The influence of seasonality on estimating return values of significant wave height

Melisa Menéndez[1*], Fernando J. Méndez [1], Cristina Izaguirre[1], Alberto Luceño[2], Inigo J. Losada[1]

[1] Ocean & Coastal Research Group, IH Cantabria, Universidad de Cantabria, Spain
[2] Dpto. Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Spain
* Corresponding author

Manuscript submitted to: Journal of Geophysical Research
Date of submission: September 13, 2007

Corresponding author address
Ocean & Coastal Research Group (Instituto de Hidráulica Ambiental "IH Cantabria"), Universidad de Cantabria.
Avda. de los Castros s/n 39005 Santander (SPAIN)
Phone: +34-942-201810
Fax: +34-942-201860
e-mail: menendezm@unican.es

Keywords

generalized extreme value, maximum likelihood, return level, seasonality, statistical model, time-dependent, wave climate

Abstract

A time-dependent generalized extreme value (GEV) model for monthly significant wave heights maxima is developed. The model is applied to several 3-hour time series from the Spanish buoy network. Monthly maxima show a clear non-stationary behavior within a year, suggesting that the location, scale and shape parameters of the GEV distribution can be parameterized using harmonic functions. To avoid a possible over-parameterization, an automatic selection model, based on the Akaike Information Criterion, is carried out. Results show that the non-stationary behavior of monthly maxima significant wave height is adequately modeled, drastically increasing the significance of the parameters involved and reducing the uncertainty in the return level estimation. The model provides new information to analyze the seasonal behavior of wave height extremes affecting different natural coastal processes.

1. INTRODUCTION

Recent advances in the extreme value theory [see Coles, 2001 and Katz et al., 2002 as general references] have appeared in the state-of-the-art allowing a better description of the natural climate variability of extreme events of geophysical variables. The modelling of the seasonality of extreme events can improve our knowledge on some important natural coastal processes such as: the seasonal distribution of benthic organisms in wave-swept environments; the seasonal variability of flow and particle distribution in nearshore seagrass meadows; the influence of wave-exposure on the growth-erosion rates and dislodgement of kelp in the surf zone or the seasonality of the sediment transport rate. The analysis of these processes may require an estimation of a given return-period level conditioned to a given season or month. Moreover, the determination of the return levels of extreme significant wave height, H_s, is vital for other purposes such as coastal management, including the analysis of coastal flooding risk and the design of maritime works. In the latter case, the definition of working time windows during the construction phase or the evaluation of the harbour operation time frames after construction during the winter season requires considering the seasonality or monthly characteristics in the estimation of the return values.

The calculation of extreme quantiles is often applied to statistical models which use annual maximum data. Due to the scarcity of this type of extreme values, some alternatives are proposed, such as the *r*-largest maxima method [*Guedes Soares and Scotto*, 2004] or the peak over threshold (POT) approach [*Goda*, 2000]. Another possibility is to fit a distribution using monthly maxima [f.i. *Panchang and Li*, 2006]. In general, these methods assume a homogeneous distribution for the extreme population data within a year. However, the hypothesis of homogeneity is not adequately satisfied, since the effects of seasonality are evident [*Holthuijsen*, 2007]. To illustrate the situation, Figure 1 shows the boxplots for monthly maxima series for five buoys of Puertos del Estado network (see locations in Figure 2). To facilitate the visualization of the extreme events, the winter season has been placed in the center of all Figures. An important modulation in the mean values as well as in the variability of the data is detected.

In an attempt to model the seasonal behavior of the maximum significant wave height within a year, *Carter and Challenor* [1981] proposed a month-to-month distribution, assuming that data are identically distributed within a given month and analyzing them separately. Subsequently, an annual distribution is obtained by combining the monthly distributions. A similar analysis is performed by *Morton et al.* [1997], applying a seasonal POT model to wind and significant wave height data.

Basically, all the aforementioned methods require the random variable to be independent and identically distributed (IID) in the block of adopted time (year, season or month). The latter hypothesis can be relaxed to incorporate smooth time variations of the random variable. Examples of this approach applied to different geophysical variables can be found in *Coles* [2001], *Katz et al* [2002] or *Mendez et al.* [2007]. Recently, *Mendez et al.* [2006] developed a time-dependent POT model for extreme significant wave height which considers the parameters of the distribution to be functions of time (harmonics within a year, exponential long-term trend, El Niño covariate, etcetera). However, that work focuses on the definition of the higher extreme events of the year (values exceeding a given threshold) and disregards the extreme events in the summer season, therefore being unable to model the entire variability within a year. The object of this article is to develop a time-dependent model based on the GEV distribution, that accounts for seasonality using independent monthly maxima events z_i observed at instants t_i, thus considering 12 maximum values per year. The non-stationary behavior of extreme significant wave height is parameterized using functions of time (harmonic functions) for the parameters of the distribution. The model is applied to five scalar buoys along the Spanish coast (see details in Table 1 and Figure 2). Through this approach, the drastic reduction of the uncertainty in the estimation of time-dependent (monthly) quantiles and the improvement in the estimation of annual return values will be shown.

The paper is organized as follows. Section 2 provides a brief description of the time-dependent generalized extreme value distribution and the parameter estimation method. Section 3 describes the regression model adopted. Next, an automatic model selection procedure is performed and explained in Section 4. The application of the model for the determination of the time-dependent quantiles and the annual quantiles is shown in Section 5. Finally, some conclusions are given in Section 6.

2. THE TIME-DEPENDENT GEV DISTRIBUTION

The monthly maximum method uses time series of block maxima for successive months, $\{Z_t = \max(H_{t1},...,H_{tN})\}$, which are called monthly maxima series (MMS), where the H_{ti}'s, for $i = 1,...,N$, are the N values of significant wave height sampled in a given month t and $t = 1,...,n$. A critical aspect with discontinuous time series (the buoys records present gaps) is the consideration of a minimum number of data per unit time to define the maxima values. This fact affects the stability of the parameter estimates for the extreme value distribution. After some tests, we adopt the criterion to reject a monthly maximum event if the percentage of gaps for that given month amount over 40% [*Mendez et al.*, 2007].

Considering the apparent temporal dependence within MMS due to seasonal effects, a way to work with MMS is to fit individually the maximum values of each month into a probability distribution and combine the monthly distributions to obtain an annual distribution [*Carter and Challenor*, 1981]. This implies to fit twelve models, thus obtaining 36 parameters (using, for example, the GEV distribution). This amount of parameters introduces a large uncertainty in the model, diminishing the validity of the results.

Another possibility to account for this temporal dependence is to use an extension of the standard models of extreme value theory for non-stationary variables [chapter 6, *Coles*, 2001]). Monthly maxima of successive months are assumed to be independent random variables, but the hypothesis of homogeneity through consecutive months is not needed (because they are not presumed to be identically distributed). We assume that the monthly maximum Z_t of the significant wave heights observed in month t follows a GEV distribution with time-dependent location parameter $\mu(t) > 0$, scale parameter $\psi(t) > 0$, and shape parameter $\xi(t)$. The cumulative distribution function (CDF) of Z_i is then given by

$$F_t(z) = \begin{cases} \exp\left\{ -\left[1 + \xi(t)\left(\dfrac{z-\mu(t)}{\psi(t)} \right) \right]_+^{-1/\xi(t)} \right\} & \xi(t) \neq 0 \\ \exp\left\{ -\exp\left[-\left(\dfrac{z-\mu(t)}{\psi(t)} \right) \right] \right\} & \xi(t) = 0 \end{cases}, \tag{1}$$

where $[a]_+ = \max\left[a, 0\right]$.

