# Multivariate Oceanographic Extremes in Time and Space

Stan Hermanus Antoine Tendijck, M.Sc.

Submitted for the degree of Doctor of
Philosophy at Lancaster University.

December 2022

# Abstract

This thesis contributes to the field of multivariate extremes. The work has been motivated by an application in oceanography to assess safety and reliability of offshore structures, vessels, and platforms, but we remark that the contents of the thesis are designed to be more generally applicable to other environmental or even financial applications.

The model that forms the foundation for this work is the conditional extremes model (Heffernan and Tawn, 2004) in which the extremes of a multivariate random variable are modelled by conditioning on one of the variables being extreme. This model is called the Heffernan-Tawn model and it is one of the most flexible models for modelling extremes of multivariate random variables.

We design a mixture model for significant wave height conditional on large wave periods in the North Sea by extending the Heffernan-Tawn model. Our extension helps with understanding the distribution of responses to offshore facilities or vessels that are dominated by resonance frequencies. A mixture model is necessary here because two types of waves are recorded in the North Sea: swell waves and wind sea waves; both of these can be associated to large wave periods.

We calculate extremal properties of a model that is widely used by engineers in oceanographical applications. This model has a simple interpretation but is not motivated by extreme value theory. This led to the development of a mathematical toolset to calculate extremal characteristics for conditional models in general. This in turn allowed us to prove a new restriction on the space of the Heffernan-Tawn model parameters.

Finally, we model the joint temporal evolution of oceanographic variables using

an extension of the Heffernan-Tawn model to increase the understanding of what a 10,000 year event would look like. This led to a generic formulation of a multivariate extremes temporal model.

# Acknowledgements

First of all, I would like to thank my supervisors without whom I surely would not have been able to complete this work. Hence, I dedicate a paragraph to each of them.

Jon Tawn, I would like to thank you so much in assisting me in completing the Ph.D. In particular, for sharing the insane amount of knowledge in the field of extreme value theory. Our weekly meetings have been super useful to me keeping me on track with the Ph.D. Moroever, I'd like to thank you for the sheer amount of work you have put in to improve my writing skills - which were pretty poor when I started my Ph.D., and I hope it has improved to acceptable over the last four years.

Phil Jonathan, I would like to thank you for being the best industrial supervisor I could wish for. I really enjoyed and learned a lot from our weekly Monday meetings. Even though,one hour per week on average seems a bit much - especially in comparison with the 15-20 minutes I spend with Jon, it always felt like it wasn't. Especially, because our meetings never started with work but always about Joel Piroe scoring again for Swansea, and me sarcastically wondering if that was in the Premier League or in the insignificant league. I do also want to thank you for the multiple days that you spent working with me in which you gave me lots of helpful comments on both my code and writing. Finally, I thank you for recommending STOR-i to me, without this recommendation I would definitely never have found this opportunity.

David Randell, I would like to thank you for being my third most important supervisor. Even though, you have been a bit more on the background compared to the other two, I always have valued our discussions a lot and I continue to learn from them. You were valuable in helping me with answering any MATLAB question I have, and I am sure that without your MATLAB skills (also skills that I hope to

III

have picked up a bit during my internship before the Ph.D.), this Ph.D. would have been a lot harder. I am looking forward to working with you at Shell in the future.

Finally, there is Emma Eastoe as my fourth supervisor. Unfortunately, due to your time constraints, you disappeared into the background after our first paper was published. I do still thank you for giving me invaluable feedback on the writing of my first paper of the Ph.D. This was particularly useful to me at that stage.

Second of all, I'd like to thank STOR-i for giving me the opportunity to join their Lancaster-based family. I have learned a lot from many of my colleagues by attending each of the weekly forums. Also, the STOR-i program (both in the Master of Research phase and the Ph.D. phase) has been invaluable giving me a taste of nearly any flavour of statistics, data science and operations research in either module form, reading group form, or master class form. There is many more aspects of STOR-i that I am grateful for but they are too many to mention all.

Last but actually most importantly, I would like to thank Livia for being the most awesome girlfriend I could wish for during my Ph.D. Because of you, the quality of my life during this period improved so much that I am for ever grateful to you. In particular, I would not have wanted to spend the corona pandemic with someone else than you.

# Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

The content in Chapter 3 has been published online as Tendijck, S., Eastoe, E., Tawn, J., Randell, D., and Jonathan, P. (2021). Modelling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2021.1996379.

The content in Chapter 4 has been published online as Tendijck, S., Tawn, J., and Jonathan, P. (2022). Extremal characteristics of conditional models. *Extremes*. DOT: 10.1007/s10687-022-00453-7.

The content in Chapter 5 is a draft paper to be submitted as Tendijck, S., Jonathan, P., Randell, D., and Tawn, J. (2023). Temporal evolution of the extreme excursions of multivariate $k$th order Markov processes with application to oceanographic data.

The word count for this thesis is approximately 62,000.

Stan Hermanus Antoine Tendijck

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation

Extreme events are the hot topic of today. Quite literally actually. At the time of writing of this thesis, record temperatures are being observed in Canada, the Netherlands, Latvia, and elsewhere in the world. The modelling of such extreme events in a statistical sense is important for many reasons related to safety, reliability and risk management. Although modelling temperature extremes is not directly the topic of this thesis, it does show how extreme events can have a direct impact on our daily lives.

When I introduce the topic of my Ph.D., I always try to visualize what it is like from the perspective of a manager on an offshore platform somewhere in some sea or ocean. For the manager, it is helpful to have an understanding of the risk whilst their employees are living and working on this structure. If the weather forecast is predicting a $10,000$ year storm to arrive next weekend, should the manager advise the workers to be evacuated or is the platform safe enough to withstand such extreme events. What about a $1,000$ year storm? Of course, the life of workers is the most important aspect of the work; on the other hand, from a business perspective, it is also very important to assess whether an offshore structure or vessels are built well enough to withstand such extreme events both from the perspective of the company and its insurers. This motivates the main aspect of the work in my thesis: modelling

oceanographic extremes, i.e., what does the ocean look like during the most extreme storms, with applications related to risk assessment. Answering the question 'how safe is it?' is not only crucial from a practical point of view, it is also very interesting from a statistical point of view because of the sheer complexity of the ocean environment.

To emphasize the importance of work in this area, we remark that there are numerous examples of offshore facility disasters in the past. Often, such events are caused by accidents and negligence (e.g., the explosions on the Piper Alpha oil platform in 1988 and the Deepwater Horizon in 2010) but many are also caused by extreme weather events (e.g., the capsizing of the Alexander L Kielland in 1980 and the collision of the Usumacinta jack-up rig with the Kab-101 in 2007); the latter of these causes being the focus of this thesis. Modelling such extreme weather events and relating them to facility failures is far from trivial. For example, it is not necessarily the case that the period of highest average waves in a storm also contains the largest wave that causes the most damage. Moreover, a combination of simultaneously extreme waves, winds, currents, etc, during a storm could have more of an impact to a structure than when one considers only the highest wave of a storm with its associate winds and current speeds.

In a statistical framework, the safety question translates to the calculation of failure probabilities. For an offshore structure, we define a generic long-term response variable $\mathbf{X}$ and a region $A$ such that $\mathbf{X} \in A$ is equivalent to a failure. We are interested in calculating probabilities of the form $\mathbb{P}(\mathbf{X} \in A)$ such that risk can be assessed accurately and offshore facilities can potentially be reinforced to reflect a company's policy. The procedure of calculating such procedures consists of 5 steps.

1. The modelling of the long-term ocean environment $\mathbf{E}$: a multivariate random variable that consists of summary statistics that summarize the state of the ocean.

2. The modelling of the short-term ocean environment $\mathbf{E}^*$ conditional on $\mathbf{E}$. This process consists of upscaling the summary statistics to estimate wave heights conditional on the average wave height, associated wave periods, wind gusts,

etc.

3. The modelling of the short-term responses $\mathbf{X}^*$ conditional on $\mathbf{E}^*$. For example, if we know the height of a wave and its period, we can calculate the wave steepness and subsequently calculate the effect of this single wave on a vessel.

4. The modelling of the long-term responses $\mathbf{X}$ conditional on $\mathbf{X}^*$, for example, the total impact of a storm on an offshore facility which can be either cumulative in nature or it can involve taking a maximum over a time period.

5. The calculation of failure probabilities $\mathbb{P}(\mathbf{X} \in A)$ by integrating out the four hierarchical models for: $\mathbf{X}|\mathbf{X}^*$, $\mathbf{X}^*|\mathbf{E}^*$, $\mathbf{E}^*|\mathbf{E}$ and $\mathbf{E}$, respectively.

All five of these steps are important and deserve an equal amount of attention. In this thesis, however, we only focus on the first step: the modelling of the long-term ocean environment $\mathbf{E}$. The main reason for this is that the other steps involve physical models. The key statistical problem lies in the first step with this requiring research development in the area of extreme value theory.

Some components of $\mathbf{E}$ that I will be using in my thesis are the following three-hourly summary statistics: significant wave height, peak wave period, wind speed, wave direction and wind direction. Significant wave height $H_S$ summarizes the amplitude of the surface elevation as the average wave height of 1/3 of the highest measured waves; the peak wave period $T_p$ is defined as the wave period that generates the most energy; wind speed $W_s$ is the average wind speed over the measured time interval; finally, the wave direction $\Theta^H$ and wind direction $\Theta^W$ are also averages of the wave and wind direction over the time period. Other features of the ocean environment that are usually part of $\mathbf{E}$ but will not be used in this thesis are current speeds at an array of depths, surge, surface pressure, precipitation, temperature, tides, etc. See Holthuijsen (2010) for more details on definitions of such oceanographic variables.

Because a failure, i.e., $\mathbf{X} \in A$, due to a weather event is generally unlikely - or at least designed to be unlikely - the failure associated weather events must be extremely unlikely as well. The modelling of such extreme weather events is at the

heart of this thesis. We note that this should definitely not be attempted using simple statistical methods - like linear regression or principal component analyses - because one learns that extrapolation beyond the observed data is a no-go in the field of statistics. However, in our case reliable records only go back for about 100 years at best, whereas companies usually aim to design facilities to only break down to a $1,000$ year storm - an observation that would be so extreme that it probably has not been observed before. Given a structure, can we additionally say if it can survive a $10,000$ year storm? If not, how can we improve or design such a structure to survive these type of extreme events? Standard statistical theory fails but the questions need answers.

The research field of probability and statistics that currently is receiving a lot of attention that does try to model such rare events is the field of extreme value theory. In short, extreme value theory methods are applicable to model tails of distributions and can assign probabilities to events never observed before. Because such extrapolation is risky from a modelling point of view, the arguments for any of these methods need to be solid. In fact, almost all univariate extreme value theory methods are derived from asymptotic probabilistic theory. So that these methods that model tails of distributions at least have a theoretical foundation to rely on.

In my Ph.D., extreme weather events are the main player, and such events are too complex to summarize with a single univariate random variable. Thus, multivariate extreme value approaches are needed for tackling the problems posed in this thesis. Such modelling techniques are much more complex than its univariate counterparts, because one cannot derive a parametric form for the extremes of any type of distribution - something that is possible for univariate random variables. So, for modelling purposes, one would need to assume a sub-class of possible limits. In the literature review, many of these methods are discussed. One of these methods, the flexible conditional extremes approach from Heffernan and Tawn (2004), in particular will be used as a building block for my newly developed methodology.

During my Ph.D. I can proudly look back on the following three major accomplishments. The first is the description of a bivariate mixture model for conditional

extremes. For this model, we have introduced two inference techniques - one of these is based on quantile regression which is similar to the inference method from Liu and Tawn (2014), and the other one is based on likelihoods which can straightforwardly be extended to model multivariate mixtures with dimension greater than 2. The impact of this work is that we now can model the dependence between significant wave height $H_S$ and peak wave period $T_p$ more accurately.

The second is a significant contribution to linking asymptotic theory with conditional models. We recognized that conditional models are used a lot in practice, e.g., Haver and Winterstein (2009) for modelling oceanographic variables. These models are usually applied without a proper understanding of what assumption such models make on the extremes: including assumptions on marginal extremes and/or extremal dependence. This is not surprising since the mathematics that link these two is far from trivial. Our major contribution is the development of a set of mathematical tools to make these calculations feasible.

Finally, our third addition to the field is the description of a temporal model for the extremes of multivariate random variables. We are not aware of anyone who has tried this before, and thus this is a completely new addition to the field of extremes. In our oceanographic application, a description for the temporal evolution of the ocean environment is important for risk assessment that is related to response variables that are cumulative of nature: for example, the cumulative damage of waves from a $10,000$ year storm onto a structure. Nota bene, this work does not just have an impact on oceanographic applications but can also be applied to a wide variety of extreme value applications: for example, a joint statistical risk analysis of river gauges data due to simultaneous snow melt and precipitation, a risk analysis of persistent heat waves at multiple locations, etc.

## 1.2 Thesis outline

In this project, we tackle three interesting extreme value related challenges.

In Chapter 2, we present an overview of the field of extreme value theory. This

includes a thorough overview of extremes for univariate random variables, stationary processes and multivariate random variables.

In Chapter 3, we present an extreme value mixture model that is applied to modelling significant wave height and wave period at a location in the northern North Sea. For this very speficic application, there exist at least two latent processes that describe the dependence between significant wave height and wave period: swell waves and wind sea waves. We recognize that standard statistical methods for modelling bivariate extremes are likely to fail to give reliable inferences in such cases. More generally, 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 our oceanographic application. 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.

In Chapter 4, we present work on the extremes of conditionally specified models. Such models are often used to describe complex multivariate data because of their simple interpretation. However, without practitioners being aware, these conditional models assume implicit structures on the extremes. Until now, we are not aware of any methodology for calculating extremal characteristics of conditional models since the copula and marginals are not expressed in closed forms. We consider bivariate conditional models that specify the distribution of $X$ and the distribution of $Y$ conditional on $X$. We provide tools to quantify implicit assumptions on the extremes of this class of models. In particular, these tools allow us to approximate the distribution of the tail of $Y$ and the coefficient of asymptotic independence $\eta$ in closed forms. We apply these methods to a widely used conditional model for wave height and wave period. Moreover, we introduce a new condition on the parameter space

for the conditional extremes model of Heffernan and Tawn (2004), and prove that the conditional extremes model does not agree with $\eta$, when $\eta < 1$.

In Chapter 5, we develop two models for the temporal evolution of extreme events of multivariate $k$th order Markov processes. The foundation of our methodology lies in the conditional extremes model of Heffernan and Tawn (2004), and it naturally extends the work of Winter and Tawn (2016, 2017) and Tendijck et al. (2019) to include multivariate random variables. We use cross-validation-type techniques to develop a model order selection procedure, and we test our models on two-dimensional met-ocean data with directional covariates for a location in the northern North Sea. We conclude that the newly-developed models perform better than a baseline historical-matching methodology for these data.

In Chapter 5, we give a summary of my main contributions to this thesis and we discuss further research opportunities.

# Chapter 2

# Literature review

## 2.1 Overview

In this chapter, an overview of the field of extreme value theory is presented. Section 2.2 introduces univariate extreme value theory for which two ideologies are detailed: peaks-over-threshold and block maxima. These ideologies were initially presented under the assumption of independent random variables. In Section 2.3, we discuss how the results change when instead of independent random variables, a stationary random process is considered. Both previous methods lay the foundation for complex multivariate extreme value theory methods in Section 2.4.

## 2.2 Univariate extreme value theory

Univariate extreme value theory lays the foundation of all multivariate extreme value theory methods. So, understanding how to model the extremes of univariate distributions is necessary before one can even think about more complex multivariate applications. In this section, we explore the extremes of a univariate random variable $X$ with distribution function $F$ that has right-upper end point $x^* := \sup\{x \in \mathbb{R} : F(x) < 1\} \in \mathbb{R} \cup \{\infty\}$. To that end, let $X_1$, $X_2$, ... be an independent and identically distributed (iid) sequence of random variables with the same distribution function $F$, and assume that data $\{x_1, x_2, \dots\}$ associated to this process are available.

There are two common methodologies for modelling the extremes of $X$: (i) block maxima, which model $\max\{X_1, \ldots, X_n\}$ as $n$ tends to infinity; and (ii) peaks-over-threshold, which models $X|X > u$ as $u$ tends to infinity. In Figure 2.2.1, these two univariate extreme value approaches are visualised. In this section, we will discuss the main theoretical results for both of these methods.
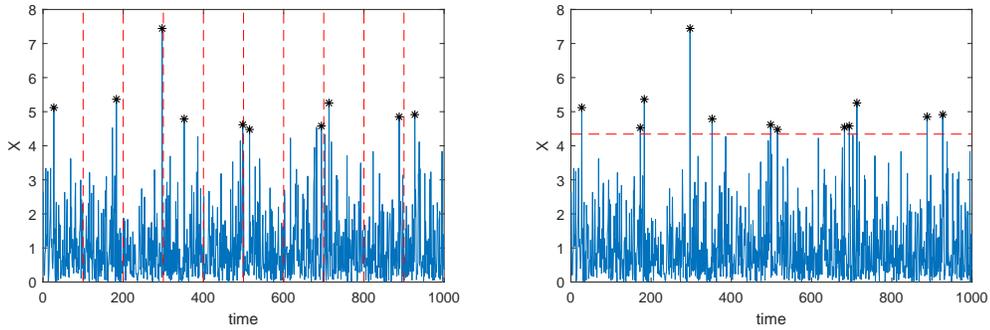


Figure 2.2.1: Two univariate approaches for modelling extremes: Left, block maxima; right, peaks-over-threshold. The starred points correspond to the maxima of the blocks (left) and the data above the threshold (right).

## 2.2.1 Block maxima

The block maxima approach evolves around modelling the maximum

$$M_n := \max\{X_1, \ldots, X_n\}, \ n \in \mathbb{N}$$

as $n$ tends to infinity. Before we get into modelling $M_n$, we remark that any theory developed for maxima can be applied to minima as well by applying the relation

$$\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}.$$

We now work out

$$\lim_{n \to \infty} \mathbb{P}(M_n \leq x) = \lim_{n \to \infty} F(x)^n = \begin{cases} 0 & \text{for } x < x^* \\ 1 & \text{for } x \geq x^*. \end{cases} \tag{2.2.1}$$

This shows that $M_n$ converges to $x^*$ in probability. In Theorem 2.2.1, we present exactly how $M_n$ converges to $x^*$.

Asymptotic theory on the convergence of $M_n$ to $x^*$ has been developed in (Fisher and Tippett, 1928), (Gnedenko, 1943) and (de Haan, 1970).

**Theorem 2.2.1.** *Let $X_i$ be an i.i.d. sequence of random variables and define $M_n := \max\{X_1, \ldots, X_n\}$ for each $n \in \mathbb{N}$. If there exists a non-degenerate random variable $G$ and constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{D}} G \tag{2.2.2}$$

*holds, then $G$ must be a generalised extreme value (GEV) random variable with shape $\xi \in \mathbb{R}$, location $\mu \in \mathbb{R}$ and scale $\sigma > 0$:*

$$P(G \leq x) = \begin{cases} \exp\left(-(1 + \xi z)^{-1/\xi}\right), & \xi \neq 0 \\ \exp(-\exp(-z)), & \xi = 0 \end{cases} \quad \text{for } 1 + \xi z > 0$$

*with $z = (x - \mu)/\sigma$, and for $1 + \xi z \leq 0$, $\mathbb{P}(G \leq x) = 0$ when $\xi > 0$, and $\mathbb{P}(G \leq x) = 1$ when $\xi < 0$.*

If we additionally assume that the distribution function $F$ is known and that $X$ has a density $f$ which is differentiable near the upper-end point $x^*$, then Smith (1987) has shown that we can pick $a_n = 1/(nf(b_n))$ and $b_n = F^{-1}(1 - 1/n)$ in Theorem 2.2.1. With these choices, the rescaled maximum converges in distribution to a generalised extreme value distribution with shape $\xi \in \mathbb{R}$, location $\mu = 0$ and scale $\sigma = 1$.

### 2.2.2 Peaks-over-threshold

A different method to model univariate extremes is the peaks-over-threshold which aims at modelling $X|X > u$ when $u$ tends to $x^*$ from below. Similar to block maxima, we scale $X|X > u$ cleverly such that in the limit the scaled version converges in distribution to a non-degenerate random variable.

**Theorem 2.2.2** (Theorem 7. in Pickands (1975)). *Let $X$ be a random variable. If Theorem 2.2.1 is applicable to a sequence of iid copies of $X$, then there exists a continuous function $c(u) > 0$ such that*

$$\frac{X - u}{c(u)} \,\bigg|\, (X > u) \xrightarrow{\mathcal{D}} V \tag{2.2.3}$$

*as $u$ tends to $x^*$ from below, where $V$ is generalised Pareto with shape $\xi \in \mathbb{R}$ and scale $\sigma > 0$:*

$$P(V \leq x) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi} & \text{for } (x > 0, \ \xi > 0) \text{ or } (x < -\sigma/\xi, \ \xi < 0) \\ 1 & \text{for } (x > -\sigma/\xi, \ \xi < 0) \\ 0 & \text{otherwise.} \end{cases}$$

## 2.2.3 Relating block-maxima to peaks-over-threshold

The peaks-over-threshold and block maxima approaches are closely related. In particular, the parameter $\xi$ in Theorem 2.2.1 is the same as the parameter $\xi$ in Theorem 2.2.2. We show this here. Define $D$ as the domain of $G$, and let $x \in D$ then

$$P(G \leq x) = \lim_{n \to \infty} F(a_n x + b_n)^n = \exp\left(\lim_{n \to \infty} n \log F(a_n x + b_n)\right)$$
$$= \exp\left(\lim_{n \to \infty} n \cdot [F(a_n x + b_n) - 1]\right).$$

The limit inside the exponent must exist. Next, let $v \in D$ such that $x > v$, then we can use this as follows:

$$\frac{-\log P(G \leq x)}{-\log P(G \leq v)} = \frac{\lim_{n \to \infty} n \cdot [1 - F(a_n x + b_n)]}{\lim_{n \to \infty} n \cdot [1 - F(a_n v + b_n)]} = \lim_{n \to \infty} \frac{1 - F(a_n x + b_n)}{1 - F(a_n v + b_n)}$$
$$= \lim_{n \to \infty} P(X - (a_n v + b_n) > a_n(x - v) \mid X > a_n v + b_n)$$
$$= \lim_{n \to \infty} P\left(\frac{X - u_n}{c(u_n)} > x - v \mid X > u_n\right)$$

with $u_n = a_n v + b_n$ and $c(u_n) := a_n$. We note that the right-hand side takes on the same form as in limit (2.2.3) and the left-hand side is exactly in the form of a generalised Pareto distribution with shape $\xi$ and scale $\sigma_v := \sigma + \xi(v - \mu)$:

$$\frac{-\log P(G \leq x)}{-\log P(G \leq v)} = \left(\frac{1 + \xi(x - \mu)/\sigma}{1 + \xi(v - \mu)/\sigma}\right)^{-1/\xi} = (1 + \xi(x - v)/\sigma_v)^{-1/\xi}.$$

So, the parameter $\xi \in \mathbb{R}$, which parameterizes the heaviness of the tails of the limiting distributions of $G$ and $V$, appears in both block maxima and peaks-over-threshold. This parameter is called the extreme value index, and historically three extreme value classes of distributions have been defined based on the value of $\xi$

1. $\xi > 0$: The Fréchet class for distributions of maxima or Pareto class for distributions of exceedances. Random variables that belong to this limiting class have a heavy tail such that moments of order $(1/\xi)$ and greater do not exist.

2. $\xi = 0$: The Gumbel class for distributions of maxima or exponential class for distributions of exceedances. Random variables that are associated with $\xi = 0$ have either a finite or infinite upper-endpoint. In either case, all moments always exist.

3. $\xi < 0$: The reverse-Weibull class for distributions of maxima, there is no name for this class associated to distributions of exceedances. Random variables that have this limiting form have a finite upper-endpoint. For example, when $\xi = -1$, the generalised Pareto distribution is the uniform distribution on $[0, \sigma]$.

Because this thesis will be centred around applications rather than theoretical results, we shall not compare these methods here via their limiting results because they will not be relevant to the rest of the thesis. For the interested reader, we refer to Cai et al. (2013); Ferreira and de Haan (2015); Dombry and Ferreira (2019).

### 2.2.4  Practical choices for inference

We do not have infinite resources and for practical applications we do not know the distribution function $F$ (and density $f$) in advance. So, to infer the extremal behaviour of an iid random process given a finite set of data $\{x_1, \ldots, x_n\}$, we need to discuss some practical choices.

First and foremost, block maxima and peaks-over-threshold are based on limiting relations. So to infer the extremal behaviour of a univariate random variable using either of these methods, we need to assume that limits (2.2.2) and (2.2.3) hold exactly at some finite level. For block maxima, this implies that we must assume there exists an $m \in \mathbb{N}$ such that $M_m := \max\{X_1, \ldots, X_m\}$ has a generalised extreme value distribution with shape $\xi \in \mathbb{R}$, location $\mu^* := a_m \mu + b_m \in \mathbb{R}$ and scale $\sigma_1^* := a_m \sigma$. For peaks-over-threshold, we must assume that there exists a $v \in \mathbb{R}$ such that $(X_1 - v)|X_1 > v$ has a generalised Pareto distribution with shape $\xi \in \mathbb{R}$ and scale

$\sigma_2^* := \sigma c(v)$. Although, $a_m, b_m, c(v), \mu, \sigma$ are not known, the unknown $\mu^*, \sigma_1^*, \sigma_2^*$ can be directly inferred as the unknown parameters of the models.

