

Modelling the extremes of bivariate mixture distributions with application to oceanographic data

Stan Tendijck

Department of Mathematics and Statistics, Lancaster University
LA1 4YW, United Kingdom

and

Emma Eastoe

Department of Mathematics and Statistics, Lancaster University
and

Jonathan Tawn

Department of Mathematics and Statistics, Lancaster University
and

David Randell

Shell Global Solutions International B.V., Amsterdam, Netherlands
and

Philip Jonathan

Shell Research Limited, London, United Kingdom

September 3, 2021

Abstract

There currently exist a variety of statistical methods for modelling bivariate extremes. However, when the dependence between variables is driven by more than one

latent process, these methods are likely to fail to give reliable inferences. We consider situations in which the observed dependence at extreme levels is a mixture of a possibly unknown number of much simpler bivariate distributions. For such structures we demonstrate the limitations of existing methods and propose two new methods: an extension of the Heffernan-Tawn conditional extreme value model to allow for mixtures and an extremal quantile-regression approach. The two methods are examined in a simulation study and then applied to oceanographic data. Finally, we discuss extensions including a subasymptotic version of the proposed model, which has the potential to give more efficient results by incorporating data that are less extreme. Both new methods outperform existing approaches when mixtures are present.

Keywords: conditional extremes, multivariate extremes, offshore wave extremes, mixture distributions, quantile-regression

1 Introduction

The dependence between multivariate response variables is often driven by their co-dependence on one or more underlying driving processes. For example, the strength of the relationship between wind speed and wave height depends on a combination of wind direction, land shadows, atmospheric pressure systems and their underlying driving processes. Often these driving processes are either unknown or unobserved, or both. If the driving processes are known, the interaction between them is likely to be highly non-linear and without specific knowledge of the physical processes that drive the response, the dependence structure can look highly complex. Consequently, building parsimonious statistical models that capture well the complex, multi-layered data generating mechanisms is difficult. One possible form of a complex dependence structure occurs when the joint distribution is a mixture of two or more simpler jointly distributed random variables. Such situations are the focus of this paper since the limiting assumptions of standard multivariate extreme value theory do not hold at non-asymptotic levels for problems of this type.

We are motivated by an oceanographic application with complex extremal dependence structures. In the design of offshore facilities it is crucial - both for safety and reliability reasons - to protect against the most extreme storms. Hence, it is necessary that the extremal dependence structures between the multiple physical hazards that may co-occur during a storm are well understood. We consider a synthetic response variable (Ross et al., 2020) that is a function of significant wave height H_S and wave period T_2 , illustrative of the response of floating offshore structures to wave loading. In particular, these synthetic response variables increase with increasing H_S and for T_2 approaching a resonance frequency. The definitions of H_S and T_2 are given in Holthuijsen (2010).

We consider (T_2, H_S) data from a location in the northern North Sea, see Figure 1. There are two different types of waves present - swell waves (relatively large wave period compared to significant wave height) and wind-generated waves (relatively small wave period compared to significant wave height). These correspond to waves generated on a large spatio-temporal scale and local wind-generated waves, respectively. Without expert knowledge, we are unable to identify which wave type each observation corresponds to. We propose and test two novel methods to make inference on data for which the dependence structure arises in this way, and for which it is either non-trivial or impossible to identify the process from which each observation has been generated.

It is standard practice in multivariate extreme value analysis to assume that observations are drawn from a common distribution and to use limit theory as a basis for extrapolation to give estimates on the probability of rare events. Traditionally, separate classes of extreme value models have been developed for bivariate data that exhibit either asymptotic dependence or asymptotic independence, see Coles et al. (1999) for an overview. Two random variables X and Y are asymptotically dependent if the probability that they are both

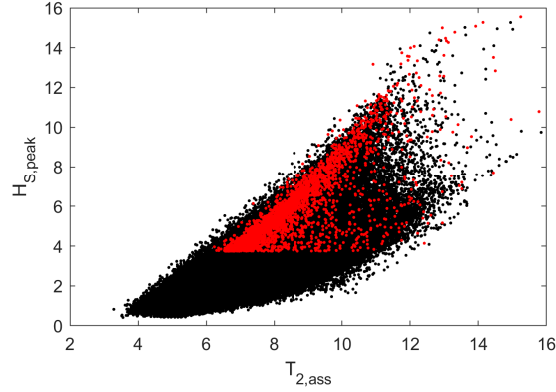


Figure 1: Significant wave height H_S and wave period T_2 from a northern North Sea location. Black dots: data for 1957 - 2018; red dots: corresponding storm peak data.

large is of the same magnitude as when one is large, i.e.,

$$\chi := \lim_{p \uparrow 1} \mathbb{P}(Y > F_Y^{-1}(p) | X > F_X^{-1}(p))$$

is such that $\chi > 0$, where F_X and F_Y denote the distribution functions of X and Y . If this limiting quantity $\chi = 0$, we say that the variables are asymptotically independent. In a multivariate setting, it is possible that some subsets of variables are asymptotically dependent and others are asymptotically independent but not completely independent, see for example Simpson et al. (2020). Many extreme value methods, for instance Coles and Tawn (1994), Joe (1994), Capéraà et al. (1997), Naveau et al. (2009), and Genest and Segers (2009) are based on the assumption of multivariate regular variation. This means that they cannot model the distribution well for parts where some variables are large and others are not. Heffernan and Tawn (2004) introduced the more flexible conditional extremes model which can handle this situation. In particular, in the bivariate case their model provides a description for $Y|X$ ($X|Y$) in the region where X (Y) is large. More precisely, when X and

Y are transformed marginally to follow standard Laplace distributions this model takes on the form of heteroscedastic regression, the parameters of which are estimated using only observations for which X (Y) is large, see e.g. Keef et al. (2013).

However, despite the flexibility of the Heffernan-Tawn model, the “vanilla” version does not provide good estimates when the data consists of non-trivial mixture structures such as that shown in Figure 1, in which we can see that significant wave height conditional on wave period grows in two different ways as wave period increases. Sometimes both variables are relatively large (wind-generated waves) at the same time, but at other times only significant wave period is high and wave height takes on more moderate values (swell waves). As will be seen in Section 2.2, this is an indication that the vanilla Heffernan-Tawn model is not likely to be suitable in practice.

Our proposed solution to this limitation is a mixture formulation of the Heffernan-Tawn model which assumes that, after a suitable transformation, $Y|X$ can be captured by a mixture of $K = 1, 2, 3, \dots$ regressions each associated with a different probability weighting. Whilst we take K to be unknown, it will not be treated as a parameter, rather we use model selection methods to select the most appropriate value for K , and do not account for its uncertainty subsequently. A second method comprising a novel quantile-regression-based mixture approach is also developed. This uses the same parametric form for the conditional quantiles as proposed in the mixture formulation of the Heffernan-Tawn model. It differs from the Heffernan-Tawn mixture formulation mainly in the sense that it is more flexible due to its semi-parametric nature. We find that for bivariate data both methods show similar performance in estimating probabilities of extreme sets, i.e., $\mathbb{P}((X, Y) \in A)$ where A is extreme in X or Y or both, with the uncertainty assessed through the use of a semi-parametric bootstrap. The main advantage of the quantile-regression approach is

that it leads to more stable fits, i.e., it gives more consistent results for small sample sizes. However, unlike the mixture formulation of the Heffernan-Tawn model, it does not extend naturally beyond bivariate data. We also discuss a subasymptotic version of the Heffernan-Tawn mixture model motivated by a theoretical example where the mixture probabilities vary with the level of extremity. However, we find that this model does not perform better in modelling the dependence structure of (T_2, H_S) than the Heffernan-Tawn mixture model.

It is worth noting that mixture models have previously been used for multivariate extremes. Boldi and Davison (2007) focus on a mixture of asymptotically dependent variables. Simpson et al. (2020), Chiapino et al. (2019), Engelke and Ivanovs (2020) investigate d -dimensional random variables and describe how to estimate on which of $2^d - 1$ subspaces the limiting spectral measure has mass. If we let $d = 2$ and condition on one variable being large, the resulting two subspaces correspond to asymptotic dependence and asymptotic independence. So, when one mixture component is asymptotically dependent and one asymptotically independent, there are strong similarities between the problem here and that considered by these authors, although only Simpson et al. (2020) consider the form of the asymptotic independence term in the mixture. However, none of the existing work considers the situation with two or more different levels of asymptotic independence.

The paper is organised as follows. In Section 2, the current conditional extremes modelling framework and its extension incorporating mixture structures are discussed. Section 3 introduces the two inference methods that exploit the framework from Section 2. A subasymptotic version of the conditional extremes mixture model is discussed in Section 4. Finally, the oceanographic application is presented in Section 5. To conserve space, a simulation study is found in the supplementary material. Code and data are published at <https://github.com/stantendijck/HTMixtureModel>.

