A multivariate statistical model of extreme events: an application to the Catalan coast

J. Lin-Ye[a,1,*], M. Garcia-Leon[a], V. Gracia[a], A. Sanchez-Arcilla[a]

[a] *Laboratory of Maritime Engineering, Barcelona Tech, D1 Campus Nord, Jordi Girona 1-3, 08034, Barcelona, Spain*

Abstract

Wave extreme events can be understood as the combination of storm-intensity, directionality and intra-time distribution. However, the dependence structure among these factors is still unclear. A methodology has been developed to model wave-storms whose components are linked together. The model is composed by three parts: an intensity module, a wave directionality module, and a intra-time distribution module. In the Storm-intensity sub-model, generalized Pareto distributions and hierarchical Archimedean copulas have been used to characterize the storm energy, unitary energy, peak wave-period and duration. In the Directionality and the Intra-time sub-models, the wave direction (at the peak of the storm) and the storm growth-decay rates are linked to the variables from the intensity model, respectively. The model is applied to the Catalan coast (NW Mediterranean). The outcomes denote spatial patterns that coincide with the state of knowledge. The proposed methodology is able to provide boundary conditions for wave and near-shore studies, saving computational time and establishing the dependence of the proposed variables. Such synthetic storms reproduce the inter-variable co-dependence of the original data.

Keywords: wave storms, Catalan coast, von Mises distribution, multivariate logit function, hierarchical Archimedean copula, generalized Pareto distribution

1. Introduction

Wave storms strongly perturb the state of coastal environments, becoming such changes concomitant with episodic coastal hazards such as coastal flooding and erosion. These extreme phenomena drive complex hydrodynamic processes whose understanding is paramount for proper infrastructure design (Goda, 2010). The conventional approach is usually based on the probabilistic definition of a single parameter, typically the wave height. Other concurrent

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

components as the duration of the storm, the storm total energy and the associated wave period influence the final response of a beach or the damage evolution of a structure (Martin-Soldevilla et al., 2015; Melby and Kobayashi, 2011). These variables are known to be semi-dependent (Salvadori et al., 2007; de Waal and van Gelder, 2005), but the classical methodology either a) assumes one variable to be stochastic and the other ones to be deterministic or, b) assumes all variables to be stochastic but completely independent. In the latter case, the lack of dependence structure hampers finding sets of physically plausible storm components, and requires expert guidance plus local knowledge to discern the suitable combinations.

A common modelling approach is to hindcast high energy events or to synthesize storms to a representative extreme sea-state, which is generally predisposed by the degree of knowledge of the area. For the latter case, dependency structures among the hydrodynamic variables pose a hurdle, as they tend to be unknown. Exploratory methods, such as 2D scatter plots, have been widely used as a rule-of-thumb for the most frequent problem, wave-height vs. wave-period. However, the interpretation of existing co-dependences among several variables is challenging. Recurrently, a wide scatter cloud can mislead about biased co-dependence structures, due to subjective criteria. Storm modelling requires to consider a multivariate analysis of storm parameters (Corbella and Stretch, 2012), as univariate analyses may oversimplify coastal processes, often leading to over or under-estimation of the storm induced damages.

Specialized statistical techniques such as copulas can be used for finding existing relationships among storm variables (Genest and Favre, 2007; Trivedi and Zimmer, 2007) with more objective criteria. Copulas were once described by Sklar (1959), for bivariate models. They were popularized in the 1990s in financial, insurance, econometrical, risk management and actuarial analyses (Cherubini et al., 2004). Applications can also be found in hydrology (De Michele and Salvadori, 2003; Salvadori and De Michele, 2004) and more recently, in coastal engineering (Corbella and Stretch (2012); Wahl et al. (2011); among others).

Corbella and Stretch (2012) employed copula based return-periods to identify the most probable combination of wave-height, wave-period, storm-duration, and water-level for a given probability of exceeding at South Africa. The threshold in the peak-over-threshold method was defined as a critical layer of multiple dimensions that prescribe both a safe and a super-critical combination of storm conditions. In the study, the extreme events were fitted to Generalized extreme value distributions (GEVD). They also noted the importance that their statistical model was constrained, to avoid unrealistic results. Hence, they proposed wave steepness as a restriction that can increase model rigidity and enhance system robustness.

Li et al. (2014) fitted maximum significant wave height, peak-wave-period and storm-duration measured in the Dutch Coast with Generalized Pareto distributions (GPD). They had used the Kolmogorov-Smirnov and the Chi-square tests to evaluate the goodness-of-fit. A similar approach had also been followed by Corbella and Stretch (2013). Salvadori et al. (2014), on the other hand, fitted the significant wave-height and the duration to a Generalized Weibull model (GW) distribution and used Akaike Information Criterion (AIC) to select the suitable copula.

Wahl et al. (2012) applied fully nested Archimedean copulas to consider both storm surge parameters (defined with the highest turning point and the intensity) and the wave height, at the German coast. Nested copulas can characterize multivariate random variables by determining *a priori* nesting architecture that composes simpler copulas structures into larger and more complex ones. Wahl et al. (2012) firstly characterized the highest turning point and intensity; and then incorporated the significant wave height.

The main objective of this paper is to propose a methodology for inferring multivariate wave storm parameters that shares a common structure. To this aim, one of the main points of the paper has been to propose a dependence structure that links the parameters that explain wave storms. The paper is divided into two steps: Model building and Applicability. The proposed wave storm model has been split into three modules: intensity, wave directionality and intra-time storm distributions. This methodology has been tested on the Catalan coast, a fetch limited environment.

The structure of the paper is as follows: Section 2 deals with the methods for building the proposed statistical model. Section 3 presents the study area and, section 4, the database used. Results are summarized in Section 5 and discussed in Section 6. Finally, Section 7 sets out the conclusions.

2. Methods

2.1. Storm definition and variables

The determination of storms has three criteria: 1) intensity definition and associated threshold, 2) minimum time-lapse between storms (D^*_{min}), and 3) minimum duration of the storm (D_{min}). Wave storms are extreme phenomena that can be dealt with the peak-over-threshold description (Embrechts et al., 1997). The threshold separates storm conditions from non-storm conditions. The D^*_{min} helps satisfy independence of the samples. The independence is one part of the "independent and equidistributed" assumption for data in many statistical techniques. D_{min} discards the storms of insufficient duration and which are, therefore, of lesser significance.

Eastoe et al. (2013) associates the threshold with the percentile 90 of the wave height. In our paper, a different approach is proposed. The occurrence in time of extreme events, for any given geographical location, follows a Poisson distribution. Therefore, it can be deduced that the time lapse between storms must be approximately an exponential distribution; if not, these events are not extreme. Appart from this, the threshold should belong to the linear segment of a mean-excess wave-height function (Ortego et al., 2012). At the same time, the events must be statistically significant in number. The wave-height threshold has been varied ranging from 1.5m to 3m, whose minimum doubles the mean wave heights (CIIRC, 2010). The finally selected value of the wave-height threshold is exposed in Section 5 and discussed in Section 6.

Turning to the independence and equal distribution of storm samples, neighbouring storms are clustered if the D^* that separates them is below D^*_{min}, which means that both episodes belong to the same storm event. After clustering, each storm can be considered to be independent from the others. On the other hand, it is assumed that the marine extreme events are generated by a limited subset of synoptic conditions (Lionello, 2012), which is true in Western Europe (Mazas et al., 2014). Therefore, the storms are regarded as identically distributed.

Three candidates for D^*_{min} are proposed: 72hrs, 48hrs, and 12hrs. $D^*_{min} =$ 72hrs is because the two sub-storms in a twin storm tend to be less than 72hrs appart. Approximately $20-30\%$ of the total storm events on the Catalan coast are twin, depending on the location (Wojtanowicz, 2010). The consideration of $D^*_{min} = 48$hrs is conceptually similar to Tolosana-Delgado et al. (2011), whereas $D^*_{min} = 12$hrs is based on direct observations of Catalan sea-storms. A sensitivity test is performed to select the most correct D^*_{min} value. The test consists of representing storms for different values of D^*_{min}. The D^*_{min} selected and the reasons leading to this choice are stated in the Section 5 and discussed in the Sub-section 6.1.

D is the duration of the event between the first and last threshold crossing (Fig. 1a). It is not to be confounded with D^*. The value of D_{min} is given in Section 5.

From each independent storm, the total storm-energy (E), the maximum storm-unitary-energy ($E_{u,p}$), the peak wave period (T_p), the duration D, the direction of the peak-wave (θ^*_p), the growth-rate and the decay-rate are obtained.

The Storm-intensity sub-model includes E, $E_{u,p}$, T_p, and D.

The E is defined as

$$E = \int_{iniT}^{endT} H^2_{m0} dt, \tag{1}$$

where H_{m0} is the spectral significant wave-height, and t is time. In case that the wave-height returns below the threshold, during the event, the duration and the energy of these low intensity periods are included in the sums of D and E.

It has been highlighted in Sánchez-Arcilla et al. (2014) that the capture with numerical models of the peak-wave-height lacks of exactitude, whereas a better skill is found for the existing temporal trend. Therefore, a new definition of the maximum wave-height (H_{max}) is proposed through the definition of $E_{u,p}$:

$$E_{u,p} = \max_i \left(mean \left(E_{u,(i-1)} + E_{u,i} + E_{u,(i+1)} \right) \right), \tag{2}$$

where E_u is the unitary storm-energy at each hour. The square root of $E_{u,p}$ is proposed, here, as an improved definition of H_{max}, and is herein called H^*_{max}.

The H^*_{max} synthesizes the energy shortly before and after the peak. The subset (see Fig. 1b) presents a) point $(t-1)$: growing to reach the peak, b) point (t): Storm peak and c) point $(t+1)$: decreasing or maintaining. The differential energy at $(t+1)$ in decreasing or maintaining the energy is a crucial assumption for point t. The reason is that Mediterranean storms usually present a sharp gradient during wave height growth and a milder one during decay. The variables E and H^*_{max} provide more complete metrics for the storm hazard rather than a representative wave height, as they describe the behaviour of the entire storm, rather than a snapshot.

The T_p relates to the frequency in which the peak of the energy from the directional wave spectrum is located (Holthuijsen, 2007). The T_p of our wave-model is the value of the T_p when E_u takes the $E_{u,p}$ value. The T_p does not vary much during each storm and its standard deviation is generally small. The reason of such reduced variation is a fetch-limited condition of the study area plus the ephemeral intensity of the storms.

The directionality is represented by the Directionality sub-model, and it is parameterized with the wave-direction of the storm-peak (θ^*_p). The value of θ^*_p is assumed to be constant throughout each individual storm-event. Both T_p and θ^*_p are values at the H^*_{max}, as interest is herein put on the behaviour of the most extreme conditions, rather than on the rest of the storms.

Milder slopes during decay have relevant consequences. For example, consider an emerged dune that collapses at the exact moment of the storm peak or maximum wave height. The after-effect (flooding/erosion) would not be the same if the energy started to decrease at the same rate as the storm growth. A sharp growth leads to collapse, defence impairment and the decay phase can lead to the real «infrastructure damage» (Gràcia et al., 2013). A parameter that considers that effect is sought in this study, whilst maintaining as much information of the peak as possible.

The storm wave evolution over threshold is modelled with either the irregular-trapezoidal or the triangular shapes (see Fig. 1c). A theoretical basis for the proposal of these two wave-height-evolution models can be found in Martin-Soldevilla et al. (2015), who conducted a shape analysis for one point at the NW Mediteranean Sea. This analysis is herein extended on a regional scale. The residuals associated with triangular and irregular-trapezoidal candidate wave-height-evolution models have been computed. The area below the hindcasted wave-height-evolution function has been compared to the area below each one of the candidate wave-height-evolution models. The area below the wave-height-evolution model is computed with the area within each figure plus the area below the threshold; the maximum wave-height considered in such calculation is H^*_{max}.

After adopting a shape, the D provides two indicators: a) the percentage of time from the beginning of the storm to the first H^*_{max} (growth-rate), and b) the percentage of time from the last H^*_{max} to the end of the storm (decay-rate). These are the ratios growth-time$/D$ and decay-time$/D$, respectively, that define the storm-shape. The growth and the decay-rates are characterized by the Intra-time-distribution sub-model.

The Storm-intensity sub-model might influence the Directionality sub-model and the Intra-time-distribution sub-model. Therefore, the three sub-models are inter-linked.

2.2. Wave-storm model building

Fig. 2 summarizes the main steps followed for the construction of the storm-model. There are three sub-models: intensity (orange boxes), wave directionality (olive green boxes) and intra-time (purple boxes). Rectangle boxes represent the inputs/outputs, whereas the parallelogram boxes represent the actions taken.

The storm components have been previously defined in sub-section 2.1.

The thresholds for the extreme variables are defined by analysing the inter-storm-time-lapse (D^*) and the location of the wave-height-threshold on a mean-excess H_{m0} plot.

In the Storm-intensity sub-model, the univariate probability distributions of E, $E_{u,p}$, T_p, and D are characterized by GPDs, whereas their joint structures, at each geographical node, are described by hierarchical Archimedean copulas. The θ_p^*, at each node (see Fig. 9), are fit to mixtures ($n \geq 2$) of von Mises distributions (Barnerjee et al., 2005; Mardia and Jupp, 2009), abbreviated hereafter as *mixture of vM*, or *movM*. From the movM at one node, the mean of each vM distribution is considered a principal direction (PD_i) of θ_p^*. These PD_i constitute categories for θ_p^*. The PD_i are linked to E, $E_{u,p}$, T_p, and D through a multivariate logistic model, then the Directional sub-model is formed.

From the event-time-description associated to the Storm-intensity sub-model, the storm growth-decay rates are defined, and linked to D, resulting in the storm Intra-time sub-model.

In summary, the Storm-intensity sub-model generates synthetic E, $E_{u,p}$, T_p, and D that, once introduced into the Storm intra-time sub-model and the Directional sub-model, generate the growth-decay rates and the wave directions, respectively. The total set of storm variables define synthetic storms that, once filtered, are ready for applications desired. Both the model and the SIMAR database (see Sec. 3) are validated/compared to the buoy records. Finally, the model-buoy validation and the SIMAR-buoy comparison are contrasted to see what kind of residual is introduced in our final model.

