AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2015JC010952

Key Points:

- Weather patterns classification is used to assess extreme wave climate variability
- Climate-emulator of extreme events

Correspondence to:
A. Rueda,
ruedaac@unican.es

Citation:
Rueda, A., P. Camus, F. J. Méndez, A. Tomás, and A. Luceño (2016), An extreme value model for maximum wave heights based on weather types, *J. Geophys. Res. Oceans*, *121*, doi:10.1002/2015JC010952.

Received 3 MAY 2015
Accepted 14 JAN 2016
Accepted article online 21 JAN 2016

An extreme value model for maximum wave heights based on weather types

Ana Rueda[1], Paula Camus[1], Fernando J. Méndez[1], Antonio Tomás[1], and Alberto Luceño[2]

[1]Environmental Hydraulics Institute IH Cantabria, Universidad de Cantabria, Cantabria, Spain, [2]Dpto de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Cantabria, Spain

Abstract Extreme wave heights are climate-related events. Therefore, special attention should be given to the large-scale weather patterns responsible for wave generation in order to properly understand wave climate variability. We propose a classification of weather patterns to statistically downscale daily significant wave height maxima to a local area of interest. The time-dependent statistical model obtained here is based on the convolution of the stationary extreme value model associated to each weather type. The interdaily dependence is treated by a climate-related extremal index. The model's ability to reproduce different time scales (daily, seasonal, and interannual) is presented by means of its application to three locations in the North Atlantic: Mayo (Ireland), La Palma Island, and Coruña (Spain).

1. Introduction

The need to understand the frequency and intensity of natural hazards and develop resilient, long-term infrastructures has promoted extreme value theory as a relevant discipline for engineering and applied science over the last century. The ultimate goal of this theory is the estimation of the probability of events larger than any on record [*Coles*, 2001].

Traditionally, extreme value theory typically provides a statistical description of the maxima of a stationary process, where stochastic properties are considered constant in time. However, in the context of environmental variables, nonstationarity (e.g., seasonality, interannual variability, long-term trends, etc.) is often found at different time scales [e.g., *Méndez et al*., 2006; *Holthuijsen*, 2007; *Menéndez et al*., 2009]. Many coastal structures have been designed on return levels derived from stationary methods, which assume no change to the frequency of extremes over time [*Klein et al*., 2009]. However, the frequency of extremes is likely to change in response to changes in climate [*IPCC*, 2007; *Milly et al*., 2008]. Therefore, statistical methods that account for the nonstationary behavior of the climate system are needed [e.g., *Parey*, 2010; *Cooley*, 2013; *Salas and Obeysekera*, 2014].

Typically, nonstationary behavior is introduced as a covariate in one or more of the parameters of the extreme value distribution, which are estimated by maximizing the likelihood function. The parameters of an extreme distribution are varied to represent nonstationary processes such as seasonal effects (modeled with harmonic functions, e.g., sine waves), long-term trends (linear and/or exponential terms), and climatic influence (covariates) such as the ENSO variability [*Katz et al*., 2002; *Méndez et al*., 2006]. Other nonstationary methods use neural networks to model the nonlinear behavior of covariates [*Cannon*, 2010]. Changes in the covariates produce changes in the distribution of extremes and thus allow analysis of projected climate variability.

Extreme events can result from the combination of high values of different components. Statistically, longer records result in smaller errors, and furthermore, the record should be long enough to encompass the range of variability in extremes [*Serafin and Ruggiero*, 2014]. This has prompted the development of probabilistic methods to simulate thousands of estimates of wave climates [*Hawkes et al*., 2002; *Méndez et al*., 2006; *Callaghan et al*., 2008; *Menéndez et al*., 2009; *Cannon*, 2010; *Wahl et al*., 2012; *Corbella and Stretch*, 2013]. Most probabilistic methods analyze the dependence between variables of interest without taking into consideration the climate [*Hawkes et al*., 2002; *Wahl et al*., 2012; *Corbella and Stretch*, 2013] or introduce the climate as a covariate which impose, normally, the use of a single variable as main driver due to the complexity involved [*Méndez et al*., 2006; *Callaghan et al*., 2008; *Menéndez et al*., 2009; *Cannon*, 2010].

This paper introduces an extreme value model that relates the nonstationary behavior of extremes to the occurrence probability of associated daily weather patterns. The proposed method is based on the availability of weather-types approaches to successfully downscale the mean [*Camus et al*., 2014] or extremes [*Garavaglia et al*., 2010] of a local variable of interest via large-scale predictors such as sea level pressure fields (SLP). The use of a weather-type approach to estimate return values of a variable of interest, in this case significant wave height, allow analysis of extreme wave climate variability at different time scales such as seasons or years and it may open the possibility to explore a multivariate analysis.

The paper is organized as follows: section 2 provides a background on weather-typing approaches and extreme value theory. Section 3 describes the methodology. Section 4 presents the application of the method to three study sites and analyzes the extreme significant wave height distribution at different time scales (annual and monthly maxima). Wave climate variability is also characterized at one of the locations by using a twentieth century hindcast of monthly significant wave height maxima. Finally, section 5 contains the summary and the conclusions.

2. Background

2.1. Weather-Type Statistical Downscaling

Statistical downscaling (SD) approaches are used to relate large-scale predictors to regional-to-local predictands. SD provides a cheap and efficient alternate to dynamical downscaling (i.e., performing a series of nested simulations). SD has proven to be a useful tool to analyze wave climate at a variety of time scales [*Giorgi et al*., 2001]. Among several SD methods described in *Camus et al*. [2014], we have chosen a weather-type approach, where a discrete number of weather patterns are classified according to synoptic similarity. To our knowledge, SD estimation of extreme daily wave events based on weather types has not been previously explored. In our model, nonstationarity at seasonal and interannual scales is introduced in the time-dependent occurrence probabilities of the weather types.

2.2. Extreme Value Theory

Basic extreme value theory assumes that realizations of a random variable are independent and identically distributed. Under the assumptions of independence and stationarity, the Generalized Extreme Value (GEV) and Generalized Pareto (GP) distributions arise as approximations for block maxima (e.g., annual maximum) and for excesses over a high threshold, respectively [*Katz*, 2013]. For the GEV and GP to hold as asymptotic approximations to the distribution of extremes, temporal independence is not strictly necessary, since a wide range of middle dependence conditions could be equally valid [*Leadbetter*, 1983; *Galambos*, 1987].

In studies of wave climate, GEV models are typically based on annual maxima [*Coles*, 2001; *Beirlant et al*., 2004]. However, employing higher sampling frequencies allows larger sample sizes and thus improved accuracy of the model as long as appropriate treatment of the dependence structure is maintained. Based on this, we analyze the feasibility of sampling daily and correcting the dependence between consecutive days by the use of extremal indices. The extremal index is defined as the inverse of the mean cluster duration [*Coles*, 2001]. Here each cluster is defined as a weather-type (WT). The WT classification, which groups the data according to similar wave-generating meteorological processes, improves the homogeneity of subsamples and, therefore, the hypothesis of "identically distributed" samples. This approach permits the downscaling of nonstationary processes taking into consideration nonlinear relationships by means of changes in the probability of occurrence of the WTs.

