Pseudo-Optimal Parameter Selection of Non-Stationary Generalized Extreme Value Models for Environmental Variables

R. Mínguez [a,*], F. J. Méndez [a], C. Izaguirre [a], M. Menéndez [a], I. J. Losada [a]

[a] *Environmental Hydraulics Institute "IH Cantabria", Universidad de Cantabria, Spain*

Abstract

Recent advances in the description of environmental and geophysical extreme events allow incorporating smooth time variations for the parameters of the GEV distribution using harmonic functions, long-term trends and covariates (North Atlantic Oscillation, El Niño, etc.). Most of the proposed models rely on the maximum likelihood estimation method for a given parameterization. However, finding the best parameter selection for each case is not an easy task, since the number of possible combinations grows exponentially with the number of possible parameters to be considered. This problem is usually overcome by assuming simplified models based on experience or using heuristic approaches, which are computationally very expensive. In this paper, a method to obtain a pseudo-optimal parameterization using the maximum likelihood method is presented. The proposed algorithm automatically selects the parameters which minimize the Akaike Information Criterion within an iterative scheme, including one parameter at a time based on a score perturbation criteria. The process is repeated until no further improvement in the objective function is achieved. The proposed method is applied for the adjustment of monthly maximum significant wave height at different locations around the Atlantic coast and results are compared with those obtained using an existing heuristic approach, showing an important reduction in computational time and comparable results in terms of fitting quality.

Key words: Automatic adjustment, Geophysical variables, Generalized extreme value, Likelihood, Non-stationary, Return period, Time-dependent

* Corresponding author. Tel.: +34 942 201852.
Email addresses: minguezsr@unican.es (R. Mínguez), mendezf@unican.es (F. J. Méndez), izaguirrec@unican.es (C. Izaguirre), menendezm@unican.es (M. Menéndez), losadai@unican.es (I. J. Losada).

1 Introduction

Recent advances in the extreme value theory (see Coles (2001); Katz et al. (2002) as general references) allow modeling the natural variability of extreme events of environmental and geophysical variables. These methods introduce time-dependent variations within a certain time scale (year, season or month), improving our knowledge on some important processes which are time dependent. Additionally, a key issue is the possibility to construct regression models to show how the variables of interest may depend on other measured covariates, for example, how maximum significant wave heights for a particular location depend on the North Atlantic Oscillation (NAO).

Examples of these kinds of models can be found in Carter and Challenor (1981), which proposes a month-to-month distribution assuming that data are identically distributed within a given month. Smith and Shively (1995) constructed a regression model for the frequency of high-level ozone exceedances in which time and meteorology are regressors. Morton et al. (1997) apply a seasonal Peak Over Threshold (POT) model to wind and significant wave height data. Analogous models but applied to different geophysical variables can be found in Coles (2001), Katz et al. (2002) or Méndez et al. (2007). Méndez et al. (2006) developed a time-dependent POT model for extreme significant wave height which considers the parameters of the distribution as functions of time (harmonics within a year, exponential long-term trend, the Southern Oscillation Index (SOI), etcetera). Brown et al. (2008) studied the global changes in extreme daily temperature since 1950 considering possible trends and the influence of the North Atlantic Oscillation (NAO). Menéndez et al. (2009a) and Izaguirre et al. (2010) developed a time-dependent model based on the GEV distribution that accounts for seasonality and interannual variablility of extreme monthly significant wave height. The non-stationary behavior is parameterized using functions of time (harmonic functions and covariates) for the parameters of the distribution.

All these models try to reproduce the behavior of environmental or geophysical extreme value variables using sophisticated parameterizations, whose parameter estimation is not an easy task because i) prior selection of significant harmonic functions and covariates is not evident and the number of possible models can be very large $(2^{n_p} - 1)$, which makes exploring all possible combinations only feasible for small values of n_p (number of parameters), ii) once the parameters are chosen, since its number can be high, appropriate estimation methods not included in standard statistical packages are required. The first problem is overcome assuming simplified models based on experience, which are tested through goodness of fit methods, or using heuristic approaches. For instance, Menéndez et al. (2009a) considers the largest parameterization possible for including non-stationary effects using sinusoidal harmonics into

a GEV distribution. The optimal parameterization is automatically selected minimizing the Akaike Information Criterion (AIC)(Akaike, 1973), the optimization problem is solved using a stepwise algorithm that combines forward selection and backward elimination procedures. The second problem is solved using nonlinear optimization subroutines for unbounded or bounded problems, which allows obtaining the maximum likelihood estimates for complex models efficiently. A similar approach is used in Menéndez et al. (2009b) where monthly maxima extreme sea levels are forecasted based on a time-dependent GEV model.

The problem of selecting the best regression model for large scale problems has been widely studied in the literature, but mostly related to linear regression techniques. Since the leaps-and-bounds algorithm (Furnival and Wilson, 1974) and forward, backward and stepwise selection techniques started by Hocking (1976), different methods have been proposed, such as, the dropping column algorithm (DCA) (Clarke, 1981), ridge regression, the non-negative garrote and the lasso (Breiman (1995), Fan and Li (2001) and Tibshirani (1996)), sequential replacement algorithms (Hastie et al., 2001), branch-and-bound techniques (Narendra and Fukunaga (1997), Roberts (1984), Somol et al. (2004), Gatu and Kontoghiorghes (2006), and Hofmann and Kontoghiorghes (2007)), and heuristic approaches (Hofmann and Kontoghiorghes, 2007). However, besides the method proposed by Menéndez et al. (2009a), we have not found in the literature methods for parameter selection related to the Generalized Extreme Value (GEV) distribution.

In this paper, a new method for automatically selecting the best parameterization and the corresponding optimal parameter estimates based on the Generalized Extreme Value (GEV) distribution is proposed. The method minimizes the Akaike Information Criterion (AIC), which establishes a compromise between obtaining a good fit and using a simple model. The main advantage of the proposed method is that it converges monotonically to the final solution incorporating one single parameter at a time. The parameter at each iteration is included based on sensitivity analysis and/or perturbation techniques from the last solution obtained. This procedure reduces drastically the number of different parameterizations checked out and it has proved to reduce considerably computational times, providing analogous results as the alternative method. Note that proposing well-founded theoretical results on how to choose the number of parameters to be included in the model is out of the scope of the paper.

The rest of the paper is structured as follows. Section 2 gives an introduction to the regression model based on the GEV distribution. In Section 3 the proposed method is explained in detail. In Section 4 its functioning is illustrated by its application to the study of the significant wave height maxima in different locations around the Atlantic coast, comparing the results with the other

existing approach. Section 5 offers some concluding remarks.

2 Regression Model based on the Time-Dependent GEV Distribution

Time-dependent methods within a certain time scale, use time series of block maxima for successive periods (x_t), which are called maxima series (MMS), where x_t is the selected maxima for a given period (e.g., month) t.

To account for this temporal dependence, Menéndez et al. (2009a) use an extension of the standard models of extreme value theory for non-stationary variables proposed by Coles (2001), with a more complex parameterization. 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). Monthly maximum X_t of the climate variable observed in month t follows a GEV distribution with time-dependent location parameter μ_t, scale parameter ψ_t, and shape parameter ξ_t, with probability density function (PDF) given by:

$$g(x_t;\mu_t,\psi_t,\xi_t)=\begin{cases}\dfrac{\exp\left\{-\left[\xi_t\dfrac{(x_t-\mu_t)}{\psi_t}+1\right]_+^{-1/\xi_t}\right\}}{\psi_t\left[\xi_t\dfrac{(x_t-\mu_t)}{\psi_t}+1\right]_+^{1+\frac{1}{\xi_t}}};\ \xi_t\neq 0,\\[2ex] \dfrac{\exp\left\{\dfrac{\mu_t-x_t}{\psi_t}-\exp\left(\dfrac{\mu_t-x_t}{\psi_t}\right)\right\}}{\psi_t};\ \xi_t=0,\end{cases} \tag{1}$$

where $[a]_+=\max(0,a)$, and the support is $x_t\leq\mu_t-\psi_t/\xi_t$ if $\xi_t<0$ (Weibull), $x_t\geq\mu_t-\psi_t/\xi_t$ if $\xi_t>0$ (Fréchet), or $-\infty<x_t<\infty$ if $\xi_t=0$ (Gumbel). Note that we consider X_t as the random variable associated with the maximum at time t, and x_t a particular instance, value or data of the corresponding random variable.

The GEV distribution includes three distribution families corresponding to the different types of the tail behavior: Gumbel family, the case $\xi_t=0$; Fréchet distribution, with $\xi_t>0$; and Weibull family, with $\xi_t<0$ and a bounded tail.

To introduce seasonality, possible long-term trends and the influence of different covariates, the model proposed by Menéndez et al. (2009a) is extended as

follows:

$$\mu_t = \beta_0 + \sum_{i=1}^{P_\mu} \left[\beta_{2i-1}\cos(iwt) + \beta_{2i}\sin(iwt)\right] + \beta_{\mathrm{LT}} t + \sum_{k=1}^{Q_\mu} \beta_k^{\mathrm{co}} n_{k,t} \quad (2)$$

$$\log(\psi_t) = \alpha_0 + \sum_{i=1}^{P_\psi} \left[\alpha_{2i-1}\cos(iwt) + \alpha_{2i}\sin(iwt)\right] + \alpha_{\mathrm{LT}} t + \sum_{k=1}^{Q_\psi} \alpha_k^{\mathrm{co}} n_{k,t} \quad (3)$$

$$\xi_t = \gamma_0 + \sum_{i=1}^{P_\xi} \left[\gamma_{2i-1}\cos(iwt) + \gamma_{2i}\sin(iwt)\right], \quad (4)$$

where t is given in years, $\log(\psi_t)$ ensures positiveness of the scale parameter ($\psi_t > 0$), β_0, α_0, and γ_0 are mean values, β_i, α_i, and γ_i are the amplitudes of harmonics considered in the model, $w = 2\pi/T$ is the angular frequency, T is one year, and P_μ, P_ψ and P_ξ are the number of sinusoidal harmonics to be considered within the year, associated with the location, scale and shape parameters, respectively. Note that it is possible to consider the effects of long-term trends and covariate influences for both location and scale parameters, through the coefficients β_{LT}, α_{LT}, β_k^{co}; $k = 1, \ldots, Q_\mu$ and α_k^{co}; $k = 1, \ldots, Q_\psi$, where Q_μ and Q_ψ are the number of covariates considered (SOI, NAO, monthly mean sea level pressure principal components, etc.) for location and scale parameters, respectively. $n_{k,t}$ is the value of covariate k at time t. We assume that long term and covariate components related to the shape parameter are negligible, nevertheless, it could be incorporated easily in the methodology. Note that the results are limited to the use of the proposed model (2)-(4), it is however complex enough to account for seasonality, covariates and long term trends.

For any model including a certain number of model parameters P_μ, P_ψ, P_ξ, Q_μ, and Q_ψ represented by the following parameter vector

$$\boldsymbol{\theta} = (\beta_0, \beta_i, \beta_{\mathrm{LT}}, \beta_k^{\mathrm{co}}, \alpha_0, \alpha_i, \alpha_{\mathrm{LT}}, \alpha_k^{\mathrm{co}}, \gamma_0, \gamma_i), \quad (5)$$

and for n_d observations of monthly maxima x_t occurring at time t, model parameters are estimated using the method of maximum likelihood (see details in Section 3.3). We advocate this approach for the following reasons:

(1) The maximum likelihood estimation is an optimization problem, and currently available optimization software is versatile, efficient and robust. Details about particular optimization methods used are given in Section 3.
(2) The iterative parameter selection is based on sensitivity analysis or perturbation techniques, which requires first order derivative evaluation of the objective function. This can be easily done for the log-likelihood function, but it is more difficult to accomplish using Bayesian methods.

Note that the estimation procedure starts once the vector of parameters (5) is selected, however the selection of which parameters are going to be incorporated in order to obtain the best model parameterization possible is not an easy task. In the following Section, the proposed method is explained in detail.

Time dependent return level quantiles $q_{t,T}$ associated with the return period T (in years) are calculated using:

$$q_{t,T} = \begin{cases} \mu_t - \frac{\psi_t}{\xi_t}\left[1 - \left(-\log\left(1 - \frac{1}{T}\right)\right)^{-\xi_t}\right], \text{ if } \xi_t \neq 0, \\ \\ \mu_t - \psi_t \log\left(-\log\left(1 - \frac{1}{T}\right)\right), \qquad \text{if } \xi_t = 0. \end{cases} \tag{6}$$

Confidence intervals are obtained, assuming approximate normality for the maximum likelihood estimators, using the delta method (Rice, 1994).

3 Automatic Regression Model Selection Based on Sensitivity Analysis

Parameterizations like (2)-(4) allow constructing complex models that better capture the characteristics of the extreme tail behavior related to environmental and geophysical variables. However, the increasing number of parameters (harmonics, trends or covariates) makes it more difficult to select the subset providing the best model.

Based on Menéndez et al. (2009a), the quality of the model is determined using the Akaike Information Criterion (AIC, Akaike (1973)), which establishes a compromise between obtaining a good fit and keeping the model as simple as possible. The best model is, therefore, selected based on minimizing the following objective function:

$$\underset{\boldsymbol{\theta}}{\text{Minimize}}\ \text{AIC} = -2\ell^*(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\theta}) + 2n_p \tag{7}$$

where $\ell^*(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\theta})$ is the maximum log-likelihood optimal objective function for given parameters $\boldsymbol{\theta}$, and n_p is the number of parameters (harmonics, trends and/or covariates) included in $\boldsymbol{\theta}$. The term $-2\ell^*(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\theta})$ measures the goodness of fit, while the number of parameters n_p favors the simplicity of the model. The proposed method remains valid if alternative criteria, such as

the ϕ-criterion (Hannan and Quinn (1979)), the Bayesian information criterion (Schwarz (1978)), George and Foster (1994) criterion or Hurvich and Tsai (1989) criterion are considered. Each of these criteria assess the quality of each model by making a compromise between goodness of fit and simplicity, so that the smaller the criterion, the better the model.

