Storm characterization and simulation for damage evolution models of maritime structures

Andrea Lira-Loarca[a], Manuel Cobos[a,*], Miguel Ángel Losada[a], Asunción Baquerizo[a]

[a]*Andalusian Institute for Earth System Research. University of Granada.*
Avda. del Mediterráneo s/n, 18006 Granada, Spain.

Abstract

This paper presents a new approach to statistically characterize and simulate the wave climate under storm conditions. The methodology includes the joint selection of the parameters that identify storm events (significant wave height threshold, minimum storm duration and minimum interarrival time between consecutive storms) by means of hypothesis testing on the distribution functions of the number of storm events and the elapsing time between storms, providing an improved characterization of the parameters that define storm events. The main wave variables and their temporal dependence are characterized by non-stationary mixture distribution functions and a vector autoregressive model. This allows to adequately reproduce the random temporal evolution of storm events, crucial for the study of damage progression in maritime structures without the use of predefined geometries. The long-term time series of storm events and calm periods is obtained using copula functions which analyze the joint dependence of storm duration and interarrival time for separate climate intervals. The model is applied to hindcast data at a location of the Mediterranean sea close to the Granada coast in Spain to show its ability to reproduce wave storm conditions accounting for the time variability of the storminess. An example of application, using a large number of simulations and a damage progression model in a maritime structure, is presented.

Keywords: threshold selection, storm characterization, storm evolution, non-stationary mixture probability models, damage progression

1. Introduction

The design of coastal infrastructures usually considers wave height as one of the main maritime variables that triggers successive stress states affecting their operationality, serviceability and reliability. Different recommendations and manuals such as the U.S. Coastal Engineering Manual (U.S. Army Corps of Engineers, 2002), the Spanish Recommendations of Maritime Works ROM (Losada, 2001), the EurOtop manual (Van der Meer et al., 2018) and the Rock Manual (CIRIA et al., 2007) focus on the analysis of the ultimate

---

limit state failure modes, which produce the collapse of the structure. The classical approach to study them usually involves the use of extreme value theory and fits an extreme probabilistic model to the annual maximum significant wave height or to the peaks over a given threshold (POT). More modern approaches gather the scientific developments regarding damage progression (Castillo et al., 2012; Melby and Kobayashi, 1999; Sousa and Santos, 2006, among others) and highlight the importance to assess the loss of functionality of the structure and the study of the so called serviceability limit states which are closely related to the storm evolution and duration.

These studies require the definition of the parameters that define a storm, in particular the value of the significant wave height threshold which is related to the minimum admissible damage of a sea state, and to simulate the temporal evolution of the storms to account for the damage progression (Losada, 2018).

Storm events are customarily defined as independent events during which the significant wave height, H_s, exceeds a certain threshold, $H_{s,u}$. Their identification for extreme data analysis is usually done by setting both the values of $H_{s,u}$ and the minimum time interval between consecutive events, δ_0, required to guarantee their independence. There are several works devoted to the selection of the threshold in different fields of expertise and methods such as graphical methods involving the stability of the model parameters, the mean residual life plot, goodness-of-fit (Bernardara et al., 2014; Coles et al., 2001), automatic or quasi-automatic computational approaches (Liang et al., 2019; Solari et al., 2017; Thompson et al., 2009), among others. In coastal engineering applications the threshold value may also depend on the particular problem under consideration (a relatively high threshold is required to study breakwater damage while smaller values have to be chosen to assess coastal retreat). Therefore, the selection needs certain expertise and engineering judgement (Goda, 2010). Regarding the selection of the minimum interarrival time between events, δ_0, it varies in literature from hours to days depending on the region and it is almost always considered a constant value, i.e. is independent from the selected threshold (Li et al., 2014; Méndez et al., 2006). In addition, when the analysis is aimed at the study of processes in which persistence is a key element, as it is the case of damage evolution assessment, the determination of storms needs the definition of a minimum storm duration d_0 to consider only significant loads. For a given sample, there might be several ($H_{s,u}$, δ_0) combinations capable of ensuring, with a reasonable reliability, storm independence and the validity of the model (Liang et al., 2019). Given the relations between parameters and their influence on the resulting extreme events, there is still need to develop methodologies for the joint characterization of the parameters that define storm events ($H_{s,u}$, d_0, δ_0) providing that it can be checked that the underlying assumptions are met.

The increased interest in recent years in damage progression models related to the loss of functionality of maritime structure requires not only a definition of the storm in terms of its duration and peak value but a characterization of its temporal evolution (storm shape). Several authors have used an idealized "storm shape" such as a triangular geometry and searched for its equivalent with a real storm by comparing their magnitude (Boccotti, 2000), power (Fedele and Arena, 2009), or duration (Corbella and Stretch, 2012).

However, the storm history evolution of the sea states in sea and swell waves is not adequately reproduced by these theoretical models and do not provide enough information for applications where the features and evolution of the metocean time series are relevant (Jäger et al., 2018; Martín-Soldevilla et al., 2015) and there is still need to account for the random temporal evolution of storm events.

Other works propose different methodologies for simulating time-dependent series for meteorological and oceanic variables. Solari and Losada (2011) proposed a methodology that used a non-stationary mixture model of the significant wave height distribution and modeled the time dependence with the aid of copulas. Jäger and Nápoles (2017) used vine-copulas to simulate significant wave height and mean periods. In these and other works, the wave direction was not included as a main variable on the study or was just studied independently justified by (i) limitations in dataset availability, (ii) waves impinging from a limited range of directions, and (iii) a weak correlation between significant wave height and wave direction (Li et al., 2014). However, for the study of certain processes, the wave direction cannot be disregarded. Jäger et al. (2018) presented a method to jointly simulate wave climate time series of the significant wave height, the mean wave period and the mean direction, this last one as a categoric variable taking two possible values. On the basis of non-stationary mixture distributions and a vector autoregressive (VAR) model Solari and Van Gelder (2011) proposed a methodology to simulate multivariate wave climate series including the mean direction fitted to a continuous variable distribution. The modeling of complete time series allows for the characterization of the evolution of metocean variables but increases the computational costs when the simulation of only storm conditions would be, most of the times, enough to address problems related to operationality, reliability and serviceability of maritime works.

The aim of this paper is the development of a simple, robust and efficient methodology for the analysis and full temporal simulation of storm events for its application in damage evolution models of maritime structures. The proposed tool is site-specific and includes the statistical analysis of historical data to define (i) a rigorous and joint selection of the parameters that allow the identification of storms which are generally defined independently and (ii) the non-stationary joint characterization of the involved random variables with emphasis on their non-stationary behavior during storm events and (iii) a methodology to simulate multivariate time series under storm conditions that jointly reproduces the significant wave height, the peak period, and the mean incoming wave direction taking into account the seasonal climate variability allowing to perform an uncertainty assessment of the related problem. The model is applied to simulate a large number of storm events time-series for their use in a probabilistic study of damage evolution of a breakwater.

The paper is organised as follows. In section 2 the methodology is presented, including the criterion proposed for the definition of the storm (§2.1), the joint statistical characterization of the wave variables (§2.2) and the simulation procedure (§2.3). In section 3, the methodology is applied to a case study in the Mediterranean coast of Spain and the verification of the results is presented in section 4. An example of the simulation of long-term time series of storm events for the probabilistic assessment of damage progression

in a breakwater is presented in section 5. Section 6 presents a discussion regarding the applicability of the methodology. Finally, section 7 summarizes the main findings derived from this work.

2. Methodology

This work proposes a methodology that uses historical wave climate data to (i) properly define the extreme events by jointly selecting the significant wave height threshold, the minimum storm duration and the minimum interarrival time, (ii) stochastically characterize the distribution functions of the significant wave height, the peak period and the mean incident wave direction and their multivariate temporal dependence and (iii) simulate long-term time series of wave climate under storm conditions. Figure 1 presents a diagram of the methodology indicating the subsection where each step is presented.
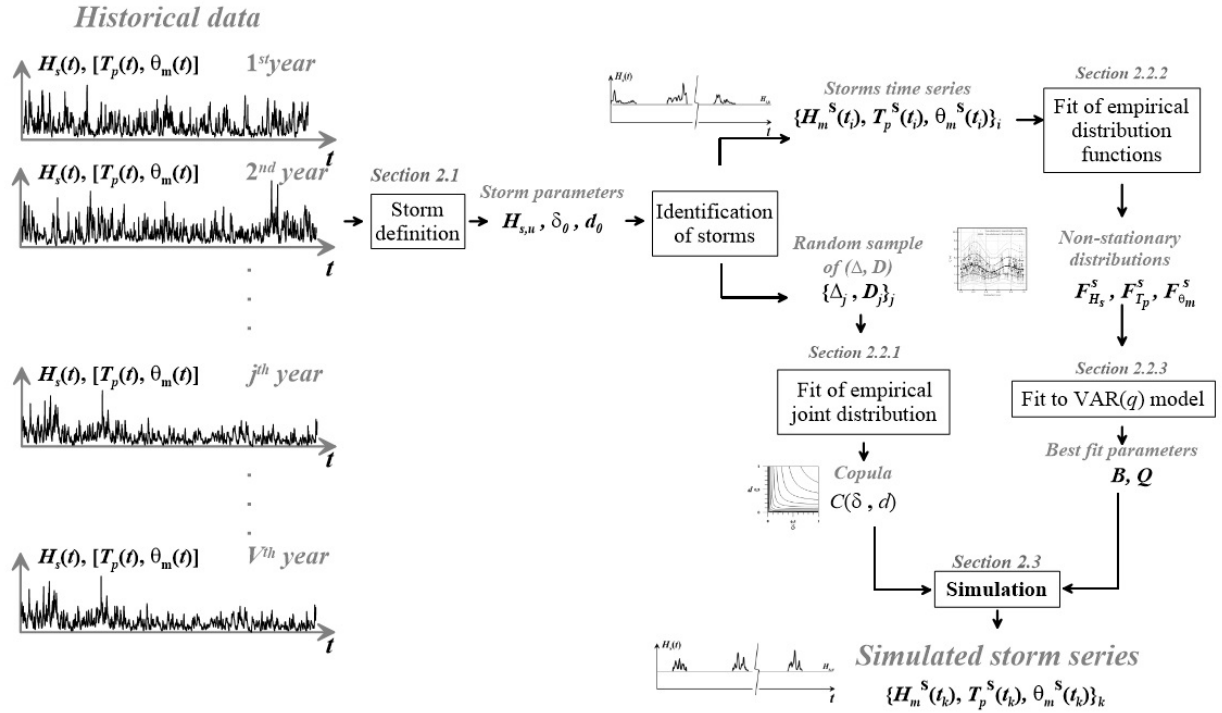

Figure 1: Flow diagram of the methodology

2.1. Criterion for storm definition

In this context, a storm is defined as a rare event that occurs when the significant wave height, H_s, exceeds a certain threshold, $H_{s,u}$. The selection of the threshold has to guarantee that the events are independent so that the counting process, N, describing the number of occurrences in a certain time interval, follows a Poisson distribution or, equivalently, the interarrival time between storms, Δ, follows an exponential distribution. Taking the year as temporal reference, the parameter of the Poisson distribution, λ, is the