2 Framework

2.1 Heffernan-Tawn model

Let (X, Y) be a real-valued bivariate random vector on standard Laplace margins. If (X, Y) do not follow such marginal distributions, transformation to standard Laplace margins is achieved using the probability integral transform (Keef et al., 2013). The Heffernan-Tawn (HT) model, from Heffernan and Tawn (2004), extrapolates the joint distribution of (X, Y) to the region of the sample space where either one, or both, of the variables is extreme. Without loss of generality, we present the model for the case in which X is extreme; the full bivariate model requires an equivalent definition for the case of Y being extreme.

The underlying principle for the model is that there exist parameters α and β with $|\alpha| \leq 1, \beta < 1$ such that the normalisation

$$Z_x := \frac{Y - \alpha X}{X^\beta} \Big| \{X = x\} \quad (1)$$

of the conditional random variable $Y | \{X = x\}$, converges in distribution to a non-degenerate random variable Z with distribution function $G(z)$ as $x \rightarrow \infty$. Under assumptions relating to convergence and existence of joint densities, Heffernan and Resnick (2007), Resnick and Zeber (2014), Wadsworth et al. (2017) show that this implies that in the limit the random variables Z and $X - u$ are independent conditional on $X > u$, i.e., for $x > 0$ and $z \in \mathbb{R}$

$$\lim_{u \rightarrow \infty} \mathbb{P} \left(\frac{Y - \alpha X}{X^\beta} \leq z, X - u > x \mid X > u \right) = G(z)e^{-x} \quad (2)$$

holds. To ensure identifiability of α and β , Keef et al. (2013) impose the condition that $\lim_{z \rightarrow \infty} G(z) = 1$, i.e., no mass of Z is at $\{+\infty\}$ but there can be mass at $\{-\infty\}$. They also impose additional constraints on α and β such that the implied distribution of Y is not inconsistent with its marginal distribution. We call these the Keef constraints.

Inference is performed by assuming that the limiting relation (2) holds above a finite threshold level u . Specifically,

$$Y|(X > u) \stackrel{\mathcal{D}}{=} \alpha X + X^\beta(\mu + \sigma \tilde{Z}), \quad (3)$$

where $\stackrel{\mathcal{D}}{=}$ represents equality in distribution, \tilde{Z} is the standardised residual random variable, i.e., $\tilde{Z} = (Z - \mu)/\sigma$, with μ the mean of Z and $\sigma > 0$ the standard deviation. The model parameters α and β are typically inferred via assuming that the distribution of Z is Gaussian. The four HT parameters $(\alpha, \beta, \mu, \sigma)$ can now be estimated using any method of statistical inference. Conditional on the estimated parameters, the distribution of the residual random variable is then estimated non-parametrically using the empirical distribution of Z_x when $x > u$. Finally, u is chosen as low as possible such that estimates for HT parameters are approximately unchanged at any higher threshold (an analogy of univariate threshold stability plots) whilst ensuring that X and Z are independent given $X > u$.

Simple interpretations of the parameters α and β exist: (i) $(\alpha, \beta) = (1, 0)$ implies asymptotic dependence between X and Y with $\chi = \int_0^\infty (1 - G(-t)) \exp(-t) dt$; (ii) $\alpha < 1$ implies asymptotic independence between X and Y ; (iii) $\beta > 0$ implies that the variability between X and Y grows as X grows; (iv) $0 < \alpha < 1$ (or $-1 < \alpha < 0$) implies positive (or negative) association between X and Y . In environmental applications, $\beta < 0$ is unrealistic, as this imposes that all the quantiles of Y conditional on X converge to the same value as X grows to infinity, so we restrict $0 \leq \beta < 1$.

We explore a variant of the Heffernan-Tawn model, which is more flexible than the HT model and in some theoretical examples improves the convergence rates to limit (2), see Lugin et al. (2021). Previously, we assumed that relation (3) holds for some finite level u . However, it is possible that for $u < \infty$, inference can be improved if we adjust the model

form. To that end, we redefine Z_x from equation (1) to

$$Z_x := \frac{Y - \alpha X - \gamma(X)}{X^\beta} | \{X = x\}. \quad (4)$$

By taking $\gamma : \mathbb{R} \rightarrow \mathbb{R}$ to be a function such that $\gamma(x) = o(x^\beta)$ for $\beta \geq 0$ as $x \rightarrow \infty$, this is equivalent to the HT model in the limit as $x \rightarrow \infty$. For reasons of parsimony, we take $\gamma(x) = \gamma \in \mathbb{R}$, and so introduce an intercept in the mean component of the HT regression model, which is identifiable if $\beta > 0$.

2.2 Extreme value distribution with asymmetric logistic dependence structure

We now give an example of a simple bivariate distribution for which the HT model is inadequate, i.e., it has limit $\lim_{z \rightarrow -\infty} G(z) \neq 0$. Here, different choices for α and β also lead to non-degenerate $G(z)$ but with $\lim_{z \rightarrow \infty} G(z) \neq 1$. The findings from this example motivate our developments in Section 2.3. Let (X_A, Y_A) denote a random variable on Laplace margins following a bivariate extreme value copula with an asymmetric logistic dependence structure (Tawn, 1988). This distribution has parameters $\theta_1, \theta_2 \in (0, 1)$, $\eta \in (0, 1)$, and distribution function

$$P(X_A \leq x, Y_A \leq y) = \exp \left\{ -\theta_1 t_x - \theta_2 t_y - \left[((1 - \theta_1) t_x)^{1/\eta} + ((1 - \theta_2) t_y)^{1/\eta} \right]^\eta \right\} \quad (5)$$

where $t_x := \log 2 - \log(2 - \exp(-x))$ for $x \geq 0$ and $t_x := \log 2 - x$ for $x < 0$, with t_y similarly defined.

In Figure 2, data simulated from this distribution, conditional on $X_A > 2$ with $\theta_1 = \theta_2 = \eta = 0.5$, show two arms centred on $y = x$ and $y = 0$ for different conditional quantiles. The heterogeneity in the two arms results in two possibly non-degenerate HT limits, each

of which fully captures the behaviour around one arm whilst treating the behaviour around the other as degenerate. The first limit models the upper arm, and has $(\alpha, \beta) = (1, 0)$ and non-degenerate $G(z)$

$$G(z) = \theta_1 + (1 - \theta_1) \cdot \left[1 + \left(\frac{1 - \theta_2}{1 - \theta_1} \right)^{1/\eta} \cdot \exp\left(-\frac{z}{\eta}\right) \right]^{\eta-1}, \quad z \in \mathbb{R}.$$

Note that $G(z)$ puts weight θ_1 on $\{-\infty\}$. The second limit corresponds to the lower arm with $(\alpha, \beta) = (0, 0)$ and $G(z) = \theta_1 e^{-tz}$ for $z \in \mathbb{R}$, which puts weight $1 - \theta_1$ on $\{+\infty\}$.

Applying the HT model to this joint distribution directly results in poor extrapolations because at any finite level the normalised random variable Z cannot capture the two different growth rates. Figure 2 (right) illustrates simulated data and implied conditional quantiles from the model we develop in Section 3 when fitted to the data shown in Figure 2 (left). The implied conditional quantiles capture the true behaviour of the top and bottom quantiles well (matching the true gradients of 1 and 0 respectively) but with a less clear distinction between the two mixture components than for the true process. This analysis may be improved by picking a different threshold.

2.3 Heffernan-Tawn mixture model

The Heffernan-Tawn model is applicable to the example from Section 2.2 with $\lim_{z \rightarrow -\infty} G(z) \neq 0$. However, no model is imposed for the part of the distribution that corresponds to G at $\{-\infty\}$. This may not be a problem if our interest lies in characterising combinations of maximum X with associated maximum Y but it might be that the other part of the distribution also corresponds to extreme scenarios. For example, in the case of H_S versus T_2 , there might be wave periods corresponding to one or more marine structural resonance frequencies which is of greater interest than the maximum wave period. Motivated by

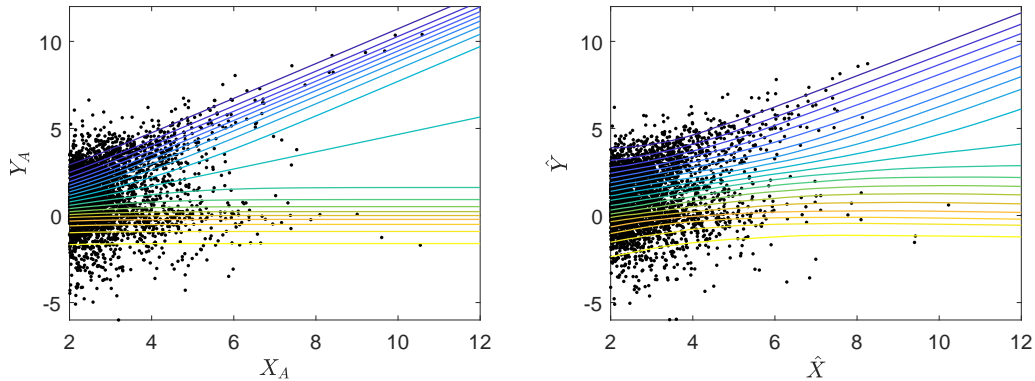


Figure 2: (Left) Data simulated from model (5), conditional on $X_A > 2$ with $\theta_1 = \theta_2 = \eta = 0.5$ including true conditional quantile functions; (Right) Data simulated from the inferred model (7), fitted to the data in the left plot, with an identical sample size, including the implied conditional quantile functions.

these considerations, we introduce the Heffernan-Tawn mixture (HTM) model to better characterise distributions such as the one discussed in Section 2.2, conditional on one of the variables being large. We allow for multiple parameter combinations such that all possible non-degenerate residual distributions are captured simultaneously.