Note that problem (7) is difficult to solve, being a mixed-integer nonlinear programming (MINLP) problem, where the binary variables are related to the inclusion or not of certain parameters in the model, and the nonlinearity is associated with the estimation problem procedure based on the maximum likelihood method. An alternative to solving this problem is exploring all the possible models, which is only feasible for small values of n_p. Several methods based on branch and cut techniques or heuristic approaches have been proposed in the linear regression literature. Menéndez et al. (2009a) propose a chromosome codification for the optimal selection of harmonics in (2)-(4) and use a stepwise algorithm that combines forward selection and backward elimination procedures. Computational time increases exponentially with the number of possible parameters to be included, which is more relevant if trends and covariates are incorporated. Note that in Menéndez et al. (2009a) approach, the maximum number of possible harmonics to be included must be fixed beforehand.

To improve the automatic model selection, we propose an iterative method based on the following assumptions:

(1) The process starts from the simplest model possible, which corresponds to the traditional stationary model whose parameter vector corresponds to $\boldsymbol{\theta} = (\beta_0, \alpha_0, \gamma_0)$.
(2) The method does not require establishing a priori the maximum number of possible parameters P_μ, P_ψ and P_ξ to be considered, it automatically includes new parameters as long as the Akaike criterion decreases.
(3) Based on sensitivity analysis information, the number of parameters is increased by one unit at each iteration, except for the harmonic case, which includes two new parameters (see equation 2).
(4) It is recognized in the literature that the shape parameter has longer time scale dependent variations than the location and scale parameters, and for this reason, the shape parameter should not vary without allowing any of the other parameters also to vary.
(5) For model selection, the order of parameter inclusion is i) harmonics $(\beta_i, \alpha_j, \gamma_k)$, that account for the non-stationary behavior (intra-annual variability), ii) covariates $(\beta_i^{\text{co}}, \alpha_j^{\text{co}}$, for inter-annual variability), and iii) tendencies $(\beta_{\text{LT}}, \alpha_{\text{LT}}$, for long-term trends). Note that the order criterion is based on physical considerations, from lower to higher time-scale effects. Numerical tests have confirmed the validity of these assumptions.
(6) The inclusion of a new parameter requires a new parameter estimation

process, which results in an optimal log-likelihood function greater than or equal to the previous value:

$$\ell^*(\boldsymbol{x}, \boldsymbol{t}; \hat{\boldsymbol{\theta}}) \geq \ell^*(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\theta}), \tag{8}$$

where $\hat{\boldsymbol{\theta}}$ includes the new parameter.

(7) Harmonics are included in the model in increasing order for location, scale and shape parameters respectively, i.e. once a new harmonic is introduced it is not removed even though new harmonics are included afterwards.

The main purpose of the methodology is to use a pseudo-steepest descent algorithm, which at every iteration and based on sensitivity analysis (first-order derivative) information, selects the best parameter to be introduce in the model that maximizes the increment in the log-likelihood function, which in this case is the parameter whose perturbation in the log-likelihood function is maximum. The algorithm continues including new parameters until no further improvement in the AIC is achieved.

Note that some could argue about the criterion used to select the best parameterization, claiming that it is over-specified, but we believe that it is not pertinent to a paper describing a fitting code. Besides, at the end of the process, standard errors of the parameter estimates could be used, and if the t-statistics of some of the parameters were not significant, this would suggest that those terms should be omitted from the model.

3.1 Maximum perturbation selection criteria

Let us consider a parameter vector $\boldsymbol{\theta}$. The parameter estimates are obtained by maximizing the log-likelihood function, which is an unconstrained non-linear optimization problem. At its optimal solution $(\boldsymbol{\theta}^*, \ell^*)$ (see Bazaraa et al. (1993) or Luenberger (1984)), the Karush-Kuhn-Tucker (KKT) first order optimality conditions reduce to:

$$\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}^*) = \mathbf{0}, \tag{9}$$

which establishes that the gradient of the objective function with respect to $\boldsymbol{\theta}$ at the optimal solution $\boldsymbol{\theta}^*$ must be zero.

Note that using the chain rule, the derivatives of the log-likelihood function with respect to parameters associated with the location parameter (μ_t), i.e. $\Omega_\mu \in \{\beta_0, \beta_j;\ j = 1, \ldots, P_\mu, \beta_{\mathrm{LT}}, \beta_k^{\mathrm{co}};\ k = 1, \ldots, Q_\mu\}$, are (see Appendix A.1

for a complete derivation):

$$\frac{\partial \ell}{\partial \Omega_\mu} = \sum_{\forall t} \frac{\partial \ell}{\partial \mu_t} \frac{\partial \mu_t}{\partial \Omega_\mu}, \tag{10}$$

where

$$\frac{\partial \mu_t}{\partial \Omega_\mu} = \begin{cases} 1 & \text{if } \Omega_\mu = \{\beta_0\} \\ t & \text{if } \Omega_\mu = \{\beta_{\text{LT}}\} \\ n_{k,t} & \text{if } \Omega_\mu = \{\beta_k^{\text{co}}\} \\ \cos\left(\frac{j+1}{2} wt\right) & \text{if } \Omega_\mu = \{\beta_j\} \text{ and } j \text{ odd} \\ \sin\left(\frac{j}{2} wt\right) & \text{if } \Omega_\mu = \{\beta_j\} \text{ and } j \text{ even.} \end{cases} \tag{11}$$

Expression (10) would allow us to evaluate the perturbation produced by a possible additional parameter set ($\hat{\boldsymbol{\theta}}$) because expressions in (11) do not depend on the new parameter values but on data related to time, covariates or harmonics. Note that (10) is the derivative of the log-likelihood function with respect to parameters, which is also known as score $U(\Omega_\mu)$. Analogous results are obtained for the scale and shape parameters. In addition and using expressions provided in Appendix A.2 the hessian of the log-likelihood function can be calculated.

For this reason once the optimal solution of the log-likelihood function is known, and considering different sets of possible parameters to be included in the model $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_\ell$, it is possible to obtain which additional parameter would produce more perturbation (maximum decrement in Akaike Information Criterion) if it is included in the model, which corresponds to the one with maximum score test statistic absolute value (Cox and Hinkley, 1974):

$$|S(\hat{\boldsymbol{\theta}}_i)| = |U(\hat{\boldsymbol{\theta}}_i)^T I^{-1}(\hat{\boldsymbol{\theta}}_i) U(\hat{\boldsymbol{\theta}}_i)|, \tag{12}$$

where $U(\hat{\boldsymbol{\theta}}_i)$ is the score associated with parameter set $\boldsymbol{\theta}_i$, and I is the observed Fisher information matrix, i.e. the negative of the log-likelihood Hessian matrix. This is considered the most influential parameter and it is the one chosen for the next iteration (see Theorem 1 in Appendix B). We use the score test statistic as selection criterion for the following reasons:

(1) It allows comparing the influence of different sets with multiple parameters, and even if the sum of scores is small detects if yet the increase in the log-likelihood is significant and the set worth including.
(2) In order to compare the different sets of parameters, they should be divided by the square root of their variance so that, for each parameters

set, its square has expectation 1 under the null hypothesis. The score test statistic has this property, i.e. is properly scaled.

For example, let us consider that our last optimal fitted model ($\boldsymbol{\theta}^*$) has i harmonics for the location, j for the scale and k for the shape parameters, and we wish to know the influence of an additional harmonic in the log-likelihood function, i.e. the inclusion of the following plausible terms in (2), (3) and (4), respectively:

$$\beta_{2(i+1)-1}\cos((i+1)wt) + \beta_{2(i+1)}\sin((i+1)wt), \tag{13}$$
$$\alpha_{2(j+1)-1}\cos((j+1)wt) + \alpha_{2(j+1)}\sin((j+1)wt), \tag{14}$$
$$\gamma_{2(k+1)-1}\cos((k+1)wt) + \gamma_{2(k+1)}\sin((k+1)wt), \tag{15}$$

which correspond to the harmonics i+1, j+1 and k+1 associated with location, scale and shape parameters. Parameter set values $\hat{\boldsymbol{\theta}}_i = \{\beta_{2(i+1)-1}, \beta_{2(i+1)}\}$, $\hat{\boldsymbol{\theta}}_j = \{\alpha_{2(j+1)-1}, \alpha_{2(j+1)}\}$, $\hat{\boldsymbol{\theta}}_k = \{\gamma_{2(k+1)-1}, \gamma_{2(k+1)}\}$ are unknown, because we have not fitted the model including these new parameters. However, using (10) and the expressions for the hessian given in Appendix A.2, the scores and the observed Fisher information matrix assuming that $\beta_{2(i+1)-1} = 0$, $\beta_{2(i+1)} = 0$, $\alpha_{2(j+1)-1} = 0$, $\alpha_{2(j+1)} = 0$, $\gamma_{2(k+1)-1} = 0$, and $\gamma_{2(k+1)} = 0$ can be calculated. The harmonic with maximum absolute value of the score test statistic:

$$\max\left(|U(\hat{\boldsymbol{\theta}}_i)^T I^{-1}(\hat{\boldsymbol{\theta}}_i)U(\hat{\boldsymbol{\theta}}_i)|; |U(\hat{\boldsymbol{\theta}}_j)^T I^{-1}(\hat{\boldsymbol{\theta}}_j)U(\hat{\boldsymbol{\theta}}_j)|; |U(\hat{\boldsymbol{\theta}}_k)^T I^{-1}(\hat{\boldsymbol{\theta}}_k)U(\hat{\boldsymbol{\theta}}_k)|\right) \tag{16}$$

is the next harmonic to be considered in the model. This is the one which can potentially produce a greater change in the log-likelihood function and thereby in the Akaike Information Criterion.

This method uses local derivatives information and is valid for small perturbations on parameters, i.e. if the resulting parameters $\boldsymbol{\theta}^*$ after fitting do not change significantly. This method allows us to make the decision about which parameters should be incorporated without the need to solve the fitting process for all possible alternatives. Note that there is no mathematical proof that the resulting parameter selection is a global optimum. However, computational tests have confirmed the effectiveness of the method to provide the same selection as the alternative, but more time consuming method, proposed by Menéndez et al. (2009a).

The process is analogous for the covariates, where the score test statistic (12) becomes:

$$S(\hat{\theta}_i) = \frac{U^2(\hat{\theta}_i)}{I(\hat{\theta}_i)}, \tag{17}$$

where $\hat{\theta}_i$ corresponds to a single parameter, and thus the covariate with higher value of the expressions:

$$\underset{i=1,\ldots,Q_\mu,\, j=1,\ldots,Q_\psi}{\text{maximum}} \left(\left| \frac{U^2(\hat{\beta}_i^{\text{co}})}{I(\hat{\beta}_i^{\text{co}})} \right| ; \left| \frac{U^2(\hat{\alpha}_j^{\text{co}})}{I(\hat{\alpha}_j^{\text{co}})} \right| \right), \tag{18}$$

is the next covariate to be considered in the model. This is the one which can potentially produce a greater change in the log-likelihood function.

3.2 Algorithm for automatic model selection

The proposed pseudo-optimal parameter selection algorithm shown in Figure 1 works as follows:

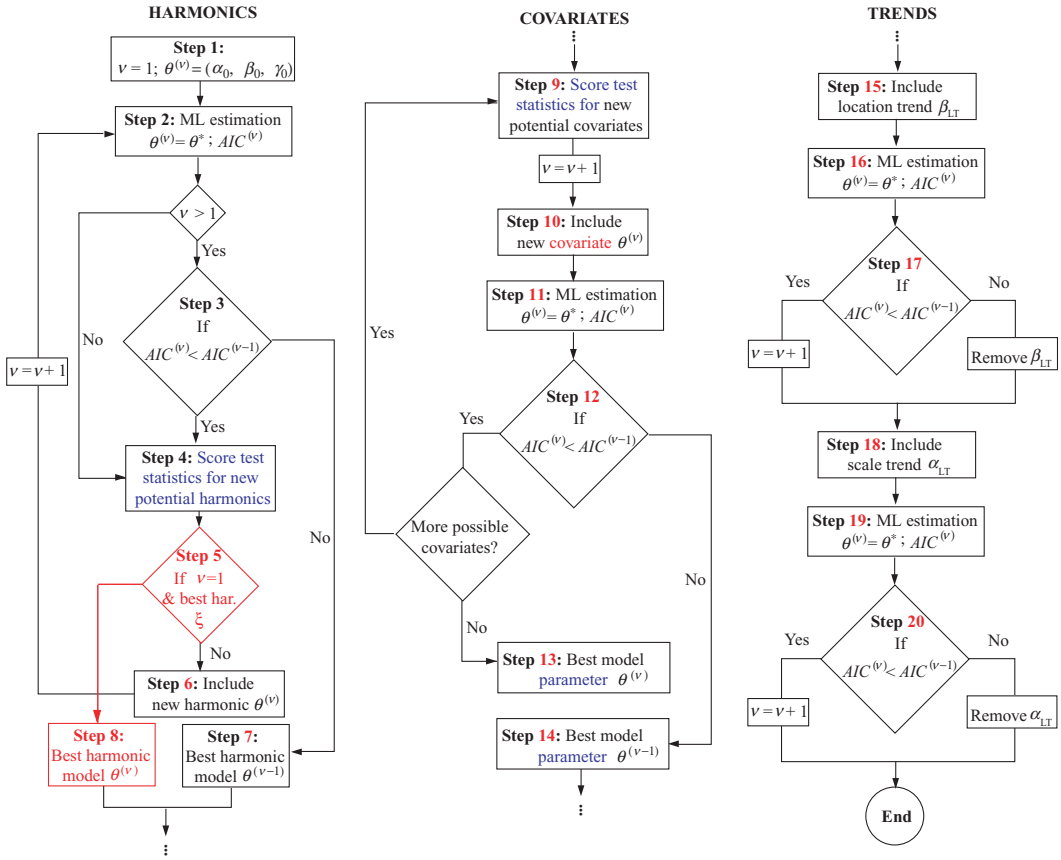

Fig. 1. Structure of the proposed pseudo-optimal parameter selection framework.