3. Methodology

Figure 1 summarizes the daily extreme SD method. The steps of the methodology are:

1. collecting historical data of predictor and predictand and preprocessing the predictor,
2. performing a regression-guided classification following *Cannon* [2012],
3. defining weather types of the synoptic circulation conditions,

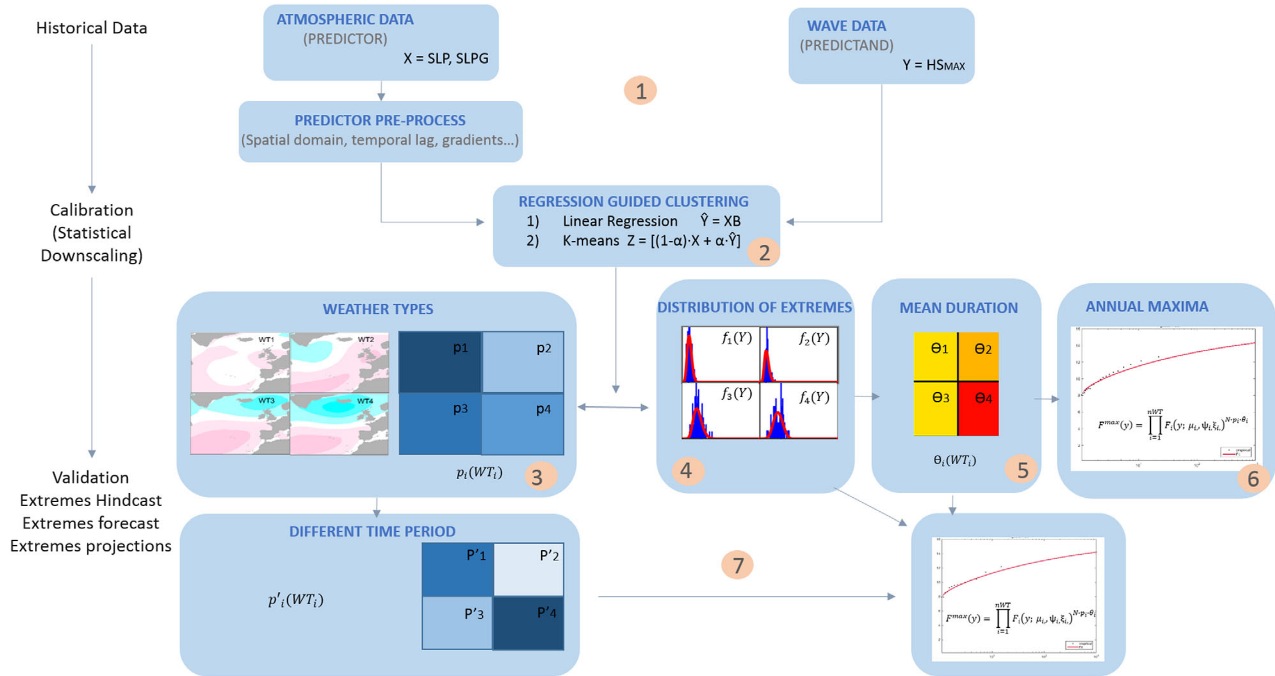

Figure 1. The proposed methodology to obtain a climate-dependent extreme model.

4. fitting a stationary extreme model (e.g., GEV) on the predictand associated to each weather type,
5. obtaining the extremal index associated to each weather type,
6. performing the convolution of the distribution functions of the weather types to obtain associated return periods, and
7. applying the model to different temporal periods.

3.1. Predictor and Predictand Definition

The first step is to obtain historical data of both the atmospheric predictor and predictand needed to ascertain the statistical model that relates them. The spatial and temporal coverage of the predictor is one of the most important consideration of the proposed methodology. Based on previous works [*Camus et al*., 2014; *Perez et al*., 2014] the predictor data must fulfill two main characteristics: (1) the spatial domain should cover the oceanic region responsible for generation of waves arriving at a location of interest and (2) the temporal coverage (recent history) should account for the wave travel time from generation to the target location. The daily predictor is defined as the sea level pressure (SLP) and the square SLP gradients (SLPG), the latter representing the geostrophic wind conditions that are derived from the SLP fields. To account for recent atmospheric conditions, responsible of the swell component, the daily predictor is defined as the mean condition of each day and the n previous days.

3.2. Regression-Guided Classification

A higher skill of the statistical downscaling method for multivariate wave climate has been achieved using a semisupervised clustering algorithm (P. Camus et al., An atmospheric-to-marine synoptic classification for statistical downscaling marine climate, submitted to *Ocean Dynamics* 2015), following the regression-guided approach proposed by *Cannon* [2012]. A better grouping of the predictand is obtained due to a stronger relation of the WTs with the local wave climate. This approach has helped to improve the classification performance also when the predictand is associated with extreme values. The classification is performed as follows: First, the dimensionality of the data is reduced by applying Principal Component Analysis (PCA). Next, a multiple linear regression linking the predictand Y (defined by the sea-state parameter of analysis, the daily maximum significant wave height, Hs) and the predictor X (defined by the PCs that explain 95% of the variance of daily SLP and SLPG fields) is performed. The linear regression is formulated as

$$Y = X \cdot B + E, \tag{1}$$

where B is the matrix of the regression coefficient to be estimated and E is the residual error matrix. The predictions from the fitted model are given by

$$\hat{Y} = X \cdot B. \tag{2}$$

Once the regression model is fitted, the atmospheric data (X) and the predictions of the local waves from the regression model ($\hat{Y}$) are concatenated and weighted using the parameter α. Next, a K-means algorithm is applied to the combined data set

$$Z = \left[(1-\alpha) \cdot X \; + \; \alpha \cdot \hat{Y} \right], \tag{3}$$

where $0 \leq \alpha \leq 1$. The two end-members are $\alpha = 0$, where only the predictor is classified corresponding to unsupervised clustering and $\alpha = 1$, where the method is driven exclusively by the prediction of the predictand, equivalent to fully supervised clustering. A random initialization of the K-means algorithm is used, but in order to ensure enough data to compute the extreme analysis, a minimum number of data points are required in each cluster to accept the classification.

3.3. Weather Types

In the third step, the number of weather types N_{WT} are calculated as the mean of the synoptic circulation conditions included in each cluster of the regression-guided classification (P. Camus et al., submitted manuscript). Each cluster of the classification will have an associated empirical probability of occurrence for the period of study.

3.4. Fitting a GEV for Each WT

In the fourth step, a stationary extreme value model is fit to the sample associated with each WT. GEV theory provides a description of the probability distribution of block maxima of a sample. Although the GEV distribution is typically fit to annual maxima, in this case the application is still appropriate as input data (or local predictand) represent maxima of daily blocks associated with each particular WT. Thus, a stationary GEV is fit for each cluster. The GEV distribution is given by

$$F(y) = \exp\left\{ -\left[1 + \; \xi \left(\frac{y - \; \mu}{\psi} \right) \right]^{\frac{-1}{\xi}} \right\}. \tag{4}$$

The model has three parameters: the location parameter, μ; the scale parameter, ψ; and the shape parameter, ξ. The GEV distribution includes three family types corresponding to the different types of the tail behavior. When $\xi > 0$, GEV corresponds to the Frechet distribution that has a heavy tail decaying polynomially; when $\xi < 0$, the GEV corresponds to the Weibull family that is characterized by a bounded tail; and when $\xi = 0$, the GEV is a Gumbel distribution having an exponentially decaying tail [*Coles*, 2001].