So, to assess the heaviness of the tail of $X$ or equivalently perform inference on $\xi$, we must additionally perform inference on $m$ or $v$. In literature, methods that infer $m$ in practice have not been addressed extensively because: (i) literature on block maxima for iid processes focuses mainly on asymptotic results, and hardly addresses a finite case because one would for example have to bound the level of model misspecification; (ii) in many applications, one cannot really compare multiple choices for $m$. For example, for weather related data, yearly maxima is either appropriate or it's not, one can usually not make the blocks slightly larger or smaller in these cases. Methods that infer $v$ on the other hand have been addressed extensively. The most common ones are: (i) rules of thumb: for example, the 10% rule (DuMouchel, 1983), or the square-root rule (Ferreira et al., 2003), or the empirical rule (Loretan and Phillips, 1994); (ii) graphical methods: for example, mean residual life plots, threshold stability plots, quantile or return level plots (Davison and Smith, 1990; Coles et al., 2001); (iii) automated methods, which do not rely on the subjectivity of a researcher (Hall, 1990; Gomes and Oliveira, 2001); (iv) composite models, which model both the tail and bulk of the distribution simultaneously (Tancredi et al., 2006); (v) theoretically motivated methods Hall and Weissman (1997); Ferreira et al. (2003); (vi) test statistics based on goodness-of-fit methods (Northrop and Coleman, 2014; Wadsworth, 2016). For an interested reader, we refer to Scarrott and MacDonald (2012) who present a very clean overview of different methods.

In all of the above, apart from the composite models, it is suggested to perform inference in a two-step fashion: First select $m$ $(v)$, then perform inference conditional on $m$ $(v)$. We now present some inference procedures for block maxima and peak-over-threshold separately.

**Block maxima** We assume that $m$ is known and for easiness of presentation we assume that $B := n/m \in \mathbb{N}$. Next, split up the data into $B$ blocks where the $i$th block consists of observations $\{x_{B \cdot (i-1)+1}, \dots, x_{B \cdot i}\}$. To infer, the distribution of $M_m$, we can now make use of $B$ observations - the maxima of each block - that are assumed

to follow a generalised extreme value distribution. The parameters of the GEV can be estimated using any preferred method of inference: for example, maximum-likelihood, probability-weighted moments or Bayesian type methods.

Two other inference methods for the block maxima approach have recently been considered in literature: sliding block maxima (SBM) (Bücher and Segers, 2018) and all block maxima (ABM) (Oorschot and Zhou, 2020). In SBM, the data are split up into $n - m$ blocks where the $i$th block contains observations $\{x_i, \ldots, x_{i+m}\}$. In ABM, the data are split up into $\binom{n-m}{B}$ blocks. In both of these methods, the maxima of the blocks can no longer be considered independent. However, the quasi-maximum likelihood estimate which is obtained by assuming that the observations are independent is still valid and even has lower asymptotic variance than the original block maxima approach.

**Peaks-over-threshold** We assume that $v$ is known, then the observations $\{x_i : x_i > v\}$ are approximated with a generalised Pareto distribution. Estimates of the parameters of this distribution can be obtained with your favourite method of inference, eg maximum likelihood, probability weighted moments or Bayesian type methods.

We remark more on the differences in the next section where we discard the independence assumption and instead consider a stationary random process.

## 2.3   Univariate extremes for stationary processes

In environmental applications, nearly any real physical process does not produce independent observations. Because the theory in the previous section has been set up for independent random variables only, it is important to understand how exactly the discarding of the independence assumption changes the theory. So, now let us assume that $X_1, X_2, \ldots$ are generated from a stationary random process. We note that for non-stationary processes, one can remove non-stationarity using standard statistical techniques and subsequently apply extreme value theory to the stationary residual process, see for example Eastoe and Tawn (2009). Below, we comment on intuition

that will be discussed more rigorously in the next subsections.

Under the assumption that temporal dependence decreases to 0 when time between random variables increases, the block maxima approach is still expected to be asymptically valid since the temporal dependence in block maxima vanishes. Thus, they can be considered independent as long as the block size tends to infinity. However, the dependence within a block cannot be ignored; so that it is not necessarily trivial how to calculate quantiles of the marginal distribution of the $X_i$ when we assume that the maxima of $n$ consecutive observations follow a GEV distribution. Moreover, in any application we cannot have block sizes growing to infinity due to finite resources.

For the peaks-over-threshold method, it is easier to understand that the method is still valid since we only consider the distribution of $X|X > u$. However, the conditional sample does not consist of independent realisations.

## 2.3.1   Block maxima for stationary processes

Mathematically, the concept of approximate independence for block maxima has been defined in Leadbetter et al. (1983) as the $D(u_n)$ condition. We do not give the exact statement of this condition here because it is not relevant for understanding the theory, however, we do like to remark that it is exactly what one expects: it defines 'near' independence by bounding the difference of the joint distribution function of a random process and the product of the marginal distribution functions.

To relate the stationary case to the independence case, they define the extremal index $\theta > 0$ of a stationary random process as follows. Firstly, let the stationary sequence $X_1, X_2, \ldots$ satisfy the $D(u_n)$ condition and let $\tilde{X}_1, \tilde{X}_2, \ldots$ be the iid sequence with the same marginal distribution as our original stationary random process, and assume that when $\tilde{M}_n := \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$ is scaled as $(\tilde{M}_n - b_n)/a_n$, we have

$$\lim_{n \to \infty} P\left(\frac{\tilde{M}_n - a_n}{b_n} \leq x\right) = P(G \leq x)$$

for a generalised extreme value random variable $G$ with shape $\xi$, location 0, and scale 1. Next, they assume that for each $x \in \mathbb{R}$, $P\left((M_n - a_n)/b_n \leq x\right)$ converges as $n$ tends to infinity, and then they derive that this implies there must exist a $\theta \in (0, 1]$ such

that for each $x \in \mathbb{R}$,

$$\lim_{n \to \infty} P \left( \frac{M_n - a_n}{b_n} \leq x \right) = P(G \leq x)^{\theta}$$

holds. Then, $\theta$ is the extremal index of the stationary random process. It is then quite straightforward to see that for $\tilde{a}_n = a_n \theta^{\xi}$ and $\tilde{b}_n = b_n - a_n(1 - \theta^{\xi})/\xi$, we must have

$$\lim_{n \to \infty} P \left( \frac{M_n - \tilde{a}_n}{\tilde{b}_n} \leq x \right) = P(G \leq x).$$

This shows that asymptotically the maximum of a stationary random process that satisfies the $D(u_n)$ condition also has a generalised extreme value distribution. So, the block maxima approach is still valid. We do remark that to estimate the marginal distribution using the block maxima approach, one needs to estimate $\theta$.

### 2.3.2   Peaks-over-threshold for stationary processes

As mentioned before, observations from the sample $\{x_i : x_i > u\}$ from the stationary process are generated from the same marginal distribution as if there was no temporal dependence. However, when there is temporal dependence, then obviously the exceedances are not independent. This needs to be taken into account in inference.

First of all, we note that probability weigthed moment estimators are applicable to temporally dependent data; so this type of estimation method is still perfectly valid in this case. Secondly we note that for evaluating likelihoods. one needs to make a modelling assumption for the dependence structure, which will not be the simple product as in the iid case since the density no longer factorizes. However, one could potentially assume that the density factorizes and maximize the product over the parameter space. This estimator is called the quasi-maximum likelihood estimator (Lee and Hansen, 1994).

A different way of modelling peaks-over-threshold to the data is by applying the methodology to cluster maxima only. Under similar conditions as in Section 2.3.1, the basic peaks-over-threshold model yields clustered Poisson processes when applied to a stationary time-series. So, Davison and Smith (1990) conclude that the asymptotic methodology still must hold when applied to cluster maxima only, see also Sec-

tion 2.3.5. Eastoe and Tawn (2012) discuss the implications of this assumption when the limiting theory is applied at finite thresholds.

### 2.3.3 Comparing peaks-over-threshold and block maxima for stationary processes

Similar to Section 2.2.3, we compare here peaks-over-threshold with block maxima approach but now for stationary processes. As described above, both methods are still applicable as described above. However, the differences between the methods are more significant compared to the iid case.

In estimating high quantiles of the marginal distribution, the peaks-over-threshold method is preferred over the block maxima approach. This is because for peaks-over-threshold, we can use a plug-in estimator that is directly derived from the generalised Pareto fit. To use block maxima to do the same, we additionally requires an estimator of the extremal index $\theta$ which induces more variability to estimates of these quantiles. We can see the latter as follows. Assume we have estimated $M_n$ with a GEV distribution that has the following parameters: location $\tilde{a}_n \in \mathbb{R}$, scale $\tilde{b}_n > 0$ and shape $\xi \in \mathbb{R}$, then we must have that

$$\mathbb{P}(X_1 \leq x) = \mathbb{P}\left(M_n \leq \tilde{a}_n + \tilde{b}_n \frac{x - \tilde{a}_n \theta^{-\xi}}{\tilde{b}_n + \tilde{a}_n(\theta^{-\xi} - 1)/\xi}\right)^{1/n}$$

where the latter is the distribution function of a GEV random variable with location $\tilde{a}_n$, scale $\tilde{b}_n$ and shape $\xi$. This expression can now be inverted to get an estimator of large quantiles of the marginal of $X_1$ in terms of $\tilde{a}_n$, $\tilde{b}_n$, $\xi$, and $\theta$.

The opposite is true for estimating large return periods for annual maxima. The block maxima approach is directly applicable, whereas peaks-over-threshold can only be used to estimate return levels if additionally the extremal index is estimated; or, in case the method is applied to cluster maxima only, an estimate for the rate of occurrence of clusters is required.

## 2.3.4 The extremal index

The extremal index of a stationary process as defined in Section 2.3.1 gives interesting information on the extremal behaviour of the process. Here, we discuss an interpretation for $\theta$, and how we can estimate it.

First of all, if $\theta = 1$, then extreme events cannot occur consecutively. So, asymptotically, it does not matter for modelling the extremes if one forgets that the process does not consist of independent observations. In practical applications, however, Eastoe and Tawn (2012) show that this assumption is not necessarily correct at finite levels, and discuss the behaviour of the sub-asymptotic version of the extremal index

$$\theta(x, m) := \mathbb{P}(\max\{X_2, \ldots, X_m\} < x \mid X_1 > x) \qquad (2.3.1)$$

for large levels $x \in \mathbb{R}$ and run length $m \in \mathbb{N}$. On the other hand, if $\theta < 1$ then extreme events will occur together.

There exist a couple of very fine interpretations of the extremal index $\theta$. Hsing et al. (1988) showed that the extremal index $\theta$ is equal to the reciprocal of the mean size of an extreme cluster, and O'Brien et al. (1987) shows that the extremal index can be interpreted as the probability of an extreme value being the last extreme value of a cluster.

Next, we describe two estimators for $\theta$: the blocks estimator and the runs estimator. Some other types of estimators do exist (Ferro and Segers, 2003; Hsing, 1993) but we do not consider these to be either relevant for the thesis or necessary to understand the broader aspects of the field of extreme value theory.

The blocks estimator for the extremal index, see Smith and Weissman (1994), is derived as follows. Let $u \in \mathbb{R}$ be some high threshold and let $N_u$ be equal to the number of times $X_i$ for $1 \leq i \leq n$ exceeded $u$. Next, split up the data up into $k$ blocks of length $n/k$, and define $Z_u$ as the number of blocks for which the maximum of the random process exceeds $u$. The blocks estimator is now defined as $\hat{\theta}_u := Z_u/N_u$.

The runs estimator is defined as follows. Let $u \in \mathbb{R}$ be some high threshold and let $N_u$ be the same as in the blocks estimator. Next, set $W_i$ equal to 1 if $X_i > u$ and 0 otherwise and let $r > 0$ be such that temporal dependence at distance $r$ can

be approximated to have vanished. Finally, set $Z_u^* := \sum_{i=1}^{n} W_i \prod_{j=1}^{r}(1 - W_{i+j})$ as
the number of last exceedances of extremal clusters with a cluster deemed to have
finished when there are $r$ consecutive values of $X_i$ below $u$. The runs estimator of the
extremal index is now defined as

$$\tilde{\theta} := Z^*/N. \tag{2.3.2}$$

### 2.3.5   Extremal clusters

Before, we mentioned the idea that cluster maxima in Section 2.3.2 are approximately
independent, and peaks-over-threshold can be applied to these cluster maxima. This
idea has been formalized in Ferro and Segers (2003). In their work, they define an
extremal cluster of a stationary process for a threshold and a separation parameter $m$
such that exceedances of this threshold that are separated with $m-1$ consecutive non-
exceedances can be considered approximately independent. Ledford and Tawn (2003)
additionally define diagnostics for assessing this independence assumption between
clusters, and they cover within-cluster dependence by bounding the subasymptotic
extremal index in equation (2.3.1) from above and below using the coefficient of
asymptotic independence, see Section 2.4.2.

## 2.4   Multivariate extreme value theory

In Section 2.2, we considered univariate iid random variables $X_1$, $X_2$, $\ldots$. In this
section, we assume $d$-dimensional data $\{\mathbf{x}_i :  i = 1, \ldots, n\}$ with $\mathbf{x}_i := (x_{i1}, \ldots, x_{id})$
that are generated from multivariate iid random variables $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$ for
$i = 1, \ldots, n$.

   The first difficulty of modelling multivariate extremes compared to univariate ex-
tremes is that there exists no natural ordering on $\mathbb{R}^d$ for $d \geq 2$, and thus it is less
clear how to define an extreme event. Barnett (1976) provides a number of options for
ordering $d$-dimensional observations of which the following two are commonly used in
modelling extremes: (i) marginal ordering; and (ii) reduced aggregate ordering.

The marginal ordering approach looks at models for component-wise maxima of a sample, see Section 2.4.3, and under reduced aggregate ordering, one applies a function that maps $\mathbb{R}^d$ onto $\mathbb{R}$ transforming the multivariate ordering problem to the univariate setting, see Coles and Tawn (1994) for examples. In Section 3.2.1, we discuss a different methodology that does not require ordering methods because it models a multivariate vector conditional on one component being large.

The second difficulty of modelling multivariate extremes is the vast array of possibilities of combinations of random variables with varying levels of dependence. To get an intuition for one of the difficulties, see Figure 2.4.1. In this plot, we show data that are simulated from three distributions with the same dependence model but on different margins. From inspection, it is not obvious that the dependence of the data is exactly the same across all three random samples; that is because marginal effects dominate our interpretation of dependence. It is for this reason that in multivariate extremes, we always model our multivariate data with the following two-step approach: (i) model each of the marginals using univariate extreme value theory, and using the probability integral transform these marginals onto standard margins - which can be uniform, exponential, Laplace, Gumbel, Fréchet, etc; (ii) model the remaining dependence with parsimonious distributions.

In Section 2.4.1, we present the theory of copulas which will prove to be useful in multivariate extremes by providing a mathematical framework for combining the two steps mentioned above. In Section 2.4.2, we define functionals that quantify the level of statistical dependence in the extremes, and in Section 2.4.5, we model the asymptotic shapes of samples generated from multivariate distributions.

## 2.4.1   Copula theory

Sklar's theorem (Sklar, 1959) forms the foundation for the theory of copulas: a statistical framework that can be used to model complex dependence models.

**Theorem 2.4.1** (Sklar (1959)). *Consider a d-dimensional distribution function $F$ with marginals $F_1, \ldots, F_d$. Then there exists a copula $C_U$ with uniform margins such*
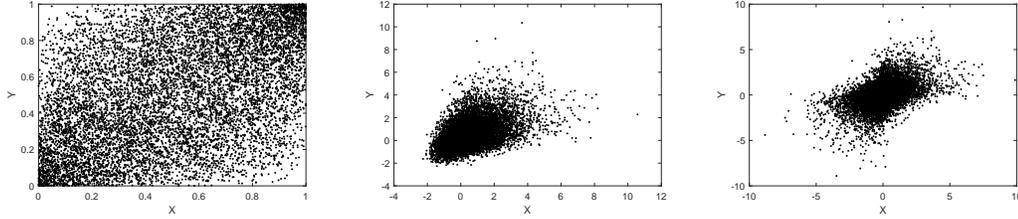
Figure 2.4.1: Data simulated from three multivariate distributions, each with the same dependence model (Gaussian) but with different marginals: standard uniform, standard Gumbel and standard Laplace (from left to right).

*that*

$$F(x_1, \ldots, x_d) = C_U(F_1(x_1), \ldots, F_d(x_d))$$

*for all* $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. *Moreover, if* $F_i$ *are continuous for all* $i = 1, \ldots, d$, *then* $C_U$ *is unique.*

In Sklar's theorem, the copula $C_U$ always has standard uniform margins. Because the extremes of $C_U$ will cluster at the boundaries of the unit cube, we extend the definition of a "copula" to include multivariate distributions with different marginals - as long as it is the same marginal distribution for each component. Copulas with Gumbel margins $C_G$ (Heffernan and Tawn, 2004) and standard Laplace margins $C_L$ (Keef et al., 2013), for example, are defined by combining the above theorem and the probability integral transform. Copula $C_G$ satisfies

$$F(x_1, \ldots, x_d) = C_G \left\{ h[F_1(x_1)], \ldots, h[F_d(x_d)] \right\}. \tag{2.4.1}$$

with $h(u) = -\log(-\log u)$ for $u \in (0, 1)$, and for copula $C_L$, equation (2.4.1) holds with $h(u) = -\mathrm{sgn}(u - 1/2) \log[1 - 2|u - 1/2|]$. In most of our applications, we use Laplace margins because it allows for identical treatment of both positive and negative dependence across components due to its symmetry, however, different parts of the theory are presented for different marginals when this simplifies our presentation. For an overview of the field of copulas, see Joe (1997).

## 2.4.2   Extremal dependence

We consider two different types of extremal dependence for multivariate random variables: asymptotic dependence (AD) and asymptotic independence (AI). In particular, for AD (AI) models, the asymptotic conditional probability $\chi \in [0, 1]$ that all components of the random variable are extreme together given that one variable is extreme is greater than 0 (equal to 0). Joe (1997) defines this concept in the bivariate case as follows.

**Definition 2.4.2.** *Let $(X_1, X_2)$ be a bivariate random vector such that $X_1 \overset{\mathcal{D}}{=} X_2$ holds. The coefficient of extremal dependence $\chi = \chi(\{1, 2\})$ is defined as*

$$\chi := \lim_{x \uparrow x^*} P(X_2 > x \mid X_1 > x),$$

*provided this limit exists, where $x^* := \sup\{x : \ \mathbb{P}(X_1 < x) < 1\}$ is the right end-point of the marginal distribution.*

So, $\chi$ is equal to the asymptotic probability of one component being extreme conditional on that the other is extreme as well. Wadsworth and Tawn (2013) extend Definition 2.4.2 to random variables $\mathbf{X} = (X_1, \ldots, X_d)$ of a higher dimension $d > 2$.

**Definition 2.4.3.** *Let $(X_1, \ldots, X_d)$ be a multivariate random vector with identically distributed components. Define $x^* := \sup\{x : \ \mathbb{P}(X_1 < x) < 1\}$ as the right end-point of the marginal distribution, and let $\mathcal{S}$ be the power set of $\{1, \ldots, d\}$. For each set $S = \{s_1, \ldots, s_k\} \in \mathcal{S}$, we define $\chi = \chi(S)$ as*

$$\chi := \lim_{x \uparrow x^*} P(X_{s_2} > x, \ \ldots, \ X_{s_k} > x \mid X_{s_1} > x),$$

*provided this limit exists.*

It is now natural to introduce the concept of strong and weak joint tail dependence (Wadsworth and Tawn, 2013), and consequently asymptotic dependence and asymptotic independence.

**Definition 2.4.4.** *Let $(X_1, \ldots, X_d)$ be a multivariate random vector with identically distributed components. Let $\mathcal{S}$ be the power set of $\{1, \ldots, d\}$, and let $S =$*

$\{s_1, \ldots, s_k\} \in \mathcal{S}$. *Then, we say that the random vector* $(X_{s_1}, \ldots, X_{s_k})$ *has strong joint tail dependence if and only if* $\chi(S) > 0$. *On the other hand, if* $\chi(S) = 0$, *then we say that this random vector has weak joint tail dependence. If* $S = \{i, j\}$, *then we use the terminology asymptotically dependent and asymptotically independent, respectively.*

The concept of asymptotic dependence cannot be compared with generic dependence. In particular, there exist strongly correlated asymptotically independent random variables, like the bivariate Gaussian with correlation $\rho < 1$ (Sibuya, 1959). In this case, the coefficient of extremal dependence $\chi = 0$ does not reflect the relative strength of dependence in the extremes. Ledford and Tawn (1996) noticed this and defined the coefficient of tail dependence $\eta \in (0, 1]$ for asymptotically independent random variables as follows

**Definition 2.4.5.** *Let* $(X_1, X_2)$ *be a bivariate random vector such that* $X_1$, $X_2$ *have standard Fréchet margins. Assume there exists a slowly varying function* $\mathcal{L} : \mathbb{R} \to \mathbb{R}_{>0}$, *i.e., for* $z > 0$, $\mathcal{L}(xz)/\mathcal{L}(x) \to 1$ *as* $x \to \infty$ *for* $z > 0$, *such that*

$$P(X_1 > x, \ X_2 > x) = \mathcal{L}(x)x^{-1/\eta} \tag{2.4.2}$$

*holds. Then* $\eta = \eta(\{1, 2\})$ *is termed the coefficient of tail dependence.*

First, we note that if $X_1$ and $X_2$ are asymptotically dependent, then $\eta = 1$, and $\chi = \lim_{x \to \infty} \mathcal{L}(x)$. Next, if $X_1$ and $X_2$ are independent, then $\eta = 0.5$ and $\mathcal{L}(x) \sim 1$ as $x \to \infty$. Finally, if $X_1$ and $X_2$ are identically distributed but do not have standard Fréchet marginals, then $\eta$ is defined as follows. First, define

$$\eta(x) := \frac{\log \mathbb{P}(X_1 > x)}{\log \mathbb{P}(X_1 > x, \ X_2 > x)}.$$

Then, $\eta := \lim_{x \uparrow x^*} \eta(x)$ with $x^* := \sup\{x : \ \mathbb{P}(X_1 < x) < 1\}$ the right upper end-point of the marginal distribution of $X_1$.

Wadsworth and Tawn (2013) extend the definition of $\eta$ for bivariate random variables to a $d$-dimensional variant for $(X_1, \ldots, X_d)$ by adapting equation (2.4.2) to

$$\mathbb{P}(X_1 > x, \ \ldots, \ X_d > x) = \mathcal{L}(x)x^{-1/\eta}$$

for a slowly varying function $\mathcal{L}$. In this case, independence between all components yields $\eta = 1/d$.

There exist many more multivariate extremal dependence measures besides $\chi$ and $\eta$. However, in our work, $\chi$ and $\eta$ are the most relevant, and these will be coming back in main parts of the thesis. For a broaded overview, we also introduce two extremal dependence measures that we will not end up using. For a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with unit-exponential margins, Wadsworth and Tawn (2013) define the angular dependence coefficient $\lambda(\boldsymbol{\omega})$ for $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$ such that $\omega_1 + \cdots + \omega_d = 1$ as follows for $v > 0$:

$$\mathbb{P}\left(X > \boldsymbol{\omega}v\right) = L(e^v; \boldsymbol{\omega})e^{-\lambda(\boldsymbol{\omega})v},$$

where $L(t; \boldsymbol{\omega})$ is regularly varying at infinity with order $0$ as function of $t$ for each fixed $\boldsymbol{\omega}$. The measures $\lambda$ and $\eta$ are linked via

$$\eta = d \cdot \lambda(1/d, \ldots, 1/d).$$

So, the angular dependence coefficient $\lambda$ is a direct extension of $\eta$ that includes the dependence of components in different directions other than the diagonal.

Simpson et al. (2020) define the regular variation coefficient $\tau_C(\delta)$ for $\delta \in [0, 1]$ and $C \subseteq \{1, \ldots, d\}$ for a multivariate random vector $\mathbf{X} = (X_1, \ldots, X_d)$ on Pareto margins. This coefficient measures the extremal dependence when simultaneously $X_i$ is large for $i \in C$ (strong joint tail depedence: $\chi(C) > 0$) and $X_j$ is small for $j \notin C$ (weak joint tail dependence: $\chi(C \cup \{j\}) = 0$ for $j \notin C$). They define the coefficient by assuming that $f_{x,y}(t) := \mathbb{P}\left(\min_{i \in C} X_i > xt, \ \max_{j \notin C} X_j \leq yt^\delta\right)$ for $x, y > 0$ as a function of $t$ is regularly varying at infinity. They define $\tau_C(\delta) = -1/k$ where $k$ is equal to the order at which $f_{x,y}$ is regularly varying.

## 2.4.3 Component-wise maxima

The simplest method for modelling multivariate extremes is to order data or random variables marginally, i.e., $M_{n,i} = \max_{j=1,\ldots,n} X_{ij}$ and investigate the behaviour of

$(M_{n,1}, \ldots, M_{n,d})$. It should be noted that observations of these component-wise maxima do not per se coincide with actual observations, when the maxima in different components do not occur simultaneously.

Similar to the univariate case, we apply an appropriate component-wise scaling such that Theorem 2.2.1 holds for all marginal components individually. So, we assume that for $i = 1, \ldots, d$ there exist constants $a_{ni} > 0$, $b_{ni} \in \mathbb{R}$ such that

$$\left( \frac{M_{n,1} - b_{n1}}{a_{n1}}, \ldots, \frac{M_{n,d} - b_{nd}}{a_{nd}} \right) \xrightarrow{\mathcal{D}} \mathbf{G}, \qquad (2.4.3)$$

for a non-degenerate random variable $\mathbf{G}$. The class of distributions that can occur in limit (2.4.3) are defined as the multivariate extreme value distributions. Unlike in the univariate case, there is no analytic form for each distribution function within this class. Nonetheless, it is still possible to specify some theoretical properties of such multivariate extreme value distributions, and to give examples of parametric submodels.

For example, a straightforward argument shows that $\mathbf{G} = (G_1, \ldots, G_d)$ needs to be max-stable, i.e., for each $i = 1, \ldots, d$, $k \geq 1$, there exist $\alpha_{ki} > 0$ and $\beta_{ki} \in \mathbb{R}$ such that for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\mathbb{P}\left( \alpha_{k1} x_1 + \beta_{k1} \leq G_1, \ldots, \alpha_{kd} x_d + \beta_{kd} \leq G_d \right)^k = \mathbb{P}(G_1 \leq x_1, \ldots, G_d \leq x_d).$$

So, each marginal $G_i$ of $\mathbf{G}$ for $1 \leq i \leq d$ must have a generalised extreme value distribution. This means that if $\mathbf{X}$ has an extreme value distribution, then limit (2.4.3) holds with equality for any $n \in \mathbb{N}$. This in turn implies that any distribution function of an extreme value distribution must have GEV marginals and must be infinitely max-divisible (Balkema and Resnick, 1977).