Algorithm 1 (Automatic model selection).

Input: *Data* ($\boldsymbol{x}$,$\boldsymbol{t}$,$\boldsymbol{n}$), *which correspond to maximum data, time and covariates.*

Output: *Vector of selected parameters which minimize (7), and their optimal values* $\boldsymbol{\theta}^*$.

- **Step 1: Harmonics.** *Set the iteration counter to* $\nu = 1$, *for this iteration the simplest model is selected* $\boldsymbol{\theta}^{(\nu)} = (\beta_0, \alpha_0, \gamma_0)$.
- **Step 2:** *Obtain the maximum-likelihood estimators for the selected parameters, i.e.* $\boldsymbol{\theta}^{(\nu)} = \boldsymbol{\theta}^*$, *and calculate the AIC objective function* $AIC^{(\nu)}$ *from (7). If* $\nu > 1$ *go to Step 3, otherwise continue in Step 4.*
- **Step 3:** *If* $AIC^{(\nu)} < AIC^{(\nu-1)}$, *go to Step 4, otherwise go to Step 7.*
- **Step 4:** *Calculate the score test statistics for possible additional harmonics related to the location, scale and shape parameters using (12) and expressions given in Appendix A.*
- **Step 5:** *If* $\nu = 1$ *and the maximum score test statistic absolute value is associated with the shape parameter* ξ *go to Step 8, otherwise go to Step 6.*
- **Step 6:** *Update the iteration counter* $\nu \rightarrow \nu + 1$, *include in the parameter vector* $\boldsymbol{\theta}^{(\nu)}$ *the harmonic, i.e. the parameters* $(\beta_{2(\nu+1)+1}, \beta_{2\nu})$, $(\alpha_{2(\nu+1)+1}, \alpha_{2\nu})$ *or* $(\gamma_{2(\nu+1)+1}, \gamma_{2\nu})$, *corresponding to the maximum value of (16) and continue in Step 2.*
- **Step 7:** *End of the harmonic checking, the optimal model parameter vector so far is* $\boldsymbol{\theta}^{(\nu-1)}$. *Go to Step 9.*
- **Step 8:** *End of the harmonic checking, the optimal model parameter vector corresponds to the stationary model, i.e.* $\boldsymbol{\theta}^{(1)}$. *Continue in Step 9.*
- **Step 9:** *Covariates. Calculate the score test statistics for possible additional covariates related to the location and scale parameters. Make use of (17) and expressions given in Appendix A.*
- **Step 10:** *Update the iteration counter* $\nu \rightarrow \nu + 1$, *and include in the parameter vector* $\boldsymbol{\theta}^{(\nu)}$ *the covariate corresponding to the maximum value of* $\left(\left| \frac{U^2(\hat{\beta}_i^{\text{co}})}{I(\hat{\beta}_i^{\text{co}})} \right|;\ \forall i, \left| \frac{U^2(\hat{\alpha}_j^{\text{co}})}{I(\hat{\alpha}_j^{\text{co}})} \right|;\ \forall j \right)$.
- **Step 11:** *Obtain the maximum-likelihood estimators for the selected parameters, i.e.* $\boldsymbol{\theta}^*$, *and calculate the AIC objective function* $AIC^{(\nu)}$ *from (7). Continue in Step 12.*
- **Step 12:** *If* $AIC^{(\nu)} < AIC^{(\nu-1)}$ *and there are additional covariates which could be incorporated in the model, go to Step 9, if there are not additional covariates go to Step 13, otherwise continue in Step 14.*
- **Step 13:** *End of the covariate checking, the optimal model parameter vector is* $\boldsymbol{\theta}^{(\nu)}$. *Continue in Step 15.*
- **Step 14:** *End of the covariate checking, the optimal model parameter vector is* $\boldsymbol{\theta}^{(\nu-1)}$. *Continue in Step 15.*
- **Step 15: Trend.** *Continue with iteration* ν, *and include in the parameter vector* $\boldsymbol{\theta}^{(\nu)}$ *the tendency for the location parameter* β_{LT}. *Go to the next step.*
- **Step 16:** *Obtain the maximum-likelihood estimators for the selected parameters, i.e.* $\boldsymbol{\theta}^*$, *and calculate the AIC objective function* $AIC^{(\nu)}$ *from (7).*
- **Step 17:** *If* $AIC^{(\nu)} < AIC^{(\nu-1)}$, *the trend with respect to the location parameter* β_{LT} *is significant, update the iteration counter* $\nu \rightarrow \nu + 1$, *otherwise the*

iteration counter is not updated and the location trend is removed from the parameter vector. For both cases continue with Step 18.

- **Step 18: Tendency.** *Include in the parameter vector* $\boldsymbol{\theta}^{(\nu)}$ *the trend for the scale parameter* α_{LT}.
- **Step 19:** *Obtain the maximum-likelihood estimators for the selected parameters, i.e.* $\boldsymbol{\theta}^*$, *and calculate the AIC objective function* $AIC^{(\nu)}$ *from (7).*
- **Step 20:** *If* $AIC^{(\nu)} < AIC^{(\nu-1)}$, *the trend with respect to the scale parameter* α_{LT} *is significant, update the iteration counter* $\nu \rightarrow \nu + 1$, *otherwise the iteration counter is not updated and the scale tendency is removed from the parameter vector. End of the algorithm.*

■

Note that the proposed algorithm has three different parts related to harmonics, covariates and trends, respectively, as shown in Figure 1. The procedure approaches the best model parameters decreasing the AIC monotonically until no further improvement is possible.

3.3 Maximum likelihood estimation

For parameter statistical inference the method of maximum likelihood is used. Note that this estimation process is repeated once a new parameter is included in the vector $\boldsymbol{\theta}$. The log-likelihood function for the parameters given by (2)-(4) is as follows:

$$\begin{aligned} \ell(\boldsymbol{x}, \boldsymbol{\theta}) &= \sum_{t=1}^{n_d} \log(g(x_t; \mu_t, \psi_t, \xi_t)) \\ &= -\sum_{t=1}^{n_d} \left\{ \log \psi_t + \left(1 + \frac{1}{\xi_t}\right) \log z_t + z_t^n \right\}, \end{aligned} \tag{19}$$

where $z_t = 1 + \xi_t \left(\dfrac{x_t - \mu_t}{\psi_t} \right)$ and $z_t^n = z_t^{-1/\xi_t}$ are auxiliary parameters to simplify the analytical expression (19) and its computational implementation. For the Gumbel case, i.e. $\xi_t = 0$, the log-likelihood function is given by:

$$\ell(\boldsymbol{x}, \boldsymbol{\theta}) = -\sum_{t=1}^{n_d} \left\{ \log \psi_t + \frac{x_t - \mu_t}{\psi_t} + \exp\left(-\frac{x_t - \mu_t}{\psi_t} \right) \right\}. \tag{20}$$

Note that the tail of the distribution may change during the periods of time considered, so that expressions (19) and (20) must be combined.

The maximization of the log-likelihood function can be done using an unconstrained nonlinear optimization routine. Additionally, a constrained optimiza-

tion method including upper and lower bounds for some parameters could be used. We have run different tests using alternative solvers to check the consistency of the obtained results and the adequacy of the different solver possibilities:

(1) Solver MINOS (Murtagh and Saunders, 1998) under GAMS (Brooke et al., 1998) including bounds on parameters to be estimated, which uses a reduced-gradient algorithm (Wolfe, 1963) combined with the quasi-Newton algorithm described in Murtagh and Saunders (1978). The gradient vector information is obtained using numerical differentiation.
(2) A Trust Region Reflective Algorithm under Matlab without bounds using function `fminunc` or with upper and lower bounds through the function `fmincon`. For details about the method see Coleman and Li (1994) and Coleman and Li (1996). In order to improve convergence properties both the gradient and hessian of the objective function are calculated analytically (see Appendix A for details).
(3) A global optimization procedure, the shuffled complex evolution (SCE) algorithm (Duan et al., 1992), which consists of an intelligent Monte Carlo minimum search for obtaining global solutions.

Based on those tests, the following comments and recommendations are pertinent:

(1) Although the theoretical possibility of multiple local maxima exists, such phenomenon is rare, and it has been speculated that they can not occur at all for simple models (stationary) (Smith, 2001). However, in this paper complex models with many covariates are used, and the possibility of multiple maxima is real. Numerical tests using different data, different solvers and different starting points converged to the same solution, which indicates that these kinds of models for environmental data are well posed from the numerical point of view. On the other hand, it has been checked that once one parameter is included in the model, its optimal value does not change significantly although new parameters are included.
(2) The trust region reflective algorithm has been chosen because i) analytical first and second order derivative information can be included, ii) upper and lower bounds on parameters can be considered easily, iii) it is computationally faster than GAMS. Note that although the MLE parameter fitting is an unconstrained maximization problem, we rather use a constrained optimization solver to including parameter bounds, which makes the estimation more robust. These bounds help avoiding parameter z_t in (19) taking negative values. However, if, for any t, the z_t value is below a small quantity (10^{-4}), its value is kept equal to this small quantity ($z_t = \min(z_t, 10^{-4})$).
(3) All Newton-type routines require the user to supply starting values, but the importance of good starting values can be overemphasized. Simple

guesses are enough (Smith, 2001). Thus, for the first iteration, initial guesses are taken as:

$$\beta_0 = \sum_{t=1}^{n_d} x_t/n_d;\ \alpha_0 = \sqrt{\sum_{t=1}^{n_d}(x_t - \beta_0)^2/n_d};\ \gamma_0 = 0.1, \tag{21}$$

where the first two parameters correspond to the sample mean and standard deviation, respectively.

Since the overall method increases the number of parameters sequentially, the best optimal parameters obtained so far are used as starting values for the parameters of the next iteration. Note that new parameters are initialized to zero.

(4) Shape parameter bounds are critical for ensuring convergence, if no information is available, they are constrained to fall between the following values:

$$\gamma_o^{\text{up}} = 0.2;\ \gamma_o^{\text{lo}} = -0.2;\ \gamma_j^{\text{up}} = 0.15;\ \gamma_o^{\text{lo}} = -0.15.$$

Note that if no convergence is achieved, or any of the optimal values equals its corresponding upper or lower bounds, the optimization procedure must be launched again modifying those bounds:

$$\gamma^{\text{up}} = \gamma^{\text{up}} + 0.05 \text{ and/or } \gamma^{\text{lo}} = \gamma^{\text{lo}} - 0.05,$$

where γ corresponds to the corresponding active bound, until convergence is achieved and the optimality conditions given by (9) hold. Note that the remaining parameters do not need to be bounded, although if for a given data set the method does not converge additional bounds would probably help in the optimization procedure.

(5) The Gumbel case is treated as a particular case of the GEV distribution. If the shape parameter absolute value $|\xi_t|$ is below a given threshold ε_{ξ_t}, we use expressions for the Gumbel case. In this particular case $\varepsilon_{\xi_t} = 10^{-8}$.

(6) Scaling data within numerical optimization procedures based on gradient information enhances the efficiency and reliability of the numerical process. Thus data must be scaled about 1.0, which means that, for example, if maximum wave height data values are about 200 centimeters, we would use the data in meters dividing by 100.

One important consideration for the justification of the functioning of the method is that the effect of including an additional parameter is a perturbation over the previous existing model, so that once one parameter is included in the model, consideration of additional parameters does not change its value significantly. From this result the following observations are pertinent:

(1) The sensitivity analysis provides the most likely relevant parameter to be included in the model.

(2) The last optimal values are appropriate starting values for fitting a new model, which includes an additional parameter.

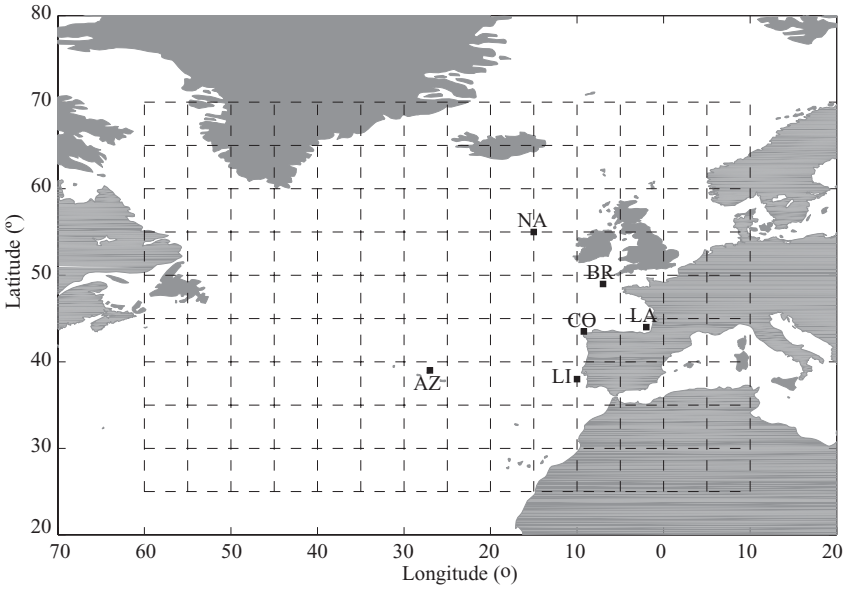

Fig. 2. Reanalysis domain and locations of the study: 15W55N (NA), Azores (AZ), Bretagne (BR), Landas (LA), Coruña (CO) and Lisbon (LI).

