Multivariate statistical modelling of future marine storms

J. Lin-Ye[a,1,*], M. García-León[a], V. Gràcia[a], M.I. Ortego[b], P. Lionello[c], A. Sánchez-Arcilla[a]

[a] *Laboratory of Maritime Engineering, Barcelona Tech, D1 Campus Nord, Jordi Girona 1-3, 08034, Barcelona, Spain*
[b] *Department of Civil and Environmental Engineering, Barcelona Tech, C2 Campus Nord, Jordi Girona 1-3, 08034, Barcelona, Spain*
[c] *Centro Euro-Mediterraneo sui Cambiamenti Climatici, Via Augusto Imperatore, 16, Lecce, Italy*

Abstract

Extreme events, such as wave-storms, need to be characterized for coastal infrastructure design purposes. Such description should contain information on both the univariate behaviour and the joint-dependence of storm-variables. These two aspects have been here addressed through generalized Pareto distributions and hierarchical Archimedean copulas. A non-stationary model has been used to highlight the relationship between these extreme events and non-stationary climate. It has been applied to a Representative Concentration pathway 8.5 Climate-Change scenario, for a fetch-limited environment (Catalan Coast). In the non-stationary model, all considered variables decrease in time, except for storm-duration at the northern part of the Catalan Coast. The joint distribution of storm variables presents cyclical fluctuations, with a stronger influence of climate dynamics than of climate itself.

Keywords: wave storm, Catalan Coast, hierarchical Archimedean copula, generalized Pareto distribution, non-stationarity, generalized additive model

1. Introduction

Extreme events characterization is a key piece of information for an efficient design and construction of any coastal infrastructure. Natural extreme events, such as hurricanes, tsunamis or earthquakes, can lead to considerable economic losses (Shi et al., 2016). From all these hazards, marine storms cause most of the damage to non-seismic coasts. This situation may eventually be aggravated as a consequence of Climate-Change, which affects the intensity and frequency of extreme wave-conditions (Wang et al., 2015; Hemer and Trenham, 2016).

Changes in climate can affect several coastal hazards: flooding (Hinkel et al., 2014; Wahl et al., 2016), erosion (Hinkel et al., 2013; Casas-Prat et al., 2016; Li

*Corresponding author
Email address: jue.lin@upc.edu (J. Lin-Ye)

et al., 2014), harbour agitation (Sánchez-Arcilla et al., 2016; Sierra et al., 2015) and overtopping (Sierra et al., 2016). A robust statistical characterization of storms is, thus, required to assess coastal risks and to forecast storm impacts (Sánchez-Arcilla et al., 2014; Gràcia et al., 2013). The stationary climate assumption, common approach in the last decades for designing infrastructures, does no longer hold valid in a context of Climate-Change. Hence, there is a pressing urge for methodologies that consider non-stationarity, not only in trends, but also in higher statistical moments such as variance.

Usual statistical distributions for extremes such as the Generalized Pareto Distribution (GPD) or the Generalized Extreme Value distribution have three parameters: location, scale and shape. Rigby and Stasinopoulos (2005) proposed a generalized additive model for these three parameters to predict river flow-data from temperature and precipitation on the Vatnsdalsa river (Iceland). Yee and Stephenson (2007) developed a methodology that allows extreme value distributions to be modelled as linear or smooth functions of covariates. One of the examples they presented was the modelling of rainfall in Southwest England. Du et al. (2015) carried out frequency analyses using meteorological variables, where they tested several combinations of co-variates with generalized additive models for location, scale and shape, and concluded that meteorological co-variates improve the characterization of non-stationary return periods. Méndez et al. (2007) used a time-dependent generalized extreme value distribution to fit monthly maxima series of a large historical tidal gauge record, allowing for the identification and estimation of time scale such as seasonality and interdecadal variability. Méndez et al. (2008) extended the former methodology to significant wave-height, while considering the effect of storm duration.

For design purposes, the most analysed variable in marine storms is the significant wave height (H_s), usually considered to be independent from other wave storm-components such as peak-period (T_p), or storm-duration (D). Nevertheless, these variables are known to be semi-dependent (De Michele et al., 2007). Univariate analyses on singular variables, such as H_s, cannot thus describe coastal processes adequately (Salvadori et al., 2014), leading to misestimation of coastal impacts and risks.

The relationship among storm variables can be modelled with statistical techniques such as parametric probability distributions (Ferreira and Soares, 2002), asymptotic theory (Zachary et al., 1998), joint modelling (Bitner-Gregersen, 2015), or copulas (Genest and Favre, 2007; Trivedi and Zimmer, 2007), among other techniques. Copulas were proposed by Sklar (1959), and have recently attracted attention from coastal engineers (Corbella and Stretch, 2012; Salvadori et al., 2015). Wahl et al. (2011) applied fully nested Archimedean copulas to wave storms off the German coast. They first characterized the highest energy point and its intensity and then incorporated the significant wave height. Complementary to these methodologies, Gómez et al. (2016) has implemented a time varying copula to analyse the relationship between air temperature and glacier discharge, which is non-constant and non-linear through time. In this case, both marginal and copula parameters depend on time, and a full Bayesian inference has been applied to obtain these parameters.

Based on this, the present work characterizes the extreme wave climate under a Representative Concentration Pathway 8.5 Climate-Change scenario (RCP8.5, i.e. an increase of the radiative forcing values by year 2100 relative to pre-industrial values of 8.5W/m^2; Stocker et al. (2013)) for a fetch-limited environment (Catalan coast). The study is based on a set of geographical nodes which are equidistant along the Catalan coast. Only eleven nodes out of the total twenty-three are used in this paper, since they represent well the main features and spatial variability of the storm distributions (see Fig. 1, red triangles). Two of the eleven nodes are in intermediate waters, while the rest are in deep waters. The subsequent analysis is performed assuming, first, stationary, and then, transient conditions.

Section 3 describes the methodology and the theoretical background. Section 2 presents the study area. Section 4 lists main results, which are discussed in Section 5. The conclusions are summarized in Section 6.

2. Study area

The Mediterranean Sea (see Fig. 1) is a semienclosed basin, constrained by the European, Asian and African continents. It has a narrow connection to the Atlantic Ocean (Gibraltar Strait), as well as an access to the Black Sea. In terms of waves, the Mediterranean Sea can be splitted into different partitions (Lionello and Sanna, 2005). This paper deals with the Catalan coast, which can be found at the northwestern Mediterranean sector. This area has, as its main morphological features, a) mountain chains which run parallel and adjacent to the coast, b) Pyrenees Mountains to the north, and c) the Ebre river valley to the south. These orographic discontinuities, along with the major river valleys, serve as channels for the strong winds that flow towards the coast (Grifoll et al., 2015).

The most frequent and intense wind in the Catalan Coast is the Tramuntana (north), appearing in cold seasons. It is the major forcing for the northern and central Catalan Coast waves. However, from latitude $41°N$ southward, the principal wind direction is the Mistral (northwest), which is formed by the winds that flow downhill the Pirinees or between the gaps of the mentioned mountains. A secondary wind, the Ponent (west), comes from the depressions in northern Europe. It is the second most frequent one, with limited intensity. Eastern winds are the ones with larger fetch for intense sheer stress, corresponding to low pressure centres over the northwestern Mediterranean. During the summer, there are southern sea-breezes and estern winds, triggered by an intense high-pressure area on the British Islands.

The northwestern Mediterranean Sea is a fetch-limited environment, primarily driven by wind-sea waves (Bolaños et al., 2009; Sánchez-Arcilla et al., 2016). The distance that waves travel, from the storm genesis to the Catalan Coast, is at most one-sixth that of a wave that reaches the Atlantic European coasts (García et al., 1993). Therefore, the corresponding wave-periods, in the northwestern Mediterranean, are much shorter.

The present climate presents a mean significant wave height $\overline{H_s}$ of 0.72m from Barcelona City nortward, and 0.78m southward. Maximum H_s ranges between 5.48m in the southern coast to 5.85m at the northern coast (Sánchez-Arcilla et al., 2008; Bolaños et al., 2009). Casas-Prat and Sierra (2013) projected future wave climate at the Catalan Coast through Regional Circulation Model outputs from the A1B scenario (IPCC, 2000) for the time-period comprising 2071-2100. Their results showed a variation compared to present of the significant wave height around ±10%, whereas the same variable for a 50year return-period exhibits rates around ±20%.

3. Proposed methodology

The methodology here developed leads to a robust assessment of storm pressures under present or future climates. Regional projections are obtained from a deterministic approach, based on the underlying physics, avoiding the computationally expensive dynamical downscaling and the oversimplification of conventional empirical downscaling. Wave storms are first characterized assuming stationarity (see Fig. 2). From here, the joint probability structure is derived and this will serve as a basis for the non-stationary model of the selected projection (in this case, under the RCP 8.5 scenario). A non-stationary model is then built, and constitutes the main part of the proposed methodology, described below.

3.1. Data and storm components

The analysis has been performed considering the wave-climate at the Catalan Coast under a RCP 8.5 Climate-Change scenario. This scenario considers a CO_2 concentration in the atmosphere close to 1250ppm in 2100, which is double that of any other scenario in the Fifth Assessment Report (Stocker et al., 2013). The modelling chain comprises the CMCC-CM (Scoccimarro et al., 2011) Global Circulation Model (see Table 1), providing boundary conditions for the Regional Circulation Model COSMO-CLM (Rockel et al., 2008). The statistical model derived from the CMCC-CM dynamical downscaling has been validated with a total of eighteen Global Circulation Models, shown in Table 1. This list includes models from the same experiment (CMIP5, Taylor et al. (2012)) and from the same Climate-Change-scenario (RCP 8.5), covering, thus, a comprehensive range of predictors. The COSMO-CLM grid, that has a resolution of $0.125° \times 0.125°$, spans the whole Mediterranean region. The next step consists of the WAM (WAMDI Group et al., 1988) wave model, where the just mentioned wind fields serve as an input, for the same domain and spatial resolution. The projections considered in all three models (Global Circulation Model, Regional Circulation Model and WAM), span the interval from year 1950 to 2100.

The nodes considered for the AR5 projections and subsequent analyses (Fig. 1, red triangles) are combined with buoy and SIMAR (Gomez and Carretero, 2005) hindcast points (green rhombuses and black dots, respectively) for validation purposes. All selected nodes (except 1 and 16) are located in deep waters, and thus the WAM model is a suitable option (Larsén et al., 2015). The application of this code to nodes 1 and 16, in intermediate waters, may present certain limitations and would, thus, require further exploration and research. The validation dataset comes from SIMAR hindcasts and Puertos-del-Estado buoy records, corresponding to the period 1990 to 2014. Storms here are clustered into storm-years. Storm-years (called "years", hereafter), which are periods of 12 months, from 1^{st} July to 30^{th} June of the next year.

Four main variables have been selected to describe the storm-intensity conditions: storm energy (E), significant wave-height at the storm-peak (H_p), peak wave-period at the storm-peak (T_p), and duration (D). The E and D are aggregated parameters, related to the total impact of the storm, whereas H_p and T_p represent the maximum intensity of the event. E, H_p, T_p and D take positive real values and, consequently, they have been log-transformed to avoid scale effects (Egozcue et al., 2006).

3.2. Pre-analysis (stationarity assumption)

Prior to the actual modelling, an explanatory analysis has been carried out with the available wave data. A set of stationary models has been built by selecting equidistant time slices from the total sample, following previous work by other authors with similar hydrodynamic variables (Muis et al., 2016; Vousdoukas et al., 2016). The three time-frames are labelled as: (i) past (PT,1950-2000); (ii) present-near-future (PRNF, 2001-2050), and far future (FF, 2051-2100). Storms have been defined using a stationary H_s threshold of $2.09m$ significant wave-height, based on previous work (Lin-Ye et al., 2016). Although the time period in Lin-Ye et al. (2016) is significantly shorter than in the present paper, this threshold should be acceptable for the three time-frames as it falls on the linear part of the excess-over-threshold plot (Fig. 3), according to methodology previously developed by Tolosana-Delgado et al. (2010).

The next step of the pre-analysis consisted in building dependograms of the selected storm variables, which were then visually inspected for non-stationary behaviour. Each variable is also presented in absolute concentration curves (ACC), where ACC1 indicates the ratio of q_{50} at a given time-frame, to the one in the PT inteval (Yitzhaki and Olkin, 1991). ACC2 denotes the same ratio, but with $(q_{75} - q_{50})$. Thus, ACC1 represents on changes in the mean, whereas ACC2 reflects on the evolution of the variance. This analysis has been performed for the energy and duration of the total events of a storm-year, E_{year} and D_{year}, as well as the mean H_s and T_p of a storm-year, $\overline{H}_{s,year}$ and $\overline{T}_{p,year}$, to assess non-stationary trends.