2.3. Storm-intensity sub-model

2.3.1. Univariate marginal distribution: GPDs

The E, $E_{u,p}$, T_p and D are sea dynamic variables that take positive real values; consequently, they can be log-transformed to avoid scale effects. One of the most widely used distributions to characterize wave peaks in a peak-over threshold (POT) approach is the GPD (Coles, 2001). It is assumed that the events are time points which have an associated random magnitude, and they also must be independent and identically distributed (Coles, 2001; Tolosana-Delgado et al., 2010). If X is the magnitude of an event and x_0 is, at the same time, a value of the support of X and a threshold, the excess over the threshold x_0 is $Y = X - x_0$, conditioned to $X > x_0$. Therefore, the support of Y is either $[0\,,\, y_{sup}]$ or a positive real line. The GPD cumulative function is

$$F_Y\left(y|\beta,\xi\right) = 1 - \left(1 + \frac{\xi}{\beta}y\right)^{-\frac{1}{\xi}}\ ,\ 0 \leq y \leq y_{sup}\,, \beta \geq 0\,, \xi \in \mathbb{R}, \tag{3}$$

and the associated probability density function is

$$f_Y\left(y|\beta,\xi\right)=\frac{1}{\beta}\left(1+\frac{\xi}{\beta}y\right)^{-\frac{1}{\xi}-1}\ ,\ 0\leq y<y_{sup}\,,\beta\geq 0\,,\xi\in\mathbb{R},\tag{4}$$

where β is the scale parameter and ξ is the shape parameter. ξ determines the domain of attraction of the distribution. For $\xi<0$, the distribution belongs to the Weibull domain of attraction, and the support of y is limited, being $\left[0,y_{sup}=-\frac{\beta}{\xi}\right)$. For $\xi>0$, the domain of attraction is Fréchet, and the support of y is $[0,+\infty)$. When $\xi=0$, the support is infinite and the distribution belongs to the Gumbel domain of attraction (Coles, 2001; Tolosana-Delgado et al., 2010). The selection of a physically justified threshold for each variable enhances tail convergence.

Thresholds have been defined for the GPD of each variable. D_{min} is 6hrs, then the threshold of D is set as D_{min}, the threshold of E is computed from $H_0^2\cdot D_{min}$, and the threshold of $E_{u,p}$ is computed from H_0^2. The thresholds for E and $E_{u,p}$ are based on their definition. The relationship of H_{m0} to the most widely used significant wave-height (H_s or $H_{1/3}$) is $H_{m0}=H_{1/3}/0.95$, (Holthuijsen, 2007). The relationship of T_p with $H_{1/3}$ can be approximated by a linear expression, defined in CIIRC (2010), so the threshold of T_p can be directly computed from the wave-height threshold.

2.3.2. Dependence structure: the Hierarchical Archimedean Copulas (HAC)

The set of storm components has passed a multivariate independence test based on the empirical copula process (Genest and Remillard, 2004). This test provides insight into inter-dependencies of any subsets of the variables. The resulting graph, the dependogram, displays the subsets on the horizontal axis and the statistic per subset (the departure from independence) on the vertical axis. A statistic (vertical line) below the threshold value (bullets) means a totally independent subset, whereas the length of the vertical line above the bullet represents the degree of co-dependence of the variables in the subset (refer to Fig. 4 for an example).

Once the semi-dependence is demonstrated, several methods are available to model multivariate distributions. Hierarchical Archimedean copulas is one of them. The copula simplifies the modelling as it estimates a multivariate distribution once the marginal distributions of each individual random variables are determined (Sklar, 1959). Pre-selected distributions separate the marginals from the dependence structure between the random variables. Consequently, the dependence modelling through copulas may be a suitable alternative for building multivariate distributions when the marginals are known and heavy tailed (de Waal and van Gelder, 2005). Heavy tails are present when extremes are much more divergent from the mean than it would be expected.

The bivariate distribution described by Sklar can be generalized into a multivariate one. For any multivariate distribution function H with margins F_j, $j\in\{1,...,d\}$, a copula C can be defined such that

$$\mathrm{H}\left(\mathbf{x}_1,...,\mathbf{x}_d\right)=\mathrm{C}\left(\mathrm{F}_1\left(\mathbf{x}_1\right),...,\mathrm{F}_d\left(\mathbf{x}_d\right)\right)\qquad,\mathbf{x}\in\mathbb{R}.\tag{5}$$

Inversely, given a copula C and univariate distribution functions F_j, $j \in \{1, ..., d\}$, an H defined by eq. 5 is a distribution function with marginals F_j, $j \in \{1, ..., d\}$. Being $u_j = F_j$, a d-dimensional copula is Archimedean if it admits the representation

$$\mathrm{C}(\mathbf{u}; \phi) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d)), \quad \mathbf{u} \in [0,1]^d, \tag{6}$$

where the generator function ϕ is continuous decreasing and convex, with $\phi(1) = 0$. An example of a generator function is the Gumbel generator function

$$\phi(u) = (-\log(u))^{\theta}, \quad \theta \in [1, \infty), \tag{7}$$

u is the storm component, and θ is the dependence parameter which indicates independence when $\theta = 1$ and total dependence when $\theta \to \infty$. The dependence parameter θ is distinguished from the peak-wave-direction θ_p^*, in this text, by adding an asterisk to the latter parameter. Other types of Archimedean copula generator functions, such as Clayton and Frank, can be referred to in Wahl et al. (2011).

Most common Archimedean copulas have constrained multivariate dependence structures, as they usually depend on a single parameter of the generator function. Moreover, they are insensitive to variable permutation, which implies that all margins of the same dimension are equal, deeming them unable to model asymmetries in the variable co-dependences(Hofert and Machler, 2011). Hierarchical Archimedean copulas (HAC, see Fig. 3 for an example) can be a useful tool to overcome these drawbacks, by nesting simple 2D-Archimedean copulas into multilayer tree structures that are fitted in a recursive way (Okhrin et al., 2013).

The hyerarchical structure of the HAC provides a series of advantages: a) it is more flexible and intuitive than the simple Archimedean copulas, b) it can model asymmetries in the variable co-dependences, unlike simple Archimedean copulas, c) there is a marginal cumulative distribution function at each node of the tree, d) it require less parameters than other kinds of copulas (e.g. elliptical copula), and e) when basing each copula on a single generator function, the copula parameters rise as the level increases, enabling simpler dependence analyses.

Different generator functions can be used to obtain the θ at each nesting level of a HAC. Extreme storms present a typical pattern of producing extreme values for most storm components, such as E, $E_{u,p}$, T_p and D above a certain threshold. Then, the most suitable HAC type is Gumbel (when a generator function is used at all the levels of nesting of a HAC, this generator function gives its name to this HAC). The Gumbel HAC includes such upper extreme dependence (Salvadori et al., 2007). Other HACs, such as the Clayton and the Frank HACs, may also be employed, as discussed in Wahl et al. (2012). Hence, although the Gumbel type is selected *a priori* for this study, goodness-of-fit-tests are also applied to Clayton and Frank HAC types, with the aim of verifying the suitability of Gumbel.

The aggregation at each nesting level depends on a parameter ε. If the absolute difference of the dependence parameters of two subsequent nodes is smaller than ε (see eq. 8),

$$|\theta_1 - \theta_2| < \varepsilon, \tag{8}$$

the aggregation method «mean», the one used here, equates the ε to the average value between the θs.

An example of a four-dimensional HAC can be

$$\mathrm{C}\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\right) = \mathrm{C}_3\left\{\mathrm{C}_2\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\right), \mathbf{u}_4\right\} = \phi_3^{-1}\left\{\phi_3 \circ \mathrm{C}_2\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\right) + \phi_3\left(\mathbf{u}_4\right)\right\}. \tag{9}$$

If the copula tree (see Fig. 3) spreads its "branches" upside down, the lowest hierarchical level would be the tip of the branches. At such lowest hierarchical level, the parameter of any pair of the given variables is estimated. The couple with the strongest dependence is aggregated and substituted by a joint pseudo-variable (Okhrin et al., 2013). For example, let E and D share a common dependence parameter $\theta_{(E,D)} = 4.44$. Let it be the highest valued dependence parameter among all the pairs of variables. The pair of variables (E, D) can be substituted by the pseudo-variable

$$\mathbf{Z}_{(E,D)} \stackrel{def}{=} \phi^{-1}_{\hat{\theta}_{(E,D)}}\left[\phi_{\hat{\theta}_{(E,D)}}\left\{\hat{\mathrm{F}}_D\left(D\right)\right\} + \phi_{\left(\hat{\theta}_{E,D}\right)}\left\{\hat{\mathrm{F}}_E\left(E\right)\right\}\right]. \tag{10}$$

At the next level, the parameter of all the pairs of variables and pseudo-variables are again evaluated. This procedure is continued until the highest hierarchical level (i.e. the root) is reached (see Fig. 3).

Several approaches can be found in the literature to determine the HAC agreement with data. Chen et al. (2004) proposed a dimension-free goodness-of-fit test which has been adpted to construct the HACs. The graphical test detailed in Okhrin and Ristig (2012) has been applied to check the goodness-of-fit at each nesting-level. It is complemented with quantitative values from a parameter k^2 (Gan et al., 1991).

Okhrin and Ristig (2012) compares the model probability-distribution with the empirical probability-distribution. The expression of an empirical copula is

$$\hat{\mathrm{C}}\left(\mathbf{u}_1, ..., \mathbf{u}_d\right) = n^{-1}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbf{I}\left\{\hat{\mathrm{F}}_j\left(\mathbf{X}_{ij}\right) \leq \mathbf{u}_j\right\}, \tag{11}$$

where n is the sample size, d is the number of variables, $\hat{F}_j\left(\mathbf{X}_{ij}\right)$ is the empirical marginal distribution function of a variable $\mathbf{X}_{ij}$, and $\mathbf{u}_j$ is a vector belonging to the interval $[0, 1]$. $\mathbf{I}$ is a unit function (it is 1 when the argumet is true, and 0, when the argument is false), so that the product represents the unit function of the AND combination of all the j conditions

$$\hat{F}_j\left(\mathbf{X}_{ij}\right) \leq \mathbf{u}_j.$$

.

Gan et al. (1991)'s k^2 quantifies the agreement of the analysis at each nesting level. Each one of these levels only has two variables, then the criterion is herein restricted to 1D dimension comparisons. k^2 takes values in $[0,1]$, the larger the number, the highest the similarity of the vectors involved.

Here, θ of different Gumbel copulas are not easily comparable, as the support of θ is semi-infinite. Thus, θ are transformed into Kendall's τ, or Kendall's rank correlation coefficient (Kendall, 1937), the support of which is $[0,1)$. The value 1 is excluded for corresponding to the infinity value in θ.

Once HAC structures are obtained for each node, $\tau_{(E,D)}$ values are obtained through ordinary kriging (OK) (Wackernagel, 2003), along the Catalan coast, in order to visually identify the spatial distribution of the co-dependences of E and D. This approximation remains valid for zones where the observed hydrodynamic patterns do not differ excessively, and offers estimations at unsampled areas.

2.4. Linking wave-direction to Storm-intensity: The wave directional sub-model

It is not possible to include the θ_p^* and the growth-decay rates into the HAC in the Storm-intensity sub-model, since these storm-components do not have a support in the space of the real numbers. However, according to results from dependograms, directionality and growth-decay rates are not entirely independent from the Storm-intensity model. Therefore, the directionality and the growth-decay rates are compelled to relate to the Storm-Intensity sub-model via a regression model, although not through a HAC structure.

The standard approach transforms a continuous variable into a predefined set of categories. Usually, the reference coordinate system (i.e. North) and some predefined bins divide the wave-rose into 16 sectors. This poses a problem when the wave-directions are near the boundaries between two sectors, and can mislead regarding contingency. It is, then, crucial to select a set of categories based on the data itself. Both reference and bin size can be established with movM distributions. This type of distributions allow a more flexible definition of the wave-direction contingency, as elementary distributions are not assumed constant over preassigned subintervals. What is more, it can be transformed into categories of principal wave-directions (PD), simplifying the prediction of wave-directions.

In this methodology, wave-directions are first characterized with movM distributions (Barnerjee et al., 2005; Mardia and Jupp, 2009), whose probability distribution function of a mixture of k elements is

$$f\left(x|\hat{\Theta}\right) = \sum_{h=1}^{k} \alpha_h f_h\left(x|\hat{\theta}_h\right), \qquad k \in \mathbb{N}, \tag{12}$$

being x a circular variable, with μ_h as the hth mean, and κ_h as the hth "standard deviation". The α_h are the mixture probabilities, they are non-negative and sum to one; by definition, the mode with the largest α_h is the principal direction. $\hat{\theta}_h = (\mu_h, \kappa_h)$ for $1 \leq h \leq k$, and $\hat{\Theta} = \left\{\alpha_1, \ldots, \alpha_k, \hat{\theta}_1, \ldots, \hat{\theta}_k\right\}$. $\hat{\Theta}$ represents the mixture probabilities, as well as the means and standard deviations of the vM distributions in the mixture. Both $\hat{\theta}$ and $\hat{\Theta}$ have hats, in order to distinguish them from the peak-wave-directions (θ_p^*) and HAC parameters (θ).

An Expectation maximization (EM) approach is used for maximizing the expectation of eq. 12. With the constraints on the vMF mean and deviance, $\mu_h^T \mu_h = 1$ and $\kappa_h \geq 0$, the expression of the mixture probabilities α_h is:

$$\alpha_h = \frac{1}{n} \sum_{i=1}^{n} p\left(h | x_i, \hat{\Theta}\right), \qquad n \in \mathbb{N}, \tag{13}$$

where n is the total number of elements in the sample, x is the angle, and $\hat{\Theta}$ is the parameter appearing in the eq. 12, and described above. $p\left(h | x_i, \hat{\Theta}\right)$ is the probability of appearance of the h vM distribution, given the angle x_i and the parameter $\hat{\Theta}$.

From the soft EM framework used here, the distribution $p\left(h | x_i, \hat{\Theta}\right)$ is given by

$$p\left(h | x_i, \hat{\Theta}\right) = \frac{\alpha_h f_h\left(x_i | \hat{\Theta}\right)}{\sum_{l=1}^{k} \alpha_l f_l\left(x_i | \hat{\Theta}\right)}, \tag{14}$$

where α_h, x_i, k, and $\hat{\Theta}$ are the same variable as in eqs. 12 and 13, and $f(x_i | \hat{\Theta})$ is the probability distribution function of x_i, given $\hat{\Theta}$. The soft EM framework, assigns soft (or probabilistic) labels to each point given by eq. 14. Other candidates can be the hard, or "winner takes all", EM, but the soft EM is selected for its flexibility, in comparison with the hard EM.

The wave-direction is decomposed into the sine and cosine of the angle, and these two elements are then fit by movM. The corresponding movM parameters can be used to generate synthetic pairs of sine-cosine that can be combined to estimate the synthetic wave-direction. The Watson's two-sample uniformity test then helps identifying the strictly necessary number of modes in the movM distribution (Pewsey et al., 2013). By doing so, it improves goodness-of-fit, whereas avoiding over-fitting. This test checks whether two groups are extracted from a common distribution. The criterion for the goodness of fit is set as the statistic U^2 to be smaller than 0.152, which corresponds to p-value$=$ 0.1. When this criterion is met, it means absence of significant difference between the empirical distribution and the model distribution.

The means μ_k of each movM are considered as principal directions (PD_k). These PD_k delimit a set of categories. Hence, the continuous wave-direction in each storm is labelled by a category that bonds the "influence area" of one of the k vM distributions in the mixture. The main advantage of this approach is that the categorization of this variable is data-dependent, so the ranks can be related to the Storm-intensity sub-model.

