JOINT STOCHASTIC SIMULATION OF EXTREME COASTAL AND OFFSHORE SIGNIFICANT WAVE HEIGHTS

BY JULIETTE LEGRAND[1,a], PIERRE AILLIOT[2,c],
PHILIPPE NAVEAU[1,b] AND NICOLAS RAILLARD[3,d]

[1]*Laboratoire des Sciences du Climat et de l'Environnement, UMR8212 CEA-CNRS-UVSQ, IPSL & Université Paris-Saclay, 91191 Gif-sur-Yvettes, France ,* [a]*juliette.legrand1@inrae.fr;* [b]*philippe.naveau@lsce.ipsl.fr*

[2]*Laboratoire de Mathématiques de Bretagne Atlantique, Université de Bretagne Occidentale, 29200 Brest, France,* [c]*pierre.ailliot@univ-brest.fr*

[3]*IFREMER, RDT, F-29280 Plouzané, France,* [d]*Nicolas.Raillard@ifremer.fr*

The characterisation of future extreme wave events is crucial because of their multiple impacts, covering a broad range of topics such as coastal flood hazard, coastal erosion, reliability of offshore and coastal structures. The main goal of this paper is to propose and study a stochastic simulator that, given offshore conditions (peak direction D_p, peak period T_p and moderately high significant wave heights H_s), produces jointly offshore and coastal extreme H_s, a quantity measuring the wave severity and which represent a key feature in coastal risk analysis. For this purpose, we rely on bivariate Peaks over Threshold and a nonparametric simulation scheme of bivariate GPD is developed. From this joint simulator, a second generator is derived, allowing for conditional simulations of extreme H_s. Finally, to take into account non-stationarities, the extended generalised Pareto model is also adapted, letting the parameters vary with specific sea state parameters T_p and D_p. The performances of the two proposed generators are illustrated on simulated data and then applied to the simulation of new extreme oceanographic conditions close to the French Brittany coast using hindcast sea state data. Results show that the proposed algorithms successfully simulate future extreme H_s near the coast in a nonparametric way, jointly or conditionally on sea state parameters from a coarser model.

1. Introduction.

French coastlines have been particularly affected by extreme maritime events in the past (Nicolae Lerma et al., 2015). Co-occurrence of high tidal coefficients, atmospheric surge conditions and specific sea states can lead to extreme maritime events. These events are particularly crucial for assessing flooding risks and their consequences (Genovese and Przyluski, 2013; Bertin et al., 2012). According to the special IPCC report (see Collins et al. (2019)), extreme wave heights, which contribute to these extreme maritime events, have increased over the past few years. The recent IPCC report (Seneviratne et al., 2021) indicates, with high confidence, an increase in the occurrence and magnitude of such coastal events in the future. More specifically, Caires and Sterl (2005) showed that the most extreme wave conditions were expected to occur in the North Atlantic, which includes the Bay of Biscay, our study area.

Sea surface elevation over a geographical area results from the superposition of waves generated by local winds and by remote swell (generated in distant regions). The characterisation of this complex surface is called a sea state and to describe it, various parameters are available. In this work, we will focus on three variables: the significant wave height denoted H_s [m], the peak period, T_p [s], and the peak direction, D_p [°] (see e.g. Holthuijsen (2007) for more details).

Keywords and phrases: Bivariate extremes, Multivariate generalised Pareto distribution, Simulation of extremes, Nonstationarity, Extended generalised Pareto distribution, Covariate effects, Significant wave heights.

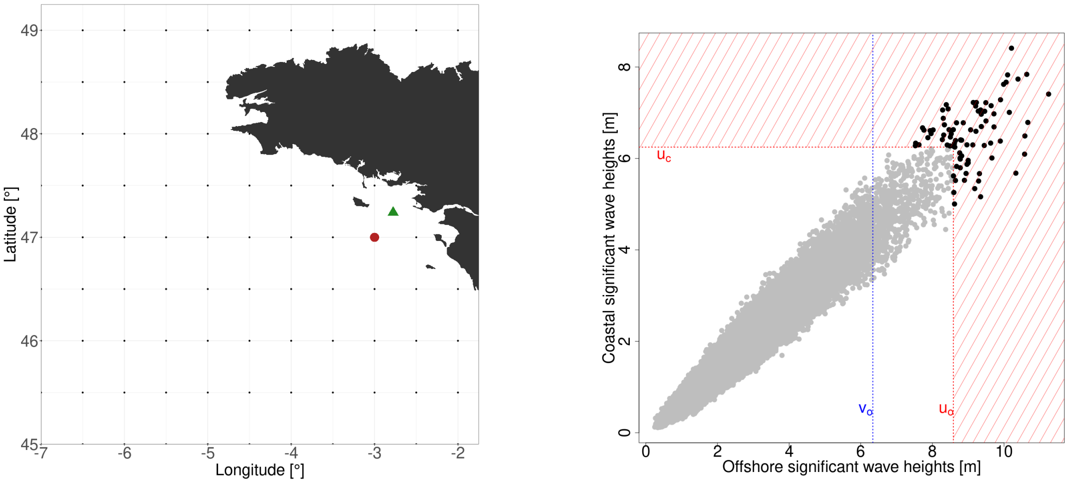


Fig 1: *(left) Portion of IOWAGA hindcast database grid, the red dot corresponds to the "offshore" point (data extracted from the IOWAGA database) and the triangular green dot corresponds to the "coastal" point (data extracted from the HOMERE database). (right) Scatter plot of the coastal significant wave heights versus the offshore significant wave heights with the different thresholds considered. The dark dots belong to the region of the data used for simulation.*

From a coastal risk point of view, a fundamental question is to determine how moderately high offshore significant wave heights can produce large coastal H_s. Peak direction and peak period influence the relationship between coastal and offshore H_s. In this context, our main goal is to propose and study a stochastic simulator that, given offshore conditions (T_p, D_p, H_s moderately high), produces jointly offshore and coastal extreme significant wave heights. Note that in this study, we assume the availability of offshore T_p and D_p, but we could also consider an appropriate bivariate generator of (T_p, D_p) (based, for example, on Heredia-Zavoni and Montes-Iturrizaga, 2019). Hereinafter, H_c (resp. H_o) will denote the coastal (resp. offshore) significant wave heights. The left-hand side map of Figure 1 shows the two locations of interest.

From such a stochastic generator (first goal), many products can be derived. In particular our second objective is to propose a conditional simulation model (second goal). The framework for each step of this study is summarised in Table 1. In order to make the two simulation models as flexible as possible, nonparametric algorithms are derived using resampling techniques (or nonparametric bootstrap (Efron, 1979)). While this study focuses on simulation of extreme H_s, and as illustrated by the numerical simulations in Section 4, the two nonparametric algorithms developed could be applied to a broad range of data.

In multivariate extreme value analysis, one is often interested in the joint behaviour of the variables as they become large. As illustrated by the right-hand side of Figure 1, which depicts a scatter plot between H_o and H_c, large values tend to occur simultaneously. For this

Table 1: *Summary of available data for each step of this study. A tick v (resp. a cross x) indicates the availability (resp. non-availability) of the data.*

		H_c	H_o	D_p	T_p
	Inference	v	v	v	v
First Goal:	Joint simulation	x	x	v	v
Second Goal:	Conditional simulation	x	v	v	v

specific type of dependence, called asymptotic dependence (Coles, 2001), models from the class of multivariate Extreme Value Theory (EVT) can be used. To achieve the two objectives (first Goal and second Goal), we will therefore propose two simulation algorithms based on multivariate Peaks over Thresholds (Sec. 8.3.1 Beirlant et al., 2004). Note that to assess whether data fall within the class of asymptotic dependence or not, summary statistics have been developed such as the dependence measures χ and $\bar{\chi}$ (Coles, Heffernan and Tawn, 1999) and, for our specific dataset, these measures do not point toward the case of asymptotic independence between H_o and H_c (see Figure 11 in Appendix A).

For independent extremes, conditional models based on Heffernan and Tawn (2004) should be favoured to deal with our second goal, see for example Towe et al. (2017); Shooter et al. (2019); Tendijck et al. (2021). In particular, if the distance between the two locations increases, the dependence is likely to decrease and other specific models should be used to address such issues (e.g. Shooter et al., 2019, 2021a).

Before modelling the joint behaviour of large values, it is necessary to model margins (Beirlant et al., 2004). To visualise this task with respect to our data, the empirical histograms displayed in Figure 2 indicate that, given $\{H_o > v_o\}$ (moderately high offshore significant wave heights), a traditional univariate extreme value approach based on fitting a generalised Pareto distribution (GPD) to the exceedances (Coles, 2001) is not appropriate. Indeed, the clear increase from zero to the mode in the left-hand histogram of Figure 2 cannot be reproduced by the probability density function (pdf) of the GPD that is a strictly decreasing function. To tackle this issue, we use the extended generalised Pareto distribution (EGPD) introduced by Naveau et al. (2016) which handles this type of setting (see also Papastathopoulos and Tawn (2013)), more details are given in Section 3.

Furthermore, like many other environmental data, extreme H_s are nonstationary with respect to covariates (Jonathan and Ewans, 2013) and marginal models need to take into account this nonstationarity (e.g. Ewans and Jonathan, 2008; Méndez et al., 2008; Casas-Prat, Wang and Sierra, 2014). To incorporate nonstationarities, Chavez-Demoulin and Davison (2005) proposed to let the parameters of an extreme value model vary as smooth functions of covariates. This has been intensively applied to oceanographic data (e.g. Feld et al., 2014;

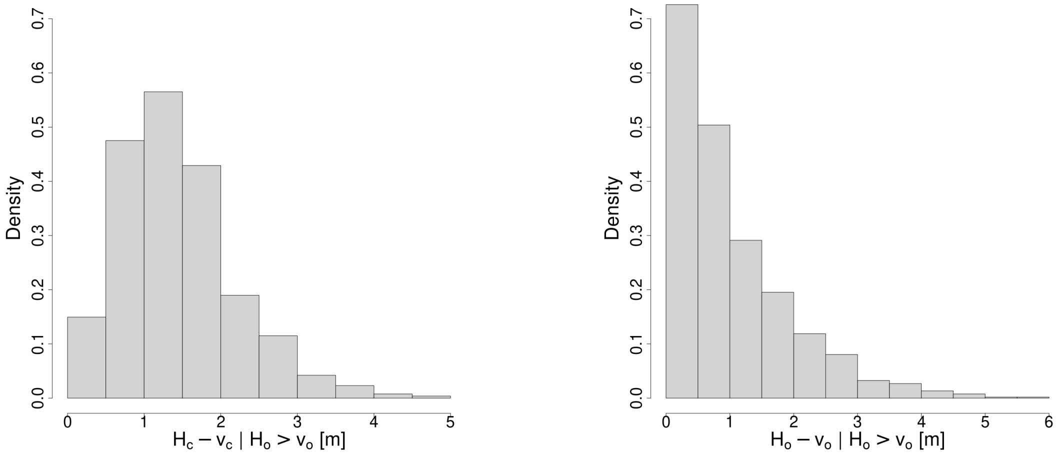


Fig 2: *(left) Empirical histogram for the coastal significant wave height threshold exceedances and (right) similarly for the offshore significant wave heights threshold exceedances, illustrating that fitting a generalised Pareto distribution (Coles, 2001) is not suitable for the coastal data. The coastal marginal threshold is defined by* $v_c := \min(H_c; H_o > v_o)$.