3.3. Stationary model

The probability distribution of each storm variable is fit by a GPD. Being $Y = X - x_0$ the excess of a magnitude X over a location-parameter x_0, conditioned to $X > x_0$, the support of Y is $[0\,,\,y_{sup}]$ (Coles, 2001). y_{sup} is the upper bound of the GPD. The GPD cumulative function is, then,

$$F_Y\left(y|\beta,\xi\right)=1-\left(1+\frac{\xi}{\beta}y\right)^{-\frac{1}{\xi}}\ ,\ 0\leq y\leq y_{sup}, \quad (1)$$

where $\beta\geq 0$ is the scale parameter and $\xi\in\mathbb{R}$ is the shape parameter. As a first approximation, the values of the location parameters x_0 obtained in Lin-Ye et al. (2016) have also been used in this case. The departure from these values is described in Sub-section 4.2.

The Hierarchical Archimedean copula (HAC) is a flexible tool that describes the dependence between variables via the nesting of a subset of 2-D copulas (Sklar, 1959; Nelsen, 2007; Okhrin et al., 2013). The Gumbel type HAC with a mean aggregation method is selected for this case of extreme events, according to Lin-Ye et al. (2016). A d-dimensional Archimedean copula has the form

$$\mathrm{C}\left(\mathbf{F};\phi\right)=\phi^{-1}\left(\phi\left(F_1\right)+\cdots+\phi\left(F_d\right)\right),\quad \mathbf{F}\in\left[0,1\right]^d, \quad (2)$$

for a given generator function ϕ. A Gumbel generator has been selected since it defines the dependence in the upper tail of the probability distribution. Note that a family of asymmetric copulas (Vanem, 2016) would include physical limitations, such as wave steepness, where high H_p cannot commute with large T_p. Due to the complexity of non-stationarity, the asymmetric copulas must be carefully introduced in a more mature future version of the proposed model.

The HAC aggregates the Gumbel generator parameters using a series of coefficients called θ, which can be transformed to Kendall's τ (Kendall, 1937; Salvadori et al., 2011). τ denotes independence when $\tau=0$, and total dependence when τ tends to 1. The goodness-of-fit of the HACs at each time-frame has been assessed by using goodness-of-fit plots of the empirical copulas (Lin-Ye et al., 2016). The κ^2 statistic (Gan et al. (1991)) serves to quantify the goodness-of-fit. It takes values in $[0,1]$, and a perfect fit happens when $\kappa^2=1$. According to our experience in the Catalan Coast, the HAC-structure in Fig. 4 should be applicable to this area. There is another approach for events where H_p is less inter-dependent with E and D (Lin-Ye et al., 2016), but this type of structure is of less interest in this study, as will be discussed later. The nesting levels in Fig. 4 start at the branching of the tree-like structure, and end at the top "root" level.

3.4. Non-stationary model

Extreme events are scarce by nature. The shorter the time-window considered, the smaller will be the available information, with larger uncertainty. This assumption means that, for the time-windows of 50years considered in the stationary model, there are fewer samples of high extreme events. Hence, the probability distribution function's upper tail estimation would not provide results reliable enough. Previous studies indicate that Climate-Change also has a non-negligible effect on extremes (Trenberth and Shepherd, 2015; Hemer and

Trenham, 2016; Du et al., 2015), so assumptions such as a stationary storm-threshold cannot be adopted. This is a first indication that non-stationarity needs to be addressed (Vanem, 2015).

In the non-stationary model, vectorial generalized additive models (VGAM, Yee and Wild (1996)) have been used to determine storminess, storm-thresholds and GPD parameters (Rigby and Stasinopoulos, 2005; Yee and Stephenson, 2007). The VGAM consists of a linear function (Fessler, 1991; Hastie and Tibshirani, 1990):

$$\eta_{i(j)} = \beta^*_{1(j)} + \mathrm{f}_{2(\mathrm{j})}\left(x_{i2}\right) + \ldots + \mathrm{f}_{\mathrm{p(j)}}\left(x_{ip}\right), \tag{3}$$

where $\eta_{i(j)}$ is the j^{th} dependent variable, x_i is the i^{th} independent variable that generates η_i. η_i is a sum of smooth functions of the individual covariates $\beta^*_{1(j)}$ and $f_{p(j)}$. In this case, β^* is not the scale parameter of the GPD. Additive models do all the smoothing in $\mathbb{R}$, avoiding the large bias introduced in defining areas in $\mathbb{R}^n$.

The mathematical assumptions for regression models are: 1) incorrelation, 2) normality, and 3) homoscedasticity of residuals. Assumption 1) is assessed with a ACF plot, assumption 2) can be assessed with a Q-Q plot against a $N\left(0, \sigma^2\right)$ distribution, where the sample standard deviation is used as σ^2. Assumption 3) can be analysed on a graph of fitted value vs. residuals. When the predicted variable is a counting one, a vectorial generalized linear model (VGLM) can be adopted (Yee and Wild, 1996). The VGLM is a particular case of VGAM. The storminess is a counting variable, and its relationship with any other factor can be approximated by a Poisson distribution.

The storm-threshold is then estimated through a VGAM that approximates its relationship with a factor by a Laplace distribution. Once storms are selected, their non-stationary GPD location-parameter x_0 is estimated through quantile regression (Koenker, 2005). The quantile regression is a specific type of VGAM, and it estimates the $100\hat{\tau}\%$ conditional quantile $y_{\hat{\tau}}\left(x\right)$ of a response variable Y as a function $u\left(x, \tau\right)$ of covariates x. The equation $l^*_u = l_u + \varrho_u R_u$ must then be minimized, where $l_u = \hat{\tau} \sum\limits_{i:r_i \geq 0} |r_i| \left(1 - \hat{\tau}\right) \sum\limits_{i:r_i < 0} |r_i|$ for residuals $r_i = y_i - u\left(x_i, \hat{\tau}\right)$. ϱ is a roughness coefficient that controls the trade-off between quality of fit to the data and roughness of the regression function; and R is a roughness penalty (Northrop and Jonathan, 2011; Jonathan et al., 2013). The above mentioned $\hat{\tau}$ has nothing to do with the τ of Kendall. Regarding the rest of the GPD parameters: ξ is assumed to remain constant; β is considered to depend on co-variates, and is estimated with VGLMs.

The option of using time as a covariate is examined in the non-stationary model, just to assess the evolution of other variables. The predicting function is a 4-degree spline (Hastie and Tibshirani, 1990). Alternative predictive parameters seems to present a greater potential. Climate-indices are eligible candidates (Rigby and Stasinopoulos, 2005), for which the linear interpolation function has been selected, advocating the principle of parsimony. Possible climate-indices are the North Atlantic Oscillation (NAO, Hurrell and Deser (2009)), the Easterly Atlantic index (EA, Barnston and Livezey (1987)), the Scandinavian oscillation (SC, Barnston and Livezey (1987)), and their first and second time derivatives. These climate-indices have been scaled to have a mean value equal to zero and a variance equal to unity, and they actually introduce time as an implicit covariate. They were computed from the monthly-averaged sea level pressure fields, from the global circulation-model listed in Table 1. In order to avoid sudden oscillations that would hinder interpretation, the time series of climate-indices have been filtered with a 2^{nd} order lowpass Butterworth filter (Butterworth, 1930), whose low-pass period was of 10years.

Different results among global circulation-models should be expected, despite the same post-processing treatment for all of them. The grid-size and physical implementations are not the same, the model with the highest resolution ($0.76° \times 0.76°$) is CMCC-CM, which is the one that has served as the calibration model. There are also slight divergences on how the model addresses the evolution of emissions (Friedlingstein et al., 2014).

Once storms events have been selected, E, D, H_p and T_p can be extracted. The effect of climate-indices as covariates is assessed at nodes 7 and 21, as these nodes represent the most distinct spatial patterns (see Sec. 2 and Fig. 1). The goodness-of-fit of the resulting VGAM with different combinations of covariates is contrasted with a likelihood-ratio test (LRT, Vuong (1989)), the Akaike information criterion (AIC, Akaike (1987)) and the Bayesian information criterion (BIC, Tamura et al. (1991)). A censorship analysis is carried out on the sample for these two nodes, corresponding to two subsets of GPDs for: a) onshore winds and b) offshore winds. For the two samples in the censorship analysis, and for the combined sample, the proposed model is calibrated with climate-indices derived from the CMCC-CM global circulation-model. The climate-indices from the other eighteen models (Figs. 5, 6, and 7) serve to predict what would be the probability distribution functions under a wide range of plausible values. In the results and discussion section, the 99^{th} quantile, a common quantile for hazard and design (Goda, 2010), has been used to inter compare these.

VGAM uses, thus, global circulation climate-indices as covariates to create time series of 99^{th} quantiles. A way of quantifying how these time series differ from the baseline (CMCC-CM), is by computing the Euclidean distance between the estimated partial autocorrelation coefficients of each time series (Galeano and Peña (2000)). This metric takes values in $[0, 1] \in \mathbb{R}$, being 0 the shortest distance (i.e. closer similarity between models), and 1, the largest one.

Regarding the joint dependence structure of the proposed model, storms are clustered into periods of 15years, under the assumption that there is stationarity in these 15years. Because of the persistence of the climate-indices considered, this is a plausible hypothesis. 15years are also the shortest time-span that provides a sufficient number of storms to determine the HAC structure. Larger time-windows would offer a greater number of storms, but with a non-stationary dependence parameter. Non-stationary HAC dependence parameters are obtained at each node, for this moving time-window of 15years. Each time-window overlaps with the former and the following ones, in half-a-year, to characterize the non-stationary effect.

The Gumbel HAC dependence structure from the stationary-model is also used in the non-stationary model. Particularly, the HAC-structure in Fig. 4 is adopted for the whole non-stationary model. The fitting criteria is the Maximum Likelihood method, where the HAC-structure in the stationary-model (see sub-section 3.3) is set as the unique structure for all nodes and for the whole simulation period. The selection of only one HAC-structure follows the principle of parsimony, being this HAC the one that better characterizes the joint-dependence at most spatial nodes during the three time-frames of the stationary model.

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992) is applied to the dependence-parameters of the HAC, to look into the stationarity of the τ time series. The p-value of such test gives the level of significance at which the null test cannot be rejected. In other words, on how likely the dependence-parameter is actually stationary.

To represent projected climatology, the probability distribution function of the H_p should resemble that of observed storm conditions (from buoys and hindcasts). The proposed model has been validated at the nodes listed on Table 2 (see Figs. 1 for node location), as follows. The SIMAR/buoy data validation nodes are denoted:

$$\{H_{p,1}, \ldots, H_{p,i}, \ldots, H_{p,n}\}, \quad i = 1 \div n, \; n \in \mathbb{R}, \tag{4}$$

and the model data (written as H_p^*, here)

$$\left\{H_{p,1}^*, \ldots, H_{p,j}^*, \ldots, H_{p,m}^*\right\}, \quad j = 1 \div n, \; m \in \mathbb{R} \tag{5}$$

They are next combined to form a joint dataset:

$$\left\{H_{p,1}, \ldots, H_{p,i}, \ldots, H_{p,n}, H_{p,1}^*, \ldots, H_{p,j}^*, \ldots, H_{p,m}^*\right\}$$

Such set is partitioned into four intervals, separated by the quartiles $\{q_0, q_{25}, q_{50}, q_{75}, q_{100}\}$. There are elements from both SIMAR/buoy H_p and AR5 projections, in each interval. The quartiles are selected as boundaries because buoy records are often interrupted due to harsh wave conditions. Then, if the selected intervals are too small, some of them might be empty, which would lead to indetermination of the distance between model and data.

Two vectors are defined as

$$vec_{obs} = \left(\sum_{q_0}^{q_{25}} p\left(H_{p,i}\right), \sum_{q_{25}}^{q_{50}} p\left(H_{p,i}\right), \sum_{q_{50}}^{q_{75}} p\left(H_{p,i}\right), \sum_{q_{75}}^{q_{100}} p\left(H_{p,i}\right)\right), \tag{6}$$

and

$$vec_{model} = \left(\sum_{q_0}^{q_{25}} p\left(H_{p,j}^*\right), \sum_{q_{25}}^{q_{50}} p\left(H_{p,j}^*\right), \sum_{q_{50}}^{q_{75}} p\left(H_{p,j}^*\right), \sum_{q_{75}}^{q_{100}} p\left(H_{p,j}^*\right)\right), \tag{7}$$

where vec_{obs} is the vector for observations, and vec_{model} is the one for projections. Each element of the vector is the summation between two quantiles of the probability distribution function. Therefore, vec_{obs} and vec_{model} are compositional data, their elements being parts of a whole (Egozcue and Pawlowsky-Glahn, 2011), and fulfilling some other properties defined in Aitchison (1982) and Egozcue et al. (2003). The distance between these two vectors can be determined with an Aitchison measure (Aitchison, 1992; Pawlowsky-Glahn and Egozcue, 2001),

$$d(\mathbf{x},\mathbf{y}) = \left| \ln \frac{\mathbf{x}(\mathbf{1}-\mathbf{y})}{\mathbf{y}(\mathbf{1}-\mathbf{y})} \right|, \quad \mathbf{x},\mathbf{y} \in (0,1) \in \mathbb{R}, \tag{8}$$

Where $\mathbf{x}$ and $\mathbf{y}$ are two compared vectors. Another measure for the distance is the Kullback-Leibler divergence (Kullback, 1997)

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}. \tag{9}$$

This function measures the extra entropy of the probability distribution Q of the model, with respect to the probability distribution P of the observations. Note that for any i, $Q(i) = 0$, must imply $P(i) = 0$, to avoid indertemination, thus ensuring that the model considers all the values that the observations show. Also, whenever $P(i) = 0$, the contribution of the i-th term is null, as $\lim_{x\to 0} x \log(x) = 0$.