The convergence characteristics of the proposed algorithm are discussed in Appendix B.

4 Maximum Significant Wave Height Adjustment in Different Locations

In order to show the functioning of the method in real cases, maximum significant wave height data from six different locations: 15W55N (NA), Azores (AZ), Bretagne (BR), Landas (LA), Coruña (CO) and Lisbon (LI) (see Figure 2) are considered. Data records consist of 528 maximum monthly values obtained from a wave reanalysis database from 1958 to 2001. We have chosen these data because of their homogeneity, both in time and in space. It is conformed by 6-hourly continuous significant wave height data records, which affects positively the stability of parameter estimates. Note that reanalysis data also include monthly mean Sea Level Pressures (SLP), which allows considering as covariates their first 10 principal components in the North Atlantic area at the times when the maximum values occur. Note that considerations about the adequacy of using reanalysis data for extreme value analysis is out of the scope of the paper, since the methodology is suitable to be applied to any set of maxima or minima data.

We have applied the proposed algorithm to get the best parameter selection for those particular sites. Results are given in Table B.1, where the best or pseudo-optimal parameter selection, the corresponding optimal maximum likelihood

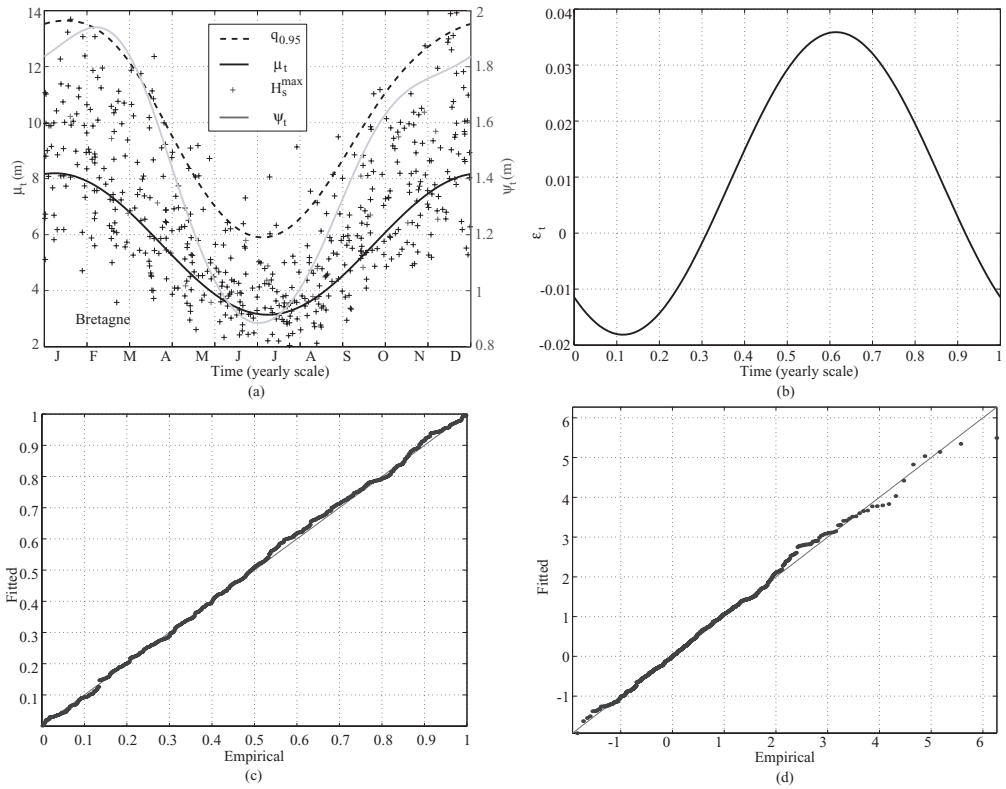

Fig. 3. Goodness of fit plots for the Bretagne data location corresponding to the best model: (a) Maximum significant wave height data, location and scale parameters of the best model and 20-year return period quantile (95%), (b) shape parameter of the best model, (c) probability plot, and (d) quantile plot.

estimators ($\boldsymbol{\theta}^*$) and their standard deviations ($\boldsymbol{\sigma}_\theta$) are provided. Additionally, the pseudo-optimal number of parameters n_p, the pseudo-optimal Akaike Information Criterion (AIC), and the optimal value of the log-likelihood function are presented.

Note that even though there is a mechanism to avoid shape parameters from varying along the year without the other parameters also varying, this never occurred within our numerical experiments, which indicates that this rarely happens in practice using real data. This confirms that the shape time scale dependent variations are longer than those for the location and scale parameters.

In Figure 3 the evolution of the location, scale, and shape parameters (solid lines) along the year for Bretagne (BR) site, and the 20-year return-period quantile (95%, dashed line) are shown. The model checking is graphically evaluated using quantile-quantile (QQ) and probability-probability plots (PP) (see Figures 3 (c) and (d)) which show very good diagnostics, with points close to the diagonal. Analogous results were obtained for the remaining sites.

In terms of location parameters, the optimal harmonic selection varies be-

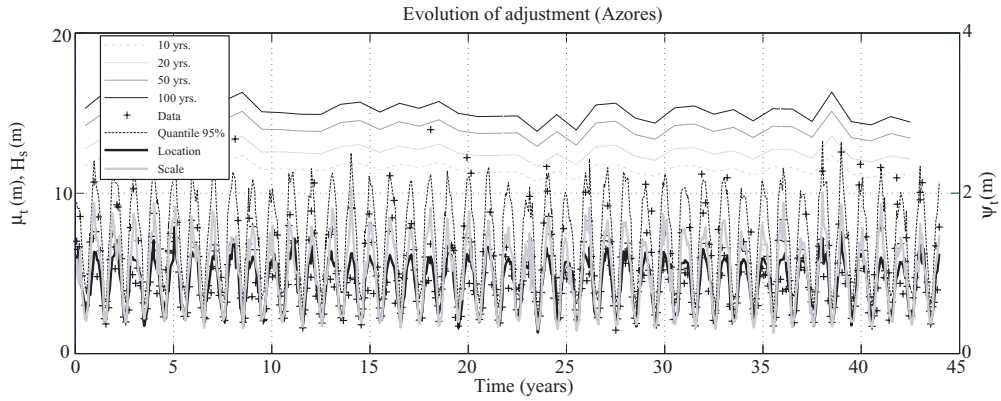

Fig. 4. Inter-annual variability induced by the covariate effects for the Azores site, showing the location, scale, 95% quantile and different aggregated quantiles evolution through different years.

tween two parameters for 15W55N, Bretagne, Landas, Lisbon and Coruña, which correspond to the annual cycle, and four parameters for Azores, which correspond to the annual plus semiannual cycle. This is enough to catch up the modulation in the mean values along the year for all sites. Note that in all cases the higher values correspond to the winter season (from November to March), whereas the lower values correspond to the summer season, reaching a minimum in July (see Figure 3 (a)). The overall behavior of the scale parameter is analogous to the location parameter, increasing the maximum variability during the winter season and decreasing during the summer. However, the variation during the year is less smooth, which is accounted for in the model with the inclusion of additional harmonics, up to four (eight parameters) for the Landas site. Finally, the shape parameter presents less variability, so that it is constant for all sites but for Bretagne and Coruña. Note that in 15W55N and Coruña present a negative shape parameter (upper bounded tail) but close to zero, which almost corresponds to a Gumbel distribution. Analogously, Azores and Landas shape parameters are positive (heavy tail, Fréchet distribution) but close to zero, whereas Lisbon distribution is Weibull. Bretagne presents a different tail behavior along the year, being Weibull from November to April, Gumbel around May, and turning to a Fréchet distribution for the rest of the year (June to October) as shown in Figure 3 (b).

In Figure 4, the influence of the covariate effects in i) the inter-annual variability of the GEV parameters, ii) the 95% quantile and iii) different aggregated return period quantiles, are shown for the Azores site.

It is interesting to observe that Bretagne and Landas optimal models introduce a positive long-term trend in the location parameter, and Bretagne also includes a negative trend in the scale parameter.

In Figure 5, the evolution of the location and scale parameters corresponding to the final selected model for all locations is shown. Note that the model is

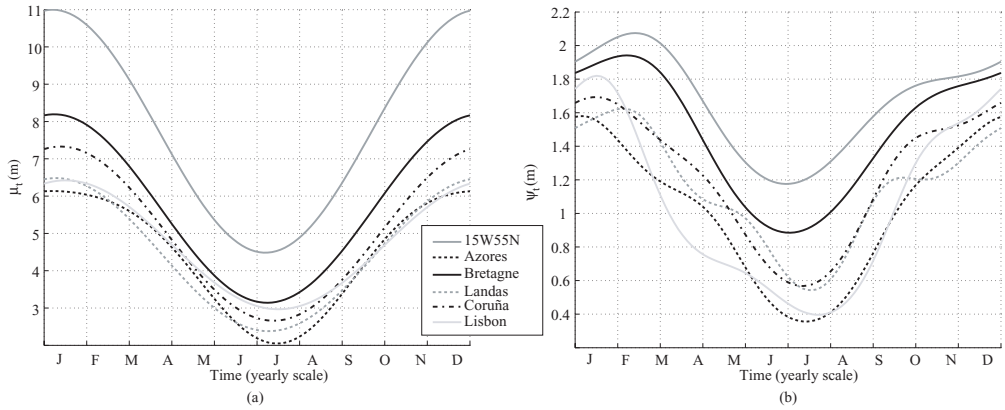

Fig. 5. Location and scale parameter seasonal evolution for the final model at all locations .

flexible enough to represent the seasonal evolution of the parameters within the year.

In Table B.2 the evolution of the model using the proposed method (see algorithm 1), for the particular site 15W55N, is provided. For this particular case, the sequence is as follows:

(1) The first model corresponds to the traditional stationary model $\boldsymbol{\theta} = (\beta_0, \alpha_0, \gamma_0)$, whose optimal values are given in column 1 of Table B.2. Note that the number of parameters is $n_p = 3$, resulting in an AIC equal to AIC = 2764.427.
(2) From all possible annual cycles to be incorporated, the higher score test statistic absolute value corresponds to location, which is included in the model. The new optimal values of the model parameters are given in column 2. Note that the number of parameters has increased by two since we introduce two additional parameters for harmonics. Note that the AIC reduces considerably with respect to the stationary model.
(3) For iterations 3 and 4, higher score test statistic absolute value corresponds to to the scale harmonic parameters, which are incorporated successively, increasing the number of parameters up to $n_p = 9$ and decreasing AIC down to AIC = 2338.251 at the end of iteration 4.
(4) For iteration 5, a new harmonic related to the shape parameter, which is the one with higher score test statistic absolute value, is introduced. However, for $n_p = 11$ parameters it results in a higher AIC = 2569.613. This indicates that the harmonic is not significant and that the harmonic process has concluded.
(5) At this step, the covariate process starts. For the best model so far (iteration 4), the sensitivities of all possible covariates (ten principal components of the SLP) both in location and scale parameters are evaluated. Since the score test statistic absolute value is maximum for the first component associated with location, it is introduced in the model resulting

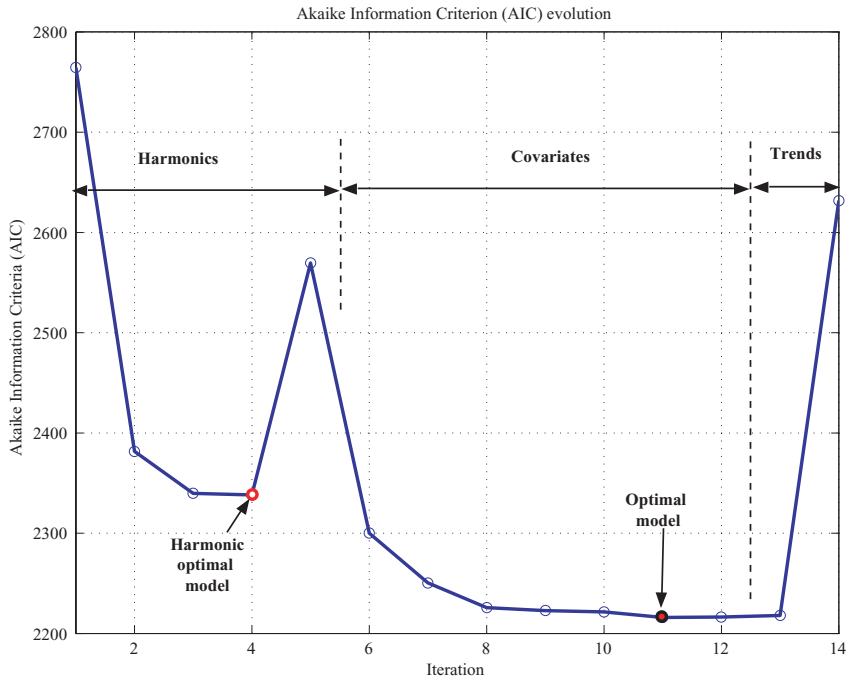

Fig. 6. Akaike Information Criterion (AIC) evolution for the automatic parameter selection process at 15W55N (NA) site.

in AIC equal to (AIC = 2299.982), which is lower than the best model so far, i.e. it is significant. Note that the first principal component of the SLP is highly correlated with the North Atlantic Oscillation (NAO) Index.

(6) The next component to be introduced is related to the second principal component of the SLP, which is correlated with the North Atlantic Oscillation (NAO) and the Eastern Atlantic (EA) Pattern. Note that the new parameter also reduces the AIC which means that it is significant. The process continues.