Now assume that $\mathbf{X}$ has standard Fréchet margins. Following Pickands (1981), we rewrite the distribution function $F_{\mathbf{G}}$ of an extreme value distribution $\mathbf{G}$ as follows

$$F_{\mathbf{G}}(\mathbf{x}) = \exp\{-V(\mathbf{x})\}$$

for $\mathbf{x} \in D := \{\mathbf{x} \in \mathbb{R}^d : F_{\mathbf{G}}(\mathbf{x}) \geq 0\}$ and some function $V : \mathbb{R}^d \to \mathbb{R}_{\geq 0} \cup \{\infty\}$. This function $V$ will be referred to as the exponent measure, and its properties are discussed below.

The first result related to the exponent measure $V$ is that any function $V$ implies that $F_{\mathbf{G}}$ is an extreme value distribution if and only if it satisfies the following conditions (de Haan and Resnick, 1977): (i) The function $V$ needs to be homogeneous of order $-1$, i.e., $V(c\mathbf{x}) = V(\mathbf{x})/c$ for any $c > 0$ and $\mathbf{x} \in D$; (ii) $V$ needs to take on the form

$$V(\mathbf{x}) = d \int_{\mathcal{S}_{d-1}} \max_{i=1,\ldots,d} \left( \frac{w_i}{x_i} \right) \, \mathrm{d}H(\mathbf{w})$$

with $H$ a distribution function on

$$\mathcal{S}_{d-1} = \left\{ w \in [0,1]^d : \sum_{i=1}^{d} w_i = 1 \right\}$$

that satisfies

$$d \int_{\mathcal{S}_{d-1}} w_i \, \mathrm{d}H(\mathbf{w}) = 1$$

for $i = 1, \ldots, d$, which ensures that the margins of $F_{\mathbf{G}}$ are standard Fréchet. In literature, $H$ is commonly referred to as the spectral measure.

**Examples of extreme value distributions**

We present two examples of extreme value distributions. The first example is the multivariate extreme value distribution on exponential margins with a logistic dependence model that is parameterised with $\alpha \in (0,1]$ (Gumbel, 1960). The exponent measure $V$ of this model is defined as

$$V(\mathbf{x}) := \left( \sum_{i=1}^{d} x_i^{-1/\alpha} \right)^{\alpha}, \tag{2.4.4}$$

where $\mathbf{x} = (x_1, \ldots, x_d)$ with $x_i > 0$, $i = 1, \ldots, d$, and a parameter $\alpha \in (0,1]$ which represents the amount of dependence within the model. More specifically, $\alpha = 1$ corresponds to the case of complete independence and $\alpha \downarrow 0$ yields perfect dependence. A sample from the bivariate version of this model with parameter $\alpha = 0.5$ and standard Gumbel margins is plotted in Figure 2.4.2 (left).

Tawn (1988, 1990) extend the dependence model (2.4.4) to a more general asymmetric logistic dependence structure. The exponent measure $V$ of their model is

$$V(\mathbf{x}) = \sum_{C \in \mathcal{S} \setminus \varnothing} \left( \sum_{i \in C} \left( \frac{\theta_{i,C}}{x_i} \right)^{1/\alpha_C} \right)^{\alpha_C} \qquad x_i > 0, \ i = 1, \ldots, d, \tag{2.4.5}$$

where $\mathcal{S}$ is the power set of $\{1, \ldots, d\}$. The parameters $\alpha_C, \theta_{i,C}$ with $C \in \mathcal{S}$ and $i \in C$ of this model must satisfy the following constraints: for all $C \in \mathcal{S}$, $i \in C$, $\alpha_C \in (0, 1]$, $\theta_{i,C} \in [0, 1]$, and $\sum_{C \in \mathcal{S}} \theta_{i,C} = 1$. Moreover, for identifiability reasons, $\alpha_C = 1$ if $C$ contains one element. A sample from the bivariate version of this model with parameters $\left(\theta_{1,\{1\}}, \theta_{2,\{2\}}, \theta_{1,\{1,2\}}, \theta_{2,\{1,2\}}, \alpha_{\{1,2\}}\right) = (0.5, 0.5, 0.5, 0.5, 0.5)$ and Gumbel margins is plotted in Figure 2.4.2 (right). This set of parameters yields a mixture between two logistic models, one with dependence and one with independence.

For more examples of parametric multivariate extreme value distributions, see Galambos (1994); Hüsler and Reiss (1989); Joe (1990); Cooley and Thibaud (2019).
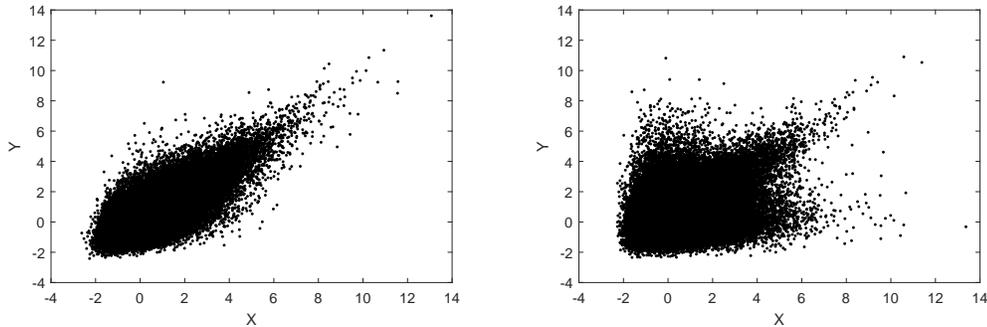


Figure 2.4.2: Bivariate extreme value distributions on Gumbel margins. $100,000$ observations from the logistic model (left) with parameter $\alpha = 0.5$ and from the asymmetric logistic model (right) on with parameters $\left(\theta_{1,\{1\}}, \theta_{2,\{2\}}, \theta_{1,\{1,2\}}, \theta_{2,\{1,2\}}, \alpha_{\{1,2\}}\right) = (0.5, 0.5, 0.5, 0.5, 0.5)$.

### 2.4.4 Conditional extremes

We introduce the conditional extreme value model of Heffernan and Tawn (2004), henceforth denoted the HT model. Their model is widely studied and applied to extrapolate multivariate models. It is a limit model and its form is motivated by derived limiting forms from numerous theoretical examples.

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with standard Laplace margins (Keef et al., 2013) and assume that its joint density exists. Define

$$\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$$

for all $i = 1, \ldots, d$. Next, assume that for each $i = 1, \ldots, d$ there exist functions $a_{|i} : \mathbb{R} \to \mathbb{R}^{d-1}$, $b_{|i} : \mathbb{R} \to \mathbb{R}^{d-1}_{>0}$ and a distribution function $H_{|i}$ on $\mathbb{R}^{d-1}$ that is non-degenerate in each margin such that for all $\mathbf{z} \in \mathbb{R}^{d-1}$ the following limits

$$\lim_{u \to \infty} \mathbb{P}\left( \frac{\mathbf{X}_{-i} - a_{|i}(X_i)}{b_{|i}(X_i)} \leq \mathbf{z}, \ X_i - u > x \mid X_i > u \right)$$

and

$$H_{|i}(\mathbf{z}) = \lim_{x \to \infty} \mathbb{P}\left( \frac{\mathbf{X}_{-i} - a_{|i}(x)}{b_{|i}(x)} \leq \mathbf{z} \mid X_i = x \right) \tag{2.4.6}$$

exist. This implies, according to l'Hôpital's rule, that for $x > 0$

$$\lim_{u \to \infty} \mathbb{P}\left( \frac{\mathbf{X}_{-i} - a_{|i}(X_i)}{b_{|i}(X_i)} \leq \mathbf{z}, \ X_i - u > x \mid X_i > u \right) = H_{|i}(\mathbf{z}) \exp(-x). \tag{2.4.7}$$

The latter in turn has the interpretation that as $u \to \infty$, $(\mathbf{X}_{-i} - a_{|i}(X_i))/b_{|i}(X_i)$ and $(X_i - u)$ are independent conditional on $X_i > u$, and are distributed as $H_{|i}$ and a standard exponential, respectively.

To apply the HT model, we assume that limit (4.4.1) holds exactly above some high threshold $u^* > 0$ so that for $u > u^*$:

$$\mathbb{P}\left( \frac{\mathbf{X}_{-i} - a_{|i}(X_i)}{b_{|i}(X_i)} \leq \mathbf{z}, \ X_i - u > x \mid X_i > u \right) = H_{|i}(z) \exp(-x). \tag{2.4.8}$$

The functions $a_{|i}$ and $b_{|i}$ are usually parameterised with $a_{|i}(x) = \boldsymbol{\alpha}_{|i} x$ and $b_{|i}(x) = x^{\boldsymbol{\beta}_{|i}}$, where $\boldsymbol{\alpha}_{|i} \in [-1, 1]^{d-1}$ and $\boldsymbol{\beta}_{|i} \in (-\infty, 1)$ satisfy the Keef et al. (2013) conditions. These constraints are there to ensure that in equation (2.4.8) the marginals of $\mathbf{X}_{-i}$ do not contradict with the initial requirement that $\mathbf{X}_{-i}$ has standard Laplace margins. These parametrisations are suitable for most generic cases but in general they can be anything as long as they satisfy some regularity conditions detailed in (Heffernan and Resnick, 2007).

A slightly different parameterization $b_{|i}(x) = 1 + \{a_{|i}(x)\}^{\boldsymbol{\beta}_{|i}}$ was considered in Wadsworth and Tawn (2013). This parameterization was motivated by the following example. Let $(X, Y)$ be a bivariate random vector with exponential margins that are linked with a Gaussian copula with correlation parameter $\rho \in [0, 1)$. Using the earlier parameterization of the conditional extremes model, we must have $a_{|i}(x) = \rho^2 x$ and $b_{|i}(x) = \sqrt{x}$ when $\rho > 0$, and $a_{|i}(x) = 0$, $b_{|i}(x) = 1$ when $\rho = 0$. They remarked

that the function $b_{|i}$ as function of $\rho$ is not continuous at 0. This discontinuity is solved with using the alternative parameterization $b_{|i}(x) = 1 + \{a_{|i}(x)\}^{\beta_{|i}}$. In our applications, however, the Keef et al. (2013) parameterizations are sufficient.

Simple interpretations of the parameters $\boldsymbol{\alpha}_{|i} := (\alpha_{1|i}, \ldots, \alpha_{i-1|i}, \alpha_{i+1|i}, \ldots, \alpha_{d|i})$ and $\boldsymbol{\beta}_{|i} := (\beta_{1|i}, \ldots, \beta_{i-1|i}, \beta_{i+1|i}, \ldots \beta_{d|i})$ are as follows:

**Case 1**: $(\alpha_{j|i}, \beta_{j|i}) = (1, 0)$ implies that $X_j$ and $X_i$ are asymptotically dependent for $i \neq j$.

**Case 2**: $\alpha_{j|i} < 1$ implies that $X_j$ and $X_i$ are asymptotically independent for $i \neq j$.

**Case 3**: $\beta_{j|i} > 0$ implies that the variability between $X_j$ and $X_i$ grows as $X_i$ grows.

**Case 4**: $0 < \alpha_{j|i} < 1$ (respectively, $-1 < \alpha_{j|i} < 0$) implies positive (negative) association between $X_j$ and $X_i$ for $i \neq j$.

For an example of how this model can capture certain dependency structures, see Figure 2.4.3 where data are simulated from the HT model with $\alpha_{|1} = 0.5$, $\beta_{|1} \in \{-0.5, 0, 0.5\}$, and $u = 3$. We note that for environmental applications $\beta_{j|i} < 0$ is unrealistic as it would imply that the uncertainty in the dependence of the most extreme events is zero. So, the restriction $\beta_{j|i} \geq 0$ can be considered, see for example Tawn et al. (2018).

### Inference

We present inference for the HT model for bivariate data $\mathcal{D} = \{(x_i, y_i) : i = 1, \ldots, n\}$, and comment on the differences for $d$-dimensional data.

The first step in the inference procedure involves around choosing a threshold $u$ and a conditioning variable. The choice of conditioning variable is usually a modelling assumption, and the choice of the threshold $u$ can be justified with parameter stability diagnostics. In this example, we condition on the first component and assume that
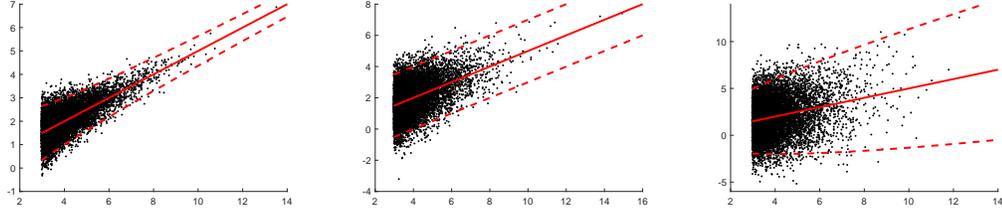
Figure 2.4.3: A sample of size 10.000 from the HT model on Laplace margins with $\alpha = 0.5$, $u = 3$, and from left to right $\beta = -0.5, 0, 0.5$ given that the first component $X_1 > 3$ of $\mathbf{X}$. Moreover, we show the conditional mean (red); and the 2.5% and 97.5% conditional quantiles (red dashed).

we know $u$. We define $\mathcal{I} := \{i : \ x_i > u\}$ as the indices of exceedances of $u$ by the first component.

In the second step, we calculate the likelihood of the model. We note that it is not possible to evaluate the likelihood without imposing any constraints on the residual distribution $H_{|1}$. We follow the approach of Heffernan and Tawn (2004) who propose to temporarily assume that the residual distribution $Z_{|1}$ is Gaussian with unknown mean $\mu_{2|1}$ and unknown variance $\sigma_{2|1}^2$. Under this assumption, we evaluate the log-likelihood $l$ as follows:

$$l\left(\alpha_{2|1}, \beta_{2|1}, \mu_{2|1}, \sigma_{2|1}^2; \mathcal{D}\right) = -\frac{|\mathcal{I}|}{2}\log(2\pi\sigma_{2|1}^2) - \sum_{i \in \mathcal{I}} \frac{\left(y_i - \alpha_{2|1}x_i - \mu_{2|1}x_i^{\beta_{2|1}}\right)^2}{2\sigma_{2|1}^2 x_i^{2\beta_{2|1}}},$$

where $|\mathcal{I}|$ is the cardinality of $\mathcal{I}$. It is now straightforward to infer the four parameters of the model under either frequentist or Bayesian ideologies. Denote with $(\hat{\alpha}_{2|1}, \hat{\beta}_{2|1}, \hat{\mu}_{2|1}, \hat{\sigma}_{2|1}^2)$ our estimates of the parameters of the HT model conditional on data $\mathcal{D}$. We now discard the Gaussianity assumption on the residual distribution and its distribution is estimated with a kernel density of the observations $\{z_i : \ i \in \mathcal{I}\}$, where

$$z_i := \frac{y_i - \hat{\alpha}_{2|1}x_i - \hat{\mu}_{2|1}x^{\hat{\beta}_{2|1}}}{\sqrt{\hat{\sigma}_{2|1}^2} x_i^{\hat{\beta}_{2|i}}}.$$

For higher dimensional HT models, inference is performed separately for each pair $(\alpha_{j|i}, \beta_{j|i})$ but is otherwise identical to the above. The only real difference in the inference procedure is the estimation of the multivariate residual distribution. Heffernan

and Tawn (2004) propose to non-parametrically estimate the multivariate residual distribution. However, when $d$ is large, there might not be enough observations for a non-parametric estimate to perform well due to the curse of dimensionality, and we might need to resort to (semi) parametric estimates: Lugrin et al. (2016) assume a mixture of Gaussian distributions, Towe et al. (2016) assume a Gaussian copula but estimate the marginals non-parametrically, and Wadsworth and Tawn (2022); Shooter et al. (2021); Richards et al. (2021) assume a Gaussian process with delta-Laplace margins.

### 2.4.5 Limit sets

In this section, we model the extremes of a random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)$ with density $f_{\mathbf{Z}}$ and distribution function $F_{\mathbf{Z}}$ by considering the geometrical shape of a sample from this random variable of size $n$ where $n$ tends to infinity. We first give some intuition and after we present the underlying theory.

For ease of presentation, we assume that each component $Z_j$ of $\mathbf{Z}$ has a standard exponential distribution - which if not arising naturally can be achieved with the probability integral transform. Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be a sequence of independent and identically distributed random variables with distribution function $F_{\mathbf{Z}}$. For each $i \geq 1$, we write $\mathbf{Z}_i := (Z_{i1}, \ldots, Z_{id})$. In this section, we present recent work that links features of the multivariate extremes of $\mathbf{Z}$ with the geometrical properties of the random shape of $\{\mathbf{Z}_i : i = 1, \ldots, n\}$ as $n$ tends to infinity.

Without any scalings, the random shape converges onto the set $\{\mathbf{z} \in \mathbb{R}^d : f_{\mathbf{Z}}(\mathbf{z}) > 0\}$. However, the question of the limiting random shape can be more revealing about the extremal dependence structure when for each $n \geq 1$, $\mathbf{Z}_i$ is scaled with $r_n := \log n$. This choice of scaling is made because $\max\{Z_{ij}/r_n : i = 1, \ldots, n\}$ converges in probability to 1 for each $j = 1, \ldots, d$, which implies that when we mathematically define the limit of the random shape of the $n$-sample point cloud $N_n := \{\mathbf{Z}_i/r_n : i = 1, \ldots, n\}$, then it must be a subset of $[0, 1]^d$ and for each component, its component-

wise maxima must be equal to 1: for each $j = 1, \ldots, d$

$$\mathbb{P}\left(\max\{Z_{ij} : \ i = 1, \ldots, n\} \le z \log n\right) = \left(1 - \exp\{-z \log n\}\right)^n$$

$$\text{as } n \to \infty \ \begin{cases} 1 & \text{for } z > 1, \\[2mm] \exp\{-1\} & \text{for } z = 1, \\[2mm] 0 & \text{for } 0 \le z < 1. \end{cases}$$

Mathematically, we define convergence of an $n$-sample point cloud onto a set as follows.

**Definition 2.4.6** (Nolde (2014)). *Let $D_L \subset \mathbb{R}^d$ be compact. We say that the $n$-sample point cloud $N_n$ converges onto $D_L$ if and only if the following two conditions hold.*

1. *For any open set $U \supset D_L$, we have that*

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}(\mathbf{Z}_i/r_n \notin U) > 0\right) = 0.$$

2. *For any $\varepsilon > 0$, $m \ge 1$, $\mathbf{p} \in D_L$, we have that*

$$\lim_{n\to\infty} \mathbb{P}\left\{\sum_{i=1}^n \mathbb{1}(\mathbf{Z}_i/r_n \in \mathbf{p} + \varepsilon B) \ge m\right\} = 1$$

*where $B$ denotes the unit Euclidean ball in $\mathbb{R}^d$.*

*$D_L$ is called the limit set of $N_n$.*

In Definition 2.4.6, Condition 1 ensures that as $n \to \infty$, no members of $N_n$ are outside any open set that is slightly bigger than the limit set. Condition 2. states that for each $\mathbf{p} \in D_L$, the number of points that are distance $\varepsilon > 0$ away from $\mathbf{p}$ grows to infinity when $n$ tends to infinity, so that the limit set is densely filled.

Limit sets can take on a variety of different forms. In line with Nolde (2014), we only consider 'nice' limit sets, which are defined by asserting that the interior of each nice limit set $D_L \subseteq \mathbb{R}^d$ can be represented by a continuous gauge function $g_{D_L}$ that is homogeneous of order 1 - so $g_{D_L}(t\mathbf{x}) = t g_{D_L}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$, $t \in \mathbb{R}$ - as follows: the interior of $D_L$ is equal to $\{\mathbf{x} \in \mathbb{R}^d : \ g_{D_L}(\mathbf{x}) < 1\}$.

Under these assumptions, it is straightforward to calculate the gauge function of the limit set of $\mathbf{Z}$ with the following result.

**Theorem 2.4.7** (Nolde (2014); Simpson (2019)). *Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be i.i.d. random vectors on $\mathbb{R}^d$ with a continuous density $f$ on standard exponential margins. Suppose now there exists a function $\tilde{g}$ on $\mathbb{R}^d_{>0}$ such that*

$$\frac{-\log f(t\mathbf{u})}{t} \to \tilde{g}(\mathbf{u}), \quad t \to \infty, \quad \mathbf{u} \in \mathbb{R}^d_{>0},$$

*then $N_n$ converges onto a limit set $D_L$ whose gauge function is given by $g_{D_L} = \tilde{g}$.*

So one can evaluate the gauge function of the limit set if the density is known. A particularly interesting property of the gauge function, is that it is straightforward to evaluate the coefficient of tail dependence $\eta$:

**Theorem 2.4.8** (Theorem 2.1 in Nolde (2014)). *Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be i.i.d. $\mathbb{R}^d$-valued random vectors with a continuous density $f$ on standard exponential margins. Assume that the $n$-sample point cloud $N_n$ converges onto a limit set $D_L$ with gauge function $g_{D_L}$. Then, the coefficient of tail dependence $\eta$ of $\mathbf{Z}_1$ is given by*

$$\eta = \min\{x > 0 : \ D_L \cap (x, \infty)^d = \varnothing\}$$

Nolde and Wadsworth (2021) show that extremal dependence measures $\lambda(\boldsymbol{\omega})$, $\tau_C(\delta)$ and the HT model parameters $\alpha$ and $\beta$ can be evaluated for random variables using the gauge function of their limit sets.

**Example**

Let $\mathbf{X}$ be a random variable on standard exponential margins that are linked with a $d$-dimensional Gaussian copula that has correlation matrix $R = (R_{ij})^d_{i,j=1}$. Then, it can be shown that the gauge function $g$ of the limit set of $\mathbf{X}$ is given by

$$g(\mathbf{x}) =: \lim_{t \to \infty} \frac{-\log f(tx_1, \ldots, tx_d)}{t} = \sqrt{\mathbf{x}'} R^{-1} \sqrt{\mathbf{x}}, \quad \text{for } \mathbf{x} \geq 0,$$

where $\sqrt{\mathbf{x}} = (\sqrt{x_1}, \ldots, \sqrt{x_d})$. In particular, if $d = 2$ and $\rho = R_{21} = R_{12}$, then the gauge function simplifies to

$$g(x, y) = \left(1 - \rho^2\right)^{-1} \left(x + y - 2\rho\sqrt{xy}\right).$$

In this case, the limit set $D_L$ of the $n$-sample point cloud is given by $\{(x, y) \in [0, \infty)^2 :$ $g(x, y) \leq 1\}$. An analytical description of the equality $g(x, y) = 1$ can be solved analytically by isolating the square-root and solving a quadratic equation. The solution $x(y, \rho)$ is given by

$$x = \rho^2(2y - 1) - y + 1 \pm 2\rho\sqrt{(1 - \rho^2)(1 - y)y}. \qquad (2.4.9)$$

For increased intuition, we plot a sample of size $n$ from a bivariate Gaussian copula with correlation parameter $\rho = 0.5$ and exponential margins that is scaled with $\log n$ in Figure 2.4.4. The boundary of the limit set $\{(x, y) \in [0, \infty)^2 : g(x, y) = 1\}$ is plotted in red. As the sample size is finite and the density is positive everywhere on $[0, \infty)^2$, it is clear that there is a chance with probability greater than 0 that the boundary of the limit set is exceeded. This is also observed for exactly 4 observations in this specific simulation. Finally, from the shape of the limit set, it is straightforward to derive that $\eta = 0.75 \ (= (1 + \rho)/2)$.
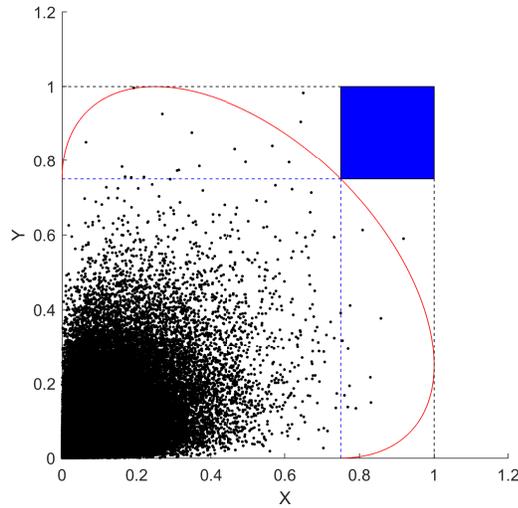


Figure 2.4.4: The $n$-sample point cloud $N_n$ with $(X, Y)$ following a Gaussian copula with exponential margins. The contour in red is the set of $(x, y)$ for which $g(x, y) = 1$ holds, where $g$ is the gauge function of the limit set, see equation (2.4.9). The coefficient of tail dependence $\eta$ is visualised graphically using the blue block in the top-right.

# Chapter 3

# Modelling the extremes of bivariate mixtures

## 3.1   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
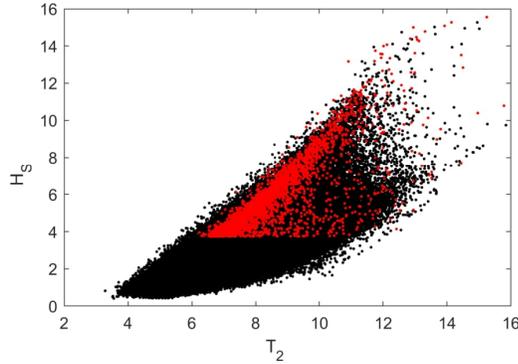
Figure 3.1.1: Significant wave height $H_S$ and wave period $T_2$ from a northern North Sea location. Black dots: data for 1957 - 2018; red dots: corresponding storm peak data.

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 3.1.1. 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 ob-

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

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

exists and 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 a 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 3.1.1, 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 3.2.2, 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 $K = 1, 2, 3, \ldots$ 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 $2^d - 1$ subsets the limiting spectral measure has mass. If we let $d = 2$ and condition on one variable being large, the resulting two subsets 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 and other supporting information is found in Appendix B. Code and data are published at

`https://github.com/stantendijck/HTMixtureModel`.