4

mean annual number of storms, that coincides with the inverse of the expected value of the corresponding exponential distribution of the interarrival time between storms, and they are usually estimated from available historical or hindcast data. In practice, for a given selected threshold, $H_{s,u}$, the identification of in-
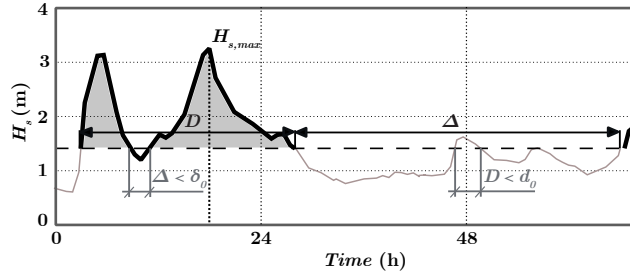

Figure 2: Definition sketch of a storm event and associated variables: storm duration, D, minimum storm duration, d_0, interarrival time between successive storms, Δ, and minimum interarrival time, δ_0.

dividual storms also requires the definition of a minimum storm duration, d_0, to avoid interpreting relatively small and short exceedances of the significant wave height as storms, and a minimum interarrival time, δ_0, between successive events, so that the independence assumption is fulfilled and the number of events follow a Poisson distribution (Figure 2). Both values depend on the threshold and cannot be set arbitrarily given that a too low δ_0 or d_0 results in an overestimation of the number of storms, whereas a too high d_0 would give an underestimated number of storms and a too high δ_0 leads to the overestimation of the storms duration. Moreover, these parameters are site specific since they are linked to physical atmospheric phenomena and depend, among others, on atmospheric wave generation conditions and fetch domain. Taking this into account, we propose a criterion for the joint selection of $H_{s,u}$, δ_0 and d_0 with the following hypotheses tests:

1. For the distribution of the number of events, N:

   Null hypothesis H_0^N: N follows a Poisson distribution

   Alternative hypothesis H_1^N: N does not follow a Poisson distribution

2. For the interarrival times, Δ:

   Null hypothesis H_0^Δ: Δ follows an exponential distribution

   Alternative hypothesis H_1^Δ: Δ does not follow an exponential distribution

To evaluate the goodness-of-fit of the annual number of storms distribution, N, a χ^2 non-parametric test is performed with samples obtained from hindcast data: $n_1, n_2, ...n_V$ with V being the number of years available in the dataset. In the case of the interarrival time Δ, a Kolmogorov–Smirnov (KS) test is used, with $\delta_1, \delta_2, ..., \delta_K$, where $K = \sum_{v=1}^{V} n_v - 1$, the number of interarrival times in the hindcast dataset. We then obtain the p-values, namely, p_N and p_Δ of the tests and check whether the conditions $p_N > \alpha$ and $p_\Delta > \alpha$ are fulfilled, with α being the chosen significance level. The nonrejection region of the null hypotheses is defined as the intersection between the regions where $p_N > \alpha$ and $p_\Delta > \alpha$. Additionally, due to the equivalence

5

of the null hypotheses H_0^N and H_0^Δ, it should be checked that $\bar{\lambda}$ and $\bar{\delta}$, indicating the estimations of the expected values of N and Δ, respectively, are related so that $\bar{\lambda} \approx 1/\bar{\delta}$.

The value of α represents the probability of incorrectly rejecting the null hypothesis, H_0, usually referred to as Type I error. The power of the test, that measures the probability of correctly rejecting the null hypothesis when it is not true, called $1 - \beta$ (where β is the so called Type II error) depends on the value of α, the size of the sample and the alternative hypothesis, H_1. Due to the definition given to H_1 (open to any other distribution than the one proposed in H_0), the power of the test cannot be estimated. It is known, however, that it is reduced as the value of α decreases.

2.2. Stochastic characterization of maritime variables and their dependence

Once the value of the tuple $(H_{s,u}, d_0, \delta_0)$ is set, the following information is obtained from the hindcast wave data: (1) samples of the annual number of storms, N, storms duration, D, and interarrival times, Δ, and (2) multivariate time series during storm conditions of the significant wave height, $H_s^s(t)$, the concomitant peak period $T_p^s(t)$ and mean incoming wave direction $\theta_m^s(t)$.

The stochastic characterization and temporal dependence of extreme events comprises (i) the use of a copula model to characterize the distribution of (D, Δ) presented in §2.2.1, (ii) the fit of univariate non-stationary mixture distributions $F_{H_s}^s$, $F_{T_p}^s$ and $F_{\theta_m}^s$ (see §2.2.2), (iii) a VAR(q) model to characterize the multivariate temporal dependence of $H_s^s(t)$, $T_p^s(t)$ and $\theta_m^s(t)$ presented on §2.2.3.

2.2.1. Interdependence of storm events via Archimedean copulas

It cannot be assumed that the duration of a storm, D, and the interarrival time with the following storm, Δ, are independent variables (Li et al., 2018). In addition, storm events usually show a seasonal variation in mid-latitudes. Therefore, in this study the characterization of (D, Δ) was done for different climatic periods that must be selected depending on location. For each climatic period, the joint distribution function F of D-Δ is obtained with a copula function, C, (Sklar, 1959) as:

$$F(d, \delta) = C(F_D(d), F_\Delta(\delta)), \tag{1}$$

where F_D and F_Δ are the marginal cumulative distribution functions (CDF) of the storms duration and interarrival times, respectively. For the selection of the best-fitting copula, given that there is not a clear procedure for selecting it (De Michele et al., 2007), we have, therefore, searched among different families and found that the Archimedean Clayton copulas provided the best visual fit to the data. This is in agreement with other studies where the Clayton, Frank, and Gumbel copulas have been used for extreme events applications (Li et al., 2018; Martín Hidalgo, 2015, and references herein). The Clayton family reads as follows:

$$C_{\theta_c}(d, \delta) = \left(\max\left[d^{-\theta_c} + \delta^{-\theta_c} - 1, 0\right]\right)^{-\frac{1}{\theta_c}}, \tag{2}$$

where $\theta_c = 2\tau/(1-\tau)$ is the dependence parameter obtained from the Kendall rank correlation coefficient, τ. The empirical CDFs of D and Δ can be calculated from historical or hindcast data.

2.2.2. Univariate non-stationary distributions

The univariate distributions of the involved random variables are estimated from the available time series. In a similar fashion to Solari and Losada (2011), the probabilistic models are considered non-stationary by characterizing each parameter a ($a = \xi, u, \sigma_i, \alpha_i, \mu_i, \cdots$) as a time-dependent function whose Fourier series expansion truncated to N_F oscillatory terms is:

$$a(t) = a_0 + \sum_{l=1}^{N_F}\left(a_l\cos(2\pi l t) + b_l\sin(2\pi l t)\right), \tag{3}$$

where a_0, a_l and b_l, $l = 1, ..., N_F$ are the coefficients of the trigonometric (Fourier) expansion of parameter a. The time is, however, not explicitly included in the notation of the distributions for the sake of simplicity.

Note that the random variable H_s is non-negative, unbounded above, and has an asymptotic probability function with a clear difference between the central body and the upper tail. Then, a non-stationary mixture probability density function, $f(H_s)$, is used as proposed by Solari and Losada (2011), capable of reproducing the statistical variability at the tails and for different time scales. The mixture distribution is composed by three functions describing the lower tail (f_m), the central regime (f_c) and the upper tail (f_M) as:

$$f(H_s) = \begin{cases} f_m(H_s)F_c(u_1) & H_s < u_1 \\ f_c(H_s) & u_1 \leq H_s \leq u_2 \\ f_M(H_s)\left(1 - F_c(u_2)\right) & H_s > u_2 \end{cases}, \tag{4}$$

where F_c is the distribution function of the central regime modeled by any parametric continuous distribution; u_1 and u_2 are the location parameters, taken as the lower and upper limits of the central regime. The tails are modeled following a Generalized Pareto Distribution (GPD). Therefore, f_m is the GPD of the minima and f_M is the GPD of the maxima, given by:

$$f_m(H_s; \xi_m, u_1, \sigma_m) = \frac{1}{\sigma_m}\left(1 - \frac{\xi_m(H_s - u_1)}{\sigma_m}\right)^{-\frac{1}{\xi_m}-1}, \tag{5}$$

$$f_M(H_s; \xi_M, u_2, \sigma_M) = \frac{1}{\sigma_M}\left(1 + \frac{\xi_M(H_s - u_2)}{\sigma_M}\right)^{-\frac{1}{\xi_M}-1}, \tag{6}$$

where ξ_α, u_α and σ_α, ($\alpha = m, M$) are the shape, location and scale parameters that, as said before, are assumed to be time-dependent. The values of thresholds u_1 and u_2 and of the parameters of the

three probability functions in Equation 4 are obtained by maximizing the negative log-likelihood function (NLLF). This logarithmic function is composed by the corresponding PDFs of the lower tail, the body and the upper tail evaluated with data in the ranges $(-\infty, u_1)$, $[u_1, u_2]$ and (u_2, ∞), respectively. The continuity is guaranteed by imposing that the CDFs estimated at u_1 and u_2 coincide with the values obtained with the corresponding expressions at the neighbouring regions. These conditions and the imposed limitation that the lower limit of the minima GPD should be zero, give the following relationships between the parameters:

$$\sigma_m = \frac{F_c(u_1)}{f_c(u_1)}; \qquad \xi_m = -\frac{\sigma_m}{u_1}; \qquad \sigma_M = \frac{1 - F_c(u_2)}{f_c(u_2)}. \tag{7}$$

As a result, the number of parameters to fit is reduced to five: u_1, u_2, the scale and location parameters of the LN, and the shape parameter of the upper tail probability model. It is worth noting that there is no relation between the thresholds u_1 and u_2 of the mixed probability function and the threshold $H_{s,u}$ used to define storm events.

