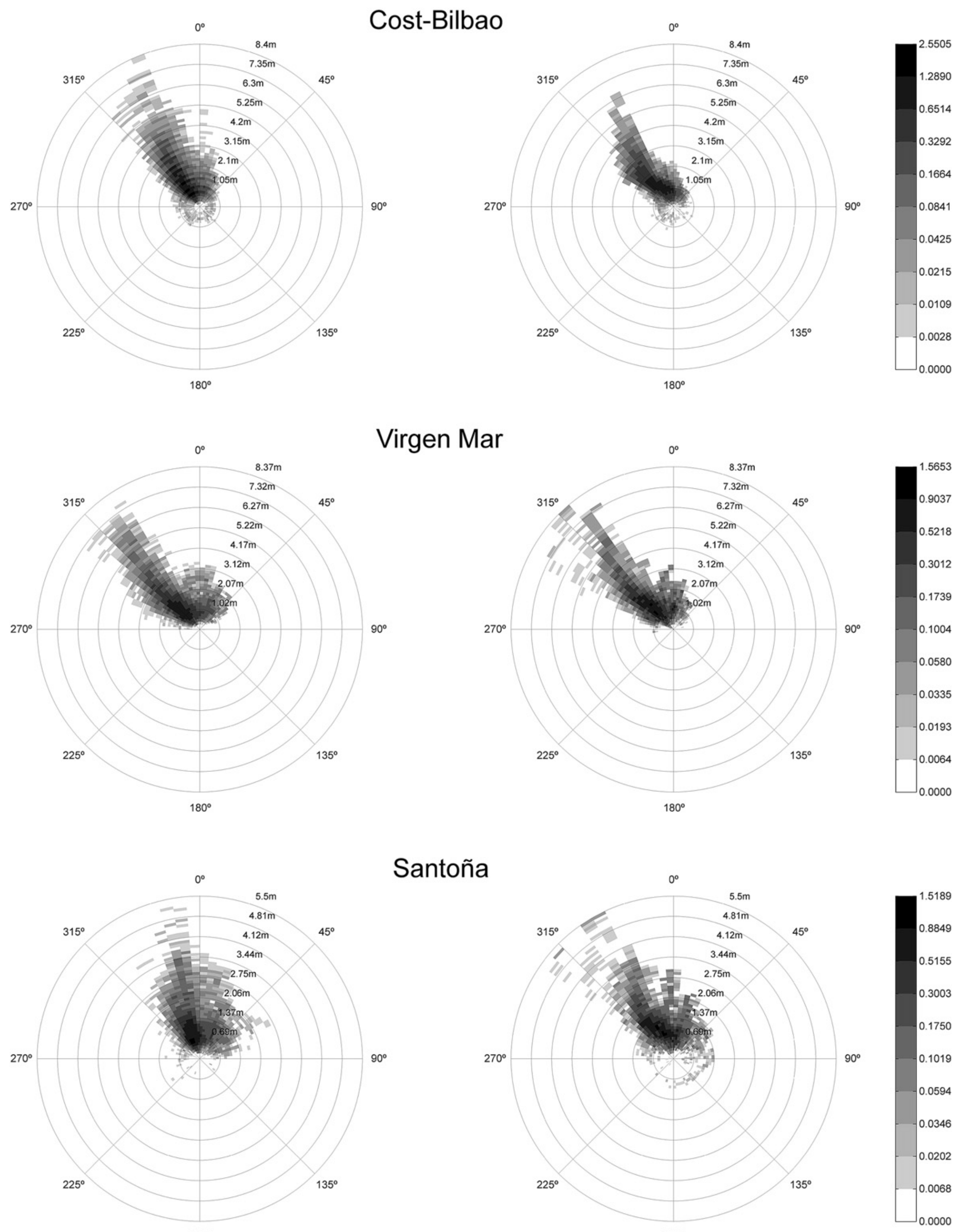

Fig. 13. Directional distribution of H_s at Cost-Bilbao buoy, Virgen Mar buoy and Santoña buoy. The instrumental distribution is displayed on the left panel while the reanalysis distribution is on the right panel.

the instrumental data from the deep water wave reanalysis. After that, the data dimension is reduced using the PCA and the number of PCs is selected in analyzing the reconstruction RMSE due to the dimension reduction, restricting the loss of information. The MDA is applied to the reduced data (d selected PCs) but the wave boundaries (defined via the spectrum) and wind fields are identified in the real space for the performance of the numerical propagations. The reconstruction is also done as a function of the wave boundaries and wind fields defined in the reduced space.

The results show the ability of the proposed methodology to reconstruct the time series of the sea state parameters H_s, T_p and θ_m in shallow water reproducing the time series structure and the different statistical parameters with high accuracy. The computational time effort of the proposed methodology, with respect to a continuous non-stationary simulation of a coastal wave reanalysis, is drastically reduced (~1000×). This methodology supposes a valuable tool in coastal engineering design purposes, especially due to its reliability in the statistical characterization of extreme events.

Acknowledgments

The authors would like to thank Puertos del Estado (Spanish Ministry of Public Works), Government of Cantabria and the Basque Agency of Meteorology (EUSKALMET) for providing the buoy data. The work was partially funded by projects "GRACCIE" (CSD2007-00067, CONSOLIDER-INGENIO 2010) from the Spanish Ministry of Science and Technology, "MARUCA" (200800050084091) from the Spanish Ministry of Public Works, "C3E" (E17/08) from the Spanish Ministry of Environment, Rural and Marine Affairs, "Regional Wave Climate Projections along the Spanish Coast" (BIA2011-28909).