The relationship between the predicted PD_k categories and the variables from the Storm-intensity sub-model ($\log E$, $\log E_{u,p}$, $\log T$, $\log D$) is built with a multinomial logistic model (Hosmer et al., 2013). A multinomial logistic model consists of a regression model where the dependent variables (i.e. PD_k) are categories and the explanatory variables can be continuous. Particularly, the predictors used in the multinomial logistic model are E, T_p and D. $E_{u,p}$ is not non-significant as a predictor. Therefore, the multinomial model predicts the probabilities that a particular PD_k can happen under certain intensity quantities, then joining directional patterns with its associated E, T_p and D.

2.5. Intra-time distribution sub-model

This sub-model is linked with the Storm-intensity sub-model via the D. A polynomial function is adopted; it predicts the growth-decay rates from a given D. Other variables from the Storm-intensity sub-model do not show clear relationship to the growth-decay rates.

A polynomial function is sufficiently flexible capturing the inner structure within D intervals vs. the growth-decay rates. What is more, a suitable relationship is a third degree polynomial function, where the independent variable is D: $f(D) = a_0 + a_1 D + a_2 D^2 + a_3 D^3$.

2.6. Wave storm generator

Once our model is built, the applicability consists of generating synthetic storms, whose parameters are related. These storms has been produced by recursive simulations that consider the nested structure of the HAC model, as well as the links between our three sub-models. The storms are generated for a given design return periods (T_r) until there is approximately a sample with more than 1000 storms, at each node. The selected tolerance for the error in joint and marginal T_r, in the storm generation, is 20%. This degree of tolerance is suggested by an estimate of observational residuals in the Catalan Sea (Sánchez-Arcilla et al., 2008a, 2014).

There is not a unique correct design T_r, since in a multidimensional space there is no single total order. There is a variety of failure modes and diverse probabilities of failure that combine the existing parameters. Several criteria exist to define a multivariate (n-variate) T_r (Salvadori and De Michele, 2010), and four representative expressions are listed below. These T_r take into consideration the various storm descriptors in the Storm-intensity sub-model.

The Kendall T_r (Salvadori et al., 2007) is:

$$Tr_k = \frac{1}{\lambda \cdot (1 - F(\mathbf{x}))}, \quad \lambda \in \mathbb{R}, \mathbf{x} = (x_1, \ldots, x_i, \ldots, x_n) \in \mathbb{R}^n, \tag{15}$$

where λ is the annual occurrence of storms, $\mathbf{x}$ is the storm components characterized by HACs, and $F(\mathbf{x})$ is

$$F(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} F(X_i < x_i), \tag{16}$$

where λ is the same concept as in the Kendall's T_r, u_i is the cumulative probability of a $1D-$variable, $\mathbf{I}$ is the unit interval $[0,1]$, the critical threshold $t \in \mathbf{I}$ is given by $t = \inf\{s \in \mathbf{I} : K_C(s) = p\} = K_C^{[-1]}(p)$, where K_C is the Kendall coefficient.

Two other possible ways to compute the joint T_r are via the mean value of the marginal T_r (eq. 17) or the geometric mean value of the marginal T_r (eq. 18):

$$Tr = \frac{1}{n}\sum_{i=1}^{n} Tr_i\left(x\right)\ ,x \in \mathbb{R}, \tag{17}$$

$$Tr = \sqrt[n]{\prod_{i=1}^{n} Tr_i}\ ,x \in \mathbb{R}, \tag{18}$$

where Tr_i is the T_r of x. x is a storm component and Tr_i is calculated by means of eq. 15.

All these different definition of T_r bring forth the need for further research into multivariate T_r, as the currently available tools are mostly statistical theoretical artefacts based on the not always true assumption that high values of variables are dangerous. All four definitions of T_r have been tested on, and, finally, the eq. 17 is selected for presenting a better approach to physical measurements. See Section 5 for results, and Section 6 for the discussion.

For a contingency study, the storm components are considered truncated. So pie-charts can be applied to represent which intervals are more frequent than others. A pie-chart leads to visually assess the different categories and the relative weights over a total simulated number of storms. For the case of of wave-height, the H_s are within $3 - 3.5$ m, and these values constitute the principal category. This visualization of the frequencies leads to a simple interpretation of the storm component interactions among themselves, thus aiding to find representative scenarios given a T_r. The 1, 2, 5, 10, 25 year return periods have been selected for synthetic data clustering, as they are routine in infrastructure design. The life-time of a hard coastal protection infrastructure (e.g. revetment, groyne, etc.) may be established as 25 years (DGP, 2001), whereas soft coastal protection (e.g. nourishment, dune building, etc.) are associated with lower T_r (5 or 10yr) (García-León et al., 2015; Sánchez-Arcilla et al., ress). Direct applications of this methodology can provide hydrodynamic loads for infrastructure design and diagnosis.

3. Study area

The Catalan coast is part of the north-western Mediterranean Sea (see Fig. 5). This water body is characterized by its semi-enclosed nature, the orographic patterns, air-sea temperature differences and the passage of low pressure centres from the Atlantic (Lionello, 2012). The main morphological features are the existence of mountain chains parallel and close to the coast, the Pyrenees Mountains to the north, and the Ebre river valley to the south. These orographic discontinuities, together with the major river valleys, allow for strong winds to be channelled down to the coast (Grifoll et al., 2015).

The Catalan coastal winds are typically low to medium, on average, ranging up to 11.05m/s (Sánchez-Arcilla et al., 2008b). The most frequent and intense wind is the Tramuntana (N), appearing from November to March. It has been observed that it is the major forcing for the northern and central Catalan coastal waves. From latitude $41°N$ southward, the principal wind direction is the Mistral (NW). It is channelled by the western Pyrenees and the Ebre valley. The NW winds are formed by the superposition of gap and downhill flows from the Pyrenees. A secondary wind mass, the Ponent, comes from the depressions in northern Europe and sweeps the entire Iberian Peninsula from west to east.

Eastern winds are frequent during the summer. They are commonly triggered by an intense high-pressure area on the British Islands. Another origin is a high level of cold air pool deepening over the Mediterranean sea, which lead to cyclo-genesis, resulting in the passage of a low off the Catalan coast (Bolaños et al., 2009; Lionello, 2012). Winds are more variable for higher intensities. Thus, some relatively large wind modulus-variability is generated during storms (Bolaños, 2004). Wave-directions are directly correlated with wind-direction, except the angle 50° of waves, which can be generated by winds in the sector NNW-ENE, approximately. This might be explained by the orientation of the coast-line, all winds, at some point, seems to create an alongshore wave-train.

The Catalan coast has a micro-tidal environment (Lionello, 2012). The slope of the bathymetry is relatively steep in the north, while it becomes milder to the south. This has a direct impact on how waves behave when reaching the coast, as the bathymetry has an effect on the type of the impacting wave, and the beach slope determines the vulnerability to flooding. Waves on the Catalan Sea also have a critical effect on sediment-transport, as the short wave-lengths do not allow the beach sediment to restore itself during summer-time.

For fetch limited environments, direct correlation has been observed between wind and wave-directions, this suggest that the local wind is the main forcing for waves at the Catalan Sea, rather than distant winds, so we stress on the difference between local (which generate wind-waves) and distant winds (which generate swell-waves). This reinforces the idea that storm-waves at the Catalan coast are driven by mesoscale processes that span the entire fetch, whereas the swell contributions can be considered as secondary.

According to Bolaños et al. (2009), who used XIOM buoy data, the largest waves come from the east, caused by the joint action of the most significant fetches and winds. In further analysis with dependograms, it can be specified that such directionality is most evident for $\overline{T_p}$, at almost the entire Catalan coast. The directionality of H_{m0} is limited to nodes N4, N5, C2, C4, S1 and S4.

The mean significant wave-height ($\overline{H_s}$) is 0.72m from Barcelona city northward (the quantile 75 of H_s is $q_{H_s,75}$ = 0.89m, $H_{s,max}$ = 5.85m), and 0.78m southward ($q_{H_s,75}$ = 0.98m, $H_{s,max}$ = 5.48m). The extreme values are approximately seven times the average values. In fact, the standard deviation is relatively high, being 30% of the mean. What can be expected is that a structure can be severely challenged by storms of higher T_r. Northern storms might be slightly more hazardous, as it is observed here that $H_{s,max}$ are 0.37m higher at northern sites than southern ones.

The mean peak-wave-period ($\overline{T_p}$) is 5.85s on the northern Catalan coast ($q_{T_p,75}$ = 6.73s, $T_{p,max}$ = 15.87s) and 5.62s on the southern Catalan coast

($q_{T_p,75} = 6.65$s, $T_{p,max} = 14.1$s) (CIIRC, 2010). In this case, standard deviation is double the mean value. However, the quantile 75, the maximum and the mean are of a similar order of magnitude. The T_p, including the mean and the maximum values, is geographically homogeneous.

The NW waves are the highest in Tortosa cape, while the eastern and southern waves are steepest in Llobregat delta (Bolaños et al., 2009). There is also a weakly linear relationship between the mean wave-period (T_z) and the H_s, that is, for each increase in 2s of T_z, H_s increases by 1m.

The study area is divided into hydro-dynamically homogeneous sectors of similar lengths (see Fig. 5). The northern sector (N-) spans the area from the border with France ($42.44°N$, $3.18°E$) to the Mataro Port ($41.53°N$, $2.44°E$), the central sector (C-) extends from the Mataro Port to the Segur de Calafell port ($41.19°N$, $1.61°E$), and the southern sector (S-) ranges from the Segur de Calafell port to the border with the Autonomous Community of Valencia ($40.53°N$, $0.52°E$). The sector boundaries are political frontiers and locations of change in beach orientation. Each sector features a mean shoreline orientation that determines "a posteriori" whether a simulated synthetic storm (see Section 2) will reach the coastline.

4. Data source, and explanatory analysis of the storms

The training set that the proposed statistical model uses comes from the SIMAR dataset (Gomez and Carretero, 2005). The data consist of wave-hindcast-simulations by WAM (WAMDI Group et al., 1988) and WAVEWATCH3 (Tolman, 2009), fed with HIRLAM wind fields (Unden et al., 2002). SIMAR provides consistent, gap-less and spatially dense time series. A series of nodes are selected to representatively cover each one of the above mentioned sectors. This results in 6-8 nodes being assigned to each sector. N1 is near Creus Cape and S7 is well below Ebre Delta (see Fig. 5). SIMAR nodes are located at -50m depth, which are intermediate waters, in this area.

The hindcast ranges from the 14^{th} January 1996 to the 25^{th} February 2013. Data in some nodes extend to the 22^{nd} January 2014. SIMAR provides a variety of wave-spectra-parameters, such as H_{m0} and T_p, among other information, including incoming wave direction and moment in time. The time resolution before June 2000 is of 3hrs and changes to 1hr thereafter. Spline-interpolation has been applied to discretize all time-series with the same temporal resolution.

Storms are obtained from the SIMAR dataset with the methodology described in Section 2. Explanatory analysis shows that the quantiles 50 of E, H^*_{max}, T_p and D are spatially uniform, whereas their quantile 85 present more geographical heterogeneity: higher values in the north, lower values in the south and in the Roses Bay (see Fig. 5); specifically, the E, D, and H_{m0} decrease approximately 25% southward and in the Roses Bay, while the T_p increases 10% in the same direction. The Northern part of the Catalan coast (above $41.2°N$) has higher waves in its strongest storms, reaching values above 4m. Storms in these locations also have a longer D, surpassing 50hrs. The T_p, on the other hand, are larger from $41.8°N$ southward. Note that the quantiles under 50, the quantiles

15 of E, H^*_{max}, T_p and D, for instance, are also spatially homogeneous, but they are ignored, as they are influenced greatly by the selected GPD thresholds.

Both PdE («Puertos del Estado» or State harbours) and XIOM buoy records (see Fig. 5 and table 1) are used for model validation. The selected buoys are located at similar positions to the SIMAR nodes. XIOM buoys provide H_{m0}, mean wave period (T_m), and date. For the sake of comparison with SIMAR dataset, the relation $T_m/T_p = 0.8$ (Goda, 2010) is considered.

5. Results

Figs. 6a through 6h, and Figs. 7a and 7b show a threshold iteration test on the nearest PdE and SIMAR nodes to the Barcelona City. This location is chosen for being the geographical centroid of the Catalan coast. The storm-threshold is named h_0. Following to the criteria mentioned in Section 2, the selected value for h_0 is 2.2m. On the other hand, the most adequate D^*_{min} is 12hr.

The numbers of storms, at each node, are listed on Table 2. The northern zone is the stormiest whereas lower number of storms were found at the south, coinciding with the state of the art (Sánchez-Arcilla et al., 2008b).

The GPD threshold of D is considered to be $D_{min} = 6$hrs, the threshold of E is $H_0^2 \cdot D_{min} = 29.4m^2\,$hr, and the threshold of $E_{u,p}$ is $H_0^2 = 4.84$m^2. The threshold of T_p corresponding to $H_{1/3} = 0.95 \cdot H_{m0}$ is 8.17s (CIIRC, 2010). E, $E_{u,p}$, T_p and D are well fit by GPD, with the selected thresholds (see parameters in table 3).

The joint structure of the Storm-intensity sub-model is compared through goodness-of-fit plots for the Gumbel, Clayton and Frank HACs. The three HACs present similar qualitative behaviour and k^2 parameter value. Then, the Gumbel type HAC is selected for being able to include upper extreme dependence. The "mean" aggregation method, in combination with the Gumbel type HAC, is adopted, for providing the best fit.

Two Gumbel HAC tree types (A and B) are observed (see Fig. 3), based on the co-dependence of $E_{u,p}$ to E and D. Type A HAC-trees differ slightly from type B HAC-trees. In type A trees, $E_{u,p}$ has a stronger relationship with E and D. There is no clear spatial pattern in how A and B trees are distributed (see table 4), but there is strong co-dependence between D, E, and $E_{u,p}$; fact that is corroborated by the dependograms (see Fig. 4). The dependence parameter of $\log D$ and $\log E$ ($\theta_{(\log E, \log D)}$), or, in other words, that of D and E ($\theta_{D,E}$), is transformed into a τ value (Kendall, 1937). This τ, which has been called $\tau_{(E,D)}$, is kriged on the -50m bathymetry (see Fig. 8). It is detected that this dependence has a tendency to decrease southward (see Fig. 8).

The contingency of θ^*_p are shown on Fig. 9 and table 5. It is observed that the principal μ is, from N1 to N6, approximately 330°-20° (except at N3). Central nodes (N7 to S2) are heavily influenced by easterly waves, whereas southern nodes (S3 to S7) suffer more heterogeneous influences. The secondary direction at N1 to S6 are eastern waves, whereas it becomes predominantly southern waves from N7 southward. The wave-contingency at N3 is similar to neighbouring nodes, only that the principal and the second directions are at the opposite direction than at node N2, for instance. It is observed that most nodes have bi-modal wave-directions, coinciding with (Alomar, 2012; Bolaños et al., 2009). The coefficients of the multivariate logit function to predict θ_p^*, from $\log E$, $\log E_{u,p}$, $\log T$ and $\log D$, are listed on table 6.