The distribution function of H_s under storm conditions, is then estimated as follows:

$$F_{H_s}^s(H_s) = \frac{F_{H_s}(H_s) - F_{H_s}(H_{s,u})}{1 - F_{H_s}(H_{s,u})}. \tag{8}$$

Unlike with H_s, where the complete dataset was used for the statistical characterization, the peak period and mean wave direction probability models were fitted with concomitant data to H_s that fulfilled $H_s \geq H_{s,u}$. The peak period under storm conditions, T_p^s, is fitted to a parametric single non-stationary continuous distribution function with the most commonly used being the Lognormal distribution. The probability model reads as follows:

$$f(T_p^s; \mu, \sigma) = \frac{1}{T_p^s \sigma \sqrt{2\pi}} e^{-\left(\ln(T_p^s) - \mu\right)^2 / 2\sigma^2}. \tag{9}$$

The mean wave direction under storm conditions, θ_m^s, is described by a sum of N_{TN} stationary truncated normal (TN) functions limited to $(0°, 360°)$ (see e.g., Fisher (1995)) in a similar fashion as (Solari and Losada, 2016), where N_{TN} is the number of main wave directions. The use of stationary functions in this case was motivated by the lack of a clear temporal variability in the wave direction. This assumption is equivalent to keeping the mean term of Eq. 3 and neglecting the expansion coefficients. The probability density function of wave direction is given by:

$$f(\theta_m^s; \alpha_i) = \sum_{i=1}^{N_{TN}} \alpha_i f_{TN_i}(\theta_m^s), \tag{10}$$

where the sum of α_i is equal to 1, and:

$$f_{TN_i}(\theta_m^s; \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(\theta_m^s - \mu_i)^2}{2\sigma_i^2}}. \tag{11}$$

8

2.2.3. Mutlivariate temporal dependence of storm events

From the series under storm conditions, $\left(H_s^s(t),\ T_p^s(t),\ \theta_m^s(t)\right)$, stationary normalized times series are obtained as:

$$Z_{H_s}(t) = \Phi^{-1}\left(F_{H_s}^s(H_s^s(t))\right)$$
$$Z_{T_p}(t) = \Phi^{-1}\left(F_{T_p}^s(T_p^s(t))\right)$$
$$Z_{\theta_m}(t) = \Phi^{-1}\left(F_{\theta_m}^s(\theta_m^s(t))\right), \tag{12}$$

where Φ^{-1} is the inverse of the normal cumulative distribution function with null mean and standard deviation equal to one. The temporal dependence between variables during storm events is characterized by means of a stationary VAR(q) model which assumes a linear relationship between the variables and their past values. The parameters of the VAR model are obtained by fitting the normalized multivariate time series $\left(Z_{H_s}(t_i),\ Z_{T_p}(t_i),\ Z_{\theta_m}(t_i)\right)_i$ to the expressions given in Appendix A.

2.3. Storm series simulations

The simulation process begins with the Montecarlo simulation of the storm duration (d_1) and the interarrival time (Δ_1) using the Copula function (§ 2.2.1). Then, we obtain a multivariate normalized time series of sea-states of duration d_1 using the VAR(q) model (§ 2.2.3). Finally, the time series describing the sea states evolution is obtained after the application of the corresponding inverse cumulative distribution functions for each variable (§ 2.2.2). This process is repeated until the end of the simulated period.

3. Application to case study

The methodology presented in this work was applied to a case study in the Mediterranean coast of Spain. This region lies in a transition zone between the Atlantic ocean and the Mediterranean sea. It is naturally protected from swell waves formed in the Atlantic ocean and its fetch is limited by the African continent. The study was carried out using hindcast wave data in coordinates 3.608° W - 36.66° N (Figure 3) located at the Alborán Sea, close to the coast of Granada, with hourly data from January 1st, 1979 until December 31st, 2018 (Mentaschi et al., 2015). The wave regimen is bimodal, with waves arriving from the WSW (35 %) and ESE (50 %). The main direction is attributable to the maximum fetch found at ESE ($\approx 112.5°$), with waves arriving obliquely to the shore. Waves are partially developed with periods ranging from 3 to 15 seconds and values of maximum significant wave height barely reaching 6.4 m.

3.1. Parameters selection for the identification of storms

For each given combination within a range of values of $H_{s,u}$, d_0 and δ_0, we obtain the series of the annual number of storms, N, and the interarrival times Δ. Then, following the methodology proposed in §2.1, the
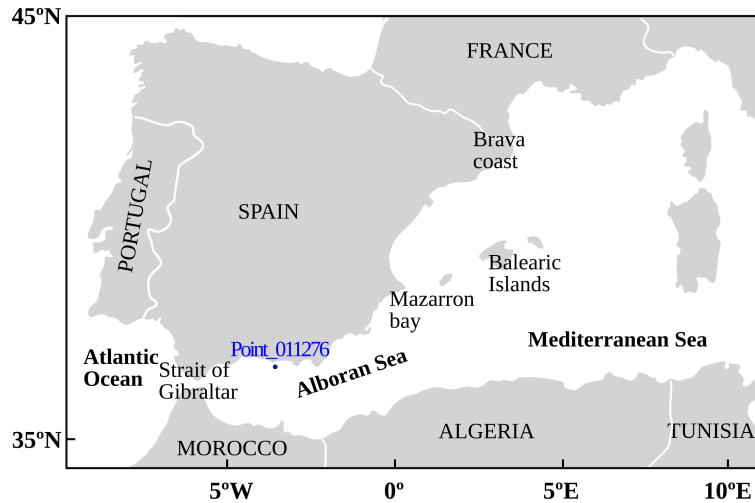
Figure 3: Location of the case study

p-values of the χ^2-test (p_N) and KS-test (p_Δ) are computed. Theoretically, the values of ($H_{s,u}$, d_0, δ_0) for the rejection of the null hypotheses define a 3D surface. For a given value of δ_0 it is possible to draw over a 2D domain ($H_{s,u}$, d_0), the nonrejection region for both tests defined as the intersection between the regions where $p_N > \alpha$ and $p_\Delta > \alpha$.

Figure 4 presents the isolines corresponding to the significance level of $\alpha = 5\%$ of the p-values for the selected range of ($H_{s,u}, d_0$) and different values of δ_0. They delimit the areas where the null hypotheses that the data follow the corresponding distributions ($p-$values> 0.05) cannot be rejected. The values of $H_{s,u}$ range from the $90th$ percentile to the $99.9th$, d_0 varies from 12 to 72 hours and the selected values of δ_0 are 12, 24, 36, 48 and 60 hours. The selected significance level of $\alpha = 5\%$ implies that there is a 5% probability of incorrectly rejecting the null hypothesis (Type I error rate).

It can be observed that, in general, the nonrejection region on the left side boundary is delimited by higher values of $H_{s,u}$ for increasing δ_0, whereas a clear pattern could not be found on the right side boundary of the region. As examples, for high values of $d_0 \approx 60$ h, for $\delta_0 = 24$ hours the valid $H_{s,u}$ ranges from $\approx 2 - 2.7$ m whereas for $\delta_0 = 60$ h, the nonrejection region comprises $H_{s,u} \approx 2.3 - 3$ m. In the case of lower values of $d_0 \approx 20$ h, the nonrejection region remains constant for $\delta_0 = 24 - 60$ h, at $H_{s,u} \approx 3 - 4$ m. Therefore, this indicates that in this case study, as the minimum storm duration decreases, $H_{s,u}$ increases and storms are identified with the highest waves possible. For a minimum storm duration of approximately 1 day, the storms are always identified to wave heights between the $99th$ and $99.8th$ percentile regardless of the minimum interarrival time. Meaning that these correspond to extreme events that happen always for interarrival times larger than 2.5 days. When the minimum storm duration is set to be at least of 2.5 days, the waves identified as extremes are significantly lower and more so when the time between storms is set to a minimum of 1 day implying that long events closely together correspond to lower waves.
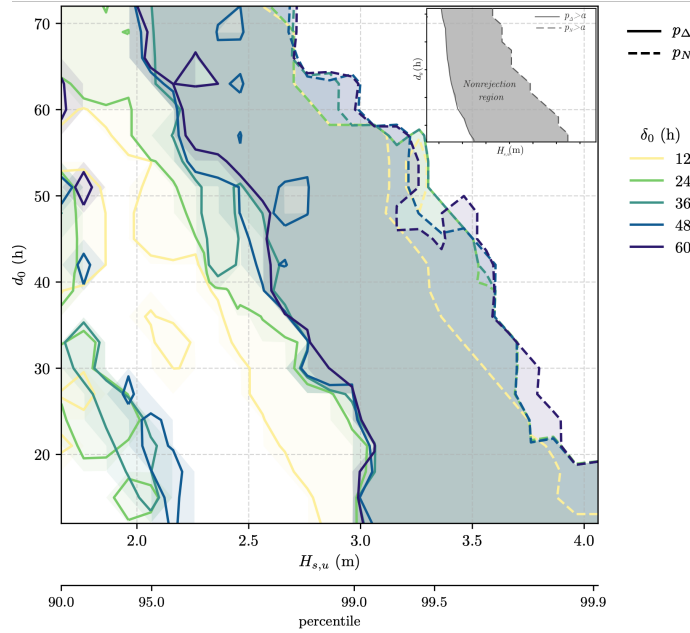
10

Figure 4: Nonrejection regions of significant wave height threshold, $H_{s,u}$, minimum storm duration, d_0, and different values of δ_0, for the case study. The nonrejection region is delimited by the isolines of the p-values of the KS-test for the distribution of the interarrival times Δ ($p_\Delta > \alpha$ - solid line) and the χ^2-test for the distribution of the number of events N ($p_N > \alpha$ - dashed lines). The different colors represent different cases of minim um interarrival time δ_0. A sketch of the nonrejection region is given in the upper right corner of the figure.

As previously mentioned the selected ($H_{s,u}$, d_0, δ_0) must fulfill the hypothesis tests, meet the condition $\bar{\lambda} \approx 1/\bar{\delta}$ and provide an adequate number of events that allows for the statistical multivariate characterization of the storm events. Figure 5 depicts the nonrejection region for different δ_0 as well as the contours of $\epsilon^* = |\bar{\lambda} - 1/\bar{\delta}|/\bar{\lambda}$ and the mean annual number of storm events, $\bar{\lambda}$. It can be observed that for $\delta_0 = 12$ h (upper left panel), most of the nonrejection region encompasses values of $\epsilon^* = 1 - 2\%$. For $\delta_0 = 24$ h (upper right) there are areas within the nonrejection region that correspond to $\epsilon^* < 0.5\%$ which delimits the optimal domain and values $\bar{\lambda} \approx 3$ where a bigger sample of storm events is obtained. For $\delta_0 > 36$ h, the nonrejection region generally comprises lower values of mean annual number of storms and higher significant wave height thresholds.