Jonathan, Ewans and Randell, 2014; Ross et al., 2017). However, there are only a few papers dealing with nonstationary EGPD (de Carvalho et al., 2022; Haruna, Blanchet and Favre, 2022). In this study, and as illustrated in Figure 3, the marginal EGPD models parameters will be described as smooth functions of the covariates T_p and D_p.

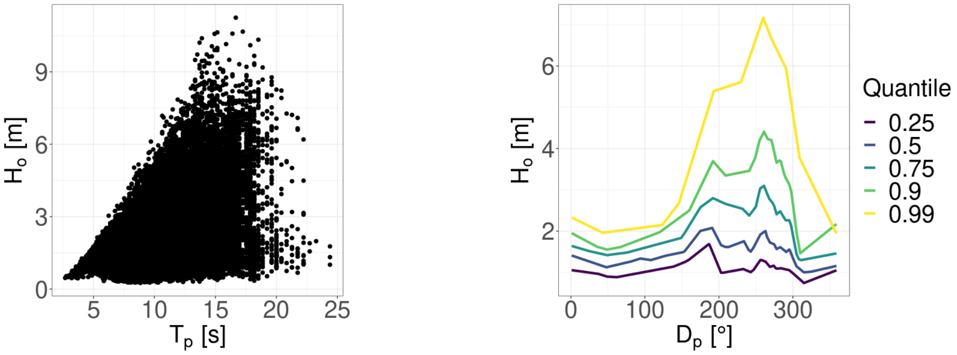


Fig 3: *Dependence of H_s with respect to T_p and D_p (left) Offshore significant wave heights H_o given peak period T_p. (right) Estimated quantiles of H_o given the peak direction D_p for different quantile levels $q \in \{0.25, 0.5, 0.75, 0.9, 0.99\}$, estimation is performed using smoothed quantile regression (Koenker, Ng and Portnoy, 1994).*

The key steps of our study are the following: (1) marginal regression modelling within the class of EGPD, (2) transformation of the data to common exponential margins, (3) modelling extremal dependence between the variables using multivariate generalised Pareto model (hereafter MGP models) (Rootzén and Tajvidi, 2006), (4) nonparametric simulation of bivariate extreme H_s within the class of MGP distributions. In our modelling scheme, different steps are novelties.To our knowledge, little attention has been paid in the literature to the modelling of multivariate nonstationary extremes using EGPD and to the nonparametric simulation within the MGP class.

Our paper is organised as follows. In Section 2, the sea state data are presented and the marginal inference incorporating covariates in the EGPD modelling is described in Section 3. In Section 4, the nonparametric method to simulate MGP vectors is presented and some numerical experiments are shown. Two algorithms are outlined, one for bivariate simulations and a second one for conditional simulations. Both algorithms are applied in Section 5 to the sea state data. The R codes along with the data used in this study can be found in the Supplementary Material A (Legrand et al. (2023)).

2. Sea state data. Our study is carried out in the northern part of the Bay of Biscay in France. The specificity of this region is that it is exposed to the Atlantic Ocean and therefore subject to complex superpositions of wind generated waves and swell. The data are extracted from two different wave hindcasts provided by IFREMER and consist of simulations of sea states by a numerical model. First, the IOWAGA database (Ardhuin and Accensi, 2014) corresponds to sea states parameters that are generated by the wave model WAVEWATCH-III and forced by CFSR winds on a Global grid (0.5° resolution grid in latitude and longitude). HOMERE is the second database (Accensi and Maisondieu, 2015), also based on WAVEWATCH-III model and forced by IOWAGA on the wet boundaries, but on an unstructured grid covering only the English Channel and the Bay of Biscay, more refined close to the coast and with the inclusion of currents and water levels.

As mentioned in the introduction, we restrict our attention to two specific locations: an offshore grid point (47°N, 3°W) from the IOWAGA database and a coastal point from the HOMERE database, near the French coast (47°24N, 2°78W) corresponding to the SEM-REV sea test site (Mouslim et al., 2009) (see Figure 1). Among all the sea states parameters, the significant wave heights $H_s[m]$ from the two locations, the peak period $T_p[s]$ and the peak direction $D_p[°]$ only from the offshore location are used. Data are available at 3-hour intervals spanning from 1994 to 2016. More precisely, the original HOMERE database has a 1-hour resolution time step but the IOWAGA database is sampled every 3 hours, so to obtain data at the same time scale a sub-sampling of the HOMERE database every 3 hours is performed. Recall that H_c (resp. H_o) correspond to the coastal (resp. offshore) significant wave heights. A scatter plot between H_c and H_o can be found in Figure 1, highlighting a strong dependence structure between the variables. Data are then split into two sets:

- Set 1 contains the first 70% of the data and is used for the inference of the marginal regression models and the preliminary steps for the simulation of H_s (see Section 5);
- Set 2 contains the remaining 30% of the data and is used for the simulation of extreme H_s.

In order to propagate the uncertainties from the marginal modelling to the bivariate simulations, a bootstrap resampling is also adopted by repeating the combined marginal and bivariate analysis for 100 resamples of the Set 1. Note that Set 2 is not involved in the bootstrap procedure.

3. Marginal regression analysis.

3.1. *Marginal regression.* In this section, only the data from Set 1 are considered. A regression model for H_c and H_o is chosen. We pre-select the extremes by considering, within the Set 1, data such that $H_o > v_o$, i.e. belonging to the right rectangular region in Figure 1, where v_o is defined as the 0.98 quantile of H_o. A common choice when someone is interested in extreme values is to work with the class of the generalised Pareto distributions (GPD) (e.g. Coles (2001)). However, this type of model always raises questions on the choice of the threshold and the GPD approximation holds true only for the very high values. In our case we want to model all the data that are to the right of the vertical blue line in Figure 1, this means that values are not necessarily extremes. To overcome such problems, Naveau et al. (2016) proposed a new class of extreme value distributions, called extended generalised Pareto distributions (EGPD). The EGPD class is suitable for modelling the entire range of data, not only the most extreme values, and avoids the need for careful threshold selection. It also ensures that both lower and upper tails are in compliance with univariate EVT. In Naveau et al. (2016), four parametric models are proposed. We restrict ourselves to the first and simplest one (corresponding to the EGP3 model introduced by Papastathopoulos and Tawn (2013)), which appears to be flexible enough, and whose cumulative distribution function is of the form

$$F(x) = \left(1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}\right)^{\kappa}. \tag{1}$$

The model has three parameters: scale $\sigma > 0$, shape $\xi \in \mathbb{R}$ and an additional parameter $\kappa > 0$ which controls the shape of the lower tail. In Naveau et al. (2016), the authors developed the EGPD for non-negative shape parameter ξ. Indeed, the main applications were the modelling of daily rainfall (e.g. Naveau et al., 2016; Tencaliec et al., 2019; de Carvalho et al., 2022; Rivoire, Martius and Naveau, 2021), which are heavy-tailed ($\xi > 0$). In our case, as reported in Jonathan and Ewans (2013), extreme H_s data are generally described by upper-bounded tail distributions. Still, the case $\xi < 0$ can be handled by model (1).

Table 2: *Estimated parameters for the regression marginal models. Point estimates and* 95% *asymptotic and bootstrap confidence intervals are given in brackets.*

Parameter	Estimate	Asymptotic CI	Bootstrap CI
ξ_C	-0.11	$[-0.15, -0.07]$	$[-0.21, -0.09]$
κ_C	4.11	$[3.57, 4.64]$	$[3.02, 4.84]$
ξ_O	-0.10	$[-0.16, -0.04]$	$[-0.17, -0.04]$
κ_O	1.16	$[1.05, 1.26]$	$[1.05, 1.28]$

As mentioned in the Introduction, H_s data are nonstationary (see Figure 3), see also Jonathan, Ewans and Randell (2014); Feld et al. (2014); De Leo et al. (2021). Therefore, we regress H_s on the peak direction D_p and the peak period T_p. We choose here to put the dependency on the scale parameter. The regression marginal models can then be written as follows

$$\begin{cases} \mathbb{P}(H_o - v_o \leq x | H_o > v_o, T_p, D_p) = \left(1 - \left(1 + \dfrac{\xi_o x}{\sigma_o(T_p, D_p)}\right)^{-1/\xi_o}\right)^{\kappa_o}, \\ \mathbb{P}(H_c - v_c \leq x | H_o > v_o, T_p, D_p) = \left(1 - \left(1 + \dfrac{\xi_c x}{\sigma_c(T_p, D_p)}\right)^{-1/\xi_c}\right)^{\kappa_c}. \end{cases} \tag{2}$$

The coastal marginal threshold v_c introduced in (2) is defined by $v_c := \min(H_c; H_o > v_o)$ so that the minimum of $H_c - v_c$ is equal to zero. This definition of v_c allows to keep a large amount of data for the simulations (see Section 5) and the use of the EGPD ensures that extreme values of H_c are still well modelled.

Note that in classical EVT, incorporating a varying threshold in a extreme value model often leads to improved marginal modelling (see for example Northrop and Jonathan (2011) for a spatially varying threshold to model extreme H_s). However, in Equation (2), the thresholds v_c and v_o should not be confused with the usual EVT thresholds. For example, going back to the left panel of Figure 2, the histogram clearly highlights a mode around 1.5. In this case, a classical EVT approach would have probably been based on varying thresholds greater than 1.5. The parameter κ of the EGPD allows us to capture the lower tail behaviour. This additional parameter appears to bring the necessary flexibility to fit well this particular dataset, see Figure 12, without the need of a varying EVT threshold. If needed, one could integrate covariate in κ (Le Carrer, 2022).