## 3.2  Framework

### 3.2.1  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 (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 $\alpha$ and $\beta$ with $|\alpha| \le 1$, $\beta < 1$ such that the normalisation

$$Z_x := \frac{Y - \alpha X}{X^\beta} \Big| \{X = x\} \tag{3.2.1}$$

of the conditional random variable $Y|\{X = x\}$, converges in distribution to a non-

degenerate random variable $Z$ with distribution function $G(z)$ as $x \to \infty$. 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 $z \in \mathbb{R}$,

$$\lim_{u \to \infty} \mathbb{P} \left( \frac{Y - \alpha X}{X^\beta} \leq z, \ X - u > x \mid X > u \right) = G(z)e^{-x}. \tag{3.2.2}$$

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

Inference is performed by assuming that the limiting relation (3.2.2) holds above a finite threshold level $u$. Specifically,

$$Y|(X > u) \stackrel{\mathcal{D}}{=} \alpha X + X^\beta(\mu + \sigma \tilde{Z}), \tag{3.2.3}$$

where $\stackrel{\mathcal{D}}{=}$ represents equality in distribution, $\tilde{Z}$ is the standardised residual random variable, i.e., $\tilde{Z} = (Z - \mu)/\sigma$, with $\mu$ the mean of $Z$ and $\sigma > 0$ the standard deviation. The model parameters $\alpha$ and $\beta$ 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 $\alpha$ and $\beta$ 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 $X = u$ converge to the same value as $u$ 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 Lugrin et al. (2021). Previously, we assumed that relation (3.2.3) 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 (3.2.1) to

$$Z_x := \frac{Y - \alpha X - \gamma(X)}{X^\beta} | \{X = x\}. \tag{3.2.4}$$

By taking $\gamma : \mathbb{R} \to \mathbb{R}$ to be a function such that $\gamma(x) = o(x^\beta)$ for $\beta \geq 0$ as $x \to \infty$, this is equivalent to the HT model in the limit as $x \to \infty$. 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 $\beta > 0$.

## 3.2.2 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 $\lim_{z \to -\infty} G(z) \neq 0$. Here, different choices for $\alpha$ and $\beta$ also lead to non-degenerate $G(z)$ but with $\lim_{z \to \infty} G(z) \neq 1$. The findings from this example motivate our developments in Section 3.2.3. 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 $\theta_1, \theta_2 \in (0, 1)$, $\kappa \in (0, 1)$, and distribution function

$$P(X_A \leq x, Y_A \leq y) = \exp\left\{-\theta_1 t_x - \theta_2 t_y - \left[((1 - \theta_1)t_x)^{1/\kappa} + ((1 - \theta_2)t_y)^{1/\kappa}\right]^\kappa\right\} \tag{3.2.5}$$

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 3.2.1, data simulated from this distribution, conditional on $X_A > 2$ with $\theta_1 = \theta_2 = \kappa = 0.5$, 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 (Papastathopoulos et al., 2017), 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 $G(z)$

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

Note that $G(z)$ puts weight $\theta_1$ on $\{-\infty\}$. The second limit corresponds to the lower arm with $(\alpha, \beta) = (0, 0)$ and $G(z) = \theta_1 e^{-t_z}$ 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 3.2.1 (right) illustrates simulated data and implied conditional quantiles from the model we develop in Section 3.3 when fitted to the data shown in Figure 3.2.1 (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.

### 3.2.3   Heffernan-Tawn mixture model

The Heffernan-Tawn model is applicable to the example from Section 3.2.2 with $\lim_{z \to -\infty} G(z) \neq 0$. 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 characteristing 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 these considerations, we introduce the Heffernan-Tawn mixture (HTM) model to better characterise distributions such as the one discussed in Section 3.2.2, conditional on one of the variables being large.
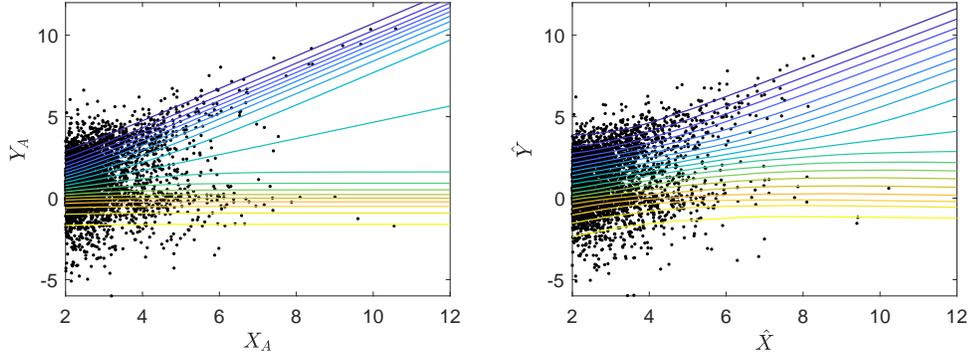
Figure 3.2.1: (Left) Data simulated from model (3.2.5), conditional on $X_A > 2$ with $\theta_1 = \theta_2 = \kappa = 0.5$ including true conditional quantile functions; (Right) Data simulated from the inferred model (3.2.7), fitted to the data in the left plot, with an identical sample size, including the implied conditional quantile functions.

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,

$$F(x, y) = \sum_{k=1}^{K} p_k F_k(x, y), \tag{3.2.6}$$

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 (3.2.2) holds for a single pair $(\alpha_k, \beta_k)$ with residual distribution $G_k$ placing no mass at $\{-\infty, +\infty\}$. Thus the asymmetric logistic copula (3.2.5) cannot be $F_k$. As we assume the distributional form (3.2.6) 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 (3.2.4) 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 (3.2.3), we define the $K$ component HTM model as

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

$$(3.2.7)$$

for large $u$, where $\tilde{Z}_k$, which only exists with probability $p_k$, follows a non-degenerate distribution for each $k$. Moreover, we assume 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 at $\{\pm\infty\}$. 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 Appendix B.5.2. We impose the Keef et al. constraints on all pairs $(\alpha_k, \beta_k)$ separately. For $k = 1, \ldots, K$, $\alpha_k$ and $\beta_k$ are such that for $x > 0$ and $z \in \mathbb{R}$

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

where

$$G_k(z) := \sum_{i=1}^{k-1} p_i + p_k \cdot H_k(z)$$

with the convention $\sum_{i=1}^{0} p_i := 0$, 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 (3.2.5) 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.

### 3.2.4 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

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

with $p \geq 1$ parameters in the parameter space $\Omega$. We are interested in a parametric form for $q(\cdot \mid \boldsymbol{\omega}_\tau)$. Motivated by the HT mixture model (3.2.7), we employ the model

$$q(x|\boldsymbol{\omega}_\tau) = \gamma(\tau) + \alpha(\tau)x + \zeta_\tau x^{\beta(\tau)}, \tag{3.2.8}$$

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 $\zeta_\tau$ is a one-to-one function of quantiles of all residual distributions $Z_k$, $k = 1, \ldots, K$, i.e.,

$$\zeta_\tau := H^{-1}_{\max\{k: \ \tau \leq \sum_{i=1}^{k} p_i\}} \left( \frac{\zeta_\tau - \sum_{i=1}^{k-1} p_i}{p_k} \right),$$

where we define $\sum_{i=1}^{0} p_i = 0$.

Let $(X, Y)$ be a bivariate random variable following model (3.2.7) and define $Y_k$, $k = 1, \ldots, K$, as the random variable representing the $k$th mixture component of the model, i.e., $Y_k|(X > u) \stackrel{\mathcal{D}}{=} \gamma_k + \alpha_k X + X^\beta(\mu_k + \sigma_k \tilde{Z}_k)$. Furthermore, assume $Y_k \perp Y_{k'}$ for $k \neq k'$. Then

$$\lim_{u \to \infty} \mathbb{P}(Y_1 < Y_2 < \cdots < Y_K \mid X > u) = 1, \tag{3.2.9}$$

i.e., the mixture components separate completely in the limit, and as $u \to \infty$, we get that the parameter functions $\gamma(\tau)$, $\alpha(\tau)$ and $\beta(\tau)$ in equation (3.2.8) are piecewise constant in $\tau$. 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 $\zeta_\tau$ is an increasing function of $\tau$ within each interval of $\tau$ where $\gamma(\tau)$, $\alpha(\tau)$ and $\beta(\tau)$ remain constant.

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

$$\Big\{(q_{\tau_1}, \ldots, q_{\tau_m}) : \mathbb{R} \to \mathbb{R}^m : \ q_{\tau_i}(x) = q(x|\boldsymbol{\omega}_{\tau_i}) \text{ for } i = 1, \ldots, m,$$

$$(\boldsymbol{\omega}_{\tau_1}, \ldots, \boldsymbol{\omega}_{\tau_m}) \in \Omega_m \subseteq \mathbb{R}^{pm}\Big\},$$

where $\Omega_m$ is the parameter space, not necessarily equal to the Cartesian product $\Omega^m$. In particular, we consider models with $\boldsymbol{\omega}_\tau = (\boldsymbol{\varphi}, \zeta_\tau)$ where $\boldsymbol{\varphi}$ is common across $\boldsymbol{\omega}_{\tau'}$ for all $\tau'$ and $\zeta_\tau$ is specific to a particular $\tau$. As an illustration, consider model (3.2.7) 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 $Z_1$.

## 3.3   Inference

### 3.3.1   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 the marignal 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 Appendix B.1-B.2.

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 $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)})$ 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 (3.2.7) 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 $J \in \{1, 2, \ldots, K\}$ 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 $Z_k$, $k = 1, \ldots, K$. 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 $k = 1, \ldots, K$. 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 $k = 1, \ldots, K$

$$\mathbb{P}\left(J = k | X = x, Y = y, \boldsymbol{\theta}\right) := P_k\left(x, y | \boldsymbol{\theta}^{(k)}\right) \Big/ \sum_{\kappa=1}^{K} P_\kappa\left(x, y | \boldsymbol{\theta}^{(\kappa)}\right), \quad (3.3.1)$$

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

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

We can now simulate index $J = j_i$ for each observation $(x_i, y_i)$ using equation (3.3.1), and calculate the augmented log-likelihood $l^{(a)}$ via

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

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

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

if $\sigma_k > 0$ and $(\alpha_k, \beta_k)$ satisfy the Keef et al. 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 3.3.2.

For each draw $\hat{\boldsymbol{\theta}}$ from the MCMC chain, we simulate $\{j_i\}_{i=1}^{n}$ each with probability (3.3.1), 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 3.2.3, 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 $k$. Appendix B.2 details estimation of probabilities of extreme sets for this model.

Selecting an optimal value of $K$ is challenging. In simulation studies discussed in Appendix B.6-B.7, 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 $\alpha_i$s are relatively distinct, this technique is reasonable.

## 3.3.2 Incorporating prior probability on asymptotic dependence

Placing a uniform prior over the parameter space (Section 3.3.1) 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 occuring together, see the simulation study in Appendix B.7 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 at the event $\{\alpha_K = 1\}$. For brevity, we consider

model (3.2.7) with $K = 1$.

We define the priors of the parameters $\gamma_1$, $\beta_1$, $\mu_1$ and $\sigma_1$ to be the (improper) uniform distribution on the parameter space with independent components. Let $\delta$ be the Dirac delta function, then the density function of the prior on $\alpha_1$ is

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

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 $\gamma_1$, $\beta_1$, $\mu_1$ and $\sigma_1$ with some fixed standard deviation $h$. We specify the proposal distribution for $\alpha_1$ 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 $\varphi$ and $\Phi$ are the density and distribution function, respectively, of a standard Gaussian. The Metropolis-Hastings acceptance ratio $\alpha_{MH}$ given a proposal $\boldsymbol{\theta}^{(t+1)}$ can now be derived using standard techniques, see Appendix B.5.1.

### 3.3.3 Inference for the quantile-regression model

We consider two data generating frameworks, namely models (3.2.3) and (3.2.7), and we show how to use quantile-regression techniques to infer the parameters without distributional assumptions. In Section 3.2.4, we parameterised the conditional quantile function $q_\tau$ via the parameter vector $\boldsymbol{\omega}_\tau$. First consider a single value of $\tau$, then $\boldsymbol{\omega}_\tau$ may be inferred as

$$\hat{\boldsymbol{\omega}}_\tau := \operatorname{argmin}_{\boldsymbol{\omega}_\tau \in \Omega} \left\{\sum_{i=1}^n \rho_\tau(y_i - q(x_i|\boldsymbol{\omega}_\tau))\right\}, \tag{3.3.2}$$

where the check function $\rho_\tau : \mathbb{R} \to \mathbb{R}$ 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 $0 < \tau_1 < \cdots < \tau_m < 1$. We infer conditional quantile functions $q_{\tau_j}$, $j = 1, \ldots, m$, jointly using

$$\hat{\boldsymbol{\omega}} = \operatorname{argmin}_{(\boldsymbol{\omega}_{\tau_1}, \ldots, \boldsymbol{\omega}_{\tau_m}) \in \Omega_m} \sum_{j=1}^{m} c_j \sum_{i=1}^{n} \rho_{\tau_j}(y_i - q(x_i \mid \boldsymbol{\omega}_{\tau_j})) \qquad (3.3.3)$$

for weights $c_j > 0$ specified by the user, see Bondell et al. (2010). We apply the methodology to equidistant $\tau_j$ on the range $[\tau_1, \tau_m] \subset (0, 1)$, for which it is natural to let $c_j := 1$ for all $j = 1, \ldots, m$. If $\tau_j$ are not equidistant, then the choice for the weights $c_j$ can be adjusted to reflect this. If the parameter space $\Omega_m$ was equal to the Cartesian product of the marginal parameter spaces $\Omega^m$, then the joint estimation procedure would be equivalent to applying equation (3.3.2) separately for each quantile. We only consider models with $\boldsymbol{\omega}_\tau = (\boldsymbol{\varphi}, \zeta_\tau)$ where $\boldsymbol{\varphi}$ is common across $\boldsymbol{\omega}_{\tau'}$ for all $\tau'$ and $\zeta_\tau$ is specific to $\tau$.

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

$$\hat{\zeta}_\tau := \operatorname{argmin}_{\zeta \in \mathbb{R}} \left\{ \sum_{i=1}^{n} \rho_\tau(y_i - q(x_i | (\hat{\boldsymbol{\varphi}}, \zeta))) : (\hat{\boldsymbol{\varphi}}, \zeta) \in \Omega \right\} \qquad (3.3.4)$$

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

The above framework requires adjustment for inferring mixture model (3.2.7) when the mixture probabilities are unknown. We discuss this for $K = 2$ but it generalises for general $K$. For $\tau_1 < \cdots < \tau_m$, 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 $(\boldsymbol{\omega}_{\tau_1}, \ldots, \boldsymbol{\omega}_{\tau_m})$ of the quantile-regression model are:

$$\boldsymbol{\omega}_{\tau_j} = \boldsymbol{\omega}_{\tau_j}(m_0) = \begin{cases} (\gamma_1, \alpha_1, \beta_1, \zeta_{\tau_j}^1) & \text{if } j \leq m_0, \\ (\gamma_2, \alpha_2, \beta_2, \zeta_{\tau_j}^2) & \text{if } j > m_0, \end{cases} \qquad (3.3.5)$$

where the parameters $\zeta_{\tau_j}^k$ increase over $j$ for $k = 1, 2$. The parametric form (3.2.8) is used for $q(x \mid \boldsymbol{\omega}_\tau)$. 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 $\zeta^1_{\tau_j} = 0$ if $\tau_j = \hat{P}_1/2$ and $\zeta^2_{\tau_j} = 0$ 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 (3.3.3) over $m_0$, i.e.,

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

where $\boldsymbol{\omega}_{\tau_j}(k)$ is as in equation (3.3.5). 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 $\hat{\boldsymbol{\omega}}_{\hat{m}_0} = (\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1, \hat{\gamma}_2, \hat{\alpha}_2, \hat{\beta}_2, \hat{\zeta}^1_{\tau_1}, \ldots, \hat{\zeta}^1_{\tau_{\hat{m}_0}}, \hat{\zeta}^2_{\tau_{\hat{m}_0+1}}, \ldots, \hat{\zeta}^2_{\tau_m})$, where each of $(\hat{\zeta}^1_{\tau_1}, \ldots, \hat{\zeta}^1_{\tau_{\hat{m}_0}})$ and $(\hat{\zeta}^2_{\tau_{\hat{m}_0+1}}, \ldots, \hat{\zeta}^2_{\tau_m})$ are increasing. Estimation of $p_1$ and probabilities of extreme sets under this model are discussed in Appendix B.5.4. As specified by our approach, inferred conditional quantile functions within a mixture component are increasing in $\tau$, since $\zeta_{\tau_j}$ 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 $\tau$ 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 (3.2.8), can result in arbitrarily large biases.

## 3.4 A subasymptotic conditional mixture model

The extension of the HT model developed in Section 3.2.3 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 $p_k(x)$, $k = 1, \ldots, K$ are varying with level $x$. In particular, if there exists a component $k$ with $p_k(x) \to 0$ as $x \to \infty$, then for model (3.2.7) 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 (3.2.7) 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

$$B \sim \text{Bernoulli}(p), \quad X \sim \begin{cases} \text{Exp}(1) & \text{if } B = 0, \\ \text{Exp}(\lambda) & \text{if } B = 1, \end{cases} \quad Y \sim \begin{cases} tX & \text{if } B = 0, \\ \lambda X & \text{if } B = 1, \end{cases} \quad (3.4.1)$$

where $\text{Exp}(\lambda)$ denotes the Exponential distribution with rate $\lambda$. Next, we 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 Appendix B.4 show that for large $X_L$,

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

Thus, for $\alpha_2$ in model (3.2.7), $\alpha_2 = \lambda > 1$, but for $x > 0$, $p_2(x) = \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)x} \to 0$ as $x \to \infty$ with $\alpha_2 = 1 - \lim_{x\to\infty} \log(p_2(x))/x$, which is explained by Theorem 3.4.1 below.

Note that we cannot model the distribution of $(X_L, Y_L)$ well under the framework of model (3.2.7). We extend model (3.2.7) by allowing $p_k$ to be a function of $X$, i.e.,

$$Y|(X > u) \overset{\mathcal{D}}{=} \gamma_k + \alpha_k X + X^{\beta_k}(\mu_k + \sigma_k \tilde{Z}_k), \qquad \text{with probability } p_k(X), \ k = 1, \ldots, K, \tag{3.4.2}$$

where $\alpha_1 < \cdots < \alpha_K$. In contrast with the Heffernan-Tawn mixture model, there could exist $2 \le k_0 \le K$ 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 (3.2.7) if $p_k(x)$ is constant for all $x > u$ and $k = 1, \ldots, K$, 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 3.4.1 provides some insight.

**Theorem 3.4.1.** *To ensure that model (3.4.2) for $Y|(X > u)$ does not contradict with $Y$ having a standard Laplace marginal distribution, for all $k = 1, \ldots, K$ we require*

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

The proof can be found in Appendix A.1. 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 $\alpha_2$ can be larger than 1 as long as it is less than $c$. More generally, we consider the class of parametric functions

$$p_k(x) = \frac{a_k x^{b_k} e^{-c_k x}}{\sum_{j=1}^{K} a_j x^{b_j} e^{-c_j x}}, \tag{3.4.3}$$

where $a_k > 0$, $b_k \in \mathbb{R}$ and $c_k \geq 0$ for all $k = 1, \ldots, K$. For identifiability reasons, we define $c_k$ for $k = 1, \ldots, K$ such that $\min\{c_k : k = 1, \ldots, K\} = 0$ 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 (3.2.7) is a special case when $a_k = p_k$, $b_k = 0$ and $c_k = 0$ for all $k = 1, \ldots, K$; (ii) model (3.4.1) 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 3.4.1 yields

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

for all $1 \leq k \leq K$. So, $|\alpha_k|$ can be larger than 1; (iv) define $\mathcal{J} := \{j = 1, \ldots, K : c_j = 0\}$, and $\tilde{\mathcal{J}} := \{j \in \mathcal{J} : b_j = \max\{b_i : i \in \mathcal{J}\}\}$, then model (3.4.2) is asymptotically equivalent to (3.2.7) 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 $u \to \infty$.

## 3.5 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 Appendix B.3. 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 3.1.1 shows a scatter plot of the original and storm peak samples.

Figure 3.1.1 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 $\text{HT}(K)$ and $\text{QR}(K)$ models with $K = 1$ and $K = 2$. We define $\text{QR}(K)$ and $\text{HT}(K)$ 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 kernel 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 $k = 1, \ldots, K$ in the HT mixture model. For trace plots of the model fits, we refer to Appendix B.3.2.

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 Appendix B.2.1 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.
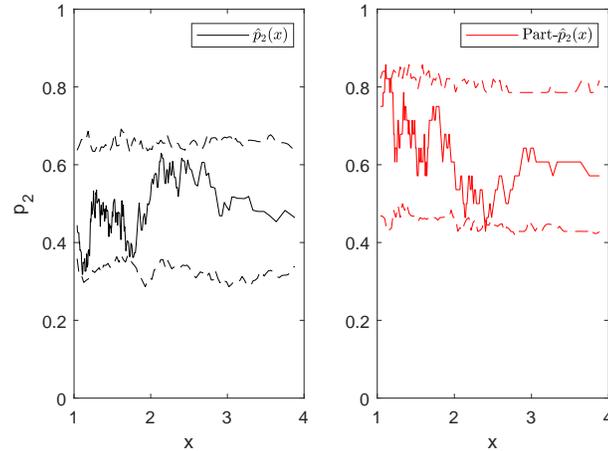
Figure 3.5.1: Estimates $\hat{p}_2(x)$ (left) and Part-$\hat{p}_2(x)$ (right) of $p_2(x)$ for the data. 95% confidence bounds are visualized with dashed lines.

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 3.4 shows this need not be the case. So, we design a goodness-of-fit diagnostic, see Figure 3.5.1, 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 (3.3.1), 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 3.5.1, 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 $\lim_{x \to \infty} p_2(x) = 0$. Hence, the HTM model should be a reasonable approach 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.

| 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 3.5.1: 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 $i \in \{1, 2\}$.

(2020),

$$R := \frac{aH_{S,peak}}{1 + b(T_{2,ass} - c)^2}, \tag{3.5.1}$$

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 (3.5.1) 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 3.5.2 and 3.5.3 and parameter estimates are given in Tables 3.5.1 and 3.5.2.

In Figure 3.5.2, we plot the inferred joint probabilities, see Appendix B.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 3.3.1.

We plot the results of the analysis for $R$ in Figure 3.5.3. The top left corner gives estimated return levels using the 6 inference methods. The remaining panels show the

| 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 3.5.2: 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 $i \in \{1, 2\}$.

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. 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 3.5.1 and 3.5.2, 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 et al. 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 $\alpha$ and $\gamma$ 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
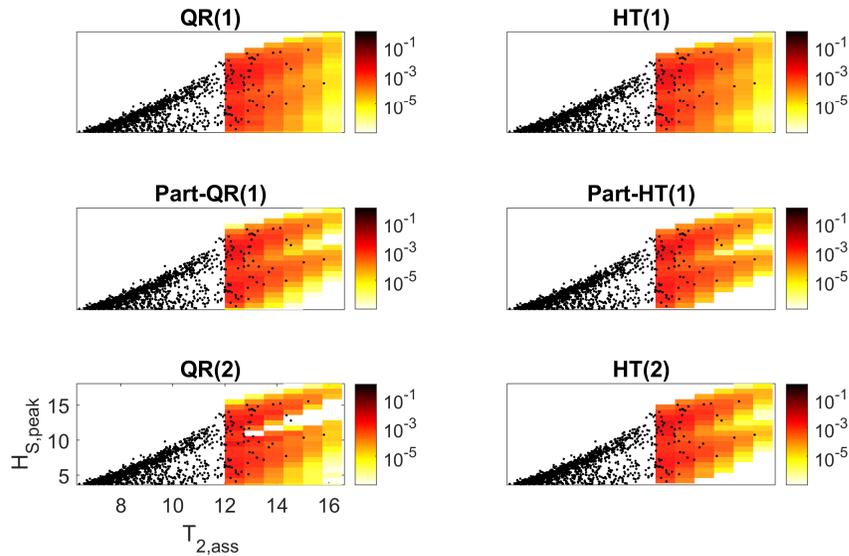
Figure 3.5.2: Estimates of probabilities of extreme sets on original margins using models: QR(1) (top left), HT(1) (top right), the partitioning method using two QR(1) models (middle left), the partitioning model using two HT(1) models (middle right), QR(2) (bottom left), and HT(2) (bottom right). Axis labels and scales are identical.

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

## ACKNOWLEDGEMENTS
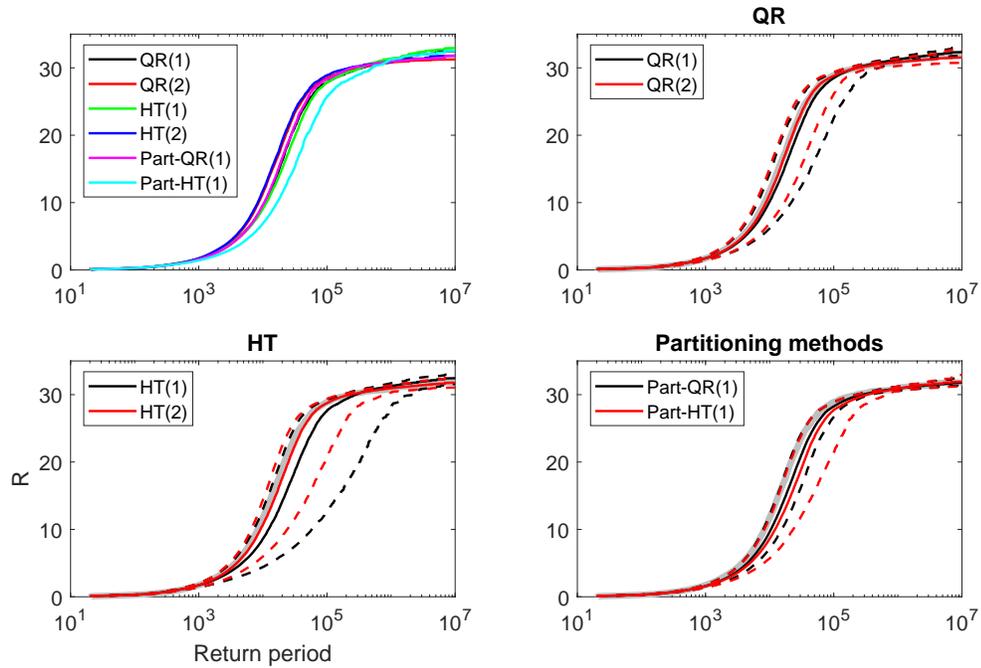
Figure 3.5.3: Top left: return level estimates of the synthetic response $R$ in years using 6 approaches to inference. Other panels: uncertainty analysis for all six methods, similar to Figure 3.5.2. The solid lines are median estimates of the bootstrap ensemble and the dashed lines are the 2.5% and 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.