The final selection of the optimal values ($H_{s,u}$, d_0, δ_0) depends on the specific problem being analized and the atmospheric processes in the site study. In this case, given that storms are produced by the passage of low pressure systems lasting at least 24 h and with minimum interarrival times of a day on average, the selected storm parameters were $H_{s,u} = 2.4$ m (97.3th percentile), $d_0 = 40$ h and $\delta_0 = 24$ h. Using these values, we obtain a series of 91 storm events with an average duration of ≈ 2 days and 13 hours and mean annual number of storms $\bar{\lambda} \approx 2.3$.
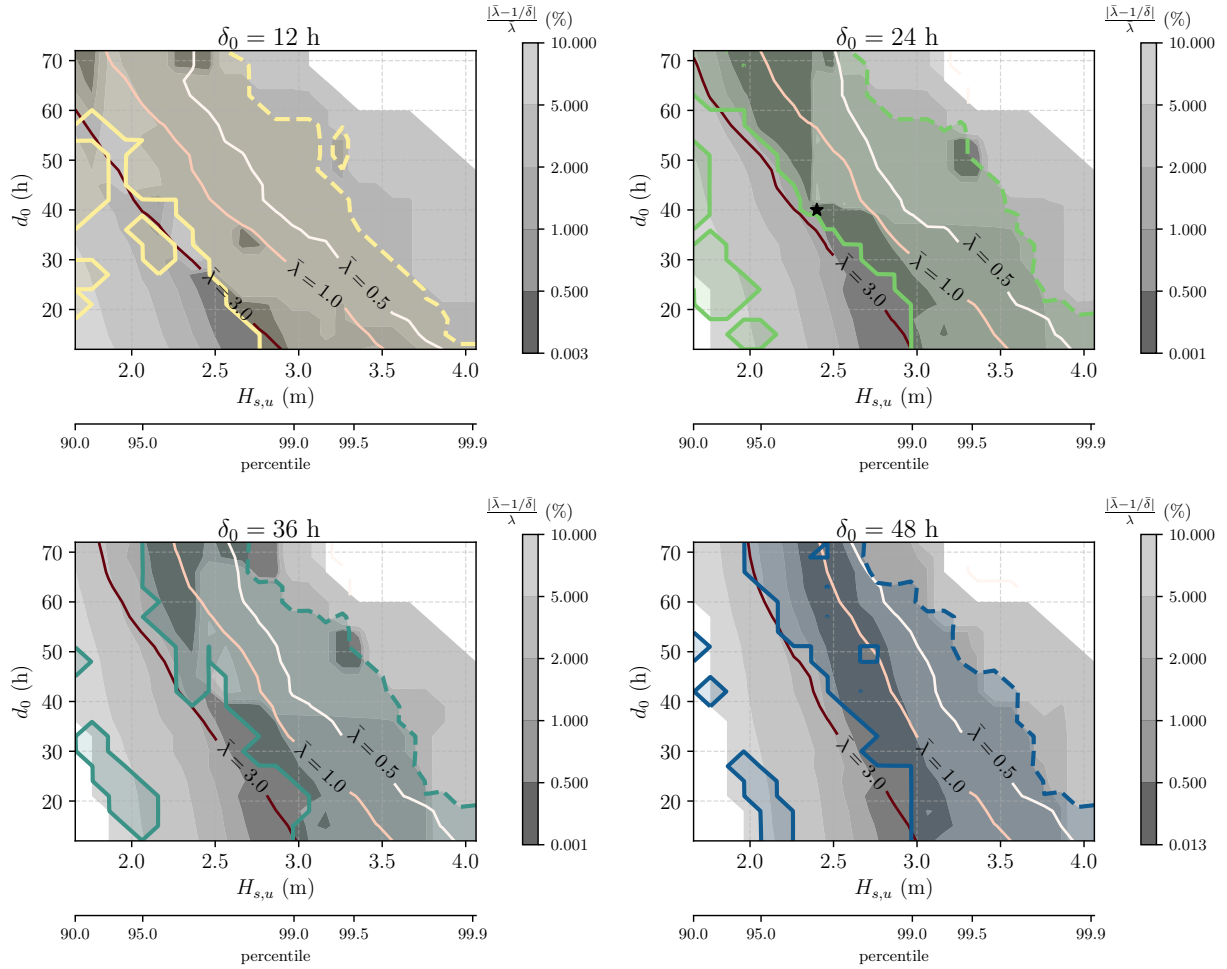
11

Figure 5: Nonrejection region of the hypothesis $p_\Delta > \alpha$ (solid colored line) and $p_N > \alpha$ (dashed colored line) for $\delta_0 = 12, 24, 36$ and 48 h. Grayscale contours of $|\bar\lambda - 1/\bar\delta|/\bar\lambda$ and isolines of the mean annual number of storm events, $\bar\lambda$ for the case study. The star symbol on the $\delta_0 = 24$ h panel (upper right) indicates the chosen value of $(H_{s,u}, d_0)$.

3.2. Interdependence of (D, Δ) via Archimedean copulas

The time series of storm events was divided into meteorological seasons to characterize its temporal distribution. Figure 6 shows the samples of the random vector (D, Δ) for Summer/Fall (June-November) and Winter/Spring (December-May) obtained from hindcast data. It can be observed that the storm events are more frequent and more extreme during Winter/Spring with higher d and $H_{s,max}$ values. These samples were used to fit copula models.

The empirical CDFs of D and Δ were calculated via Kernel Density Estimation (KDE). The use of non-parametric methods such as KDE models is adequate when there is not an established recommended distribution or when an acceptable fit cannot be obtained. This is the reason why we applied a KDE for the storm duration, D, and for simplicity, we also extended its use for the interarrival time, Δ. A bandwidth

12

needs to be chosen for the KDE. A compromise solution has to be adopted since a very narrow window ($b_w \sim 0.01$) perfectly represents the marginal functions but does not allow the extrapolation to values slightly outside the range. Conversely, a very large bandwidth does not adequately represent the marginal distributions. In this case, an Epanechnikov kernel with unitary bandwidth was selected. The use of a unitary bandwidth (b_w) is standard in several statistics programs. This choice ensures that there are no differences in the tails between the hindcast and simulated data. It also avoids the simulation of data too far from the range of the original data values. Figure 7 shows the copula function of storms duration and interarrival times from hindcast data and simulation. As it can be observed in Figure 7 the empirical joint distribution function of both time series have similar patterns. The differences in the diagonal are mainly due to the KDE bandwidth as previously discussed.
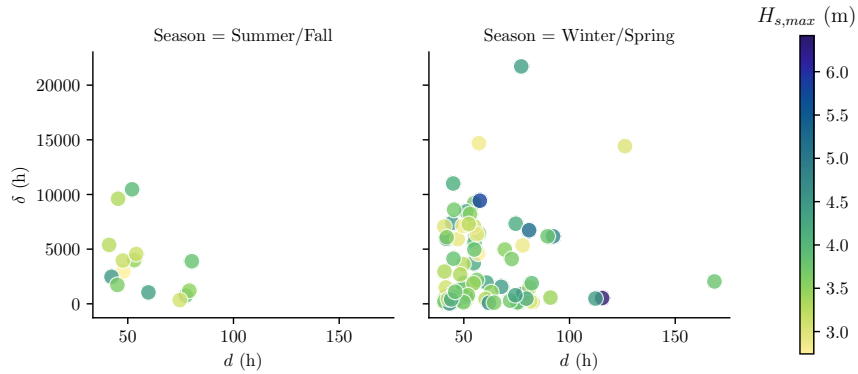

Figure 6: Scatter plots of storm duration, d, and interarrival time, δ, of storm events for the two selected climatic periods. The colors represent the $H_{s,max}$ of each event.

The value of θ_c was obtained following a Canonical Maximum Likelihood (CML) method. The Kendall's τ, the dependency parameter for the Clayton copula θ_c related to it ($\theta_c = 2\tau/(1-\tau)$), and the Spearman's ρ_s, are given in the Table 1. As it can be observed, $\theta_c \approx 0$ which suggests a weak association for winter/spring. During summer/fall, a significant negative association is found which indicates that as the rank on one of the varibles increases, the other one decreases. The absence of storms during some seasons can difficult the analysis. In those cases, it is better to extend the climatic periods from seasons to semi-annual or yearly periods.

3.3. Univariate non-stationary probabilistic models

The empirical distribution of the significant wave height F_{H_s} is fitted to a non-stationary mixture distribution composed by a Lognormal distribution in the central regime and a Generalized Pareto Distribution for each tail (Eqs. 5-6). Under storm conditions, $F_{H_s^s}$ is calculated as described by Equation 8. $F_{T_p}^s$ is fitted to a non-stationary Lognormal distribution (Eq. 9) and $F_{\theta_m}^s$ is given by a stationary mixture distribution of 2-Truncated-Normal distributions (Eqs. 10-11) with the aim at reproducing the clear bimodal wave

13

Season	θ_c	τ	ρ_s
Summer/Fall	-0.4310	-0.2747	-0.4549
Winter/Spring	0.0172	0.0085	0.0045

Table 1: Parameters of the copula fit θ_c, Kendall's τ, Pearson ρ_p, and Spearman ρ_s, for the selected climatic periods.
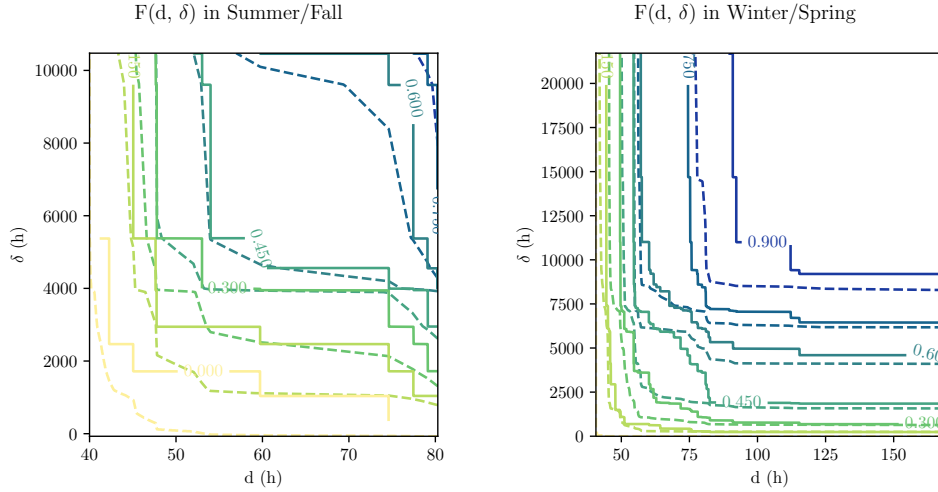

Figure 7: Bivariate Empirical Cumulative Distribution Function (ECDF) (solid lines) and copula fit (dashed lines) of D and Δ for the selected climatic periods.

climate observed in this case. The parameters of the expansion of non-stationary distributions as Fourier series truncated to the fourth order, are estimated by the negative log-likelihood function and the Bayesian Information Criteria. The values of the obtained coefficients are presented in Appendix B.