Both eq. 8 and 9 are distances, and thus take values in $\mathbb{R}_0^+$. The module of the vector is a particular case of both distances (Egozcue and Pawlowsky-Glahn, 2011), and thus both can be compared to the vectorial module, in Euclidean space, of $\mathbf{x}$ and $\mathbf{y}$, which should be of order 1.

4. Results

4.1. Pre-analysis (stationarity assumption)

The dependograms, which do not vary for the different time-frames, show inter-dependence of T_p and the other variables (E, H_p, D), except at node 1 in the FF. ACC1 and ACC2 ratios are represented in Figs. 1 to 3 of the Supplementary material. E and D decrease in PRNF and FF (see Supplementary material, Fig. 1). $ACC1_{H,prnf}$, $ACC1_{H,ff}$, $ACC1_{T,prnf}$ and $ACC1_{T,ff}$ are equal to one for the entire Catalan Coast (figures not shown). $ACC1_{E,prnf}$ is slightly below 1, being specially low in bays or similar local coastal domains. $ACC1_{E,ff}$ is approximately 1.05 in the northern sector (Girona). $ACC1_{D,prnf}$ and $ACC1_{D,ff}$ are high in apexes like the Creus cape (near node 22), and low in bays like the Tarragona one (see Fig. 1). All the ACC2 ratios are slightly below one in the PRNF (see Supplementary material, Fig. 2), and get closer to one in the FF (see Supplementary material, Fig 3). The temporal evolution of E_{year}, $\overline{H}_{s,year}$, $\overline{T}_{p,year}$ and D_{year} are presented in Figs. 4 to 7 of the Supplementary material. The E_{year} are only autocorrelated at node 22 and 12, with a lag of 9years in PT, and are not autocorrelated for larger lags. $\overline{H}_{s,year}$ is autocorrelated at nodes 6, 12, 16, 17, 20, 22 and 23, at different time-frames, and

$\overline{T}_{p,year}$ is autocorrelated along the entire Catalan coast. D_{year} is autocorrelated at node 22, in PT, with a lag of 5years, and at node 1 in PRNF, with a lag of 2years.

4.2. Stationary model

After defining the GPD parameters x_0 and β, each storm-intensity variable is fit by a GPD, of discontinuous support. T_p has required an increase of its location-parameter (10% in FF, at nodes 20 and 22), before fitting GPD. Depending on location, differences may appear within storm-parameters, possibly due to wave propagation effects and the control of land winds at the northermost and southernmost sectors. Unlike for SIMAR hindcasts, the HAC-structure in Fig. 4 is the only one present at all nodes and for all time-frames. The goodness-of-fit of the HAC are represented in Figs. 8 to 10 of the Supplementary material. The k^2 parameter and the graph show a good fit of the Gumbel-HAC, as observed in Lin-Ye et al. (2016).

4.3. Non-stationary model

Two different kinds of non-stationary model have been built: a) using time as the single covariate (NS-T hereafter); and b) implementing large scale climate-indices as covariates (NS-CI hereafter). By using time alone as a covariate to storminess, the storm threshold and GPD parameters, whenever NS-T shows a clear time-dependent behaviour, the non-stationary model NS-CI is applicable. Figures 8, 9, and 10 show the temporal evolution of the HAC dependence-parameters for NS-T. The KPSS test (Kwiatkowski et al., 1992) is applied on τ for the NS-T model, and the outcome is that the null hypothesis of stationarity cannot be rejected in $1-4\%$ of the cases. That is, τ is highly non-stationary.

Regarding storminess, the SIMAR-dataset and the available buoy-records confirm higher storminess-indices (λ) at the northern coast (Figs. 11 and 12). Figure 11 shows that λ decreases with time, but the stationary model can only capture this trend via the predefined time-blocks. This supports using a non-stationary model to improve the representation of the extreme wave-climate. A sensitivity analysis has been carried out on the covariates, at nodes 7 and 21. In the censorship analysis within this sensitivity analysis, the subset with on-shore winds has presented better fit with NAO as covariate, whereas the subset with offshore-winds has done the same with SC. However, an additional test on the rest of nodes has not shown better performance, and for the sake of consistency and parsimony, the uncensored sample has been applied in all nodes. In the uncensored sample, the maximum likelihood estimation indices are smallest for NAO and SC, meaning that these are the covariates that mostly influence λ. The LRT, in turn, denotes that the combination of the two do not provide significantly more information than each of these factors by themselves. What is more, the AIC and the BIC are lowest for the NAO. Therefore, the NAO is selected as the sole covariate for the Poisson-VGAM. Figure 12 shows that λ increases with negative NAO.

NAO, EA, SC (see Figs. 5, 6, and 7) and their first and second derivatives are also used as covariates in the NS-CI VGAM to predict the storm-thresholds and the GPD parameters. The normality and homoscedasticity assumptions of the VGAM (Rigby and Stasinopoulos, 2005) cannot be rejected for the storm-threshold and the GPD parameters x_0 and β. The incorrelation assumption is similarly not rejected for the GPD parameters x_0 and β, but should be rejected for the storm-threshold. The latter non-conformity should be considered when examining the final results.

The statistical model derived from the CMCC-CM (CMCC-A) global circulation-model is, then, compared to the eighteen other models, in the Supplementary material, Figs. 11 to 18 show the similarity of CMCC-CM results to other global circulation-models. For nodes 7 through 23, the distance between each pair of climate-index models is relatively short for most cases, except MIROC-ESM-CHEM (MIR-B) and MIROC5 (MIR-C). The Aitchison and the Kullback-Leibler distances between vec_{obs} and vec_{model} are shown on Table 2. The location-parameters of the GPD are presented in Figs. 13 and 14. τ from the NS-CI HAC-structures are presented in Figs. 15 a 16.

5. Discussion

5.1. Pre-analysis (stationarity assumption)

The decrease in E and D denote loss of energy and duration of storms in future climates. D presents more drastic temporal changes in the northern Catalan Coast. The $ACC2$ increase in the FF, faster than in the PRNF, suggesting that storm-components will present a larger variance over time. $ACC2_E$ does not behave like $ACC2_D$. Possibly, H_p has a certain role in lowering the variance of E. The northward decrease in variance of T_p, observed in Figs. 2 and 3 of the Supplementary material, was also reported for SIMAR hindcasts, in Lin-Ye et al. (2016). This phenomenon occurs when T_p depends heavily on fetch and origin, rather than being a function of wind pulse characteristics.

As for E_{year}, $\overline{H}_{s,year}$, $\overline{T}_{p,year}$ and D_{year} (see Supplementary material, Figs. 4 to 7), E_{year} and $\overline{H}_{s,year}$ fluctuate from PRNF on, whereas they have been considerably stationary in PT (see Supplementary material, Fig. 4 and 5). The general trend in E_{year} is a high in the first quarter of the XXI$^{\text{st}}$ century, followed by approximately 25years of low E_{year}, and another quarter of century of high E_{year}. $\overline{H}_{s,year}$ has a cyclicity of approximately 50years. $\overline{T}_{p,year}$ has the same cyclicity as $\overline{H}_{s,year}$, but it presents stationarity in the PRNF, instead of presenting it in the PT. The time derivatives, dE_{year}/dt, $d\overline{H}_{s,year}/dt$, $d\overline{T}_{p,year}/dt$, dD_{year}/dt fluctuate periodically, but no clear cycles are detectable (not shown here). The reasons behind the clusterings of E_{year}, $\overline{H}_{s,year}$, $\overline{T}_{p,year}$ and D_{year} peaks need further atmospheric analysis (see Sub-section 5.3), but the consequences can be outlined.

D_{year}, behaves similarly to E_{year}. E_{year} becomes less stable from PRNF onward. D_{year} and E_{year} behave similarly, due to the definition of E, which includes D. The low D_{year} and the high E_{year} at the Ebre-Delta in the midst of the XXI$^{\text{st}}$ century may lead to more sediment mobility and a loss of resilience of the area, which is already highly erosive (CIIRC, 2010). The fact that E_{year}

depends more on a summation of small storms than a great one elevates the importance of the smaller storms with 1 to 5years of return period. Low lifetime solutions such as Transient Defence Measures (Sánchez-Arcilla et al., 2016) would be a plausible solution for these periods. What can be expected is that these two seasonal features are not going to be as predictable in the PRNF and FF as in PT, but there are some remarkable periods in the second half of the XXI[st] century, when extreme events are present. From the fluctuations of E_{year}, $\overline{H}_{s,year}$, $\overline{T}_{p,year}$ and D_{year}, it can be perceived that a non-stationary approximation is needed.

5.2. Stationary model

The fact that the HAC-structure in Fig. 4 is predominant in the AR5-projections might be due to H_p being more dependent of E-D in these AR5 projections than in the SIMAR hindcasts (Lin-Ye et al., 2016). This means a remarkable difference between AR5 and SIMAR data. Apparently, the AR5 waves have a lower variability on H_p than the SIMAR data, thus leading to this phenomenon. E and D are averaged values, and a higher correlation can be expected with data that have lower variability values. In other words, SIMAR data might be more heteroschedastic than AR5 data, and this affects the copula definition. Here, the goodness-of-fit of the Gumbel-type HAC with a "mean"-type aggregation-method should be acceptable (see Supplementary material, Figs. 8 to 10).

The dependence of H_p with the subset E-D increases southward due to the proximity of node 1 to the coast (see Fig. 1). The fact that H_p, E and D have milder values in south-Barcelona and in Tarragona (not shown here), indicate that storms in the south are less energetic and durable than at northern locations. Also, E and D is the strongest related components in all storms, so the more energy a storm has, the more time it needs to be dissipated, as expected.

T_p becomes independent from the rest of the variables (E, H_p and D) in the FF. It is observed that, at nodes 1 and 2, E, H_p and D decrease in the second half of the XXI[st] century. However, the time series of T_p does not present any trend. Also, except T_p, the rest of the variables consistently depend on D; as D decreases in the second half of the XXI[st] century, the other variables behave in the same manner. The values of H_p, D and E are closely inter-connected. T_p, on the other hand, is fetch limited, and can hardly surpass 12s, as the most frequent wave direction is related to a fetch of 550km (García et al., 1993; Sánchez-Arcilla et al., 2008), several orders of magnitude lower than Atlantic coasts. The limitation by fetch can also be observed on the H_p data, for all time-frames. The temporal and spatial variability of H_p are greater, however, than those of T_p. The main storm impact is thus reduced to isolated energetic events, with no previous warning nor further replicas. The isolated nature of such events will make storm forecasting a fundamental management tool in the future, based on causal factors, rather than warning signals of the surrounding environment.

5.3. Non-stationary model

The storm-thresholds of the non-stationary model, in all the nodes, fall on the linear part of the excess-over-treshold graphs for PT, PRNF, and FF (see Fig. 3). Therefore, these thresholds are defining extreme events (Tolosana-Delgado et al., 2010).

According to Fig. 12, λ increases with negative NAO. This contradicts Nissen et al. (2014), who stated that positive NAO are more favourable for cyclone intensification, opposite to the findings here. Hence, further research is needed to help revise the relationship between λ and NAO, and since NAO is strongly related to temperature changes, Climate-Change indirectly affects storminess at the Catalan Coast.

In the censorship analysis at nodes 7 and 21, cases with on-shore and off-shore winds have presented better metrics that the general model herein presented. When the model is built with the whole storm sample, the interaction of the covariates leads to more variability among the global circulation-models. This analysis has also reinforced the initial hypothesis that onshore winds are correlated with NAO and offshore winds with SC, which is plausible for the study area. Regarding the uncensored sample, the most influencing covariates for storm-threshold are: NAO, d^2EA, and SC. The covariates mostly affecting the GPD location parameter x_0 of each storm-intensity variable are: dSC for the E; SC for H and T_p; and EA, for D. The most influencing factors on the GPD scale-parameter β of each storm-intensity variable are: d^2EA for the E; d^2EA and d^2SC for H; NAO for T_p, and dSC for D. From all the possible combinations with climate-indices and their time derivatives, the abovementioned covariates have been the ones that presented minimum AIC and BIC, plus lower p-values of LRT. The suitability of these covariates strongly suggests that storms are more affected by the dynamics (sea level pressure gradients) of climate-indices than the climate-indices themselves. In other words, gradients in atmospheric change can lead to an outcome different from that of regular shifts of atmospheric states.