The GEV distribution includes three distribution families corresponding to the different form of the tail behavior: Gumbel family in the case of null shape parameter, with a light tail decaying exponentially; Fréchet distribution with $\xi > 0$ and a heavy tail decaying polinomially and Weibull family with $\xi < 0$ and a bounded tail (note that this Weibull for maxima distribution differs from the commonly used Weibull for minima distribution adopted in the POT method for some engineering applications [*Goda*, 2000]). The probability density function (PDF) of Z_i is obtained by differentiating (1) with respect to *z*, so that,

$$f_t(z) = \begin{cases} \dfrac{1}{\psi(t)}\left[1 + \xi(t)\left(\dfrac{z-\mu(t)}{\psi(t)} \right) \right]_+^{-(1+1/\xi(t))} \exp\left\{ -\left[1 + \xi(t)\left(\dfrac{z-\mu(t)}{\psi(t)} \right) \right]_+^{-1/\xi(t)} \right\}, & \xi(t) \neq 0 \\ \dfrac{1}{\psi(t)} \exp\left(-\dfrac{z-\mu(t)}{\psi(t)} \right) \exp\left[-\exp\left(-\dfrac{z-\mu(t)}{\psi(t)} \right) \right], & \xi(t) = 0 \end{cases} \tag{2}$$

For notational purposes, the time-dependent GEV pdf of Equation (2) will be expressed using the following identity:

$$f_t(z) \equiv f_t(z;\theta), \tag{3}$$

where θ encompasses the three parameters $\mu(t)$, $\psi(t)$ and $\xi(t)$ as indicated later.

In Section 3, a representation of $\mu(t)$, $\psi(t)$ and $\xi(t)$ by means of harmonic functions (annual cycle, semiannual cycle, etc) will be used. For illustration, Figure 3 shows the time-dependent PDF of equation (2) for the best model found for the Gijon buoy, which uses $\mu(t) = 3.11 + 1.14\cos(2\pi t) + 0.15\sin(2\pi t)$ for the location parameter, $\psi(t) = 0.79 + 0.31\cos(2\pi t) + 0.07\sin(2\pi t)$ for the scale parameter and $\xi(t) = -0.12$ for the shape parameter. Time t is in years, with origin t = 0 at the beginning of the year, and the location and scale parameters are in meters. Note how the seasonality affects the shape of the GEV probability density function, which takes its maximum value for small wave heights in summer (t = 0.5 and 1.5 years) but for large wave heights in winter (t = 0, 1 and 2 years).

3. REGRESSION MODEL

After a visual inspection of the box-plots of Figure 1, it seems reasonable to allow for seasonality in the model considering harmonic functions within a year. In order to further support this evidence, we examine graphically the variability within a year of the location, scale and shape parameters of the stationary GEV distribution fitted for every month individually. The maximum likelihood estimates (MLEs) $\hat{\mu}$, $\hat{\psi}$ and $\hat{\xi}$ of these parameters over a year (from July to June) for La Coruña buoy are shown in Figure 4. The location parameter shows a single peak, possibly due to the highest extreme events in December-January. The scale parameter is also modulated along a year with a maximum value in the winter season. The results for the shape parameter are not so evident due to the uncertainty in the estimation of this parameter with only one observation per year for every month. Figure 4 also shows the regression fit for the first two harmonics. The good quality of the fit suggests the use of harmonic functions within a year to approximate the seasonal behavior. Mathematically, this model can be expressed as

$$\mu(t) = \beta_0 + \sum_{i=1}^{P_\mu} \left[\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t) \right] \tag{4}$$

$$\psi(t) = \alpha_0 + \sum_{i=1}^{P_\psi} \left[\alpha_{2i-1} \cos(2i\pi t) + \alpha_{2i} \sin(2i\pi t) \right] \tag{5}$$

provided $\psi(t) > 0$, and

$$\xi(t) = \gamma_0 + \sum_{i=1}^{P_\xi} \left[\gamma_{2i-1} \cos(2i\pi t) + \gamma_{2i} \sin(2i\pi t) \right] \tag{6}$$

where β_0, α_0 and γ_0 are mean values; β_i, α_i and γ_i ($i > 0$) are the amplitudes of the harmonics; P_μ, P_ψ, and P_ξ are the number of sinusoidal harmonics in a year; and t is given in years.

The parameters of a possible model (see the example of Figure 3 applied to Gijon buoy) can be packed into the vector $\theta = (\beta_0, \beta_1, \beta_2, \alpha_0, \alpha_1, \alpha_2, \gamma_0)$. In this particular case, the fitted model contains annual cycles for the location and scale parameters, $\mu(t) = \beta_0 + \beta_1 \cos(2\pi t) + \beta_2 \sin(2\pi t)$ and $\psi(t) = \alpha_0 + \alpha_1 \cos(2\pi t) + \alpha_2 \sin(2\pi t)$, and a constant value for the shape parameter $\xi(t) = \gamma_0$. The number of significantly non-null regression parameters is therefore $p = 7$.

For any of the candidate models, represented by its vector parameter θ, and for m observations of monthly maxima Z_{ti} ocurring at instants t_i, we estimate the model parameters $\hat{\theta}$ using the method of maximum likelihood. Approximate standard errors, $\mathrm{se}(\hat{\theta})$, for the estimators, and 95% confidence intervals, $\left(\hat{\theta}_i - 1.96\mathrm{se}(\hat{\theta}_i), \hat{\theta}_i + 1.96\mathrm{se}(\hat{\theta}_i)\right)$, for the regression parameters, are obtained using standard likelihood theory (see details in Appendix A).

4. MODEL SELECTION

4.1. Codification

In this work, we consider the largest parameterization with two sinusoidal harmonics ($P_\mu = 2$, $P_\psi = 2$ and $P_\xi = 2$). Moving from the simplest model $\theta^1 = (\beta_0, \alpha_0)$ (that is, a homogeneous Gumbel distribution) to the most complex one $\theta^b = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$, we have a large variety of models to choose from, many of them having different degrees of freedom. Following the genetic algorithms nomenclature [*Goldberg*, 1989], we adopt a binary codification to represent each model, according to the involved factors. Therefore, every model is encoded using a binary chromosome, $c = [g_1 g_2 . g_3 g_4 . g_5 . g_6 g_7]$, where g_i are binary genes which represent given factors. Each gene g_i has two possible values, $g_i = 1$ if the ith factor is switched on and $g_i = 0$ if it is switched off. Gene g_1 represents the annual cycle (β_1, β_2) for the location parameter, g_2 (β_3, β_4) the semiannual cycle for the location parameter, g_3 (α_1,α_2) the annual cycle for the scale parameter, g_4 (α_3,α_4) the semiannual cycle for the scale parameter, g_5 is a gene that includes a constant non-zero shape parameter (γ_0), $g_6\,(\gamma_1,\gamma_2)$ allow for the annual cycle for the shape parameter and g_7 (γ_3,γ_4) the semiannual cycle for the shape parameter.

For example, the simplest model $\theta^1=(\beta_0,\alpha_0)$ has a binary chromosome $c^1=[00.00.0.00]$ and the saturated model θ^b has a binary chromosome $c^b=[11.11.1.11]$. The model shown previously for Gijon buoy has the vector parameter $\theta=(\beta_0,\beta_1,\beta_2,\alpha_0,\alpha_1,\alpha_2,\gamma_0)$, so that its binary chromosome is $c=[10.10.1.00]$.

4.2.Fitness criteria

The quality of a particular model *i*, with a binary chromosome c^i (and consequently a vector parameter θ^i) is assessed by using a penalized function based on the Akaike information criterion [*Akaike*, 1973],

$$\mathrm{AIC_i}=-2\ell(\hat{\theta}^i\,|\,t_j,z_j)+2p_i, \tag{7}$$

where p_i is the number of parameters, and $l(\hat{\theta}^i\,|\,t_j,z_j)$ is the maximum of the log-likelihood resulting from model "*i*" for the sample $\{t_j,z_j\}$. Equation (7) establishes a compromise between obtaining a good fit, which is measured by how small the resulting $-2l(\hat{\theta}^i\,|\,t_j,z_j)$ term is, and using a simple model, where simpler models use less parameters than complex models. Therefore, the smaller the criterion, the better the model.