Special consideration is required in the estimation of the shape parameter both because it is usually correlated with the location and scale estimations and because it exhibits a strong influence on the estimate of large (return period) quantiles. Therefore, some analysis is performed to provide a more reliable fit of the distribution. The first step is to determine the suitability of the fitted distribution based on a Chi-squared test and finally a weighted average of the shape parameter with the four immediate neighbors in the PCs space is performed. Therefore, the GEV parameters obtained to each WT are based on the associated historical data, only a smoothing of the shape parameter is performed via its average within similar WTs.

3.5. Climate-Based Extremal Index

The extremal index Θ [*Coles*, 2001] is used to approximately account for the dependence between records of the same sample. In this approach, we have defined a finite number of samples corresponding with the number of WTs. Persistent circulation patterns or WTs may have an interdaily dependence which can be overcome with the climate-based extremal index $\{\Theta_i, i = 1,.,N_{WT}\}$. The extremal index is estimated by calculating the mean duration $\bar{d}_i$ of persistent conditions at each WT, so that the larger the

Table 1. Geographical Characteristics of Each Study Site

Location	Lon (°)	Lat (°)	Water Depth (m)	Distance From Coastline (km)
Mayo	−10.5	54.5	179.08	35
Coruña	−9	43.5	204.24	21
La Palma	−18	29	2707	25

duration of the WT, the larger the dependence among successive observations and smaller the extremal index.

3.6. Monthly and Annual Distributions

The statistical relationship between the predictor and the predictand is established in the sixth step. The cumulative distribution function (CDF) at a monthly or annual scale of the peak sea-state parameter, y, for the whole study period can be inferred as

$$F^{\max}(y)=\prod_{i=1}^{N_{WT}} F_i(y;\mu_i,\psi_i,\xi_i)^{N\cdot p_i\cdot\theta_i}\quad , \tag{5}$$

where $F_i(y;\mu_i,\psi_i,\xi_i)$ is the cumulative distribution function for the corresponding predictand of $\mathrm{WT_i}$, p_i is the probability (monthly or annual) of the ith cluster in the studied period, N is the number of block maxima per month (N = 30 days/month) or per year (N = 365 days/year), and θ_{i} is the extremal index associated to $\mathrm{WT_i}$. Assuming a constant extremal index associated to each WT imposes a mean persistence of each WT. Although with this simplification, certain natural variability is probably not being modeled, it simplifies the simulation and still provides accurate estimates. This methodology bears similarities to the procedure proposed by *Challenor* [1982] using twelve distributions for monthly maxima and by *Morton et al.* [1997] using four seasonal distributions. However, in our method, each weather type has its own occurrence probability p_i, which must be considered when combining the corresponding distributions.

3.7. Using the Model in Different Time Periods

Extreme wave height distributions for different time periods are estimated based on the new probabilities of the clusters $\{p'_i,\ i=1,\ldots,\ N_{WT}\}$ and the corresponding distribution function of the predictand obtained previously with equation (4) for each weather type (or cluster). The model is presented by calculating return levels for annual and monthly maxima and a reconstruction of significant wave maxima over the twentieth century. Since the method presented here is a nonstationary model, the quantiles are time-dependent.

4. Application

The methodology is applied to three regions of the North Atlantic with different wave climates: Coruña, Spain [9°W, 43.5°N], La Palma, Canary island, Spain [18°W, 29°N], and Mayo, Ireland [10.5°W, 54.5°N]. Different wave climates are found among these sites despite their common location in the North Atlantic. The northern locations (Mayo and Coruña) are more exposed to Atlantic storms, thus receiving higher energetic conditions. In addition, the bathymetry and coastline configuration varies at each location, for example the Canary Islands rise from very deep waters where as Ireland has a shallow continental shelf (see Table 1 and Figure 2 for more information of each site).

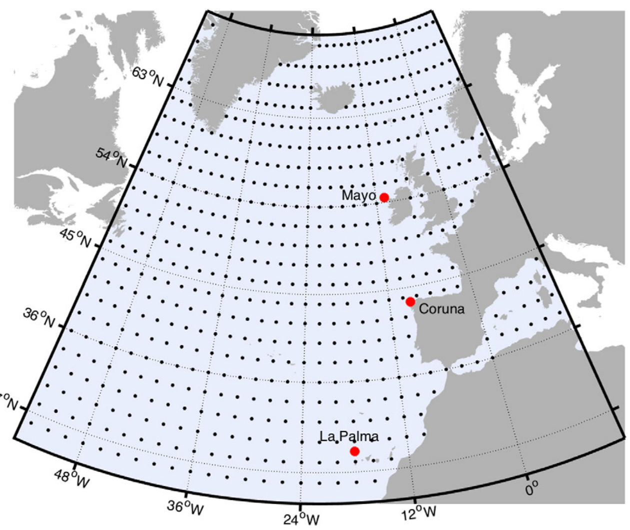

Figure 2. Selected spatial domain of SLP predictor (black points). Red points show study locations (analyzed predictands).

4.1. Data

4.1.1. Predictor

The global SLP fields of the Climate Forecast System Reanalysis (CFSR) [*Saha*, 2010] are used to define the predictor of the SD model. The temporal coverage spans from 1979 to 2013 with hourly temporal resolution and 0.5° spatial resolution.