Regarding the residuals associated with the triangular and irregular-trapezoidal candidate wave-height-evolution models, both the overestimation and underestimation residuals are well below $3\text{m} \cdot \text{hr}$ (that is considerably inferior to the area below the H_{m0} time-series' curve) and ranges from a quantile 10 of $20.20\text{m} \cdot \text{hr}$ to a quantile 90 of $157.65\text{m} \cdot \text{hr}$. The trapezoidal model overestimates in 0-1m $\cdot$ hr more than the triangular model, and the triangular underestimates in $0{-}1\text{m} \cdot \text{hr}$ more than the trapezoidal model. Therefore, the trapezoidal model is selected as overestimation has been considered to be less harmful than underestimation, assuming that both residuals are of the same order of magnitude.

The growth-decay rates are assessed with heat-maps, whose "affection areas" are defined with a bandwidth of $radius = 5\text{hrs}$ (see Fig. 10). When several points are inside the "affection area" of one point, the frequency for such pairing is higher and the area becomes "darker". The coefficients of third degree polynomial that relates D and growth are shown on table 7.

Our model has been validated by buoy data (see Figs. 13 and 14). Figs. 13 and 14 are then contrasted with Figs. 11 and 12. The amount of residuals present in our model is comparable to the one present in the SIMAR database. T_p shows a poorer fit (see Figs. 13c and 14c). The same poor fit is present in Figs. 11c and 12c. This behaviour can be explained because the wave-model (WAM and WAVEWATCH) considers *a priori* a parametrized wave-spectra. Such spectra has a predefined shape that does not necessarily represent the real sea state (Pallarés et al., 2014; Alomar et al., 2014). The method of representing the wave-contingency with the principal directions seems to be useful to represent the wave-contingency (see Fig 11g). Regarding the SIMAR model, wave-directions from node N5 seems to differ significantly from the records of the nearest buoy, which suggests sensitivity of the wave-direction registry to the location of the node. The predicted growth and decay suffer rotation from the perfect fit, in the Q-Q plot, that is, central values are better fit than extreme ones (see Figs. 13e, 13f, 14e and 14f). Nonetheless, this better fit of the central values is also present for the node N5 in the SIMAR model (see Figs. 11e and 11f). Ergo, the SIMAR E, $E_{u,p}$, T_p, D and θ_p^* are well validated by the buoy datasets (see Figs. 11 and Fig. 12).

Storms simulated from the statistical model developed herein have been classified according to T_r (eq. 17), and represented in a series of pie-charts along the coast. It can be observed, for example, that E for a T_r of 5years is mainly of the highest values at nodes N1 through N4 (except at Roses Bay, N3), whereas the more southern coastal tracts present less E (see Fig. 15). Similar gradation occurs to D (see Fig. 15d), whereas a milder one occurs to H_{max}^* (see Fig. 15b) and none is observed in T_p (see Fig. 15c). In general, the same spatial gradations are observed at each respective storm component for any one of the

T_r from 1 to 25 years.

6. Discussion

The discussion section will be divided into two parts: the first one will discuss the results from the proposed methodology (Model Building and Validation, in Fig. 2), whereas the second one will focus on the Wave-storm-generator.

6.1. The statistical model

This paper has proposed a statistical model that feeds upon a dataset from a wave-model (Section 3) which reproduces the main processes within the study area (Lionello, 2012). The Mediterranean Sea is characterized by local constraints, such as mountain chains that funnel wind fluxes in a manner that limits the storm-pattern modes (Sánchez-Arcilla et al., 2008b). The Balearic Islands also trigger wave transformation-processes. At the south-most part of the central sector, the beach shoreline orientation induces a sheltering effect from northerly and easterly waves. It can be seen from Fig. 5 that the north-most part of the central sector is not sheltered from wave-storms. Strong forcings from the north and east directions cause the wind to exchange bursts of momentum with waves. The north direction has shorter fetch, while the east direction has different fetches depending on the location of the cyclo-genesis.

In a further consideration, the role of the sea level within a storm, especially when dealing with its consequences, is undeniable. Some authors (Masina et al., 2015) detected a considerable positive correlation between the peak water level (PWL) and the H_s. However, other authors (Mendoza et al., 2011) support the premise that the sea water level is independent from the storm conditions. This paper is based on the definition of storm-waves, therefore, it has focused only on storm-wave components, neglecting the effects of the water level.

The perception threshold in the Catalan Sea is $H_s = 2.0$m ($H_{m0} \approx 2.1$m) (Bolaños and Sánchez-Arcilla, 2006; DGP, 1992) and is introduced as an initial value in the iteration. The goodness-of-fit of observations to exponential models yield residuals to be analyzed. In Figs. 6a and 6e, as the threshold is low, these residuals are large, meaning that the corresponding D^* does not belong to exponential distributions. When the threshold rises, as observed in Figs. 6c, 6d, 6g, and 6h, the residuals are minimized.

Bernardara et al. (2014) discussed that a limitation to this rise in threshold is the statistical significance in number of events over the threshold. It is observed in Figs. 7a and 7b that it is not recommended to go further than $H_{m0} =$ 3m. Model validation has served to refine the value to $H_{m0} = 2.2$m. This result intends to complement Sánchez-Arcilla et al. (2008a), which proposed $H_s = 2.0$m based on mean-excess plots and Kolmogorov-Smirnov goodness-of-fit tests. The threshold $H_{m0} = 2.2$m is adequate because a) the associated D^* is close to be exponentially distributed, b) the threshold falls in the linear part of the mean-excess-graphs, and c) the resulting storms are statistically significant in number. Please note that the fit to the exponential distribution is not perfect, so the excess-over-treshold plot has been crucial in the selection of the storm-threshold.

On the other hand, the sensitivity test on D^* has shown that 12hrs is the most adequate value, since 48 or 72hrs lead to unrealistic storms that differ from field observations. Once storms are defined, it can be perceived that, in general, the northern Catalan coast is stormier than the southern one (see Table 2). N3 behaves differently as it is located inside the Creus Cape (see Fig. 5), which shelters the area from cyclonic activity.

The validation of our model by the buoy records helps identify the sources of residuals in our model. For instance, the lesser similarity of T_p in our model to the buoy recorded T_p partly comes due to the difficulties of modelling this parameter with state-of-the art wave-models (WISE Group, 2007; Pallarés et al., 2014). Another possible explanation is that, for a given H_{m0}, the T_p depends heavily on fetch length and its origin. However, the influence of the T_p is not filtered by the intensity threshold.

Residuals in the growth-decay rates come from two main sources: physical and numerical. A physical source of residuals appears as offshore and onshore winds show distinct growth-decay rates, depending on remarkable differences in fetch extension. These differences can be compensated by uneven wind intensities, but their effect remains in the growth and decay rates.

The numerical residuals in the growth-decay rates come from the third-grade polynomial, used to link growth-decay rates to D, and from the SIMAR dataset. The limitations of SIMAR datasets in representing growth-decay rates might be due to the fact that wave-models usually introduce residuals when reproducing sharp gradients (Cavaleri, 2009; Sánchez-Arcilla et al., 2014). This limitation may be partly alleviated with the novel terms for the wave-action-balance equation (Zieger et al., 2015), that show better agreement with recent measurements. Also, at the study area, storm-wave patterns can be affected by current intensifications originated in the joint action of sustained winds from the NE-SE plus a shelf narrowing effect (Mestres et al., 2016). Thus, coupling the wave-model with a high resolution circulation model may improve the results. The shortcoming of the third-degree polynomial is that it has difficulties reflecting a link of the growth-decay rates for a D below 100hrs, where a dense cloud of values is present (see Fig. 10); further research on the intra-time distribution module is on-going. Apart from these issues, the statistical model reproduces the prominent features at the study area, and the storm components show agreement with the buoy records.

It can be inferred from the HAC results (see table 4) that the strongest dependent variables are $\log D$ and $\log E$. This dependency structure is consistent with physical observations, as the most enduring storms are usually those which have higher hydrodynamic forcings. It can be argued that, as E is integrated over D, that the correlation between them has to be the most prominent. The outcomes also show that, despite some dependence exists between $E_{u,p}$ and the E or the D, the dependence among $E_{u,p}$ and (E, D) is weaker. This behaviour can be explained due to the point-based definition of $E_{u,p}$ that presents more variability than the integrated values of E and D, that features lower variability. It can be observed how $\tau_{(E,D)}$ increases northward (Fig. 8), implying more correlation between durations and northern storm magnitudes. At nodes where type A trees are prevalent, not only E, but also H^*_{max} is co-dependent on D.

Please regard that θ^*_p is the direction of the storm-peak, and therefore represents the storm at its peak, rather than being a mean direction of the event. The East is one of the principal θ^*_p, and the main effective θ^*_p at a great part of the Catalan coast. Waves that blow northward from the Gulf of Lyon tend to veer counter-clockwise and do not impact at the Catalan coast (Bolaños et al., 2009). The coastline orientation (from N6 northward) is the reason, as despite having more recorded storms at the SIMAR points, the effective storms obtained with synthetic simulations were not as significant in number than the other southern points. Due to larger fetch, from N6 northward, northern θ^*_p are dominating. From N7 southward, the southern waves gain importance. The buoy used to validate either SIMAR or our data should be as close as possible to the node in the model to validate, as θ^*_p is considerably sensible to location.

The intercept of the growth-rate is, generally, 0.46, as well as the intercept of the decay-rate (see table 7). Both growth and decay-rates are considerably independent of D for durations under 100hrs. However, for $D > 100$hrs, while the growth-rate become asymptotic to 0.8, the decay-rate becomes asymptotic to 0.2. That is, under this condition of D, more durable storms tend to also present higher growth-rates and lower decay-rates. Such large growth-rate and small decay-rate contradict the common phenomenon. The high T_r events recorded at the Catalan coast (November 2001 , October 2003 and December 2008) are scarce, but reflect this sharp gradient response, veered by the pulsative wind momentum.

The eq. 17 and 18, of T_r, by being arithmetic and geometric averages, respectively, set physical constraints on each marginal variable. This equalizes the marginal T_r of each variable to the total T_r of the storm, as real maritime storms present such equivalence between marginal and total T_r. For example, when the T_r of a storm is 10yrs, the storm should not have a H_{m0} of $T_r = 50$yrs and a D of $T_r = 1$yrs. The T_r from eq. 17, in particular, provides the best constraints to the T_r of each integrating marginal storm component.

E and D can reach significantly large values with increasing T_r at the North (see Figs. 15a and 15d). Eastern storms generated at the Ligurian Sea are the most energetic and lasting storms due to the fetch distance (near 600 km). For $T_r = 5$yrs (see Fig. **??** and the section below), larger D can significantly affect E, as H_{m0} appears to be more spatially uniform along the Catalan coast.

6.2. Application

In order to visualize the potential of the methodology used, an example of the characterization of storms for a $T_r = 5$years is presented. The 5-year T_r has been selected because it is an extreme condition in which a) SIMAR dataset has a representative number of samples and b) the order of magnitude of such category has been analysed in detail for the study area (Mendoza et al., 2011; Sánchez-Arcilla et al., 2008b). As to provide suitable data for elements on the coast, the land originated storms (non-effective storms) are filtered from the set of synthetic storms. Note that, as the principal directions at some nodes might be land-generated, the number of effective storms decrease considerably after the filtering, compared to other nodes.

Our model provides joint combinations of E, $E_{u,p}$, T_p, D, θ_p^* and growth-decay rates. The outcomes of the model can be examined at Fig. 15. The seven predicted variables are summarized in pie-charts, the categories of which describe the differences and principal patterns that appear on a particular node. One of the main findings of this paper is that, rather than a single value that represents a particular category (i.e. a T_r) for a specific wave component, a range of plausible values can be considered, instead. Note, however, that within this plausible range, there may be various intervals of disparate frequency (i.e. particular intervals shown in the pie-charts). The seven variables are linked via statistical models and it appears that a wide range of possibilities satisfy the clustering criteria. A description of the general study area is provided, whereas numeric outcomes are given for an example-node, N5.

The Storm-intensity sub-model provides the first variables of the synthetic storms generated by our model. Fig. 15b shows that the H_{max}^* can range from 2.2m (by definition) to over 8m. The highest waves are located in the northern coast-sector, and decreases southward, just as described in Sec. 5. $\mathrm{mode}\,(\mathrm{H}_{\mathrm{max}}^*)$ at node N5 is $(6, 7.5]\,\mathrm{m}$ ($\mathrm{mode}\left(H_{1/3}\right) = (5.7, 7.1]\,\mathrm{m}$). Fig. 15c shows that T_p is independent from the location along the coast. The $\mathrm{mode}\,(T_p)$ at node N5 is $(11, 12.5]\,\mathrm{s}$. D presents a clear boundary at node C2: southward of node C2, storms generally span 48hrs (2days) of duration (see Fig. 15d). The $\mathrm{mode}\,(D)$ is > 96hr. Fig. 15a shows a geographical distribution that is clearly the result of a combination of the effects of both H_{m0} and D. The $\mathrm{mode}\,(E)$ at node N5 is $> 2000\mathrm{m}^2\,\mathrm{hr}$. The above-mentioned large values for $\mathrm{mode}\,(D)$ and $\mathrm{mode}\,(E)$ are due to the effects of the GPD extreme value functions and the Gumbel HAC, and they surpass physical constraints to such storm components, so the values of 96hrs and $2000m^2\,\mathrm{hr}$ are to be used for mean-D ($\overline{D}$) and mean-E ($\overline{E}$), respectively. These values reinforce the existing idea that storms magnitudes at the northern part of the coast are higher than at the rest of the coast.

The Directionality sub-model specifies that the θ_p^* along the Catalan coast are mainly eastern directions (see Fig. 15e). At node N5, in particular, the principal peak-wave direction is $76.27°$ (see Table 5); this is the PC2 at node N5, but regard that PC1 is not an effective wave-direction.

The Intra-time distribution sub-model reproduce higher growth-rates than decay ones (see Figs. 15f and 15g). The exception is at the Northern nodes, where longer fetches exist and thus, a wider variety of wave ages can be found. The growth-decay rates are geographically uniform, although this is due to an above-mentioned limitation of the SIMAR model and the Intra-time distribution sub-model. The growth-rate to consider at node N5 is $(0.5, 0.6]$, and the decay-rate is $(0.3, 0.4]$.

The results from our model are compared to the conventional engineering approach, where, given a T_r and a location, a H_s is obtained, followed by the T_p. The conventional method presents the following 90% confidence interval, for Tr = 5yrs, $H_s = (4.3, 5)\,m$, and $T_p = (12.4, 12.8)\,\mathrm{s}$ (CIIRC, 2010). The

D is usually considered as 24hrs in the Catalan Sea. The storm wave-height evolution is usually modelled by an isosceles triangle where the height is the maximum H_s. In this case, the conventional $\overline{E}$ is $(491.7, 726)\, m^2\,$h, and other information such as incoming wave-direction can be obtained from contingency tables in the literature.