Regarding the 99^{th} quantile in Figs. 11 to 18 of the Supplementary material, both amplitude, phase and trend of the signals present similar patterns in all global circulation-models, although the oscillations do not necessarily coincide among themselves (summarized in Figs. 11 to 18 of the Supplementary material). Stronger disagreement at nodes 1 and 5 can also be understood, because of the strong bimodality that exists on the southern part of the Catalan Coast (García et al., 1993; Grifoll et al., 2016). The Aitchison and Kullback-Leibler distances between vec_{obs} and vec_{model} 2 are of order 1, which is the order of magnitude of the module of the vectors, in all the validating nodes. This indicates that the proposed model has been well validated.

The obtained results do not indicate that Climate-Change is the main contributor to the switch in storm-patterns. It is not certain to what extent this is related to natural variability of large scale indices and how it is affected by the anthropogenic footprint (Trenberth and Shepherd, 2015). Such an explanatory analysis denotes that in this time period, the CMCC-CM global circulation-model presents a climate in which the superposition of both natural variability and greenhouse gases will lead to this change. Regardless of each component's contribution, this information can be useful to tackle problematic seasons in the future.

The trends of the GPD location-parameters of storm-intensity variables (see Figs. 13 and 14) determine their general behaviour. So that where the location-parameters of E, H_p and T_p decrease in time, there should also be a linear decrease of the variables. There is much noise for all variables except T_p. The trends of the GPD location-parameters x_0 of E, H_p, and T_p are either constant or downward. D clearly increases in time at the northern Catalan Coast. This increase may have a relevant impact on harbours, which would require adaptive engineering to face switches in storm-wave patterns and sea-level-rise (Burcharth et al., 2014; Sánchez-Arcilla et al., 2016). Meanwhile, the trend of D is negative at the southern Catalan Coast. The decrease in E has been suggested in Subsection 5.1, but the increase in D at the northern Catalan Coast is a new information that has only been clarified by the non-stationary model.

As for the semi-dependence among storm-components, τ (see Figs. 15 to 16) values are more constant at the north coast than near the Ebre Delta (south coast), where water depths are shallower. That is to say that, wave conditions present more variability in shallower waters. $\tau_{(E,D)}$ has a considerable upward trend at all nodes. This might be explained by a decreasing role of wave-height, and a predominant role of D as the local storm feature. There also seems to be a cyclical variation in dependence among variables, whose cause should be explored in future work. It can also be noted that the peak of $\tau_{((E,D),H)}$ in the period 2000-2050 shows a particular dependence of H_p with respect to D, hinting a concurrence of extreme conditions for wave-height and storm-duration.

6. Conclusions

The extreme wave-climate under a RCP8.5 Climate-Change scenario has been characterised for a fetch-limited environment (Catalan Coast). For this purpose, a non-stationary model for the extreme wave-climate in the period 1950-2100 has been built. The pre-analysis under the stationary assumption provides a first assessment of the AR5 projected storms. It suggests that wave-storms might be dependent on time, stressing the importance of a non-stationary approach. In addition, the stationary model suggests a HAC-structure for this non-stationary approach.

The non-stationary model establishes two types of covariates: a) time and b) climate-indices. The first type indicates the necessity of a non-stationary approach, whereas b) analyses the effects of climate-indices, and their first and second time-derivatives. Storminess appears to depend specially on NAO, as the negative NAO may be associated with storm intensification. Regarding storm-thresholds and the parameters of the GPDs, they are most influenced by the dynamics of climate-indices, rather than by the value of the indices. Location-parameters decrease with time for all variables, except for storm duration (D) at the northern part of the Catalan Coast. HAC dependence-parameters (τ) between storm energy (E) and duration (D) present a considerable upward trend in time. Also, the peak of $\tau_{((E,D),H)}$ in the period 2000-2050 can be translated as a climatic co-existence (under present conditions) of extreme conditions for wave-height (H_p) and storm duration, D.

Funding. This paper has been supported by the European project CEASELESS (H2020-730030-CEASELESS), the Spanish national projects PLAN-WAVE (CTM2013-45141-R) and the MINECO FEDER Funds co-funding CODA-RETOS (MTM2015-65016-C2-2-R). As a fellow group, we would also like to thank the Secretary of Universities and Research of the department of Economics of the Generalitat de Catalunya (Ref. 2014SGR1253, 2014SGR551). The second author acknowledges the Ph.D. scholarship from the Generalitat de Catalunya (DGR FI-AGAUR-14).

Acknowledgements. The support of the Puertos del Estado, in providing the buoy data and the SIMAR model outputs, is also duly appreciated.

Aitchison, J.: 1982, 'The statistical analysis of compositional data'. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 139–177.

Aitchison, J.: 1992, 'On criteria for measures of compositional difference'. *Mathematical Geology* **24**(4), 365–379.

Akaike, H.: 1987, 'Factor analysis and AIC'. *Psychometrika* **52**(3), 317–332.

Barnston, A. G. and R. E. Livezey: 1987, 'Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns'. *Monthly Weather Review* **115**(6), 1083–1126.

Bitner-Gregersen, E. M.: 2015, 'Joint met-ocean description for design and operations of marine structures'. *Applied Ocean Research* **51**, 279 – 292.

Bolaños, R., G. Jorda, J. Cateura, J. Lopez, J. Puigdefabregas, J. Gomez, and M. Espino: 2009, 'The XIOM: 20 years of a regional coastal observatory in the Spanish Catalan coast'. *Journal of Marine Systems* **77**, 237–260.

Burcharth, H. F., T. L. Andersen, and J. L. Lara: 2014, 'Upgrade of coastal defence structures against increased loadings caused by climate change: A first methodological approach'. *Coastal Engineering* **87**, 112 – 121.

Butterworth, S.: 1930, 'On the theory of filter amplifiers'. *Wireless Engineer* **7**(6), 536–541.

Casas-Prat, M., K. L. McInnes, M. A. Hemer, and J. P. Sierra: 2016, 'Future wave-driven coastal sediment transport along the Catalan coast (NW Mediterranean)'. *Regional Environmental Change* pp. 1–12.

Casas-Prat, M. and J. Sierra: 2013, 'Projected Future Wave Climate in the NW Mediterranean Sea'. *Journal of Geophysical Research: Oceans* **118**, 3548–3568.

CIIRC: 2010, 'Estat de la zona costanera a Catalunya'. (Departament de Política Territorial i Obres Públiques) Generalitat de Catalunya.

Coles, S.: 2001, *An introduction to Statistical modeling of extreme values*. Springer.

Corbella, S. and D. Stretch: 2012, 'Multivariate return periods of sea storms for coastal erosion risk assessment'. *Natural Hazards and Earth System Sciences* **12**, 2699–2708.

De Michele, C., G. Salvadori, G. Passoni, and R. Vezzoli: 2007, 'A multivariate model of sea storms using copulas'. *Coastal Engineering*.

Du, T., L.-H. Xiong, C.-Y. Xu, C. J. Gippel, S. Guo, and P. Liu: 2015, 'Return period and risk analysis of nonstationary low-flow series under climate change'. *Journal of Hydrology* **527**, 234 – 250.

Egozcue, J. and V. Pawlowsky-Glahn: 2011, 'Evidence information in Bayesian updating'. In: *Proceedings of the 4th International Workshop on Compositional Data Analysis*. pp. 1–13.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelo-Vidal: 2003, 'Isometric logratio transformations for compositional data analysis'. *Mathematical Geology* **35**(3), 279–300.

Egozcue, J. J., V. Pawlowsky-Glahn, M. I. Ortego, and R. Tolosana-Delgado: 2006, 'The effect of scale in daily precipitation hazard assessment'. *Natural Hazards and Earth System Sciences* **6**(3), 459–470.

Ferreira, J. and C. G. Soares: 2002, 'Modelling bivariate distributions of significant wave height and mean wave period'. *Applied Ocean Research* **24**(1), 31 – 45.

Fessler, J.: 1991, 'Nonparametric fixed-interval smoothing with vector splines'. *Signal Processing, IEEE Transactions on* **39**(4), 852–859.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti: 2014, 'Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks'. *Journal of Climate* **27**(2), 511–526.

Galeano, P. and D. Peña: 2000, 'Multivariate Analysis in Vector Time Series'. *Resenhas do Instituto de Matemática e Estadística da Universidade de Sao Paulo* **4**(4), 383–403.

Gan, F., K. Koehler, and J. Thompson: 1991, 'Probability Plots and Distribution Curves for Assessing the Fit of Probability Models'. *The American Statistician* **45**(1), 14–21.

García, M., A. Sánchez-Arcilla, J. Sierra, J. Sospedra, and J. Gómez: 1993, 'Wind waves off the Ebro Delta, NM Mediterranean'. *Marine Systems* **4**, 235–262.

Genest, C. and A.-C. Favre: 2007, 'Everything you always wanted to know about copula modeling but were afraid to ask'. *Journal of hydrologic engineering* **12**(4), 347–368.

Goda, Y.: 2010, *Random Seas and Design of Maritime Structures*, Vol. Advanced Series on Ocean Engineering Vol. 33. World Scientific, 3rd edition.

Gomez, M. and J. Carretero: 2005, 'Wave forecasting at the Spanish coasts'. *Journal of Atmospheric and Ocean Science* **10**(4), 389–405.

Gómez, M., M. Concepción Ausín, and M. Carmen Domínguez: 2016, 'Seasonal copula models for the analysis of glacier discharge at King George Island, Antarctica'. *Stochastic Environmental Research and Risk Assessment* pp. 1–15.

Gràcia, V., M. García, M. Grifoll, and A. Sánchez-Arcilla: 2013, 'Breaching of a barrier under extreme events. The role of morphodynamic simulations'. *Journal of Coastal Research* **65**, 951–956.

Grifoll, M., A. L. Aretxabaleta, and M. Espino: 2015, 'Shelf response to intense offshore wind'. *Journal of Geophysical Research: Oceans* **120**(9), 6564–6580.

Grifoll, M., J. Navarro, E. Pallares, L. Ràfols, M. Espino, and A. Palomares: 2016, 'Ocean–atmosphere–wave characterisation of a wind jet (Ebro shelf, NW Mediterranean Sea)'. *Nonlinear Processes in Geophysics* **23**(3), 143–158.

Hastie, T. J. and R. J. Tibshirani: 1990, *Generalized additive models*, Vol. 43. CRC Press.

Hemer, M. A. and C. E. Trenham: 2016, 'Evaluation of a CMIP5 derived dynamical global wind wave climate model ensemble'. *Ocean Modelling* **103**, 190 – 203.

Hinkel, J., D. Lincke, A. T. Vafeidis, M. Perrette, R. J. Nicholls, R. S. J. Tol, B. Marzeion, X. Fettweis, C. Ionescu, and A. Levermann: 2014, 'Coastal flood damage and adaptation costs under 21st century sea-level rise'. *Proceedings of the National Academy of Sciences* **111**(9), 3292–3297.

Hinkel, J., R. Nicholls, R. Tol, Z. Wang, J. Hamilton, G. Boot, A. Vafeidis, L. McFadden, A. Ganopolski, and R. Klein: 2013, 'A global analysis of erosion of sandy beaches and sea-level rise: An application of DIVA'. *Global and Planetary Change* **111**, 150–158.

Hurrell, J. W. and C. Deser: 2009, 'North Atlantic climate variability: The role of the North Atlantic Oscillation'. *Journal of Marine Systems* **78**(1), 28 – 41.

IPCC: 2000, 'Summary for policymakers. Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel of Climate Changefried'. Technical report, IPCC.

Jonathan, P., K. Ewans, and D. Randell: 2013, 'Joint modelling of extreme ocean environments incorporating covariate effects'. *Coastal Engineering* **79**, 22 – 31.

Kendall, M.: 1937, 'A new measure of rank correlation'. *Biometrika* **6**, 83–93.

Koenker, R.: 2005, *Quantile Regression*, Econometric Society Monographs. Cambridge University Press.

Kullback, S.: 1997, *Information theory and statistics*. Courier corporation.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin: 1992, 'Testing the null hypothesis of stationarity against the alternative of a unit root'. *Journal of Econometrics* **54**(1), 159 – 178.

Larsén, X. G., C. Kalogeri, G. Galanis, and G. Kallos: 2015, 'A statistical methodology for the estimation of extreme wave conditions for offshore renewable applications'. *Renewable Energy* **80**, 205 – 218.