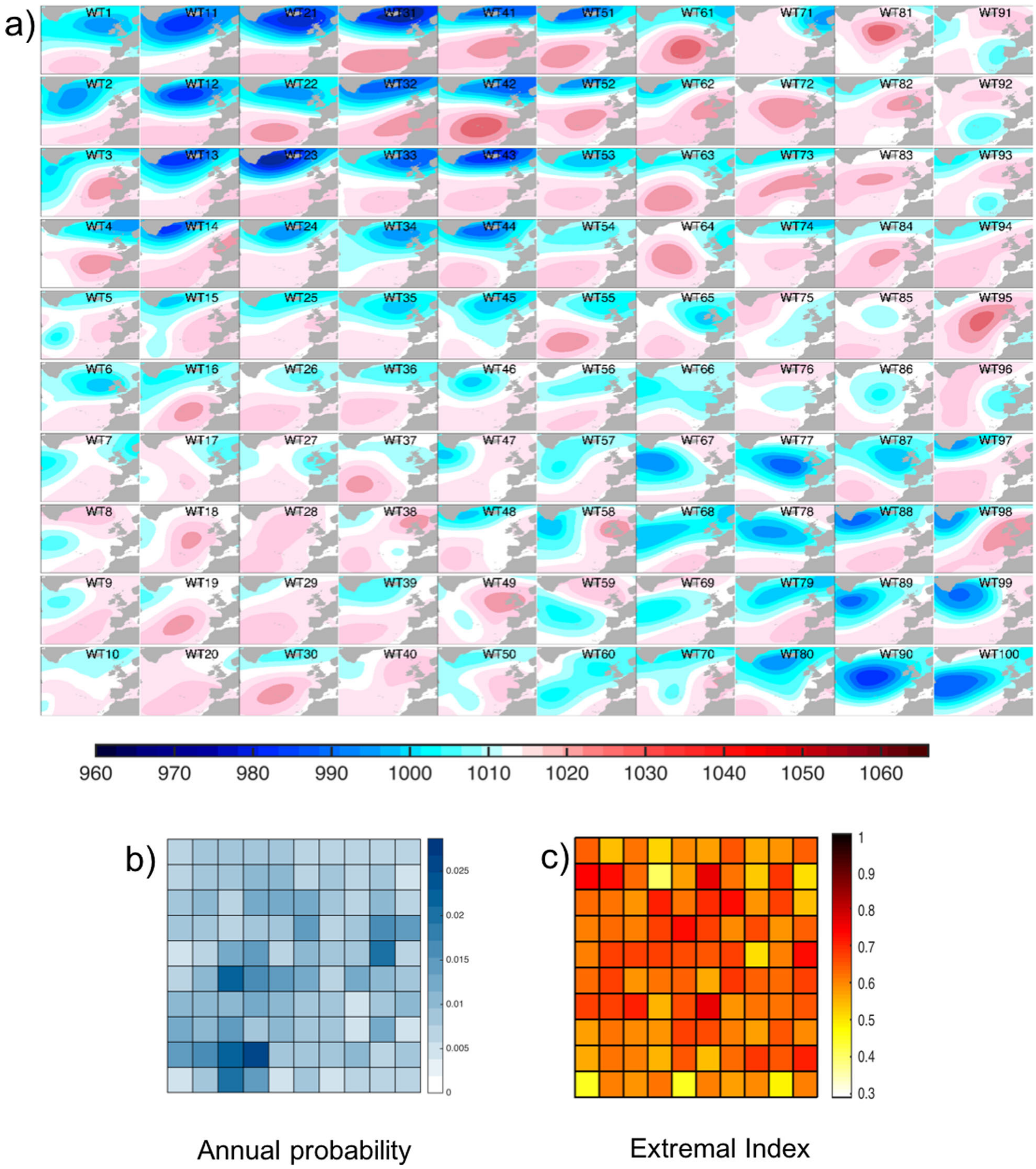

Figure 3. (a) Weather-types (WT) classifications represented by SLP fields (hPa) corresponding to the predictor-to-predictand classification obtained for Mayo, Ireland. (b) Occurrence probability (p_i, in blue scale) and (c) associated extremal index (θ_i, in red scale).

4.1.2. Predictand

Long-term, continuous, and spatially resolved records of data are needed in the construction of SD models. Therefore, wave reanalyses often provide the preferred data source due to their homogeneity and temporal coverage. In this work, the wave hindcast of 1979–2013 by *Perez et al.* [2015] with hourly resolution and 0.125° spatial resolution in the continental shelf provides the historical significant wave height (*Hs*) data.

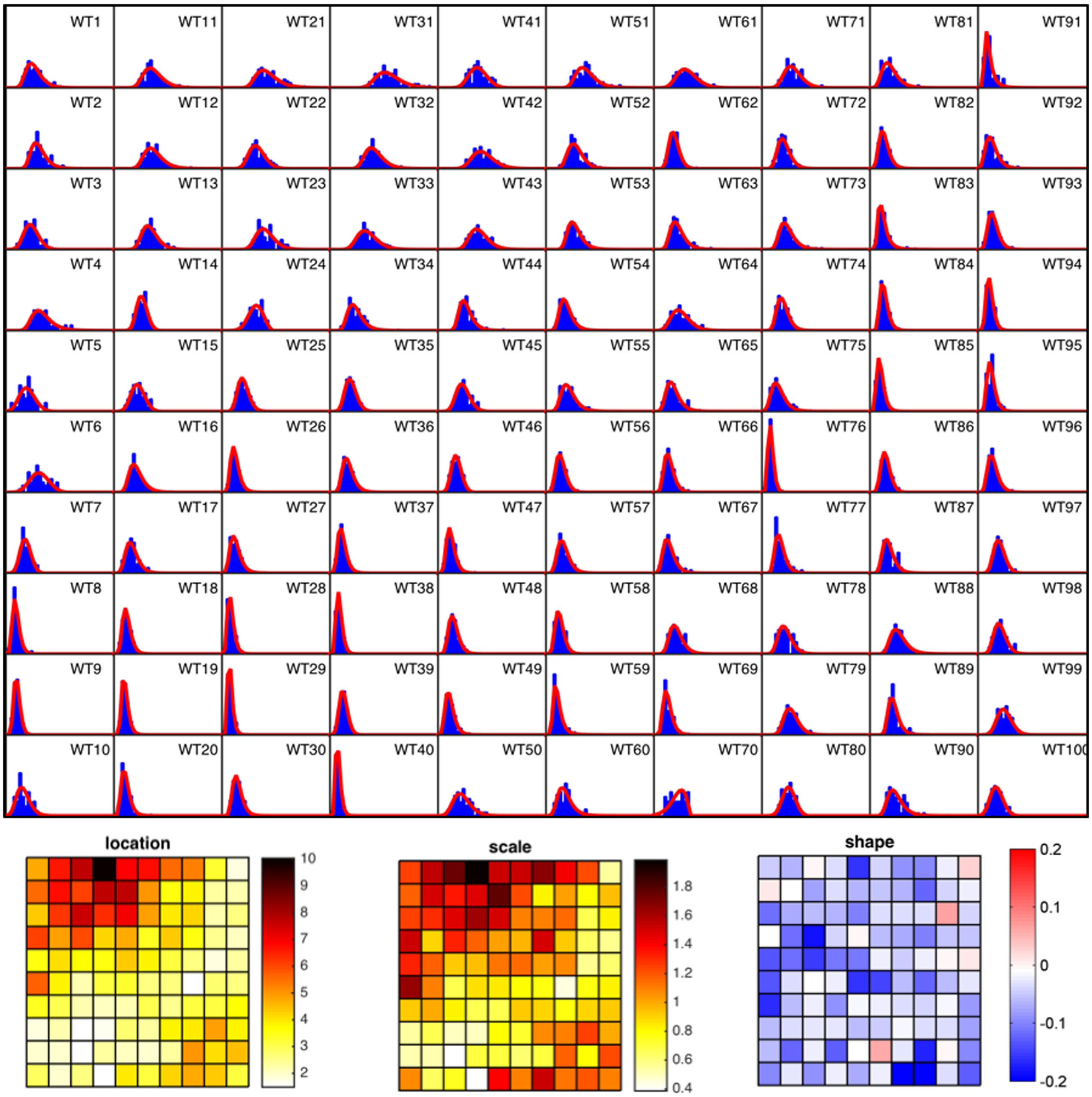

Figure 4. Associated histogram and GEV fitted probability distributions for daily maxima significant wave height in meters at each WT at Mayo, Ireland. The corresponding parameter estimates of each distribution are illustrated in the bottom plots.

4.2. Statistical Downscaling Method for Daily Maxima

Based on the method for "Evaluating the Source and Travel time of the wave Energy reaching a local Area," namely ESTELA [*Perez et al.*, 2014], a common generation area for all the locations is selected. Figure 2 shows the area selected in the North Atlantic basin 24°N–70°N and 54°W–10°E as the spatial domain of the predictor. In order to account for the wave propagation time from generation to destination and based on the ESTELA information, the predictor is defined as the three-daily mean SLP and three-daily mean SLPG. PCA reduces the data redundancy and keeps 95% of the data variance using 95 PCs. The predictand y is the maximum significant wave height (Hs_{max}) every 24 h at the target location.