The regression marginal models (2) are estimated using the R package `gamlss` (Stasinopoulos, Rigby and Akantziliotou, 2008) with the EGPD family (Le Carrer, 2022). The inference is performed using maximum penalised likelihood estimation (note that the model fitting is achieved with the CG algorithm (Cole and Green, 1992)). In our model (2), we assume that the parameters σ_o and σ_c vary smoothly with T_p and D_p. This is achieved using tensor product of cubic regression splines. The parameter estimates for the regression marginal models are reported in Table 2. Both asymptotic 95% confidence intervals (CI), derived from the asymptotic variance-covariance matrix of the fitted models, and bootstrap 95% CI for each parameter are given in brackets. Both procedures produce similar results.

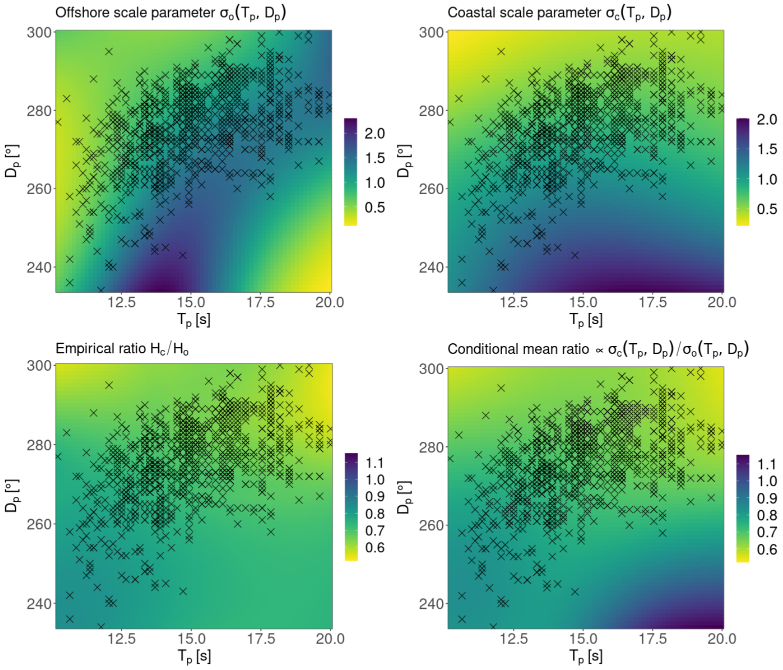
Both estimated shape parameters are negative but close to zero, which suggests light-tailed or bounded distributions. This is in accordance with previous studies and the physical behaviour of wave heights in shallow waters (Castillo and Sarabia, 1992; Vanem and Fazeres-Ferradosa, 2022). Regarding the lower tail parameter κ, looking at Equation (1) one can check that less mass is put at zero for large values of κ. Thus looking at the empirical histograms in Figure 2, the estimates of κ seem reasonable. Goodness-of-fit plots of the model (2) are shown in Appendix B.

3.2. *Covariate effects.* The covariate effects on the scale parameters can be seen on the top panels in Figure 4. Regarding the effect of D_p, we find that the biggest storms come from the WSW and with strong similarities between the coastal and the offshore model. There are more differences between the two locations if we look at the effect of T_p, which might be due to a greater loss of energy during the propagation of waves with large period. This is in particular the case for swells generated well offshore, corresponding to a large peak period, and coming from the NW, which are filtered at the coast due to the islands. To visualise in an alternative way the effect of the covariates on the scale parameters, we also choose to consider the theoretical expectation of the fitted EGPD models. As can be seen from Equation (3) (and similarly for H_o), the theoretical expectation of model (2) is directly proportional to the scale parameter (see Naveau et al. (2016)):

$$\mathbb{E}(H_c|H_o > v_o, T_p, D_p) = \sigma_c(T_p, D_p)\frac{1}{\xi_c}\left[\kappa_c B(\kappa_c, 1-\xi_c) - 1\right], \tag{3}$$

where B denotes the Beta function defined by

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt.$$


Offshore scale parameter $\sigma_o(T_p, D_p)$

Coastal scale parameter $\sigma_c(T_p, D_p)$

Empirical ratio H_c/H_o

Conditional mean ratio $\propto \sigma_c(T_p, D_p)/\sigma_o(T_p, D_p)$

Fig 4: *(top left) Estimated offshore scale parameter conditionally to the offshore peak direction D_p and the offshore peak period T_p. (top right) Identical to the top left panel but for the coastal scale parameter. (bottom left) Interpolated ratio of empirical extreme coastal significant wave heights H_c and extreme offshore significant wave heights H_o. The interpolated surface between the data points is performed using local polynomial interpolation of degree 2. (bottom right) Ratio of the predicted conditional expectations of extreme coastal significant wave heights H_c and extreme offshore significant wave heights H_o, conditionally to the offshore peak direction D_p and the offshore peak period T_p. On the four plots, observed data points are superimposed.*

The bottom left panel in Figure 4 represents the empirical ratio H_c/H_o given T_p and D_p values. Local polynomial interpolation (LOESS, Cleveland and Devlin (1988)) is performed between the observed data points to get values on a regular grid of D_p and T_p, ranging from 230 to 300 degrees for the peak direction and from 10 to 20 seconds for the peak period. This first panel is then compared to the estimated ratio of the two conditional expectations (bottom right panel in Figure 4), which, from Equation (3), is proportional to

$$\hat{\sigma}_c\left(T_p, D_p\right)/\hat{\sigma}_o\left(T_p, D_p\right).$$

This ratio can give us an idea of the propagation of the wave energy from the offshore to the coast as a function of the covariates. These results can be physically interpreted: the loss of wave energy between the offshore and the coast is lower for small periods but also for waves coming from the SW rather than the NW due to the bathymetry (see the map on Figure 1).

Using the estimated $\hat{\sigma}_c(T_p, D_p)$ and $\hat{\sigma}_o(T_p, D_p)$, the H_s data are then transformed to common exponential scale using the probability integral transform

$$\begin{aligned} H_o^E &:= -\log\left\{1 - \hat{F}_o[(H_o - v_o)/\hat{\sigma}_o(T_p, D_p)]\right\}, \\ H_c^E &:= -\log\left\{1 - \hat{F}_c[(H_c - v_c)/\hat{\sigma}_c(T_p, D_p)]\right\}, \end{aligned} \tag{4}$$

where $\hat{F}_o$ (resp. $\hat{F}_c$) is the fitted $EGPD(\hat{\xi}_o, \hat{\kappa}_o, 1)$ cdf's (resp. $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$) from Equation (2).

4. Multivariate Pareto model. In this section, the threshold exceedances of H_s transformed to common exponential margins are modelled. This vector is denoted $\boldsymbol{H^E} := (H_c^E, H_o^E)$ in the following. For that, we adapt the definition of Rootzén and Tajvidi (2006) of bivariate threshold exceedances given as

$$\left[\boldsymbol{H}^E - \boldsymbol{u} | \boldsymbol{H}^E \nleq \boldsymbol{u}\right] \tag{5}$$

where $\boldsymbol{u} := (u_c, u_o) \in \mathbb{R}_+^2$ and $\boldsymbol{H}^E \nleq \boldsymbol{u}$ means that $H_c^E > u_c$ and/or $H_o^E > u_o$, that is to say we are extreme in at least one of the two components. Then multivariate EVT theory states that (5) can be well approximated by a multivariate generalised Pareto (MGP) distribution (Rootzén and Tajvidi, 2006). Note that there are different equivalent definitions for multivariate threshold exceedances (see Section 8.3 of Beirlant et al. (2004)).

Rootzén, Segers and Wadsworth (2018) derived a stochastic representation of standard MGP vectors considering that a bivariate random vector $\boldsymbol{Z}$ follows a standard MGP distribution if, and only if,

$$\boldsymbol{Z} = E + \boldsymbol{T} - \max(\boldsymbol{T}), \tag{6}$$

with $\boldsymbol{T}$ a random vector and E a unit exponential random variable independent of $\boldsymbol{T}$. In the above equation, the addition should be interpreted component-wise and the scalars E and $\max(\boldsymbol{T})$ are repeated twice if $\boldsymbol{Z}$ is a bivariate random vector. Note that the standard MGP distribution is supported by the set $L := \{\boldsymbol{x} \in \mathbb{R}^d; \boldsymbol{x} \nleq 0\}$.

In our study, Equation (6) is adapted to $\boldsymbol{Z} = (Z_1, Z_2)$ defined as

$$\begin{cases} Z_1 := H_o^E - u_o | H_o^E > u_o \text{ or } H_c^E > u_c, \\ Z_2 := H_c^E - u_c | H_o^E > u_o \text{ or } H_c^E > u_c. \end{cases} \tag{7}$$

Algorithm 1 Nonparametric bootstrap MGP simulation

```
input A sample $(Z_{1,i}, Z_{2,i})_{1\leq i\leq n}$ from a MGP distribution
output A simulated sample $(Z_{1,k}^{(m)}, Z_{2,k}^{(m)})_{1\leq k\leq m}$, potentially with $m \neq n$
procedure
    Define $\Delta_i := Z_{1,i} - Z_{2,i}$ for $1 \leq i \leq n$
    Generate $m$ realisations $E_k^{(m)} \sim Exp(1)$, independently of $(\Delta_i)_{1\leq i\leq n}$, for $1 \leq k \leq m$
    Bootstrap $m$ realisations $\Delta_k^{(m)}$, $1 \leq k \leq m$, from $(\Delta_1, \dots, \Delta_n)$
end procedure
return $Z_{1,k}^{(m)} := E_k^{(m)} + \Delta_k^{(m)} \mathbb{1}_{\Delta_k^{(m)}<0}$ and $Z_{2,k}^{(m)} := E_k^{(m)} - \Delta_k^{(m)} \mathbb{1}_{\Delta_k^{(m)}>0}$, for $1 \leq k \leq m$
```

4.1. *Simulation of bivariate standard generalised Pareto distributed vectors.* Kiriliouk et al. (2019) established several parametric MGP models by setting explicit densities for $\boldsymbol{T}$ in a multivariate setting. In the following, we consider only vectors of dimension 2, *i.e.* $\boldsymbol{Z} = (Z_1, Z_2)$ and $\boldsymbol{T} = (T_1, T_2)$. To bypass the choice of the underlying distribution for $\boldsymbol{T}$, we start from the following rewriting of Equation (6),

$$\begin{cases} Z_1 = E + \Delta \mathbb{1}_{\Delta<0}, \\ Z_2 = E - \Delta \mathbb{1}_{\Delta\geq 0}, \end{cases} \tag{8}$$

where $\Delta := Z_1 - Z_2 = T_1 - T_2$ and $\mathbb{1}_A$ denotes the indicator function, equals to 1 if A is true and 0 otherwise.