4.3. Automatic selection

For the particular codification proposed in this work, we have $2^7 = 128$ models to select the best one. Instead of naive procedures, such as exploring all the possible models, we use a stepwise algorithm that combines forward selection and backward elimination procedures. In addition to the steps performed in the forward selection algorithm, all non-zero genes are tested backward to see if their contributions are significant after a new gene has been switched on. This may lead to the elimination of an already selected gene if its factor has become superfluous because its effects may be represented by other factor. The criterion to incorporate a given factor is based on the AIC statistic given by Equation (7). Table 2 and Figure 5 show the application of this automatic selection procedure to the data set of the Valencia buoy. In Table 2, we show the binary chromosome, maximum likelihood estimates, log-likelihood function ℓ, number of parameters *p* and AIC value. Figure 5 shows the location parameter (solid lines), the scale parameter (dashed lines) and the 20-year return period time-dependent quantile (bold lines). The starting model is $c^1 = [00.00.0.00]$ and the final model after 6 steps is $c^6 = [11.11.1.00]$ (two harmonics for the location and scale parameters and a constant non-zero value for the shape parameter).

For this particular case, the incorporation factor sequence is: (step 2) annual cycle for the scale parameter; (3) annual cycle for the location parameter; (4) semiannual cycle for the scale parameter; (5) semiannual cycle for the location parameter; and (6) a constant non-zero value for the shape parameter. Note how the location and scale parameters as well as the quantile progressively improve the fit. Particularly interesting is the bimodal behavior of monthly maxima in Valencia buoy, suggesting two extreme seasons (Fall and beginning of Spring) throughout the year.

4.4. Model diagnostic

The model-checking of the best model obtained is evaluated graphically by means of quantile-quantile (QQ) and probability-probability (PP) plots. We standardize the maximum Z_t using

$$\overline{Z}_t = \frac{1}{\hat{\xi}(t)} \log\left[1 + \hat{\xi}(t)\left(\frac{Z_t - \hat{\mu}(t)}{\hat{\psi}(t)}\right)\right], \tag{8}$$

so that $\overline{Z}_t$ would follow a standard Gumbel distribution if the model and parameter values were exactly true. Probability and quantile plots for the sample of computed values $\overline{z}_t$ can be obtained using Eq. (8). If $\overline{z}_{(1)},...,\overline{z}_{(m)}$ are the corresponding sample order statistics, the plotting points (e.g., empirical vs model) for the probability plot are $\{i/(m+1), \exp(-\exp(-\overline{z}_{(i)}))\}$ whilst the plotting points for the quantile plot are $\{-\log(-\log(i/(m+1))), \overline{z}_{(i)}\}$ for $i = 1,...,m.$ We can see in Figure 6 that, for the Valencia data set and for the best model $c^6 = [11.11.1.00]$, the PP and QQ plots show very good diagnostics, with points close to the diagonal. Similar plots have been obtained for the remaining locations, although they are no shown for space limitations.

5. INFERENCE FOR RETURN LEVELS

5.1. Time-dependent quantiles

For non-stationary or time-dependent GEV parameters, the calculation of "effective" design value quantiles can be carried out using

$$z_q(t,\theta) = z_q(\mu(t),\psi(t),\xi(t)) = \begin{cases} \mu(t) - \dfrac{\psi(t)}{\xi(t)}\left[1-\left\{-\log(1-q)\right\}^{-\xi(t)}\right] & \xi(t)\neq 0 \\ \mu(t)-\psi(t)\log\left\{-\log(1-q)\right\} & \xi(t)=0 \end{cases}, \tag{9}$$

where probability q is given by $F_t(z)=1-q$ and the quantile estimate $\hat{z}_q(t,\hat{\theta})$ is the time-dependent return level associated with the return period $1/q$. Therefore, the quantity varies depending on the time of the year [*Méndez et al.*, 2007]. Confidence intervals can be obtained by the delta method [*Rice*, 1994], assuming approximate normality for the maximum likelihood estimators.

Figure 7 shows, for the Bilbao data set, the comparison between the stationary model applied month-to-month and the best time-dependent model, both with 20-year return period quantiles and 95% confidence intervals. Note how the confidence intervals are reduced with the time-dependent model and how the point estimates are more consistent with each other. The month-to-month approach may lead to unreliable estimates, as seen in the months of January and February. This is due to the fact that, instead of using a sample of 14 – 19 maximum values for the estimation month-to-month, we are using a sample of 190 values when modeling the seasonal behavior within a year. This results in a better explanation of data variability, a reduction in the uncertainty of the quantile estimates and a better estimation for return values in any month or season, as this methodology uses all the surrounding information throughout the year.

We have applied the methodology to the five buoys, obtaining for each particular case the model that best fits the data ($c^{BI} = c^{GI} = [10.10.1.00]$, $c^{CO} = [10.11.1.00]$, $c^{CA} = [11.11.0.10]$ and $c^{VA} = [11.11.1.00]$). Numerical values of the maximum likelihood estimates of the best model for each buoy are shown in Table 3. Figure 8 shows, for each selected model, the location parameter (solid line), the time-dependent 20-year return period quantile (bold-solid line) and the 95% confidence interval (dashed lines). Observed values of monthly maxima significant wave heights are indicated by crosses. One can see that the fit is remarkably good for all the cases, each of them presenting a different parameterization. To further assess the goodness of fit we have evaluated the proportions of data for the studied MMS falling below $\mu(t)$ or $H_{20}(t)$; these proportions show an averaged absolute deviation of 0.0136 with regard to the theoretical value of 0.37 for $\mu(t)$ and 0.0046 with regard to the theoretical value of 0.95 for $H_{20}(t)$.

The results obtained for Bilbao and Gijon present an annual cycle for the location and the scale parameter and a negative value for the scale parameter (upper bounded tail), clearly pointing at the winter season (from November to March). In the case of La Coruña, a semiannual cycle in the scale parameter is also significant. The western location of La Coruña buoy causes a wider winter season than that seen in the Gijon and

Bilbao buoys, that being more exposed to west winter cyclones. The data set from Cadiz is the only one presenting a significative annual cycle for the shape parameter; resulting in a short tailed behavior in the summer season, due to local storms, and a long tailed one in the winter season (November to February), due to long-fetched storms. The modulation of seasonality showed in the Valencia buoy is a well-known pattern along the Spanish Mediterranean coast. It presents two peaks of maxima significant wave heights, one resulting from '*gota fria'* phenomenon in the fall season (warming of Mediterranean Sea together with cyclones) and another in the spring season. The knowledge of the seasonal distribution of the location, scale and shape parameters allows estimating quantitatively the intrannual variability of extreme wave climate. The location parameter (which coincides with the mode of the distribution) varies in a range of about two meters in the Atlantic buoys (for the case of La Coruña is almost three meters), whereas the range of Valencia buoy, in the Mediterranean Sea, is less than half meter. The scale parameter shows the highest range (almost one meter) in Cadiz buoy and is greatly influenced by the semiannual cycle in La Coruña and Valencia buoys.

5.2. Annual quantiles

The calculation of extreme significant wave height return values for a given period (year, season or month) requires a more complex approach. With time-dependent GEV parameters, determinating a return period involves combining probabilities that differ depending on the day within a year (see chapter 7 in *Coles*, [2001]). A similar procedure for mixed populations was proposed by *Carter and Challenor* [1981] using twelve distributions for monthly maxima, and by *Morton et al.* [1997] using four seasonal POT distributions.