The weather-type classification is performed at each location. As in *Camus et al.* [2014] a number of $N_{WT} = 100$ synoptic WTs are used. A lesser number of WTs was tested with poorer results and a larger number of WTs diminishes the number of data points at each WT. Although individual classification is obtained

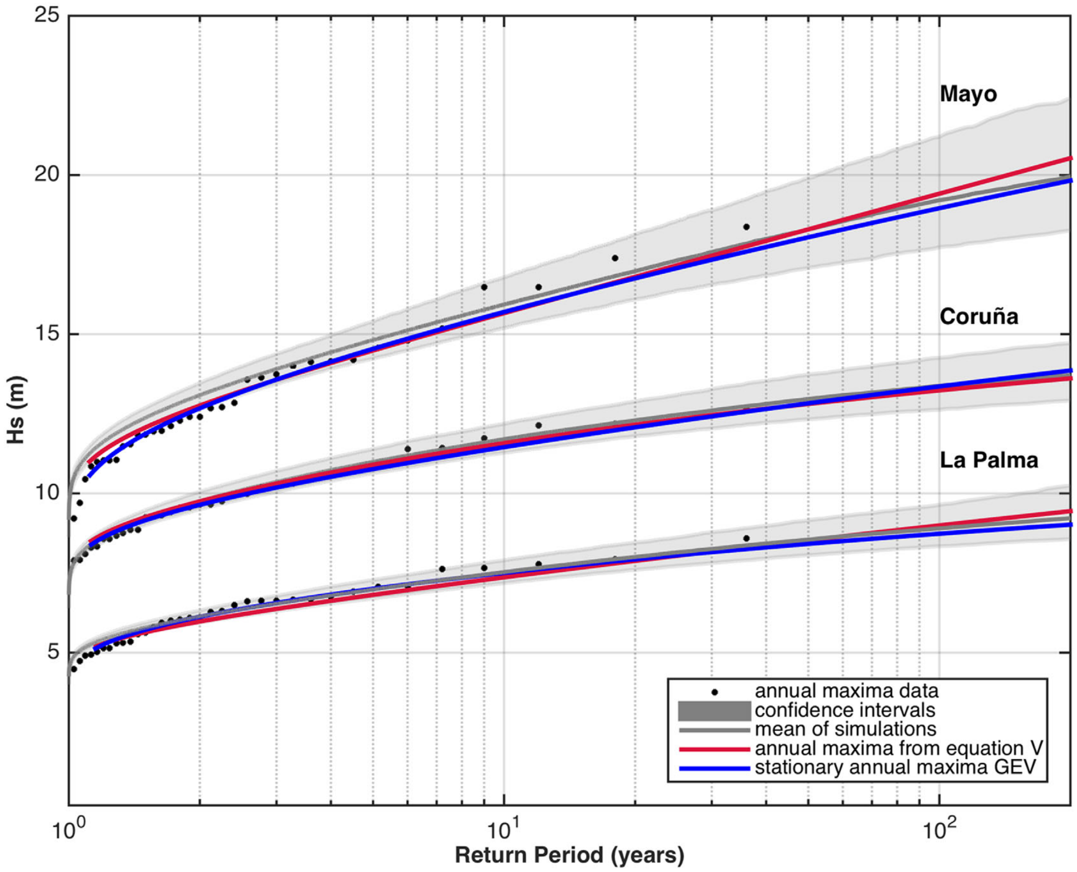

Figure 5. Annual return period significant wave height for the three study sites. The shaded areas represent the 95% confidence intervals of the annual significant wave heights obtained via Monte Carlo simulation based on WTs annual occurrence probabilities during the calibration period (1979–2013).

at each study site, only the classification at Mayo on the Northwest Irish coast is shown. However, the results for each site are discussed in the text.

In Figure 3a, the WTs corresponding to Mayo location are illustrated in a 2-D lattice using a similarity criterion to provide an intuitive visualization of the classification. Low-pressure systems (below an averaged sea level pressure of 1013 hPa) are displayed on blue color scale and high-pressure systems on a red color scale. Two different groups of low-pressure systems are found. The first group, associated with the WTs located on the upper left corner of the lattice, is related to the positive phase of the North Atlantic Oscillation (NAO), which is characterized by intense low-pressure system located over Greenland and the high over the Azores islands. The second group, located on the bottom right corner, exhibits a similar dipole but displaced south-eastward and can be associated with the East Atlantic (EA) positive phase [*Izaguirre et al.*, 2011]. Figure 3b shows the occurrence probability of the WTs during the historical period (1979–2013) in a blue color scale, and Figure 3c the associated extremal index in a red color scale. Lower values of the extremal index are found for the WTs representing anticyclonic situations over the study site, since their mean persistence is usually larger than in storm conditions. Similar weather patterns are found in the other locations, however, different optimal factors (α) for the classification are chosen at each site to obtain the most accurate extreme analysis. The final optimal factors chosen at each site were $\alpha = 0.6$ for Coruña, $\alpha = 0.3$ for Mayo, and $\alpha = 0.2$ for La Palma. The differences between optimal factors are attributable to the geographic location of each study site. In particular, because only one parameter (*Hs*) is used in the regression-guided algorithm the predictand does not provide enough information to infer the origin of the waves. Notably, this effect and the lower values of the optimal factor are found for the Mayo and La Palma sites.

Figure 4 shows the histogram of daily significant wave height maxima, the fitted GEV distribution and the parameter estimates for each WT. In Figure 4, weather types with more intense low-pressure systems, responsible for wave generation in the Atlantic basin correspond to larger values of the location parameter. This pattern