Equation (8) is the basis for our simulation algorithms. From this equation, we see that we need to simulate values of Δ and E independently, instead of (T_1, T_2). Generating independent, and identically distributed, unit exponentials is trivial, so the main difficulty is to simulate Δ. This can be achieved by bootstrapping (see Efron (1979)). Our approach is then described in Algorithm 1 and a theoretical proof can be found in Appendix C.

4.2. *Numerical experiments.* In the following, we simulate MGP vectors $\boldsymbol{Z} = (Z_1, Z_2)$ from the representation (6) with different parametric models on (T_1, T_2) and we compare with our simulation algorithm. The different experiments are reported in Table 3 and some graphical results are shown in Figure 5 which displays for each model a scatter plot of the data, the measure of extremal dependence $\chi(u)$ for increasing values of u, and the marginal quantile-quantile plots. We use the measure $\chi(u)$ which gives a measure of asymptotic dependence between two variables X and Y (for more details see *e.g.* Coles, Heffernan and Tawn (1999)) and which is defined by

$$\chi(u) := \mathbb{P}\left(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)\right), \ u \in (0,1). \tag{9}$$

In Table 3, (a) and (b) are two bivariate Gaussian models with same correlation coefficient $\rho < 1$ but with $\mu_1 \neq \mu_2$ for (b), leading to asymmetry. Model (c) corresponds to the Type I bivariate logistic distribution proposed by Gumbel (1961). For model (d) we consider two independent Gumbel distributed variables with different scale parameters. And lastly, (e) corresponds to a bivariate exponential distribution as defined in Marshall and Olkin (1967).

As seen in Figure 5, Algorithm 1 successfully simulates draws from the parametric simulations in terms of the marginals Z_1 and Z_2, but also recovers well the dependence structure when looking at the measure of dependence $\chi(u)$.

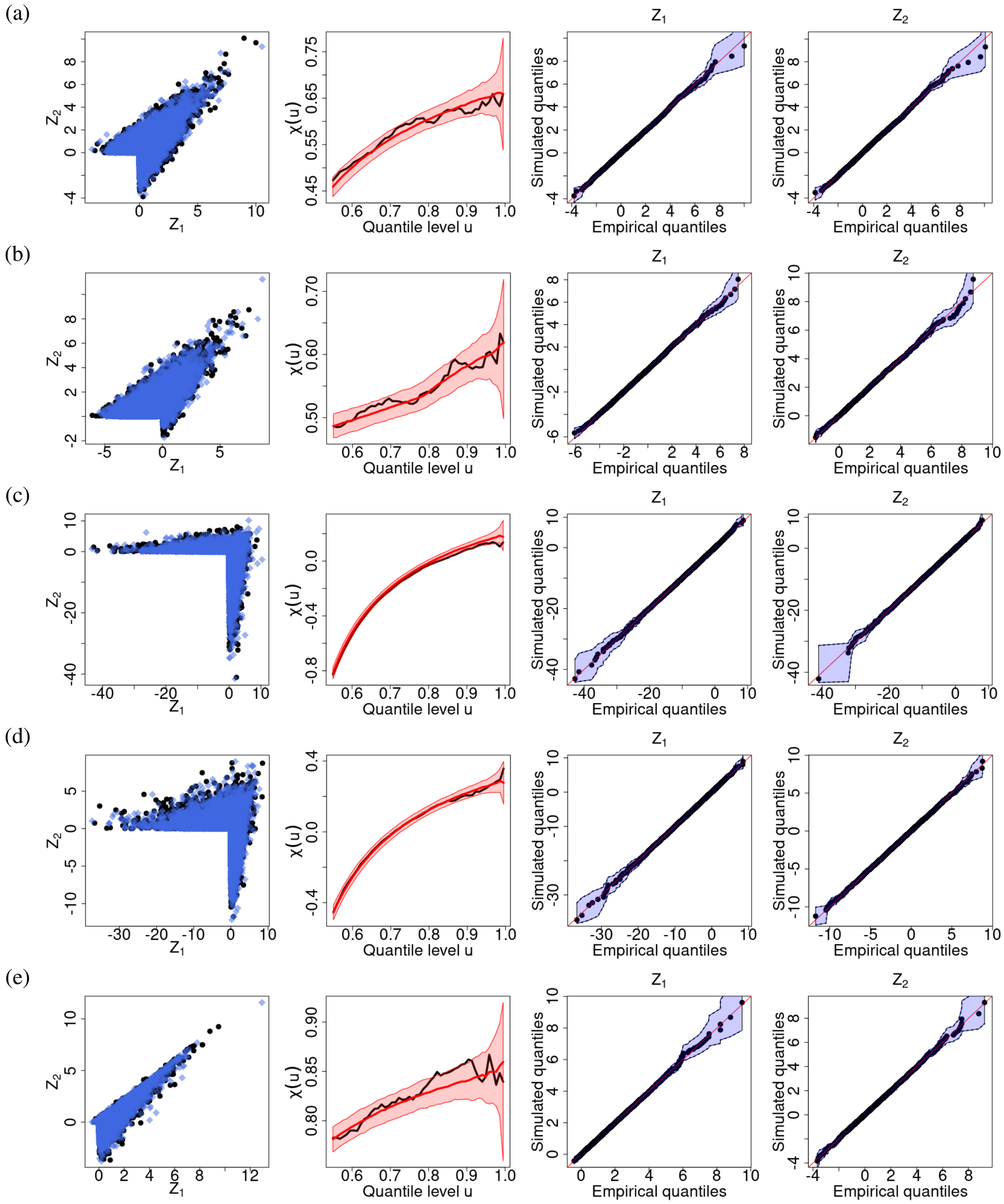


Fig 5: *Each panel line corresponds to one of the parametric model* (a) *to* (e)*, and shows from left to right:* (1) *Scatter plot of simulated data with the parametric model with sample size* $n = 10000$ *(black dots) and sampled data from one simulation using Algorithm 1 with sample size* $m = 10000$ *(blue diamond-shaped dots);* (2) *Empirical estimates of the measure of asymptotic dependence* $\chi(u)$ *for the simulated data with the parametric model (black line) and for the sampled data from Algorithm 1 (red line), with associated* 95% *pointwise confidence intervals based on* 1000 *bootstrap replications;* (3) *and* (4) *Quantile-quantile plots for* Z_1 *and* Z_2 *with associated* 95% *pointwise confidence intervals based on* 1000 *bootstrap replications.*

Table 3: *Overview of the different experiments carried-out given the joint distribution of $\boldsymbol{T}$. For each, we give the joint distribution $F(x_1, x_2)$ when it can be written easily or the survival function $S(x_1, x_2)$. In the third column, we give the different parameters values used in the numerical experiments.*

Bivariate model	Joint distribution of $\boldsymbol{T}$	Parameters
(a) Gaussian symmetric	$\mathcal{N}\left((\mu_1, \mu_2), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$	$\mu_1 = 0$ $\mu_2 = 0$ $\rho = 0.4$
(b) Gaussian asymmetric	$\mathcal{N}\left((\mu_1, \mu_2), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$	$\mu_1 = 0$ $\mu_2 = 2$ $\rho = 0.4$
(c) Logistic	$F(x_1, x_2) = \left(1 + e^{-x_1/\sigma_1} + e^{-x_2/\sigma_2}\right)^{-1}$	$\sigma_1 = 1$ $\sigma_2 = 5$
(d) Gumbel	$F(x_1, x_2) = \exp\left[-\exp\{-x_1/\sigma_1\}\right] \exp\left[-\exp\{-x_2/\sigma_2\}\right]$	$\sigma_1 = 1$ $\sigma_2 = 4$
(e) Exponential	$S(x_1, x_2) = \exp\{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_3 \max(x_1, x_2)\}$	$\lambda_1 = 2$ $\lambda_2 = 10$ $\lambda_3 = 1$

4.3. *Conditional simulation within the MGP class.* From an application perspective, we also want to be able to simulate conditionally on one of the two variables. In this section we describe the conditional simulation algorithm for the MGP model.

From Equation (6) we can derive

$$Z_2 = Z_1 + T_2 - T_1 = Z_1 - \Delta. \tag{10}$$

From Equation (10), we can design a simulation strategy but caution is required because Δ and Z_1 are not necessarily independent. But for some values of Z_1, this will be the case. To see this, one can compute the conditional distribution of Δ given $Z_1 = z_1$ starting from the joint distribution function of (Z_1, Z_2) which is given by

$$f_{(Z_1,Z_2)}(z_1, z_2) = e^{-\max(z_1,z_2)} f_\Delta(z_1 - z_2) \mathbb{1}_{(z_1,z_2)\in L}\ , \text{ for } z_1, z_2 \in \mathbb{R},$$

where f_Δ denotes the distribution function of Δ.

First case: If $z_1 > 0$. In this case, noting that if $z_1 > 0$ then $\mathbb{1}_{(z_1,z_2)\in L} = 1$, the marginal distribution of Z_1 is given as follows

$$f_{Z_1}(z_1) = e^{-z_1} K,$$

where $K := \int_{-\infty}^{0} e^u f_\Delta(u) du + \int_0^\infty f_\Delta(u) du$ and does not depend on z_1.

Therefore, the conditional distribution of Z_2 given $Z_1 = z_1$ when $z_1 > 0$ is given by

$$f_{Z_2|Z_1}(z_2 \mid z_1) = \frac{1}{K}\left[f_\Delta(z_1 - z_2)\mathbb{1}_{z_1 \geq z_2} + e^{z_1 - z_2} f_\Delta(z_1 - z_2)\mathbb{1}_{z_1 < z_2}\right]$$

From this, the conditional distribution of Δ given $Z_1 = z_1 > 0$ can then be deduced:

$$f_{\Delta|Z_1}(\delta \mid z_1) = \frac{1}{K}\left[f_\Delta(\delta)\mathbb{1}_{\delta \geq 0} + e^\delta f_\Delta(\delta)\mathbb{1}_{\delta < 0}\right]. \tag{11}$$

This shows that, conditionally on $Z_1 > 0$, Δ does not depend on Z_1.