# Chapter 4

# Extremal characteristics of conditional models

## 4.1 Introduction

Extreme value theory is a topic of growing interest because of its many important applications in for example risk management (Embrechts et al., 1999) or ocean engineering (Castillo et al., 2005). For instance, in the design or assessment of offshore facilities it is crucial to understand the distribution of extreme sea states. Such extreme sea states are quantified in terms of extreme wave heights, wave periods possibly associated with resonant frequencies, and extreme wind speeds. In risk management, it is important to identify which stocks are likely to suffer extreme losses simultaneously, and to which extent this might happen. In general, we need to use well-estabilished extreme value methods to model such events. Traditionally, such multivariate extreme value methods are composed of marginal models and a dependence copula, each having parametric forms for the tails.

In other areas of statistics, however, it is common to use conditional models for multidimensional data. Intuitively, this is the most sensible approach. We observe $X$ that partially explains $Y$. So, we define a model for $X$ and a model for $Y$ conditional on $X$. There exist many examples in the literature of models within this conditional framework with applications in extremes, e.g., the conditional extreme

value model (Heffernan and Tawn, 2004; Fougeres and Soulier, 2012), the Weibull-log normal distribution (Haver and Winterstein, 2009, henceforth the Haver-Winterstein distribution), and hierarchical models (Eastoe, 2019). Although conditional models are easy to interpret, it can be rather difficult to study the extremes of both $Y$ and $(X, Y)$ within this class. Recently, Engelke and Hitz (2020) developed graphical models for extremes. However, we do not know of any literature that links existing conditional models directly to extremal dependence measures.

There are two extremal dependence measures that are key in identifying and measuring the degree of asymptotic dependence or asymptotic independence (Coles et al., 1999). Identifying the correct asymptotic dependence class is important since extrapolation of models from different classes is different. To define asymptotic dependence, we first define $\chi \in [0, 1]$, with

$$\chi := \lim_{p \uparrow 1} \chi(p) := \lim_{p \uparrow 1} \mathbb{P} \left\{ Y > F_Y^{-1}(p) \mid X > F_X^{-1}(p) \right\}, \tag{4.1.1}$$

assuming this limit exists, where $F_X$ and $F_Y$ denote the marginal distribution functions of $X$ and $Y$. We say that these random variables are asymptotically dependent if $\chi > 0$, i.e., when the joint probability that both random variables are large is of the same magnitude as when one is large. If the coefficient of asymptotic dependence $\chi = 0$, we say that the variables are asymptotically independent. In this case, $\chi$ does not give us information on the level of asymptotic independence. So, we additionally define the coefficient of asymptotic independence $\eta \in (0, 1]$ (Ledford and Tawn, 1996). This coefficient describes the rate of decay to zero of the joint exceedance probability $\mathbb{P}\{X > F_X^{-1}(p), \, Y > F_Y^{-1}(p)\}$ as $p$ tends to 1. More specifically, $\eta$ is defined to satisfy

$$\mathbb{P} \left\{ X > F_X^{-1} \left[ F_E(u) \right], \, Y > F_Y^{-1} \left[ F_E(u) \right] \right\} \sim \mathcal{L} \left( e^u \right) e^{-u/\eta} \tag{4.1.2}$$

as $u \to \infty$, where $F_E(u) = 1 - \exp(-u)$ is the distribution function of a standard exponential, and where $\mathcal{L}$ is a slowly varying function. Here, we write $f(x) \sim g(x)$ as $x \to \infty$ when $f(x)/g(x) \to 1$ as $x \to \infty$. We rewrite definition (4.1.2) as

$$\eta := \lim_{p \uparrow 1} \eta(p) := \lim_{p \uparrow 1} \frac{\log(1 - p)}{\log \left[ (1 - p)\chi(p) \right]}. \tag{4.1.3}$$

If the variables are asymptotically dependent, then $\eta = 1$; if the variables are asymptotically independent, then $\eta \in (0, 1)$ or $\eta = 1$ and $\mathcal{L}(u) \to 0$ as $u \to \infty$.

Evaluating $\chi$ for a bivariate random variable $(X, Y)$ is relatively straightforward. First, define for each $z \in \mathbb{R}$,

$$\overline{H}(z) := \lim_{p \uparrow 1} \mathbb{P}\left( \log\left( \frac{1 - F_X(X)}{1 - F_Y(Y)} \right) > z \,\Big|\, F_X(X) > p \right).$$

Although this formulation looks complex, it is simply an analogue of the spectral measure (Engelke and Hitz, 2020) in Fréchet margins but here it is expressed as a representation in exponential margins, see Section 4.4. We then apply the dominated convergence theorem to get

$$\chi = \int_0^\infty \overline{H}(-x)e^{-x}\, \mathrm{d}x.$$

In particular, $\chi > 0$ if and only if $\lim_{z \to -\infty} \overline{H}(z) > 0$.

Additionally calculating $\eta$ is straightforward for distributions when the joint distribution function is specified parametrically, e.g., a bivariate extreme value distribution (Ledford and Tawn, 1996), or when the joint density function is specified parametrically (Nolde and Wadsworth, 2021), e.g., a multivariate normal distribution. In this paper, we consider models specified within the conditional framework. For these cases, it is hard to calculate $\eta$ analytically, and numerical estimation can be difficult since convergence of $\eta(p)$ to $\eta$ can be exceptionally slow. We set up methodology to calculate $\eta$ in closed form within this framework and demonstrate the techniques on two widely used examples specified below. We support these limiting results using numerical integration.

First, we consider the model described in Haver and Winterstein (2009), used to explain the dependence between extreme significant wave height and their associated wave periods. Secondly, we investigate the model of Heffernan and Tawn (2004). This is a conditional model which describes the distribution of $Y \mid X$ for large $X$, where both $X$ and $Y$ are on standard margins. As the Heffernan-Tawn model focusses on normalising the distribution of $Y|X = x$ as $x \to \infty$ to give a non-degenerate limit, it asymptotically focusses on a different aspect of the joint distribution to the

events which determine $\eta$, i.e., $\{X > x,\ Y > x\}$ as $x \to \infty$, when the variables are asymptotically independent. As a consequence, it seems reasonable to expect that the upper tail of $Y|X = x$ for large $x$ does not give $\eta$. We will show by giving an example that there exist distributions that share the same Heffernan-Tawn normalization but do not share the same $\eta$. More theoretical examples, like $Y := X^\beta Z$ and $Y := |Z|^{|X|}$ where $Z$ is some random variable independent of $X$, can be found in Appendix D.5.

The layout of the article is as follows. In Section 4.2, we demonstrate novel techniques for calculating the coefficient of asymptotic independence $\eta$ and illustrate the techniques with some examples. In Sections 4.3 and 4.4, we apply these techniques to the Haver-Winterstein model and the Heffernan-Tawn model, respectively. Proofs are found in Appendix C-D.

## 4.2   Methodology

### 4.2.1   Motivation

We aim to investigate the extremal properties of the bivariate distribution of $(X, Y)$, for which the distribution of $X$ and the distribution of $Y \mid X$ are specified. In particular, we aim to investigate the tail of the distribution of $Y$ and joint extremes of $X$ and $Y$ via the coefficient of asymptotic independence $\eta$. Deriving such extremal quantities in closed form within this class is not trivial. In this section, we provide a set of tools, derived from the Laplace approximation, to calculate such properties for any conditional model.

First, we consider the tail of the distribution of $Y$. Because the distributions of $X$ and $Y \mid X$ are specified, it is natural to write

$$1 - F_Y(y) := \mathbb{P}(Y > y) = \int_{-\infty}^{\infty} \mathbb{P}(Y > y \mid X = x) f_X(x)\, \mathrm{d}x,$$

where $f_X$ is the density of $X$. In general, this integral is analytically intractable. In Section 4.2.2, we present the tools with which we can derive the asymptotic properties of this integral as $y$ tends to the upper end point of the distribution of $Y$.

To derive the coefficient of asymptotic independence, we additionally need the inverse distribution $F_Y^{-1}(p)$ for values of $p$ close to 1, and

$$\mathbb{P}(X > F_X^{-1}(p),\ Y > F_Y^{-1}(p)) = \int_{F_X^{-1}(p)}^{\infty} \mathbb{P}(Y > F_Y^{-1}(p) \mid X = x) f_X(x)\, \mathrm{d}x.$$

This integral is also intractable in general; the tools from Section 4.2.2 can again be applied to derive the asymptotic decay to 0 as $p$ tends to 1.

## 4.2.2 Extension to the Laplace approximation

Here we present our theory to calculate asymptotic rates of decay of integrals, that can be used to compute extremal properties, such as $\eta$, of conditional models. We first recall the Laplace approximation, a technique commonly used in Bayesian inference for approximating intractable integrals. This asymptotic approximation forms the basis of our main result. We then state that result, and illustrate key differences with the Laplace approximation by comparing examples.

**Proposition 4.2.1** (Laplace approximation). *Let $a < b$. Suppose $g : [a,b] \to \mathbb{R}$ is twice continuously differentiable and assume there exists a unique $x^* \in (a,b)$ such that $g(x^*) = \max_{x \in [a,b]} g(x)$ and $g''(x^*) < 0$. Then*

$$\int_a^b e^{ng(x) - ng(x^*)}\, \mathrm{d}x \cdot \sqrt{n(-g''(x^*))} \sim \sqrt{2\pi}$$

*as $n \to \infty$.*

The main disadvantage of the Laplace approximation is that it can only be used to approximate integrals where the integrands are of the form $f(x)^n$, where $f(x) = e^{g(x)}$ is a positive function. However, we are interested in calculating integrals with integrand $f_n(x) = e^{g_n(x)}$, for some sequence of functions $\{g_n\}_{n \in \mathbb{N}}$. Now we extend the Laplace approximation under the assumptions that: (i) the analogue $x_n^*$ of $x^*$ is allowed to depend on $n$; (ii) $x_n^*$ can be equal to either $a$ or $b$; (iii) $g_n''(x_n^*)$ does not need to be negative.

**Proposition 4.2.2.** *Let $I \subseteq \mathbb{R}$ be connected with non-zero Lebesgue mass, $k_0 \geq 1$ an integer, and $g_n \in C^{k_0}(I)$ a sequence of real-valued (at least) $k_0$-times continuously*

*differentiable functions defined on $I$. For $1 \le i \le k_0$, we define $g_n^{(i)}$ as the ith derivative of $g_n$. We assume that for all $n \in \mathbb{N}$, there exists a unique $x_n^* \in I$ such that $g_n(x_n^*) > g_n(x)$ for all $x \in I \setminus \{x_n^*\}$. Moreover, we assume that $k_0$ is the smallest integer such that $g_n^{(k_0)}(x_n^*) < 0$ and $\lim_{n\to\infty} g_n^{(i)}(x_n^*)[-g_n^{(k_0)}(x_n^*)]^{-i/k_0} = 0$ for all $1 \le i < k_0$. Additionally, assume that there exists a $\delta > 0$ for which there exists an $\varepsilon > 0$ such that for all $|x| < \delta$*

$$
\lim_{n\to\infty} \frac{g_n^{(k_0)}\left\{ x_n^* + x \left[ -g_n^{(k_0)}(x_n^*) \right]^{-\frac{1}{k_0}} \right\}}{g_n^{(k_0)}(x_n^*)} < 1 + \varepsilon.
$$

*Then, there exists $N \in \mathbb{N}$ such that for $n > N$, there exists a constant $C_1 > 0$ such that*

$$
\int_I e^{g_n(x) - g_n(x_n^*)} \, \mathrm{d}x \cdot \left[ -g_n^{(k_0)}(x_n^*) \right]^{\frac{1}{k_0}} \ge C_1.
$$

The proof of Proposition 4.2.2 can be found in Appendix C.1. One disadvantage of our extension is that it only gives an asymptotic lower bound. In many practical applications, an upper bound can be found directly. For example, in Section 4.3, we can use inequality (4.3.5).

Functions for which Proposition 4.2.2 is applicable include functions $g_n$ with a single mode $x_n^*$ that are approximated well with a Taylor expansion of some order on a large enough neighbourhood of the mode. For example, for $g_n(x) = -|x|^p$ with $p \in \mathbb{R}$, the proposition is applicable if and only if $p \in \mathbb{Z}_{\ge 2}$. We specify further that the first set of assumptions ensures that the $k_0$th order Taylor approximation of $g_n$ around $x_n^*$ has at most two significant terms (the 0th and the $k_0$th term) by setting a limit on the size of the $i$th terms in this Taylor approximation, where $1 \le i \le k_0 - 1$. The second set of assumptions defines if the Taylor approximation is good enough on a neighbourhood of $x_n^*$, see the second example in Section 4.2.3

## 4.2.3   Examples

We demonstrate the use of Proposition 4.2.2 in three cases. Firstly, let $g_n(x) = -nx^m$ for $n \in \mathbb{N}$, $m \in \mathbb{Z}_{\ge 1}$ and $I = [0, \infty)$. It is then valid to apply Proposition 4.2.2 with

$x_n^* = 0$ and $k_0 = m$. Applying the proposition yields a constant $C_1 > 0$ such that for sufficiently large $n$,

$$n^{\frac{1}{m}} \int_0^\infty e^{-nx^m} \, dx \geq C_1.$$

We remark that $C_1$ is not exactly the same as in Proposition 4.2.2 as it has absorbed some constants that do not depend on $n$ from the left hand side. This lower bound is tight for each $m \geq 1$. We verify this by using the variable transformation $y = nx^m$ to give

$$n^{\frac{1}{m}} \int_0^\infty e^{-nx^m} \, dx = \frac{1}{m} \int_0^\infty y^{\frac{1}{m}-1} e^{-y} \, dy = \Gamma\left(\frac{1}{m} + 1\right).$$

After recognizing that the integral over $[0, \infty)$ is equal to half of the integral over $\mathbb{R}$, we see that Proposition 4.2.1 is applicable only when $m = 2$. In this case, Proposition 4.2.1 additionally gives as $n \to \infty$

$$\int_0^\infty e^{-nx^2} \, dx = \frac{1}{2} \int_{-\infty}^\infty e^{-nx^2} \, dx \sim \frac{\sqrt{\pi}}{2\sqrt{n}}.$$

Secondly, let $g_n(x) = -x - nx^2$ and $I = [0, \infty)$. Now Proposition 4.2.1 is not applicable since no function $g(x)$ exists for which $g_n(x) = ng(x)$ holds. Note that Proposition 4.2.2 is also not applicable with $k_0 = 1$, since $x_n^*$ has to be equal to 0 and for $x \neq 0$

$$\lim_{n \to \infty} \frac{g_n'(0 + x \cdot n)}{g_n'(0)} = \lim_{n \to \infty} 1 + 2n^2 x = \infty,$$

contradicting one of the assumptions. Proposition 4.2.2 is applicable with $k_0 = 2$, yielding a constant $C_2 > 0$ such that for sufficiently large $n$,

$$\sqrt{n} \int_{-\infty}^\infty e^{-x - nx^2} \, dx \geq C_2.$$

Similar to our first example, this lower bound is tight since we can also directly calculate as $n \to \infty$

$$\sqrt{n} \int_{-\infty}^\infty e^{-x - nx^2} \, dx = \sqrt{n} \int_{-\infty}^\infty e^{-n\left(x + \frac{1}{2n}\right)^2 + \frac{1}{4n}} \, dx \sim \sqrt{\pi}.$$

Finally, let $\alpha_n > 0$, $\beta_n > 0$ for $n \in \mathbb{N}$ and assume $\liminf \alpha_n > 0$. Define $g_n(x) = \alpha_n \log x - \beta_n x$. Using an argument similar to that in the second example, we see

that Proposition 4.2.1 is not applicable. However Proposition 4.2.2 is applicable with $k_0 = 2$, yielding a constant $C_3 > 0$ such that for sufficiently large $n$,

$$\alpha_n^{-\alpha_n - \frac{1}{2}} \beta_n^{\alpha_n + 1} e^{\alpha_n} \int_0^\infty x^{\alpha_n} e^{-\beta_n x} \, \mathrm{d}x \geq C_3.$$

This bound is also tight, which can be seen from recognizing the density of a gamma distribution in the expression above, and applying limit results for the gamma function.

## 4.3 Haver-Winterstein model

Haver and Winterstein (2009) introduce the Haver-Winterstein (HW) distribution for significant wave height $H_S$ and wave period $T_p$ in the North Sea. Their model is set up in the conditional framework: they specify a class of distributions for $H_S$ and a class of distributions for $T_p \mid H_S$. Variations of this approach have been widely applied in ocean engineering with over 150 citations, 25 of which correspond to 2021, see for example Drago et al. (2013). However we are not aware of any literature quantifying $\chi$ and $\eta$ in closed form for the HW distribution; we now show how to calculate these.

For easiness of notation, let $X = T_p$ and $Y = H_S$. The marginal distribution of the HW is formulated as

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\alpha x} \exp\left\{-\frac{(\log x - \theta)^2}{2\alpha^2}\right\}, & \text{for } 0 < x \leq u, \\ \frac{k}{\lambda^k} x^{k-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, & \text{for } x > u. \end{cases} \tag{4.3.1}$$

where $u, \alpha, k, \lambda > 0$ and $\theta \in \mathbb{R}$. In particular, the parameters are constrained such that $f_X$ is continuous at $u$ and integrates to 1. Secondly, they take $Y \mid X$ to be conditionally log-normal

$$f_{Y|X}(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma(x)y} \exp\left\{-\frac{(\log y - \mu(x))^2}{2\sigma(x)^2}\right\}, \quad \text{for } x, y > 0, \tag{4.3.2}$$

where $\mu(x) := \mu_0 + \mu_1 x^{\mu_2}$ and $\sigma(x) := [\sigma_0 + \sigma_1 \exp(-\sigma_2 x)]^{1/2}$ with $\mu_0 \in \mathbb{R}$, $\mu_1, \mu_2, \sigma_0, \sigma_1, \sigma_2 > 0$.

Model-based parameter estimates (Haver and Winterstein, 2009) from data observed in the northern North Sea are given in Appendix D.2.1. For ease of presentation, we make two assumptions about the parameter space of the HW distribution

that are consistent with parameter estimates $(\hat{\mu}_2, \hat{k}) = (0.225, 1.55)$ from Haver and Winterstein (2009). Specifically, we make the following restrictions: $0 < \mu_2 < 0.5$ and $2\mu_2 < k$. These assumptions reduce the number of cases to be considered significantly whilst including realistic domains for the parameters as considered by practitioners.

We now show how to use Proposition 4.2.2 to calculate the extremal dependence measures $\chi$ and $\eta$ for the bivariate random vector $(X, Y)$ distributed according to the HW distribution in the restricted parameter space. Calculation of $\eta$ is split into two steps. In the first step, we calculate the distribution function $F_Y$ of $Y$ and in the second we evaluate the rate of decay of joint probabilities $\mathbb{P}\{X > F_X^{-1}[F_E(u)], Y > F_Y^{-1}[F_E(u)]\}$ as $u$ tends to infinity.

We have

$$\mathbb{P}(Y > y) = \int_0^\infty \mathbb{P}(Y > y \mid X = x) f_X(x) \, \mathrm{d}x = \int_0^\infty \overline{\Phi}\left(\frac{\log y - \mu(x)}{\sigma(x)}\right) f_X(x) \, \mathrm{d}x, \tag{4.3.3}$$

where $\overline{\Phi}$ is the survival function of a standard Gaussian. This integral is analytically intractable but we can calculate its limiting leading order behaviour as $y \to \infty$ in closed form. Proposition 4.2.2 gives a lower bound and an upper bound of the same order as the lower bound is then found directly. For ease of notation, we denote the integrand by

$$g_y(x) := \overline{\Phi}\left(\frac{\log y - \mu(x)}{\sigma(x)}\right) f_X(x) \tag{4.3.4}$$

for $x > 0$. In Figure 4.3.1, we plot $g_y$ for various values of $y$. From the figure, we note that $g_y$ has two local maxima for suffiiciently large $y$. These are $x_y^*$, which converges to zero, and $x_y^{**}$, which diverges to infinity. This observation implies that we cannot apply Proposition 4.2.2 directly in this case. We therefore proceed as follows: (i) calculate $x_y^*$ and $x_y^{**}$; (ii) partition the interval of integration into intervals $I_1$ and $I_2$, where $x_y^* \in I_1$ and $x_y^{**} \in I_2$, such that the conditions of Proposition 4.2.2 hold for both intervals, and then apply the proposition on each interval; (iii) combine the two lower bounds found to get a lower bound for integral (4.3.3); (iv) derive a limiting upper bound for integral (4.3.3) of the same order as the lower bound.
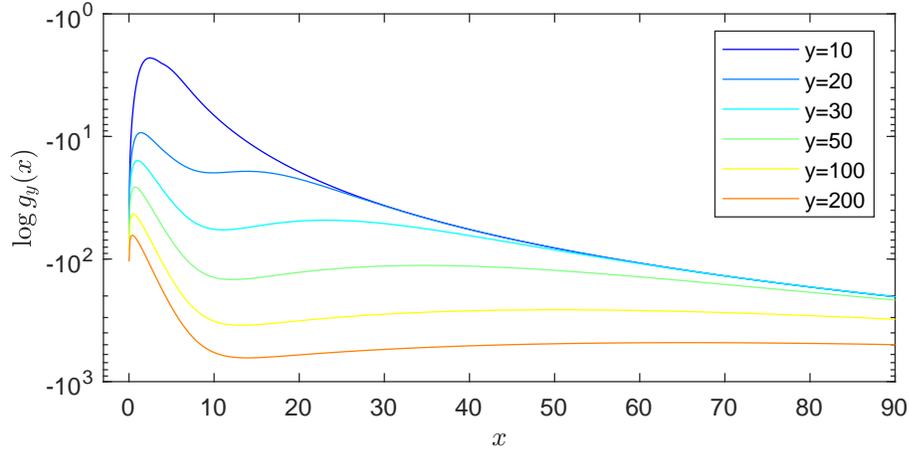
Figure 4.3.1: The function $\log g_y$ from equation (4.3.4) for $y = 10$, 20, 30, 40, 50, 100 with parameters as reported in Haver and Winterstein (2009), see Appendix D.2.1.

In Appendix D.3.1-D.3.2, we derive that as $y \to \infty$

$$x_y^* \sim \left( \frac{\sigma_1 \sigma_2 \cdot \log y}{2\mu_1 \mu_2 (\sigma_0 + \sigma_1)} \right)^{-\frac{1}{1-\mu_2}} \quad \text{and} \quad x_y^{**} \sim \left( \frac{\lambda^k \mu_1 \mu_2 \cdot \log y}{k\sigma_0} \right)^{\frac{1}{k-\mu_2}},$$

where in the calculation of $x_y^*$ we use $0 < \mu_2 < 0.5$. From Figure 4.3.1, we recognize that $g_y(x_y^*) > g_y(x_y^{**})$ as $y \to \infty$. We show that this holds analytically in Appendix D.3.3 when $2\mu_2 < k$. We now apply Proposition 4.2.2 and find that $k_0 = 2$ is appropriate. The proposition then gives a lower bound for integral (4.3.3) around $x_y^*$ as $y \to \infty$ of

$$\mathbb{P}(Y > y) \geq \exp \left\{ -\frac{\log^2 y}{2(\sigma_0 + \sigma_1)} + O(\log y) \right\}.$$

Finally, since $g_y(x_y^*) > g_y(x_y^{**})$, it is straightforward to show as $y \to \infty$ that

$$\mathbb{P}(Y > y) \leq \exp \left\{ -\frac{\log^2 y}{2(\sigma_0 + \sigma_1)} + O(\log y) \right\}$$

using the inequality

$$\mathbb{P}(Y > y \mid X = x) f_X(x) \leq g_y(x_y^*) \mathbb{1}\{x \in [0, x_y^{**}]\} + f_X(x) \mathbb{1}\{x > x_y^{**}\}. \tag{4.3.5}$$

We now can calculate $\eta$ and show that $\chi = 0$. To that end, we first need to calculate the inverse probability integral transform, transforming $Y$ to standard exponential margins; i.e., we need $F_Y^{-1}[F_E(u)]$. Next, we need to evaluate the asymptotic behaviour of $\mathbb{P}\{Y > F_Y^{-1}[F_E(u)], X > F_X^{-1}[F_E(u)]\}$ as $u \to \infty$. To evaluate $F_Y^{-1} \circ F_E$,

we first calculate for $y \to \infty$

$$F_E^{-1}(F_Y(y)) = -\log(1 - F_Y(y)) = \frac{\log^2 y}{2(\sigma_0 + \sigma_1)} + O(\log y).$$

We invert this expression by solving $F_E^{-1}(F_Y(y)) = u$ for $\log y$. This yields $\log y = \sqrt{2(\sigma_0 + \sigma_1)u} + O(1)$ as $u \to \infty$. We can now write down an asymptotic expression for $\chi(u)$ as $u \to \infty$

$$\begin{aligned}
\chi(u) &:= \mathbb{P}\left\{F_E^{-1}\left[F_Y(Y)\right] > u, \ F_E^{-1}\left[F_X(X)\right] > u\right\} \\
&= \mathbb{P}\left\{\log Y > \sqrt{2(\sigma_0 + \sigma_1)u} + O(1), \ (X/\lambda)^k > u\right\} \\
&= \int_{\lambda u^{1/k}}^{\infty} \overline{\Phi}\left(\frac{\sqrt{2(\sigma_0 + \sigma_1)u} + O(1) - \mu(x)}{\sigma(x)} \mid X = x\right) \cdot \frac{kx^{k-1}}{\lambda^k} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\} \, dx.
\end{aligned}$$

In Appendix D.3.4, we show that Proposition 4.2.2 is applicable for this integral with $k_0 = 1$ and $x_u^* = \lambda u^{1/k}$. Moreover, we derive directly an upper bound of the same order, obtaining

$$\chi(u) = \exp\left\{-\left(2 + \frac{\sigma_1}{\sigma_0}\right)u + O\left(u^{1/2 + \mu_2/k}\right)\right\}$$

as $u \to \infty$. Hence, $\chi = 0$ and

$$\eta = \left(2 + \frac{\sigma_1}{\sigma_0}\right)^{-1}.$$

In particular, for the parameter estimates from Haver and Winterstein (2009), the value of $\eta \in (0, 1/2)$ implies that the distribution exhibits negative asymptotic independence (Ledford and Tawn, 1996). This contrasts with the positive correlation of the Haver-Winterstein distribution, which might lead practitioners to assume falsely that the positive correlation also exists in the extremes of the Haver-Winterstein model; this is far from the truth.