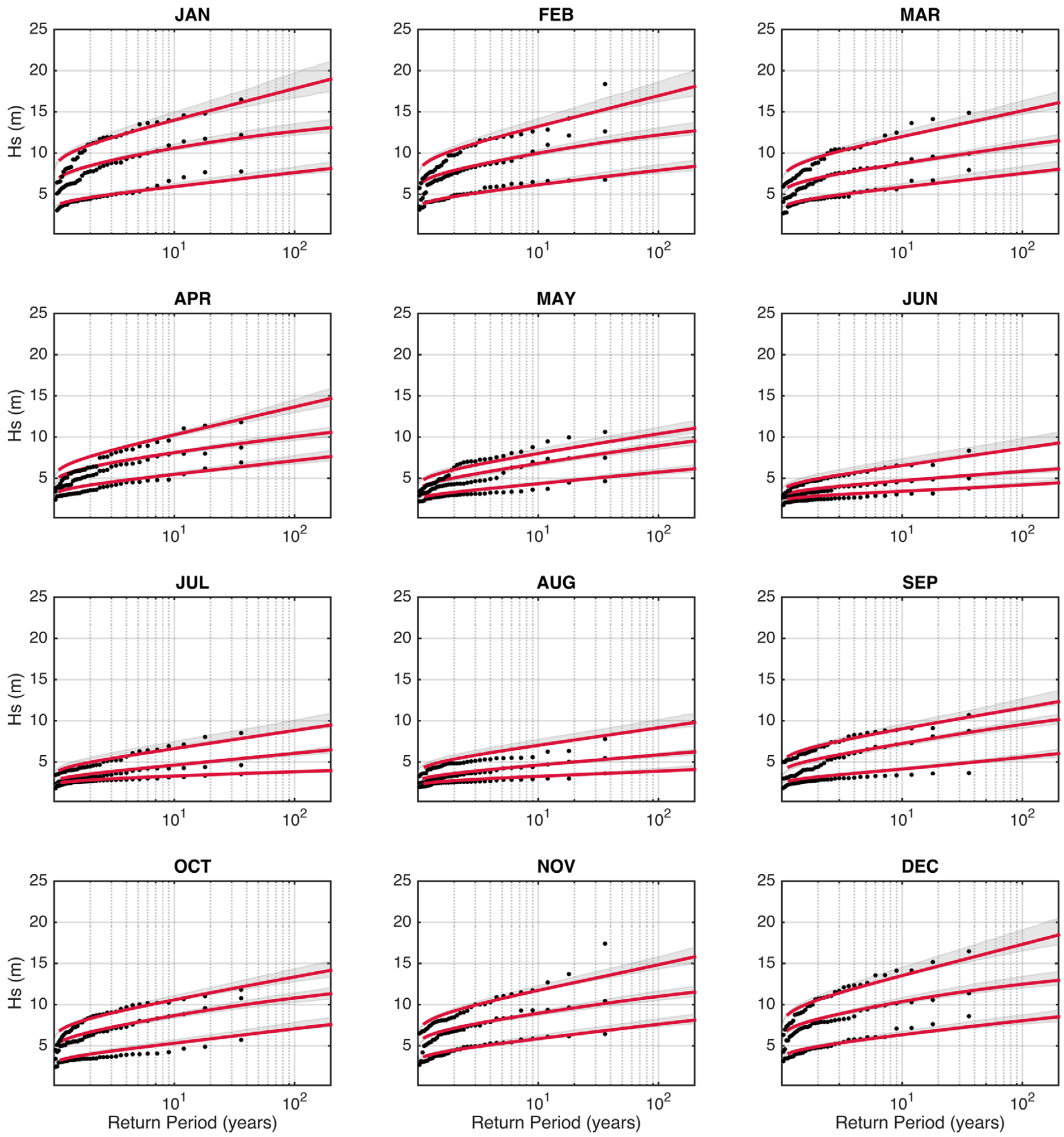

Figure 6. Monthly return period significant wave heights for the three study sites. The shaded area represents the 95% confidence intervals for the annual significant wave heights obtained via Monte Carlo simulation based on WTs monthly occurrence probabilities during the calibration period (1979–2013).

is observed for all the locations. The shape estimate parameter in Mayo takes slightly positive values (Frechet family type), at some WTs that represent high-pressure systems possibly due to the rare occurrence of large waves associated to these WTs. This effect is consistent with typical behavior of summer season [*Menéndez et al.*, 2009]. No positive values of the shape parameter are found at La Palma or Coruña sites for any of the WTs.

Figure 5 shows annual return periods of significant wave heights based on the proposed method (red line) for the three study sites including 95% confidence intervals (shaded areas) obtained via a Monte Carlo simulation. One thousand realizations of 3000 years at daily scale are used to calculate the confidence intervals. In each realization, different values $\{\mu_i,\psi_i,\xi_i,\ i=1,\ldots,\ N_{WT}\}$ are employed considering the statistical distribution of the parameter estimates. In order to account for the interdaily dependence in the Monte Carlo simulation, the parameter estimates associated to each WT are modified according to the associated extremal index [*Coles*, 2001]. Different behavior is observed at the different locations: Coruña and La Palma have bounded tails (Weibull-family behavior) while Mayo has a heavy tail (Frechet-family behavior) and wider confidence intervals. Therefore, for example, the estimates of the 100 year event for the different locations are Mayo 19.2 ($\pm$1.9) m, Coruña 13.37 ($\pm$0.9) m, and La Palma 8.9 ($\pm$0.7) m. For comparison purposes, also