Second case: If $z_1 < 0$. Then, noting that $\mathbb{1}_{(z_1,z_2)\in L} = \mathbb{1}_{z_2 > 0}$ if $z_1 < 0$, we get

$$f_{Z_1}(z_1) = e^{-z_1} K(z_1),$$

Algorithm 2 Nonparametric conditional MGP simulation

```
 1: input A sample $(\Delta_i)_{1\leq i\leq n}$ ; a realisation $z_1$ of $Z_1$
 2: output A simulated sample $(Z_{2,k}^{(m)})_{1\leq k\leq m}$ conditionally on $Z_1 = z_1$, potentially with $m \neq n$
 3: procedure
 4:     if $z_1 > 0$ then
 5:         Define $\Delta_{|Z_1^+}$ the subset of $(\Delta_i)_{1\leq i\leq n}$ such that $Z_1 > 0$
 6:         Bootstrap $m$ realisations $\Delta_k^{(m)}$, $1 \leq k \leq m$, from $\Delta_{|Z_1^+}$ independently of $Z_1$
 7:     else
 8:         for $1 \leq k \leq m$ do
 9:             Sample one realisation $\Delta_k^{(m)}$ from $(\Delta_i)_{1\leq i\leq n}$ independently of $Z_1$
10:             Generate a random number $u \in [0,1]$
11:             while $u > \exp(\Delta_k^{(m)})\mathbb{1}_{\Delta_k^{(m)}<z_1}$ do
12:                 Repeat steps 9 and 10
13:         end for
14: end procedure
15: return $Z_{2,k}^{(m)} := z_1 - \Delta_k^{(m)}$ for $1 \leq k \leq m$
```

where $K(z_1) := \int_{-\infty}^{z_1} e^u f_\Delta(u)du$. And we can derive the conditional distribution of Z_2 given $Z_1 = z_1 < 0$ as follows

$$f_{Z_2|Z_1}(z_2 \mid z_1) = \frac{1}{K(z_1)} e^{z_1 - z_2} f_\Delta(z_1 - z_2)\mathbb{1}_{z_2>0}.$$

The conditional distribution of Δ given $Z_1 = z_1 < 0$ is then given by

$$f_{\Delta|Z_1}(\delta \mid z_1) = \frac{1}{K(z_1)} e^{\delta} f_\Delta(\delta)\mathbb{1}_{\delta<z_1}. \tag{12}$$

From Equations (11) and (12) we derive the conditional simulation algorithm described in Algorithm 2, where the simulation procedure is split into two cases:

1. If $z_1 > 0$, we can sample values of Δ independently of Z_1,
2. otherwise, if $z_1 < 0$, we use a rejection sampling approach to approximate the targeted conditional density in Equation (12).

Similarly, we could also derive a simulation scheme of Z_1 given $Z_2 = z_2$.

4.4. *Numerical experiment continued.* As for the bivariate simulations, we can illustrate Algorithm 2 with numerical experiments. We choose here to show the results only for Model (a) (Symmetric Gaussian) since for this specific model we have an explicit form for the theoretical distribution of Δ. The results are presented in Figure 6 where we simulated the conditional distribution of Z_2 for eight different conditioning values. The sampled and theoretical conditional distributions appear to be in close conformity.

5. Application to extreme significant wave height.

The methodology presented in Section 4 is applied to the joint and the conditional simulations of extreme significant wave heights. For that, the sample of bivariate threshold exceedances (Z_1, Z_2) defined in Equation (7) is used as input data for Algorithm 1 or Algorithm 2. The thresholds u_o and u_c in (7) are defined as the 0.8 quantile of H_o^E, or equivalently of H_c^E. This quantile has been selected following the stability plot approach proposed in Kiriliouk et al. (2019).

Recall that in both cases, simulations are performed on the exponential scale. A final step of back transformation is then necessary to get simulations of H_s on the original scale. This final step corresponds to part 3 (resp. 5) in the following procedure for the joint (resp. conditional) simulation of H_s. For the sake of clarity we now divide the joint and the conditional simulation scheme of H_s in two separate sections. Note that in the following two sections, simulations are performed given the marginal parameters estimates. The quantification of the uncertainty propagation from the marginal inference step to the bivariate estimation is postponed to Section 5.3.

5.1. *Joint simulation of significant wave heights.* The joint simulation scheme for extreme H_s is described hereafter. In the following we fix the pair value $(t_p, d_p) \in \mathbb{R}^2$ which may be taken from Set 2.

1. Compute $\hat{\sigma}_o(t_p, d_p)$ and $\hat{\sigma}_c(t_p, d_p)$ from the marginal EGPD models fitted on Set 1 (see Section 3).
2. Simulate m pairs of (z_1, z_2) applying Algorithm 1 with input data (Z_1, Z_2) as defined in (7). We therefore obtain m simulated pairs $((z_{1,1}, z_{2,1}), \dots, (z_{1,m}, z_{2,m}))$ for a fixed value (t_p, d_p).
3. Transform the simulated values to the original scale

$$h_{o,i} := \hat{\sigma}_o(t_p, d_p)\hat{F}_o^{-1}(1 - e^{-(z_{1,i}+u_o)}) + v_o \in \mathbb{R}^m,$$
$$h_{c,i} := \hat{\sigma}_c(t_p, d_p)\hat{F}_c^{-1}(1 - e^{-(z_{2,i}+u_c)}) + v_c \in \mathbb{R}^m,$$

   where $\hat{F}_o^{-1}$ (resp. $\hat{F}_c^{-1}$) is the inverse cdf of the $EGPD(\hat{\xi}_o, \hat{\kappa}_o, 1)$ (resp. $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$) estimated in Section 3.

This procedure is then applied to four selected pairs (t_p, d_p) from the Set 2 corresponding to the four largest H_c of this dataset. Figure 7 depicts simulated pairs of offshore and coastal H_s with simulation sample size $m = 1000$.

Figure 7 shows that for these specific conditions, Algorithm 1 successfully generates extreme H_c and H_o. Note that since the four points considered are the largest observations, they are expected to be among the extremes of the simulated distributions.

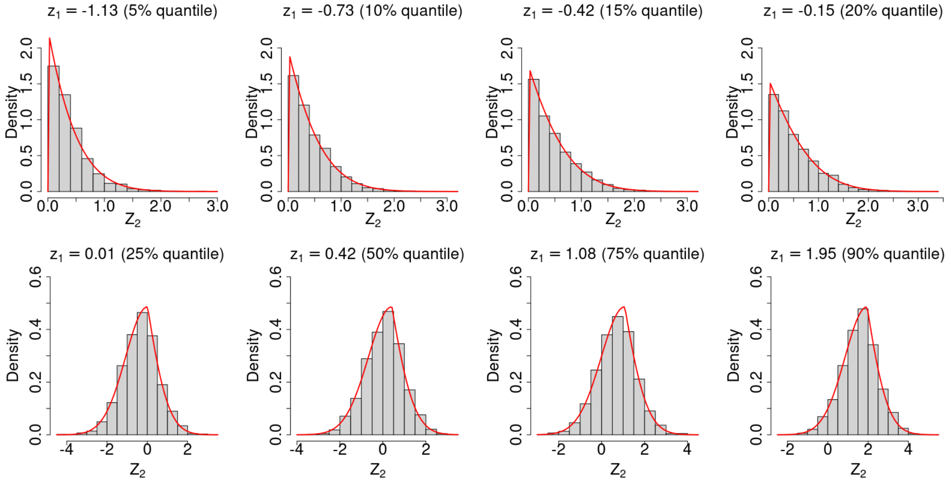


Fig 6: *Sampled conditional distribution of Z_2 given $Z_1 = z_1$ using Algorithm 2 for the asymmetric Gaussian model with sample size $n = 10000$. Eight experiments are presented for different quantiles of Z_1 whose values are reported in each panel title. The sample size for each simulation is $m = 10000$. The theoretical conditional density is superimposed in red.*

Table 4 describes the eight largest h_c of Set 2, giving the time event and the corresponding h_o, t_p and d_p values. The last column of Table 4 gives the empirical estimate of the joint survival probability defined as $1/m \sum_{i=1}^{m} \mathbb{1}\{h_{o,i} > h_o, h_{c,i} > h_c\}$, applying the above procedure with simulation sample size $m = 1.10^6$. This estimated probability quantifies the observations made with Figure 7: for the most extreme events, the associated joint probabilities are expected to be lower.

5.2. *Conditional simulation of coastal significant wave heights.* The conditional simulation scheme for extreme coastal H_s is described below. Note that the procedure is symmetrical for simulating offshore H_s. In the following we fix the triplet value $(h_o, t_p, d_p) \in \mathbb{R}^3$ which may be taken from Set 2.

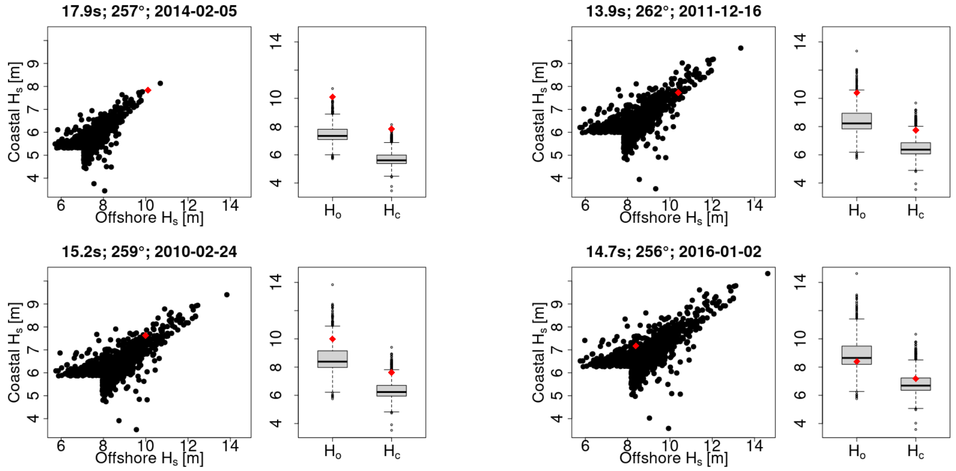


Fig 7: *Sampled values of coastal versus offshore significant wave heights from Algorithm 1,* $m = 1000$ *pairs of points are simulated. Each panel corresponds to a fixed value of* (t_p, d_p) *corresponding to the four largest* H_c *from Set 2 in decreasing order (recall that Set 2 contains the last* 30% *of the original dataset that have not yet been used). Next to each scatter plot, marginal distribution of simulated* H_o *and* H_c *are depicted with boxplots. On both the scatter plots and the boxplots, the red diamond-shaped points represent the true values of the coastal and offshore significant wave heights for the four pairs.*