Let $K \geq 1$ be an integer and let (X, Y) have standard Laplace margins such that the joint distribution is a mixture of K copulas. We assume that for $x > u$, where u is large,

$$F(x, y) = \sum_{k=1}^K p_k F_k(x, y), \quad (6)$$

where F is the cumulative distribution function of (X, Y) and for all k , F_k is a distribution function on Laplace margins, $p_k \in (0, 1]$, and $\sum_{k=1}^K p_k = 1$. We also assume that each F_k has a copula formulation such that the associated limit (2) holds for a single pair (α_k, β_k) with residual distribution G_k placing no mass at $\{-\infty, +\infty\}$. Thus the asymmetric logistic

copula (5) cannot be F_k . As we assume the distributional form (6) only for $x > u$, this condition holds for standard mixture distributions, but can apply for more complex models which, when in an extremal state (i.e., $X > u$), approximate to a mixture form.

Central to the HTM model is the assumption that the HT model, with the intercept extension proposed in equation (4) is applicable to F_k , with parameters $\gamma_k \in \mathbb{R}$, $\alpha_k \in [-1, 1]$, $0 \leq \beta_k < 1$, $\mu_k \in \mathbb{R}$ and $\sigma_k > 0$, for all k . Using the notation of equation (3), we define the K component HTM model as

$$Y|(X > u) \stackrel{D}{=} \gamma_k + \alpha_k X + X^{\beta_k}(\mu_k + \sigma_k \tilde{Z}_k), \quad \text{with probability } p_k, \quad k = 1, \dots, K, \quad (7)$$

for large u , where \tilde{Z}_k , which only exists with probability p_k , follows a non-degenerate distribution for each k . Moreover, we assert that \tilde{Z}_k is independent of both X and $\tilde{Z}_{\tilde{k}}$ for $k \neq \tilde{k}$. We further impose that $\mathbb{E}[\tilde{Z}_k] = 0$ and $\text{Var}(\tilde{Z}_k) = 1$, which implies that the distribution function of $Z_k := \mu_k + \sigma_k \tilde{Z}_k$ does not put mass on $\{\pm\infty\}$. For identifiability reasons, $\alpha_k > \alpha_{k'}$ for all pairs $k > k'$. A model formulation which allows $\alpha_k = \alpha_{k'}$ ($k \neq k'$) is discussed in the supplementary material. We impose the Keef constraints on all pairs (α_k, β_k) separately. For $k = 1, \dots, K$, α_k and β_k are such that for $x > 0$ and $z \in \mathbb{R}$

$$\lim_{u \rightarrow \infty} \mathbb{P} \left(\frac{Y - \gamma_k - \alpha_k X}{X^{\beta_k}} \leq z, \quad X - u > x \mid X > u \right) = G_k(z)e^{-x},$$

where $G_k(z) = \sum_{i=1}^{k-1} p_i + p_k \cdot H_k(z)$ and H_k is the distribution function of Z_k , so G_k has mass $\sum_{i=1}^{k-1} p_i$ at $\{-\infty\}$ and $\sum_{i=k+1}^K p_i$ at $\{+\infty\}$.

The Heffernan-Tawn model is a special case of this model with $K = 1$. Distribution (5) corresponds to the case $K = 2$ with $(\alpha_1, \beta_1, p_1, \alpha_2, \beta_2, p_2) = (1, 0, 1 - \theta_1, 0, 0, \theta_1)$. When $K = 2$ other classes of distributions that fall into this mixture formulation with $(\alpha_1, \beta_1, \alpha_2, \beta_2) = (1, 0, 0, 0)$ are bivariate max-stable distributions with some mass of the spectral measure

on the boundaries of the simplex, and max-linear models (De Haan and Ferreira, 2007, Chapter 6) where different innovation variables control each marginal variable.

2.4 The quantile-regression model

Quantile-regression is used to model the inverse of the distribution function of Y conditional on $X = x$, i.e., to find $q_\tau(x) := F_{Y|X=x}^{-1}(\tau)$ for some non-exceedance probability $\tau \in (0, 1)$ over a range of x . It is common practice to define a parametric class of functions

$$\{q_\tau : \mathbb{R} \rightarrow \mathbb{R} : q_\tau(x) := q(x|\boldsymbol{\omega}_\tau), \boldsymbol{\omega}_\tau = (\omega_{\tau 1}, \dots, \omega_{\tau p}) \in \Omega \subseteq \mathbb{R}^p\}$$

with $p \geq 1$ parameters in the parameter space Ω . We are interested in a parametric form for $q(\cdot | \boldsymbol{\omega}_\tau)$, that is motivated by the HT mixture model (7),

$$q(x|\boldsymbol{\omega}_\tau) = \gamma(\tau) + \alpha(\tau)x + \zeta_\tau x^{\beta(\tau)}, \quad (8)$$

where $\boldsymbol{\omega}_\tau = (\alpha(\tau), \beta(\tau), \gamma(\tau), \zeta_\tau)$ such that $(\alpha(\tau), \beta(\tau)) \in [-1, 1] \times [0, 1)$ satisfy the Keef et al. (2013) constraints, $\gamma(\tau), \zeta_\tau \in \mathbb{R}$, and ζ_τ is a one-to-one function of quantiles of all residual distributions Z_k , $k = 1, \dots, K$.

Let (X, Y) be a bivariate random variable following model (7) and define Y_k , $k = 1, \dots, K$, as the random variable representing the k th mixture component of the model, i.e., $Y_k|(X > u) \stackrel{D}{=} \gamma_k + \alpha_k X + X^\beta(\mu_k + \sigma_k \tilde{Z}_k)$. Further, let $Y_k \perp Y_{k'}$ for $k \neq k'$, then

$$\lim_{u \rightarrow \infty} \mathbb{P}(Y_1 < Y_2 < \dots < Y_K | X > u) = 1, \quad (9)$$

i.e., the mixture components separate completely in the limit, and as $u \rightarrow \infty$, we get that the parameter functions $\gamma(\tau)$, $\alpha(\tau)$ and $\beta(\tau)$ in equation (8) are piecewise constant in τ . More precisely, $\gamma(\tau) = \gamma_k$ for $\sum_{i=1}^{k-1} p_i < \tau \leq \sum_{i=1}^k p_i$; $\alpha(\tau)$ and $\beta(\tau)$ behave similarly.

The parameter function ζ_τ is an increasing function of τ within each interval of τ where $\gamma(\tau)$, $\alpha(\tau)$ and $\beta(\tau)$ remain constant.

To consider multiple conditional quantiles jointly, let $m \in \mathbb{N}$ and $0 < \tau_1 < \dots < \tau_m < 1$ be m non-exceedance probabilities. We assume that the vector conditional quantile function $(q_{\tau_1}(x), \dots, q_{\tau_m}(x))$ belongs to the following parametric class of functions

$$\{(q_{\tau_1}, \dots, q_{\tau_m}) : \mathbb{R} \rightarrow \mathbb{R}^m : q_{\tau_i}(x) = q(x|\boldsymbol{\omega}_{\tau_i}) \text{ for } i = 1, \dots, m, (\boldsymbol{\omega}_{\tau_1}, \dots, \boldsymbol{\omega}_{\tau_m}) \in \Omega_m \subseteq \mathbb{R}^{pm}\},$$

where Ω_m is the parameter space, not necessarily equal to the Cartesian product Ω^m . In particular, we consider models with $\boldsymbol{\omega}_\tau = (\boldsymbol{\varphi}, \zeta_\tau)$ where $\boldsymbol{\varphi}$ is common across $\boldsymbol{\omega}_{\tau'}$ for all τ' and ζ_τ is specific to a particular τ . As an illustration, consider model (7) with $K = 1$, then $\boldsymbol{\varphi} = (\gamma_1, \alpha_1, \beta_1)$ and $\zeta_\tau = F_{Z_1}^{-1}(\tau)$ where F_{Z_1} is the distribution function of the residuals Z_1 .

3 Inference

3.1 The Heffernan-Tawn mixture model

We focus our discussion on fitting the HTM model given a known number of mixture components K . Inference for K and accounting for its uncertainty are generic in mixture modelling and a reversible jump MCMC mixture model could be applied, see Richardson and Green (1997). Instead, we adapt a simple pseudo-Bayesian inference approach. As the focus of this paper is on the extremal dependence structure, we first estimate the marginal distributions and ignore their inference uncertainty subsequently. We describe a heuristic to select K , and also examine the sensitivity of inferences to this selection. We then proceed to make inferences using standard Bayesian methods. Consequently, unlike in reversible jump MCMC mixture models, we do not account for uncertainty in K in

subsequent inference. Details of outline code for inference for model parameters and for the probabilities of extreme events are given in the supplementary material.