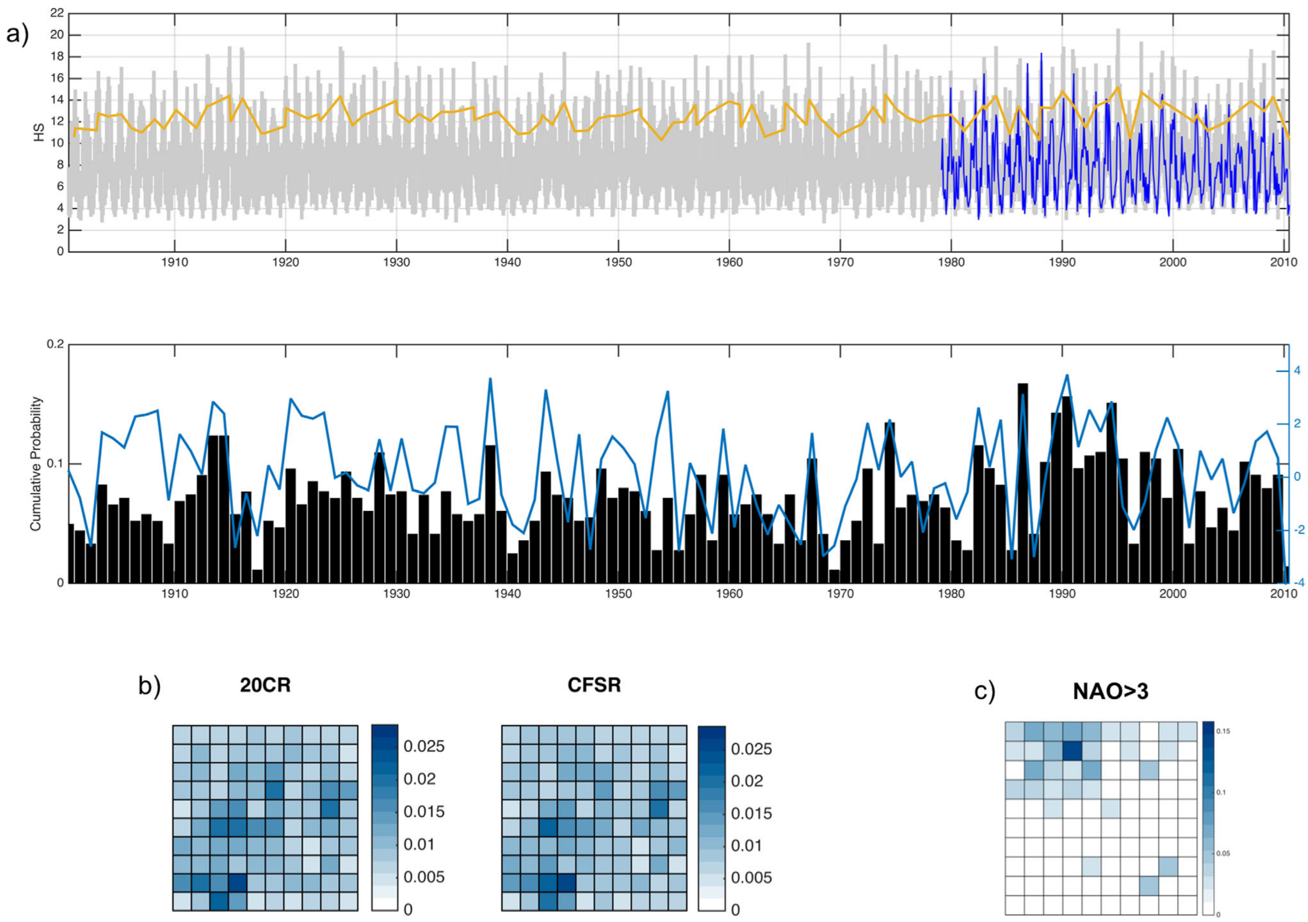

Figure 7. (a) Top: monthly maxima temporal evolution at Mayo, Ireland (blue line) from GOW reanalysis (1979–2013) and 2.5 and 97.5 quantiles of simulated Hs (shaded area) based on monthly WTs occurrence probabilities from 20CR (1950–2010); orange line corresponds to the mean annual simulated maxima. Bottom: WTs annual occurrence probability of those WT associated with large values of NAO index (≥3) and annual NAO index [*Hurrell et al*., 2003]. (b) Probabilities of occurrence for the present conditions (1979–2010) of the WTs (classification shown in Figure 3) from CFSR reanalysis and 20CR reanalysis. (c) WTs occurrence probabilities associated with NAO index ≥3.

a stationary GEV (blue line) is fit to the annual maxima *Hs* (black dots) and shown in Figure 5. The stationary GEV provides similar results to the current WTs method; however, the stationary method is constrained to the available historical data and thus is limited in its application to longer time periods.

The monthly GEV distributions of significant wave height are shown in Figure 6. Larger waves occur at Mayo for every month compared to Coruña and La Palma. To obtain the annual maxima distribution associated to each month, the monthly WTs occurrence probabilities are used for both the annual return period estimation based on equation (5) (red line) as well as for the Monte Carlo simulation. The most severe conditions, with the largest monthly quantiles, occur during the winter months (December, January, and February) for the three locations. The statistical model is able to reproduce the extreme monthly behavior with a few exceptions at Mayo in February or November, where it does not replicate two exceptionally large wave events. In general, the model has poorer skills to reproduce the lower values of the distribution, however on average the upper tail of the distribution is adequately modeled.

4.3. Assessing Climate Variability

The statistical model developed here is able to reproduce extreme wave heights at different time periods outside the calibration period, owing to the statistical relationships established between the local extreme conditions and the atmospheric forcings. As an example, a reconstruction of monthly significant wave height maxima for the whole twentieth century is presented in Figure 7.

The long-term reconstruction of maxima significant wave height is derived from SLP data from the twentieth century atmospheric reanalysis (20CR) [*Compo et al*., 2011], spanning from 1871 to 2010. The use of 20CR reanalysis as a predictor is validated with the comparison of the occurrence rates of the defined WTs for the two reanalysis (20CR and CFSR) for the common period (1979–2010) (Figure 7b). In Figure 7a (top), the historical monthly maxima significant wave height is shown for the wave reanalysis time period (blue), as well as the reconstruction of 95% intervals of simulated monthly maxima based on WTs probabilities obtained with SLP fields from 20CR. Figure 7b shows good agreement between hindcasts and high skill of the model to reproduce wave height maxima outside of the calibration period. Application of the current method to Global Climate Models (GCMs) under different climate scenarios to estimate future extreme events is straightforward. However, due to the large uncertainty associated with GCMs we have focused the current study on the twentieth century.

Variability in atmospheric patterns characterized by the NAO index may have a large influence on wave extremes [*Izaguirre et al*., 2010]. Therefore, in order to test how the model is able to reproduce climate variability, we have analyzed the evolution of the occurrence probability of WTs that are related to the positive phase of NAO [*Hurrell et al*., 2003] at the Mayo location. The selected WTs correspond with those located on the upper left corner of the mesh of Figure 3, which as mentioned in section 4.2 have higher probability of occurrence with positive values of NAO (Figure 7c). Figure 7a (bottom) demonstrates a clear correlation between the occurrence probability of these WTs and the NAO index (0.67 Pearson correlation). Large values of NAO index were found during the 1990s which correspond to larger probabilities of occurrence of the selected WTs and therefore larger values of wave height maxima (Figure 7a, top). Similarly, the probability of occurrence of the selected WTs (and significant wave height maxima) diminishes with lower values of the NAO index. This pattern is perceptible by the simulated mean annual maxima, the orange line of Figure 7a (top). These results are in agreement with previous works [*Bertin et al*., 2013; *Camus et al*., 2014], where the influence of the NAO pattern on mean wave height in northern Europe was investigated.

The analysis performed here reveals the importance of large-scale climatic patterns in the behavior of extreme significant wave heights. In addition, if we calculate the 100 year return period event in the Irish location estimated using the whole twentieth century and only the last 34 years of wave reanalysis data it is found that they differ by almost 1 m (not shown). This difference may be due to the persistence of positive phase of NAO during most of the wave reanalysis time period which coincides with the occurrence of larger wave events. Therefore, it is possible that the largest recorded event at the Mayo location (18.37 m) had a much longer return period that would be expected to occur during the wave reanalysis period.

5. Summary and Conclusions

A statistical model to downscale and analyze the variability of daily maxima of significant wave height is presented. The model is based on the ability of a predictor-to-predictand synoptic classification model to group observations according to similar generating meteorological processes, namely WTs. A stationary extreme value model is fit to each WT. The associated distribution for a certain period of time is obtained via a combination of the GEV distributions of the 100 WTs used here to classify the whole population. Nonstationarity is introduced in the model through the occurrence probability of each WT as a function of time, i.e., probabilities of each WT at a particular month, season or year. The interdaily dependence is treated by a climate-based extremal index, which specifically models the persistence at each particular WT.

The model is applied to three locations in the Atlantic basin with different wave climates. Differences were found in the optimal regression-guided classification, mainly in the factor α, that defines the relative influence of the predictor and predictand in the classification. Some differences were also found in the parameter estimates of the extreme value distributions of each WT, reflecting the particularities of the extreme distributions for each site. The results of the model provide useful information to identify which WTs are related to the more extreme events and to explain the interannual variability. The influence of large-scale patterns, such as those described by the NAO index, has been explored by analyzing the time evolution of the occurrence probabilities of certain WTs. A reconstruction of monthly maxima estimates for the twentieth century has been performed using the 20CR atmospheric reanalysis [*Compo et al*., 2011]. Although it is beyond the scope of this work the importance of understanding large-scale patterns such as NAO is highlighted due to its influence on extreme wave height variability.

The model provides new ways to gain insights about climate variability of extreme events. This work has focused on a univariate model given the complexity and novelty introduced. However, future research will focus on modeling the daily-to-interannual sequence of weather patterns and the extension to multivariate extreme analysis by considering other sea-state parameters such as wave period, direction or 10 m wind intensity.