Table 4: *Empirical joint survival probability of exceeding the observed extreme significant wave heights* h_c *and* h_o *for the eight largest coastal significant wave heights of Set 2. Only events from different storms are given (i.e. events separated by more than* 3 *hours). Estimation is performed using* $m = 10^6$ *simulated pairs* $(h_{o,i}, h_{c,i})_{1 \leq i \leq m}$ *for each largest event with Algorithm 1.*

	Date-Time	H_c [m]	H_o [m]	T_p [s]	D_p [°]	Joint probability
1	2014-02-05 12:00:00 GMT	7.83	10.1	17.86	257	0.002
2	2011-12-16 03:00:00 GMT	7.74	10.4	13.89	262	0.045
3	2010-02-24 15:00:00 GMT	7.62	10.0	15.15	259	0.057
4	2016-01-02 06:00:00 GMT	7.18	8.4	14.71	256	0.277
5	2013-12-24 06:00:00 GMT	7.08	8.4	14.71	253	0.388
6	2014-02-14 21:00:00 GMT	7.06	8.3	14.08	248	0.691
7	2015-01-15 09:00:00 GMT	6.45	7.9	13.70	260	0.489
8	2016-03-28 03:00:00 GMT	6.30	7.6	14.08	256	0.762

1. Compute $\hat{\sigma}_o(t_p, d_p)$ and $\hat{\sigma}_c(t_p, d_p)$ from the marginal EGPD models fitted on Set 1 (see Section 3).
2. Transform h_o to the standardised space using the probability integral transform:
$$h_o^E = -\log\left\{1 - \hat{F}_o[(h_o - v_o)/\hat{\sigma}_o(t_p, d_p)]\right\}$$
where $\hat{F}_o$ is the $EGP(\hat{\xi}_o, \hat{\kappa}_o, 1)$ cdf's.
3. Set $z_1 := h_o^E - u_o$, where u_o is the threshold on the offshore H_s on the exponential scale.
4. Simulate z_2 applying Algorithm 2, only in the case $z_1 > 0$. Here $\Delta_{|Z_1^+}$ is defined from Set 1 through $\Delta_{|Z_1^+} := (Z_1 - Z_2)\mathbb{1}\{Z_1 > 0\}$, with Z_1, Z_2 as defined in (7), and bootstrapped m times. We therefore obtain m simulations of $z_2 = (z_{2,1}, \dots, z_{2,m})$ for a fixed triplet (h_o, t_p, d_p), given $z_1 > 0$.
5. Transform the predicted values to the original scale
$$h_{c,i} := \hat{\sigma}_c(t_p, d_p)\hat{F}_c^{-1}(1 - e^{-(z_{2,i}+u_c)}) + v_c$$
where $\hat{F}_c^{-1}$ is the inverse cdf of the $EGPD(\hat{\xi}_c, \hat{\kappa}_c, 1)$ (see Section 3).

Note that in Step 4 above, the simulation is restricted to the case when $z_1 > 0$ for convenience, since our focus is on the simulation of extreme H_s.

The pseudo-algorithm described above is applied with all the triplet values (h_o, t_p, d_p) from Set 2, with simulation sample size $m = 1000$. These conditional simulations of coastal significant wave heights are then compared to the true values of H_c from Set 2. The overall coverage probability (*i.e.* the number of times the actual value of coastal H_s is within the 95% range of the predicted distribution) is equal to 95% and the simulations are shown in Figure 8. The simulations and the true H_c values (red dots) are most of the time in good agreement. Since no declustering approach has been adopted, consecutive observations, which belong to the same storm event, are kept. They are depicted with identical colour, highlighting a temporal dependence structure between each storm cluster.

The effect of the covariates T_p and D_p in the conditional simulations is depicted in Figure 9, showing that the simulation model is able to simulate both the most intense and the more moderate H_c. This plot also highlights for which sea state conditions the simulations are far

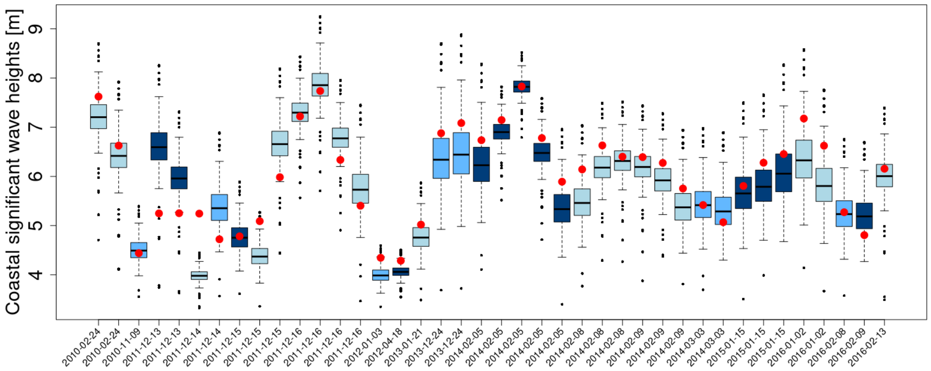


Fig 8: *Boxplot of predicted H_c conditionally on (H_o, T_p, D_p) using Algorithm 2. The simulation sample size is equal to $m = 1000$ for each observation. Red dots represents the observed H_c values from Set 2. The alternating colours depict different storms: consecutive boxes with same colour correspond to observations that belong to the same storm (i.e. separated by less than 3 hours).*

from the observed values. It appears that the two predictions such that the observed H_c value does not fall within the 95% simulation range correspond to small H_c and H_o.

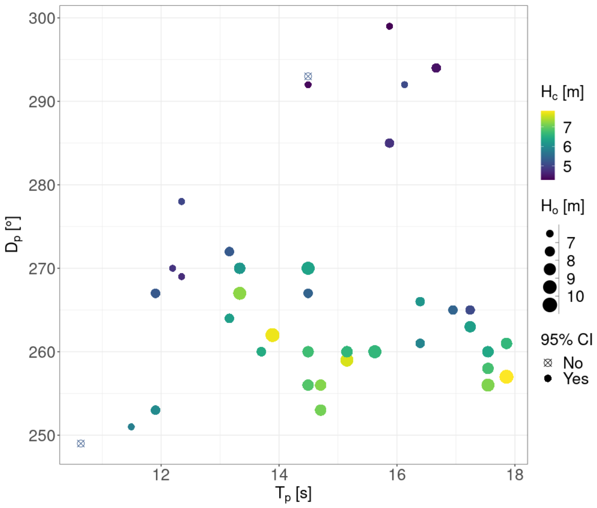


Fig 9: *Scatter plot from Set 2 with the peak direction D_p on the y-axis and the peak period T_p on the x-axis. A dot's colour corresponds to the value of the coastal significant wave heights H_c. The size of the dots corresponds to the value of the offshore significant wave heights H_o. The shape indicates if the observed H_c falls within the 95% range of the predicted distribution from the conditional model where the simulation sample size is equal to $m = 1000$ for each observation.*

5.3. *Uncertainty quantification.* As mentioned in Section 2, in addition to these simulations, we also combined the marginal and the dependence analysis in a bootstrap scheme in order to accommodate for marginal uncertainties. Both Algorithm 1 and Algorithm 2 have been applied within the bootstrap procedure, leading to 100 bootstrap sample of joint and conditional simulations of H_s. Results of each iteration can be found in the Supplementary Material B (Legrand et al. (2023)), showing the reproduction of Figures 7 and 8 for each bootstrap sample. Overall, it seems that the performances are reasonably good within each resampling iteration. Also, regarding the conditional simulation algorithm, the mean coverage probability over the 100 bootstraps is equal to 93%.

6. Discussion and conclusions. Simulation of extreme events in a multivariate setting is of great interest to capture not only the statistical behaviour of the extremes, but also the dependence between large values of complex processes. Based on the multivariate EVT, this work presents two nonparametric simulation algorithms of bivariate generalised Pareto distributed variables, without assuming any specific parametric shape for the MGP model. Thanks to Algorithm 1, one can simulate joint extremes. As for Algorithm 2, it allows the simulation of conditional extremes. Both methods have been validated with numerical simulations.

We would like to point out that in the context of bivariate extremes, other simulation algorithms have been developed. For example Marcon, Naveau and Padoan (2017) proposed a simulation method with a semi-parametric structure for the extremal dependence function, but it was not based on a MGP model and did not cover nonstationarities. Michel (2006) derived a nonparametric simulation framework of bivariate generalised Pareto variables using a different representation of MGP vectors than the one used in this paper.

For application purposes and as a by-product, a nonstationary marginal modelling with the EGPD was also developed, adding covariate effects on the scale parameter of the EGPD using smoothing splines.

We applied this work to the simulation of extreme significant wave heights near the Brittany coast given specific offshore sea state conditions (T_p, D_p) with compelling results. In both joint and conditional settings, thanks to the presented algorithms, we are able to simulate realistic extreme H_s events. Regarding the marginal models, other studies have considered extreme value models for H_s with both varying scale and shape parameters (e.g. Jonathan and Ewans, 2007; Feld et al., 2014). Incorporating effects of T_p and D_p on both parameters could certainly improve the models, but at the cost of computational limitations. Moreover, one could argue that for our specific dataset, the sea states could be considered as homogeneous since D_p has a constrained domain between $[234°, 300°]$.

Note that in possible extensions of this work to climate projections, it is assumed that data will not be available at the coastal location but only on a coarse grid, similar to the IOWAGA Global hindcast. This argument, which is illustrated in Table 1, motivates the first pre-selection of the H_s data through $\{H_o > v_o\}$ for the marginal regression analysis (see Section 3).

Extensions to the multivariate case will be the subject of future works. Considering more than two locations raises different modelling issues. It would also be interesting to apply this methodology to other locations in order to ensure the proper generalisation of the methodology.