In particular, the annual quantile return level $\overline{z}_q[t_1,t_2]$ corresponding to a given probability $1-q$ and an interval $[t_1,t_2]$, equal or larger than one month, can be obtained by iteratively solving the equation:

$$1-q=\exp\left\{-k_m\int_{t_1}^{t_2}\left[1+\xi(t)\left(\frac{\overline{z}_q[t_1,t_2]-\mu(t)}{\psi(t)}\right)\right]_+^{-1/\xi(t)}dt\right\}, \tag{10}$$

where $1/k_m$ is the length of the block maxima, that is, one month ($1/k_m=1/12$ year).

Figure 9 shows the monthly and annual return level plot for Gijon's best model. This figure is obtained by successively taking the interval $[t_1,t_2]$ in equation (10) equal to each month from January to December and subsequently taking $[t_1,t_2]$ equal to the whole year. The figure shows, for example, that significant wave height attains a value of 8 meters once every 31 years, but only once every 77 Januaries, once every 105 Decembers, and so on. A bounded tail behavior (Weibull family) is detected. Note that the winter months (December, January and February) present the highest monthly quantiles whereas the summer months show the lowest monthly quantiles.

Although not required by our method, comparison with alternative formulations may be facilitated by expressing equation (10) in terms of the rescaled annual maxima GEV parameters $\mu^*(t)$, $\psi^*(t)$, $\xi^*(t)$ [*Dixon and Tawn*, 1994]. Thus it may be easily verified that the right hand side of equation (10) is exactly equal to

$$\exp\left\{-k_y \int_{t_1}^{t_2} \left[1+\xi^*(t)\left(\frac{\overline{z}_q[t_1,t_2]-\mu^*(t)}{\psi^*(t)}\right)\right]_+^{-1/\xi^*(t)} dt\right\},$$

(11)

where $1/k_y = 1$ year, $\mu^*(t) = \mu(t) + \frac{\psi(t)}{\xi(t)}\left[\left(\frac{k_m}{k_y}\right)^{\xi(t)} - 1\right]$, $\psi^*(t) = \psi(t)\cdot\left(\frac{k_m}{k_y}\right)^{\xi(t)}$ and $\xi^*(t) = \xi(t)$.

6. CONCLUSIONS

A statistical model to analyze the seasonality of monthly maxima significant wave heights is presented. The model is based on the time-dependent generalized extreme value distribution for independent monthly maxima series of significant wave heights. Non-stationarity is introduced in the model using cosine functions that represent the annual and semiannual cycles. These factors are included in the location, scale and shape parameters of the probability distribution of extreme significant wave height. The inclusion of seasonal variabilities substantially reduces the residuals of the fitted model. To avoid over-parameterization, an automatic model selection procedure based on the Akaike Information Criterion is carried out.

The developed time-dependent methodology provides more reliable results than the stationary model of individually month-to-month analysis. For a particular month, we are using not only the maximum values from this month but also the information of the neighbour months, thus including the natural climate variability. Therefore, we believe that the large discrepancy in the month-to-month method is due to unduly sampling variation.

The model is applied to five buoys of Puertos del Estado network (Bilbao, Gijon, La Coruña, Cadiz and Valencia) showing seasonality characteristics of the wave climate of these particular sites. Obtained results show that the model provides a tool for quantitatively examining the long-term seasonal distribution, using monthly maxima of significant wave heights. The methodology provides time-dependent and annual return-period values and their confidence intervals. The information obtained in this study can be useful to better understand several issues governed by ocean waves such as distribution of organisms in wave-swept environments, coastal management or the design of maritime works.

The applicability of our methods to various Spanish buoys with different wave climate conditions and the flexibility of the time-dependent GEV distribution (that allows modeling the seasonality and varied upper tail behaviors), reinforces the possibility of using the same methods to analyze other geophysical variables such as sea surface temperature, wind velocity or sea level.

We believe that the model provides a new way to gain further insights in our knowledge of climate variability of extreme events. We have analyzed just seasonal processes but the methodology should be able to deal with different time scales, such as long-term trends or interannual variability (North Atlantic Oscillation, Southern Oscillation, etcetera).

APPENDIX A. Parameter estimation

We use the method of maximum likelihood to estimate the model parameters. The location, scale and shape parameters $\mu(t)$, $\psi(t)$ and $\xi(t)$ are expressed in terms of harmonic functions whose amplitudes are regression parameters that must be estimated [*Coles*, 2001]. The complete vector of p regression parameters is denoted by θ. The likelihood function of the parameters for any given sample $\{(t_1, z_1), ..., (t_m, z_m)\}$ of the periods t_i at which the maxima z_i are attained, is provided by

$$L(\theta \mid t_i, z_i) = \prod_{i=1}^{m} f_t(z_i, \theta) = \prod_{i=1}^{m} f_t(z_i; \mu(t_i), \psi(t_i), \xi(t_i)) \tag{12}$$

where $f_t(z; \mu(t), \psi(t), \xi(t))$ is the time-dependent GEV pdf given by Equation (2). The log-likelihood function is

$$\ell(\theta \mid t_i, z_i) = -\sum_{i=1}^{m} \left\{ \log \psi(t_i) + (1 + 1/\xi(t_i)) \log \left[1 + \xi(t_i) \left(\frac{z_i - \mu(t_i)}{\psi(t_i)} \right) \right]_+ + \left[1 + \xi(t_i) \left(\frac{z_i - \mu(t_i)}{\psi(t_i)} \right) \right]_+^{-1/\xi(t_i)} \right\} \tag{13}$$

provided that $\psi(t_i) > 0$ for $i = 1, ..., m$. For every value of $\xi(t_i)$ that equals zero, the appropriate limiting form must be used, replacing the GEV by the Gumbel (equation (1) for $\xi = 0$) log-likelihood function,

$$\ell(\theta \mid t_j, z_j) = -\log\psi(t_j) - \frac{z_j - \mu(t_j)}{\psi(t_j)} - \exp\left[-\frac{z_j - \mu(t_j)}{\psi(t_j)}\right]. \tag{14}$$

Maximization of (13) and/or (14) yields to $\ell(\hat{\theta} \mid t_j, z_j)$ and the maximum likelihood estimate of θ, denoted by $\hat{\theta}$. A global optimization procedure, namely the shuffled complex evolution (SCE) algorithm [*Duan et al.* 1992] is used to compute the parameter estimates.

An advantage of adopting maximum likelihood for parameter estimation is that a widely applicable approximation for standard errors and confidence intervals is available based on asymptotic properties of maximum likelihood estimators. Solving log-likelihood equations, we can evaluate the observed information matrix at $\theta = \hat{\theta}$. Assessing the inverse of this matrix and then the square roots of the diagonal entries, we obtain approximate values for the asymptotic standard errors of the parameters estimates, abbreviated $\mathrm{se}(\hat{\theta}_i)$. Confidence intervals for θ_i can be obtained in the form $\left[\hat{\theta}_i - z_\alpha \mathrm{se}(\hat{\theta}_i), \hat{\theta}_i + z_\alpha \mathrm{se}(\hat{\theta}_i)\right]$, where $\mathrm{se}(\hat{\theta}_i)$ is the standard error of the ML estimator $\hat{\theta}_i$ and $z_{0.95} = 1.96$ gives approximate confidence intervals of 95%.

Acknowledgements

The authors would like to thank "Puertos del Estado" for providing the information from the instrumental observation networks. The work was partially funded by projects ENE2004-08172 and CGL2005-05365/CLI from the Spanish Ministry of Educación y Ciencia. Melisa Menéndez and Fernando J. Méndez are indebted to the Spanish Ministry of Educación y Ciencia for funding them through the FPI and "Ramón y Cajal" program, respectively. Alberto Luceño acknowledges the support of the Spanish Dirección General de Investigación under Grant MTM2005-00287.

REFERENCES

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov and F. Csáki, eds. Proc. 2nd Internat. Symp. on Information Theory (Akadémia Kiadó, Budapest) pp. 267–281.

Carter, D. J. T., and P. G. Challenor (1981) Estimating return values of environmental variables. *Quart. J. Roy. Meteor. Soc.*, 107, 259-266.