Figure 8.a presents a range of percentiles of the non-stationary empirical and fitted CDF of the significant wave height. It can be observed that the selected theoretical distribution adequately reproduces the behavior of the hindcast data. Figures 8.b and 8.c present the empirical and fitted CDFs under storm conditions $F^s_{T_p}$ and $F^s_{\theta_m}$, respectively. The mean direction was found to be adequately reproduced with a stationary mixture distribution. Figure 9 presents the Q-Q plots of the analyzed wave variables. As it is observed in panels (a) and (b), the time-dependent data is classified by months where the non-stationary probability models are assumed to be stationary. The time period of this classification depends on the Fourier order used in the analysis. In this case, the fourth order was selected meaning that the probability model can capture from semiannual variabilities up until three weeks. As it is observed, good fits between the empirical distribution functions and the selected mixed probability models are obtained. For θ^s_m a larger variability is observed in the tails than for H_{m0} and T^s_p.
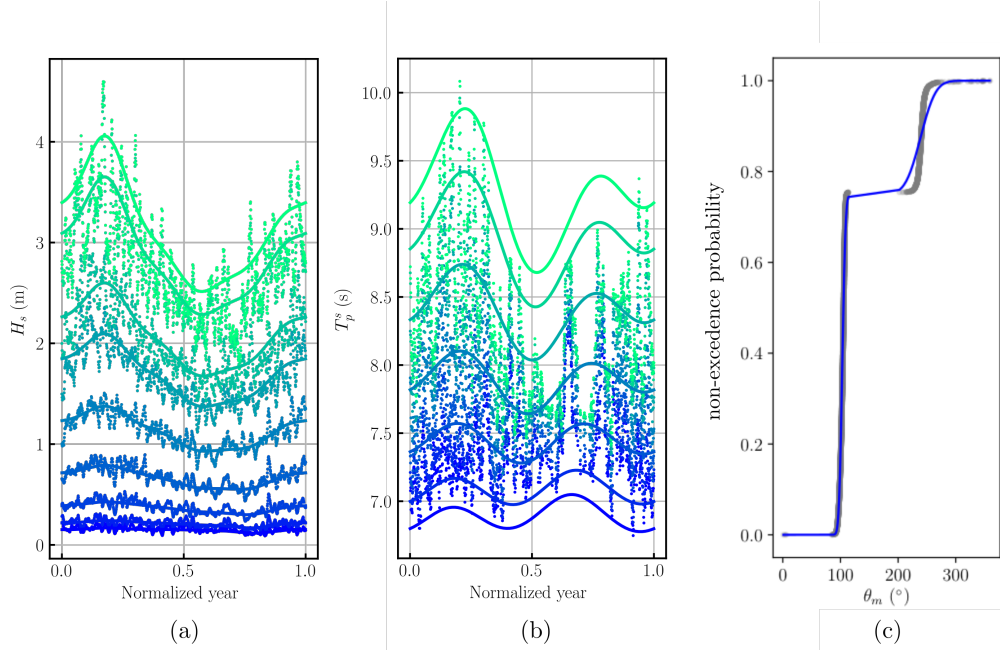
Figure 8: Cumulative Distribution Functions of H_s, T_p^s and θ_m^s. *(a)* Iso-probability percentiles (5, 10, 25, 50, 75, 90, 95, 99, 99.5th) of the empirical and fitted non-stationary distributions for H_s. *(b)* Iso-probability percentiles (5, 10, 25, 50, 75, 90, 95th) of the ECDF and theoretical fit $F_{T_p}^s$. *(c)* ECDF and theoretical fit $F_{\theta_m}^s$.
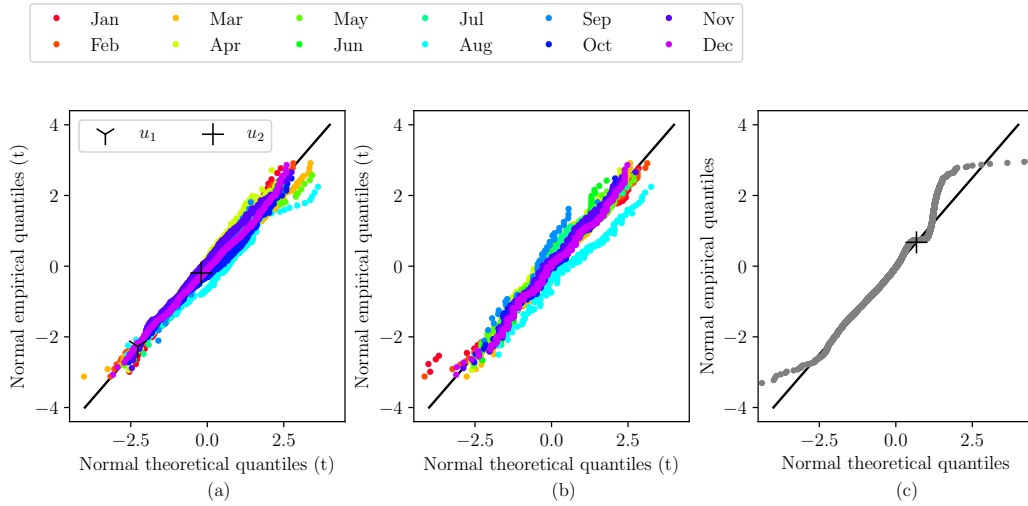

Figure 9: Q-Q plots of H_s, T_p^s and θ_m^s. The Q-Q plots for the non-stationary distributions (H_s and T_p^s) were computed monthly assuming stationarity. *(a)* H_s. The thresholds u_1 and u_2 are established at the limits of the lower tail, body and upper tail of the mixed distribution. *(b)* T_p^s and *(c)* θ_m^s. The plus symbol indicates the limit between the corresponding truncated normal distributions.

3.4. Multivariate temporal dependence

The time series $H_s^s(t)$, $T_p^s(t)$ and $\theta_m^s(t)$ are transformed with Equation 12. The stationary and normalized series $Z_{H_s}(t)$, $Z_{T_p}(t)$ and $Z_{\theta_m}(t)$ allow the calculation of the parameters of the VAR model used to characterize the multivariate temporal dependence of the series.

4. Verification of storm shape, magnitude, potential damage and occurrence

In the previous section we have presented statistical validations to assess whether the different fits proposed in the methodology were adequate, i.e. storm definition, bivariate ECDFs and non-stationary probabilistic models. This section tackles the issue of verifying the suitability of this methodology to reproduce the storm time series taking into account the storm shape, magnitude, potential damage and the probability of occurrence of events.

To further investigate whether the model is capable to reproduce the intravariability of the multivariate storm events, a data reduction analysis based on clustering, as in De Leo et al. (2019), can be performed to the series of hindcast and simulated normalized storms. In this work, instead of using a clustering and classification space to distinguish patterns, we followed the approach of Baquerizo and Losada (2008) which uses Empirical Orthogonal functions (EOF) as a standard procedure for data reduction in the euclidean space (see Appendix C). Each individual storm was nondimensionalized with respect to its corresponding d, $H_{s,max}$, $T_{p,max}$ and $\theta_{m,max}$. Figure 10 presents the mean function and the first five eigenfunctions for H_s^*, T_p^* and θ_m^*. It can be observed that the 5th eigenfunction explains more than 80% of the variability for the hindcast data (α_d) and more than 70% of the variability for the simulation (α_s). Even though the explained variabilities (α_d and α_s) are not exactly the same, they are relatively close so that the eigenfunctions are represented together and it is possible to see their similitudes. Also presented in Figure 10 are the probability distribution functions (CDF) of the coefficients or scores for which similitudes can be observed. It is observed that the shapes of the mean functions are quite similar between the hindcast and simulated data. Moreover, the eigenfunctions also reproduce the behavior of the deviation from them. As an example, the 1st eigenfunction captures approximately the 30% of the deviations from the mean H_s, for the hindcast data and simulation. In the case of the mean direction of the hindcast data, it can be observed that the first three eigenfunctions already account for the 97% of the deviation.

The fact that the analysis performed with hindcast and simulation data, (i) shows similar eigenfunctions that explain about relatively similar percentages of variability and (ii) the CDFs of the coefficients are alike, is an indication that the model is capable to reproduce storms whose shapes diverge from the triangular or trapezoidal shape.

The model does not only reproduce the intravariability of each of the time series under consideration but also their joint behavior. In Figure 11, the joint distribution of (H_s, T_p) and (H_s, θ_m) obtained from
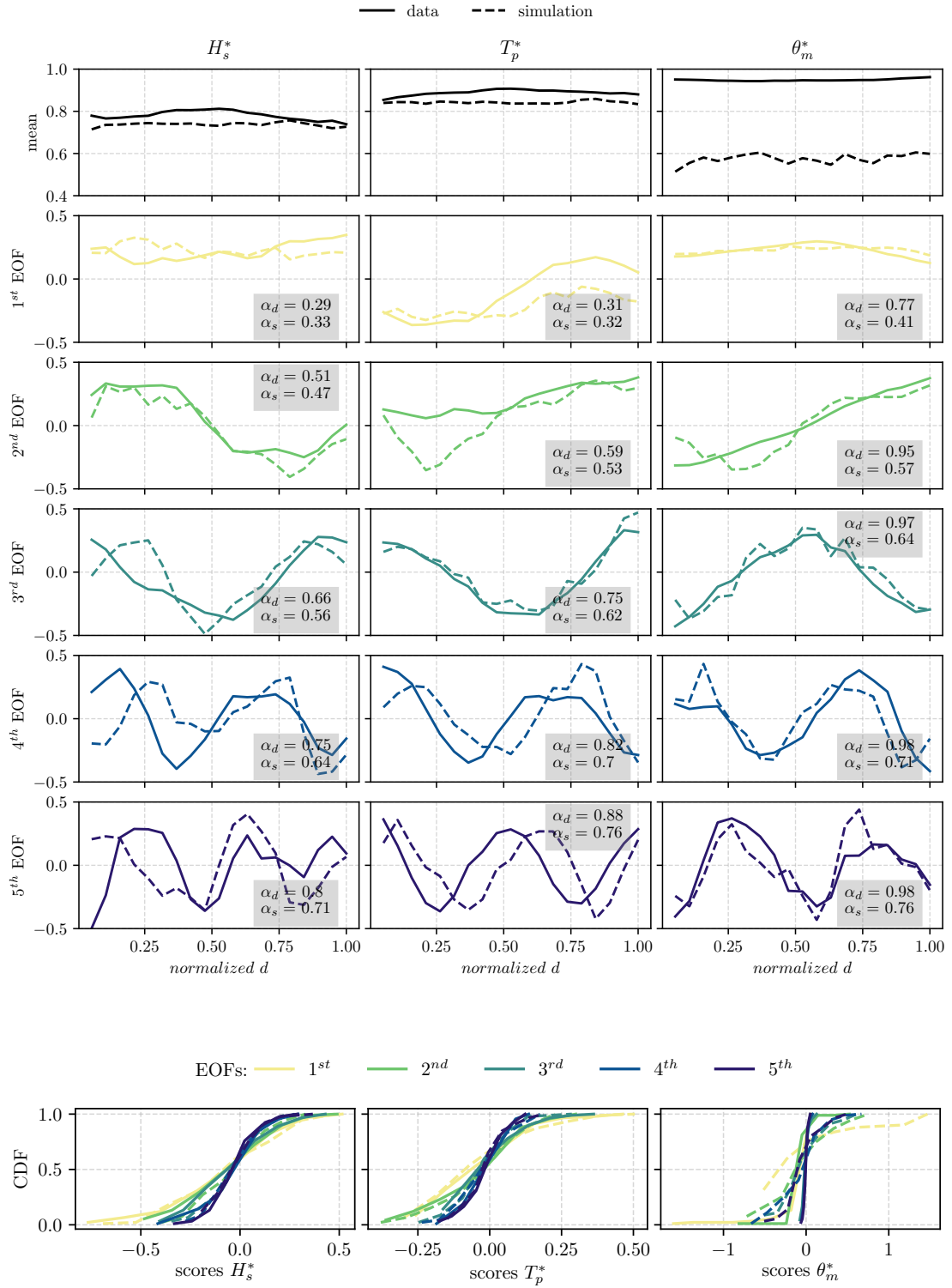
16

Figure 10: Empirical Ortogonal Functions for the normalized storm events (top) and CDFs of the corresponding coefficients or scores.