Finally, as already mentioned, a classical approach regarding extreme significant wave height simulations is the conditional extreme model of Heffernan and Tawn (2004). For example, Shooter et al. (2021b, 2022) applied this model in a spatial setting to characterise the behaviour of the extremal dependence structure between different metocean variables. One strength of our bivariate model (hereafter referred to as MGPD approach) is that a conditional sampling scheme can be easily produced from simulations of joint extremes. So, we can compare the MGPD approach to the traditional conditional model of Heffernan and Tawn (2004) for our data at hand. We consider hereinafter the conditional generation of extreme H_s on the standardised scales. In other words, we generated standardised coastal significant wave heights on the exponential scale with Algorithm 2 on the one hand and with the conditional extreme value model of Heffernan and Tawn (2004) on the other hand. We simulated $n_{\text{sim}} = 1000$ samples of same length as the observation exceedances $h_o^E > u_o$. Figure 10 depicts one of these n_{sim} simulations of coastal H_s versus observed offshore H_s in the exponential space. One can see that for this specific simulated sample, the conditional extreme approach tends to simulate less extreme values of H_c than the MGPD approach. A first comparison has been made computing the mean absolute error between the observed data and the simulations for both models. Results are depicted in the right panel of Figure 10, showing slightly better results for the MGPD model.

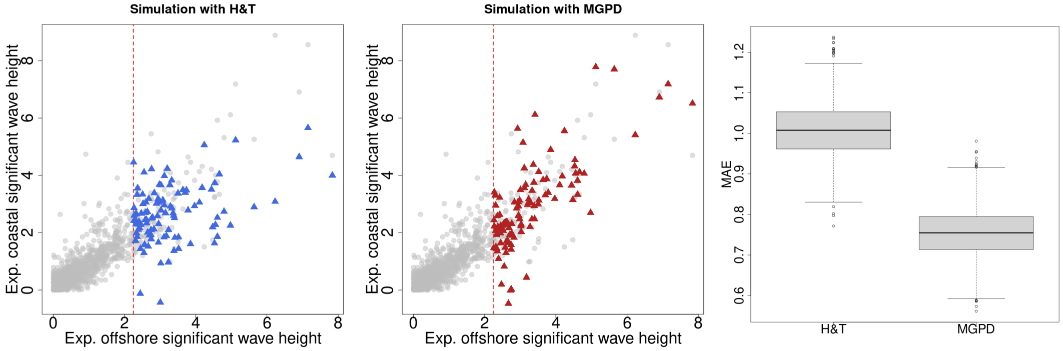


Fig 10: *Scatter plots between coastal H_s and offshore H_s in the standardised space (exponential margins). The coloured triangle shaped points correspond to one simulation run with the conditional extreme model of Heffernan and Tawn (2004) (left) and with our MGPD approach (middle). The vertical dashed line corresponds to u_o. (right) Mean absolute error over the $n_{sim} = 1000$ simulations for both models: Heffernan and Tawn (2004) on the left - MGPD approach on the right.*

APPENDIX A: TAIL DEPENDENCE BETWEEN H_s DATA

Figure 11 depicts the estimated tail dependence measures $\chi(u)$ (see (9)) and $\bar{\chi}(u)$ between the offshore and the coastal significant wave heights. Recall that the dependence measure $\bar{\chi}(u)$ (e.g. Coles, Heffernan and Tawn, 1999) between two variables X and Y is define as

$$\bar{\chi}(u) = \frac{2\log(1-u)}{\log \mathbb{P}\left(F_X(X) > u, F_Y(Y) > u\right)}, \; u \in (0,1).$$

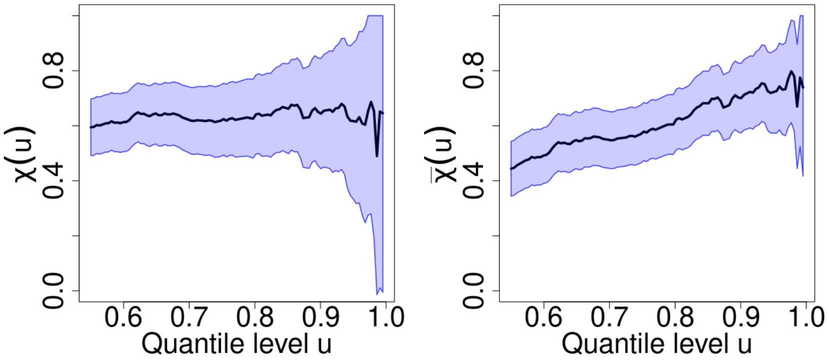


Fig 11: *Empirical estimates of the measure of asymptotic dependence $\chi(u)$ (left) and $\overline{\chi}(u)$ (right) between H_o and H_c, with 95% pointwise confidence intervals.*

APPENDIX B: MARGINAL REGRESSION MODELLING

We show here the goodness of fit for the marginal regression models defined in Equation (2). As our models depend on some covariates, the diagnostic plots presented here are built for the *standardised* H_s exceedances which are defined as $(H_c - v_c)/\sigma_c(T_p, D_p)$ (and similarly for H_o). From Figure 12, one can see that the fits seem to be fairly good, a slight discrepancy

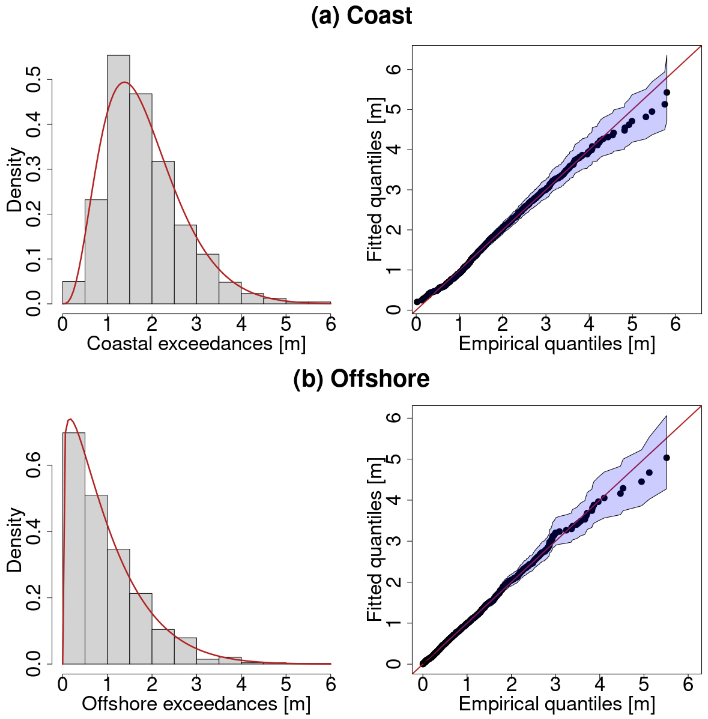


Fig 12: *(left) Empirical histogram (grey) of the standardised extreme significant wave heights exceedances (a) at the coast and (b) offshore. The fitted EGPD density is superimposed. (right) The corresponding quantile-quantile plots with associated* 95% *pointwise confidence intervals computed using parametric bootstrap.*

in the lower values can be noticed for the coastal model but this is not a major issue as the interest lies mainly in the larger values.

APPENDIX C: PROOF OF ALGORITHM 1

With the same notations as in Algorithm 1, let F be the common distribution function of $\Delta_1,\dots,\Delta_n$ and $F_{nm}^{(m)}$ be the empirical distribution function of the bootstrap sample $\Delta_1^{(m)},\dots,\Delta_m^{(m)}$.

LEMMA 1. *If $F_{nm}^{(m)}$ converges in distribution to F, as n and m tend to infinity, then $(Z_{1,k}^{(m)},Z_{2,k}^{(m)})_{1\le k\le m}$ converge in distribution to a bivariate GPD G where G is the common distribution function of the sample $(Z_{1,i},Z_{2,i})_{1\le i\le n}$.*

PROOF. As $P(E\le u)=1-\min(1,\exp(-u))$ for any $u\in\mathbb{R}$ if $E\sim Exp(1)$, the bivariate distribution function of $(Z_1^{(m)},Z_2^{(m)})$ is equal to

$$\mathbb{P}\left[(Z_1^{(m)},Z_2^{(m)})\le(z_1,z_2)\right]=1-\mathbb{E}\left[\min\left(1,e^{-\min(z_1-\Delta^{(m)}\mathbb{1}_{\Delta^{(m)}<0},z_2+\Delta^{(m)}\mathbb{1}_{\Delta^{(m)}>0})}\right)\right]$$
$$=1-\mathbb{E}\left[\min\left(1,e^{-\min(z_1-\Delta^{(m)},z_2)-\max(\Delta^{(m)},0)}\right)\right],$$

for any $(z_1,z_2)\in\{\boldsymbol{x}\in\mathbb{R}^2;\boldsymbol{x}\not\leq 0\}$.

Then, one can show that the function $x \mapsto \min(1,x)$, defined for $x \geq 0$, is Lipschitz and bounded by 1. And applying the Portmanteau theorem, we have, letting $\min(n,m) \to \infty$,

$$\begin{aligned}\mathbb{P}\left[(Z_1^{(m)}, Z_2^{(m)}) \leq (z_1, z_2) \mid \Delta_1, \ldots, \Delta_n\right] &\to 1 - \mathbb{E}\left[\min\left(1, e^{-\min(z_1-\Delta, z_2)-\max(\Delta,0)}\right)\right] \\ &= 1 - \mathbb{E}\left[\min\left(1, e^{\max(T_1-z_1, T_2-z_2)-\max(T_1,T_2)}\right)\right].\end{aligned}$$

This is the cumulative distribution function of the MGP vector (Z_1, Z_2) as defined in Rootzén, Segers and Wadsworth (2018) (Prop. 8). □

The assumption in Lemma 1 is linked to bootstrap asymptotic theory (e.g. Bickel and Freedman (1981)).

Acknowledgments. The authors would like to thank Noémie Le Carrer for helpful discussion on the implementation of the EGPD family within the `gamlss` package.

The authors are indebted to the Laboratoire d'Océanographie Physique et Spatiale, IFREMER for providing the data used in this study.

Funding. The third author was supported in part by the DAMOCLES-COST-ACTION on compound events, the French national program (FRAISE-LEFE/INSU and 80 PRIME CNRS-INSU), and the European H2020 XAIDA (Grant agreement ID: 101003469. The authors also acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project), the ANR-Melody (ANR-19-CE46-0011).

SUPPLEMENTARY MATERIAL

Supplement A: R code and data
This supplement contains the R code and the data used in this study.

Supplement B: Bootstrap simulations
This supplement file contains two movies showing the results of each $n_{\text{boot}} = 100$ bootstrap for both algorithms iteration by reproducing Figure 7 and Figure 8 of the main paper.