Coles, S. G. (2001) *An introduction to statistical modelling of extreme values.* London: Springer, 208 pp.

Dixon, M. J. and J. A. Tawn (1994) Extreme sea-levels at the U.K. A class-sites: site by site analysis. Proudman Oceanographic Laboratory Internal Document 65, 228 pp.

Duan, Q., S. Sorooshian, and V. Gupta (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, 28, 1015-1031.

Goda, Y. (2000) *Random seas and design of maritime structures*. Advanced Series on Ocean Engineering , Vol. 15. World Scientific. Ed. P. L.-F. Liu., 443 pp.

Goldberg, D.E. (1989) *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, Mass, 372 pp.

Guedes Soares, C., and M. G. Scotto (2004) Application of the r-largest order statistics for long-term predictions of significant wave height. *Coast. Eng.*, 51, 387-394.

Holthuijsen, L.H. (2007) Waves in Oceanic and Coastal Waters, Cambridge University Press. 404 pp.

Katz, R. W., M. B. Parlange, and P. Naveau (2002) Statistics of extremes in hydrology. *Adv. in Water Resour.*, 25, 1287-1304.

Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J. (2006) Estimation of the long-term variability of extreme significant wave height using a time-dependent POT model, *Journal of Geophysical Research*, Vol. 111, C07024, doi: 10.1029/2005JC003344.

Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J. (2007) Analyzing monthly extreme sea levels with a time-dependent GEV model, *Journal of Atmospheric and Oceanic Technology*, 24, 894-911, doi: 10.1175/JTECH2009.1.

Morton, I.D., J. Bowers, and G. Mould (1997) Estimating return period wave heights and wind speeds using a seasonal point process model, Coastal Eng., 31, 305-326.

Panchang, V.G. and D. Li (2006) Large waves in the Gulf of Mexico caused by Hurricane Ivan, *Bulletin of the American Meteorological Society*, 481-489.

Rice, J. (1994) *Mathematical Statistics and Data Analysis*, 2nd ed., Duxbury, Pacific Grove, Calif, 651pp.

	Water depth (m)	Period of measurement	Number of monthly maxima
Bilbao (BI)	50	1985-2003	190
Gijón (GI)	23	1984-2002	199
Coruña (CO)	50	1984-2003	179
Cádiz (CA)	22	1984-2002	174
Valencia (VA)	20	1985-2003	170

Table 1. Descriptive issues of the buoys used in the analysis

Step	1	2	3	4	5	6
Chromosome	[00.00.0.00]	[00.10.0.00]	[10.10.0.00]	[10.11.0.00]	[11.11.0.00]	[11.11.1.00]
β_0 (cm)	157.4 (4.3)	150.8 (3.7)	158.6 (3.4)	156.9 (2.96)	158.6 (4.9)	161.6 (4.9)
β_1 (cm)	-	-	22.8 (4.1)	21.8 (2.9)	18.1 (5.3)	19.0 (5.5)
β_2 (cm)	-	-	0.07 (4.3)	-0.38 (5.4)	1.5 (9.2)	1.2 (6.7)
β_3 (cm)	-	-	-	-	-11.2 (4.1)	-12.1 (5.7)
β_4 (cm)	-	-	-	-	2.8 (7.5)	2.0 (5.9)
α_0 (cm)	53.9 (3.3)	52.6 (1.4)	52.3 (2.2)	51.8 (2.9)	51.6 (3.7)	53.7 (3.7)
α_1 (cm)	-	17.4 (2.6)	23.8 (4.0)	18.4 (3.0)	17.9 (5.9)	18.2 (4.2)
α_2 (cm)	-	3.2 (3.0)	2.7 (3.6)	3.1 (4.8)	3.1 (4.2)	2.6 (4.7)
α_3 (cm)	-	-	-	-8.7 (4.2)	-10.6 (5.2)	-12.4 (4.5)
α_4 (cm)	-	-	-	-3.7 (4.1)	-3.2 (4.3)	-6.0 (4.4)
γ_0	-	-	-	-	-	-0.106 (0.06)
γ_1	-	-	-	-	-	-
γ_2	-	-	-	-	-	-
γ_3	-	-	-	-	-	-
γ_4	-	-	-	-	-	-
ℓ	-948.54	-942.60	-935.18	-932.54	-930.36	-929.21
p	2	4	6	8	10	11
AIC	1901.1	1893.2	1882.4	1881.1	1880.7	1880.4

Table 2. Summary of the results for the stepwise evolution of the Valencia buoy: chromosome, maximum likelihood estimates of the parameters with standard errors (in parentheses), maximum of the log-likelihood function (ℓ), number of parameters involved (p) and Akaike Information Criteria statistic (AIC).

Buoy	BI	GI	CO	CA	VA
Chromosome	[10.10.1.00]	[10.10.1.00]	[10.11.1.00]	[11.11.0.10]	[11.11.1.00]
β_0 (cm)	325.3 (6.9)	310.7 (6.6)	409.8 (9.0)	216.8 (6.4)	161.6 (4.8)
β_1 (cm)	99.5 (8.6)	114.2 (8.3)	142.9 (11.1)	94.4 (10.3)	19.0 (5.5)
β_2 (cm)	13.5 (8.1)	15.4 (7.7)	16.4 (12.2)	11.7 (6.6)	1.2 (6.7)
β_3 (cm)	-	-	-	19.6 (5.8)	-12.1 (5.6)
β_4 (cm)	-	-	-	-16.7 (5.3)	2.0 (5.8)
α_0 (cm)	80.8 (4.9)	79.1 (4.6)	105.2 (6.6)	62.4 (4.7)	53.7 (3.6)
α_1 (cm)	29.4 (6.1)	31.4 (6.3)	24.6 (7.9)	59.0 (7.7)	18.2 (4.2)
α_2 (cm)	9.0 (5.6)	6.7 (4.9)	-0.2 (8.9)	0.9 (5.2)	2.6 (4.6)
α_3 (cm)	-	-	-15.5 (8.1)	15.8 (4.7)	-12.4 (4.5)
α_4 (cm)	-	-	-3.7 (9.0)	-9.1 (4.0)	-6.0 (4.4)
γ_0	-0.13 (0.05)	-0.12 (0.05)	-0.14 (0.05)	-	-0.11 (0.06)
γ_1	-	-	-	-0.17 (0.08)	-
γ_2	-	-	-	0.03 (0.07)	-
γ_3	-	-	-	-	-
γ_4	-	-	-	-	-
p	7	7	9	12	11

Table 3. Summary of the final results for the time-dependent model for the studied buoys: final chromosome, maximum likelihood estimates for the location, scale and shape parameters (with standard errors) and number of involved parameters.

FIGURE CAPTIONS

Figure 1. Boxplots for monthly maxima significant wave height in Bilbao, Gijon, La Coruña, Cadiz and Valencia buoys. Trapezoidal boxes have lines at the lower quartile, median and upper quartile values. The whiskers extend to the 1.5 interquartile range or to the range of the data, whichever is shorter, and crosses show unusual values.

Figure 2. Location of the buoys (BI, GI, CO, CA and VA stands for Bilbao, Gijon, La Coruña, Cadiz and Valencia, respectively)

Figure 3. Time-dependent GEV probability density function for Gijon with the final parameterization.

Figure 4. Scatter plots of annual stationary GEV parameter estimates (each for a given month) along a year for La Coruña buoy. Regression fit to one (grey line) and two (black line) harmonics is also plotted.

Figure 5. Evolution of the stepwise method for Valencia buoy, starting from $c_1 = [00.00.0.00]$ (stationary case) and ending in $c_6 = [11.11.1.00]$ (two harmonics for the location and scale parameters and a constant non-zero value for the shape parameter). Location parameter is indicated in solid lines, scale parameter in dashed lines and 20-year time-dependent return level in bold lines.