available data and simulation is represented. The similarities between them show that the model captures the joint dependence of the variables that define the storms. In particular, the significant wave height, H_s,
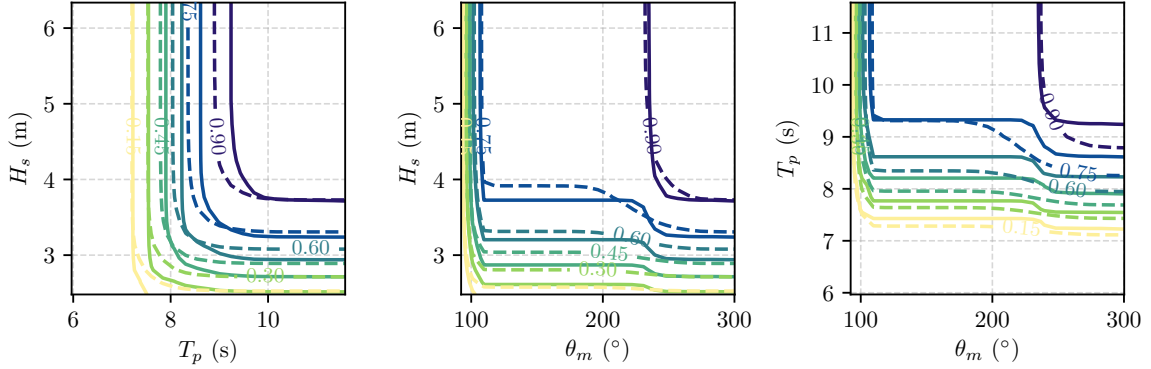

Figure 11: Bivariate ECDF of hindcast data (solid lines) and simulation (dashed lines) for H_s (m), T_P (s) and θ (°) during storm conditions.

and the storm duration D, deserve a special reference, as they constitute key variables in the statistical behavior of the storms' energetic content M, (Eq. 13).

$$M = \int_0^d (H_s^2(t) - H_{s,u}^2(t))dt, \tag{13}$$

where d is the duration of the storm. In this work, the formula of M used by other authors as storm magnitude (De Michele et al., 2007) has been changed to take into account the energy content (proportional to the square of the significant wave height) throughout the storm duration. Figure 12 shows that the energy content of the simulated data is equivalent to the registered storms.

The linear wave power is an instantaneous measure of the storm and, therefore, quantifies the interdependence of H_s with T_p (Eq. 14) in a single value (Corbella and Stretch, 2013). We also extend this concept including the storm duration and the incident wave direction as:

$$P = \int_0^d E(t) \cdot c_g(t) \, dt = \int_0^d \left(\frac{1}{16} \rho g H_s(t)^2 \right) \frac{g T_p(t)}{4\pi} \cos(\theta_n) \, dt, \tag{14}$$

where g is the gravitational acceleration, ρ is the water density and θ_n is the angle with respect to the normal direction to the shore. In this case we have considered a W-E shore orientation and therefore, only the waves within $\theta_m \in (\pi/2, 3\pi/2)$ were considered. Figure 12 shows that the simulated storms yield a similar behavior to the historical data.

5. Example of a damage evolution model

The proposed simulation model has been used to study the damage evolution in a rubble-mound breakwater triggered by an initial damage, D_0^*, measured at a certain stage. This process is stochastic in nature,
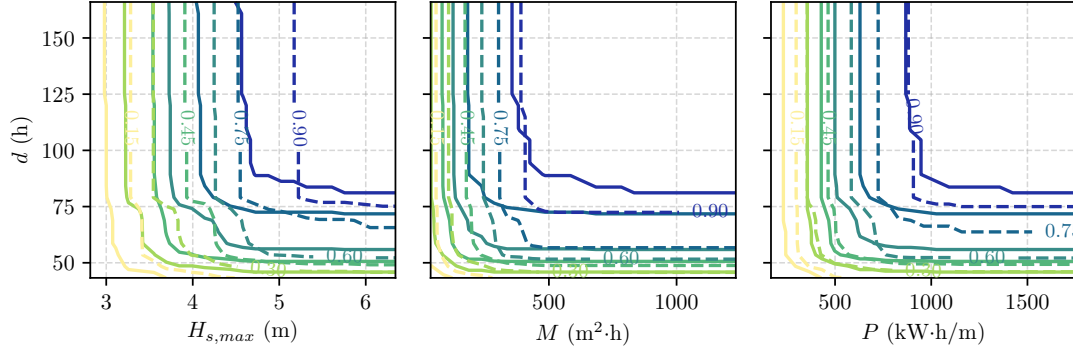
Figure 12: Bivariate ECDF of hindcast data (solid lines) and simulation (dashed lines) for d, $H_{s,max}$, M and P.

as both the loadings and the armor conditions are random themselves. Castillo et al. (2012) described the essential elements in a damage progression model. They also proposed a stochastic approach for damage accumulation in ruble-mound breakwaters under certain probabilistic assumptions.

In this work, for the sake of simplicity, we adopt a law describing the mean damage progression obtained from the best fit to empirical data. It is also hypothesized that every time the breakwater suffers any damage, it attains an equilibrium position. This means that only the same and more energetic sea states than those that produced the damage are capable of increasing it.

5.1. Damage produced by a sequence of sea states

The damage, D^*, is measured in terms of the dimensionless eroded cross-section area, $D^* = A_e/D_{50}^2$, where A_e is the eroded area and D_{50} is the nominal stone diameter. We used a cumulative damage curve for irregular waves proposed by Melby and Kobayashi (1999) to describe the damaged produced on the structure during a sea state with normal incidence (of duration dt and characterized by the values of its significant wave height, H_s, and peak period, T_p) as a function of the previous damage, $D^*(t)$, the stability number, N_s, and the number of waves in the sea state, t_w:

$$D^*(t + dt) = \left[D^*(t)^{1/b} + \left(a \cdot N_s^5 \right)^{1/b} \cdot t_w \right]^b \quad \text{if} \quad N_s \geq N_{s0}, \qquad (15)$$

where a and b are empirical coefficients, the number of waves in the sea state is estimated as $t_w = dt/T_p$ and the stability number is given by:

$$N_s = \frac{H_s}{\Gamma \cdot D_{50}}, \qquad (16)$$

with Γ being the relative excess of specific weight, $\Gamma = (\gamma_a - \gamma_w)/\gamma_w$, where γ_a and γ_w are, respectively, the specific weights of the armor and the water. N_{s0} is the value of the stability number that induced damage during previous sea states.

19

This model, built as indicated by Castillo et al. (2012), involves a set of independent dimensionless variables according to the Pi-Buckingham theorem and guarantees that the expression fulfills a compatibility condition which ensures that the damage produced during two loading cycles of given durations is equal to the sum of the damages produced during their individual durations.

5.2. Damage produced by a sequence of storm events

In order to describe the damage progression during a series of storms, based on Losada (2001) and Baquerizo and Losada (2008), we consider every storm as a sequence of sea states. Then, the accumulated damage, $DS_n^*(t)$, at the n-th storm given by n_s sea states, $(H_s(t_i), T_p(t_i))$ for $i = 1, \ldots, n_s$ is:

$$DS_n^* = \left[\left(DS_{n-1}^* \right)^{1/b} + \sum_{i=1}^{n_s} \left(\left(a \cdot N_s(t_i)^5 \right)^{1/b} \cdot t_w(t_i) \right) \right]^b \tag{17}$$

where DS_{n-1}^* is the damage caused by the previous storm.

5.3. Breakwater damage progression

We consider an ideal breakwater located at a water depth equal to 14 m in the coast of Motril (Granada), with $D_{50} = 1.16$ m, slope $1:2$ and $\Gamma = 1.66$. These are the same characteristics as the structure studied by Sousa and Santos (2006) who used the empirical formula given by Melby and Kobayashi (1999) with $a = 0.011$ and $b = 0.5$.

Following the methodology presented in this work, we obtained a large number of simulations (500) of a 200-years time series of storm events and propagated them to the coast by means of linear theory. For each simulation, we estimated the time-evolving damage of the breakwater as described by Equation 17.

Based on the empirical results of Melby and Kobayashi (1999), we adopted $D_A^* = 6$ as the admissible level of failure and $D_D^* = 12$ as the destruction level. Figure 13 shows, for three different simulations, the accumulated damage evolution. It can be observed that for the three cases, the admissible damage is achieved during the first seven storms. The time of destruction varies in these examples from the beginning of the 7th year, with the destruction occurring during the 14th storm, to the end of the 22th year on the 60th storm and the beginning of the 38th year, after the breakwater was loaded by 86 storms. Figure 14 shows the probability distribution functions of the times of occurrence of admissible damage, T_A, and destruction, T_D. It can be observed that the probability that the time of admissible damage occurs in the first seven years is approximately 0.6. Also represented in Figure 14, is the exceedance probability of time to destruction if a repair or maintenace strategy is implemented every 7 years, where the exceedance probability of time to destruction, T_D, is periodically reset to zero. Therefore, the analysis of damage using a probabilistic approach and taking into account the temporal evolution of the storms provides rich information for the different stages of the breakwater ranging from no damage to serviceability and ultimate limit states. It
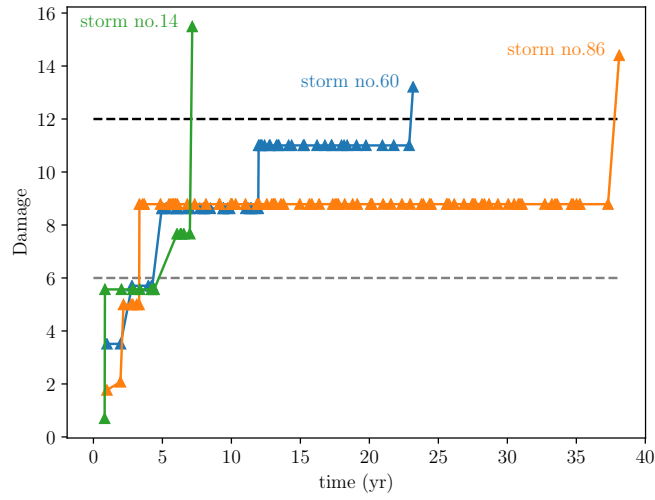
20

Figure 13: Accumulated damage evolution for three 200-years time series simulation of storm events.
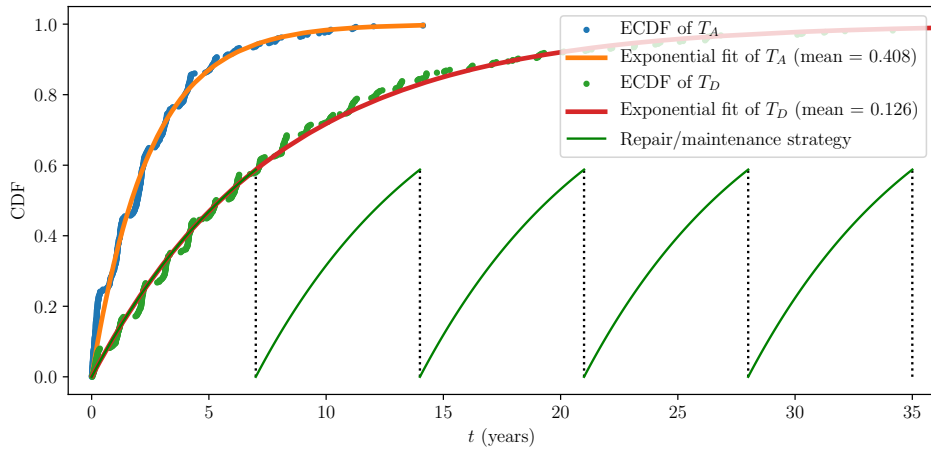

Figure 14: Empirical Cumulative Distribution Function (ECDF) and fit of T_A and T_D and exceedance probability of time to destruction under a given repair/maintenance strategy.

allows to design a breakwater considering the possibility of implementing management decisions and repair strategies when the breakwater attains a certain prefixed level of damage.