What we learn from our work is not necessarily that the Haver-Winterstein model should not be used - we can derive this conclusion in many simpler ways than with this paper. Instead, we can use this example to understand how a conditional model makes complex assumptions on the dependence structure: imposing a positive correlation overall but a highly negative correlation in the extremes.

## 4.4 Heffernan-Tawn model

In multivariate extreme value theory, the conditional extreme value model of Heffernan and Tawn (2004), henceforth denoted the HT model, is widely studied and applied to extrapolate multivariate data. The HT model has been cited over 600 times, and is applied e.g. in oceanography (Ross et al., 2020), finance (Hilal et al., 2011), and spatio-temporal extremes (Simpson and Wadsworth, 2021). The HT model is a limit model and its form is motivated by derived limiting forms from numerous theoretical examples.

Let $(X, Y)$ be a bivariate random variable with standard Laplace margins (Keef et al., 2013) and assume that its joint density exists. Next, assume there exist parameters $\alpha \in [-1, 1]$, $\beta < 1$ and a non-degenerate distribution function $H$ such that for $x > 0$, and for all $z \in \mathbb{R}$ the following limits

$$\lim_{u \to \infty} \mathbb{P} \left( \frac{Y - \alpha X}{X^\beta} \leq z, \ X - u > x \mid X > u \right)$$

and

$$H(z) = \lim_{x \to \infty} \mathbb{P} \left( \frac{Y - \alpha x}{x^\beta} \leq z \mid X = x \right) \tag{4.4.1}$$

exist. This implies, according to l'Hopital's rule, that

$$\lim_{u \to \infty} \mathbb{P} \left( \frac{Y - \alpha X}{X^\beta} \leq z, \ X - u > x \mid X > u \right) = H(z) \exp(-x). \tag{4.4.2}$$

The latter in turn has the interpretation that as $u \to \infty$, $(Y - \alpha X) X^{-\beta}$ and $(X - u)$ are independent conditional on $X > u$, and are distributed as $H$ and as a unit exponential, respectively. As is common practice in extreme value theory, the limit results are assumed to hold above some high threshold. So here, the HT model assumes that the corresponding limiting family in (4.4.1) holds exactly at a finite level $u$ and beyond.

Now, if we additionally assume that a $u > 0$ exists such that for all $x > u$

$$\mathbb{P}(Y > y \mid X = x) = \overline{H} \left( \frac{y - \alpha x}{x^\beta} \right) \tag{4.4.3}$$

holds for all $y \in \mathbb{R}$ where $\overline{H} = 1 - H$ is some non-degenerate survival function. Then, we say that $(X, Y)$ is modelled with the exact version of the HT model.

In this case study, we assume that $(X, Y)$ is modelled with the exact version of the HT model with the additional assumption that $\alpha, \beta \in [0, 1)$. We consider two cases for $H$, corresponding to finite and infinite upper end points. If $H$ has a finite upper end point $z^H$, calculations for $\eta$ are trivial. Indeed, when $X = x$, $Y$ cannot be larger than $\alpha x + x^\beta z^H$. Thus, as $u \to \infty$, $Y > u$ implies $X > u/\alpha + o(u)$. So, as $u \to \infty$

$$\mathbb{P}(X > u, Y > u) \sim \mathbb{P}\left\{X > u, X > u/\alpha + O(u^\beta)\right\}$$
$$\sim \mathbb{P}\left\{X > u/\alpha + O(u^\beta)\right\}$$
$$= \exp\left\{-u/\alpha + O(u^\beta)\right\}.$$

Therefore, $\eta = \alpha$ when $\alpha > 0$ and otherwise does not exist.

Now assume that $H$ has an infinite upper end point. To make calculations tractable, we parameterise $\overline{H}$ as

$$\overline{H}(z) = \exp\left\{-\gamma z^\delta + o\left(z^\delta\right)\right\} \mathbb{1}\{z > 0\} + \mathbb{1}\{z \leq 0\} \tag{4.4.4}$$

for $\gamma > 0$, $\delta \geq 1$. For simplicity, we do not consider potential negative arguments for $\overline{H}$ since the precise form of its lower tail is not relevant to the current work. Parameterisation (4.4.4) covers most non-trivial light-tailed cases for the upper tail including Gaussian, Weibull and exponential tails; see examples in Heffernan and Tawn (2004). It is also the tail model of the delta-Laplace (generalised Gaussian) distribution used in the spatial conditional extremes model from Wadsworth and Tawn (2022). Moreover if the tail of $\overline{H}$ is heavier than that of the exponential, $Y$ cannot possibly follow a standard Laplace distribution. This links to the restricton $\delta \geq 1$. For illustration, we set $o(z^\delta) = 0$ in equation (4.4.4). The resulting Weibull survival function is a suitable choice for $\overline{H}$, since it has an extreme value tail index of 0, but a varying tail thickness controlled by $\delta$.

**Proposition 4.4.1.** *If $(X, Y)$ follows distribution (4.4.3) with $H$ as in (4.4.4) with $o(z^\delta) = 0$, then $\delta \geq (1 - \beta)^{-1}$.*

The proof of Proposition 4.4.1 is found in Appendix C.1. Following similar arguments to those used in the proof of Proposition 4.4.1, we calculate $\eta$ for any combination

| $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\eta$ |
|---|---|---|---|---|
| $(0,1)$ | $[0,1)$ | $(0,\infty)$ | $((1-\beta)^{-1},\infty)$ | $\alpha$ |
| $(0,1)$ | $(0,1)$ | $(0,\infty)$ | $(1-\beta)^{-1}$ | $\left(\frac{\gamma(1-\alpha c)^{\delta}}{c^{\delta-1}}+c\right)^{-1}$ |
| $(0,1)$ | $0$ | $(1/\alpha,\infty)$ | $1$ | $\alpha$ |
| $(0,1)$ | $0$ | $(0,1/\alpha]$ | $1$ | $1/(\gamma+1-\gamma\alpha)$ |
| $0$ | $(0,1)$ | $(0,\infty)$ | $((1-\beta)^{-1},\infty)$ | Not defined |
| $0$ | $(0,1)$ | $(0,(1-\beta)/\beta]$ | $(1-\beta)^{-1}$ | $1/(\gamma+1)$ |
| $0$ | $(0,1)$ | $[(1-\beta)/\beta,\infty)$ | $(1-\beta)^{-1}$ | $\gamma^{-1/\delta}(\delta-1)^{1-1/\delta}/\delta$ |

Table 4.4.1: Values of $\eta$ for model (4.4.3) with $\overline{H}$ as in (4.4.4) for different ranges of parameter combinations, where $c = \max\{1, c_0\} \in [1, 1/\alpha)$ for $c_0$ given in equation (4.4.5).
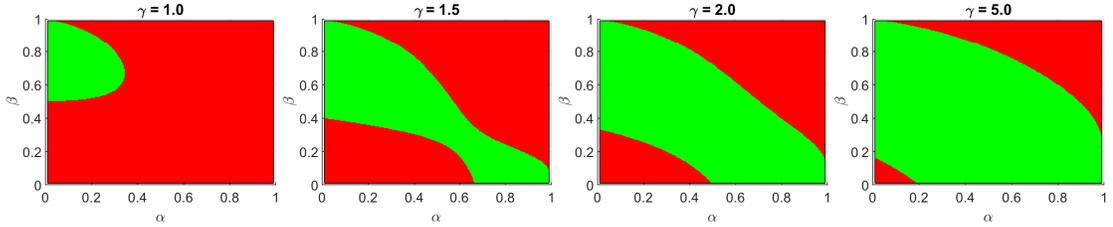


Figure 4.4.1: Visualisation of $c_0$ from equation (4.4.5) for $\gamma = 1$, 1.5, 2, 5 and $\delta = (1-\beta)^{-1}$. The region corresponding to $c_0 \in (0,1)$ is shown in red; the region corresponding to $c_0 \in (1, 1/\alpha)$ is shown in green.

of the parameters $(\alpha, \beta, \delta, \gamma)$ in their specified parameter space. We collect results in Table 4.4.1. In Appendix D.4, we only give details of the $\eta$ calculations when $\alpha, \beta \in (0,1)$, $\gamma > 0$ and $\delta = (1-\beta)^{-1}$. For the other five cases in Table 4.4.1, we state results without proof. In particular, the argument underpinning the $\eta$ calculation when $\delta > (1-\beta)^{-1}$ is similar to the argument used when $\overline{H}$ has a finite upper end point. In this case, $\eta = \alpha$ when $\alpha > 0$ and when $\alpha = 0$, $\eta$ is not defined.

In Table 4.4.1, it is convenient to refer to $c = \max\{1, c_0\} \in [1, 1/\alpha)$ where $c_0 \in (0, 1/\alpha)$ satisfies

$$\gamma(1-\alpha c_0)^{\delta-1}(\delta - 1 + \alpha c_0) = c_0^{\delta}. \tag{4.4.5}$$
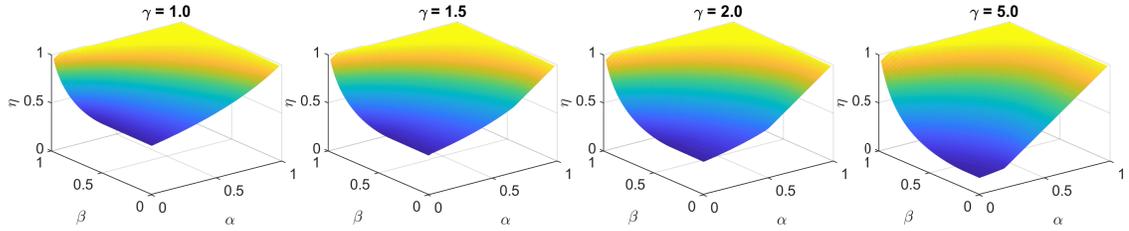
Figure 4.4.2: The value of $\eta$ as a function of $\alpha$, $\beta$ and $\gamma$ with $\delta = (1 - \beta)^{-1}$ from the HT model (4.4.3) and (4.4.4).

To give some intuition on the value of $c$, in Figure 4.4.1 we sketch the region of the parameter space corresponding to $c = 1$ (in red) for different values of $\gamma$. Finally in Figure 4.4.2 we visualise $\eta$ for a set of different parameter combinations with $\delta = (1 - \beta)^{-1}$.

We note the following interesting findings. The parameter $\eta$ is non-decreasing with increasing $\alpha$ and with increasing $\beta$. Parameter combinations $(\alpha, \beta, \gamma, \delta)$ exist for which $\alpha, \beta > 0$ but $\eta < 0.5$. Hence, there are cases for which $Y$ increases with $X$ but the extremes of $(X, Y)$ are negatively associated as measured by $\eta$ (Ledford and Tawn, 1996).

Finally we note that the Heffernan-Tawn model is not $\eta$ invariant, i.e., there exist models that asymptotically follow the same conditional Heffernan-Tawn representation but have different $\eta$. We illustrate this result below with an example, but first we comment on its implications. Our finding implies that if $X$ and $Y$ are asymptotically independent, then there do not exist asymptotically consistent Heffernan-Tawn model-based estimators for probabilities $\mathbb{P}(Y > X > v)$ and $\mathbb{P}(X > v, \ Y > v)$ where $v$ is large. This in turn provides an interesting insight in the lack of self-consistency of the Heffernan-Tawn model with regard to the choice of conditioning variable, see Liu and Tawn (2014).

To illustrate our claim, we consider two bivariate random variables $(X, Y)$ and $(X_{HT}, Y_{HT})$. Let $(X, Y)$ follow an inverted bivariate extreme value distribution with a logistic dependence structure (Ledford and Tawn, 1996) on Laplace margins with

parameter $\xi \in (0, 1]$, such that

$$\mathbb{P}(X > x,\ Y > y) = \exp\left\{-\left[t_x^{1/\xi} + t_y^{1/\xi}\right]^\xi\right\}, \tag{4.4.6}$$

where $t_x := \log 2 - \log[2 - \exp(x)]$ for $x < 0$ and $t_x := \log 2 + x$ for $x > 0$, with $t_y$ similarly defined. It is straightforward to derive that in the limit, the Heffernan-Tawn model (4.4.3) is applicable to $(X, Y)$ with $\overline{H}$ as in equation (4.4.4) and $o(z^\delta) = 0$. Specifically,

$$\lim_{x \to \infty} \mathbb{P}\left(YX^{\xi-1} > z \mid X = x\right) = \exp\left(-\xi z^{1/\xi}\right).$$

Now let $(X_{HT}, Y_{HT})$ be distributed following the exact version of the HT model associated with $(X, Y)$. That is, for $X_{HT} < u$, we have $(X_{HT}, Y_{HT}) = (X, Y)$, and for $X_{HT} \geq u$, $X_{HT} - u$ is a standard exponential and $Y_{HT} \mid X_{HT}$ follows model (4.4.3) with $\overline{H}$ as in (4.4.4) with parameters $(\alpha,\ \beta,\ \gamma,\ \delta) = (0,\ 1 - \xi,\ \xi,\ 1/\xi)$ and $o(z^\delta) = 0$. In this case $\gamma < (1 - \beta)/\beta$, and Table 4.4.1 implies that the coefficient of asymptotic independence $\eta_{HT}$ of $(X_{HT}, Y_{HT})$ is equal to $1/(\xi + 1)$. In contrast, it is straightforward to derive directly from definition (4.4.6) that $\eta$ of $(X, Y)$ is equal to $2^{-\xi}$. Hence $\eta_{HT} \neq \eta$ when $\xi \in (0, 1)$.

Finally we illustrate numerically the differences between $\eta$, $\eta_{HT}$ and their finite level counterparts $\eta(p)$ and $\eta_{HT}(p)$ for $p \in (0, 1)$. For definiteness, we let $(X, Y)$ follow distribution (4.4.6) with $\xi = 0.35$. We simulate a sample $\{(x_i, y_i) :\ i = 1, \ldots, n\}$ of size $n = 10,000$. First we empirically estimate $\eta(p)$ from equation (4.1.3) for $p \in (0, 1)$ and calculate pointwise 95% confidence intervals using the binomial distribution. Next we note that $\eta(p) = \eta$ for $p \in (0.5, 1)$. Finally we calculate the corresponding $\eta_{HT}(p)$ for $p$ near 1 using numerical integration.

Results are shown in Figure 4.4.3. Left and right hand plots are the same except for the scale of the $x$-axis, illustrating the behaviour of $\eta_{HT}(p)$ for $p$ near 1. Reassuringly, the true $\eta$ of the underlying model (red dashed) falls within the 95% confidence interval for its empirical counterpart $\hat{\eta}(p)$ (blue). Further, $\eta_{HT}(p)$ (black dashed) converges to $\eta_{HT}$ (green dashed). We note that $\eta_{HT}(p)$ varies as a function of $p$ and only seems to asymptote for $p > 1 - \exp(-50)/2 \approx 1 - 9.6 \cdot 10^{-23}$. Finally, since $\eta_{HT} < \eta$, we would expect that $\eta_{HT}(p)$ would underestimate $\eta$, but it turns out this is only the case for
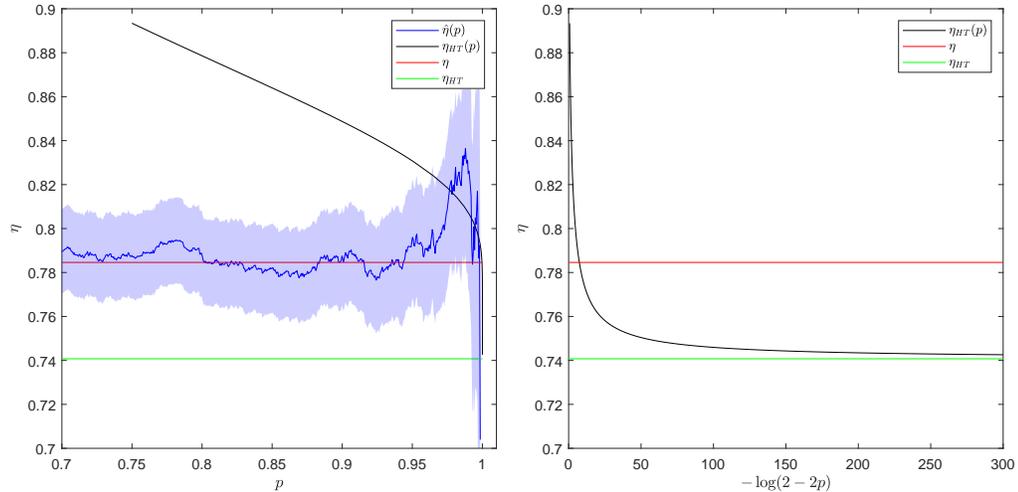
Figure 4.4.3: Coefficients of asymptotic independence $\eta$ (red dashed) for distribution (4.4.6) with $\xi = 0.35$, and the corresponding value for the exact limiting HT model $\eta_{HT}$ (green dashed), and its finite level counterpart $\eta_{HT}(p)$ (black dashed). Empirical estimates $\hat{\eta}(p)$ for a sample of size $10,000$ with pointwise confidence intervals are shown in blue. Left and right hand panels are the same except for the scale of the $x$-axis, set on the right to illustrate the behaviour of $\eta_{HT}(p)$ for $p$ near 1.

$p > 1 - \exp(-7.5)/2 \approx 0.9997$.

**Acknowledgments**

# Chapter 5

# Extreme excursion of multivariate processes

## 5.1 Introduction

Farmers, stock brokers and sailors have one thing in common: they or their businesses are most heavily affected by extreme events like droughts and rainfall, stock market crashes, or extreme winds and waves, respectively. Understanding the statistical behaviour of such events as a whole is crucial for risk analyses. To make this more precise, if we let $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ be a $d$-dimensional random process of interest, then we seek to model excursions of the process into a set $E \subset \mathbb{R}^d$, i.e., the behaviour of

$$\{\mathbf{X}_i : \; i = a, \ldots, b; \; \mathbf{X}_i \in E; \; \mathbf{X}_{a-1}, \mathbf{X}_{b+1} \notin E\}, \tag{5.1.1}$$

where $E$ is associated with extreme events of $\mathbf{X}$. Moreover, we assume that the random process consists of multiple components that can be extreme. To solve this task, we assume that the multivariate random process is a realisation of a $k$th order Markov chain.

We use extreme value theory, a subfield of statistics, to characterise excursions. There is considerable recent attention to this area in the literature, but most of extreme value theory for stationary Markov chains dates back over 20 years. Rootzén (1988) and Perfekt (1997) develop limiting results for univariate Markov chains and

multivariate Markov chains, respectively. Smith (1992) calculates the extremal index (Leadbetter et al., 1983) for a univariate Markov chain and Smith et al. (1997) use parametric bivariate transition distributions to model the extremes of a univariate first order Markov process. Finally, Yun (2000) develops asymptotic theory for functionals of univariate $k$th order Markov extreme events. All of these papers derive their results under the assumption of asymptotic dependence (Joe, 1997), i.e., for a stationary process $(X_t)_{t \in \mathbb{Z}}$ satisfying suitable long-range mixing conditions, they derive their results under the assumption that for any lag $l = 1, 2, \ldots$

$$\lim_{u \to x^*} \mathbb{P}(X_{t+l} > u | X_t > u) > 0$$

where $x^*$ is the right upper end point of the distribution of $X_t$. This early work doesn't consider what happens when asymptotic independence is present, i.e., when this limiting probability converges to 0 for some $l$. The first paper which considers such processes is Bortot and Tawn (1998) who assume at first order Markov model, with Ledford and Tawn (2003) considering a general framework for the modelling of asymptotic independent processes, with key recent probabilistic developments given by Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2020).

Randell et al. (2015) speculate that a statistical model for the evolution of (multivariate) storm trajectories would be a valuable enhancement of current methodology. The first statistical work the current authors are aware of, that defines a model for the distribution of all observations during an excursion is Winter and Tawn (2016), who assume a univariate first order Markov process exhibiting either asymptotic independence or asymptotic dependence across lags. Winter and Tawn (2017) extend this work to include $k$th order Markov processes with $k > 1$. Finally, Tendijck et al. (2019) extend this latest model to a $k$th order univariate Markov process with a directional covariate. We remark that their work cannot be considered to model the extremes of bivariate Markov processes since the associated directional covariate cannot be considered to take on extreme values. An honourable mention goes to Feld et al. (2015) who use a sophisticated covariate model for the most extreme observation (the most extreme value of the dominant variable) in an excursion, combined with a baseline

historical-matching approach for the intra-excursion trajectory; see also Section 5.3.2 in which we discuss a version of this model to our case study. Finally, we mention well-established literature on multivariate time series, e.g., Tiao and Tsay (1989), which is not directly applicable to modelling environmental extremes because such models are only designed to model typical behaviours.  Financial time-series models, e.g., Bauwens et al. (2006), are also not applicable because these are specifically tailored to model financial data with tail switching during extreme events (Bortot and Coles, 2003).

In this work, we present a natural extension to Tendijck et al. (2019) by defining two multivariate $k$th order Markov models that exhibit both asymptotic (in)dependence across variables and/or at some lags.  The work is motivated by our case study in which we model excursions of meteorological-oceanographic (met-ocean) data: significant wave height, wind speed, and their associated directions, for a location in the northern North Sea.

We use the following set up.  Assume that at each time $t \in \mathbb{Z}$, the distribution of the $d$-dimensional random variable $\mathbf{X}_t$ is stationary through time; that is, $\mathbf{X}_t$ has the same distribution as some $\mathbf{X} = (X_1, \ldots, X_d)$ with distribution function $F_{\mathbf{X}}$. For $1 \le j \le d$, write $F_{X_j}$ as the $j$th marginal distribution of $F_{\mathbf{X}}$.  The distribution functions $F_{X_j}$ are unknown and must be estimated. For extreme arguments, we use univariate extreme value theory to motivate a class of parametric tail forms.  More precisely, we assume that for each $1 \le j \le d$, the excesses tail above some high level $u_j \in \mathbb{R}$ of the marginal distribution $F_{X_j}$ are approximated with a generalised Pareto distribution (Davison and Smith, 1990).  For non-extreme arguments $x < u_j$ of the function $F_{X_j}$, an empirical model usually suffices.

In multivariate extreme value theory, it is common to consider the marginals and the dependence of random variables separately, such that the usually-dominant marginal effect does not influence the modelling of a possibly complex dependence structure.  So given the marginal models as discussed above, we transform the random process $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ onto standard Laplace margins $(\mathbf{Y}_t)_{t \in \mathbb{Z}}$ using the transformation: $X_j \mapsto Y_j := F_L^{-1}(F_{X_j}(X_j))$, where $F_L^{-1}$ is the inverse of the standard Laplace distribu-

tion function. Here the choice of Laplace margins is made to allow for the modelling of potential negative dependence at certain lags or components (Keef et al., 2013).

For multivariate random processes, there are uncountably many ways of defining an extreme event. In our case study, we take the met-ocean variable significant wave height $H_S$ as the excursion-defining component. We follow Winter and Tawn (2017) and Tendijck et al. (2019) in adopting the conditional extremes model of Heffernan and Tawn (2004), see also Section 5.3.1, as the foundation of our approach. Without loss of generality, we first define the component $X_1$ of $\mathbf{X}$ as the defining variable for the extreme events. So, we set our excursion set $E = E_u := (F_{X_1}^{-1}\{F_L(u)\}, \infty) \times \mathbb{R}^{d-1}$ for some high threshold $u \in \mathbb{R}$ and rewrite our definition of an excursion as

$$\{\mathbf{Y}_i : \ i = a, \ldots, b; \ Y_{i,1} > u; Y_{a-1,1} \leq u, \ Y_{b+1,1} \leq u\} \qquad (5.1.2)$$

for $a, b \in \mathbb{Z}$, indices for the start and the end time points of the excursion, respectively. In shorthand, the excursion is then $\mathbf{Y}_{a:b}$. We remark that in this definition, we accept that multiple excursions can occur close together in time, and thus these cannot be considered independent. The reason for this choice is that imposing a minimal separation of excursions would complicate the modelling significantly. We recognize that this is a feature of the current approach which can be improved.

The remaining part of this paper is organised as follows. In Section 5.2, we present our strategy for modelling excursions by defining time intervals corresponding to so-called peak, pre-peak and post-peak periods. In Section 5.3.1, we introduce the conditional extremes model, and in Section 5.3.2 we present a baseline historical-matching model. In Sections 5.3.3-5.3.4, we define two novel $k$th order Markov models for the evolution of the multivariate time-series during the pre-peak and post-peak periods. Combined with the conditional extremes model from Section 5.3.1 for the period of the peak, we obtain two models for extreme excursions. In Section 5.4, we apply the two introduced Markov models to met-ocean data for a location in the northern North Sea. We compare the model performance with the baseline historical-matching approach by assessing their respective performance in estimating the tails of the distributions of complex structure variables (Coles and Tawn, 1994), corresponding to

approximations of the response of hypothetical offshore or coastal facilities to extreme met-ocean environments. We find that in general the new models are preferred.