The storms from our model are consistent with the values provided in Mendoza et al. (2011); Sánchez-Arcilla et al. (2008b). The H_s and the T_p in our model are slightly larger than in the conventional methodology, in this case, without significant physical implications. $\overline{E}$ and $\overline{D}$ from our model, although considerably larger, are possibly more accurate than their classical counterparts, and the same applies to the growth-decay rates. Also, $D = 24$hrs is an average duration, while $D = (114, 168]\,$hrs derives from the SIMAR dataset. θ_p^* is an extra information provided here and which is not so much considered in the conventional approach. Most importantly, the conventional methodology can hardly reflect the probable behaviour of the storm, mainly because it ignores the variable interactions and feedbacks.

7. Conclusions

The statistical wave-storm model proposed is composed by three sub-modules: a) Storm-intensity, b) Wave-directionality and c) Intra-time distribution. In these sub-modules, waves have been defined by a set of storm-components (E, $E_{u,p}$, T_p, D, θ_p^* and growth-decay rates), representing their nature in a more accurate manner. Our model is well validated by buoy records, whereas main sources of residuals are related to growth-decay rates.

Storms have been defined with a threshold of $H_{m0} = 2.2$m, which has been obtained after testing on D^*, plus H_{m0} excess-over-threshold plots.

In the Intensity sub-model, the marginal distributions of each variable are characterized by GPDs, whereas dependences among the variables are represented by HACs. The best fitting HAC type is Gumbel. It is observed that the strongest dependence may be between E and D. Two HAC structures are observed along the Catalan coast: type A and type B, depending on the degree of semi-dependence between $E_{u,p}$ and (E, D). The semi-dependence parameter $\tau_{(E,D)}$ increases northward. Therefore, northern E and D present more correlation.

Wave-directions are described via movM. The movM distribution is selected using a statistic from the Watson test as convergence criteria. The principal peak-wave incoming-direction, θ_p^*, at N1 to N6 are, by decreasing order of importance, North and East; whereas eastern and southern directions are predominant from N7 to S7.

The most appropriate model for wave-height evolution is the irregular-trapezoidal model. On the other hand, the growth-decay rates are related to the rest of the storm components through a polynomial relationship with D. A mean behaviour of D for $D < 100$hrs is reproduced by the model, although for greater D the model tends to predict higher growth rates and lower decay rates.

One feature of our model is its ability to generate synthetic storm conditions and to classify them by T_r; these storms are evaluated in the form of pie-charts. In general, for a T_r of 5yrs, storms at the northern Catalan coast have greater E, D, and H_{m0}; while T_p are similar to central or southern Catalan coasts. Also, the principal θ_p^* is eastern and the growth and decay rates approximate 0.55 and 0.35, respectively.

8. Acknowledgements

This article has been supported by the EU-projects iCOAST (Grant Agreement 661ECHO/SUB/2013/009) and RISES-AM (FP7-ENV-693396), and the Spanish national project PLAN-WAVE (CTM2013-45141-R). As a group, we would like to thank the Secretary of Universities and Research of the department of Economics of the Catalan Generalitat (Ref. 2014SGR1253). The second author acknowledges the PhD scholarship from the Government of Catalonia (DGR FI-AGAUR-14). The support of the Puertos del Estado and XIOM buoy network is duly appreciated. The authors would also like to thank Prof. Dr. Ostap Okhrin and Mr. Alexander Ristig for their invaluable guidance on the use of the HAC R package; to Dr. Maribel Ortego and Prof. Dr. Juan Jose Egozcue, for their comments on the statistical aspects of the work; Dr. Bettina Grün, Prof. Peter Jupp and Prof. Alan Gelfand for their patient guidance on directional statistics. Special thanks are also acknowledged to two anonymous peer-reviewers whose genuinely constructive comments have enriched this paper.

9. References

Alomar, M.: 2012, 'Improving wave forecasting in variable wind conditions. The effect of resolution and growth rate for the Catalan coast'. Ph.D. thesis, Barcelona-Tech.

Alomar, M., A. Sánchez-Arcilla, R. Bolaños, and A. Sairouni: 2014, 'Wave growth and forecasting in variable, semi-enclosed domains'. *Continental Shelf Research* **87**, 28–40.

Barnerjee, A., I. Dhillon, J. Ghosh, and S. Sra: 2005, 'Clustering on the Unit Hypersphere using von Mises-Fisher Distributions'. *Journal of Machine Learning Research* **6**, 1345–1382.

Bernardara, P., F. Mazas, X. Kergadallan, and L. Hamm: 2014, 'A two-step framework for over-threshold modelling of environmental extremes'. *Natural Hazards and Earth System Sciences* **14**, 635–647.

Bolaños, R.: 2004, 'Tormentas de oleaje en el Mediterraneo: Física y predicción'. Ph.D. thesis, BarcelonaTech.

Bolaños, R., G. Jordà, J. Cateura, J. López, J. Puigdefàbregas, J. Gómez, 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.

Bolaños, R. and A. Sánchez-Arcilla: 2006, 'A note on nearshore wave features: implications for wave generation'. *Progress in Oceanography* **70**, 168–180.

Cavaleri, L.: 2009, 'Wave Modeling - Missing the Peaks'. *Journal of Physical Oceanography* **39**(2757-2778).

Chen, X., Y. Fan, and A. Patton: 2004, 'Simple tests for models of dependence between multiple financial time series, with applications to US equity returns and exchange rates'. *London Economics Financial Markets Group Working Paper* –(483).

Cherubini, U., E. Luciano, and W. Vecchiato: 2004, *Copula methods in finance*. John Wiley & Sons Ltd.

CIIRC: 2010, 'State of the Catalan coastline'. Technical report, CIIRC.

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.

Corbella, S. and D. D. Stretch: 2013, 'Simulating a multivariate sea storm using Archimedean copulas'. *Coastal Engineering* **76**, 68–78.

De Michele, C. and G. Salvadori: 2003, 'A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas'. *Journal of Geophysical Research* **108**(D2), 4067.

de Waal, D. J. and P. H. A. J. M. van Gelder: 2005, 'Modelling of extreme wave heights and periods through copulas'. *Extremes* **8**, 345–356.

DGP: 1992, 'ROM 0.3-91 Recommendations for waves and atlas of maritime climate on Spanish coast'. Direccion General de Puertos.

DGP: 2001, 'ROM 0.0 General procedures and computations basics. On maritime and harbour projects. Part I'. Direccion General de Puertos.

Eastoe, E., S. Koukoulas, and P. Jonathan: 2013, 'Statistical measures of extremal dependence illustrated using measured sea surface elevations from a neighbourhood of coastal locations'. *Ocean Engineering* **62**, 68–77.

Embrechts, P., C. Klüppelberg, and T. Mikosch: 1997, *Modelling extremal events, volume 33 of Applications of Mathematics*. New York. Springer-Verlag, Berlin.

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-León, M., V. Gràcia, L. Robichaux, A. Kroger, J. Gault, and A. Sánchez-Arcilla: 2015, *Evaluation of transient defence measures against storms*, Chapt. 166. World Scientific.

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.

Genest, C. and B. Remillard: 2004, 'Test of independence and randomness based on the empirical copula process'. *Test* **13**, 335–369.

Goda, Y.: 2010, *Random Seas and Design of Maritime Structures*, 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.

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

Hofert, M. and M. Machler: 2011, 'Nested Archimedean Copulas Meet R–The nacopula Package'. *Journal of Statistical Software*.

Holthuijsen, L.: 2007, *Waves in oceanic and coastal waters*. Cambridge University Press.

Hosmer, D., S. Lemeshow, and R. Sturdivant: 2013, *Applied Logistic Regression, 3rd Edition*. Wiley. 528 pp.

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

Li, F., P. van Gelder, R. Ranasinghe, D. Callaghan, and R. Jongejan: 2014, 'Probability modelling of extreme storms along the Dutch coast'. *Coastal Engineering* **86**, 1–13.

Lionello, P. (ed.): 2012, *The climate of the Mediterranean region*. Elsevier.

Mardia, K. and P. Jupp: 2009, *Directional statistics*, Vol. 494. John Wiley & Sons.

Martin-Soldevilla, M., M. Martin-Hidalgo, V. Negro, J. Lopez-Gutierrez, and P. Aberturas: 2015, 'Improvement of theoretical storm characterization for different climate conditions'. *Coastal Engineering* **96**, 71–80.

Masina, M., A. Lamberti, and R. Archetti: 2015, 'Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves'. *Coastal Engineering* **97**, 37 – 52.

Mazas, F., X. Kergadallan, P. Garat, and L. Hamm: 2014, 'Applying POT methods to the revised joint probability method for determining extreme sea levels'. *Coastal Engineering* **91**, 140–150.

Melby, J. and N. Kobayashi: 2011, 'Stone armor damage initiation and progression based on the maximum wave momentum flux'. *Journal of Coastal Research* **27**, 110–119.

Mendoza, E. T., J. Jimenez, and J. Mateo: 2011, 'A coastal storms intensity scale for the Catalan sea (NW Mediterranean)'. *Natural Hazards and Earth System Science* **11**(9), 2453–2462.

Mestres, M., M. Grifoll, and A. Sánchez-Arcilla: 2016, 'Analysis of current intensification in the Northwest Mediterranean shelf'. *Continental Shelf Research* **114**, 29 – 40.

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

Okhrin, O. and A. Ristig: 2012, 'Hierarchical Archimedean copulae: the HAC package'. Humboldt-Universitat zu Berlin.

Ortego, M., R. Tolosana-Delgado, J. Gibergans-Báguena, J. Egozcue, and A. Sánchez-Arcilla: 2012, 'Assessing wavestorm hazard evolution in the NW Mediterranean with hindcast and buoy data'. *Climatic Change* **113**, 713–731.

Pallarés, E., A. Sánchez-Arcilla, and M. Espino: 2014, 'Wave energy balance in wave models (SWAN) for semi-enclosed domains–appication to the Catalan coast'. *Continental Shelf Research* **87**, 41–53.

Pewsey, A., M. Neuhäuser, and G. Ruxton: 2013, *Circular statistics in R*. Oxford University Press.

Salvadori, G. and C. De Michele: 2004, 'Frequency analysis via copulas: Theoretical aspects and applications to hydrological events'. *Water Resources Research* **40**(12), W12511.

Salvadori, G. and C. De Michele: 2010, 'Multivariate multiparameter extreme value models and return periods: A copula approach'. *Water Resources Research* **46**.

Salvadori, G., C. De Michele, N. Kottegoda, and R. Rosso: 2007, *Extremes in nature: An approach using copulas*. Springer.

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., J. G. Aguar, J. Egozcue, M. Ortego, P. Galiatsatou, and P. Prinos: 2008a, 'Extremes from scarce data: The role of Bayesian and scaling techniques in reducing uncertainty'. *Journal of Hydraulic Research* **46**(sup2), 224–234.

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., M. García-León, V. Gràcia, R. Devoy, A. Stanica, and J. Gault: (in press), 'Managing coastal environments under climate change: pathways to adaptation'. *Science of the Total Environment*.

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

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

Tolman, H.: 2009, 'User manual and system documentation of WAVEWATCH III version 3.14.'. Technical report, NOAA / NWS / NCEP / MMAB.

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.

Tolosana-Delgado, R., M. Ortego, J. Egozcue, and A. Sánchez-Arcilla: 2011, 'Checking model-data weather hazard occurrence fit in the context of climate change'. In: *Proceedings of the 2011 Annual Conference of the International Association for Mathematical Geosciences*. pp. 870–877.

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

Unden, P., L. Rontu, H. Jarvinen, P. Lynch, J. Calvo, G. Cats, J. Cuixart, K. Eerola, C. Fortelius, J. A. Garcia-Moya, C. Jones, G. Lenderlink, A. McDonald, R. McGrath, B. Navascues, N. W. Nielsen, V. Odergaard, E. Rodrigues, M. Rummukainen, R. Room, K. Shattler, B. H. S. nd H. Savijarvi, B. Schreur, R. Sigg, H. The, and A. Tijm: 2002, 'HIRLAM-5 Scientific Documentation'. Technical report, HIRLAM-5 Project, c/o Per Unden SMHI, S-601 76 Norrkoping, SWEDEN. 144 p.

Wackernagel, H.: 2003, 'Ordinary kriging'. In: *Multivariate Geostatistics*. Springer Berlin Heidelberg, pp. 79–88.

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., 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'. *Natural Hazards and Earth System Science* **12**(2), 495–510.

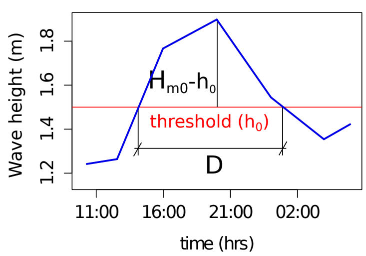
WAMDI Group, S. H., K. Hasselmann, P. A. E. M. Janssen, G. J. Komen, L. Bertotti, P. Lionello, A. Guillaume, V. C. Cardone, J. A. Greenwood, M. Reistad, L. Zambresky, and J. A. Ewing: 1988, 'The WAM model: a third-generation ocean wave predicition model'. *Journal of Physical Oceanography* **18**, 1775–1810.

WISE Group: 2007, 'Wave modelling-the state of the art'. *Prog. Oceanogr.* **75**, 603–674.

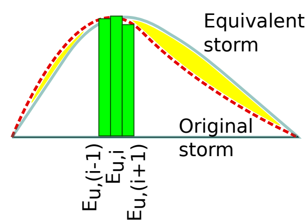
Wojtanowicz, A. M.: 2010, 'Twin Storm Events: A probability analysis and risk evaluation of twin storm occurrences along the coast of Catalonia'. Master's thesis, UPC Barcelona-Tech.

Zieger, S., A. V. Babanin, W. E. Rogers, and I. R. Young: 2015, 'Observation-based source terms in the third-generation wave model {WAVEWATCH}'. *Ocean Modelling* **96, Part 1**, 2 – 25. Waves and coastal, regional and global processes.

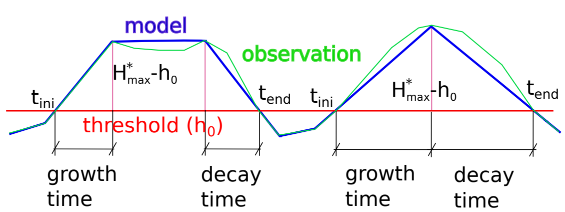
10. Figure Captions



(a)



(b)



(c)

Figure 1: a) Definition of variables for a single peak storm, where H_{m0} is the wave-height, D is the storm duration, b) definition of the peak-unitary-storm energy, $E_{u,p}$, where $E_{u,i}$ are the unitary-storm-energies at each hour (the red dashed line is the actual storm and the green line is an equivalent storm without the skewness problem), c) proposed storm shapes (irregular-trapezoid and triangular), where the parameters are initial time (t_{ini}), ending time (t_{end}), and our model's maximum wave-height (H^*_{max}).