For ease of notation, we define $\boldsymbol{\theta}^{(k)} = (\gamma_k, \alpha_k, \beta_k, \mu_k, \sigma_k, p_k)$ to be the vector of parameters of component $1 \leq k \leq K$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)})$ the vector containing all parameters of the model. Henceforth, we assume that we have data $\{(x_i, y_i) : x_i > u\}$ of size n generated by model (7) for some $K \ll n$. Moreover, we assume that $\alpha_i < \alpha_j$ for $i < j$.

In mixture modelling, it is common to introduce a latent random variable $J \in \{1, 2, \dots, K\}$ denoting the mixture component associated with the random pair (X, Y) . In performing Bayesian inference, we need to calculate the likelihood of parameters given the data. So, we need to make a computationally convenient assumption on the HTM residual distributions Z_k , $k = 1, \dots, K$. We follow the same arguments as Heffernan and Tawn (2004), and assume for parameter estimation purposes that all residual distributions are Gaussian, i.e., $\tilde{Z}_k \sim \mathcal{N}(0, 1)$ for all $k = 1, \dots, K$. This technique based on a (false) assumption is equivalent to estimating equations methods, which are known to yield asymptotically consistent estimators (Zeger and Liang, 1986). We stress that the Gaussian assumption is discarded once $\boldsymbol{\theta}$ is estimated. So, we get for $k = 1, \dots, K$

$$\mathbb{P}(J = k | X = x, Y = y, \boldsymbol{\theta}) := P_k(x, y | \boldsymbol{\theta}^{(k)}) / \sum_{\kappa=1}^K P_\kappa(x, y | \boldsymbol{\theta}^{(\kappa)}), \quad (10)$$

where $J = k$ is the event that random variable (X, Y) is drawn from component k , where

$$P_k(x, y | \boldsymbol{\theta}^{(k)}) = \frac{1}{\sigma_k x^{\beta_k} \sqrt{2\pi}} \exp \left\{ -\frac{(y - \gamma_k - \alpha_k x - \mu_k x^{\beta_k})^2}{2\sigma_k^2 x^{2\beta_k}} \right\}.$$

We can now simulate index $J = j_i$ for each observation (x_i, y_i) using equation (10), and calculate the augmented log-likelihood $l^{(a)}$ via

$$l^{(a)}(\boldsymbol{\theta} | \{(x_i, y_i, j_i)\}_{i=1}^n) = \sum_{k=1}^K l_k(\boldsymbol{\theta}^{(k)} | \{(x_i, y_i, j_i)\}_{i=1}^n),$$

where the log-likelihood l_k for data from mixture component k is given by

$$l_k \left(\boldsymbol{\theta}^{(k)} | \{(x_i, y_i, j_i)\}_{i=1}^n \right) = \sum_{i=1}^n \mathbb{1}\{j_i = k\} \cdot \log P_k \left(x_i, y_i | \boldsymbol{\theta}^{(k)} \right),$$

if $\sigma_k > 0$ and (α_k, β_k) satisfy the Keef constraints (Keef et al., 2013), otherwise $l_k = -\infty$. Since we now have an expression for the likelihood, we can infer the parameters of the model. In particular, we use an adaptive MCMC algorithm similar to Roberts and Rosenthal (2009). In this algorithm, we do not estimate the mixture probability parameters p_k since estimates can be inferred from the estimates of the remaining parameters. A uniform prior over the whole parameter space is used to illustrate performance with no expert knowledge. Other prior choices may be more appropriate for some applications, see Section 3.2.

For each draw $\hat{\boldsymbol{\theta}}$ from the MCMC chain, we simulate $\{j_i\}_{i=1}^n$ each with probability (10), and define the residuals of the k th component $\hat{z}_{ki} := (y_i - \hat{\gamma}_k - \hat{\alpha}_k x_i) / x_i^{\hat{\beta}_k}$ for all $1 \leq i \leq n$, $1 \leq k \leq K$ with $j_i = k$. The distribution function $H_k(z)$, defined in Section 2.3, is now estimated using the empirical distribution function of $\{\hat{z}_{ki} : 1 \leq i \leq n \text{ with } j_i = k\}$ and $\hat{p}_k := n_k/n$, where n_k denotes the number of observations allocated to component k . The supplementary material details estimation of probabilities of extreme sets for this model.

Selecting an optimal value of K is challenging. In simulation studies discussed in the supplementary material, rather than attempt to fix K , we explore the sensitivity of our inference to the value of K . If an estimate of K were required, we suggest the following two-step heuristic: (i) fit the model as described above with $K = 1$; (ii) The number of modes of the residual distribution conditional on $X > v$ for some $v > u$ is our heuristic estimator for K . We use $v > u$ since the mixture components might not yet have separated at u . Simulations show that when the α_i s are relatively distinct, this technique is reasonable.

3.2 Incorporating prior probability on asymptotic dependence

Placing a uniform prior over the parameter space (Section 3.1) implies that the posterior does not put positive weight on the class of models corresponding to asymptotic dependence (i.e., $\alpha_K = 1$). This is because the subset of the parameter space which corresponds to asymptotic dependence is a null set with respect to the uniform prior. Hence, this procedure consistently under-estimates the risk of extreme events occurring together, see the simulation study in the supplementary material and discussion in Coles and Pauli (2002).

Here, we show how to sample from the posterior distribution of the parameters of the Heffernan-Tawn mixture model using a prior which puts mass on both asymptotic dependence and asymptotic independence. We will not discuss how to calculate the likelihood given a set of parameters as this is similar to before. Instead, we focus on making good MCMC proposals and calculating the MCMC acceptance probability when the prior puts a positive mass on the event $\{\alpha_K = 1\}$. For brevity, we consider model (7) with $K = 1$.

We define the priors of the parameters γ_1 , β_1 , μ_1 and σ_1 to be the (improper) uniform distribution on the parameter space with independent components. Let δ be the Dirac delta function, then the density function of the prior on α_1 is

$$f_\omega(x) = \omega \cdot \delta(1 - x) + \frac{1 - \omega}{2} \cdot \mathbb{1}\{-1 \leq x < 1\},$$

i.e., the prior is a mixture which puts weight $\omega \in [0, 1]$ on $\{\alpha_1 = 1\}$ and weight $(1 - \omega)$ on a uniform $(-1, 1)$, similar to that in Coles and Pauli (2002).

We choose a Metropolis-Hastings Gaussian random walk update for the parameters γ_1 , β_1 , μ_1 and σ_1 with some fixed standard deviation h . We specify the proposal distribution for α_1 at iteration t with current value $\alpha_1^{(t)}$ to be given by $\min\{1, \alpha_1^{(t)} + hZ\}$ for standard Gaussian distribution Z such that there is a positive probability of proposing the asymptotic

dependent model. The proposal density $g(\cdot|\alpha_1^{(t)})$ is thus given by

$$g(x|\alpha_1^{(t)}) := \left(1 - \Phi\left(\frac{1 - \alpha_1^{(t)}}{h}\right)\right) \cdot \delta(1 - x) + \frac{1}{h}\varphi\left(\frac{x - \alpha_1^{(t)}}{h}\right)$$

for $x \in \mathbb{R}$, where φ and Φ are the density and distribution function, respectively, of a standard Gaussian. The Metropolis-Hastings acceptance ratio α_{MH} given a proposal $\boldsymbol{\theta}^{(t+1)}$ can now be derived using standard techniques, see the supplementary material.

3.3 Inference for the quantile-regression model

We consider two data generating frameworks, namely models (3) and (7), and we show how to use quantile-regression techniques to infer the parameters without distributional assumptions. In Section 2.4, we parameterised the conditional quantile function q_τ via the parameter vector $\boldsymbol{\omega}_\tau$. First consider a single value of τ , then $\boldsymbol{\omega}_\tau$ may be inferred as

$$\hat{\boldsymbol{\omega}}_\tau := \operatorname{argmin}_{\boldsymbol{\omega}_\tau \in \Omega} \left\{ \sum_{i=1}^n \rho_\tau(y_i - q(x_i|\boldsymbol{\omega}_\tau)) \right\}, \quad (11)$$

where the check function $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}$ is defined as $\rho_\tau(z) = z(\tau - \mathbb{1}\{z < 0\})$, see Koenker and Hallock (2001) and Koenker (2005). The check function is locally linear hence the estimator is robust to outliers.

Now let $m \in \mathbb{N}$ and $0 < \tau_1 < \dots < \tau_m < 1$. We infer conditional quantile functions q_{τ_j} , $j = 1, \dots, m$, jointly using

$$\hat{\boldsymbol{\omega}} = \operatorname{argmin}_{(\boldsymbol{\omega}_{\tau_1}, \dots, \boldsymbol{\omega}_{\tau_m}) \in \Omega_m} \sum_{j=1}^m c_j \sum_{i=1}^n \rho_{\tau_j}(y_i - q(x_i | \boldsymbol{\omega}_{\tau_j})) \quad (12)$$

for weights $c_j > 0$ specified by the user, see Bondell et al. (2010). We apply the methodology to equidistant τ_j on the range $[\tau_1, \tau_m] \subset (0, 1)$, for which it is natural to let $c_j := 1$ for all

$j = 1, \dots, m$. If τ_j are not equidistant, then the choice for the weights c_j can be adjusted to reflect this. If the parameter space Ω_m was equal to the Cartesian product of the marginal parameter spaces Ω^m , then the joint estimation procedure would be equivalent to applying equation (11) separately for each quantile. We only consider models with $\omega_\tau = (\varphi, \zeta_\tau)$ where φ is common across $\omega_{\tau'}$ for all τ' and ζ_τ is specific to τ .