(7) From iterations 8 to 11, the following covariate parameters are introduced: β_3^{co}, α_3^{co}, α_6^{co}, and β_7^{co}, where the AIC reduces monotonically. Finally, at iteration 12, α_{10}^{co} is introduced but resulting in an increase of the AIC. This ends the process concluding that the best model is given at iteration 11, resulting in a pseudo-optimal value of the log-likelihood function of (AIC* = 2215.984).

(8) At iteration 13, the location trend parameter β_{LT} is introduced, optimal parameter values are in column 13. Since the AIC is higher than at iteration 11 (AIC = 2217.979 > 2215.984) this parameter is not significant and it is removed for the remaining iterations.

(9) Analogously, at iteration 7, the scale tendency parameter α_{LT} is introduced. Since the AIC is higher than at iteration 4 (AIC = 2636.716 > 2215.984) this parameter is also not significant and it is removed for the remaining iterations.

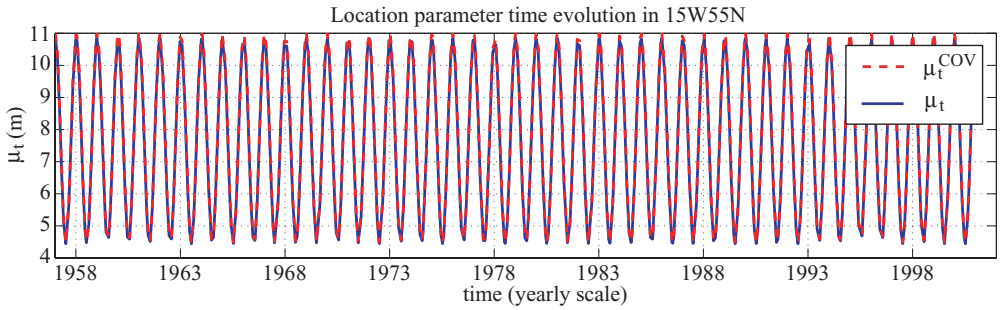

Fig. 7. Location parameter time evolution considering i) harmonics accounting for seasonality, and ii) seasonality and covariate effects(COV) for the 15W55N data location.

In Figure 6 it is shown how the Akaike Information Criterion (AIC) evolves during the iteration process. It can be observed that it is clearly divided in three stages corresponding to harmonics, covariates and trends, respectively. For each of these regions the objective function decreases monotonically until an increase occurs, which means that the corresponding stage has concluded. Analogous graphs are obtained for the remaining sites. Note that the order criterion: harmonics, covariates and trends is based on physical considerations, from lower to higher time-scale effects, which makes the method organization clearer and easier to implement. Numerical tests have confirmed the validity of these assumptions for most cases. For this reason we consider our solution as pseudo-optimal instead of optimal.

Note that the proposed algorithm assumes that the process converges monotonically to the optimal parameter selection. This hypothesis is justified since consideration of subsequent parameters introduces a perturbation around the existing model, so that new optimal values of existing parameters do not change significantly between consecutive iterations. Once the optimal parameter values for a given parameterization are obtained, the method relies on information about the maximum perturbation produced by additional parameters evaluated at those optimal values. This derivative information is only true locally and valid for small variations of the optimal parameter values. Note that once a new parameter is introduced, the only way to know its real influence on the log-likelihood function and the new parameter estimates is fitting the model again. Since we do not fit all possible alternatives, our result would be valid if the new optimal parameter values associated with the already existing parameters do not change significantly. Figure 7 shows both the location parameters including i) harmonics and ii) harmonics and covariates; note that they are almost indistinguishable. This justifies the good computational behavior of the proposed method.

4.1 Comparison with an existing method

In order to compare the results with an existing method, models proposed in the previous section for the six locations have also been automatically fitted using the method proposed in Menéndez et al. (2009a). In that work, based on genetic algorithm nomenclature, a binary codification to represent each model is adopted. Genes g_i are related to cycles or harmonics (annual, semiannual, and so on), covariates and trends, where $g_i = 1$ indicates that the corresponding factor is included. Afterwards, an automatic selection procedure based on the AIC is performed using 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 backwards to see if their contributions are significant after a new gene has been switched on.

The set up for the Stepwise Forward selection and Backward elimination (SFB) method is more complex than for the proposed method, since the maximum number of genes must be fixed beforehand, and appropriate limits for all parameters must be provided for the Shufflex Complex Evolution Optimization method to work appropriately. For this particular example, the following genes are considered: two for the location harmonics (4 parameters), two for the scale harmonics (4 parameters), no harmonics for the shape parameter, 10 genes associated to the location covariates, 10 genes related to the scale covariates, and two additional genes for the trends.

In Table B.3 the comparison between both methods is provided. For each location, optimal parameter values obtained through both methods are shown in columns. If the method does not include a particular parameter, its corresponding cell is set to $(-)$. Additionally, in the last rows, the number of parameters of the final model, the optimal log-likelihood function obtained, and the number of different models checked to attain the best solution are provided. Note that using (7) the higher the log-likelihood function (which is negative) the better the GEV fit, and the lower the AIC the better the compromise between fitting and model complexity (number of parameters). From this table, the following observations are pertinent:

(1) For locations 15W55N, Azores, Bretagne, and Coruña the proposed method incorporates more parameters than SFB, resulting in a higher value of the log-likelihood function (better GEV fit) but also a higher value of the AIC (worst compromise), however the differences are small. Note that in terms of obtaining the best AIC value, the SFB defeats our algorithm since our model always gets a better fit but worse AIC values (overparameterization). This is an expected result since the number of different parameter combinations using SFB is about $\circ(n_p^2)$, comparatively much

higher than the the $\circ(n_p)$ parameter combinations using our approach. In all cases, the computational time using our method decreases considerably being, for these particular cases, about 100 times faster.

(2) For locations Lisbon and Landas the proposed method also incorporates more parameters to the final fitted distribution; however, in these cases both the log-likelihood function and AIC improve, i.e., higher and lower values than the SFB solution respectively. This result is due to the limitation of the SFB method, which fixes the maximum number of possible harmonics to be incorporated at the beginning of the process to reduce the number of combinations of the algorithm.

(3) Related to covariates, the more significant covariates (i.e., higher absolute values of their corresponding parameters) are essentially the same for both models, while those covariates that are different are associated with comparatively lower parameter values. For example, for Bretagne, the proposed model incorporates two additional scale covariates with respect to SFB, α_1^{co} and α_9^{co}. However, their corresponding optimal parameters are -0.036 and 0.01. Relatively small values compared with the maximum covariate parameter value 0.132.

(4) Both methods incorporate the same trends with similar magnitudes, however for the Landas site long term trends related to the location parameters are of different sign and magnitude. This result is not surprising because our method incorporates two additional harmonics with respect to SFB, and the trend obtained by SFB is trying to explain the contribution not considered through those harmonics. Note that our result is more reliable since both the log-likelihood and AIC criteria are better.

5 Conclusions

Within the context of natural variability of extreme events associated with environmental (geophysical) variables, the main contribution of the paper is to establish valuable guidelines for practical applications, which can be used for any scientist to make extreme value analysis from data sets. This work provides an automatic method for parameter selection which minimizes the Akaike Information Criterion (AIC) within an iterative scheme. It is based on an iterative method which converges monotonically to the pseudo-optimal selection. Incorporation of additional parameters at every iteration is based on sensitivity analysis and score test statistic information. The method provides an efficient and robust way for automatic calibration which clearly outperforms an approach existing in the literature, in terms of the ratio of computational time versus optimal objective function.

This work opens the possibility of an automatic fit for time-dependent extreme value models considering different time scales of interest for any environmental

variable. Note that Matlab functions for a) estimating the parameters of the GEV distribution function based on the maximum likelihood method for a given parameterization, and b) automatic selection and estimation minimizing the AIC are provided.

The utilization of the perturbation criterion for the selection of parameters brings new possibilities for using alternative criteria to the AIC, in terms of assessing the quality of each model by making a compromise between goodness of fit and simplicity. This constitutes a subject for further research.

6 Acknowledgments

The authors would like to thank "Puertos del Estado" (Spanish State Port) for providing the information from the wave reanalysis data base. R. Mínguez is indebted to the Spanish Ministry of Science and Innovation (MCI) for the funding provided within the "Ramon y Cajal" program. This work was partly funded by the project "GRACCIE" (CSD2007-00067, Programa Consolider-Ingenio 2010) from MCI, by project C3E (200800050084091) from the Spanish Ministry of Environment and by project MARUCA (E17/18) from the Spanish Ministry of Public Works.

A Jacobian and Hessian of the log-likelihood function

For the best parameter selection at every iteration and for maximization of the log-likelihood function, first and second order derivatives of the log-likelihood function with respect to parameters $\boldsymbol{\theta}$ at the optimum must be obtained. The analytical derivation for all required matrices is provided below. Note that all derivations are based on the chain rule. Hereinafter index i refers to maximum data, j is related to harmonics of the location, scale and shape parameters, and k is associated with possible covariates included in the model (SLP principal components, NAO, SOI, etc.).

A.1 First order derivatives of the log-likelihood function

First order derivatives of the log-likelihood function with respect to location, scale and shape parameters evaluated at point x_t (time t) are:

$$\left.\frac{\partial \ell}{\partial \mu_t}\right|_{x_t} = \frac{1+\xi_t - z_t^n}{\psi_t z_t} \tag{A.1}$$

$$\left.\frac{\partial \ell}{\partial \psi_t}\right|_{x_t} = -\frac{1-x_t^n(1-z_t^n)}{\psi_t z_t} \tag{A.2}$$

$$\left.\frac{\partial \ell}{\partial \xi_t}\right|_{x_t} = \frac{z_t^n\left[x_t^n\left(1-\frac{1+\xi_t}{z_t^n}\right)+\frac{z_t(-1+1/z_t^n)\log(z_t)}{\xi_t}\right]}{\xi_t z_t} \tag{A.3}$$

where $x_t^n = \frac{x_t-\mu_t}{\psi_t}$ is the standardized maximum parameter value, which are valid for the case $\xi_t \neq 0$ (Weibull or Fréchet), and

$$\left.\frac{\partial \ell}{\partial \mu_t}\right|_{x_t} = \frac{1-\exp\left(-x_t^n\right)}{\psi_t} \tag{A.4}$$

$$\left.\frac{\partial \ell}{\partial \psi_t}\right|_{x_t} = \frac{x_t^n - 1 - x_t^n \exp\left(-x_t^n\right)}{\psi_t} \tag{A.5}$$

for the Gumbel case $\xi = 0$.

Note that from the practical point of view the Gumbel distribution is a particular case of the generalized extreme value distribution for a small value of the shape parameter ξ. For those cases, equations (A.1) and (A.2) are equivalent to (A.4) and (A.5), respectively, and the partial derivative with respect to the shape parameter is null. Analogously, hereinafter if $\xi_t = 0$ then all derivatives related to parameters associated with the shape parameter are null.

Using the chain rule, the derivatives of the log-likelihood function with respect to the harmonic amplitudes, including the constant terms, are:

$$\frac{\partial \ell}{\partial \beta_0} = \sum_t \frac{\partial \ell}{\partial \mu_t}\frac{\partial \mu_t}{\partial \beta_0}\,;\; \frac{\partial \ell}{\partial \beta_j} = \sum_t \frac{\partial \ell}{\partial \mu_t}\frac{\partial \mu_t}{\partial \beta_j};\ \forall j \tag{A.6}$$

$$\frac{\partial \ell}{\partial \alpha_0} = \sum_t \frac{\partial \ell}{\partial \psi_t}\psi_t\frac{\partial \log(\psi_t)}{\partial \alpha_0}\,;\; \frac{\partial \ell}{\partial \alpha_j} = \sum_t \frac{\partial \ell}{\partial \psi_t}\psi_t\frac{\partial \log(\psi_t)}{\partial \alpha_j};\ \forall j \tag{A.7}$$

$$\frac{\partial \ell}{\partial \gamma_0} = \sum_t \frac{\partial \ell}{\partial \xi_t}\frac{\partial \xi_t}{\partial \gamma_0}\,;\; \frac{\partial \ell}{\partial \gamma_j} = \sum_t \frac{\partial \ell}{\partial \xi_t}\frac{\partial \xi_t}{\partial \gamma_j};\ \forall j. \tag{A.8}$$

The corresponding derivatives with respect to possible tendencies or covariates, if they are considered, are:

$$\frac{\partial \ell}{\partial \beta_{\mathrm{LT}}} = \sum_t \frac{\partial \ell}{\partial \mu_t}\frac{\partial \mu_t}{\partial \beta_{\mathrm{LT}}}\,;\; \frac{\partial \ell}{\partial \beta_k^{\mathrm{co}}} = \sum_t \frac{\partial \ell}{\partial \mu_t}\frac{\partial \mu_t}{\partial \beta_k^{\mathrm{co}}};\ \forall k \tag{A.9}$$

$$\frac{\partial \ell}{\partial \alpha_{\mathrm{LT}}} = \sum_t \frac{\partial \ell}{\partial \psi_t} \psi_t \frac{\partial \log(\psi_t)}{\partial \alpha_{\mathrm{LT}}}\,; \frac{\partial \ell}{\partial \alpha_k^{\mathrm{co}}} = \sum_t \frac{\partial \ell}{\partial \psi_t} \psi_t \frac{\partial \log(\psi_t)}{\partial \alpha_k^{\mathrm{co}}};\ \forall k \tag{A.10}$$

where index k refers to covariates.