Li, F., P. van Gelder, J. Vrijling, D. Callaghan, R. Jongejan, and R. Ranasinghe: 2014, 'Probabilistic estimation of coastal dune erosion and recession by statistical simulation of storm events'. *Applied Ocean Research* **47**, 53 – 62.

Lin-Ye, J., M. Garcia-Leon, V. Gracia, and A. Sanchez-Arcilla: 2016, 'A multivariate statistical model of extreme events: An application to the Catalan coast'. *Coastal Engineering* **117**, 138 – 156.

Lionello, P. and A. Sanna: 2005, 'Mediterranean wave climate variability and its links with NAO and Indian Monsoon'. *Climate Dynamics* **25**, 611–623.

Méndez, F. J., M. Menéndez, A. Luceño, and I. J. Losada: 2007, 'Analyzing monthly extreme sea levels with a time-dependent GEV model'. *Journal of Atmospheric and Oceanic Technology* **24**(5), 894–911.

Méndez, F. J., M. Menéndez, A. Luceño, R. Medina, and N. E. Graham: 2008, 'Seasonality and duration in extreme value distributions of significant wave height'. *Ocean Engineering* **35**(1), 131–138.

Muis, S., M. Verlaan, H. C. Winsemius, J. C. Aerts, and P. J. Ward: 2016, 'A global reanalysis of storm surges and extreme sea levels'. *Nature Communications* **7**.

Nelsen, R.: 2007, *An introduction to copulas*. Springer Science & Business Media.

Nissen, K., G. Leckebusch, J. Pinto, and U. Ulbrich: 2014, 'Mediterranean cyclones and windsttorms in a changing climate'. *Regional Environmental Change* **14**(5), 1873–1890.

Northrop, P. J. and P. Jonathan: 2011, 'Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights'. *Environmetrics* **22**(7), 799–809.

Okhrin, O., Y. Okhrin, and W. Schmid: 2013, 'On the structure and estimation of hierarchical Archimedean copulas'. *Journal of Econometrics* **173**, 189–204.

Pawlowsky-Glahn, V. and J. J. Egozcue: 2001, 'Geometric approach to statistical analysis on the simplex'. *Stochastic Environmental Research and Risk Assessment* **15**(5), 384–398.

Rigby, R. A. and D. M. Stasinopoulos: 2005, 'Generalized additive models for location, scale and shape'. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(3), 507–554.

Rockel, B., A. Will, and A. Hense: 2008, 'The Regional Climate Model COSMO-CLM (CCLM)'. *Meteorologische Zeitschrift* **17**, 347–348.

Salvadori, G., C. De Michele, and F. Durante: 2011, 'On the return period and design in a multivariate framework'. *Hydrology and Earth System Sciences* **15**, 3293–3305.

Salvadori, G., F. Durante, G. Tomasicchio, and F. D'Alessandro: 2015, 'Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering'. *Coastal Engineering* **95**, 77–83.

Salvadori, G., G. Tomasicchi, and F. d'Alessandro: 2014, 'Practical guidelines for multivariate analysis and design in coastal and off-shore engineering'. *Coastal Engineering* **88**, 1–14.

Sánchez-Arcilla, A., M. García, and V. Gràcia: 2014, 'Hydro-morphodynamic modelling in Mediterranean storms–errors and uncertainties under sharp gradients'. *Nat. Hazards Earth Syst. Sci.* **14**, 2993–3004.

Sánchez-Arcilla, A., D. Gónzalez-Marco, and R. Bolaños: 2008, 'A review of wave climate and prediction along the Spanish Mediterranean coast'. *Nat. Hazard. Earth Sys.* **8(6)**, 1217–1228.

Sánchez-Arcilla, A., J. P. Sierra, S. Brown, M. Casas-Prat, R. J. Nicholls, P. Lionello, and D. Conte: 2016, 'A review of potential physical impacts on harbours in the Mediterranean Sea under climate change'. *Regional Environmental Change* pp. 1–14.

Scoccimarro, E., S. Gualdi, A. Bellucci, A. Sanna, P. G. Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra: 2011, 'Effects of Tropical Cyclones on Ocean Heat Transport in a High-Resolution Coupled General Circulation Model'. *Journal of Climate* **24**(16), 4368–4384.

Shi, P., X. Yang, J. Fang, J. Wang, W. Xu, and G. Han: 2016, 'Mapping and ranking global mortality, affected population and GDP loss risks for multiple climatic hazards'. *Journal of Geographical Sciences* **26**(7), 878–888.

Sierra, J. P., I. Casanovas, C. Mösso, M. Mestres, and A. Sánchez-Arcilla: 2016, 'Vulnerability of Catalan (NW Mediterranean) ports to wave overtopping due to different scenarios of sea level rise'. *Regional Environmental Change* **16**(5), 1457–1468.

Sierra, J. P., M. Casas-Prat, M. Virgili, C. Mösso, and A. Sánchez-Arcilla: 2015, 'Impacts on wave-driven harbour agitation due to climate change in Catalan ports'. *Natural Hazards and Earth System Sciences* **15**(8), 1695–1709.

Sklar, A.: 1959, *Fonctions dé repartition à n dimension et leurs marges*. Université Paris 8.

Stocker, T., D. Qin, G.-K. Plattner, M. tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley: 2013, 'IPCC, 2013: Summary for Policymakers'. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sánchez-Arcilla, A., M. García-León, V. Gracia, R. Devoy, A. Stanica, and J. Gault: 2016, 'Managing coastal environments under climate change: Pathways to adaptation'. *Science of The Total Environment* pp. –.

Tamura, Y., T. Sato, M. Ooe, and M. Ishiguro: 1991, 'A procedure for tidal analysis with a Bayesian information criterion'. *Geophysical Journal International* **104**(3), 507–516.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl: 2012, 'An Overview of CMIP5 and the Experiment Design'. *Bulletin of the American Meteorological Society* **93**(4), 485–498.

Tolosana-Delgado, R., M. Ortego, J. Egozcue, and A. Sánchez-Arcilla: 2010, 'Climate change in a Point-over-threshold model: an example on ocean-wave-storm hazard in NE Spain'. *Advances in Geosciences* **26**, 113–117.

Trenberth, K.E., F. J. and T. Shepherd: 2015, 'Attribution of climate extreme events'. *Nature Climate Change*.

Trivedi, P. K. and D. M. Zimmer: 2007, *Copula modeling: an introduction for practitioners*. Now Publishers Inc.

Vanem, E.: 2015, 'Non-stationary extreme value models to account for trends and shifts in the extreme wave climate due to climate change'. *Applied Ocean Research* **52**, 201–211.

Vanem, E.: 2016, 'Joint statistical models for significant wave height and wave period in a changing climate'. *Marine Structures* **49**, 180–205.

Vousdoukas, M. I., E. Voukouvalas, A. Annunziato, A. Giardino, and L. Feyen: 2016, 'Projections of extreme storm surge levels along Europe'. *Climate Dynamics* pp. 1–20.

Vuong, Q.: 1989, 'Likelihood ratio tests for model selection and nonnested hypothesis'. *Econometrica* **57**(2), 307–333.

Wahl, T., J. Jensen, and C. Mudersbach: 2011, 'A multivariate statistical model for advanced storm surge analyses in the North Sea'. *Coastal Engineering Proceedings* **1**(32), 19.

Wahl, T., N. G. Plant, and J. W. Long: 2016, 'Probabilistic assessment of erosion and flooding risk in the northern Gulf of Mexico'. *Journal of Geophysical Research: Oceans* **121**(5), 3029–3043.

WAMDI Group, S. H., K. Hasselmann, P. Janssen, G. Komen, L. Bertotti, P. Lionello, A. Guillaume, V. Cardone, J. Greenwood, M. Reistad, L. Zambresky, and J. Ewing: 1988, 'The WAM model: a third-generation ocean wave predicition model'. *Journal of Physical Oceanography* **18**, 1775–1810.

Wang, X. L., Y. Feng, and V. R. Swail: 2015, 'Climate change signal and uncertainty in CMIP5-based projections of global ocean surface wave heights'. *Journal of Geophysical Research: Oceans* **120**(5), 3859–3871.

Yee, T. W. and A. G. Stephenson: 2007, 'Vector generalized linear and additive extreme value models'. *Extremes* **10**(1), 1–19.

Yee, T. W. and C. J. Wild: 1996, 'Vector Generalized Additive Models'. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(3), 481–493.

Yitzhaki, S. and I. Olkin: 1991, 'Concentration indices and concentration curves'. *Stochastic orders and decision under risk* pp. 380–392.

Zachary, S., G. Feld, G. Ward, and J. Wolfram: 1998, 'Multivariate extrapolation in the offshore environment'. *Applied Ocean Research* **20**, 273–295.

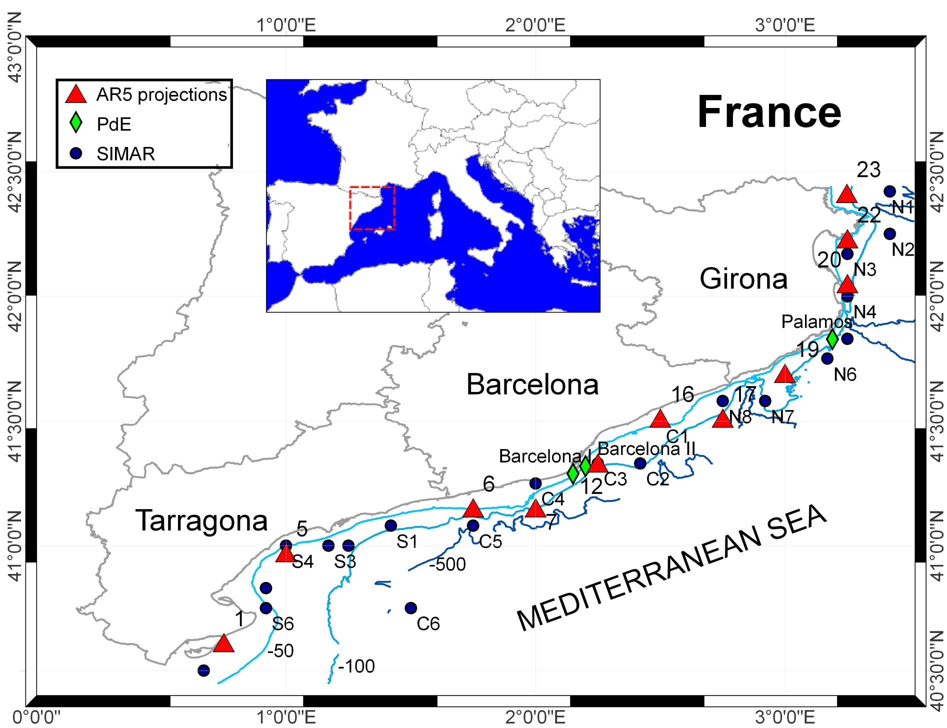


Figure 1: Map of the Catalan Coast, area located in the northwestern Mediterranean. The bathymetry is in meters, showing how all nodes where the proposed model applies (AR5 nodes) are in deep water, except nodes 1 and 16. AR5 nodes are represented by red triangles, buoy (PdE) nodes are green rhombuses, and SIMAR nodes are solid black points.