Figure 6. Probability (left) and quantile (right) plots for the best model $c^6 = [11.11.1.00]$ for the Valencia data set.

Figure 7. Comparison between the stationary model applied month-to-month (the midpoint are the 20-year return level and the vertical lines are the 95% confidence interval) and the time-dependent model for Bilbao data set (solid line is the 20-year return period quantile and dashed lines 95% confidence interval).

Figure 8. 20-year return period quantiles (bold lines) within a year and 95% confidence intervals (dashed lines). The lower lines show the time-dependent location parameter. Observed values of monthly maxima significant wave height indicated by crosses. Results are for the best model for the five buoys.

Figure 9. Monthly and annual return level plots for Gijon buoy.

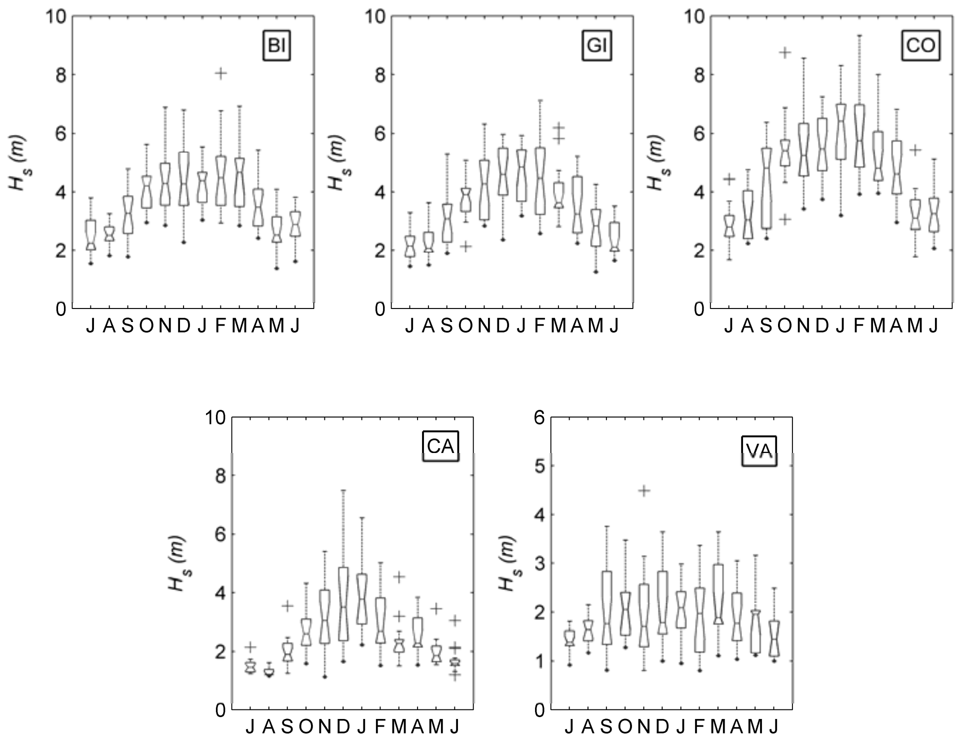

Figure 1. Boxplots for monthly maxima significant wave height in Bilbao, Gijon, La Coruña, Cadiz and Valencia buoys. Trapezoidal boxes have lines at the lower quartile, median and upper quartile values. The whiskers extend to the 1.5 interquartile range or to the range of the data, whichever is shorter, and crosses show unusual values.

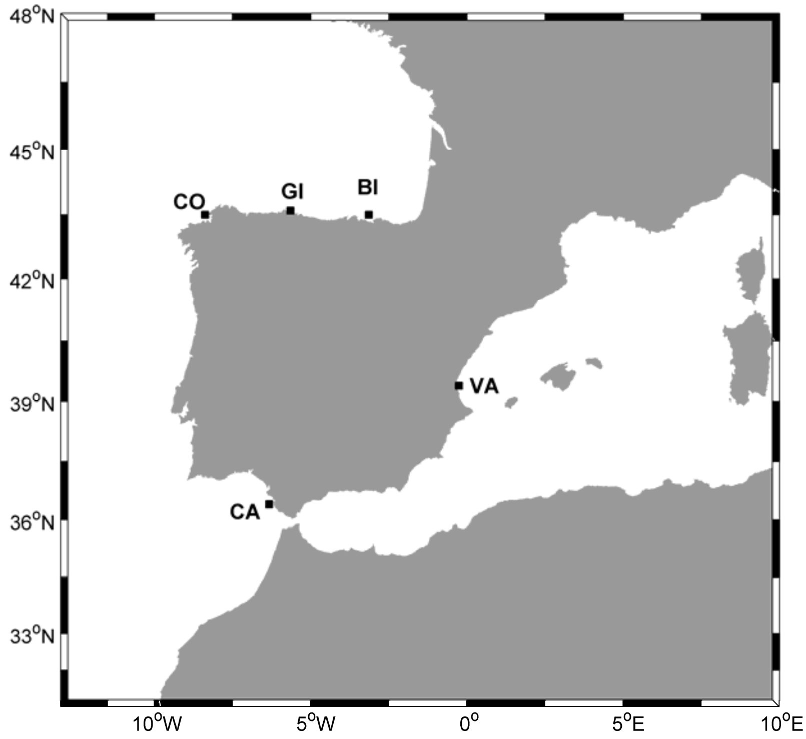

Figure 2. Location of the buoys (BI, GI, CO, CA and VA stands for Bilbao, Gijon, La Coruña, Cadiz and Valencia, respectively)

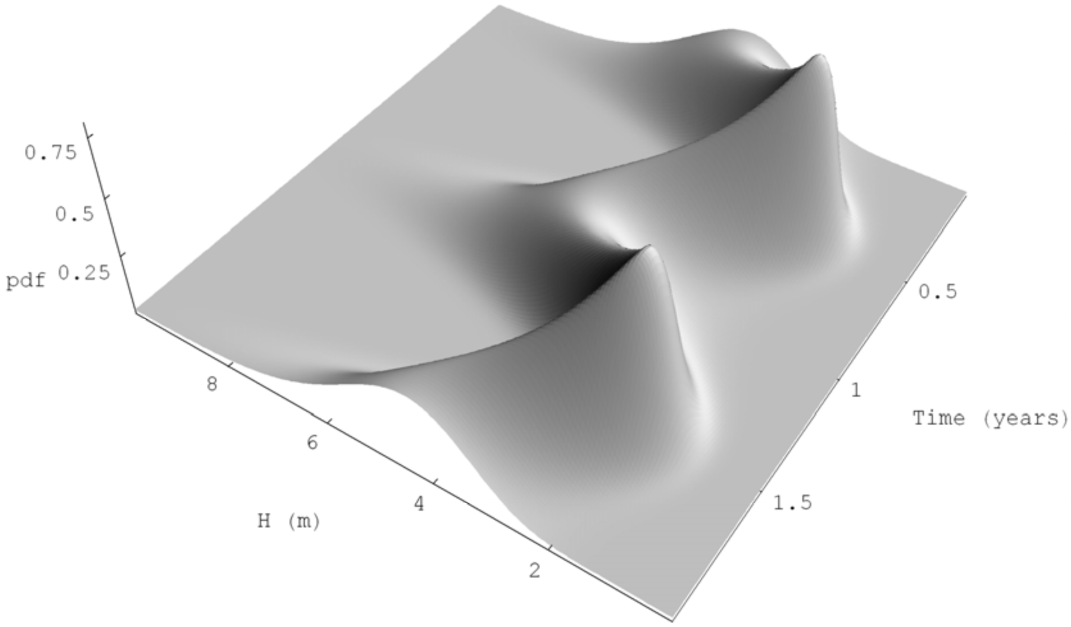

Figure 3. Time-dependent GEV probability density function for Gijon with the final parameterization.