## 5.2   Modelling strategy

To model excursions as in definition (5.1.2), two types of approaches have been proposed in the literature of univariate extremes: a forward model (Rootzén, 1988) and a peak model (Smith et al., 1997). Both of these are two-step approaches by nature. The forward model first describes the distribution of a random exceedance $Y_t > u$ with a univariate extremes model and a conditional model for the distribution of $Y_{t+j}|(Y_{t+i-1} = y_{t+j-2}\ i = 1, \dots, j)$ where $y_{t+j-2} > u$ for any $j \geq 1$. Even though this approach does not directly model the univariate equivalent of excursions in equation (5.1.2) (because the first exceedance of an excursion does not have the same distribution as a random exceedance), estimates of some extremal properties of the process $(Y_t)_{t \geq 1}$, such as the extremal index (Leadbetter et al., 1983), can still be obtained by allowing the excursion threshold to be significantly lower than the cluster threshold used in extremal index estimators. Notably, Winter and Tawn (2016, 2017) use the forward approach in their work.

The peak model, on the other hand, does model excursions as defined here. This method relies on a univariate extremes model for the largest observation of an excursion, e.g., Eastoe and Tawn (2012), and a conditional model for observations before and after the excursion maximum. Winter and Tawn (2016) use this approach for their first order model but not for their $k$th order model (Winter and Tawn, 2017). They avoid this method explicitly because of difficulties that arise in preserving model characteristics in forward and backward simulations near the excursion maximum (i.e., the time point at which the defining variate $X_1$ achieves its maximum value during the excursion).

Tendijck et al. (2019) use the peak method, but they do not address the issues associated with forward and backward simulation under the method. Because the excursion maximum is usually the most important observation of an excursion for risk

assessments, we also use the peak method in the current work, but with consideration of backward and forward models. We separate the modelling of excursions into three stages: the modelling of the period of the peak, and the modelling of the pre-peak and post-peak periods; see Figure 5.2.1 in which the three time periods are illustrated for $k = 3$. Without loss of generality, let $t = 0$ be the time point at which the first component $Y_{t,1}$ takes its maximum value within an excursion such that $Y_{0,1} > u$ for the threshold $u$. The period of the peak $\mathcal{P}_0^k$ of an excursion of a $k$th order model is then defined as the set of $2k - 1$ observations: $\mathcal{P}_0^k := \{\mathbf{Y}_t : -(k - 1) \leq t \leq k - 1\}$ with $Y_{0,1} > u$. The pre-peak $\mathcal{P}^{\mathrm{pre}}$ and post-peak $\mathcal{P}^{\mathrm{post}}$ periods are defined as the sets of observations that include the excursion maximum and the observations before and after, respectively:

$$\mathcal{P}^{\mathrm{pre}} := \{\mathbf{Y}_t : t' \leq t \leq 0, \text{ with } t' = \min\{s < 0 : \min_{i=s,\dots,0}\{Y_{i,1}\} > u\}\}$$

and

$$\mathcal{P}^{\mathrm{post}} := \{\mathbf{Y}_t : 0 \leq t \leq t', \text{ with } t' = \max\{s > 0 : \min_{i=0,\dots,s}\{Y_{i,1}\} > u\}\},$$

so each of them intersects with $\mathcal{P}_0^k$. We remark that the length of $\mathcal{P}_0^k$ can be longer or shorter than the length of an excursion if the excursion ends within the period of the peak. We choose to define the period $\mathcal{P}_0^k$ in this manner so that the pre-peak and post-peak parts of the excursion are both initialized with $k$ observations.

We then model an excursion as follows: (i) we model the excursion maximum $Y_{0,1}$ using a generalised Pareto distribution; (ii) we model the period of the peak $\mathcal{P}_0^k$ conditional on the storm maximum $Y_{0,1}$ using the model described in Section 5.3.1; (iii-a) if $\min_{j=1,\dots,k-1} Y_{j,1} < u$ ($\min_{j=1,\dots,k-1} Y_{-j,1} < u$), then the period $\mathcal{P}^{\mathrm{post}}$ ($\mathcal{P}^{\mathrm{pre}}$) of the excursion has ended; (iii-b) if $\min_{j=1,\dots,k-1} Y_{j,1} \geq u$ ($\min_{j=1,\dots,k-1} Y_{-j,1} \geq u$), then the remaining part of the excursion is modelled with our time-series models from Sections 5.3.3-5.3.4 until there exist a $j_1, j_2 > 0$ such that $Y_{j_1,1} < u$ and $Y_{-j_2,1} < u$.

In the next sections, we discuss forwards models that are applicable to model the post-peak period $\mathcal{P}^{\mathrm{post}}$. We model the pre-peak period $\mathcal{P}^{\mathrm{pre}}$ using the forwards models applied to $(Y_{-t})_{t\in\mathbb{Z}}$ (with potentially different parameters, although these would be the

same if the process was time reversible). Importantly, we do not impose consistency in the forwards and backwards models to yield a kth order Markov chain, e.g., in the case of asymptotic dependent Markov chains the precise dependence conditions between the forward and backward hidden tail chains are given by Janßen and Segers (2014). We make this choice for two reasons: (i) for environmental applications, such as in this work, the pre-peak and post-peak period have different distributions, see for example the asymmetry in Figure 5.4.5. (ii) the assumption of $k$th order Markov is an approximation for the process that generates our data. Thus, imposing forward and backward consistency for a $k$th order Markov chain is likely to yield worse results for our application. So, we consider the violating of this assumption to yield more flexible descriptions of excursions as a benefit more than a limitation.
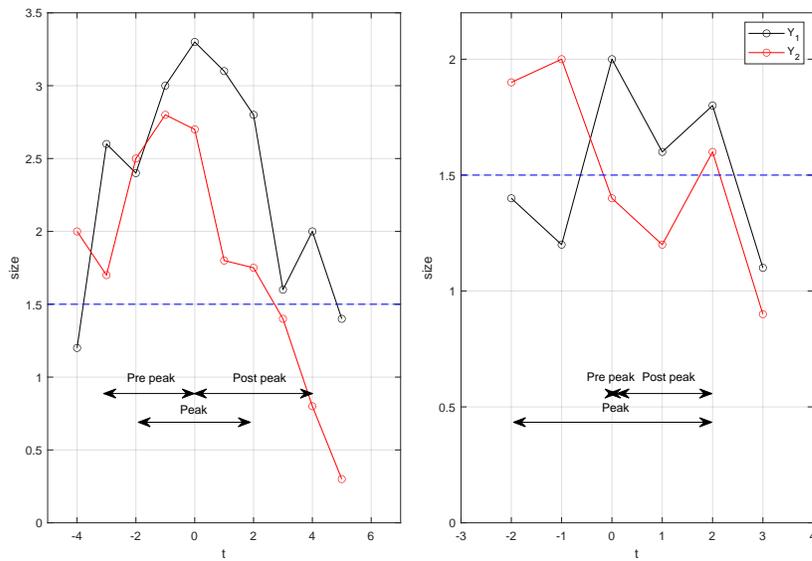


Figure 5.2.1: Illustration of the periods of the peak, pre-peak, and post-peak periods for two excursions from a Markov model with order $k = 3$.

## 5.3 The models

### 5.3.1 Model 0: The conditional extremes model

We introduce the conditional extreme value model of Heffernan and Tawn (2004), henceforth denoted the HT model, with notation specific to modelling the period of the peak $\mathcal{P}_0^k$. The HT model is widely studied and applied to extrapolate tails of multivariate distributions, e.g., in oceanography (Ross et al., 2020), finance (Hilal et al., 2011), spatio-temporal extremes (Simpson and Wadsworth, 2021), and multivariate spatial extremes (Shooter et al., 2022). The HT model is a limit model and its form was originally motivated by deriving possible limiting forms for numerous theoretical examples.

Let

$$\mathbf{Y}_{-(k-1):(k-1)} := \begin{pmatrix} Y_{-(k-1),1} & \cdots & Y_{-(k-1),d} \\ \vdots & & \vdots \\ Y_{k-1,1} & \cdots & Y_{k-1,d} \end{pmatrix}$$

be a random matrix on $\mathbb{R}^{(2k-1)\times d}$ with standard Laplace margins (Keef et al., 2013). We define $\underline{\mathbf{Y}}$ as $\mathbf{Y}_{-(k-1):(k-1)}$ without the $(k,1)$th element $Y_{0,1}$. Additionally, we assume that the joint density of $\mathbf{Y}_{-(k-1):(k-1)}$ exists. For ease of presentation, we use this notation more general, i.e., we define the irregular matrix $\underline{\mathbf{x}} \in \mathbb{R}^{(2k-1)d-1}$ as follows:

$$\underline{\mathbf{x}} = \begin{pmatrix} x_{-k+1,1} & x_{-k+1,2} & \cdots & x_{-k+1,d} \\ \vdots & \vdots & & \vdots \\ x_{-1,1} & x_{-1,2} & \cdots & x_{-1,d} \\ & x_{0,2} & \cdots & x_{0,d} \\ x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ \vdots & \vdots & & \vdots \\ x_{k-1,1} & x_{k-1,2} & \cdots & x_{k-1,d} \end{pmatrix},$$

that is, $\underline{\mathbf{x}}$ does not contain the $(k,1)$th element. Equivalently, one could write $\underline{\mathbf{x}} = \mathbf{x}_{-(k,1)}$ for $\mathbf{x} \in \mathbb{R}^{(2k-1)\times d}$.

The conditional extremes model for $\underline{\mathbf{Y}}$ conditional on $Y_{0,1}$ assumes that irregular

parameter matrices $\underline{\boldsymbol{\alpha}} \in [-1, 1]^{(2k-1)d-1}$, $\underline{\boldsymbol{\beta}} \in (-\infty, 1)^{(2k-1)d-1}$ and a distribution function $H$ with non-degenerate marginals on $\mathbb{R}^{(2k-1)d-1}$ (the space of irregular matrices) exist, such that for all irregular matrices $\underline{\mathbf{z}} \in \mathbb{R}^{(2k-1)d-1}$ the limit

$$H(\underline{\mathbf{z}}) = \lim_{y \to \infty} \mathbb{P} \left( \frac{Y_{i,j} - \alpha_{i,j} y}{y^{\beta_{i,j}}} \leq z_{i,j} \ \forall \ (i,j) \in \mathscr{I} \mid Y_{0,1} = y \right) \qquad (5.3.1)$$

exists, where $\mathscr{I} := \big( \{-(k-1), \ldots, (k-1)\} \times \{1, \ldots, d\} \big) \setminus \{(0,1)\}$ and where $\alpha_{i,j}$, $\beta_{i,j}$ and $z_{i,j}$ are the $(i,j)$th elements of $\underline{\boldsymbol{\alpha}}$, $\underline{\boldsymbol{\beta}}$ and $\underline{\mathbf{z}}$, respectively. Limit (5.3.1) implies, according to l'Hopital's rule, that for $y > 0$, $\underline{\mathbf{z}} \in \mathbb{R}^{(2k-1)d-1}$

$$\lim_{u \to \infty} \mathbb{P} \left( \frac{\mathbf{Y} - \boldsymbol{\alpha} Y_{0,1}}{Y_{0,1}^{\boldsymbol{\beta}}} \leq \underline{\mathbf{z}}, \ Y_{0,1} - u > y \mid Y_{0,1} > u \right) = H(\underline{\mathbf{z}}) \exp(-y), \qquad (5.3.2)$$

assuming component-wise operations. Limit (5.3.2) in turn has the interpretation that as $u$ tends to infinity, $(\mathbf{Y} - \boldsymbol{\alpha} Y_{0,1}) Y_{0,1}^{-\boldsymbol{\beta}}$ and $(Y_{0,1} - u)$ are independent conditional on $Y_{0,1} > u$, and are distributed as $H$ and a standard exponential, respectively.

In practice, we exploit these results by assuming they hold exactly above some high finite threshold $u > 0$. So, we approximate the conditional distribution of $\mathbf{Y}|Y_{0,1} = y$ for $y > u$, $\underline{\mathbf{y}} \in \mathbb{R}^{(2k-1)d-1}$ as

$$\mathbb{P}(\mathbf{Y} \leq \underline{\mathbf{y}} \mid Y_{0,1} = y) = H \left( \frac{\underline{\mathbf{y}} - \boldsymbol{\alpha} y}{y^{\underline{\boldsymbol{\beta}}}} \right), \qquad (5.3.3)$$

and we assume independence of $(\mathbf{Y} - \boldsymbol{\alpha} Y_{0,1}) Y_{0,1}^{-\boldsymbol{\beta}}$ and $Y_{0,1}$. There is no finite dimensional parametric form for $H$, so non-parametric methods are typically applied. However, we remark that there are applications of the conditional extreme value model where the copula $H$ is assumed to be Gaussian (Towe et al., 2019) or a Bayesian semi-parametric model is used (Lugrin et al., 2016). For inference, see Section 5.3.5.

## 5.3.2   Model 1: Historical-Matching

An empirical method for simulating excursions is described in Feld et al. (2015) and termed historical-matching (HM) in this work. They model trajectories of significant wave height, wave direction, season and wave period during extreme events. The key assumption they make is that storm trajectory (or excursion) profiles are not independent of storm maximum conditions. This approach serves as our baseline method

for the simulation of storm trajectories of significant wave height, wind speed, wave direction and wind direction, and is compared against our proposed Markov models in Sections 5.3.3-5.3.4. Specifically, the HM approach is a composition of four models: (i) a model for storm maximum wave direction; (ii) a model for storm maximum significant wave height conditional on storm maximum wave direction; (iii) a model that selects at random a historical storm trajectory with similar storm maximum characteristics to that simulated; (iv) a model that adjusts the historical storm trajectory by matching storm maximum characteristics of simulated and historical storms.

For step (i), we simply sample at random from the observed wave directions associated with storm maximum significant wave height (excursion maximum). In step (ii), storm maximum significant wave height are modelled as generalised Pareto distributions conditional on the sampled storm maximum wave direction using a generalised additive model with the parameters as B-splines conditional on directional covariates (Chavez-Demoulin and Davison, 2005). In step (iii), we use a distance measure to calculate the dissimilarity between pairs of storm maximum significant wave heights and storm maximum wave directions for simulated and historical trajectories. Here, we use the heuristic recommended by Feld et al. (2015) ensuring that a difference of 5 degrees in storm maximum wave direction corresponds to the same dissimilarity as $0.5m$ of difference in storm maximum significant wave height; one of the closest 20 matching storms is then selected at random for associated with the simulated storm maximum. In step (iv), we match the variables of the chosen historical trajectory as follows: (a) the historical significant wave height series are multiplied by the ratio of the simulated maximum significant wave height to the maximum of the historical significant wave height; (b) the historical wave directions are shifted such that the storm maximum wave directions of simulated and historical trajectories coincide; (c) the associated historical wind directions are rotated in the exact same way as wave direction; (d) for the full set of historical storm maxima, storm maximum associated wind speed $W_s^M$ (namely the value of wind speed at the time point corresponding to the storm maximum event) conditional on storm maximum significant wave height

$H_S^M$ is described using linear regression with parameters $\beta_0, \beta_1 \in \mathbb{R}$, $\sigma > 0$:

$$W_s^M | H_S^M = \beta_0 + \beta_1 H_S^M + \sigma\varepsilon$$

with $\varepsilon$ standard normal; (e) wind speed for the selected historical trajectory is scaled linearly such that it agrees with the storm maximum associated wind speed from (d).

Perhaps the main deficiencies of the HM approach are that for levels beyond that observed in historical excursions, it does not provide a means for modelling the extremal temporal dependence characteristics of excursions, and the extremal dependence between different components of the time-series, and does not provide a natural framework for the assessment of model fit or uncertainty propagation.

### 5.3.3 Model 2: Multivariate Markov Extremal Model

For ease of presentation, we present the multivariate Markov extremal model (MMEM) of order $k$ only for a two-dimensional time-series $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$ such that $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2})$ in the notation of Section 5.1, i.e., $\mathbf{Y}_t$ has standard Laplace margins. We only describe a forwards model that is applicable to the post-peak period $\mathcal{P}^{\mathrm{post}}$. As mentioned in Section 5.2, we apply a different forwards MMEM model to $(\mathbf{Y}_{-t})_{t\in\mathbb{Z}}$ to yield the backwards model for the pre-peak period $\mathcal{P}^{\mathrm{pre}}$. Concisely put, the MMEM exploits the HT model to estimate the distribution for $\mathbf{Y}_{t+k}$ conditional on $(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t+k-1})$ when $Y_{t,1} > u$ for a large threshold $u > 0$. Similar to in Section 5.3.1, for each $t \in \mathbb{Z}$, we define $\tilde{\mathbf{x}}_t \in \mathbb{R}^k \times \mathbb{R}^{k+1}$ to be an irregular matrix with $k+1$ rows and 2 columns without the element that is on the first row and first column:

$$\tilde{\boldsymbol{x}}_t = \begin{pmatrix} & x_{t,2} \\ x_{t+1,1} & x_{t+1,2} \\ \vdots & \vdots \\ x_{t+k,1} & x_{t+k,2} \end{pmatrix}.$$

Then, we assume that for a large threshold $u > 0$, there exist parameters $\tilde{\boldsymbol{\alpha}}_0 \in [-1,1]^k \times [-1,1]^{k+1}$, $\tilde{\boldsymbol{\beta}}_0 \in (-\infty, 1)^k \times (-\infty, 1)^k$, and a residual random variable $\tilde{\varepsilon}_t$ on $\mathbb{R}^k \times \mathbb{R}^{k+1}$ with non-degenerate marginals such that for $t \in \mathbb{Z}$

$$\tilde{\mathbf{Y}}_t | (Y_{t,1} > u) = \tilde{\boldsymbol{\alpha}}_0 Y_{t,1} + Y_{t,1}^{\tilde{\boldsymbol{\beta}}_0} \tilde{\boldsymbol{\varepsilon}}_t.$$

Similar to Winter and Tawn (2017), for $t \in \mathbb{Z}$, $j \geq 1$ when $Y_{t+j,1} > u$, we then get

$$[Y_{t+k+j,1} \ Y_{t+k+j,2}]|(\mathbf{Y}_{t+j:t+k+j-1}, Y_{t+j,1} > u) = [\alpha_{k,1}, \ \alpha_{k,2}]Y_{t+j,1} + Y_{t+j,1}^{[\beta_{k,1}, \ \beta_{k,2}]} \cdot \boldsymbol{\varepsilon}_{k,1:2}^C,$$

where $\boldsymbol{\varepsilon}_{k,1:2}^C$ is short-hand notation for $[\varepsilon_{k,1}, \varepsilon_{k,2}]$ conditional on $(\varepsilon_{1:k-1,1}, \varepsilon_{0:k-1,2})$. For inference, we refer to Section 5.3.5.

### 5.3.4 Model 3: extremal vector autoregression

Here, we introduce extremal vector autoregression (EVAR) for the extremes of the process $(\mathbf{Y}_t)_{t \geq 1}$. This model combines the HT model with a vector autoregression model for the joint evolution of the time-series at large levels. Here we focus on the post-peak period, but note that the pre-peak period is modelled with MMEM in the same spirit as EVAR. We define an EVAR model of order $k$ with parameters $\Phi^{(i)} \in \mathbb{R}^d \times \mathbb{R}^d$ for $i = 1, \ldots, k$ and $\mathbf{B} \in (-\infty, 1)^d$ as follows

$$\mathbf{Y}_{t+k}|(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t+k-1}) = \sum_{i=1}^{k} \Phi^{(i)} \mathbf{Y}_{t+k-i} + y^{\mathbf{B}} \boldsymbol{\varepsilon}_t, \tag{5.3.4}$$

with $Y_{t,1} = y$ for $y > u$, where $u > 0$ is a large threshold and $\boldsymbol{\varepsilon}_t$ is a $d$-dimensional multivariate random variable that has non-degenerate margins and is independent of $(\mathbf{Y}_t, \ldots, \mathbf{Y}_{t+k-1})$. Usually for a vector autoregressive model, parameter constraints would be imposed so that the resulting process is stationary. In the current extreme value context, stationarity is not of concern to us, since (i) we reject trajectories that exceed the excursion maximum, and (ii) we stop the process once the first component dips below a threshold. We define EVAR$_0$ as a special case of EVAR corresponding to $\mathbf{B} = \mathbf{0}$. EVAR$_0$ therefore has clear similarities with a regular vector autoregressive model (Tiao and Box, 1981), yet we emphasise that there is considerable difference between the two, since (i) the parameters of EVAR$_0$ do not need to yield a stationary process; (ii) the parameters of EVAR$_0$ are estimated using only extreme observations. To estimate the EVAR model, we adopt the same approach as that used to estimate the HT model, see Section 5.3.5. As explained in Appendix E.1, the resulting parameter estimators $\hat{\Phi}^{(i)}$ are highly correlated, and a reparameterisation is introduced to reduce the correlation.

For practical applications, an advantage of EVAR over MMEM is a lower dimension of the residual distribution when $k > 1$: $d$ and $kd$, respectively. So, the estimates of the residual distribution are less affected by the sparsity of high dimensional applications. As a consequence, a drawback of EVAR is that it might oversimplify a complex dependence.

### 5.3.5   Inference for conditional models

We discuss inference for models 0, 2 and 3 with parameter vector $\boldsymbol{\theta}$. We discuss these together because they can be summarized in the same form. Specifically, let $\mathbf{W} = (W_1, \ldots, W_d)$ be a $d$-dimensional random variable and assume that for some high threshold $u > 0$,

$$\mathbf{W}_{2:d}|(W_1 > u) = \mathbf{g}_1(W_1; \boldsymbol{\theta}) + \mathbf{g}_2(W_1; \boldsymbol{\theta})\boldsymbol{\varepsilon}. \tag{5.3.5}$$

for some parametric functions $\mathbf{g}_1(\,\cdot\,; \boldsymbol{\theta}) : \mathbb{R} \to \mathbb{R}^{d-1}$ and $\mathbf{g}_2(\,\cdot\,; \boldsymbol{\theta}) : \mathbb{R} \to \mathbb{R}^{d-1}_{>0}$, where for $x \in \mathbb{R}$,

$$\mathbf{g}_1(x, \boldsymbol{\theta}) := (g_{1,2}(x, \boldsymbol{\theta}), \ldots, g_{1,d}(x, \boldsymbol{\theta})) \text{ and } \mathbf{g}_2(x, \boldsymbol{\theta}) := (g_{2,2}(x, \boldsymbol{\theta}), \ldots, g_{2,d}(x, \boldsymbol{\theta})),$$

and where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)$ is a $d$-dimensional multivariate random variable that is non-degenerate in each margin and that is independent of $W_1$. As an example, for MMEM, $g_{1,j}(x) = \alpha_j x$ for some $\alpha_j$ and $g_{2,j}(x) = x^{\beta_j}$ for some $\beta_j$.

Next, assume that we have $n$ observations $\mathcal{D} := \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ of the conditional random variable $\mathbf{W}|W_1 > u$, where $\mathbf{w}_i = (w_{i1}, \ldots, w_{id})$ with $w_{i1} > u$ for $i = 1, \ldots, n$. We then infer $\boldsymbol{\theta}$ by calculating the likelihood of model (5.3.5) by temporarily assuming that the $\boldsymbol{\varepsilon}$ has a multivariate normal distribution with unknown mean $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_d)$ and unknown diagonal covariance matrix $\Sigma = \boldsymbol{\sigma}^2 I$ where $\boldsymbol{\sigma}^2 = (\sigma_2^2, \ldots, \sigma_d^2)$. These assumptions imply that the mean and the variance of $\boldsymbol{\varepsilon}$ are estimated simultaneously with the model parameters. The likelihood $L \equiv L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\sigma}; \mathcal{D})$ is then evaluated as

$$L = \prod_{i=1}^{n} \prod_{j=2}^{d} \frac{1}{\sqrt{2\pi}\sigma_j g_{2,j}(w_{i1}; \boldsymbol{\theta})} \exp\left\{ -\frac{1}{2\sigma_j^2} \left( \frac{w_{ij} - g_{1,j}(w_{i1}) - \mu_j g_{2,j}(w_{i1}; \boldsymbol{\theta})}{g_{2,j}(w_{i1}; \boldsymbol{\theta})} \right)^2 \right\}.$$

Finally, the parametric assumption on the distribution of $\boldsymbol{\varepsilon}$ is discarded and estimated conditional on the parametric estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, with a kernel density $\hat{h} := \hat{h}_{1:d}$ using the 'observations' $\{\boldsymbol{\varepsilon}_i : \ i = 1, \ldots, n\}$ where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i2}, \ldots, \varepsilon_{id})$ and

$$\varepsilon_{ij} := \frac{w_{ij} - \hat{g}_{1,j}(w_1; \hat{\boldsymbol{\theta}})}{\hat{g}_{2,j}(w_1; \hat{\boldsymbol{\theta}})}$$

for $i = 1, \ldots, n$, $j = 2, \ldots, d$.

In case of models 2 and 3, we additionally require estimates for the density of a conditional random variable $\boldsymbol{\varepsilon}_{1:l|l+1:d} = (\varepsilon_1, \ldots, \varepsilon_l)|(\varepsilon_{l+1}, \ldots, \varepsilon_d)$ for some $l \in \{1, \ldots, d-1\}$. Given the same set of observations as above, we estimate its conditional density $h_{1:l|l+1:d}$ as follows

$$\hat{h}_{1:l|l+1:d}(\varepsilon_1, \ldots, \varepsilon_l|\varepsilon_{l+1}, \ldots, \varepsilon_d) = \frac{\hat{h}_{1:d}(\varepsilon_1, \ldots, \varepsilon_d)}{\hat{h}_{l+1:d}(\varepsilon_{l+1}, \ldots, \varepsilon_d)}$$

where the right-hand side is estimated using kernel densities similar to $\hat{h}_{1:d}$.

## 5.4 Case Study - Met-Ocean in the Northern North Sea

We apply MMEM, EVAR and HM to characterise excursions of significant wave height $H_S$ and wind speed $W_s$ with directional covariates for a location in the northern North Sea. Our goal is to estimate predictive models for the joint evolution of $H_S$ and $W_s$ time-series conditional on $H_S$ being large whilst keeping the model complexity as simple as possible.