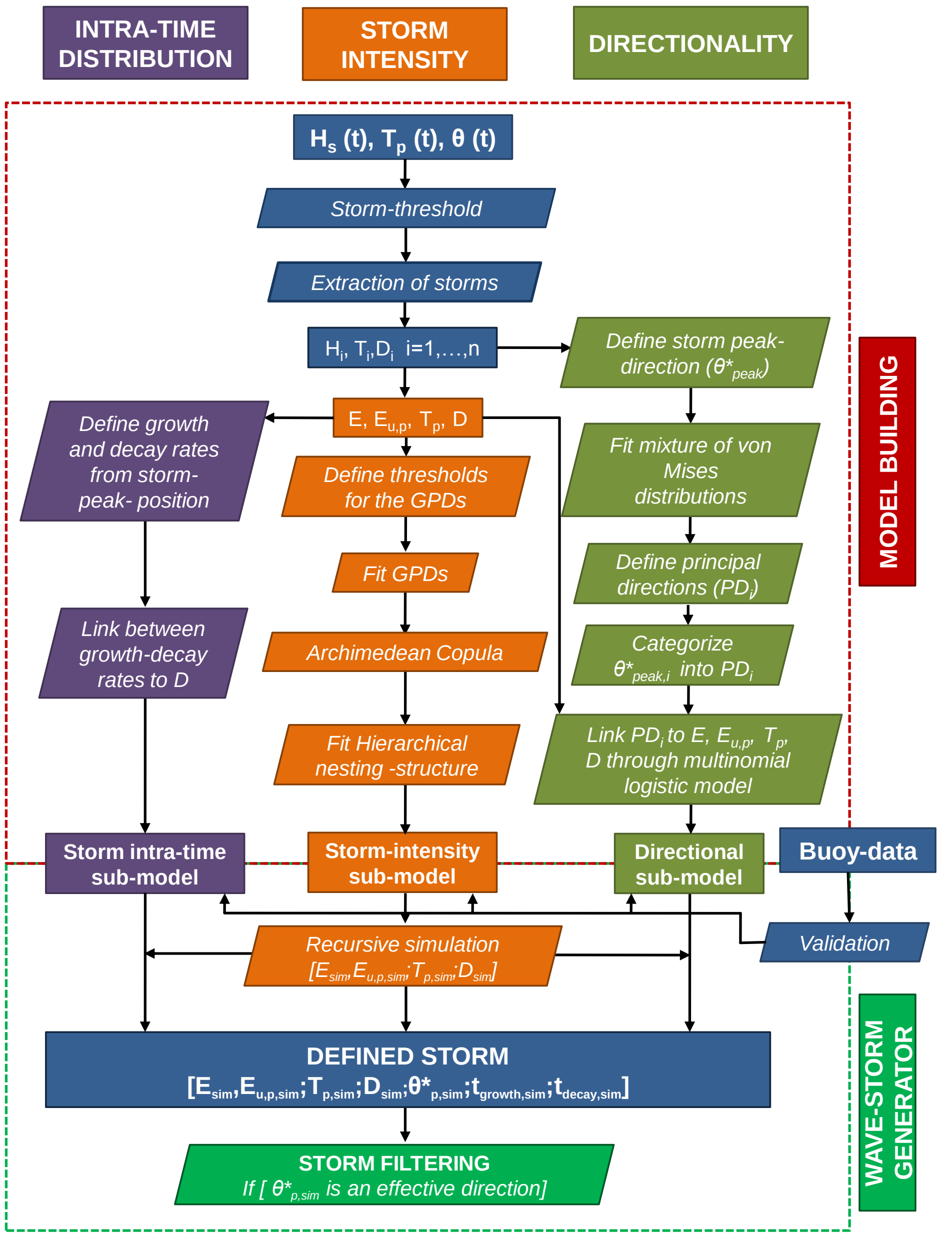


Figure 2: Flow-chart of the methodology used to construct the statistical storm model. The model is composed by three sub-models: intensity (orange), wave directionality (olive green) and intra-time (purple). Rectangle boxes represent input/output data whereas the parallelograms represent the actions taken.

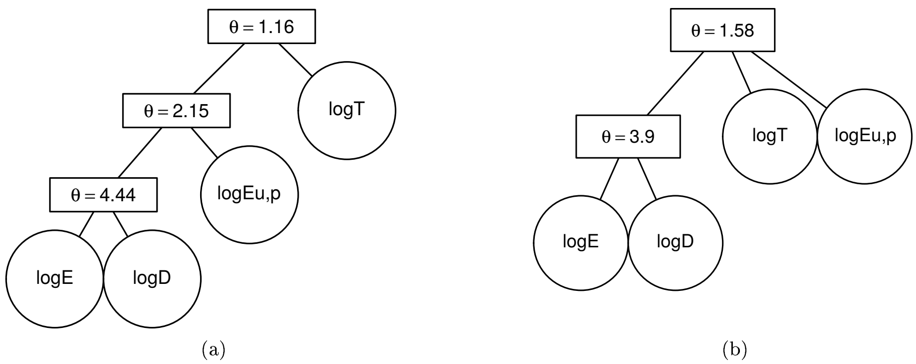


Figure 3: Types of HAC trees obtained for the Catalan Sea. a) Type-A: HAC structured with 3 levels of variable dependencies (at node N1), b) type-B: HAC structured with 2 levels of variable dependencies (at node N7). The upmost level is the «root». The variables sequentially cluster according to their dependence (θ) with other variables.

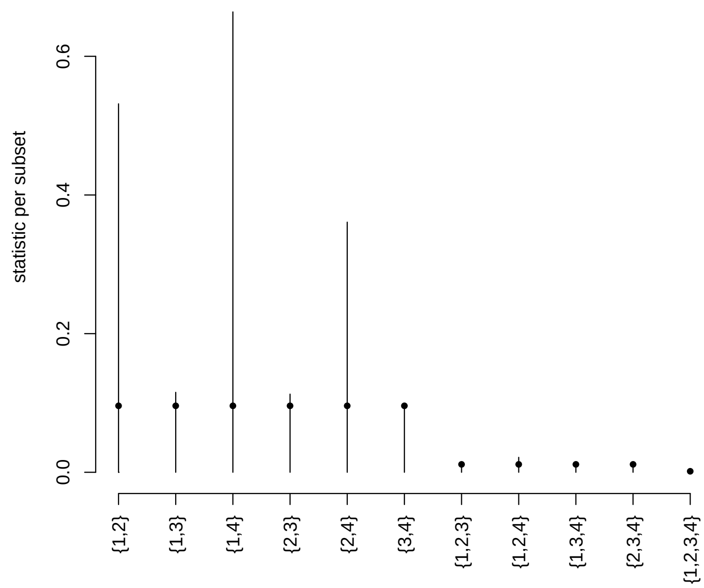


Figure 4: Dependogram: dependence among variables $\log E$ (1), $\log E_{u,p}$ (2), $\log T$ (3), and $\log D$ (4), at node C3. The length of the bar (statistic) exceeding the bullet (critical value) represents the degree of dependence. E and D present the greatest dependence, followed by the subsets $\{E, E_{u,p}\}$ and $\{E_{u,p}, D\}$.

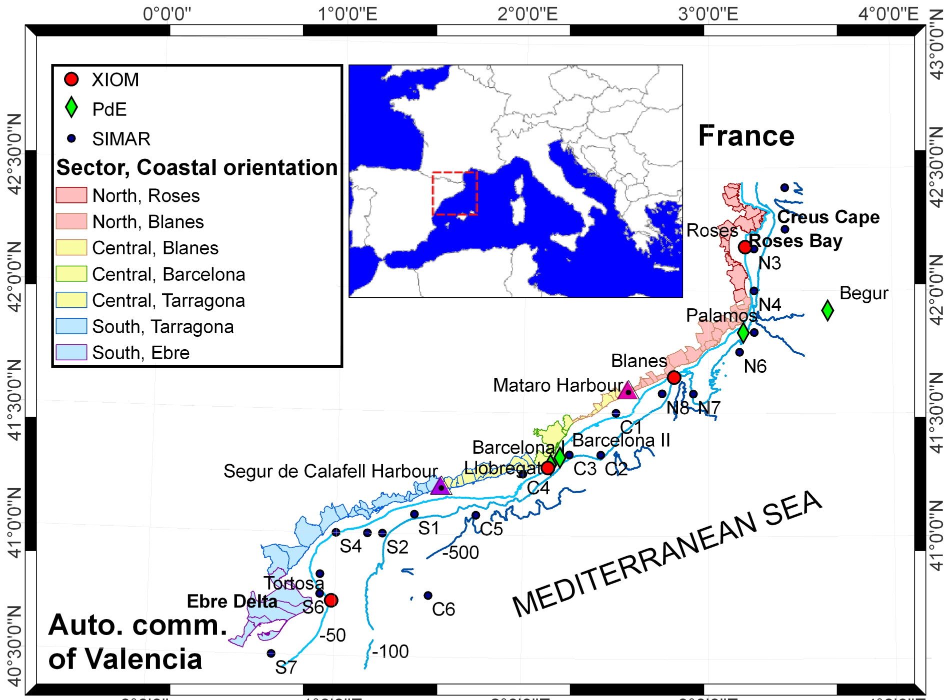
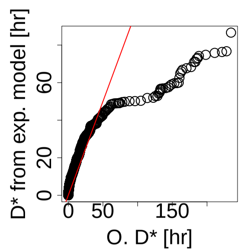
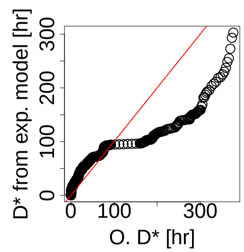


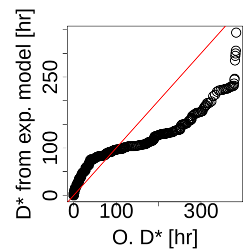
Figure 5: Map of the study area showing wave measurement networks (XIOM and PdE), and the SIMAR nodes. The colour lines of the regions (red, orange, green, blue and purple) and the coloured areas (red, yellow and blue) cluster the coast into the three sectors: North (France to Mataro harbour), Central (Mataro harbour to Segur de Calafell harbour) and South (Segur de Calafell harbour to the Autonomous Community of Valencia).



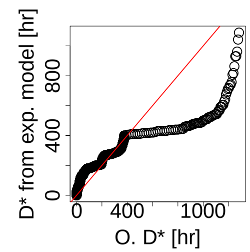
(a)



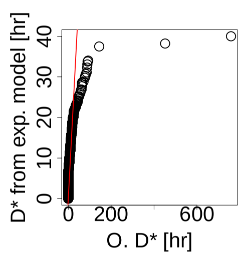
(b)



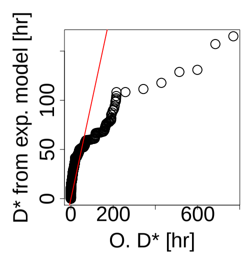
(c)



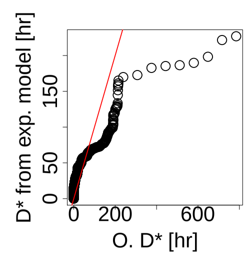
(d)



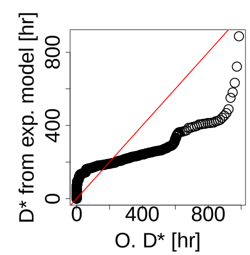
(e)



(f)

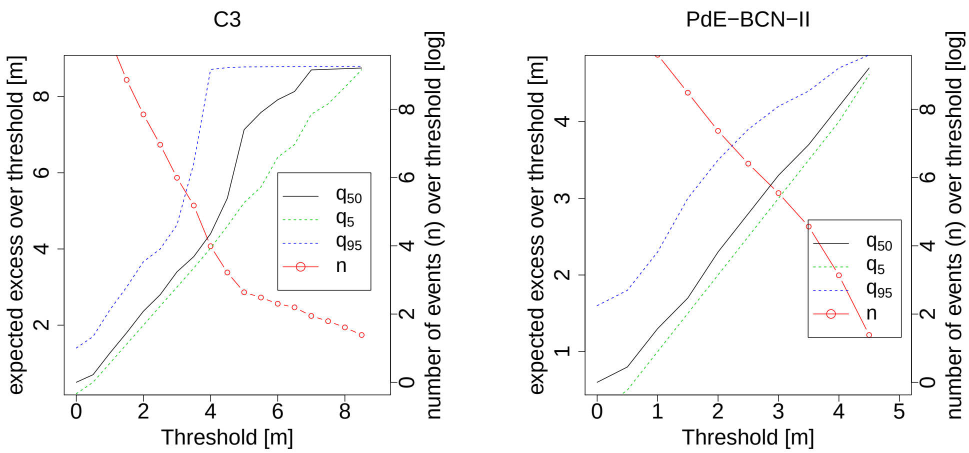


(g)



(h)

Figure 6: Q-Q plot of the observed D^* (x-axis) and D^* from the corresponding best-fit exponential model (y-axis). The red line represents null residuals (1:1 ratio), while the circles are the scatter points. From a) to d), the time lapses are obtained using a storm-threshold of: 1.5m, 2.1m, 2.2m, and 3m, respectively, for C3 node. From e) to h), it is the same for the PdE-Barcelona-II buoy node.



(a) (b)

Figure 7: Mean-excess-plot of H_{m0} for the a) SIMAR node C3 and b) PdE-BCN-II buoy node. The red line represents the log-transformed number of events over a given threshold, while q_{50}, q_5 and q_{95} are the quantiles 50, 5 and 95, of H_{m0}.

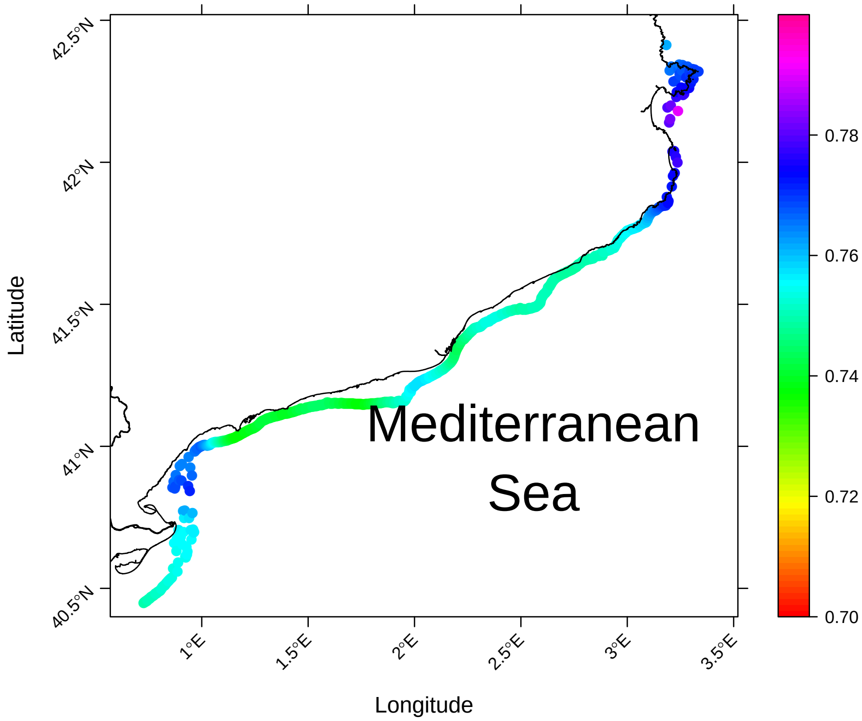
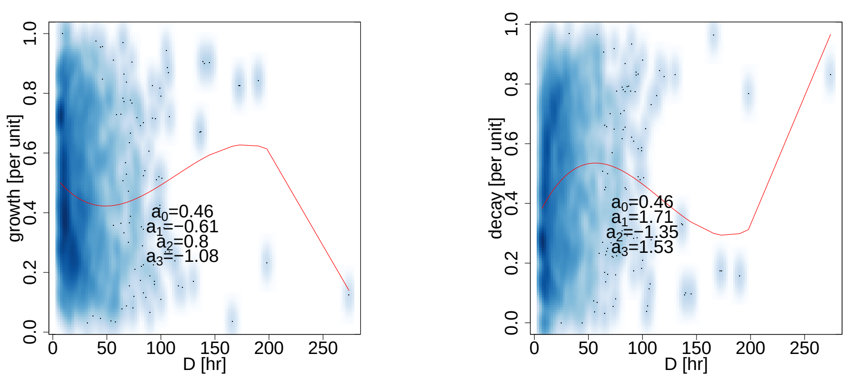


Figure 8: Spatial distribution of the Kendall's rank correlation coefficient (τ) between E and D ($\tau_{(E,D)}$). $\tau \in [0, 1)$, where 0 is total independence and 1 is total dependence.