Now we have jointly estimated conditional quantiles $q_{\tau_1}(x), \dots, q_{\tau_m}(x)$ for fixed quantile levels τ_i for $i = 1, \dots, m$, we give an estimator for $q_\tau(x)$ for any $\tau \in (0, 1) \setminus \{\tau_1, \dots, \tau_m\}$. To do this, we only need to estimate ζ_τ since $\hat{\varphi}$ is already available. We estimate ζ_τ by

$$\hat{\zeta}_\tau := \operatorname{argmin}_{\zeta \in \mathbb{R}} \left\{ \sum_{i=1}^n \rho_\tau(y_i - q(x_i | (\hat{\varphi}, \zeta))) : (\hat{\varphi}, \zeta) \in \Omega \right\} \quad (13)$$

and so $\hat{q}_\tau(x) := q(x | (\hat{\varphi}, \hat{\zeta}_\tau))$ for $q_\tau(x)$.

The above framework requires adjustment for inferring mixture model (7) when the mixture probabilities are unknown. We discuss this for $K = 2$ but it generalises for general K . For $\tau_1 < \dots < \tau_m$, we define m_0 as the index for which $\tau_{m_0} < p_1 < \tau_{m_0+1}$, where p_1 is the mixture probability corresponding to the first component, and separate our presentation into cases where m_0 is known (falsely) and unknown.

If m_0 is known, the parameters $(\omega_{\tau_1}, \dots, \omega_{\tau_m})$ of the quantile-regression model are:

$$\omega_{\tau_j} = \omega_{\tau_j}(m_0) = \begin{cases} (\gamma_1, \alpha_1, \beta_1, \zeta_{\tau_j}^1) & \text{if } j \leq m_0, \\ (\gamma_2, \alpha_2, \beta_2, \zeta_{\tau_j}^2) & \text{if } j > m_0, \end{cases} \quad (14)$$

where the parameters $\zeta_{\tau_j}^k$ increase over j for $k = 1, 2$. The parametric form (8) is used for $q(x | \omega_\tau)$. To improve stability of parameter estimates, we constrain the medians of Z_1 and Z_2 to be equal to 0. To impose this constraint, we fix $\zeta_{\tau_j}^1 = 0$ if $\tau_j = \hat{P}_1/2$ and $\zeta_{\tau_j}^2 = 0$ if $\tau_j = (1 + \hat{P}_1)/2$, where $\hat{P}_1 = (\tau_{m_0} + \tau_{m_0+1})/2$ is a crude estimator for p_1 .

When m_0 is unknown, its natural estimator \hat{m}_0 minimizes (12) over m_0 , i.e.,

$$\hat{m}_0 := \operatorname{argmin}_{k=1, \dots, m-1} \left\{ \min_{(\omega_{\tau_1(k)}, \dots, \omega_{\tau_m(k)}) \in \Omega_m} \sum_{j=1}^m c_j \sum_{i=1}^n \rho_{\tau_j}(y_i - q(x_i | \omega_{\tau_j(k)})) \right\},$$

where $\omega_{\tau_j}(k)$ is as in equation (14). The optimisation is over a set of $m-1$ possibilities with a non-convex objective function. The estimated parameters of the model for \hat{m}_0 are denoted by $\hat{\omega}_{\hat{m}_0} = (\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1, \hat{\gamma}_2, \hat{\alpha}_2, \hat{\beta}_2, \hat{\zeta}_{\tau_1}^1, \dots, \hat{\zeta}_{\tau_{\hat{m}_0}}^1, \hat{\zeta}_{\tau_{\hat{m}_0+1}}^2, \dots, \hat{\zeta}_{\tau_m}^2)$, where each of $(\hat{\zeta}_{\tau_1}^1, \dots, \hat{\zeta}_{\tau_{\hat{m}_0}}^1)$ and $(\hat{\zeta}_{\tau_{\hat{m}_0+1}}^2, \dots, \hat{\zeta}_{\tau_m}^2)$ are increasing. Estimation of p_1 and probabilities of extreme sets under this model are discussed in the supplementary material. As specified by our approach, inferred conditional quantile functions within a mixture component are increasing in τ , since ζ_{τ_j} have to be ordered. However, this is not necessarily true across different components.

In fitting the model we have chosen not to impose the condition that the inferred conditional quantile functions are increasing in τ for each x for the following reasons. Simulation studies showed that conditional quantile functions usually do not cross if there is little to no overlap of the different components. When they do cross, the intersection is near the threshold u and not at large values. Hence, if non-crossing is of interest, this can be achieved by increasing the threshold sufficiently. The trade-off is that less data will be used for inference and the variance of the estimates will increase. Moreover, if there exists a significant amount of overlap, then requiring quantile-functions to not cross whilst also requiring that they take on the form in equation (8), can result in arbitrarily large biases.

4 A subsasymptotic conditional mixture model

The extension of the HT model developed in Section 2.3 assumes that mixture probabilities on the conditional distribution $Y|X$ are constant with respect to X , when X takes an

extreme value. This model adds significant flexibility over the HT model, however, it does not include the case where the mixture probabilities $p_k(x)$, $k = 1, \dots, K$ are varying with level x . In particular, if there exists a component k with $p_k(x) \rightarrow 0$ as $x \rightarrow \infty$, then for model (7) to fit well, the threshold u needs to be raised until $p_k(x) \approx 0$ for $x > u$. This is potentially very inefficient, given the subsequent loss of data to fit the model. So, here we extend model (7) to a non-standard subasymptotic version of the Heffernan-Tawn mixture model, where the mixture probabilities change with the level of extremity of X .

We are motivated by the following theoretical example. Let, $\lambda > 1$, $0 < t < 1$ and

$$Z \sim \text{Bernoulli}(p), \quad X \sim \begin{cases} \text{Exp}(1) & \text{if } Z = 0, \\ \text{Exp}(\lambda) & \text{if } Z = 1, \end{cases} \quad Y \sim \begin{cases} tX & \text{if } Z = 0, \\ \lambda X & \text{if } Z = 1, \end{cases} \quad (15)$$

where $\text{Exp}(\lambda)$ denotes the Exponential distribution with rate λ . Define $X_L = F_L^{-1}(F_X(X))$ and $Y_L = F_L^{-1}(F_Y(Y))$, where F_L^{-1} is the inverse cumulative distribution function of a standard Laplace, F_X and F_Y are the distribution functions of X and Y . Calculations in the supplementary material show that for large X_L ,

$$Y_L | X_L \stackrel{\mathcal{D}}{=} \begin{cases} tX_L + [t \log(2(1-p)) - \log(2p)] & \text{with probability } 1 - \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)X_L} \\ \lambda X_L + [\lambda \log(2(1-p)) - \log(2p)] & \text{with probability } \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)X_L}. \end{cases}$$

Thus, for α_2 in model (7), $\alpha_2 = \lambda > 1$, but for $x > 0$, $p_2(x) = \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)x} \rightarrow 0$ as $x \rightarrow \infty$ with $\alpha_2 = 1 - \lim_{x \rightarrow \infty} \log(p_2(x))/x$, which is explained by Theorem 1 below.

Note that we cannot model the distribution of (X_L, Y_L) well under the framework of model (7). We extend model (7) by allowing p_k to be a function of X , i.e.,

$$Y|(X > u) \stackrel{\mathcal{D}}{=} \gamma_k + \alpha_k X + X^{\beta_k}(\mu_k + \sigma_k \tilde{Z}_k), \quad \text{with probability } p_k(X), \quad k = 1, \dots, K, \quad (16)$$

where $\alpha_1 < \dots < \alpha_K$. In contrast with the Heffernan-Tawn mixture model, there could exist $2 \leq k_0 \leq K$ such that $\alpha_{k_0-1} < 1 < \alpha_{k_0}$, as illustrated in the motivating example,

where $k_0 = K = 2$. This model is equivalent to (7) if $p_k(x)$ is constant for all $x > u$ and $k = 1, \dots, K$, but it is different for generic $p_k(x)$. With this new model form, we capture non-constant mixture probability. It is not trivial to determine the valid parameter space of this new model, however, Theorem 1 provides some insight.