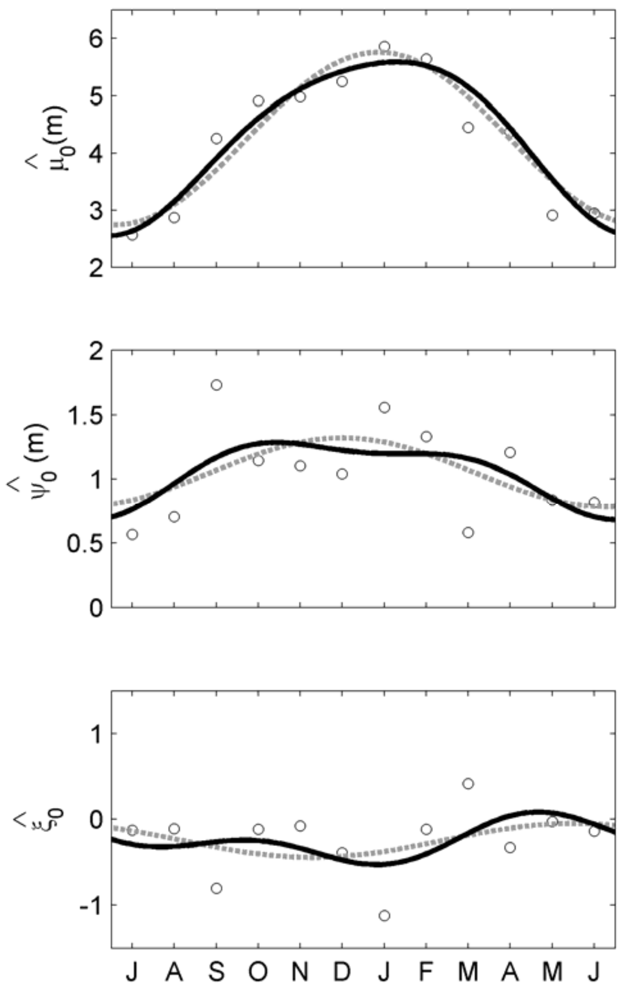

Figure 4. Scatter plots of annual stationary GEV parameter estimates (each for a given month) along a year for La Coruña buoy. Regression fit to one (grey line) and two (black line) harmonics is also plotted.

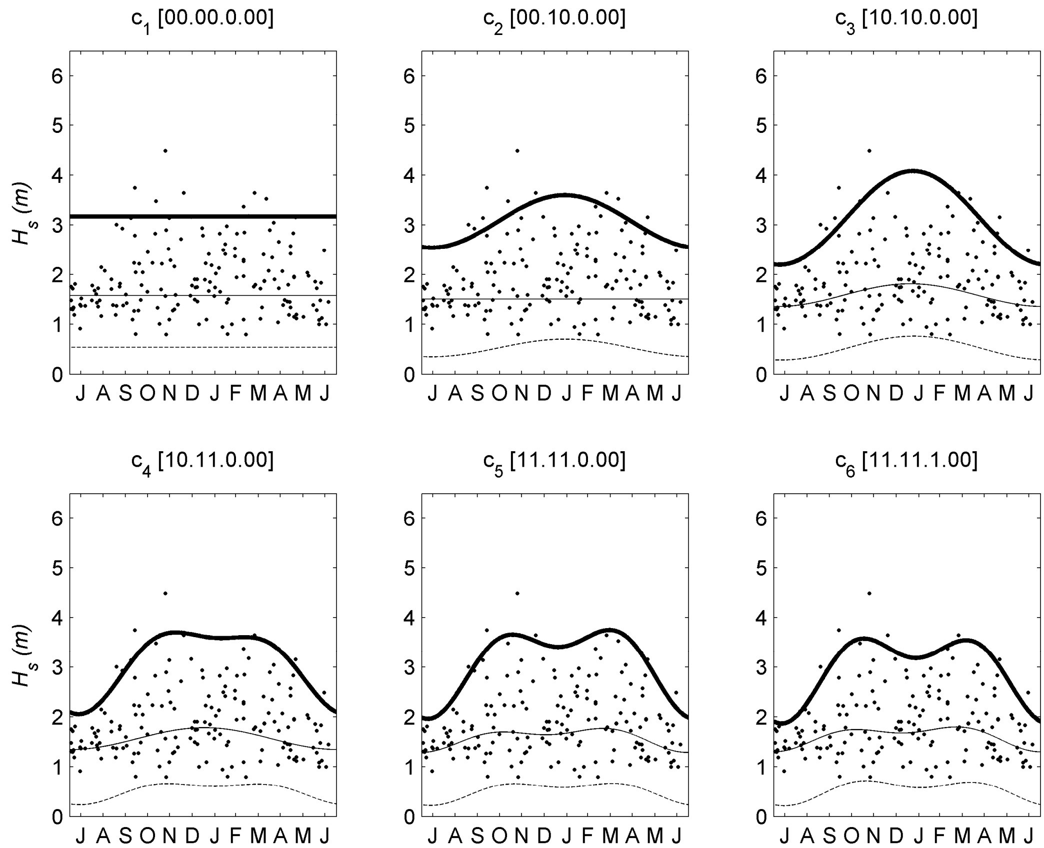

Figure 5. Evolution of the stepwise method for Valencia buoy, starting from $c_1 = [00.00.0.00]$ (stationary case) and ending in $c_6 = [11.11.1.00]$ (two harmonics for the location and scale parameters and a constant non-zero value for the shape parameter). Location parameter is indicated in solid lines, scale parameter in dashed lines and 20-year time-dependent return level in bold lines.

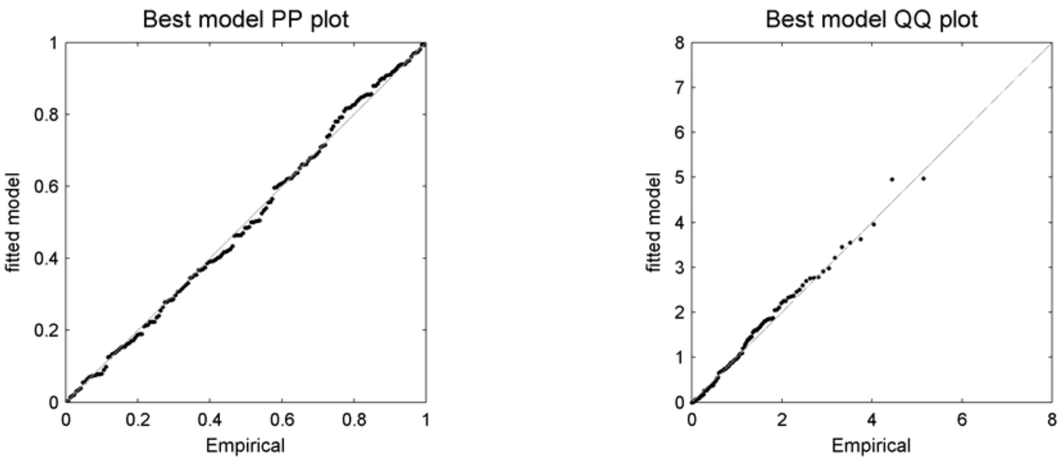

Figure 6. Probability (left) and quantile (right) plots for the best model $c^6 = [11.11.1.00]$ for the Valencia data set.