6. Discussion

The proposed methodology is based on the assumption that storms are rare events and, therefore, the number of occurrences follows a Poisson distribution. A constant parameter λ is adopted, however, its value may vary along the year. The consideration of seasons to characterize the random vector (D, Δ) that is used to redistribute storms along the year, allows to take into consideration that, in fact, λ varies in time with higher values during the more severe seasons and is as a nonhomogeneous Poisson process. The ability of the model to reproduce this behavior can be seen in Figure 15 where monthly number of storms are shown for hindcast and simulated data. In the case under analysis two climate seasons were chosen to characterize
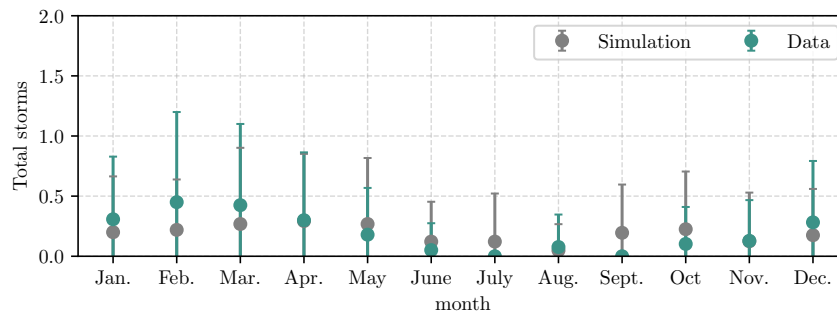

Figure 15: Monthly average number of storms for hindcast and simulated data including confidence intervals (standard deviation).

the time variability of the random vector (D,Δ). This choice responds to the need of having enough data in each considered season and also to the fact that climate variability in Mediterranean semiarid zones does not really follow the traditional four seasons climate. The selection of the climate periods under analysis affects the distribution of events throughout the year. As it can be observed in Figure 15, during the months of June until November (Summer/Fall panel of Figures 6 and 7) the simulation generally presents a slightly higher number of events and deviation than the hindcast data. This is due to the fact that the copula model is being characterized with 14 storm events and a higher sample would be needed to get a better copula fit.

In addition, during the other analyzed season (from December until May), the simulation generally presents slightly lower mean number of events and lower deviations than the hindcast data. Indeed, the hindcast data shows a large deviation during this period suggesting a strong interannual variability during the Winter/Spring period which, given the length of the hindcast dataset and the methodology proposed in this work, the model is not able to capture.

This study proposes the long-term simulation of extreme events and their temporal evolution. The statistical characterization of the different steps is done with the extracted storms from hindcast data and therefore the sample size is reduced considerably from the initial time series. This hinders the capacity of the model to capture temporal variability beyond the seasonal variation. Longer time series would be needed to account for interannual, decadal or multidecadal temporal variability.

At this selected site, two mean incoming sea state directions alternate (Félix et al., 2012) with values close to 100° and 250° and no clear differences in trends can be observed along the year. At other locations, non-stationary univariate models may be more appropriate (see Mendonça et al. (2012)).

Regarding the storm definition, the methodology presented in this work helps the user in the selection of the values ($H_{s,u}$, d_0, δ_0) that define the storm events. Its application, however, needs the knowledge of the climatological processes on the study site and the expert vision on the particular problem that is being studied. For the example, in order to analyze breakwater damage evolution, a relatively high value of $H_{s,u}$, close to values capable to initiate the damage is required, while for coastal applications such as beach retreat assessment, smaller values are chosen. This also applies to the chosen values of d_0 and δ_0 where management strategies need to be taken into account.

The threshold $H_{s,u} = 2.4$ m selected in the results, corresponds to the $97.3th$ percentile of all the data and it is most of the year above the $95th$ percentile. This value, obtained with a methodology that allows to narrow down the possible valid values, is within the range of the different percentile values found in literature ($90 - 99.5th$). If the selection would have been done in the traditional way e.g. as the $95th$ percentile of the mean regime, a threshold close to 2 m, that is out of the nonrejection region for the majority the range of tested values of d_0 and δ_0 (see Figure 4), would have been obtained. This example brings to light that commonly used criteria to fix the threshold to define storm events (i.e. a given percentile value) does not guarantee that the statistical assumptions are fulfilled.

The method has also been applied to another location in the Andalusian Atlantic Ocean (AAO; 6.50°W - 36.50°N) where sea climate is rather severe, with values of the significant wave height above 2 m ($90th$ percentile). Under these circumstances, for $\delta_0 = 24$ h, the nonrejection region where $\epsilon^* = |\bar{\lambda} - 1/\bar{\delta}|/\bar{\lambda} > 0.5$ is very limited. In this case only for relatively high thresholds (e.g. $H_{s,u} \approx 2.9$ m - $97.7th$ percentile and $d_0 = 51$ h, $H_{s,u} \approx 3.5$ m - $98.5th$ percentile and $d_0 = 33$ h) the hypothesis cannot be rejected, which leads to low values of λ (close to one) and, therefore, to small size samples that reduce the confidence of the estimation of the model parameters (Figure 16).

7. Conclusions

This work proposes a site specific methodology to define storm events and to characterize and simulate multivariate wave series under storm conditions from historical or hindcast sea state data.
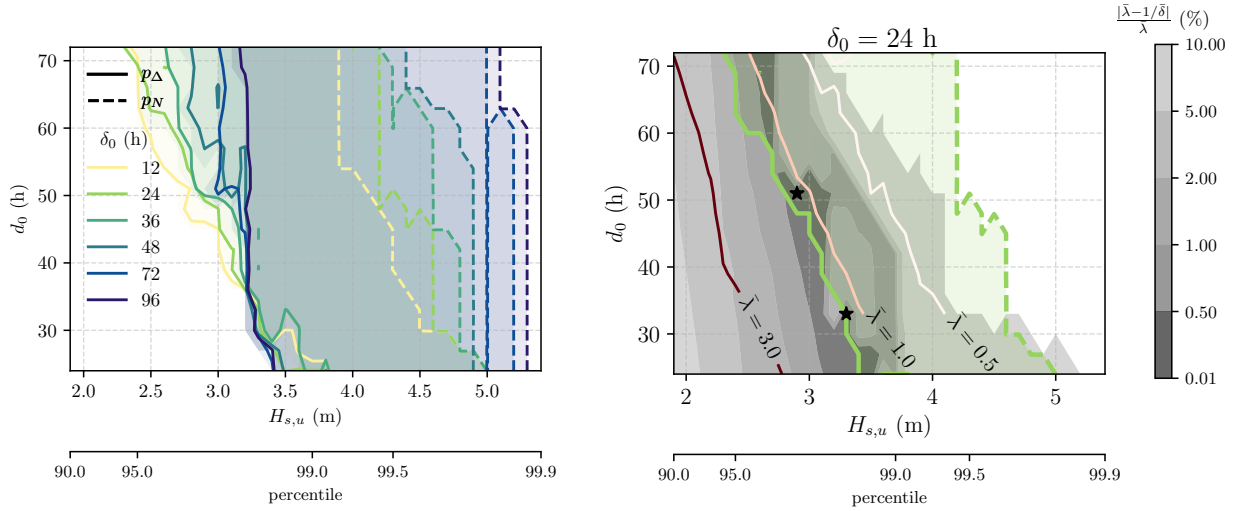
Figure 16: The left-hand panel presents the nonrejection regions of the hypothesis $p_\Delta > \alpha$ (solid colored line) and $p_N > \alpha$ (dashed colored line) for a range of significant wave height thresholds, $H_{s,u}$, and minimum storm durations, d_0, for the case study AA0. The different colors represent different cases of minimum interarrival time δ_0. The right-side panel depicts the nonrejection region of the hypothesis $p_\Delta > \alpha$ (solid green line) and $p_N > \alpha$ (dashed green line) for $\delta_0 = 24$ h. Grayscale contours of $|\bar{\lambda} - 1/\bar{\delta}|/\lambda$ and isolines of the mean annual number of storm events, $\bar{\lambda}$ for the case study AA0. The star symbols indicate possible valid values ($H_{s,u}$, d_0).

The procedure to select storm events is based on the joint definition of the parameters ($H_{s,u}, d_0, \delta_0$) and ensures that the hypotheses underlying the analysis of rare events, namely that the number of occurrence of events is a Poisson process and that the elapsing time between events follows an exponential distribution, are fulfilled. This methodology entails an advance in threshold selection by introducing the dependencies with the minimum storm duration and interarrival times and allows to narrow down the optimal values that fulfill the underlying assumptions.

The stochastic characterization of wave climate during storm conditions accounts for: (i) the frequency, persistence and elapsing time between consecutive storms, which is reproduced fairly well using a copula model and where the storm events are indirectly treated as a nonhomogeneous Poisson process; (ii) the intensity of the storms, whose probability at the tails and body is adequately reproduced by non-stationary mixture probability models; and, (iii) the shape of the storms, where the multivariate dependencies and their intravariability are achieved with a vector autoregressive model. Given the adequate characterization and simulation of wave climate, the model is able to reproduce other dependent non-linear variables such as the energetic content and the wave power and therefore, it can be applied to perform a probabilistic analysis of damage progression in maritime structures.

The methodology was applied as an example to assess the damage progression of a breakwater located in the coast of Granada, Spain. The results show that the proposed methodology allows for the assessment

of the damage evolution in the structure and provides a tool to efficiently check different management
and repair strategies which can significantly improve the decision-making process of coastal engineers and
455 managers.

Appendix A. Vector Autoregressive model

We denote the values of the variables at time t_i as $y_i^1 = Z_{H_s}(t_i)$, $y_i^2 = Z_{T_p}(t_i)$, $y_i^3 = Z_{\theta_m}(t_i)$ and
$Y_i = \left(y_i^1 \, y_i^2 \, y_i^3\right)^T$ where T stands for the vector transposition. The dependence in time and between wave
data variables in the VAR(q) model is given by:

$$Y_i = c + A^1 Y_{i-1} + A^2 Y_{i-2} + ... + A^q Y_{i-q} + e_i, \tag{A.1}$$

460 where $c = (c_1 \, c_2 \, c_3)^T$ contains the mean values of the variables, A^m, $m = 1, ..., q$ are the 3×3 coefficients
matrices and $e_i = \left(e_i^1 \, e_i^2 \, e_i^3\right)^T$ is the vector with the white noise error terms. Using equation A.1 to relate
data at an instant t_i to their previous q values, for $i = q + 1, ..., N$, where N is the length of the record, we
obtain $Y = AX + E$, where $Y = (Y_{q+1} \, Y_{q+2}... Y_N)$, $X = (X_{q+1} \, X_{q+2}... X_N)$, with $X_i = (1 \, Y_{i-1}^T ... Y_{i-q}^T)^T$,
$A = (A^1 \, A^2 ... A^q)$ and $E = (e_{q+1} \, e_{q+2}... e_N)$.