Acknowledgments
Atmospheric data from CFSR are available online at https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr. Marine data from global reanalysis are provided by IHCantabria and are available by request for research purposes (ihdata@ihcantabria.com). NAO Index Data are provided by the Climate Analysis Section, NCAR, Boulder, USA [*Hurrell et al.*, 2003] and was accessed on 28 October 2015. The work has been partially funded by project "2013/S 122-208379—Assessment of climate impacts on coastal systems in Europe" from the European Commission, JRC, Institute for prospective Technological Studies (IPTS). A.R. and F.J.M. acknowledge the support of the Spanish "Ministerio de Economia y Competitividad" under grant BIA2014-59643-R. The authors would like to thank the anonymous reviewers and S. Vitousek from United States Geological Survey Santa Cruz for their helpful comments that have significantly improved the the final paper.

References

Beirlant, J., Y. Geogebeur, J. Segers, and J. Teugels (2004), *Statistics of Extremes. Theory and Applications*, John Wiley, Chichester, U. K.
Bertin, X., E. Prouteau, and C. Letetrel (2013), A significant increase in wave height in the North Atlantic Ocean over the 20th century, *Global Planet. Change*, *106*, 77–83.
Callaghan, D., P. Nielsen, A. Short, and R. Ranasinghe (2008), Statistical simulation of wave climate and extreme beach erosion, *Coastal Eng.*, *55*(5), 375–390.
Camus, P., M. Menéndez, C. Izaguirre, A. Espejo, V. Cánovas, J. Pérez, A. Rueda, I. J. Losada, and R. Medina (2014), A weather-type statistical downscaling framework for ocean wave climate, *J. Geophys. Res. Oceans*, *119*, 7389–7405, doi:10.1002/2014JC010141.
Cannon, A. J. (2010), A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology, *Hydrol. Processes*, *24*, 673–685, doi:10.1002/hyp.7506.
Cannon, A. J. (2012), Regression-guided clustering: A semi-supervised method for circulation-to-environment synoptic classification, *J. Appl. Meteorol. Climatol.*, *51*, 185–190.
Challenor, P. G. (1982), A new distribution for annual extremes of environmental variables, *Q. J. R. Meteorol. Soc.*, *108*(458), 975–980.
Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, London, U. K.
Compo, G. P., et al. (2011), The twentieth century reanalysis project, *Q. J. R. Meteorol. Soc.*, *37*(654), 1–28.
Cooley, D. (2013), Return periods and return levels under climate change, in *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, vol. 65, edited by A. AghaKouchak, D. Easterling, and K. Hsu, in Chapter 4, Extreme in a Changing Climate, Springer, N. Y.
Corbella, S., and D. D. Stretch (2013), Simulating a multivariate sea storm using Archimedean copulas, *Coastal Eng.*, *76*, 68–78.
Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger, Malabar, Fla.
Garavaglia, F., J. Gailhard, E. Paquet, M. Lang, R. Garcon, and P. Bernardara (2010), Introducing a rainfall compound distribution model based on weather patterns sub-sampling, *Hydrol. Earth Syst. Sci. Discuss.*, *14*, 951–964.
Giorgi, F., B. Hewitson, J. Christensen, M. Hulme, H. Von Storch, P. Whetton, R. Jones, L. Mearns, and C. Fu (2001), Regional climate information: Evaluation and projections, in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the IPCC*, edited by J. T. Houghton et al., chap. 10, pp. 739–768, Cambridge Univ. Press, Cambridge, U. K.
Hawkes, P. J., B. P. Gouldby, J. A. Tawn, and M. W. Owen (2002), The joint probability of waves and water levels in coastal engineering design, *J. Hydraul. Res.*, *40*(3), 241–251.
Holthuijsen, L. H. (2007), *Waves in Oceanic and Coastal Waters*, 387 pp., Cambridge Univ. Press, Cambridge, U. K.
Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (2003), An overview of the north Atlantic oscillation, in The north Atlantic oscillation: Climatic significance and environmental impact, edited by J. W. Hurrell, et al., American Geophysical Union, Washington, D. C. doi:10.1029/134GM01.
IPCC (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Chapter 1, Historical overview of climate change science, edited by S. Solomon, 996 pp., Cambridge Univ. Press, Cambridge, U. K.
Izaguirre, C., F. J. Méndez, M. Menéndez, A. Luceño, and I. J. Losada (2010), Extreme wave climate variability in southern Europe using satellite data, *J. Geophys. Res.*, *115*, C04009, doi:10.1029/2009JC005802.
Izaguirre, C., F. J. Méndez, M. Menéndez, and I. J. Losada (2011), Global extreme wave height variability based on satellite data, *Geophys. Res. Lett.*, *38*, L10607, doi:10.1029/2011GL047302.
Katz, R. W. (2013), Statistical methods for nonstationary extremes, Chapter 2, Extremes in a changing climate: Detection, analysis and uncertainty, vol. 65, edited by A. AghaKouchak, D. Easterling, and K. Hsu, Spinger, N. Y.
Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304, doi:10.1016/S0309-1708(02)00056-8.
Klein, T. A. M. G., F. W. Zwiers, and X. Zhang (2009), Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, *WMO-TD 1500*, 56 pp., World Meteorol. Organ., Geneva, Switzerland.
Leadbetter, M. R. (1983), Extremes and local dependence in stationary sequences, *Z. Wahrsch. Verw. Gebiete*, *65*, 291–306.
Méndez, F. J., M. Menéndez, A. Luceño, and I. J. Losada (2006), Estimation of the long term variability of extreme significant wave height using a time-dependent POT model, *J. Geophys. Res.*, *111*, C07024, doi:10.1092/2005JC003344.
Menéndez, M., F. J. Méndez, C. Izaguirre, and I. J. Losada (2009), The influence of seasonality on estimating return values of significant wave height, *Coastal Eng.*, *56*(3), 211–219.
Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Climate change: Stationarity is dead: Whither water management?, *Science*, *319*(5863), 573, doi:10.1126/science.1151915.
Morton, I. D., J. Bowers, and G. Mould (1997), Estimating return period wave heights and wind speeds using a seasonal point process model, *Coastal Eng.*, *31*(1), 305–326.
Parey, S. (2010), Different ways to compute temperature return levels in the climate change context, *Environmetrics*, *21*, 698–718.
Perez, J., M. Menéndez, F. J. Méndez, and I. J. Losada (2014), ESTELA: A method for evaluating the source and travel-time of the wave energy reaching a local area, *Ocean Dyn.*, *64*, 1025–1038.
Perez, J., M. Menéndez, P. Camus, F. J. Méndez, and I. J. Losada (2015), Statistical multi-model climate projections of surface ocean waves in Europe, *Ocean Modell.*, *96*, 161–170, doi:10.1016/j.ocemod.2015.06.001.
Saha, S., et al. (2010), The NCEP climate forecast system reanalysis, *Bull. Am. Meteorol. Soc*. *91*(8), 1015–1057, doi:10.1175/2010BAMS3001.1.
Salas, J., and J. Obeysekera (2014), Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events, *J. Hydrol. Eng.*, *19*(3), 554–568.
Serafin, K. A., and P. Ruggiero (2014), Simulating extreme total water levels using a time-dependent, extreme value approach, *J. Geophys. Res. Oceans*, *119*, 6305–6329, doi:10.1002/2014JC010093.
Wahl, T., C. Mudersbach, and J. Jensen (2012), Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: A multivariate statistical approach based on Copula functions, *Nat. Hazards Earth Syst. Sci.*, *12*(2), 495–510.