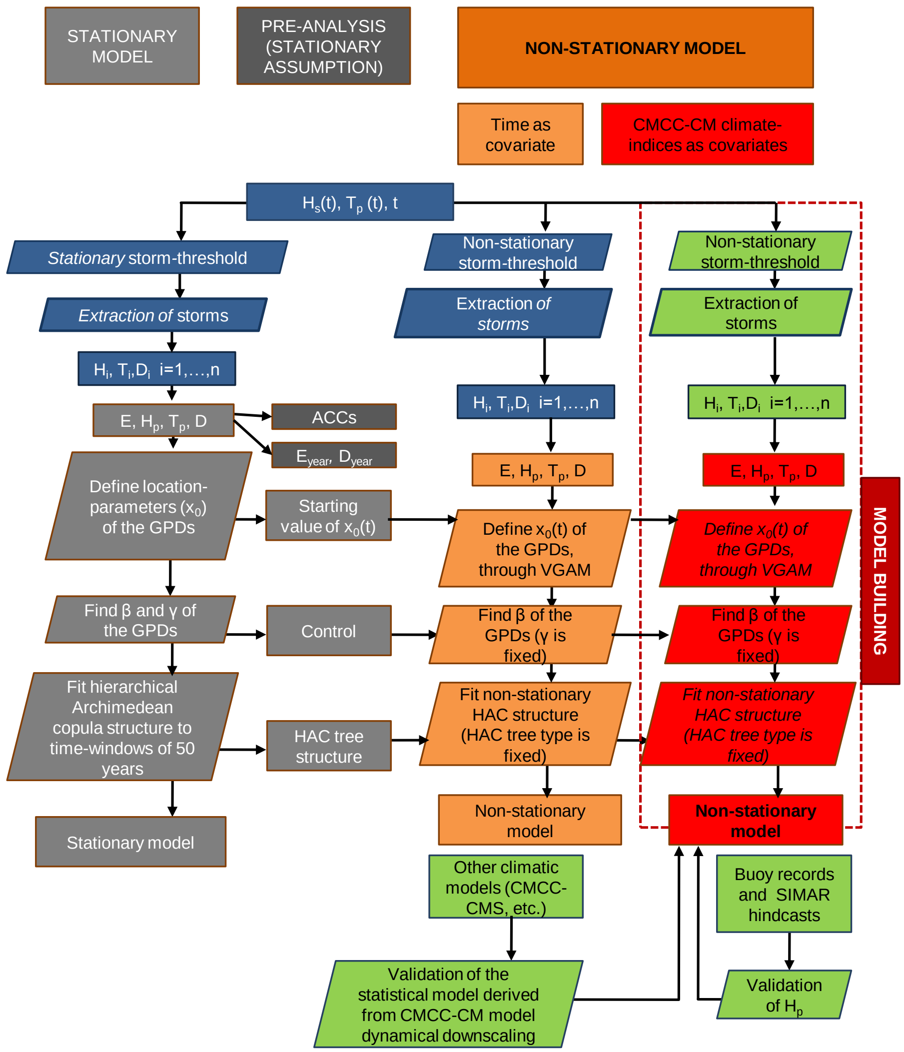
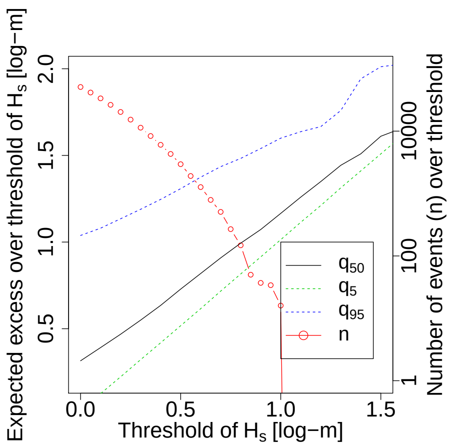
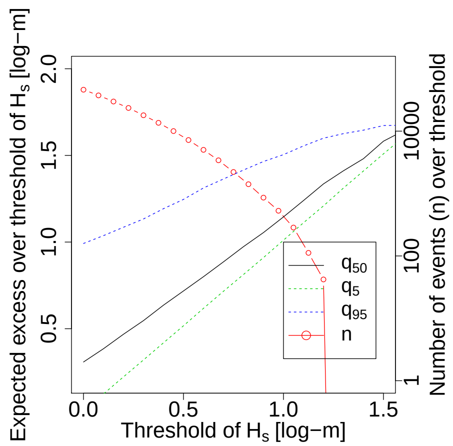


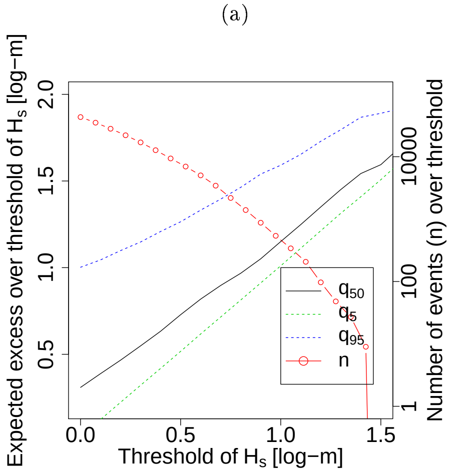
Figure 2: Flow-chart of the methodology applied in this paper.



(a)



(b)



(c)

Figure 3: Excess-over-threshold plots at node 12, in a) past (PT), b) present-near-future (PRNF), and c) far-future (FF) time frames. The red line denotes the number of events (n) over the threshold.

843

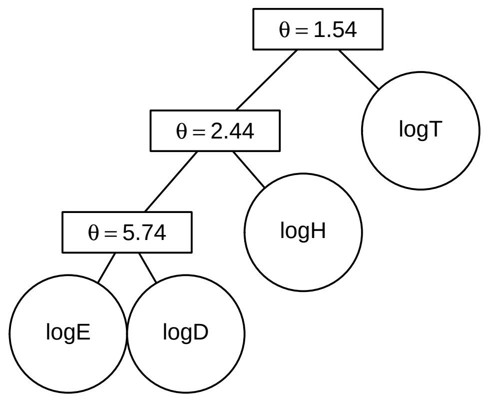


Figure 4: Example of HAC-structure, at node 12, in past (PT). The circles enclose the analysed storm variables, and the θ is the HAC-dependence-parameter.

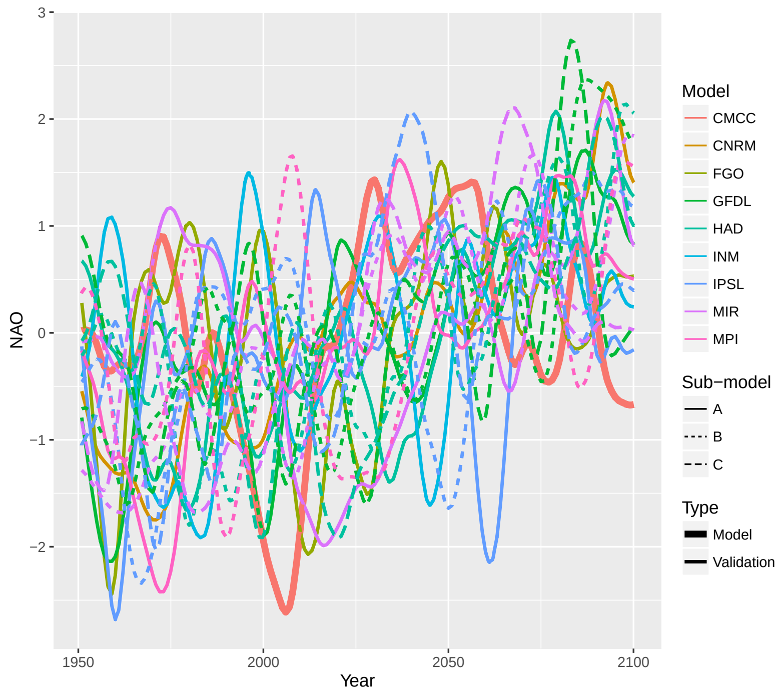


Figure 5: Temporal evolution of NAO index from the global circulation-model monthly outputs (see Table 1). NAO is represented by an adimensional index, scaled to have a mean value equal to zero and a variance equal to unity.

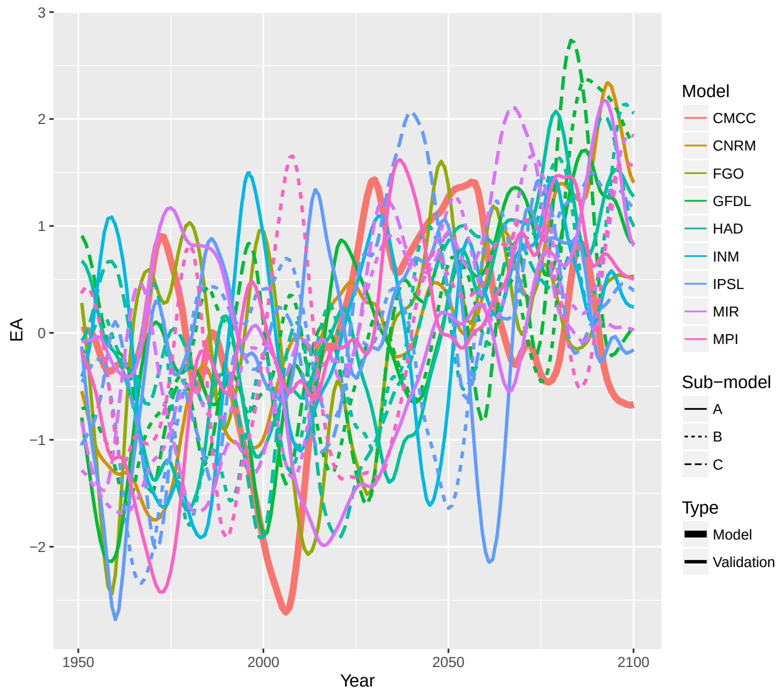


Figure 6: Temporal evolution of EA index from the global circulation-model monthly outputs (see Table 1). EA is represented by an adimensional index.

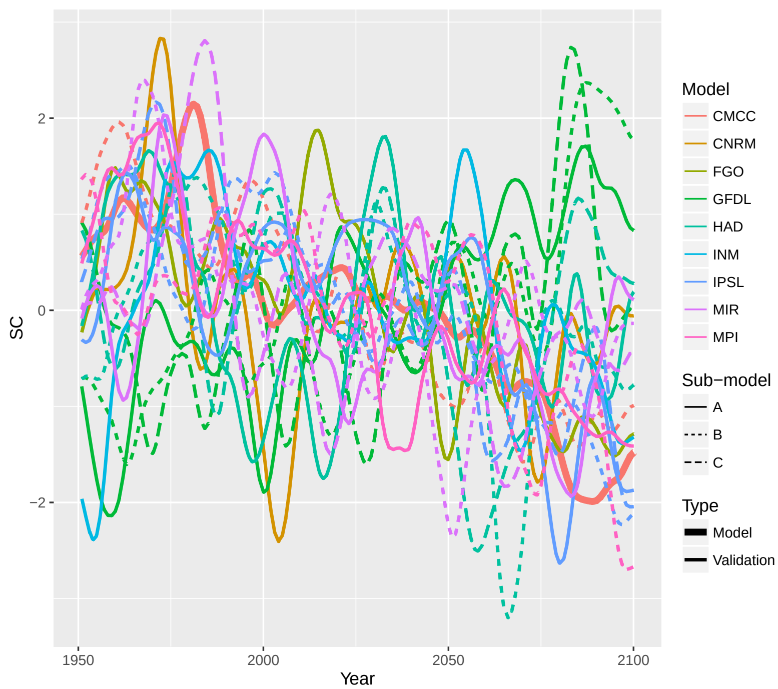


Figure 7: Temporal evolution of SC index from the global circulation-model monthly outputs (see Table 1). SC is represented by an adimensional index.

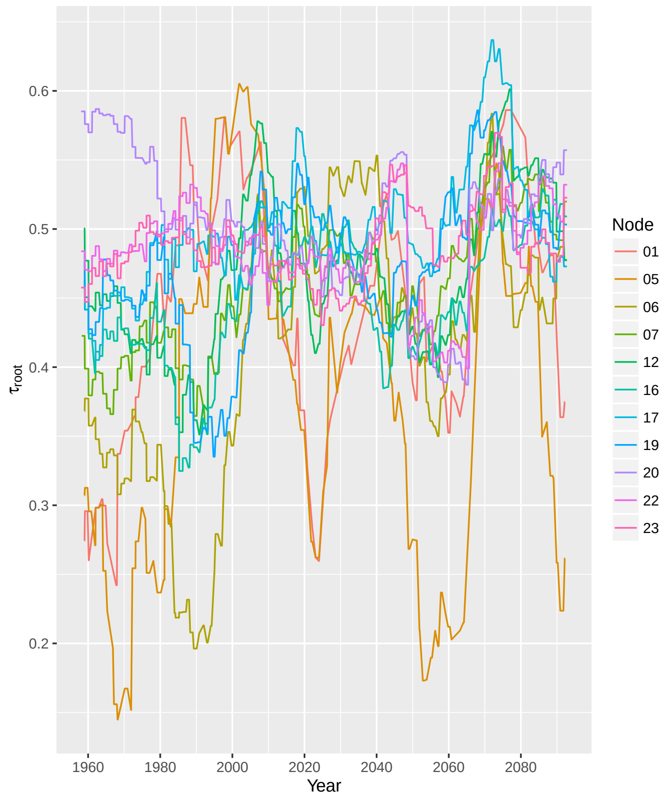


Figure 8: Non-stationary τ_{root} dependence parameter (Kendall, 1937) at the root nesting level of the HAC structure. The marginal distributions are fitted with the VGAM, with time as the sole covariate (NS-T). The colours represent different nodes.