The derivatives of the location, logarithm of the scale and shape parameters, given by (2)-(4), the harmonic amplitudes, including the constant terms, are:

$$\begin{aligned}
&\tfrac{\partial \mu_t}{\partial \beta_0} = 1; \qquad \tfrac{\partial \mu_t}{\partial \beta_j} = \begin{cases} \cos(\tfrac{j+1}{2} wt) & \text{if } j \text{ odd} \\ \sin(\tfrac{j}{2} wt) & \text{if } j \text{ even} \end{cases}; \forall j \\
&\tfrac{\partial \log(\psi_t)}{\partial \alpha_0} = 1; \tfrac{\partial \log(\psi_t)}{\partial \alpha_j} = \begin{cases} \cos(\tfrac{j+1}{2} wt) & \text{if } j \text{ odd} \\ \sin(\tfrac{j}{2} wt) & \text{if } j \text{ even} \end{cases}; \forall j \\
&\tfrac{\partial \xi_t}{\partial \gamma_0} = 1; \qquad \tfrac{\partial \xi_t}{\partial \gamma_j} = \begin{cases} \cos(\tfrac{j+1}{2} wt) & \text{if } j \text{ odd} \\ \sin(\tfrac{j}{2} wt) & \text{if } j \text{ even} \end{cases}; \forall j.
\end{aligned} \tag{A.11}$$

Analogously, the corresponding derivatives with respect to possible tendencies or covariates, if they are considered, are:

$$\begin{aligned}
&\tfrac{\partial \mu_t}{\partial \beta_{\mathrm{LT}}} = t; \qquad \tfrac{\partial \mu_t}{\partial \beta_k^{\mathrm{co}}} = n_{k_t};\ \forall k, \\
&\tfrac{\partial \log(\psi_t)}{\partial \alpha_{\mathrm{LT}}} = t; \tfrac{\partial \log(\psi_t)}{\partial \alpha_k^{\mathrm{co}}} = n_{k_t};\ \forall k.
\end{aligned} \tag{A.12}$$

A.2 Second order derivatives of the log-likelihood function

Second order derivatives of the log-likelihood function with respect to location, scale and shape parameters evaluated at point x_t (time t) are:

$$\begin{aligned}
\tfrac{\partial^2 \ell}{\partial \mu_t^2} &= \tfrac{(1+\xi_t)(\xi_t - z_t^n)}{(\psi_t z_t)^2} \\
\tfrac{\partial^2 \ell}{\partial \psi_t^2} &= \tfrac{-z_t^n x_t^n \left((1-\xi_t)x_t^n - 2\right) + \left[1 - 2x_t^n - \xi_t \left(x_t^n\right)^2\right]}{(\psi_t z_t)^2}
\end{aligned} \tag{A.13}$$

$$\begin{aligned}
\tfrac{\partial^2 \ell}{\partial \xi_t^2} = -z_t^n \Bigg[&\tfrac{x_t^n \left(x_t^n [1+3\xi_t] + 2 + [-2 - \xi_t(3+\xi_t) x_t^n] z_t^{1/\xi_t} \right)}{(\xi_t z_t)^2} \\
&\tfrac{+\frac{z_t}{\xi_t^2} \log(z_t) (2\xi_t (-x_t^n (1+\xi_t) - 1 + z_t^{1+1/\xi_t}) + z_t \log(z_t))}{(\xi_t z_t)^2} \Bigg]
\end{aligned} \tag{A.14}$$

$$\frac{\partial^2 \ell}{\partial \mu_t \partial \psi_t} = \frac{-[1 + \xi_t - (1 - x_t^n) z_t^n]}{(\psi_t z_t)^2} \tag{A.15}$$

$$\frac{\partial^2 \ell}{\partial \mu_t \partial \xi_t} = -z_t^n \frac{\xi_t \left[-(1+\xi_t)x_t^n - \xi_t(1-x_t^n)/z_t^n\right] + z_t \log(z_t)}{\psi_t \left[\xi_t z_t\right]^2} \tag{A.16}$$

$$\frac{\partial^2 \ell}{\partial \psi_t \partial \xi_t} = x_t^n \frac{\partial^2 \ell}{\partial \mu_t \partial \xi_t} \tag{A.17}$$

where $x_t^n = \frac{x_t - \mu_t}{\psi_t}$ is a standardized maximum data value, which are valid for the case $\xi \neq 0$ (Weibull or Fréchet), and

$$\frac{\partial^2 \ell}{\partial \mu_t^2} = \frac{-\exp\left(-x_t^n\right)}{\psi_t^2} \tag{A.18}$$

$$\frac{\partial^2 \ell}{\partial \psi_t^2} = \frac{1 - 2x_t^n + \exp\left(-x_t^n\right)(2 - x_t^n)x_t^n}{\psi_t^2} \tag{A.19}$$

$$\frac{\partial^2 \ell}{\partial \mu_t \partial \psi_t} = \frac{-1 + \exp\left(-x_t^n\right)(1 - x_t^n)}{\psi_t^2} \tag{A.20}$$

for the Gumbel case $\xi_t = 0$.

For the definition of the second order derivatives for the Hessian matrix considering location, scale and shape parameters, two auxiliary variables Ω and Φ are considered. These variables may represent any of the following parameters $\boldsymbol{\theta} = (\beta_0, \beta_i, \beta_{\text{LT}}, \beta_k^{\text{co}}, \alpha_0, \alpha_i, \alpha_{\text{LT}}, \alpha_k^{\text{co}}, \gamma_0, \gamma_i)$, which can be equal (for second order derivatives) or different (for second order cross derivatives). Considering also that if Ω is equal to any of $(\beta_0; \beta_i; \beta_{\text{LT}}; \beta_k^{\text{co}})$ then $f_\Omega = \mu_t$, if Ω is equal to any of $(\alpha_0; \alpha_j; \alpha_{\text{LT}}; \alpha_k^{\text{co}})$ then $f_\Omega = \psi_t$, if Ω is equal to any of $(\gamma_0; \gamma_i)$ then $f_\Omega = \xi_t$, and analogously for Φ, where f_Ω and f_Φ are auxiliary functions to simplify analytical derivations. The second order derivatives can be calculated as follows:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \Omega \partial \Phi} &= \frac{\partial}{\partial \Omega}\left(\sum_t \left[\frac{\partial \ell}{\partial \mu_t}\frac{\partial \mu_t}{\partial \Phi} + \frac{\partial \ell}{\partial \psi_t}\frac{\partial \psi_t}{\partial \Phi} + \frac{\partial \ell}{\partial \xi_t}\frac{\partial \xi_t}{\partial \Phi}\right]\right) = \\
&\sum_t \left[\left(\frac{\partial^2 \ell}{\partial \mu_t^2}\frac{\partial \mu_t}{\partial \Omega} + \frac{\partial^2 \ell}{\partial \mu_t \partial \psi_t}\frac{\partial \psi_t}{\partial \Omega} + \frac{\partial^2 \ell}{\partial \mu_t \partial \xi_t}\frac{\partial \xi_t}{\partial \Omega}\right)\frac{\partial \mu_t}{\partial \Phi} + \frac{\partial \ell}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial \Omega \partial \Phi}\right. \\
&+ \left(\frac{\partial^2 \ell}{\partial \psi_t \partial \mu_t}\frac{\partial \mu_t}{\partial \Omega} + \frac{\partial^2 \ell}{\partial \psi_t^2}\frac{\partial \psi_t}{\partial \Omega}\frac{\partial^2 \ell}{\partial \psi_t \partial \xi_t}\frac{\partial \xi_t}{\partial \Omega}\right)\frac{\partial \psi_t}{\partial \Phi} + \frac{\partial \ell}{\partial \psi_t}\frac{\partial^2 \psi_t}{\partial \Omega \partial \Phi} \\
&\left.\left(\frac{\partial^2 \ell}{\partial \xi_t \partial \mu_t}\frac{\partial \mu_t}{\partial \Omega}\frac{\partial^2 \ell}{\partial \xi_t \partial \psi_t}\frac{\partial \psi_t}{\partial \Omega}\frac{\partial^2 \ell}{\partial \xi_t^2}\frac{\partial \xi_t}{\partial \Omega}\right)\frac{\partial \xi_t}{\partial \Phi} + \frac{\partial \ell}{\partial \xi_t}\frac{\partial^2 \xi_t}{\partial \Omega \partial \Phi}\right].
\end{aligned} \tag{A.21}$$

Expression (A.21) simplifies as follows:

$$\frac{\partial^2 \ell}{\partial \Omega \partial \Phi} = \sum_t \left[\frac{\partial^2 \ell}{\partial f_\Phi \partial f_\Omega}\frac{\partial f_\Phi}{\partial \Phi}\frac{\partial f_\Omega}{\partial \Omega}\right], \tag{A.22}$$

if any f_Φ and f_Ω do not correspond to parameter ψ, or

$$\frac{\partial^2 \ell}{\partial \Omega \partial \Phi} = \sum_t \left[\frac{\partial^2 \ell}{\partial f_\Phi \partial f_\Omega} \frac{\partial f_\Omega}{\partial \Omega} f_\Phi \frac{\partial \log(f_\Phi)}{\partial \Phi} \right], \tag{A.23}$$

if f_Φ corresponds to parameter ψ. This is due to the use of $\log(\psi)$ in (3), or

$$\frac{\partial^2 \ell}{\partial \Omega \partial \Phi} = \sum_t \left[\left(\frac{\partial^2 \ell}{\partial f_\Phi^2} f_\Phi + \frac{\partial \ell}{\partial f_\Phi} \right) f_\Phi \frac{\partial \log(f_\Phi)}{\partial \Phi} \frac{\partial \log(f_\Omega)}{\partial \Omega} \right], \tag{A.24}$$

if both f_Φ and f_Ω correspond to parameter ψ, where $\frac{\partial^2 \ell}{\partial f_\Phi \partial f_\Omega}$ is given in (A.13)-(A.15), whereas derivatives $\frac{\partial f_\Phi}{\partial \Phi}$, $\frac{\partial \log(f_\Phi)}{\partial \Phi}$ and $\frac{\partial f_\Omega}{\partial \Omega}$ are equal to:

$$\frac{\partial f_\Phi}{\partial \Phi} \text{ or } \frac{\partial \log(f_\Phi)}{\partial \Phi} \text{ or } \frac{\partial f_\Omega}{\partial \Omega} = \begin{cases} 1 & \text{if } (\Phi \text{ or } \Omega) = \{\beta_0, \alpha_0, \gamma_0\} \\ t & \text{if } (\Phi \text{ or } \Omega) = \{\beta_{\text{LT}} \text{ or } \alpha_{\text{LT}}\} \\ n_{k_t} & \text{if } (\Phi \text{ or } \Omega) = \{\beta_k^{\text{co}} \text{ or } \alpha_k^{\text{co}}\} \\ \cos(\frac{j+1}{2} wt) & \text{if } (\Phi \text{ or } \Omega) = \{\beta_j, \alpha_j, \gamma_j\} \ \& \ j \text{ odd} \\ \sin(\frac{j}{2} wt) & \text{if } (\Phi \text{ or } \Omega) = \{\beta_j, \alpha_j, \gamma_j\} \ \& \ j \text{ even} \end{cases} \tag{A.25}$$

where $\frac{\partial \log(f_\Phi)}{\partial \Phi}$ is meaningful if f_Φ corresponds to ψ.

A.3 Hessian

Once all second order derivatives of the log-likelihood function with respect to all parameters to be estimated are obtained, the Hessian matrix ($\boldsymbol{H}_{\boldsymbol{\theta}\boldsymbol{\theta}}$) has the structure shown in Figure A.1.

B Convergence Discussion

Convergence of the method is based on the following proposition:

Proposition 1 *Once the optimal log-likelihood estimators for a given parameter vector $\boldsymbol{\theta}^*$ are obtained, the new optimal estimators including a new parameter set $\hat{\boldsymbol{\theta}}$ in the model do not change significantly, since they produce perturbations around the previous model.*

Based on this proposition, the following theorem is derived:

Theorem 1 *For given sets of new possible parameters $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_m\}$, which can be introduced in model (1)-(4), the maximum decrement in the AIC*

$$
\boldsymbol{H}_{\boldsymbol{\theta\theta}} = \left(\begin{array}{cccccccccccccc}
\frac{\partial^2\ell}{\partial^2\beta_0} & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_{P_\mu}} & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_{\mathrm{LT}}} & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_0\partial\alpha_0} & \frac{\partial^2\ell}{\partial\beta_0\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_0\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\beta_0} & \frac{\partial^2\ell}{\partial\beta_0\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_0\partial\gamma_0} & \frac{\partial^2\ell}{\partial\beta_0\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_0\partial\gamma_{P_\xi}} \\
 & \frac{\partial^2\ell}{\partial\beta_1^2} & \cdots & \frac{\partial^2\ell}{\partial\beta_1\partial\beta_{P_\mu}} & \frac{\partial^2\ell}{\partial\beta_1\partial\beta_{\mathrm{LT}}} & \frac{\partial^2\ell}{\partial\beta_1\partial\beta_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_1\partial\alpha_0} & \frac{\partial^2\ell}{\partial\beta_1\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_1\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\beta_1} & \frac{\partial^2\ell}{\partial\beta_1\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_1\partial\gamma_0} & \frac{\partial^2\ell}{\partial\beta_1\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_1\partial\gamma_{P_\xi}} \\
 & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & & & \frac{\partial^2\ell}{\partial\beta_{P_\mu}^2} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\beta_{\mathrm{LT}}} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\beta_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\alpha_0} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\beta_{P_\mu}} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_{P_\mu}\partial\gamma_{P_\xi}} \\
 & & & & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}{}^2} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\beta_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\alpha_0} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\beta_{\mathrm{LT}}} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_{\mathrm{LT}}\partial\gamma_{P_\xi}} \\
 & & & & & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}2}} & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\alpha_0} & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\beta_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_k^{\mathrm{co}}\partial\gamma_{P_\xi}} \\
 & & & & & & \frac{\partial^2\ell}{\partial^2\alpha_0} & \frac{\partial^2\ell}{\partial\alpha_0\partial\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_0\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\alpha_0} & \frac{\partial^2\ell}{\partial\alpha_0\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\alpha_0\partial\gamma_0} & \frac{\partial^2\ell}{\partial\alpha_0\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_0\partial\gamma_{P_\xi}} \\
 & & & & & & & \frac{\partial^2\ell}{\partial^2\alpha_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_1\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\alpha_1} & \frac{\partial^2\ell}{\partial\alpha_1\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\alpha_1\partial\gamma_0} & \frac{\partial^2\ell}{\partial\alpha_1\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_1\partial\gamma_{P_\xi}} \\
 & & & & & & & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & & & & & & & & & \frac{\partial^2\ell}{\partial^2\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\alpha_{P_\psi}} & \frac{\partial^2\ell}{\partial\alpha_{P_\psi}\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\alpha_{P_\psi}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\alpha_{P_\psi}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_{P_\psi}\partial\gamma_{P_\xi}} \\
 & & & & & & & & & & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}{}^2} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\alpha_k^{\mathrm{co}}} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_{\mathrm{LT}}\partial\gamma_{P_\xi}} \\
 & & & & & & & & & & & \frac{\partial^2\ell}{\partial\alpha_k^{\mathrm{co}2}} & \frac{\partial^2\ell}{\partial\alpha_k^{\mathrm{co}}\partial\gamma_0} & \frac{\partial^2\ell}{\partial\alpha_k^{\mathrm{co}}\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\alpha_k^{\mathrm{co}}\partial\gamma_{P_\xi}} \\
 & & & & & & & & & & & & \frac{\partial^2\ell}{\partial^2\gamma_0} & \frac{\partial^2\ell}{\partial\gamma_0\partial\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\gamma_0\partial\gamma_{P_\xi}} \\
 & & & & & & & & & & & & & \frac{\partial^2\ell}{\partial^2\gamma_1} & \cdots & \frac{\partial^2\ell}{\partial\gamma_1\partial\gamma_{P_\xi}} \\
 & & & & & & & & & & & & & & \ddots & \vdots \\
 & & & & & & & & & & & & & & & \frac{\partial^2\ell}{\partial^2\gamma_{P_\xi}}
\end{array}\right)
$$

Fig. A.1. Hessian matrix of the log-likelihood function.

is produced by the parameter whose score test statistic absolute value is maximum, i.e.,

$$\hat{\boldsymbol{\theta}}^{*} = \boldsymbol{\theta} / \underset{\boldsymbol{\theta} \in \hat{\boldsymbol{\Theta}}}{\textit{Minimum}}\ AIC(\boldsymbol{\theta}^{*}, \boldsymbol{\theta}) - AIC(\boldsymbol{\theta}^{*}) = \underset{\boldsymbol{\theta} \in \hat{\boldsymbol{\Theta}}}{\textit{Maximum}}\ |S(\boldsymbol{\theta})| \tag{B.1}$$

where $AIC(\boldsymbol{\theta}^{*}, \boldsymbol{\theta})$ is the AIC criterion (7) for the parameter vector $\boldsymbol{\theta}^{*}$ including the new parameter set $\boldsymbol{\theta}$.

Proof 1 *Considering the optimal log-likelihood estimators for a given parameterization $\boldsymbol{\theta}^{*}$ with n_p parameters, assuming that a new set of parameters $\hat{\boldsymbol{\theta}} = \mathbf{0}$ of length m is included in the model and defining the vector $\boldsymbol{\theta} = [\boldsymbol{\theta}^{*}; \hat{\boldsymbol{\theta}}]$, the second order Taylor series expansion of the log-likelihood function is given by:*

$$\ell(\boldsymbol{\theta} + \boldsymbol{\delta\theta}) = \ell(\boldsymbol{\theta}) + \boldsymbol{U}(\boldsymbol{\theta})^{T} \boldsymbol{\delta\theta} + \frac{1}{2} \boldsymbol{\delta\theta}^{T} \boldsymbol{H}_{\boldsymbol{\theta\theta}} \boldsymbol{\delta\theta}, \tag{B.2}$$

where $\boldsymbol{U}(\boldsymbol{\theta})$ is the score vector whose components are equal to $U(\theta_j) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j}$; $j = 1, \ldots, n_p + m$.

The perturbation which produce the maximum or minimum change in the log-likelihood function can be obtained forcing the derivative of the increment with respect the perturbation to be equal to zero as follows:

$$\frac{\partial \left(\ell(\boldsymbol{\theta} + \boldsymbol{\delta\theta}) - \ell(\boldsymbol{\theta}) \right)}{\partial \left(\boldsymbol{\delta\theta} \right)} = \boldsymbol{U}(\boldsymbol{\theta}) + \boldsymbol{H}_{\boldsymbol{\theta\theta}} \boldsymbol{\delta\theta} = \mathbf{0}, \tag{B.3}$$

where the optimal perturbation becomes:

$$\boldsymbol{\delta\theta}^{*} = \left(-\boldsymbol{H}_{\boldsymbol{\theta\theta}} \right)^{-1} \boldsymbol{U}(\boldsymbol{\theta}). \tag{B.4}$$

The corresponding maximum or minimum change in the log-likelihood function can be obtained replacing (B.4) in (B.2), which neglecting second order terms becomes:

$$\Omega \left(\ell(\boldsymbol{\theta} + \boldsymbol{\delta\theta}) - \ell(\boldsymbol{\theta}) \right) = \boldsymbol{U}(\boldsymbol{\theta})^{T} \boldsymbol{\delta\theta}^{*} = \boldsymbol{U}(\boldsymbol{\theta})^{T} \left(-\boldsymbol{H}_{\boldsymbol{\theta\theta}} \right)^{-1} \boldsymbol{U}(\boldsymbol{\theta}), \tag{B.5}$$

where function $\Omega(\bullet)$ corresponds to the maximum or minimum function depending on the Hessian, and expression (B.5) coincides with the score test statistic $S(\boldsymbol{\theta})$.

Due to the optimality condition (9) it is known that $\boldsymbol{U}(\boldsymbol{\theta}^{*}) = \mathbf{0}$ and therefore parameter set $\hat{\boldsymbol{\theta}}$ is the only one contributing to the change in the log-likelihood

function. (B.5) becomes:

$$\Omega\left(\ell(\boldsymbol{\theta}+\boldsymbol{\delta\theta})-\ell(\boldsymbol{\theta})\right)=\boldsymbol{U}(\hat{\boldsymbol{\theta}})^T\boldsymbol{I}^{-1}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}\boldsymbol{U}(\hat{\boldsymbol{\theta}})=S(\hat{\boldsymbol{\theta}}), \tag{B.6}$$

where matrix $\boldsymbol{I}^{-1}_{\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}}$ is the diagonal block matrix from the inverse of the observed Fisher information matrix associated with parameters belonging to $\hat{\boldsymbol{\theta}}$.

Finally, since it is known that at the optimal solution of new parameters $\hat{\boldsymbol{\theta}}^*$ the difference in the maximum possible change in (B.6) is always positive, but the score test statistic could be negative, the maximum change would correspond to the absolute value of the score test statistic, i.e.:

$$\max\left(\ell(\boldsymbol{\theta}+\boldsymbol{\delta\theta})-\ell(\boldsymbol{\theta})\right)=|S(\hat{\boldsymbol{\theta}})|. \tag{B.7}$$

Thus, if different parameter sets $\hat{\boldsymbol{\Theta}}=\{\hat{\boldsymbol{\theta}}_1,\ldots,\hat{\boldsymbol{\theta}}_m\}$ can be introduced in the model, the set that produces the maximum change in the log-likelihood function corresponds to the parameter whose score test statistic absolute value $|S(\hat{\boldsymbol{\theta}}_i)|$ is maximum. Note that using the chain rule the maximum derivative of the AIC coincides, since:

$$\frac{\partial AIC(\boldsymbol{\theta}^*,\hat{\boldsymbol{\theta}})}{\partial\hat{\boldsymbol{\theta}}}=-2\frac{\partial\ell(\boldsymbol{\theta}^*,\hat{\boldsymbol{\theta}})}{\partial\hat{\boldsymbol{\theta}}}+2. \tag{B.8}$$

This completes the proof.

Numerical tests have demonstrated the validity of these assumptions for environmental variables.

Note also that the optimal estimators solution of best selected model at a given iteration is an appropriate starting point for the log-likelihood maximization when an additional parameter is introduced, the new parameter is taken equal to zero. Since the solution does not change significantly, Newton method converges to the optimal solution (see Bazaraa et al. (1993)).

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Csáki, F. (Eds.), Proceedings 2nd International Symposium on Information Theory. Akadémia Kiadó, Budapest, pp. 267–281.

Bazaraa, M. S., Sherali, H. D., Shetty, C. M., 1993. Nonlinear Programming. Theory and Algorithms, 2nd Edition. John Wiley & Sons, New York.

Breiman, L., 1995. Better subset regression using the nonnegative garrote. Technometrics 37, 373–384.

Brooke, A., Kendrick, D., Meeraus, A., Raman, R., 1998. GAMS: A user's guide. GAMS Development Corporation, Washington.

Brown, S. J., Caesar, J., Ferro, A. T., 2008. Global changes in extreme daily temperature since 1950. Journal of Geophysical Research 113, d05115, doi:10.1029/2006JD008091.

Carter, D. J. T., Challenor, P. G., 1981. Estimating return values of environmental variables. Quart. J. Roy. Meteor. Soc. 107, 259–266.

Clarke, M. A., 1981. Statistical algorithms: algorithm as 163: a gvens algorithm for moving from one linear model to another without going back to the data. J. Roy. Statist. Soc. Ser. C Appl. Statist. 30, 198–203.

Coleman, T. F., Li, Y., 1994. On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds. Mathematical Programming 67 (2), 189–224.

Coleman, T. F., Li, Y., 1996. An interior, trust region approach for nonlinear minimization subject to bounds. SIAM Journal on Optimization 6, 418–445.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics.

Cox, D. R., Hinkley, D. V., 1974. Theoretical Statistics. Chapman and Hall, London.

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

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96, 1348–1360.

Furnival, G., Wilson, R., 1974. Regression by leaps and bounds. Technometrics 16, 499–511.

Gatu, C., Kontoghiorghes, E. J., 2006. Branch-and-bound algorithms for computing the best subset regression models. J. Comput. Graph. Statist. 15, 139–156.

George, E. I., Foster, D. P., 1994. The risk inflation criterion for multiple regression. Ann. Statist. 22, 1947–1975.

Hannan, E. J., Quinn, B. G., 1979. The determination of the order of an autoregression. J. R. Statist. Soc. B 41, 190–195.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York.

Hocking, R. R., 1976. The analysis and selection of variables in linear regression. Biometrics 32, 1–49.

Hofmann, M., Kontoghiorghes, E. J., 2007. Pipeline givenssequences for computing the QR decomposition on a EREW PRAM. Parallel Comput. 32, 222–230.

Hurvich, C. M., Tsai, C. L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.

Izaguirre, C., Méndez, F. J., Menéndez, M., Luceño, A., Losada, I. J., 2010.

Extreme wave climate variability in southern europe using satellite data, J. of Geophys. Res., in press (doi:10.1029/2009JC005802).
Katz, R. W., Parlange, M. B., Naveau, P., 2002. Statistics of extremes in hydrology. Advanced Water Resources 25, 1287–1304.
Luenberger, D. G., 1984. Linear and Nonlinear Programming, 2nd Edition. Addison-Wesley, Reading, Massachusetts.
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. J. of Geophys. Res. 111, 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. J. Atmos. Ocean. Technol. 24, 894–911.
Menéndez, M., Méndez, F. J., Izaguirre, C., Losada, I. J., 2009a. The influence of seasonality on estimating return values of significant wave height. Coastal Engineering 56 (3), 211–219.
Menéndez, M., Méndez, F. J., Losada, I. J., 2009b. Forecasting seasonal to interannual variablility in extreme sea levels. ICES Journal of Marine Science 66, 1490–1496.
Morton, I. D., Bowers, J., Mould, G., 1997. Estimating return period wave heights and wind speeds using a seasonal point process model. Coastal Engineering 31, 305–326.
Murtagh, B. A., Saunders, M. A., 1978. Large-scale linearly constrained optimization. Mathematical Programming 14, 41–72.
Murtagh, B. A., Saunders, M. A., 1998. MINOS 5.5 Users Guide. Report SOL 83-20R SOL 83-20R, Department of Operations Research, Stanford University, Stanford, California.
Narendra, P. M., Fukunaga, K., 1997. A branch and bound algorithm for feature subset selection. IEEE Trans. Comput. 26, 917–922.
Rice, J., 1994. Mathematical Statistics and Data Analysis. Duxbury Press, Boston.
Roberts, S. J., 1984. Statistical algorithms: algorithm AS 199: a branch and bound algorithm for determining the optimal feature subset of a given size. Appl. Statist. 33, 236–241.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Smith, R. L., July 2001. Environmental Statistics. Department of Statistics, University of North Carolina.
Smith, R. L., Shively, T. S., 1995. A point process approach to modeling trends in tropospheric ozone. Atmospheric Environmental 29, 3489–3499.
Somol, P., Pudil, P., Kittler, J., 2004. Fast branch and bound algorithms for optimal feature selection. IEEE Trans. Pattern Anal. Mach. Intell. 26, 900–912.
Tibshirani, R. J., 1996. Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B (Stat. Methodol.) 58, 267–288.
Wolfe, P., 1963. Methods of nonlinear programming. In: Graves, R. L., Wolfe,