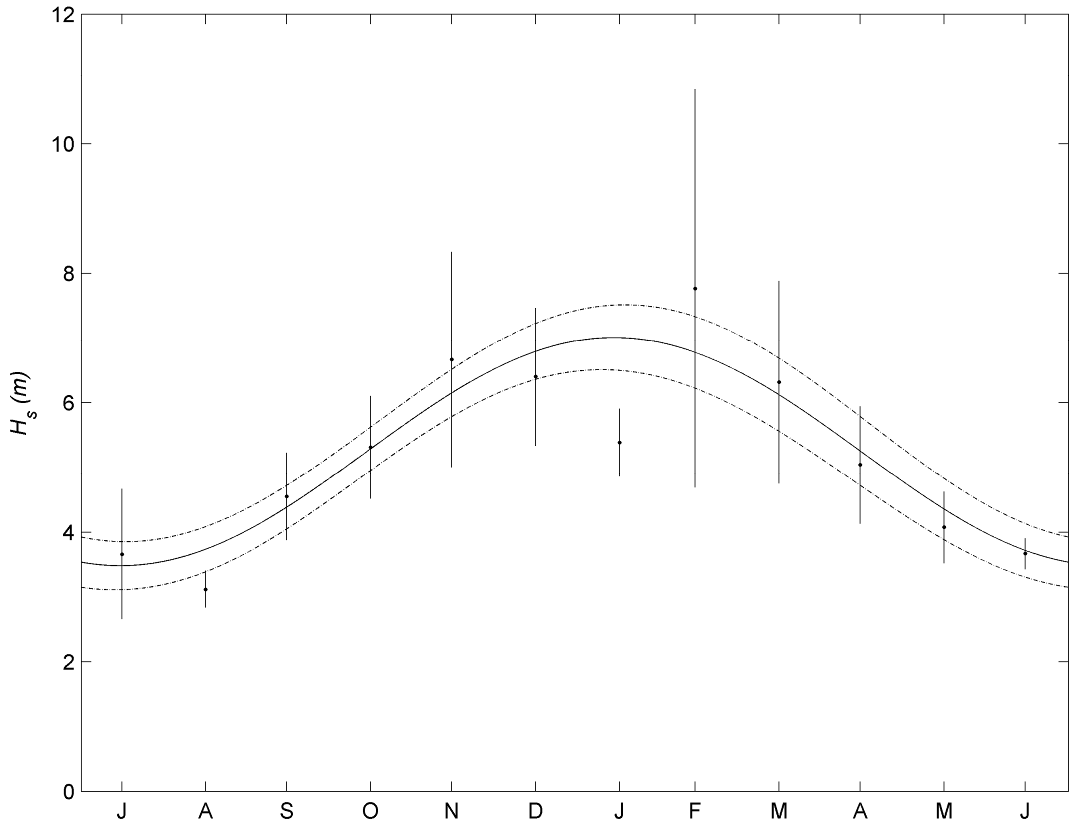

Figure 7. Comparison between the stationary model applied month-to-month (the midpoint are the 20-year return level and the vertical lines are the 95% confidence interval) and the time-dependent model for Bilbao data set (solid line is the 20-year return period quantile and dashed lines 95% confidence interval).

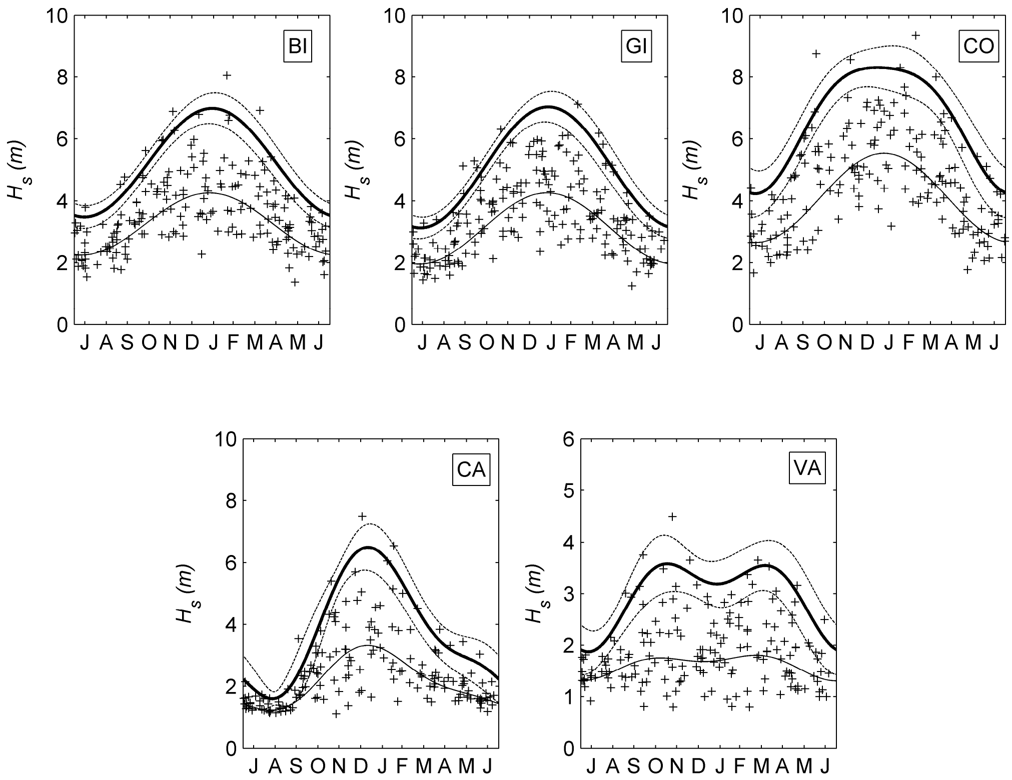

Figure 8. 20-year return period quantiles (bold lines) within a year and 95% confidence intervals (dashed lines). The lower lines show the time-dependent location parameter. Observed values of monthly maxima significant wave height indicated by crosses. Results are for the best model for the five buoys.

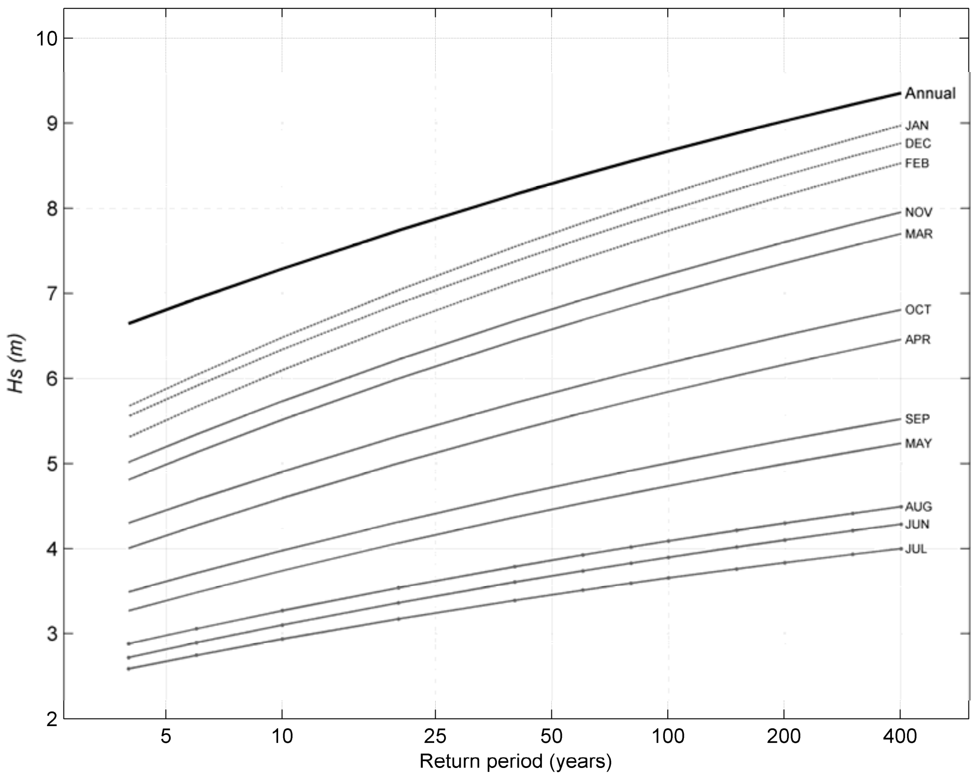

Figure 9. Monthly and annual return level plots for Gijon buoy.