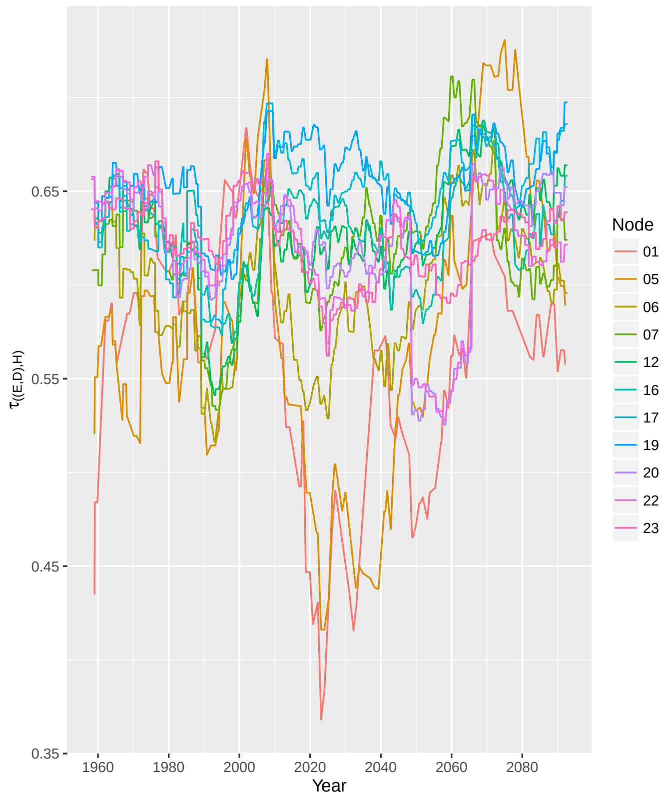


Figure 9: Non-stationary $\tau_{((E,D),H)}$ dependence parameter at the$((E,D),H)$ nesting level of the HAC structure. The marginal distributions are fitted with the VGAM, with time as the sole covariate (NS-T).

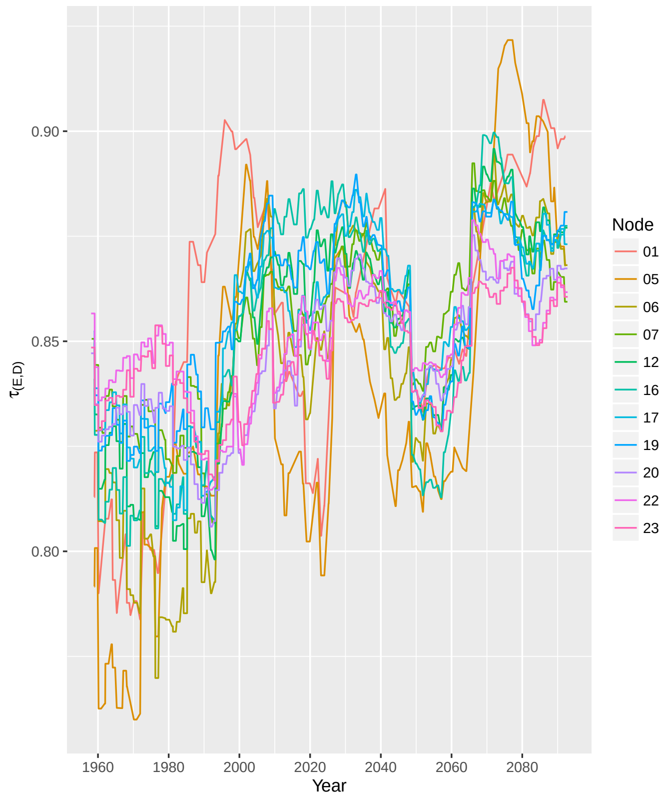


Figure 10: Non-stationary $\tau_{((E,D))}$ dependence parameter at the (E,D) nesting level of the HAC. The marginal distributions are fitted with the VGAM, with time as the sole covariate (NS-T).

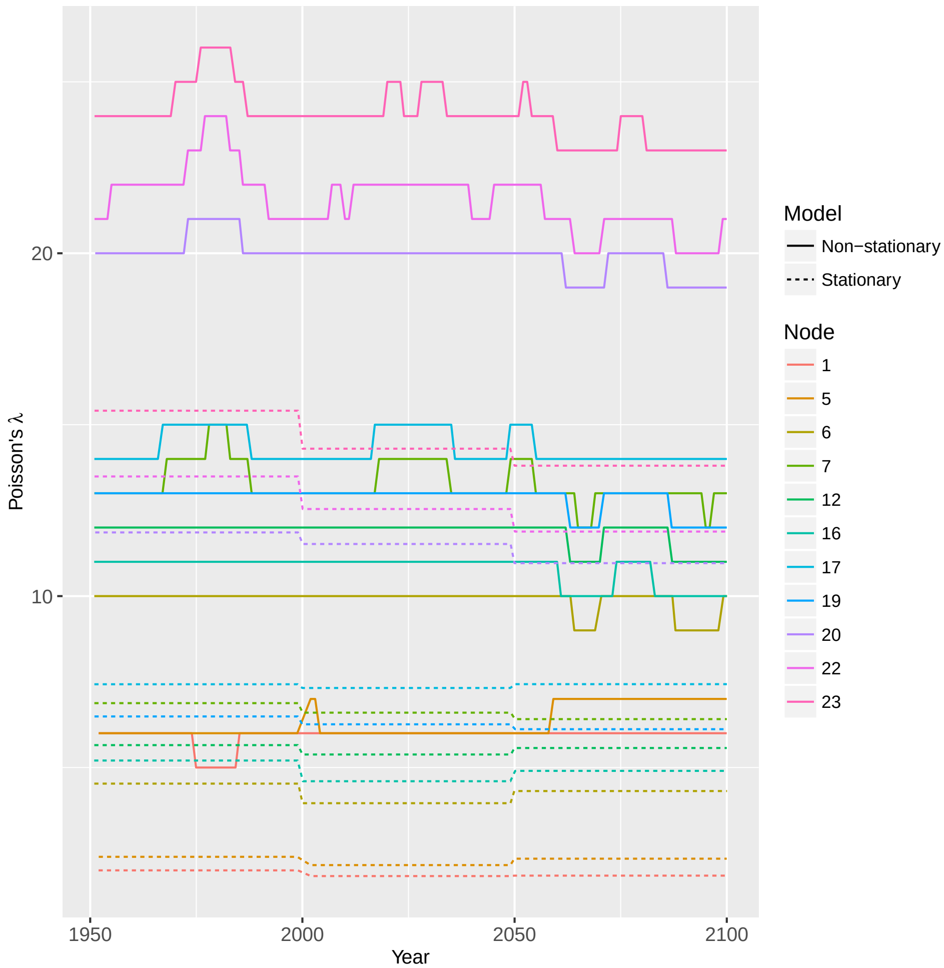


Figure 11: Storminess-index function (λ) for the stationary and non-stationary models, the latter using time as covariate (NS-T).

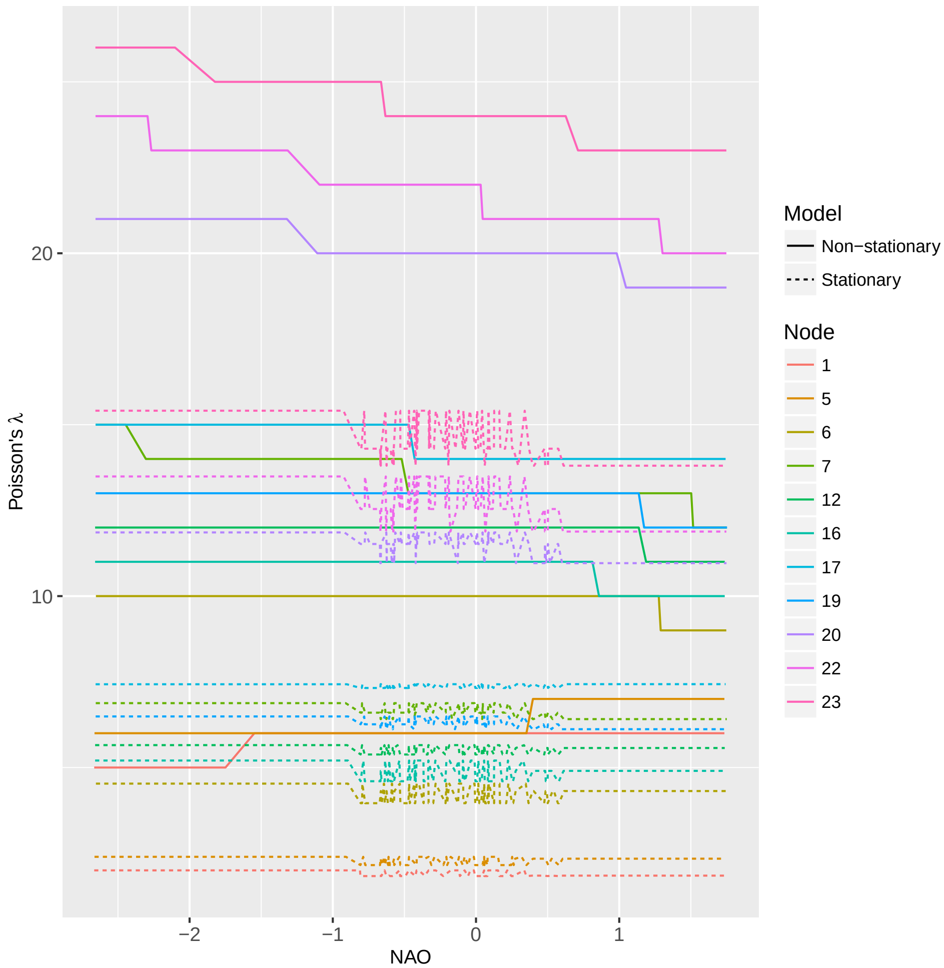
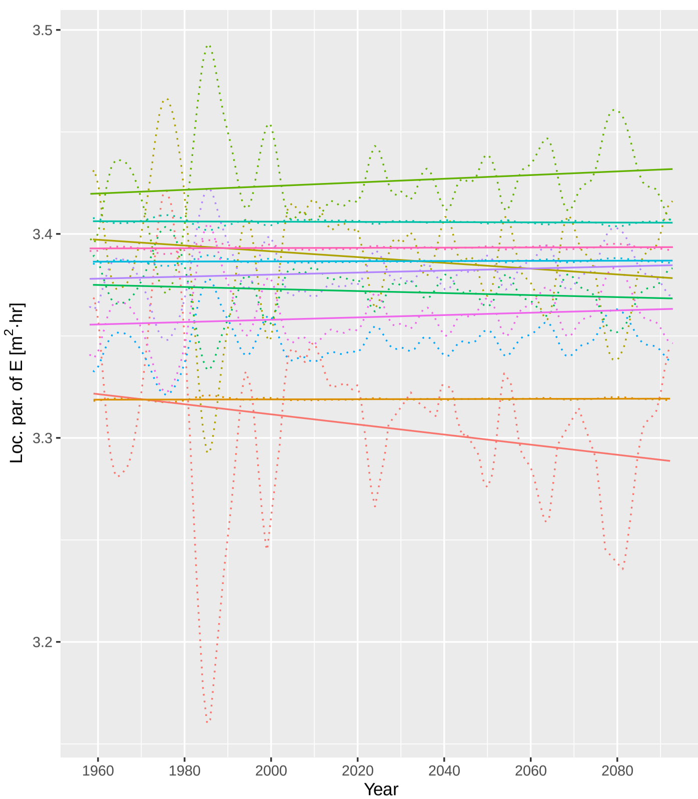
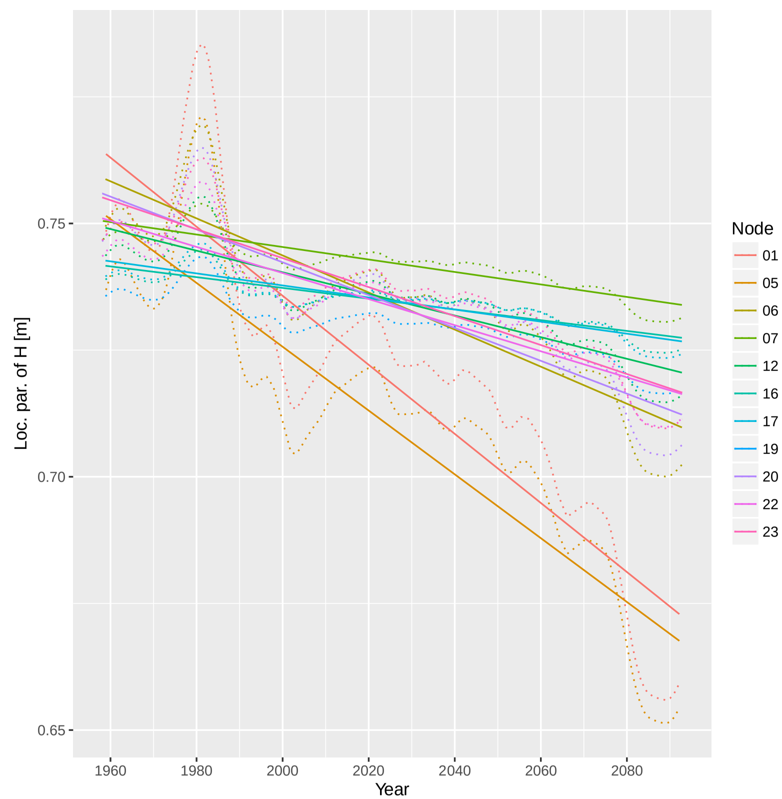


Figure 12: Storminess-index function (λ) for stationary and non-stationary models, the latter using NAO as covariate (from the CMCC-CM, or CMCC-A, model, NS-CI).