References

Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions. Part I: model description and validation. Journal of Geophysical Research 104 (C4), 7649–7666.

Breivik, Ø., Gusdal, Y., Furevik, B.R., Aarnes, O.J., Reistand, M., 2009. Nearshore wave forecasting and hindcasting by dynamical and statistical downscaling. Journal of Marine Systems 78, S235–S243.

Browne, M., Castelle, B., Strauss, D., Tomlinson, R., Blumenstein, M., Lane, C., 2007. Near-shore swell estimation from a global wind–wave model: spectral process, linear and artificial neural network models. Coastal Engineering 54, 445–460.

Camus, P., Méndez, F.J., Medina, R., Cofiño, A.S., 2011a. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coastal Engineering 58 (6), 453–462.

Camus, P., Méndez, F.J., Medina, R., 2011b. A hybrid efficient method to downscale wave climate to coastal areas. Coastal Engineering 58 (9), 851–862.

Chini, N., Stansby, P., Leake, J., Wolf, J., Robert-Jones, J., Lowe, J., 2010. The impact of sea level and climate change on inshore wave climate: a case study for East Anglia (UK). Coastal Engineering 57, 973–984.

Galiskova, L., Weisse, R., 2006. Estimating near-shore wave statistics from regional hindcasts using downscaling techniques. Ocean Dynamics 56, 26–35.

Groeneweg, J., van Ledden, M., Zijlema, M., 2007. Wave transformation in front of the Dutch Coast. In: McKee Smith, Jane (Ed.), Proc. of International Conference Coastal Engineering, ASCE, pp. 552–564.

Herman, A., Kaiser, R., Niemeyer, H.D., 2009. Wind–wave variability in shallow tidal sea — spectral modelling combined with neural network methods. Coastal Engineering 56, 759–772.

Kalnay, E.M., Kanamitsu, R., Kistler, W., Collins, D., Deaven, L., Gandin, M., Iredell, S., Saha, G., White, J., Woollen, Y., Zhu, M., Chelliah, W., Ebisuzaki, W., Higgins, J., Janowiak, K.C., Mo, C., Ropelewski, J., Wang, A., Leetmaa, R., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437–470.

Kalra, R., Deo, M.C., Kumar, R., Agarwal, V.K., 2005. Artificial neural network to translate offshore satellite wave to data to coastal locations. Ocean Engineering 32, 1917–1932.

Kennard, R.W., Stone, L.A., 1969. Computer aided design experiments. Technometrics 11, 137–148.

Mínguez, R., Espejo, A., Tomás, A., Méndez, F.J., Losada, I.J., 2011a. Directional calibration of wave reanalysis databases using instrumental data. Journal of Atmospheric and Oceanic Technology. http://dx.doi.org/10.1175/JTECH-D-11-00008.1.

Mínguez, R., Reguero, B.G., Luceño, A., Méndez, F.J., Losada, I.J., 2011b. Regression models for outlier identification (hurricanes and typhoons) in wave hindcast databases. Journal of Atmospheric and Oceanic Technology. http://dx.doi.org/10.1175/JTECH-D-11-00059.1.

NOAA, National Geophysical Data Center, 2006. Etopo2v2 Global Gridded 2-minute Database.

Polinsky, A., Feinstein, R.D., Shi, S., Kuki, A., 1996. Librain: software for automated design of exploratory and targeted combinatorial libraries. In: Chaiken, I.M., Janda, K.D. (Eds.), Molecular Diversity and Combinatorial Chemistry: Libraries and Drug Discovery. American Chemical Society, Washington, D.C., pp. 219–232.

Reguero, B.G., Menéndez, M., Méndez, F.J., Mínguez, R., Losada, I.J., 2011. A global ocean wave (GOW) calibrated reanalysis from 1948 onwards. Coastal Engineering 65, 38–55.

Rippa, S., 1999. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Advances in Computational Mathematics 11, 193–210.

Rogers, W.E., Kaihatu, J.M., Hsu, L., Jensen, R.E., Dykes, J.D., Holland, K.T., 2007. Forecasting and hindcasting waves with the SWAN model in the Southern California Bight. Coastal Engineering 54, 1–15.

Rusu, L., Pilar, P., Guedes Soares, C., 2008. Hindcast of the wave conditions along the west Iberian coast. Coastal Engineering 55, 906–919.

Snarey, M., Terrett, N.K., Willet, P., Wilton, D.J., 1997. Comparison of algorithms for dissimilarity-based compound selection. Journal of Molecular Graphics and Modelling 15, 372–385.

Stansby, P., Zhou, J., Kuang, C., Walkden, M., Hall, J., Dickson, M., 2007. Long-term prediction of nearshore wave climate with an application to cliff erosion. In: McKee Smith, Jane (Ed.), Proc. of International Conference Coastal Engineering, ASCE, pp. 616–627.

Tolman, H.L., 2002. User manual and system documentation of WAVEWATCH III version 2.22. NOAA / NWS / NCEP Technical Note.

Willet, P., 1996. Molecular diversity techniques for chemical databases. Information Research 2 (3) (Available at: http://informationr.net/ir/2-3/paper19.html).