(a)

(b)

Figure 10: Heat map of a) dimensionless growth-rate vs. D, and b) decay-rate vs. D, at node C3. Greater density is represented by darker blue colour.

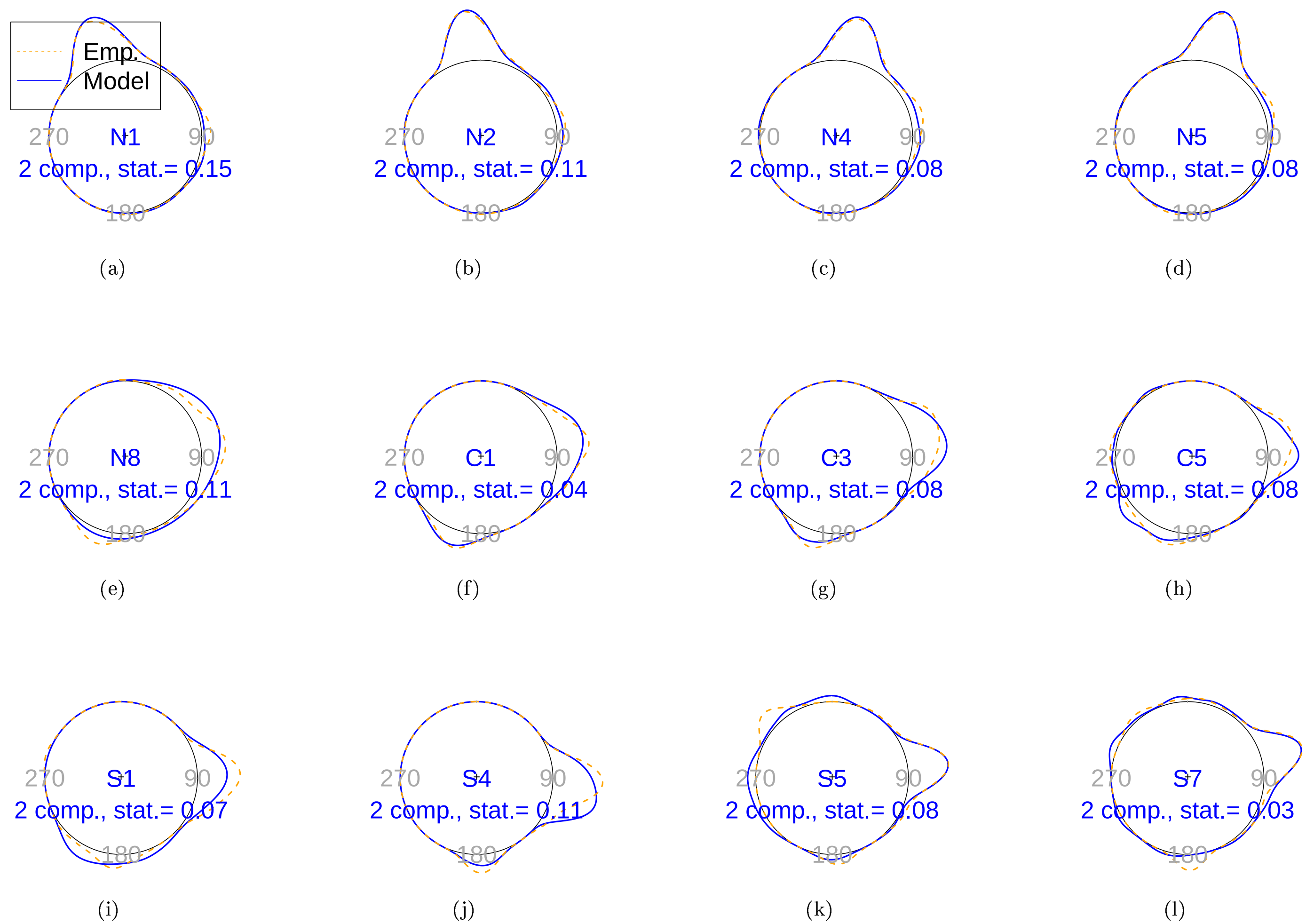


Figure 9: Fitting of movM distributions to θ_p^*. In dotted orange, the empirical distribution for each node; in continuous blue, the same for the movM distributions. The node name, the number of movM components and the value of the U^2 statistic are shown, for each node. The North is in the upper part of the directional distribution and the wave directions follow the Nautical convention.

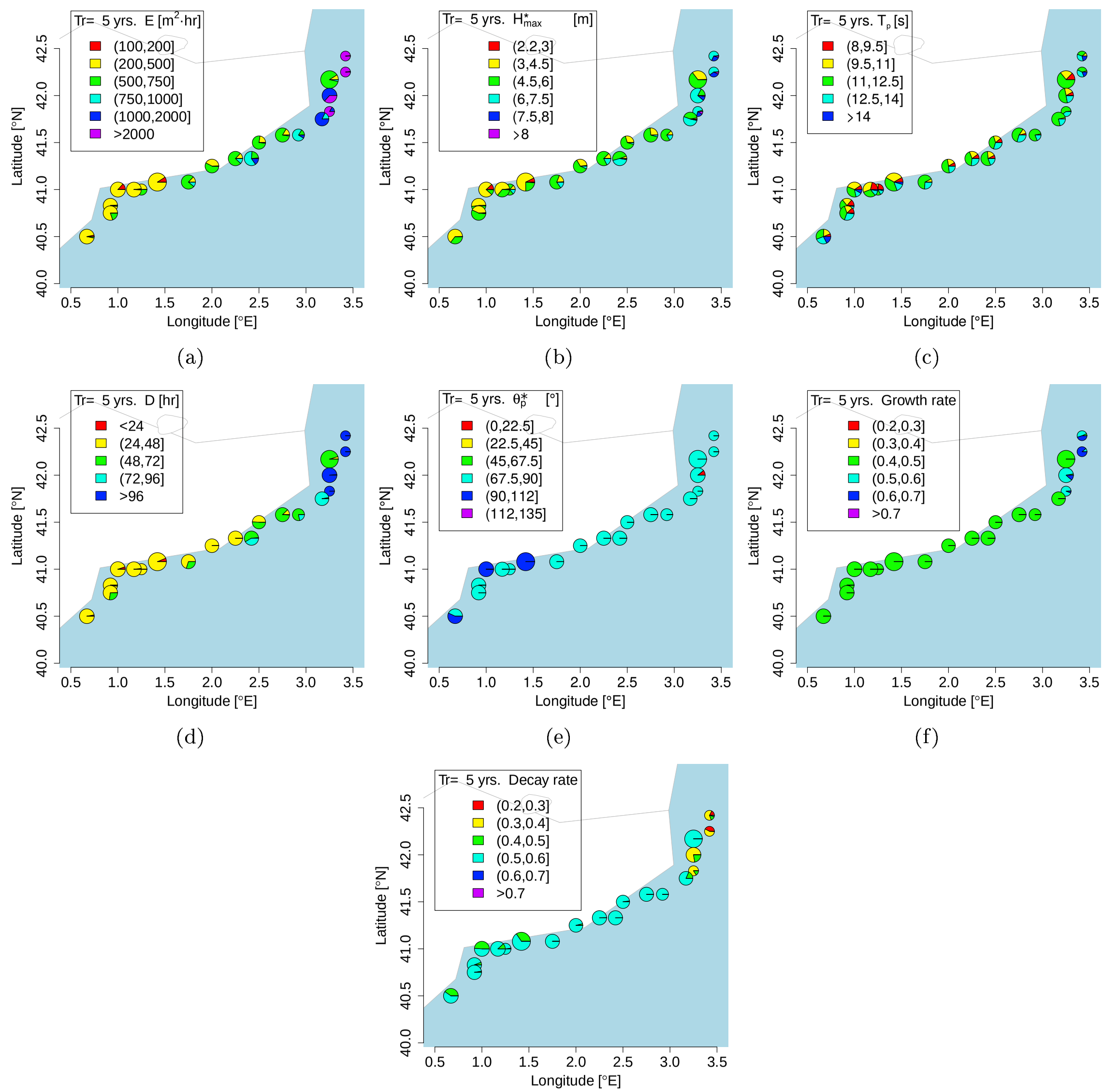


Figure 15: Storm components predicted by the proposed model, for T_r of 5years, and at each studied node: a) energy (E), b) maximum wave-height (H^*_{max}), c) peak-period (T_p), d) duration (D), e) wave-direction at peak (θ^*_p), f) growth-rate, and g) decay-rate. Pie-charts represents the frequency of a subset of intervals for each storm component.

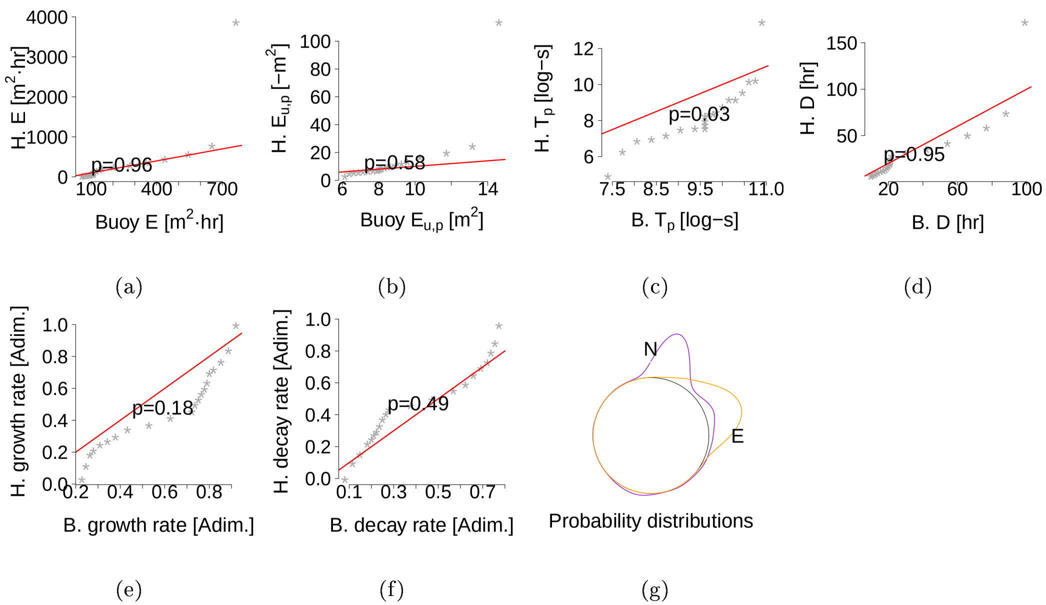


Figure 11: Q-Q plots of PdE-Palamos buoy vs. the SIMAR node N5. The x-axis is buoy data, and the y-axis is the hindcasted data (SIMAR). The orange lines in g) is the buoy θ_p^*, whereas the purple line is the SIMAR θ_p^*. The red straight line in the rest of the plots represents the perfect fit. “p” is the p-value, the higher it is, the better the fit.

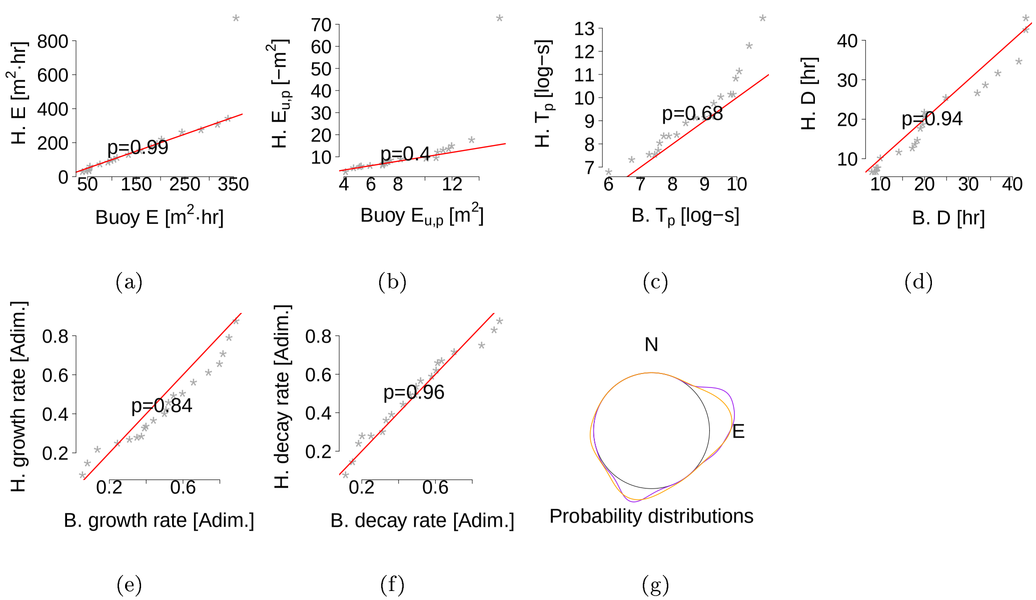


Figure 12: Q-Q plots of PdE-Barcelona-II buoy vs. the SIMAR node C3. The graphs are represented with the same elements (e.g. line colour) than for the PdE-Palamos buoy.