Theorem 1. *To ensure that model (16) for $Y|(X > u)$ does not violate the properties of the marginal distribution of Y , for all $k = 1, \dots, K$ we require*

$$|\alpha_k| \leq \liminf_{x \rightarrow \infty} \left(-\frac{\log p_k(x)}{x} \right) + 1.$$

The proof can be found in Appendix A. If $K = 2$ and $p_2(x) = e^{-(c-1)x}$ for $c \geq 1$, this result implies that a necessary condition is $\alpha_2 \leq c$. So α_2 can be larger than 1 as long as it is less than c . More generally, we consider the class of parametric functions

$$p_k(x) = \frac{a_k x^{b_k} e^{-c_k x}}{\sum_{j=1}^K a_j x^{b_j} e^{-c_j x}}, \quad (17)$$

where $a_k > 0$, $b_k \in \mathbb{R}$ and $c_k \geq 0$ for all $k = 1, \dots, K$. For identifiability reasons, we define c_k for $k = 1, \dots, K$ such that $\min\{c_k : k = 1, \dots, K\} = 0$ without loss of generality. We consider this class of parametric functions for $p_k(x)$ to be flexible as it satisfies the following four properties: (i) model (7) is a special case when $a_k = p_k$, $b_k = 0$ and $c_k = 0$ for all $k = 1, \dots, K$; (ii) model (15) is also a special case with $a_1 = 1$, $b_1 = 0$, $c_1 = 0$, $a_2 = p\lambda/(2^{\lambda-1}(1-p)^\lambda)$, $b_2 = 0$, and $c_2 = (\lambda - 1)$; (iii) Theorem 1 yields

$$|\alpha_k| \leq \liminf_{x \rightarrow \infty} \left(-\frac{\log(a_k x^{b_k} e^{-c_k x}) - \log\left(\sum_{j=1}^K a_j x^{b_j} e^{-c_j x}\right)}{x} \right) + 1 = c_k + 1$$

for all $1 \leq k \leq K$. So, $|\alpha_k|$ can be larger than 1; (iv) define $\mathcal{J} := \{j = 1, \dots, K : c_j = 0\}$, and $\tilde{\mathcal{J}} := \{j \in \mathcal{J} : b_j = \max\{b_i : i \in \mathcal{J}\}\}$, then model (16) is asymptotically equivalent to (7) with $K = |\tilde{\mathcal{J}}|$ and $p_j = a_j / \sum_{i \in \tilde{\mathcal{J}}} a_i$ for $i \in \tilde{\mathcal{J}}$ as $u \rightarrow \infty$.

5 Oceanographic data analysis

We investigate the oceanographical variables T_2 and H_S from the NORA10 hindcast dataset of Reistad et al. (2011). These variables are 3 hourly summary statistics that characterise the ocean environment; T_2 is the wave period and H_S the significant wave height. We apply the methods introduced above to data from a site in the northern North Sea, see e.g. Figure 1. from Konzen et al. (2021). This site is scientifically interesting as it displays seasonal and directional variability in the ocean environment.

To avoid issues with temporal and directional dependence in the observed data, we preprocess the data by combining an established peak-picking method of Randell et al. (2015) and keeping observations that are associated with storms originating from the Atlantic ocean, see supplementary material. This method is used to identify a subset of storm peak observations of H_S and associated values of T_2 which are approximately temporally independent. The main idea behind this method is the underlying assumption that consecutive storms are independent events. From the preprocessing, we reduce 176,765 observations recorded over the period 1957–2018 to 1597, and we denote the observations corresponding to significant wave height and wave period with the labels $H_{S,peak}$ and $T_{2,ass}$, respectively. Figure 1 shows a scatter plot of the original and storm peak samples.

Figure 1 shows that conditional on T_2 , H_S either takes on relatively small or relatively large values, whereas intermediate values are rare suggesting a mixture model with at least two components. We compare 4 mixture models applied to these data: the HT(K) and QR(K) models with $K = 1$ and $K = 2$. We define QR(K) and HT(K) to be abbreviations that correspond to the fits of the HTM model using the quantile-regression model and the Heffernan-Tawn mixture model, respectively, with a fixed number K of components.

We also consider two intuitive “partitioning” methods that accommodate the mixture

structure by partitioning the data into swell waves and wind-generated waves. We allocate observations to either component using wave steepness, a quantity proportional to $H_{S,peak}/T_{2,ass}^2$. This approach is physically well-motivated and often adopted by ocean engineering practitioners. Observations with steepnesses below a threshold value are allocated to the swell component, the remainder to the wind-generated component. For our data, the marginal distribution of steepness is bimodal, the lower mode corresponding to swell and the upper to wind-generated waves. It is therefore natural to choose the threshold as the value that lies between the modes and has a minimal estimated density. After partitioning, we fit single component models to swell and wind wave subsamples using the quantile-regression and Heffernan-Tawn models. We denote these models as Part-QR(1) and Part-HT(1), respectively. Finally, the joint distribution is given by a mixture of the two inferred single component distributions, with mixture weights determined empirically. Here, the context of the data allows us to undertake this method of partitioning. In general, this method is application dependent and requires expert knowledge.

Since the models introduced in this paper assume data on standard Laplace margins, we transform the data to standard Laplace scale using standard extreme value techniques (Coles and Tawn, 1994). Specifically, we proceed as follows: (i) select a marginal threshold; (ii) below the threshold, transform using the empirical probability integral transform; (iii) above the threshold, use the generalised Pareto distribution. Here we take the 70% quantiles of the marginal distributions of $T_{2,ass}$ and $H_{S,peak}$ although the fitted generalised Pareto distribution was found not to be overly sensitive to the choice of threshold. Both estimated generalised Pareto shape parameters were negative, implying finite upper endpoints of the marginal distributions of $T_{2,ass}$ (95% confidence interval $[15.4s, 21.3s]$) and $H_{S,peak}$ (95% confidence interval $[16.8m, 24.4m]$) for our data. We also took a threshold u for the HT

model as the 80% quantile of the standard Laplace distribution. Inferences were found to be relatively insensitive with respect to this choice. Similarly to the simulation studies, we set $\gamma_k = 0$ for all $k = 1, \dots, K$ in the HT mixture model. For trace plots of the model fits, we refer to the supplementary material.

We fix the originally fitted marginal models to avoid introducing further marginal uncertainty into inferences considering dependence. The uncertainty within the procedures is estimated via the following semi-parametric bootstrap procedure. We simulate a dataset of the same size as the original data using the inferred HT(2) model, see the supplementary material for details. Next, we fit the HT(K) and QR(K) models for different values of k and estimate the distribution of the response variable by simulating a large number of observations from the inferred models. Then, we transform the generated sample to original margins using the original inferred fixed marginal models.

One of the assumptions of the HTM model is that the mixture probability is constant as a function of the conditioning random variable although Section 4 shows this need not be the case. So, we design a goodness-of-fit diagnostic, see Figure 3, in which we plot estimates for the mixture probability p_2 as function of x . Both estimators in this plot are calculated using a sliding window approach with a fixed number of 30 observations per sliding window. The estimator $\hat{p}_2(x)$ is defined as the average of the allocation probabilities, see equation (10), over its corresponding bin. Additionally, we plot Part- $\hat{p}_2(x)$, estimated under the partitioning method, for which allocation to groups is deterministic. Finally, we obtain confidence bounds for both estimators using a semi-parametric bootstrap. From Figure 3, we argue that both estimators provide enough evidence to assume that the true mixture probability is constant as a function of x for $x > u$. Moreover, there is no evidence to assume that $\lim_{x \rightarrow \infty} p_2(x) = 0$. Hence, the HTM model should be a reasonable approach

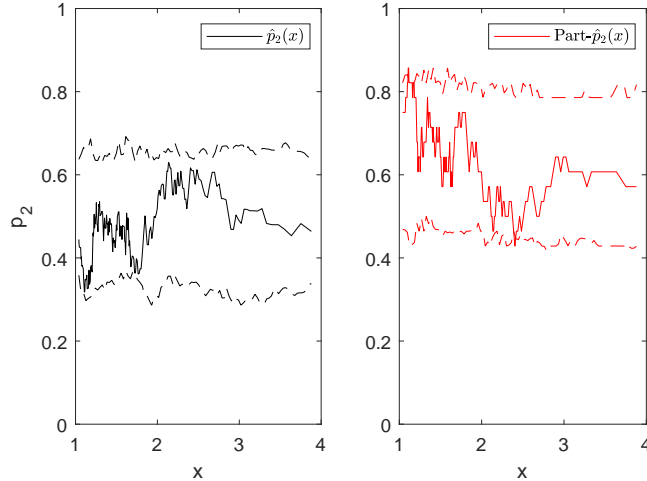


Figure 3: Estimates $\hat{p}_2(x)$ (left) and $\text{Part-}\hat{p}_2(x)$ (right) of $p_2(x)$ for the data. 95% confidence bounds are visualized with dashed lines.

in modelling the dependence structure of significant wave height and wave period, and there is no need to use the subasymptotic extension of the HTM model.

We compare the different model fits via the inferred joint distribution functions and the distributions of synthetic structure response variable R , illustrative of the response of floating offshore structures to wave loading, considered by Ross et al. (2020),

$$R := \frac{aH_{S,peak}}{1 + b(T_{2,ass} - c)^2}, \quad (18)$$

where a , b and c are structure-specific parameters with c being a resonant frequency. We take one response variable that is parameterised using equation (18) with $(a, b, c) = (2, 1, 16)$. This choice corresponds to structural responses with resonant frequencies in the near and far tail of the distribution of wave period.