P. (Eds.), Recent Advances in Mathematical Programming. McGraw-Hill, New York, pp. 76–77.

	15W55N		Azores		Bretagne		Landas		Coruña		Lisbon	
	$\boldsymbol{\theta}^*$	$\boldsymbol{\sigma}_\theta^{(1)}$	$\boldsymbol{\theta}$	$\boldsymbol{\sigma}_\theta$	$\boldsymbol{\theta}$	$\boldsymbol{\sigma}_\theta$	$\boldsymbol{\theta}$	$\boldsymbol{\sigma}_\theta$	$\boldsymbol{\theta}$	$\boldsymbol{\sigma}_\theta$	$\boldsymbol{\theta}$	$\boldsymbol{\sigma}_\theta$
β_0	7.740	8.239	4.410	5.212	5.666	11.649	4.431	9.390	4.994	5.516	4.694	5.434
β_1	3.232	10.586	1.970	7.192	2.497	8.183	2.022	7.694	2.266	7.009	1.642	7.006
β_2	0.402	10.763	0.526	6.705	0.378	8.031	0.350	7.060	0.561	7.141	0.537	5.340
β_3	–	–	-0.251	5.896	–	–	–	–	–	–	–	–
β_4	–	–	-0.208	5.866	–	–	–	–	–	–	–	–
α_0	0.489	3.620	-0.099	3.635	0.349	7.932	0.085	3.804	0.014	6.863	-0.082	3.452
α_1	0.241	7.344	0.636	3.285	0.365	5.213	0.407	4.879	0.474	4.346	0.698	3.600
α_2	0.050	7.473	0.125	2.645	0.052	5.268	0.115	4.976	0.065	4.441	0.046	2.882
α_3	-0.086	6.593	-0.153	3.245	-0.107	4.411	-0.111	4.187	-0.147	4.800	-0.026	3.234
α_4	0.033	3.620	-0.123	3.497	0.018	4.268	-0.029	3.695	-0.055	4.577	-0.090	2.959
α_5	–	–	0.070	3.635	–	–	0.035	4.639	0.043	3.884	-0.035	3.452
α_6	–	–	0.063	5.229	–	–	0.110	4.934	0.057	6.863	0.120	4.808
α_7	–	–	–	–	–	–	-0.005	3.804	–	–	–	–
α_8	–	–	–	–	–	–	-0.059	5.629	–	–	–	–
γ_0	-0.009	3.365	0.012	3.356	0.089	3.551	0.063	3.748	-0.041	3.328	-0.087	2.939
γ_1	–	–	–	–	-0.020	5.015	–	–	0.034	5.161	–	–
γ_2	–	–	–	–	-0.018	5.099	–	–	-0.018	4.548	–	–
β_{LT}	–	–	–	–	0.012	0.404	0.007	0.335	–	–	–	–
β_1^{co}	0.654	7.333	-0.311	3.712	–	–	-0.101	4.934	-0.115	4.578	–	–
β_2^{co}	0.490	7.344	–	–	0.606	5.214	0.496	4.879	0.508	5.560	0.222	3.907
β_3^{co}	-0.378	7.473	–	–	-0.529	5.269	-0.446	4.976	-0.225	4.347	-0.168	3.600
β_4^{co}	–	–	0.063	3.245	–	–	0.130	4.639	0.148	4.800	–	–
β_5^{co}	–	–	0.234	3.285	-0.114	4.411	-0.205	4.187	–	–	-0.088	2.882
β_6^{co}	–	–	–	–	–	–	–	–	0.112	4.441	0.085	3.234
β_7^{co}	0.177	6.593	–	–	–	–	–	–	-0.078	3.884	–	–
β_8^{co}	–	–	0.079	2.645	-0.163	4.268	-0.111	3.695	–	–	-0.070	2.959
β_9^{co}	–	–	–	–	-0.121	4.234	–	–	–	–	–	–
β_{10}^{co}	–	–	0.064	3.497	–	–	–	–	–	–	–	–
α_{LT}	–	–	–	–	-0.008	7.932	–	–	-0.005	6.863	–	–
α_1^{co}	0.076	3.443	–	–	–	–	-0.067	3.748	-0.048	3.457	–	–
α_2^{co}	–	–	–	–	0.132	3.317	0.108	4.073	–	–	–	–
α_3^{co}	–	–	–	–	-0.036	3.551	-0.097	3.725	–	–	–	–
α_4^{co}	–	–	0.094	3.625	–	–	0.096	3.668	–	–	0.080	2.939
α_5^{co}	–	–	–	–	–	–	–	–	–	–	–	–
α_6^{co}	-0.066	3.365	–	–	-0.118	3.713	-0.096	3.578	-0.078	3.328	–	–
α_7^{co}	–	–	0.088	3.765	0.049	3.762	–	–	–	–	–	–
α_8^{co}	–	–	–	–	–	–	–	–	–	–	–	–
α_9^{co}	–	–	–	–	–	–	–	–	–	–	–	–
$\alpha_{10}^{\mathrm{co}}$	–	–	-0.057	3.356	–	–	–	–	–	–	–	–
n_p	15	–	21	–	22	–	25	–	22	–	17	–
ℓ	-1092.992	–	-795.797	–	-931.254	–	-907.180	–	-839.768	–	-772.777	–
AIC	2215.984	–	1633.594	–	1906.507	–	1864.359	–	1723.538	–	1579.555	–

(1) Standard deviations $\boldsymbol{\sigma}$ are multiplied by 10^{-2}.

Table B.1

Pseudo-optimal automatic parameter selection for the studied locations. Units are in meters for α_i and β_i, and γ is dimensionless.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
β_0	7.302	7.646	7.657	7.645	7.064	7.707	7.719	7.740	7.718	7.719	7.740	7.747	7.731	7.496
β_1	–	2.991	3.192	3.181	3.977	3.255	3.256	3.279	3.211	3.209	3.232	3.242	3.232	3.316
β_2	–	0.276	0.365	0.371	0.339	0.385	0.393	0.409	0.405	0.393	0.402	0.401	0.402	-0.011
α_0	1.079	0.698	0.639	0.631	0.505	0.606	0.534	0.507	0.498	0.491	0.489	0.487	0.489	0.740
α_1	–	–	0.322	0.331	0.392	0.256	0.290	0.265	0.258	0.255	0.241	0.241	0.241	0.221
α_2	–	–	0.095	0.097	0.069	0.085	0.084	0.073	0.066	0.059	0.050	0.049	0.050	-0.276
α_3	–	–	–	-0.106	-0.150	-0.100	-0.078	-0.086	-0.082	-0.081	-0.086	-0.088	-0.086	-0.001
α_4	–	–	–	0.010	0.093	0.005	0.012	0.022	0.024	0.019	0.033	0.026	0.033	-0.324
γ_0	-0.089	-0.053	-0.044	-0.044	-0.086	-0.060	-0.023	-0.019	-0.011	-0.004	-0.009	-0.009	-0.010	-0.190
β_{LT}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_1^{co}	–	–	–	–	–	0.522	0.557	0.592	0.619	0.617	0.654	0.645	0.654	0.800
β_2^{co}	–	–	–	–	–	–	0.551	0.511	0.485	0.496	0.490	0.499	0.489	0.293
β_3^{co}	–	–	–	–	–	–	–	-0.405	-0.396	-0.384	-0.378	-0.377	-0.379	-0.241
β_4^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_5^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_6^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_7^{co}	–	–	–	–	–	–	–	–	–	–	0.177	0.179	0.179	0.337
β_8^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_9^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
β_{10}^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_{LT}	–	–	–	–	–	–	–	–	–	–	–	–	–	0.018
α_1^{co}	–	–	–	–	–	–	–	–	0.079	0.083	0.076	0.076	0.077	-0.114
α_2^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_3^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_4^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_5^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_6^{co}	–	–	–	–	–	–	–	–	–	-0.062	-0.066	-0.063	-0.066	0.116
α_7^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_8^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
α_9^{co}	–	–	–	–	–	–	–	–	–	–	–	–	–	–
$\alpha_{10}^{\mathrm{co}}$	–	–	–	–	–	–	–	–	–	–	–	-0.045	–	–
n_p	3	5	7	9	11	10	11	12	13	14	15	16	16	16
ℓ	-1379.214	-1185.838	-1162.853	-1160.125	-1273.806	-1139.991	-1114.233	-1100.854	-1098.386	-1096.751	-1092.992	-1092.171	-1092.990	-1299.858
AIC	2764.427	2381.677	2339.706	2338.251	2569.613	2299.982	2250.466	2225.709	2222.771	2221.502	2215.984	2216.342	2217.979	2631.716

Table B.2
Automatic parameter selection for 15W55N location.

	15W55N		Azores		Bretagne		Landas		Coruña		Lisbon	
	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$	$\boldsymbol{\theta}$	$\boldsymbol{\theta}^{\text{SFB}}$
β_0	7.740	7.739	4.410	4.424	5.666	5.674	4.431	4.476	4.994	4.984	4.694	4.676
β_1	3.232	3.235	1.970	1.991	2.497	2.506	2.022	2.028	2.266	2.282	1.642	1.653
β_2	0.402	0.406	0.526	0.531	0.378	0.374	0.350	0.340	0.561	0.562	0.537	0.518
β_3	–	–	-0.251	-0.222	–	–	–	–	–	–	–	–
β_4	–	–	-0.208	-0.193	–	–	–	–	–	–	–	–
α_0	0.489	0.490	-0.099	-0.087	0.349	0.353	0.085	0.122	0.014	0.126	-0.082	-0.099
α_1	0.241	0.239	0.636	0.626	0.365	0.352	0.407	0.396	0.474	0.466	0.698	0.626
α_2	0.050	0.051	0.125	0.123	0.052	0.043	0.115	0.107	0.065	0.048	0.046	0.066
α_3	-0.086	–	-0.153	-0.137	-0.107	-0.106	-0.111	-0.116	-0.147	-0.138	-0.026	–
α_4	0.033	–	-0.123	-0.112	0.018	0.023	-0.029	-0.013	-0.055	-0.054	-0.090	–
α_5	–	–	0.070	–	–	–	0.035	–	0.043	–	-0.035	–
α_6	–	–	0.063	–	–	–	0.110	–	0.057	–	0.120	–
α_7	–	–	–	–	–	–	-0.005	–	–	–	–	–
α_8	–	–	–	–	–	–	-0.059	–	–	–	–	–
γ_0	-0.009	0.005	0.012	0.007	0.089	-0.036	0.063	0.036	-0.041	0.012	-0.087	-0.096
γ_1	–	–	–	–	-0.020	–	–	–	0.034	–	–	–
γ_2	–	–	–	–	-0.018	–	–	–	-0.018	–	–	–
β_{LT}	–	–	–	–	0.012	0.012	0.007	-0.049	–	–	–	–
β_1^{co}	0.654	0.656	-0.311	-0.306	–	–	-0.101	-0.100	-0.115	-0.104	–	–
β_2^{co}	0.490	0.499	–	0.061	0.606	0.620	0.496	0.500	0.508	0.529	0.222	0.246
β_3^{co}	-0.378	-0.376	–	–	-0.529	-0.521	-0.446	-0.454	-0.225	-0.235	-0.168	-0.172
β_4^{co}	–	–	0.063	0.077	–	–	0.130	0.131	0.148	0.147	–	0.074
β_5^{co}	–	–	0.234	0.240	-0.114	-0.116	-0.205	-0.227	–	–	-0.088	-0.085
β_6^{co}	–	–	–	–	–	–	–	–	0.112	0.108	0.085	0.068
β_7^{co}	0.177	0.166	–	–	–	–	–	–	-0.078	-0.073	–	–
β_8^{co}	–	–	0.079	0.089	-0.163	-0.164	-0.111	-0.118	–	–	-0.070	-0.050
β_9^{co}	–	–	–	–	-0.121	-0.125	–	–	–	–	–	0.048
β_{10}^{co}	–	–	0.064	0.062	–	–	–	–	–	–	–	–
α_{LT}	–	–	–	–	-0.008	-0.008	–	–	-0.005	–	–	–
α_1^{co}	0.076	0.078	–	–	--	–	-0.067	-0.080	-0.048	-0.063	–	–
α_2^{co}	–	–	–	–	0.132	0.130	0.108	0.109	–	–	–	–
α_3^{co}	–	–	–	–	-0.036	–	-0.097	-0.099	–	–	–	–
α_4^{co}	–	–	0.094	0.093	–	–	0.096	0.084	–	–	0.080	0.072
α_5^{co}	–	–	–	–	–	–	–	–	–	–	–	–
α_6^{co}	-0.066	-0.069	–	–	-0.118	-0.100	-0.096	–	-0.078	-0.050	–	–
α_7^{co}	–	–	0.088	0.091	0.049	0.053	–	–	–	–	–	–
α_8^{co}	–	–	–	–	–	–	–	–	–	-0.005	–	–
α_9^{co}	–	–	–	–	–	–	–	–	–	–	–	–
α_{10}^{co}	–	–	-0.057	–	–	–	–	–	–	–	–	–
n_p	15	13	21	19	22	19	25	20	22	18	17	15
ℓ	-1092.992	-1094.980	-795.797	-797.684	-931.254	-932.084	-907.180	-910.991	-839.768	-841.803	-772.777	-778.429
AIC	2215.984	2215.960	1633.594	1633.368	1906.507	1902.2	1864.359	1865.982	1723.538	1719.605	1579.555	1586.857
n_m	14	240	18	348	19	375	21	456	19	348	15	294
Time[(1)]	5.6	537.5	11.5	1120.1	10.9	1221.8	15.3	1676.9	14.5	1116.3	7.0	781.0

(1) Time in seconds.

Table B.3
Comparison between the proposed method (parameters $\boldsymbol{\theta}$) and the Stepwise Forward selection and Backward elimination (SFB) procedure (parameters $\boldsymbol{\theta}^{\text{SFB}}$).