REFERENCES

ACCENSI, M. and MAISONDIEU, C. (2015). HOMERE. `https://doi.org/10.12770/cf47e08d-1455-4254-955e-d66225c9dc90`.

ARDHUIN, F. and ACCENSI, M. (2014). IOWAGA. `https://sextant.ifremer.fr/record/c87f6f24-63b4-46ec-b40e-f185a61dc672/`.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. L. (2004). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley.

BERTIN, X., BRUNEAU, N., BREILH, J.-F., FORTUNATO, A. B. and KARPYTCHEV, M. (2012). Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay. *Ocean Modelling* **42** 16–30. https://doi.org/10.1016/j.ocemod.2011.11.001

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics* **9** 1196 – 1217. https://doi.org/10.1214/aos/1176345637

CAIRES, S. and STERL, A. (2005). 100-Year Return Value Estimates for Ocean Wind Speed and Significant Wave Height from the ERA-40 Data. *Journal of Climate* **18** 1032–1048. https://doi.org/10.1175/JCLI-3312.1

CASAS-PRAT, M., WANG, X. L. and SIERRA, J. P. (2014). A physical-based statistical method for modeling ocean wave heights. *Ocean Modelling* **73** 59–75. https://doi.org/10.1016/j.ocemod.2013.10.008

CASTILLO, E. and SARABIA, J. M. (1992). Engineering Analysis of Extreme Value Data: Selection of Models. *Journal of Waterway, Port, Coastal, and Ocean Engineering* **118** 129–146. https://doi.org/10.1061/(ASCE)0733-950X(1992)118:2(129)

CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005). Generalized Additive Modelling of Sample Extremes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **54** 207–222.

CLEVELAND, W. S. and DEVLIN, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* **83** 596–610. https://doi.org/10.1080/01621459.1988.10478639

COLE, T. J. and GREEN, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* **11** 1305–1319. https://doi.org/10.1002/sim.4780111005

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer-Verlag London. https://doi.org/10.1007/978-1-4471-3675-0

COLES, S., HEFFERNAN, J. E. and TAWN, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes* **2** 339–365. https://doi.org/10.1023/A:1009963131610

COLLINS, M., SUTHERLAND, M., BOUWER, L., CHEONG, S. M., FRÖLICHER, T., COMBES, H. J. D., ROXY, M. K., LOSADA, I., MCINNES, K., RATTER, B., RIVERA-ARRIAGA, E., SUSANTO, R. D., SWINGEDOUW, D. and TIBIG, L. (2019). Extremes, Abrupt Changes and Managing Risk. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H. O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama and N. M. Weyer, eds.) Cambridge University Press, Cambridge, UK and New York, NY, USA.

DE CARVALHO, M., PEREIRA, S., PEREIRA, P. and DE ZEA BERMUDEZ, P. (2022). An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails. *Journal of Agricultural, Biological and Environmental Statistics*. https://doi.org/10.1007/s13253-021-00469-9

DE LEO, F., BESIO, G., BRIGANTI, R. and VANEM, E. (2021). Non-stationary extreme value analysis of sea states based on linear trends. Analysis of annual maxima series of significant wave height and peak period in the Mediterranean Sea. *Coastal Engineering* **167** 103896. https://doi.org/10.1016/j.coastaleng.2021.103896

EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7** 1 – 26. https://doi.org/10.1214/aos/1176344552

EWANS, K. and JONATHAN, P. (2008). The Effect of Directionality on Northern North Sea Extreme Wave Design Criteria. *Journal of Offshore Mechanics and Arctic Engineering* **130**.

FELD, G., RANDELL, D., WU, Y., EWANS, K. and JONATHAN, P. (2014). Estimation of Storm Peak and Intra-Storm Directional-Seasonal Design Conditions in the North Sea. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE* **4**. https://doi.org/10.1115/OMAE2014-23157

GENOVESE, E. and PRZYLUSKI, V. (2013). Storm surge disaster risk management: the Xynthia case study in France. *Journal of Risk Research* **16** 825–841. https://doi.org/10.1080/13669877.2012.737826

GUMBEL, E. J. (1961). Bivariate Logistic Distributions. *Journal of the American Statistical Association* **56** 335–349.

HARUNA, A., BLANCHET, J. and FAVRE, A. C. (2022). Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation. *Hydrology and Earth System Sciences* **26** 2797–2811. https://doi.org/10.5194/hess-26-2797-2022

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

HEREDIA-ZAVONI, E. and MONTES-ITURRIZAGA, R. (2019). Modeling directional environmental contours using three dimensional vine copulas. *Ocean Engineering* **187** 106102. https://doi.org/10.1016/j.oceaneng.2019.06.007

HOLTHUIJSEN, L. H. (2007). *Waves in Oceanic and Coastal Waters*. Cambridge university press.

JONATHAN, P. and EWANS, K. (2007). The effect of directionality on extreme wave design criteria. *Ocean Engineering* **34** 1977-1994. https://doi.org/10.1016/j.oceaneng.2007.03.003

JONATHAN, P. and EWANS, K. (2013). Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering* **62** 91–109. https://doi.org/10.1016/j.oceaneng.2013.01.004

JONATHAN, P., EWANS, K. and RANDELL, D. (2014). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics* **25** 172–188. https://doi.org/10.1002/env.2262

KIRILIOUK, A., ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2019). Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions. *Technometrics* **61** 123–135. https://doi.org/10.1080/00401706.2018.1462738

KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. https://doi.org/10.1093/biomet/81.4.673

LE CARRER, N. (2022). egpd4gamlss. `https://github.com/noemielc/egpd4gamlss`.

LEGRAND, J., AILLIOT, P., NAVEAU, P. and RAILLARD, N. (2023). Supplement to "Joint stochastic simulation of extreme coastal and offshore significant wave heights.". https://doi.org/[provided by typesetter]

MARCON, G., NAVEAU, P. and PADOAN, S. (2017). A semi-parametric stochastic generator for bivariate extreme events. *Stat* **6** 184–201. https://doi.org/10.1002/sta4.145

MARSHALL, A. W. and OLKIN, I. (1967). A Multivariate Exponential Distribution. *Journal of the American Statistical Association* **62** 30–44. https://doi.org/10.1080/01621459.1967.10482885

MICHEL, R. (2006). Simulation and estimation in multivariate generalized Pareto models, PhD thesis, Universität Würzburg.

MOUSLIM, H., BABARIT, A., CLÉMENT, A. and BORGARINO, B. (2009). Development of the French Wave Energy Test Site SEM-REV. In *Proceedings of the 8th European wave and tidal energy conference* 31 – 35.

MÉNDEZ, F. J., MENÉNDEZ, M., LUCEÑO, A., MEDINA, R. and GRAHAM, N. E. (2008). Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering* **35** 131–138. https://doi.org/10.1016/j.oceaneng.2007.07.012

NAVEAU, P., HUSER, R., RIBEREAU, P. and HANNART, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52** 2753–2769. https://doi.org/10.1002/2015WR018552

NICOLAE LERMA, A., BULTEAU, T., LECACHEUX, S. and IDIER, D. (2015). Spatial variability of extreme wave height along the Atlantic and channel French coast. *Ocean Engineering* **97** 175–185. https://doi.org/10.1016/j.oceaneng.2015.01.015

NORTHROP, P. J. and JONATHAN, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics* **22** 799-809. https://doi.org/10.1002/env.1106

PAPASTATHOPOULOS, I. and TAWN, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference* **143** 131–143. https://doi.org/10.1016/j.jspi.2012.07.001

RIVOIRE, P., MARTIUS, O. and NAVEAU, P. (2021). A Comparison of Moderate and Extreme ERA-5 Daily Precipitation With Two Observational Data Sets. *Earth and Space Science* **8** e2020EA001633. https://doi.org/10.1029/2020EA001633

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis* **165** 117–131. https://doi.org/10.1016/j.jmva.2017.12.003

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917 – 930. https://doi.org/10.3150/bj/1161614952

ROSS, E., RANDELL, D., EWANS, K., FELD, G. and JONATHAN, P. (2017). Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. *Ocean Engineering* **142** 315–328.

SENEVIRATNE, S. I., ZHANG, X., ADNAN, M., BADI, D. C. W., DI LUCA, A., GHOSH, S., ISKANDAR, I., KOSSIN, J., LEWIS, S., OTTO, F., PINTO, I., SATOH, M., VICENTE-SERRANO, S. M. and WEHNER, M. (2021). Weather and Climate Extreme Events in a Changing Climate. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou, eds.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

SHOOTER, R., ROSS, E., TAWN, J. A. and JONATHAN, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics* **30**.

SHOOTER, R., TAWN, J., ROSS, E. and JONATHAN, P. (2021a). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes* **24**. https://doi.org/10.1007/s10687-020-00389-w

SHOOTER, R., ROSS, E., RIBAL, A., YOUNG, I. R. and JONATHAN, P. (2021b). Spatial dependence of extreme seas in the North East Atlantic from satellite altimeter measurements. *Environmetrics* **32** e2674. https://doi.org/10.1002/env.2674

SHOOTER, R., ROSS, E., RIBAL, A., YOUNG, I. R. and JONATHAN, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering* **247** 110647. https://doi.org/10.1016/j.oceaneng.2022.110647

STASINOPOULOS, M., RIGBY, B. and AKANTZILIOTOU, C. (2008). *Instructions on how to use the gamlss package in R Second Edition*.

TENCALIEC, P., FAVRE, A. C., NAVEAU, P., PRIEUR, C. and NICOLET, G. (2019). Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics* **31** e2582. e2582 env.2582. https://doi.org/10.1002/env.2582

TENDIJCK, S., EASTOE, E. F., TAWN, J. A., RANDELL, D. and JONATHAN, P. (2021). Modeling the Extremes of Bivariate Mixture Distributions With Application to Oceanographic Data. *Journal of the American Statistical Association* **0** 1–12. https://doi.org/10.1080/01621459.2021.1996379

TOWE, R., EASTOE, E. F., TAWN, J. A. and JONATHAN, P. (2017). Statistical downscaling for future extreme wave heights in the North Sea. *The Annals of Applied Statistics* **11** 2375–2403. https://doi.org/10.1214/17-AOAS1084

VANEM, E. and FAZERES-FERRADOSA, T. (2022). A truncated, translated Weibull distribution for shallow water sea states. *Coastal Engineering* **172** 104077. https://doi.org/10.1016/j.coastaleng.2021.104077