Results are summarised in Figures 4 and 5 and parameter estimates are given in Tables 1 and 2. In Figure 4, we plot the inferred joint probabilities, see supplementary material

Model	$\hat{\alpha}$	$\hat{\beta}$
HT(1)	0.42 (0.03, 0.69)	0.39 (0.23, 0.53)
HT ₁ (2)	0.39 (0.16, 0.62)	0.05 (0.00, 0.25)
HT ₂ (2)	0.84 (0.21, 0.96)	0.24 (0.12, 0.58)
Part-HT(1)-Sea	0.74 (0.53, 0.85)	0.40 (0.31, 0.52)
Part-HT(1)-Swell	0.77 (0.24, 0.97)	0.13 (0.01, 0.45)

Table 1: Posterior median parameter estimates for the fitted HT(K) ($K = 1, 2$) and Part-HT(1) models including 95% credible intervals. We write HT _{i} (2) to denote the i th component of the HT(2) model, where $i \in \{1, 2\}$.

for details, for each rectangle on a discrete grid of rectangles covering the $(T_{2,ass}, H_{S,peak})$ domain $[12, 16.5] \times [3.1, 15.1]$. The partitioning methods, QR(2) and HT(2) models generate similar estimates showing two distinct arms in the dependency structure with increasing $T_{2,ass}$. In contrast, the HT(1) and QR(1) models do not capture the mixture structure as well. However, closer inspection of the HT(1) and QR(1) estimates also provides evidence for two arms in the dependency structure. This is due to the fact that the corresponding residual distributions are bimodal, itself suggesting the need for a two component mixture model, i.e., $K \geq 2$, see the end of Section 3.1.

We plot the results of the analysis for R in Figure 5. The top left corner gives estimated return levels using the 6 inference methods. The remaining panels show the semi-parametric bootstrap uncertainty of these estimates. We note that for a return period less than 1,000,000 years, the one component models estimate a smaller return level of the response R compared to the two component models; and higher return level for a larger return period. Moreover, the one component models tend to have wider confidence bounds.

Model	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$
QR(1)	0.69 (0.37, 1.07)	0.56 (0.31, 0.75)	0.48 (0.31, 0.55)
QR ₁ (2)	0.16 (−0.99, 1.03)	0.25 (0.10, 0.59)	0.00 (0.00, 0.45)
QR ₂ (2)	0.90 (0.68, 1.19)	0.85 (0.55, 1.00)	0.24 (0.00, 0.46)
Part-QR(1)-Sea	0.27 (0.08, 0.46)	0.81 (0.70, 0.95)	0.36 (0.11, 0.47)
Part-QR(1)-Swell	−0.09 (−0.58, 0.49)	0.91 (0.54, 1.00)	0.16 (0.00, 0.43)

Table 2: Parameter estimates for the fitted QR(K) ($K = 1, 2$) and Part-QR(1) models including 95% confidence intervals calculated via bootstrap. We write QR _{i} (2) to denote the i th component of the QR(2) model, where $i \in \{1, 2\}$.

Finally, the grey curve (representing the distribution calculated using the data-generating HT(2) model) lies within the 95% confidence bounds of all of the methods.

From Tables 1 and 2, we note that the parameter estimates for the HT(K) and QR(K) models are similar across the different methods. The confidence intervals tend to be wider for the one component models which is explained by a combination of model misspecification and the Keef constraints. The model parameters for the Part-QR(1) and Part-HT(1) models are significantly different from the QR(2) and HT(2) models. This is an expected feature due to the separate marginal transformations of the partitioned dataset. These parameter estimates for α and γ are strongly negatively correlated, so representations to obtain orthogonality of these parameters may be helpful.

We conclude that the two component models should be used when applicable since the confidence bounds on the return level estimates are tighter. There seems to be little difference between QR(2) and HT(2), hence either could be used. The estimates that are generated using the partitioning methods, especially Part-QR(1), have an even smaller

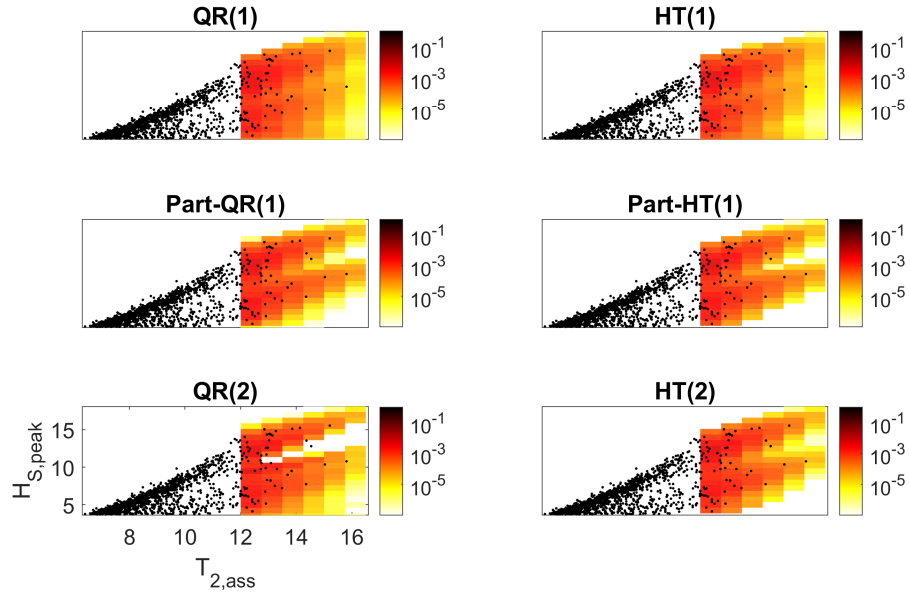


Figure 4: Estimates of probabilities of extreme sets on original margins using models: QR(1) (top left), HT(1) (top right), the partitioning method using two QR(1) models (middle left), the partitioning model using two HT(1) models (middle right), QR(2) (bottom left), and HT(2) (bottom right). Axis labels and scales are identical.

variance when compared to QR(2). However, the partitioning methods rely on considerable prior understanding of the underlying physical processes.

Acknowledgments

This paper is based on work completed while Stan Tendijck was part of the EPSRC funded STOR-i centre for doctoral training (EP/L015692/1), with part-funding from Shell Research Ltd.

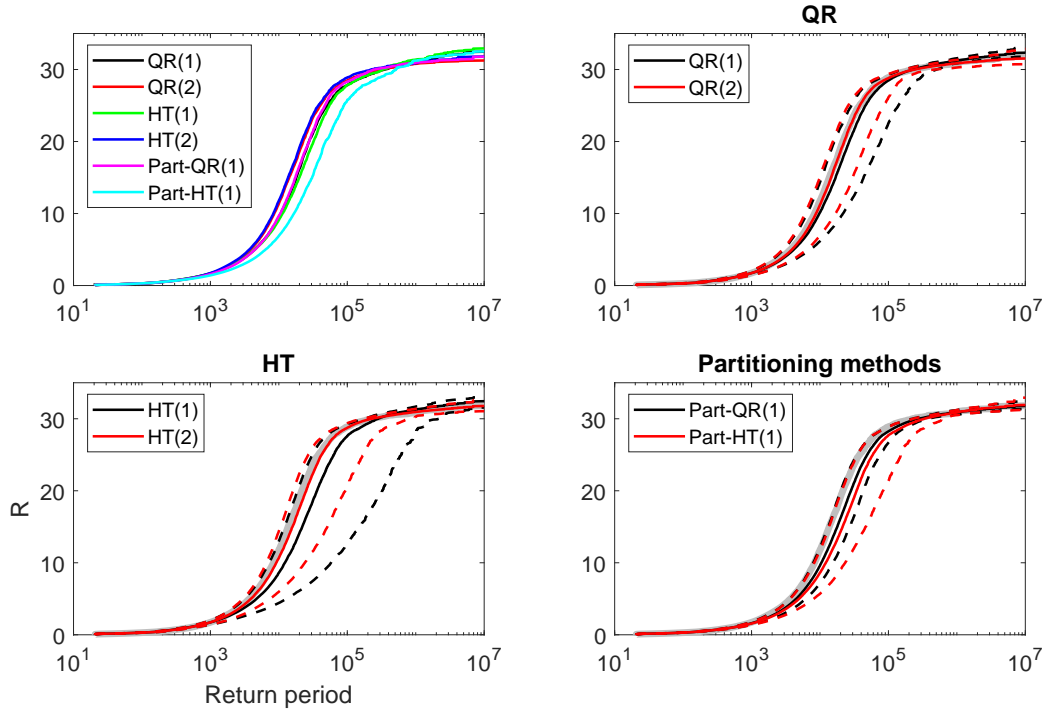


Figure 5: Top left: return level estimates of the synthetic response R in years using 6 approaches to inference. Other panels: uncertainty analysis for all six methods, similar to Figure 4. The solid lines are median estimates of the bootstrap ensemble and the dashed lines are the 2.5% and 97.5% quantiles. The light grey line shows the originally inferred HT(2) model and is common across all subplots. Axes are identical over plots.