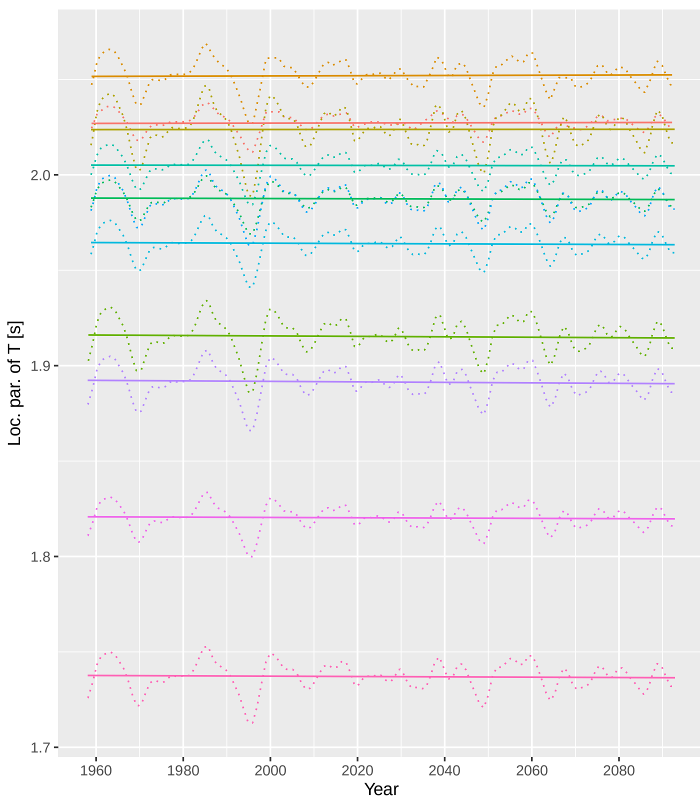


(a)

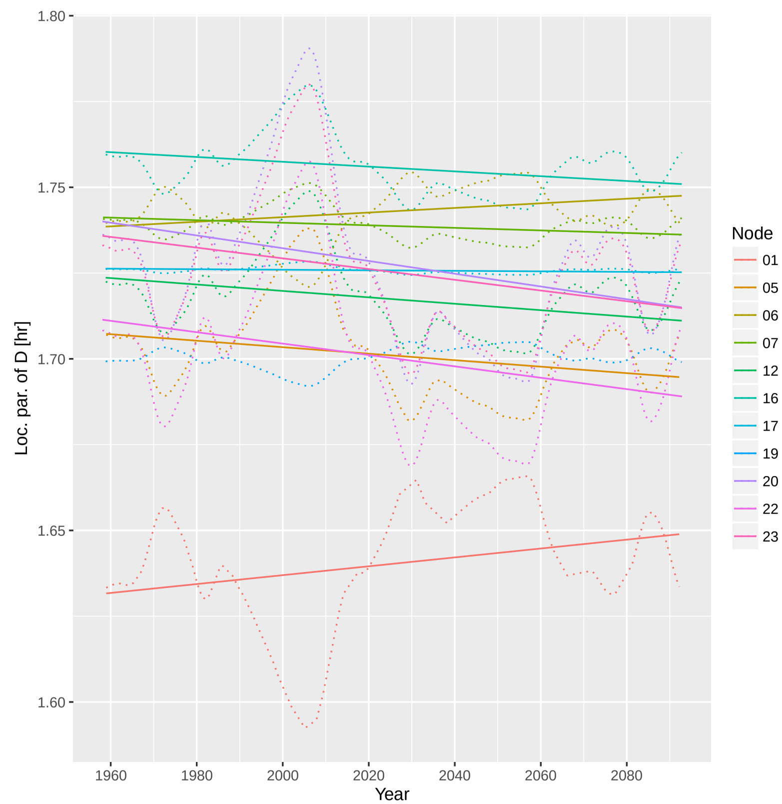


(b)

Figure 13: Non-stationary GPD location-functions (x_0) for a) wave energy (E) and b) significant wave-height at the peak (H_p) using VGAM (GPD distribution) with climate-indices as covariates: $E \sim \left(GPD\left(\mu\left(dSC\right), \sigma\left(d^2EA\right), \xi\right)\right)$ and $H_p \sim \left(GPD\left(\mu\left(SC\right), \sigma\left(d^2EA, d^2SC\right), \xi\right)\right)$. The discontinuous lines show the time variation of the location-parameter and the solid lines represent their linear trend. The colours represent different nodes (see Fig. 1).

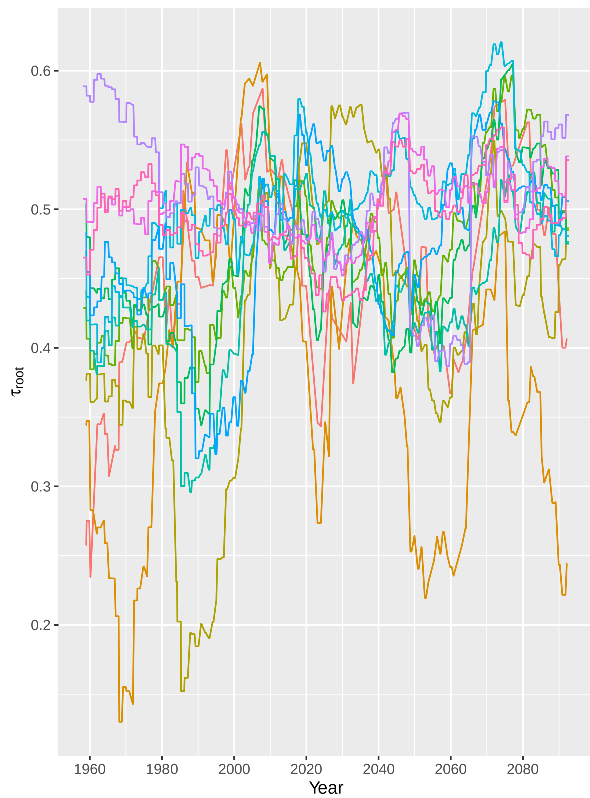


(a)

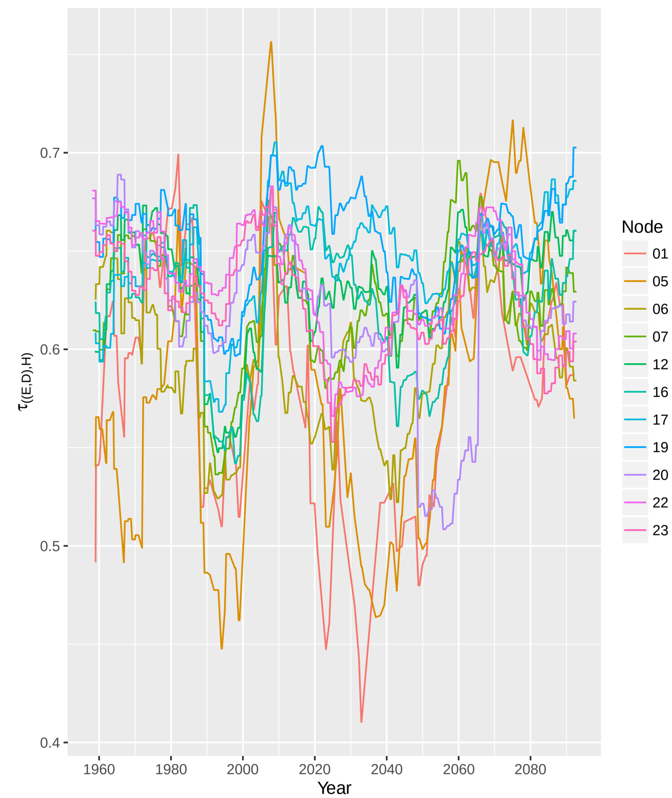


(b)

Figure 14: Non-stationary GPD location-parameters (x_0) for a) peak-period (T_p) and b) storm-duration (D) using VGAM (GPD distribution) with climate-indices as covariates: $T_p \sim (GPD\,(x_0\,(SC)\,,\beta\,(NAO)\,,\xi))$ and $D \sim (GPD\,(x_0\,(EA)\,,\beta\,(dSC)\,,\xi))$. The discontinuous lines show the time variation of the location function and the solid lines represent their linear trend. The colours represent different nodes (see Fig. 1).



(a)



(b)

Figure 15: Non-stationary τ_{root} and $\tau_{((E,D),H)}$ dependence-parameter (Kendall, 1937) at the root and $((E, D), H)$ nesting levels of the HAC structure. The marginal distributions are fitted with the VGAM with climate-indices as covariates (NS-CI). The colours represent different nodes (see Fig. 1).

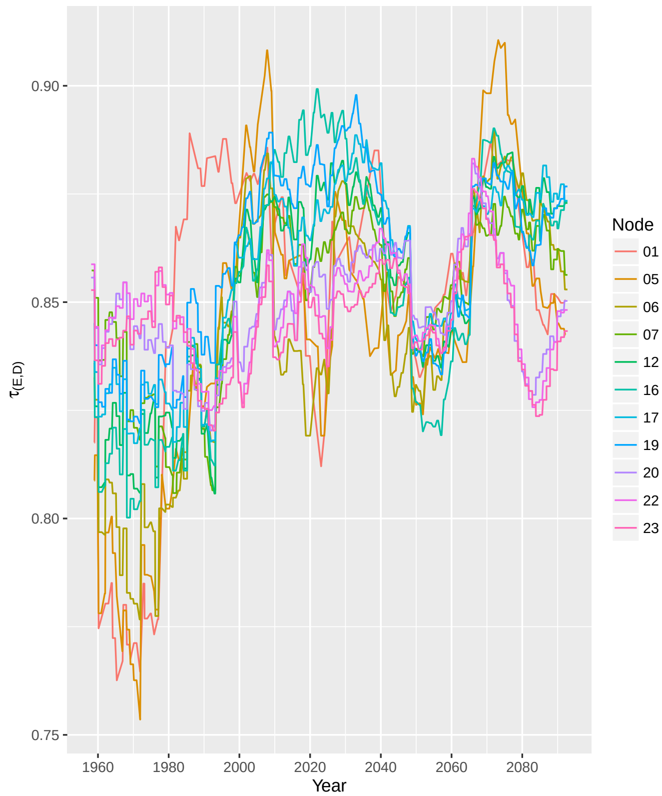


Figure 16: Non-stationary $\tau_{((E,D))}$ dependence parameter at the (E,D) nesting level of the HAC.

Table 2: Validation of the proposed model by computing the Aitchison and the Kullback-Leibler distances between vec_{obs} and vec_{model} (see eqs. 6 and 7).

SIMAR/buoy node	AR5 node	Ait.dist(vec_{obs}, vec_{model}) (Aitchison distance)	KL.dist(vec_{obs}, vec_{model}) (Kulback-Leibler distance)
N1	23	0.52	0.07
N3	22	0.81	0.16
N4	20	0.18	0.01
N7	19	0.45	0.05
N8	17	0.54	0.07
C1	16	0.20	0.01
C3	12	0.26	0.02
C4	07	0.26	0.02
C5	06	0.96	0.24
S4	5	1.31	0.30
S7	1	0.98	0.23
PdE-Begur	20	0.96	0.24
PdE-BCN-I	12	1.31	0.41

Table 1: Global circulation-models from CMIP5 experiment (Taylor et al., 2012) that are considered in this study. The latitude and longitude columns denote the grid size.

Acronym	Global circulation-model	Latitude grid size (°)	Longitude grid size(°)
CMCC_A	CMCC-CM	0.7484	0.75
CMCC_B	CMCC-CMS	3.7111	3.75
CNRM_A	CNRM-CM5	1.4008	1.40625
FGO_A	FGOALS-G2	2.7906	2.8125
GFDL_A	GFDL-CM3	2	2.5
GFDL_B	GFDL-ESM2G	2.0225	2
GFDL_C	GFDL-ESM2M	2.0225	2.5
HAD_A	HadGEM2-AO	1.25	1.875
HAD_B	HadGEM2-CC	1.25	1.875
HAD_C	HadGEM2-ES	1.25	1.875
INM_A	INM-CM4	1.5	2
IPSL_A	IPSL-CM5A-LR	1.8947	3.75
IPSL_B	IPSL-CM5B-LR	1.8947	3.75
IPSL_C	IPSL-CM5A-MR	1.2676	2.5
MIR_A	MIROC-ESM	2.7906	2.8125
MIR_B	MIROC-ESM-CHEM	2.7906	2.8125
MIR_C	MIROC5	1.4008	1.40625
MPI_A	MPI-ESM-LR	1.8653	1.875
MPI_B	MPI-ESM-MR	1.8653	1.875