465 The solution is obtained by means of minimum least square errors as $A = YX^T(XX^T)^{-1}$, where $E = Y - AX$. A detailed description can be found e.g. in (Lütkepohl, 2005).

Appendix B. Parameters of the fitted distribution functions

Tables B.2 and B.3 show the information about the estimated parameters of the marginal distributions
of H_s and T_p^s, respectively. The Fourier order of every variable was selected according to the minimum BIC.
470 In the case of H_s (Table B.2), the normalized and non-stationary thresholds for the separation of the
tails and the central regime are $z_1 = -2.2699$ and $z_2 = -0.1962$. The dimensional thresholds are retrieved
applying $u_i = F_{LN}^{-1}(\Phi(z_i))$ for i the lower and upper threshold.   The mean wave direction over the
threshold, θ_m^s, was fitted using two stationary truncated normal distributions where μ, σ and α are mean,
standard deviation and weight of every distribution function ($\mu_{TN_1} = 1.8004$, $\sigma_{TN_1} = 0.0763$, $\alpha_1 = 0.75$,
475 $\mu_{TN_2} = 4.1702$, $\sigma_{TN_2} = 0.3615$, and $\alpha_2 = 0.25$).

Appendix C. Empirical orthogonal functions as best approach functions

The Emprical Ortogonal Functions is a methodology based on linear algebra. Its application to the
analysis of a series of N observations of a discrete function y, evaluated at M values $(t_1, \cdots, t_M)$, $y_j = (y_j(t_1), y_j(t_2), \cdots, y_j(t_M))$ $(j = 1,...,N)$ whose mean values are $c = (c_1, \cdots, c_M)$ can be interpreted as

	ξ_{LN}		μ_{LN}		σ_{LN}		ξ_{GP_M}	
N_F	a_l	b_l	a_l	b_l	a_l	b_l	a_l	b_l
0	0.0830	—	-0.0296	—	0.6909	—	-0.0123	—
1	0.0495	-0.0209	9.12e-5	-0.0170	0.0625	0.1190	0.0071	0.0014
2	-0.0214	0.0068	-4.42e-4	0.0106	4.27e-4	-0.0101	-0.0125	0.0125
3	-0.0097	0.0141	-0.0038	0.0021	0.0013	-0.0098	-0.0672	-0.0033
4	0.0048	-0.0205	0.0012	-0.0107	-0.0064	-2.09e-4	—	—

Table B.2: Parameters of marginal fit for H_s. The mixture model comprises a the Generalized Pareto distribution (GP) ($N_F = 3$) for minima and maxima and a Lognormal distribution (LN) ($N_F = 4$) for the central regime. The estimated parameters of the distributions are: (i) shape, ξ_{LN}; (ii) location, μ_{LN}; and, (iii) and scale, σ_{LN} of the LN distribution, while ξ_{GP_M} is the shape parameter of the maxima. The coefficients a_l and b_l represent the Fourier expansion as shown in Eq. 3.

	ξ_{LN}		μ_{LN}		σ_{LN}	
N_F	a_l	b_l	a_l	b_l	a_l	b_l
0	0.1991	—	4.2685	—	3.6195	—
1	-0.0015	0.0485	-0.4138	0.5317	0.5000	-0.5000
2	-0.0161	0.0012	0.0143	0.2500	-0.1705	-0.0196

Table B.3: Parameters of marginal fit for T_p^s. The model was fitted with a LN distribution.

the search of best approach M discrete functions (vectors in $\mathbb{R}^M$) $\{E_k\}$ that allow to express any of the observations y_j as a linear combination of them:

$$y_j = c + z_j^1 E_1 + \cdots + z_j^M E_M, \tag{C.1}$$

where z_j^k is the coefficient (also called score) of vector E_k to reproduce the jth observation. $()^T$ stands for the transpose vector.

The discrete functions $\{E_k\}$ are obtained as the eigenvectors of the matrix AA^T where $A = Y/\sqrt{N}$ and $Y = (y_j(t_i) - c_i)$ is the $M \times N$ matrix that contains the demeaned values of the observations. For simplicity of notation we are assuming that the eigenfunctions E_k of AA^T are sorted according to their respective eigenvalues $\{\lambda_j\}$ with $\lambda_1 > \lambda_2 > ... > \lambda_M$. We call $y_j^m = c + z_j^1 E_1 + \cdots + z_j^m E_m$ to the sum (Equation C.1) truncated to the first m values. The total mean squared error made by using this approach is:

$$\varepsilon_m = \lambda_{m+1} + \cdots + \lambda_M. \tag{C.2}$$

Calling $V_m = (\lambda_{m+1} + \cdots + \lambda_M)/(\lambda_1 + \cdots + \lambda_M)$, the value $1 - V_m$ measures the deviation of data to the approximation made, that is, what cannnot be represented with the first m eigenfunctions and, therefore, $\alpha_m = 100 \times Vm$ express the percentage of variability explained with them.

The values $\{z_j^k\}_k$ can be considered as a sample of the random variable Z_j, that represent the value of the coefficient of E_j in the approximation. By randomly choosing their values it is possible to simulate the dimensionless shape of a storm, in a similar way to that of Baquerizo and Losada (2008).

Baquerizo, A., Losada, M.A., 2008. Human interaction with large scale coastal morphological evolution. an assessment of the uncertainty. Coastal Engineering 55, 569 – 580.

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

Boccotti, P., 2000. Wave mechanics for ocean engineering. volume 64. Elsevier.

Castillo, C., Castillo, E., Fernández-Canteli, A., Molina, R., Gómez, R., 2012. Stochastic model for damage accumulation in rubble-mound breakwaters based on compatibility conditions and the central limit theorem. Journal of Waterway, Port, Coastal, and Ocean Engineering 138, 451–463.

CIRIA, CUR, CETMEF, 2007. The Rock Manual. The use of rock in hydraulic engineering.

Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An introduction to statistical modeling of extreme values. volume 208. Springer.

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

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

De Leo, F., Solari, S., Besio, G., 2019. Extreme waves analysis based on atmospheric patterns classification: an application along the Italian coast. Natural Hazards and Earth System Sciences Discussions 2019, 1–20.

De Michele, C., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coastal Engineering 54, 734–751.

Fedele, F., Arena, F., 2009. The equivalent power storm model for long-term predictions of extreme wave events, in: ASME 2009 28th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers. pp. 401–411.

Fisher, N.I., 1995. Statistical analysis of circular data. Cambridge University Press.

Félix, A., Baquerizo, A., Santiago, J., Losada, M., 2012. Coastal zone management with stochastic multi-criteria analysis. Journal of Environmental Management 112, 252 – 266.

Goda, Y., 2010. Random seas and design of maritime structures. volume 33. World Scientific Publishing Company.

Jäger, W., Nagler, T., Czado, C., McCall, R., 2018. A statistical simulation method for joint time series of non-stationary hourly wave parameters. Coastal Engineering .

Jäger, W.S., Nápoles, O.M., 2017. A vine-copula model for time series of significant wave heights and mean zero-crossing periods in the north sea. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3, 04017014.

Li, F., van Gelder, P.H.A.J.M., Ranasinghe, R., Callaghan, D.P., Jongejan, R.B., 2014. Probabilistic modelling of extreme storms along the Dutch coast. Coastal Engineering 86, 1–13.

Li, F., Zhou, J., Liu, C., 2018. Statistical modelling of extreme storms using copulas: A comparison study. Coastal Engineering 142, 52 – 61.

Liang, B., Shao, Z., Li, H., Shao, M., Lee, D., 2019. An automated threshold selection method based on the characteristic of extrapolated significant wave heights. Coastal Engineering 144, 22–32.

Losada, M.A., 2001. ROM 0.0 General procedure and requirements in the design of harbor and maritime structures. Part I. Technical Report. Puertos del Estado.

Losada, M.A., 2018. Recommendations for Maritime Works ROM 1.1-18 (Articles). Technical Report. Puertos del Estado.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media.

Martín Hidalgo, M., 2015. Caracterización multivariada de los temporales para su aplicación en el dimensionamiento del manto principal. Ph.D. thesis. Caminos.

Martín-Soldevilla, M.J., Martín-Hidalgo, M., Negro, V., López-Gutiérrez, J., Aberturas, P., 2015. Improvement of theoretical storm characterization for different climate conditions. Coastal Engineering 96, 71–80.

Van der Meer, J., Allsop, N., Bruce, T., De Rouck, J., Kortenhaus, A., Pullen, T., Schüttrumpf, H., Troch, P., Zanuttigh, B., 2018. EurOtop: Manual on wave overtopping of sea defences and related structures.

Melby, A., Kobayashi, N., 1999. Damage progression and variability on breakwater trunks, in: Coastal Structures, pp. 309–316.

Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J., 2006. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. Journal of Geophysical Research: Oceans 111, 561.

Mendonça, A., Losada, M., Solari, S., Neves, M., Reis, M.T., 2012. Incorporating a risk assessment procedure into submarine outfall projects and application to portuguese case studies, in: Coastal Engineering Proceedings.

Mentaschi, L., Besio, G., Cassola, F., Mazzino, A., 2015. Performance evaluation of Wavewatch III in the Mediterranean Sea. Ocean Modelling 90, 82–94.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris 8, 229–231.

Solari, S., Egüen, M., Polo, M.J., Losada, M.Á., 2017. Peaks over threshold (POT): A methodology for automatic threshold estimation using goodness-of-fit p-value. Water Resources Research .

Solari, S., Losada, M.Á., 2011. Non-stationary wave height climate modeling and simulation. Journal of Geophysical Research: Oceans 116, C09032.

Solari, S., Losada, M.Á., 2016. Simulation of non-stationary wind speed and direction time series. Journal of Wind Engineering and Industrial Aerodynamics 149, 48–58.

Solari, S., Van Gelder, P., 2011. On the use of vector autoregressive (var) and regime switching var models for the simulation of sea and wind state parameters. Marine Technology and Engineering 1, 217–230.

Sousa, I., Santos, J., 2006. Risk-based maintenance of rubble-mound breakwaters, in: Safety and reliability for Managing Risks. European Safety and Reliability Conference.

Thompson, P., Cai, Y., Reeve, D., Stander, J., 2009. Automated threshold selection methods for extreme wave analysis. Coastal Engineering 56, 1013 – 1021.

U.S. Army Corps of Engineers, 2002. Coastal Engineering Manual (CEM), Engineer Manual 1110-2-1100.