SUPPLEMENTARY MATERIAL

Supplementary material: includes results of simulation studies, code and data, and other supporting material as referred to in the main text (.zip file)

A Proof of Theorem 1

Let $X, Y \sim \text{Laplace}(1)$. We consider the following simplified version of model (16)

$$Y|X \sim \alpha_k X, \text{ with probability } p_k(X) \quad (19)$$

with $0 \leq \alpha_k < \alpha_{k'}$ for all $1 \leq k \leq k' \leq K$, and where $p_k : \mathbb{R} \rightarrow [0, 1]$ are functions such that for all x $\sum_{k=1}^K p_k(x) = 1$. We will find a necessary condition on α_k given p_k such that the model formulation for Y does not contradict the assumption that Y is marginally distributed as a standard Laplace random variable. We need $2 \cdot \mathbb{P}(Y > y) = e^{-y}$ for all $y > 0$ as this implies that for all $u \in \mathbb{R}$ and $y > 0$, $\mathbb{P}(Y > y, X > u) \leq \frac{1}{2}e^{-y}$.

Using model (16), we simplify this probability as follows

$$\begin{aligned} \mathbb{P}(Y > y, X > u) &= \sum_{k=1}^K \mathbb{P}(Y > y, X > u, J = k) = \sum_{k=1}^K \mathbb{P}(\alpha_k X > y, X > u, J = k) \\ &= \sum_{k=1}^K \mathbb{P}(X > \max\{y/\alpha_k, u\}, J = k) = \sum_{k=1}^K \int_{\max\{y/\alpha_k, u\}}^{\infty} \mathbb{P}(J = k|X = x) f_X(x) dx \\ &= \frac{1}{2} \sum_{k=1}^K \int_{\max\{y/\alpha_k, u\}}^{\infty} p_k(x) e^{-x} dx \\ &= \frac{1}{2} e^{-\max(y/\alpha_1, u)} + \frac{1}{2} \sum_{k=2}^K \int_{\max\{y/\alpha_k, u\}}^{\max\{y/\alpha_{k-1}, u\}} \left(\sum_{i=k}^K p_i(x) \right) e^{-x} dx, \end{aligned} \quad (20)$$

from writing $p_1(x) = 1 - \sum_{i=2}^K p_i(x)$. We note that if $\alpha_k \leq 1$ for all $1 \leq k \leq K$, then by bounding $p_k(x) \leq 1$, we get $\mathbb{P}(Y > y, X > u) \leq e^{-y}/2$. Moreover, if $\alpha_1 > 1$, then we

trivially have $\mathbb{P}(Y > y, X > u) \geq e^{-y}/2$. We will assume from here onwards that we have an index $1 \leq j < K$ such that $\alpha_1 < \alpha_2 < \dots < \alpha_j \leq 1 < \alpha_{j+1} < \dots < \alpha_K$.

We are now in a position to prove the theorem by contradiction. To that end, assume that there exists an $1 \leq i_0 \leq K$ such that

$$\alpha_{i_0} = \liminf_{x \rightarrow \infty} -\frac{\log p_{i_0}(x)}{x} + 1 + \varepsilon \quad (21)$$

for some $\varepsilon > 0$. Note that this expression excludes $i_0 \leq j$ as $\liminf_{x \rightarrow \infty} -\frac{\log p_{i_0}(x)}{x} + 1 + \varepsilon > 1$. Hence, we deduce that $i_0 > j$. Now, define $A_{i_0}(x) = p_{i_0}(x)e^{(\alpha_{i_0}-1)x}$. Equation (21) implies

$$\liminf_{x \rightarrow \infty} A_{i_0}(x) = \liminf_{x \rightarrow \infty} p_{i_0}(x)e^{\left(-\frac{\log p_{i_0}(x)}{x} + 1 + \varepsilon - 1\right)x} = \liminf_{x \rightarrow \infty} e^{\varepsilon x} = \infty.$$

Hence, there exists an x' such that for all $x > x'$, $A_{i_0}(x) > \alpha_{i_0}$. For $y > \max\{\alpha_K x', \alpha_K u\}$, we get that equation (20) simplifies as follows

$$\begin{aligned} \mathbb{P}(Y > y, X > u) & \quad (22) \\ &= \frac{1}{2}e^{-y/\alpha_1} + \frac{1}{2} \sum_{k=2}^K \int_{y/\alpha_k}^{y/\alpha_{k-1}} \left(\sum_{i=k}^K p_i(x) \right) e^{-x} dx \geq \frac{1}{2}e^{-y/\alpha_1} + \frac{1}{2} \sum_{k=2}^{i_0} \int_{y/\alpha_k}^{y/\alpha_{k-1}} \left(\sum_{i=k}^K p_i(x) \right) e^{-x} dx \\ &\geq \frac{1}{2}e^{-y/\alpha_1} + \frac{1}{2} \sum_{k=2}^{i_0} \int_{y/\alpha_k}^{y/\alpha_{k-1}} p_{i_0}(x) e^{-x} dx = \frac{1}{2}e^{-y/\alpha_1} + \frac{1}{2} \sum_{k=2}^{i_0} \int_{y/\alpha_k}^{y/\alpha_{k-1}} (A_{i_0}(x) e^{-(\alpha_{i_0}-1)x}) e^{-x} dx \\ &> \frac{1}{2}e^{-y/\alpha_1} + \frac{1}{2} \int_{y/\alpha_{i_0}}^{y/\alpha_1} \alpha_{i_0} e^{-\alpha_{i_0}x} dx = \frac{1}{2} (e^{-y/\alpha_1} - e^{-\alpha_{i_0} \cdot y/\alpha_1}) + \frac{1}{2}e^{-y}, \end{aligned} \quad (23)$$

where in the last line we used that $A_{i_0}(x) > \alpha_{i_0}$ and $\bigcup_{k=2}^{i_0} \left[\frac{y}{\alpha_k}, \frac{y}{\alpha_{k-1}} \right] = \left[\frac{y}{\alpha_{i_0}}, \frac{y}{\alpha_1} \right]$. Since $\alpha_{i_0} > 1$ there exists a $y' > \max\{\alpha_K x', \alpha_K u\}$ such that for all $y > y'$, we have $e^{-y/\alpha_1} - e^{-\alpha_{i_0} \cdot y/\alpha_1} > 0$. Then for all $y > y'$ and $u \in \mathbb{R}$, $\mathbb{P}(Y > y, X > u) > \frac{1}{2}e^{-y}$. Thus, $\liminf_{x \rightarrow \infty} A_{i_0}(x) > \alpha_{i_0}$

contradicts with the marginal distribution of Y . We conclude that for all $i \geq j + 1$, we need to have $\liminf_{x \rightarrow \infty} p_i(x)e^{(\alpha_i - 1)x} \leq \alpha_i$, which rearranged gives

$$\alpha_i \leq \liminf_{x \rightarrow \infty} -\frac{\log p_i(x) - \log \alpha_i}{x} + 1 = \liminf_{x \rightarrow \infty} -\frac{\log p_i(x)}{x} + 1.$$

A symmetrical argument gives the same bound for $-\alpha_i$, concluding the proof.

References

- Boldi, M.-O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Methodology)*, 69(2):217–229.
- Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.
- Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577.
- Chiapino, M., Sabourin, A., and Segers, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes*, 22(2):193–222.
- Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.
- Coles, S. G. and Pauli, F. (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89(1):183–196.
- Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):1–31.

- De Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: an Introduction*. Springer Science & Business Media.
- Engelke, S. and Ivanovs, J. (2020). Sparse structures for multivariate extremes. *Annual Review of Statistics and its Application*, 8:241–270.
- Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17:537–571.
- 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 (Methodology)*, 66(3):497–546.
- Holthuijsen, L. H. (2010). *Waves in Oceanic and Coastal Waters*. Cambridge University Press.
- Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, 22(1):47–64.
- Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

- Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.
- Konzen, E., Neves, C., and Jonathan, P. (2021). Modeling nonstationary extremes of storm severity: Comparing parametric and semiparametric inference. *Environmetrics*, page e2667.
- Lugrin, T., Davison, A. C., and Tawn, J. A. (2021). Sub-asymptotic motivation for new multivariate extremes models. *To appear in Stat.*
- Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.
- Randell, D., Feld, G., Ewans, K., and Jonathan, P. (2015). Distributions of return values for ocean wave characteristics in the south china sea using directional–seasonal extreme value analysis. *Environmetrics*, 26(6):442–450.
- Reistad, M., Breivik, Ø., Haakenstad, H., Aarnes, O. J., Furevik, B. R., and Bidlot, J.-R. (2011). A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea. *Journal of Geophysical Research: Oceans*, 116(C5).
- Resnick, S. I. and Zeber, D. (2014). Transition kernels and the conditional extreme value model. *Extremes*, 17:263–287.
- Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Methodology)*, 59(4):731–792.

- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.
- Ross, E., Astrup, O. C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., and Vanem, E. (2020). On environmental contours for marine and coastal design. *Ocean Engineering*, 195:106194.
- Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.
- Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society Series B*, 79:149–175.
- Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130.