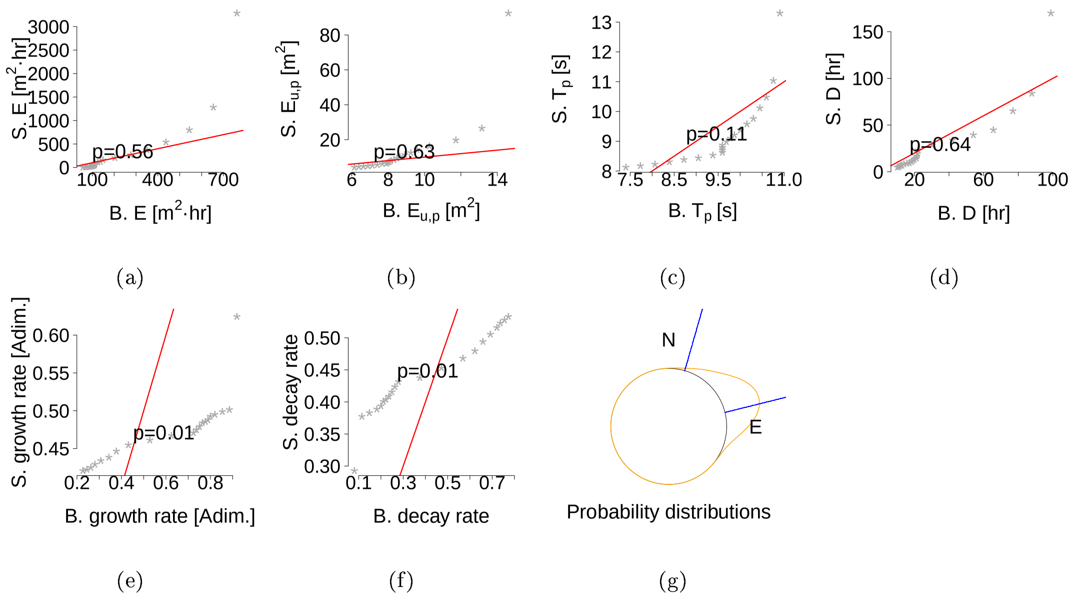


Figure 13: Validation of the node N5 our model by the PdE-Palamos buoy data. The x-axis is the buoy data, and the y-axis is the model generated data. In Fig. g), the blue line represents the model θ_p^* and the orange line represents the buoy θ_p^*. “p” is the p-value, the highest it is, the better the fit.

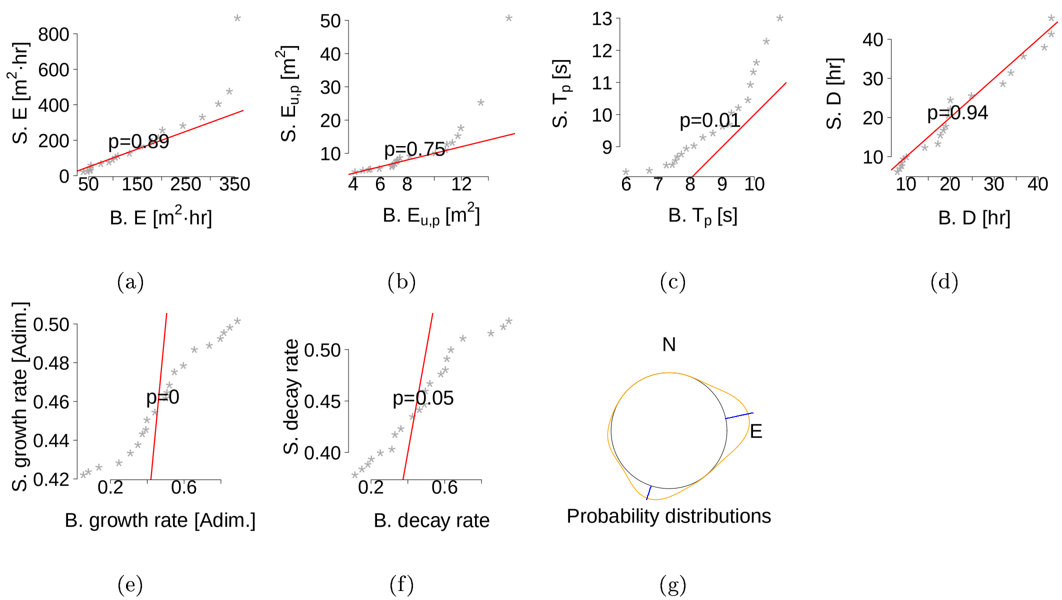


Figure 14: Validation of the node C3 our model by the PdE-Barcelona-II buoy data. Graphs are represented with exactly the same elements (e.g. line colour) than for the N5 case.

11. Tables

Table 1: Buoy location and data availability. All the considered buoys are directional.

Buoy	Longitude ($^{\circ}E$)	Latitude ($^{\circ}N$)	Depth (m)	Data availability
PdE-Palamos	3.19	41.83	90	26/03/2010 to 30/06/2011
XIOM-Blanes	2.82	41.65	74	13/07/2007 to 31/12/2012
PdE-Barcelona I	2.15	41.29	50	08/03/2004 to 22/12/2013
PdE-Barcelona II	2.20	41.32	68	08/03/2004 to 30/11/2011
XIOM-Llobregat	2.14	41.28	45	05/02/2004 to 31/12/2012
XIOM-Tortosa	0.98	40.72	60	15/06/1990 to 31/12/2012

Table 2: Number of storms per node.

Node	Storms	**Node**	Storms	**Node**	Storms	**Node**	Storms
N1	471	**N6**	201	**C3**	75	**S3**	44
N2	467	**N7**	134	**C4**	49	**S4**	31
N3	88	**N8**	62	**C5**	77	**S5**	59
N4	255	**C1**	60	**S1**	42	**S6**	73
N5	348	**C2**	99	**S2**	65	**S7**	52

Table 3: Parameters of the GPD adjusted to each SIMAR node: location (μ), scale (σ), and shape (ξ). The E is the storm energy, $E_{u,p}$ is the maximum unitary storm energy, T_p is the peak-wave-period associated to H^*_{max}, and D is the storm duration. The $h_0 = 2.2$m is the wave height threshold. The $D_{min} = 6$hrs is the required minimum storm duration or duration threshold. The $T_{min} = 8.17$s is the T_p threshold, obtained from CIIRC (2010).

	GPD parameters							
	$\log E$ ($\mu = D_{min} \cdot H_0^2$)		$\log E_{u,p}$ ($\mu = H_0^2$)		$\log T$ ($\mu = T_{min}$)		$\log D$ ($\mu = D_{min}$)	
Node	σ	ξ	σ	ξ	σ	ξ	σ	ξ
N1	2.65	-0.54	1.01	-0.34	0.10	-0.00	1.94	-0.50
N2	2.57	-0.52	0.98	-0.32	0.10	-0.02	2.00	-0.57
N3	2.42	-0.72	0.71	-0.30	0.33	-0.79	1.91	-0.76
N4	2.32	-0.50	0.81	-0.24	0.15	-0.24	1.83	-0.54
N5	2.37	-0.48	0.91	-0.27	0.14	-0.23	1.86	-0.55
N6	2.27	-0.55	0.81	-0.24	0.17	-0.29	2.08	-0.76
N7	2.36	-0.63	0.81	-0.26	0.25	-0.53	1.88	-0.72
N8	2.54	-0.75	0.81	-0.27	0.28	-0.53	1.77	-0.68
C1	2.31	-0.68	0.79	-0.25	0.31	-0.59	1.43	-0.56
C2	2.32	-0.61	0.85	-0.24	0.29	-0.61	1.72	-0.62
C3	2.20	-0.62	0.83	-0.25	0.27	-0.48	2.02	-0.99
C4	2.21	-0.64	0.81	-0.22	0.26	-0.47	1.87	-0.90
C5	2.24	-0.63	0.82	-0.24	0.21	-0.34	1.90	-0.87
S1	2.07	-0.76	0.66	-0.22	0.16	-0.12	1.53	-0.75
S2	2.20	-0.68	0.76	-0.25	0.17	-0.21	1.99	-0.95
S3	2.23	-0.76	0.71	-0.25	0.14	-0.08	1.78	-0.86
S4	2.04	-0.74	0.67	-0.23	0.16	0.01	1.99	-1.09
S5	1.87	-0.61	0.64	-0.20	0.28	-0.48	1.50	-0.68
S6	1.87	-0.59	0.68	-0.23	0.24	-0.38	1.45	-0.62
S7	1.64	-0.49	0.65	-0.20	0.16	0.00	1.31	-0.65

Table 4: Parameters of HACs. The selected copula type is Gumbel-HAC, and the aggregation method is “mean”. These parameters can be used to compare different locations.

Node	Tree type	$\theta_{(E,D)}$	$\theta_{((E,D),E_{u,p})}$	θ_{root}	Node	Tree type	$\theta_{(E,D)}$	$\theta_{((E,D),E_{u,p})}$	θ_{root}
N1	A	1.16	2.15	4.44	C3	A	1.27	2.03	3.79
N2	A	1.22	2.23	4.47	C4	A	1.29	1.97	4.17
N3	A	1.14	1.87	4.89	C5	B	1.62		3.54
N4	A	1.27	2.10	4.54	S1	B	1.48		3.74
N5	A	1.45	2.10	4.66	S2	A	1.18	1.81	3.74
N6	B	1.67		4.08	S3	A	1.33	1.92	3.69
N7	B	1.58		3.90	S4	B	1.55		4.30
N8	A	1.43	1.98	3.92	S5	A	1.22	2.01	4.47
C1	A	1.23	1.94	3.95	S6	A	1.29	1.84	4.11
C2	A	1.23	2.03	4.36	S7	A	1.59	2.19	3.89

Table 5: The wave directions at each node derive into pairs of $(\sin, \cos)$. The set of sines and the set of cosines are characterized by seldom movM distributions. Means (μ) of the movM distributions are provided for each principal principal direction (PD).

	Mean (μ) [°]				Mean (μ) [°]	
Node	PD1	PD2	PD3	**Node**	PD1	PD2
N1	344	84		**C3**	78	198
N2	353	76		**C4**	81	196
N3	73	353		**C5**	81	220
N4	11	78		**S1**	91	195
N5	15	76		**S2**	85	203
N6	23	88		**S3**	183	88
N7	74	33	205	**S4**	94	176
N8	81	200		**S5**	82	320
C1	84	198		**S6**	334	77
C2	70	205		**S7**	74	109

Table 7: Parameters of the function $f(D) = a_0 + a_1 D + a_2 D^2 + a_3 D^3$, where D is storm duration, and $f(D)$ is either growth or decay-rate.

	growth rate				decay rate			
	a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3
N1	0.48	-0.52	0.10	-0.39	0.45	0.85	-0.29	0.51
N2	0.48	-0.48	0.25	-0.57	0.45	1.09	-0.56	0.78
N3	0.48	-0.46	0.24	-0.55	0.45	1.17	-0.62	0.81
N4	0.47	-0.45	0.40	-0.68	0.45	1.35	-0.91	1.07
N5	0.47	-0.52	0.61	-0.82	0.45	1.48	-1.11	1.21
N6	0.47	-0.58	0.67	-0.90	0.45	1.50	-1.15	1.30
N7	0.47	-0.50	0.61	-0.86	0.45	1.47	-1.13	1.31
N8	0.47	-0.47	0.62	-0.89	0.45	1.46	-1.13	1.33
C1	0.47	-0.45	0.61	-0.90	0.45	1.45	-1.12	1.34
C2	0.46	-0.47	0.64	-0.92	0.45	1.47	-1.15	1.36
C3	0.46	-0.47	0.65	-0.94	0.46	1.49	-1.18	1.40
C4	0.46	-0.47	0.65	-0.94	0.46	1.50	-1.19	1.41
C5	0.46	-0.51	0.69	-0.97	0.46	1.56	-1.24	1.46
S1	0.46	-0.52	0.70	-0.98	0.46	1.59	-1.27	1.48
S2	0.46	-0.53	0.71	-1.00	0.46	1.59	-1.27	1.49
S3	0.46	-0.52	0.71	-1.00	0.46	1.61	-1.29	1.52
S4	0.46	-0.52	0.72	-1.02	0.46	1.63	-1.31	1.53
S5	0.46	-0.55	0.75	-1.04	0.46	1.66	-1.33	1.54
S6	0.46	-0.60	0.79	-1.07	0.46	1.70	-1.35	1.55
S7	0.46	-0.61	0.80	-1.08	0.46	1.71	-1.35	1.53

Table 6: Coefficients of the multivariate logit function to predict θ_p^* from $\log E$, $\log E_{u,p}$, $\log T$ and $\log D$. If there are three principal direction (PD), the function always uses the most principal component as a reference and displays the prediction for the other components (see column "PD#").

Node	PD#	Intercept	$a_{\log E}$	$a_{\log E_{u,p}}$	$a_{\log T,1}$	$a_{\log T,2}$	$a_{\log T,3}$	$a_{\log D,}$,	$a_{\log D,2}$	$a_{\log D,3}$
N1	2	-1.26	25.38	-63.57	92.02	14.70	-4.37	-21.49	-8.15	3.84
N2	2	-1.10	3.37	-45.61	78.22	13.03	-5.25	-7.70	-10.36	0.96
N3	2	-0.10	-21.12	0.05	-19.67	-0.31	3.74	25.05	-1.53	1.15
N4	2	-0.47	-11.27	5.63	15.35	4.40	-0.57	3.48	0.04	1.80
N5	2	-0.74	-4.10	-8.34	17.40	11.11	0.83	-1.63	-4.04	-1.53
N6	2	-0.34	28.49	-5.24	-0.10	7.76	-0.79	-21.77	-6.14	4.94
N7	2	-0.10	-61.42	17.76	-1.56	0.21	-3.80	45.05	0.46	-0.08
N7	3	-0.99	-63.15	23.87	-13.30	-9.98	-5.55	49.31	-3.11	0.66
N8	2	-1.66	-36.08	12.13	-6.08	-9.84	-8.40	25.42	-3.25	1.99
C1	2	-1.12	-36.46	12.11	1.74	-4.65	2.91	24.59	-1.72	-1.32
C2	2	-1.36	-35.21	13.17	-1.91	-8.35	-1.57	23.79	-5.83	-1.29
C3	2	-1.62	-37.34	12.22	7.13	-9.50	6.30	22.78	-2.36	-6.66
C4	2	-1.77	-6.53	5.34	0.69	-9.34	1.90	4.30	0.40	-3.56
C5	2	-0.19	-21.52	12.48	-16.53	2.32	-14.11	11.23	3.30	-6.14
S1	2	-1.22	21.23	-8.58	-3.61	4.73	-8.45	-16.31	-1.53	-4.63
S2	2	1.05	56.13	-22.29	-29.22	23.29	-10.70	-39.64	3.25	-0.43
S3	2	-1.80	-22.87	14.49	35.72	-29.01	9.78	18.48	3.96	7.55
S4	2	-2.75	26.24	-6.02	-4.28	-5.24	-5.97	-31.02	-13.96	-10.79
S5	2	4.01	40.45	-21.28	-64.86	43.36	-13.80	-35.83	-2.55	-2.51
S6	2	-3.65	-13.77	4.41	60.06	-31.89	15.36	26.01	8.89	9.07
S7	2	3.83	55.13	-17.94	-65.67	50.19	-25.26	-37.77	4.19	2.63