In Section 5.4.1, we describe the available met-ocean data. In Section 5.4.2, we outline a model for the evolution of storm direction that is needed for simulation under our time-series models. In Section 5.4.3, we introduce structure variable responses that approximate fluid drag loading on a marine structure such as a wind turbine, or coastal defence. Finally, in Section 5.4.4, we compare the predictive performance of MMEM and EVAR for different orders of the associated Markov models relative to that of the HM method in terms of estimating structure variables for withheld intervals of time-series.

### 5.4.1    Data

We have 53 years of hindcast data

$$\mathcal{D} := \{(H_{S,i}, W_{s,i}, \theta_i^H, \theta_i^W) : \; i \in \mathcal{T}\},$$

consisting of time-series for four three-hourly met-ocean summary statistics at a location in the northern North Sea (Reistad et al., 2009): significant wave height ($H_{S,i}$ in metres), wind speed ($W_{s,i}$ in metres per second), wave direction ($\theta_i^H$ in degrees) and wind direction ($\theta_i^W$ in degrees) for each $i \in \mathcal{T}$. To use MMEM and EVAR, we transform significant wave height and wind speed onto Laplace marginals: $H_{S,i}|\theta_i^H \mapsto H_{S,i}^{\mathrm{L}}$ and $W_{s,i}|\theta_i^W \mapsto W_{s,i}^{\mathrm{L}}$ using directional marginal extreme value models (Chavez-Demoulin and Davison, 2005), but ignoring seasonal covariates. This part of the analysis has been reported on numerous occasions, see for example Randell et al. (2015). Because the marginal transformation includes direction as a covariate and because direction is not constant during an excursion, we also establish a model for the directional evolution of excursions in order to transform them between standard and original margins, see Section 5.4.2.

Let $D^L$ be the collection of the transformed data

$$\mathcal{D}^L := \{(H_{S,i}^L, W_{s,i}^L, \theta_i^H, \theta_i^W) : \; i \in \mathcal{T}\}.$$

To define excursions in $\mathcal{D}^L$, we set the excursion threshold $u$ equal to the 95% percentile of a standard Laplace distribution, i.e., $u \approx 2.3$. This choice yields $1,467$ observations of extreme excursions $\mathcal{E}_u$.

Figure 5.4.1 shows four intervals of the time-series chosen to contain the 100%, 95%, 90% and 85% sample percentile of significant wave height on original margins. We additionally show the associated time-series on standard Laplace margins with its directional covariates. From this figure, we observe typical profiles centred around extreme events. First, as expected, we note a large dependence of $H_S$ and $W_s$ on both original and standard margins. Moreover, we observe that the variables associated to significant wave height, i.e., $H_S$, $H_S^L$ and $\theta^H$, are much smoother than their wind speed counterparts. Additionally, the directional covariates $\theta^H$ and $\theta^W$ centre around each other with no large deviations of each other during extreme events.

In Figure 5.4.2, we visualize the (cross) dependence of key Laplace-scale variables $H_S^L$ and $W_s^L$ at time lags up to lag 4. In this figure, we observe the complex dependence structure of the bivariate time-series of significant wave height and wind speed on Laplace margins. As expected, we observe (slow) convergence to an independent variable model as lag increases. Most notably, we observe a similar level of dependence of $(H_{S,t}^L, W_{s,t+4}^L)$ and $(W_{s,t}^L, W_{s,t+4}^L)$.

In Figure 5.4.3, we plot (cross) correlation functions for these variables, and also for the change in directional covariates at various lags in terms of hours, recalling that a lag of 1 in time stamps is equal to a three hours in real time. These show that the dependence of $(H_{S,t}^L, H_{S,t+\tau}^L)$ decays relatively slowly as $\tau$ grows to infinity, and that indeed the cross dependence between $(H_{S,t}^L, W_{s,t+\tau}^L)$ is larger than the dependence of $(W_{s,t}^L, W_{s,t+\tau}^L)$ for large $\tau$. Finally, the correlation plot of the change in directional covariates on the right shows that a first order model for these covariates is appropriate since the correlations nearly vanish at lag 2 or 6 hours. In this case, the cross dependence is hardly different from zero.

## 5.4.2   Directional Model

We model wave direction $\theta_i^H$ in a similar fashion as Tendijck et al. (2019). Let $\mathcal{I} \subset \mathcal{T}$ be the set of indices of the original data that correspond to all observations of any excursion. Next, let $\{d(\theta_{i+1}^H, \theta_i^H) : i \in \mathcal{I}\}$ be the set of changes in wave directions, where $d(\theta, \theta') = (\theta - \theta' + 180; \mod 360) - 180 \in [-180, 180)$ denotes the circular difference of $\theta$ and $\theta'$ in degrees. We remark here that in our application, the set of changes in wave directions during excursions do not contain values close to $-180$ or $180$. In particular, all of the observed changes centre around 0.

For $i \in \mathcal{I}$, we transform observations $d(\theta_{i+1}^H, \theta_i^H) \mapsto \delta_i := \Phi^{-1}(\hat{F}(d(\theta_{i+1}^H, \theta_i^H)))$ on Gaussian margins, where $\hat{F}$ denotes the empirical distribution function of the set of changes in wave directions. Assume that $\{\delta_i : i \in \mathcal{I}\}$ are realisations of the random variable $\Delta$. We estimate the following autoregressive model for $\Delta_t$ of order
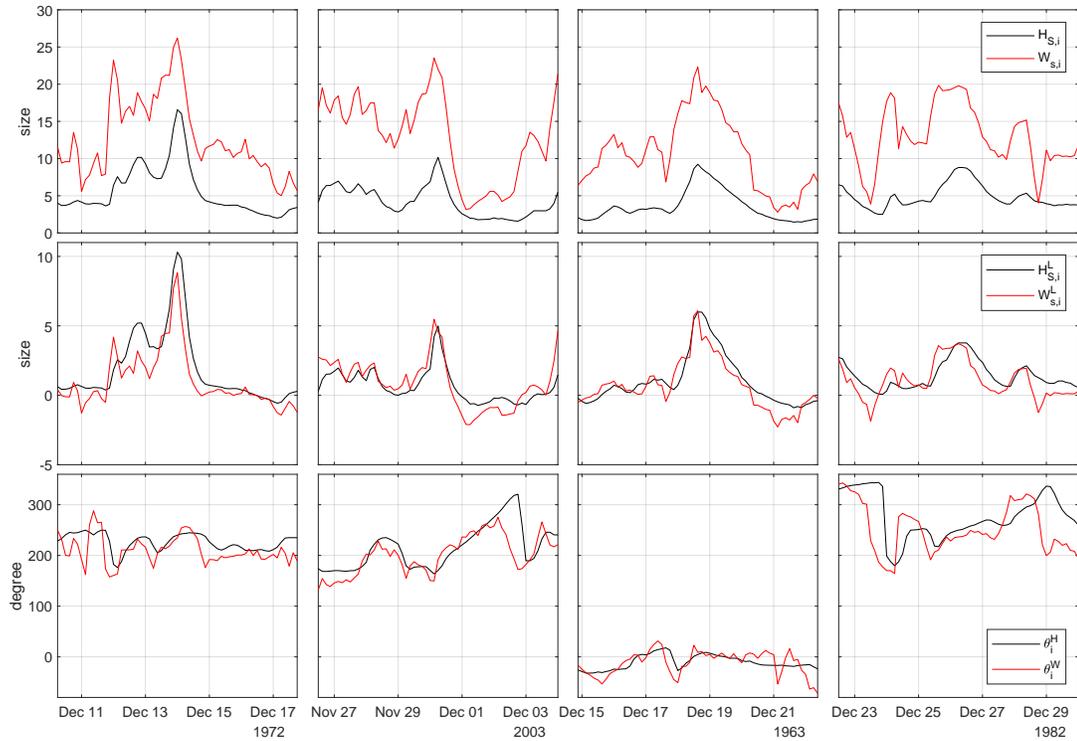
Figure 5.4.1: Intervals of oceanographic time-series: (top) key variables: significant wave height $H_{S,i}$ and wind speed $W_{s,i}$ on original margins; (middle) on Laplace margins; (bottom) covariates: wave direction $\theta_i^H$ and wind direction $\theta_i^W$. The four columns correspond to time periods that contain the 100%, 95%, 90% and 85% empirical percentiles of $H_{S,i}$, respectively.
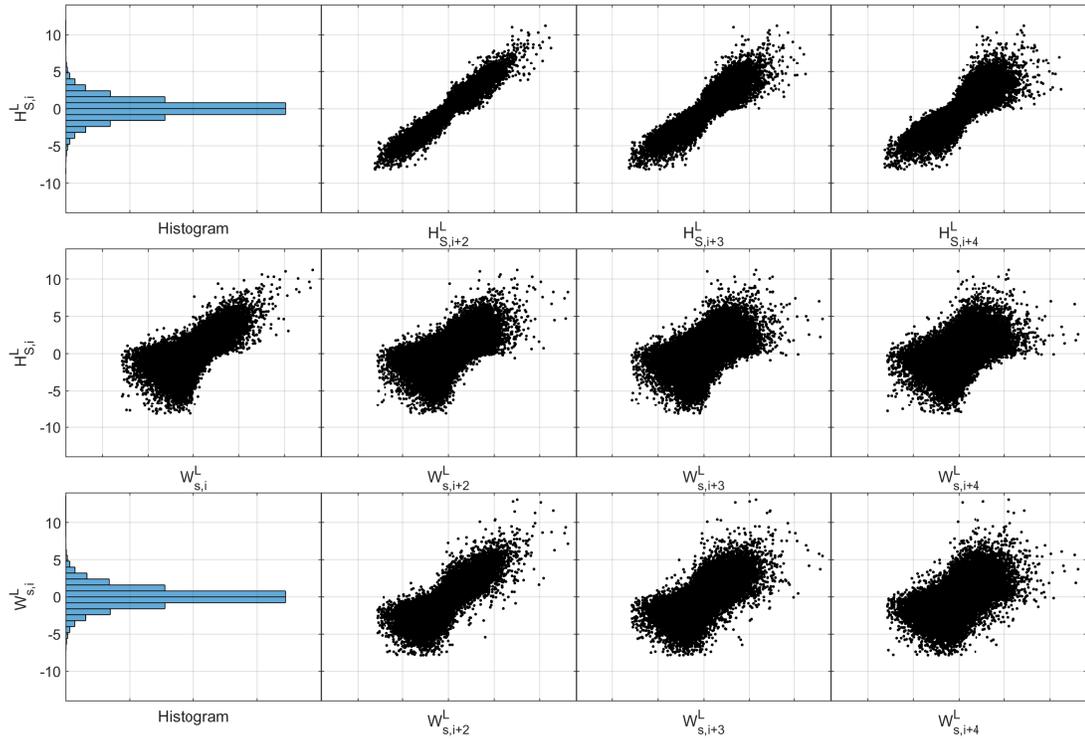
Figure 5.4.2: Matrix plot of observed $H_{S,i}^L$ and $W_{s,i}^L$ at various time lags up to lag 4 including cross dependece.
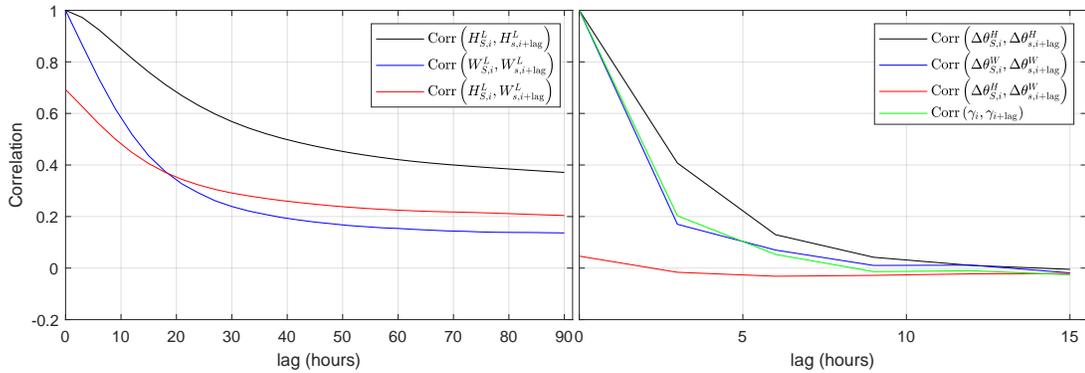


Figure 5.4.3: Estimated correlation and cross-correlation at various time lags of: (left) the key variables on Laplace margins: $H_{S,i}^L$ and $W_{s,i}^L$; (right) the covariates: delta wave direction $\Delta\theta_i^H := \theta_{i+1}^H - \theta_i^H \mod 360$, delta wind direction $\Delta\theta_i^W := \theta_{i+1}^W - \theta_i^W \mod 360$ and $\gamma_i$, see definition (5.4.2).

$p_1 = 1, 2, 3, \ldots$ with parameters $\varphi_j^{\mathrm{H}} \in \mathbb{R}$ for $j = 1, \ldots, p_1$ as

$$\Delta_t = \sum_{j=1}^{p_1} \varphi_j^{\mathrm{H}} \Delta_{t-j} + \zeta(H_{S,t}) \varepsilon_t, \tag{5.4.1}$$

where $\varepsilon_t$ is a standard Gaussian random variable, and standard error $\zeta(h)$ is given by

$$\zeta^2(h) = \lambda_1 + \lambda_2 \exp(-\lambda_3 h)$$

with $\lambda_{j'} > 0$ for $j' = 1, 2, 3$, see Tendijck et al. (2019). In particular, the standard error $\zeta(h)$ decays as $h$ grows due to the significantly larger amounts of energy needed to change the direction of large waves compared to small waves. The parameters of this model are inferred with maximum likelihood, and as opposed to the inference discussed in Section 5.3.5, we do not reject the assumption that $\varepsilon_t$ is a standard Gaussian. In practice, we use $p_1 = 1$ in line with Tendijck et al. (2019).

Given model (5.4.1) for $\theta_t^H$, we propose the following model for wind direction $\theta_t^W$ conditional on wave direction $\theta_t^H$:

$$\theta_t^W = \theta_t^H + \gamma_t \quad \mathrm{mod}\ 360, \tag{5.4.2}$$

where $\gamma_t$ is a zero-mean stationary $\mathrm{AR}(p_2)$ process, i.e., there exist parameters $\varphi_j^{\mathrm{W}} \in \mathbb{R}$, $1 \leq j \leq p_2$, and a non-degenerate residual distribution $\tilde{\varepsilon}_t$ independent of $\gamma_{t-j}$ for $j \geq 1$, such that

$$\gamma_t = \sum_{j=1}^{p_2} \varphi_j^{\mathrm{W}} \gamma_{t-j} + \tilde{\varepsilon}_t$$

and such that the polynomial $1 - \sum_{j=1}^{p_2} \varphi_j^{\mathrm{W}} z^j$ has roots outside the unit circle. The model parameters and the distribution of $\tilde{\varepsilon}_t$ are inferred as described in Section 5.3.5 conditional on the model order $p_2$, which is selected by investigating the correlation function in Figure 5.4.3 and an unpublished figure of the partial autocorrelation function of $\gamma_t$. In our application, we conclude that $p_2 = 1$ is sufficient.

### 5.4.3   Response variable

To measure the practical impact of extreme met-ocean excursions, we define structure response variables for a simple hypothetical offshore facility. We consider a platform

in the form of a unit cube standing above the water, supported by thin rigid legs, with vertical cube faces aligned with cardinal directions. Only wave and wind impact on the cube itself is of interest to us, and we neglect the effects of swell, surge, tide, etc., and potential climate non-stationarity. For simplicity, we assume that when $H_S < h$, for some threshold $h > 0$, the wave impact on the structure is negligible, and structural response is dominated by wind. When $H_S \geq h$, we assume that wave impact increases cubically with $H_S$ and quadratically with $W_s$. Hence, see Morison et al. (1950) and Ma and Swan (2020) for supporting literature, the impact of an extreme excursion on the facility is defined by the instantaneous response variable $R$

$$
R(H_S, W_s, \theta^H, \theta^W; c, h) = \begin{cases} c \cdot I_W^2(W_s, \theta^H - \theta^W) & \text{if } H_S < h, \\ c \cdot I_W^2(W_s, \theta^H - \theta^W) + A(\theta^H) \cdot (H_S - h) \cdot H_S^2 & \text{if } H_S \geq h, \end{cases}
$$

where $I_W : \mathbb{R}_{>0} \times [-180, 180) \to \mathbb{R}$ is the inline wind-speed, see below, $A : [-180, 180) \to [1, \sqrt{2}]$ is the exposed cross-sectional area of the cube, see below, and the parameter $c > 0$ is specified such that both significant wave height and wind speed have an approximately equal contribution to the largest values of $R$.

The exposed cross sectional area $A(\theta) \in [1, \sqrt{2}]$ of the cube is given by

$$
A(\theta^H) := 1/\cos([(\theta^H + 45; \mod 90) - 45] \cdot \pi/180),
$$

for a given wave direction $\theta$. The inline wind-speed $I_W$ is the component of the wind speed in the direction of the wave and is given by

$$
I_W(W_s, \theta^H - \theta^W) = W_s \cos((\theta^H - \theta^W) \cdot \pi/180).
$$

To simplify notation, we write $R_i(c, h) := R(H_{S,i}, W_{s,i}, \theta_i^H, \theta_i^W; c, h)$ for $i \in \mathcal{T}$.

To define a structure response for a complete excursion $\mathcal{E}_u$, we write

$$
\mathcal{E}_u := \{(H_{S,i}, W_{s,i}, \Theta_i^H, \Theta_i^W) : a \leq i \leq b\}
$$

for some $a < b$ such that for a threshold $u > 0$ (on Laplace margins) $H_{S,i}^L > u$ for $a \leq i \leq b$ and $H_{S,a-1}^L, H_{S,b+1}^L \leq u$. Next, let $i^* := i^*(\mathcal{E}_u)$ be the time of the excursion maximum, i.e., $H_{S,i^*}$ is the maximum of $H_{S,i}$ over $\mathcal{E}_u$.

We define two natural structure response variables representing the maximum impact of an excursion $\max_{\{a \leq i \leq b\}} R_i(c, h)$, and the cumulative impact of an excursion $\sum_{\{a \leq i \leq b\}} R_i(c, h)$, respectively. For our application, we consider slight alterations $R^{\max}(c, h, \mathcal{E}_u)$ and $R^{\text{sum}}(c, h, \mathcal{E}_u)$

$$R^{\max}(c, h, \mathcal{E}_u) := \max_{\{a \leq i \leq b, \ |i - i^*| > 2\}} R_i(c, h), \qquad R^{\text{sum}}(c, h, \mathcal{E}_u) := \sum_{\{a \leq i \leq b, \ |i - i^*| > 2\}} R_i(c, h),$$

i.e., we consider versions that do not depend directly on the characteristics of the excursion near the excursion maximum. This choice is made only to exaggerate the dependence of the structure variables on the pre-peak and post-peak periods compared to the period of the peak, and hence the importance of estimating good models for the pre-peak and post-peak periods using MMEM or EVAR. Moreover, we define $R^{\max}(c, h)$ and $R^{\text{sum}}(c, h)$ as the random variables of the structure responses related to random excursions.

### 5.4.4   Model comparisons

Here, we use our time-series models to characterise extreme excursions for the met-ocean data of Section 5.4.1 with structure responses $R^{\max}$ and $R^{\text{sum}}$. First, we describe our model comparison procedure, and then, assess model performance using a visual diagnostic.

To compare MMEM and EVAR with each other and with HM, we take a similar approach to Gandy et al. (2022). We select at random 25% of the observed excursions for our training sample; the remaining 75% forms our test sample. We calculate performance statistics, which we derive below, by averaging over 50 such random partitions of the sample.

For training, we fit EVAR, $\text{EVAR}_0$ and MMEM with model orders $k = 1, 2, \ldots, 6$. The fitting of these 18 models is a two-stage procedure. In the first part, we fit (six) conditional extremes models for the period of the peak $\mathcal{P}_0^k$ for each $k$. In the second part, we fit $2 \cdot 18 = 36$ models to the pre-peak $\mathcal{P}^{\text{pre}}$ and post-peak $\mathcal{P}^{\text{post}}$ periods. For each of the 18 models and HM, we simulate $20,000$ excursions, calculate structure response variables $R^{\max}$ and $R^{\text{sum}}$, and compare distributions of simulated

structure response variables with those corresponding to the withheld test data. This is achieved by defining a distance function $D$ that measures the level of difference in tails of distribution functions. We select 20 equidistant percentiles $p_1, \ldots, p_{20}$ ranging from 97% to 99.9% corresponding to moderately extreme to very extreme levels with respect to the (smaller) training sample but not too extreme for the (larger) withheld data. Mathematically, we define distance $D$ of distribution functions $F_M$ (of model $M$) and $F_E$ (an empirical distribution function) as the mean absolute relative error over these percentiles, i.e.,

$$D(F_M, F_E; p_1, \ldots, p_{20}) = \frac{1}{20} \sum_{j=1}^{20} \left| \frac{F_E^{-1}(p_j) - F_M^{-1}(p_j)}{F_E^{-1}(p_j)} \right|.$$

We remark that in the above definition, we never divide by zero because we only calculate $D$ to measure the distance of the distributions of positive random variables.

In Figure 5.4.4, we show the results for the 50 random partitions of the original sample by plotting the average distance $D$ (with 80% confidence intervals) of each model together with HM for four different structure response variables corresponding to two choices of $c$ and $h$ for each of $R^{\mathrm{max}}$ and $R^{\mathrm{sum}}$. Note that similar studies for other values of $c$ and $h$ for $R^{\mathrm{max}}$ and $R^{\mathrm{sum}}$ were examined, and general findings are consistent with those illustrated in Figure 5.4.4. For readability, we omit the confidence bands for $\mathrm{EVAR}_0$ since the difference with EVAR is minimal. Our model selection procedure now involves selecting the model that yields the smallest average dissimilarity distance $D$ whilst keeping the model order as low as possible.

We make a number of observations. For the $R^{\mathrm{max}}$ response, EVAR and MMEM clearly outperform HM whatever order of the Markov chain is taken for $k = 1, \ldots, 6$. However, for the $R^{\mathrm{sum}}$ response, high order (e.g., $k = 4, 5, 6$) EVAR and MMEM are necessary to be competitive with HM. We observe also that performance of EVAR and MMEM does not significantly improve or worsen for $k > 4$. This finding is further supported with an unpublished study with Markov model orders of $k \leq 10$. On our measures of fit, large values for $k$ do not induce overfitting. For this response, model performance for MMEM is relatively insensitive to the choice of model order $k$, whereas for EVAR, a model order of at least $k = 4$ is necessary for competitive

performance. This said, illustration of excursions in Appendix F demonstrate that MMEM with $k = 1$ does not explain the variability of the pre-peak and post-peak periods well. Finally, we remark that for different choices of $c$ and $h$, we obtain broadly the same type of results.

By looking at the average relative errors in $R^{\max}$ and $R^{\mathrm{sum}}$ of our proposed selection of methods, we conclude that a third order MMEM and a fourth order EVAR are competitive models within their class. Since these models have similar performance, we prefer EVAR(4) over MMEM(3) because of its simpler two-dimensional residual distribution.
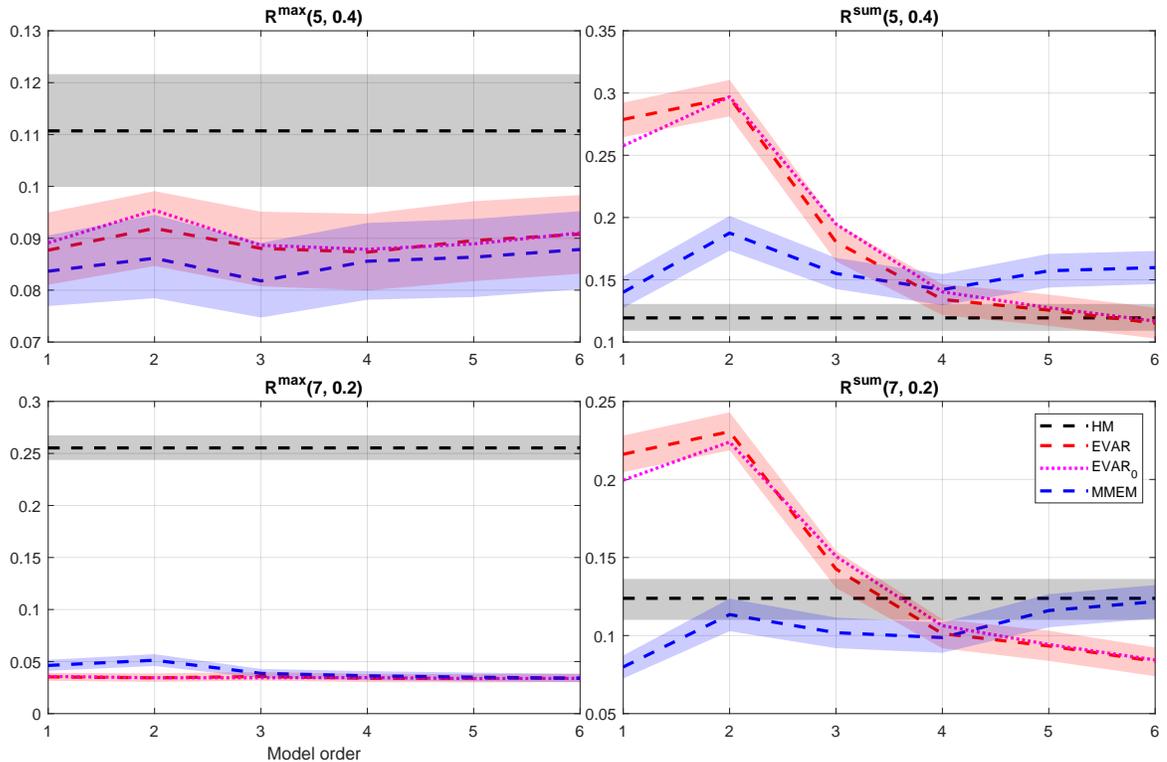


Figure 5.4.4: Average mean relative errors of HM, EVAR, $\mathrm{EVAR}_0$ and MMEM (dashed/dotted) and 80% confidence regions (shaded) for estimating the distribution of structure responses using 25% of data for training and 75% of data for testing. For details, see the text.

Next, we fit the models to $\mathcal{D}$ and illustrate model characteristics for EVAR(4) in Figure 5.4.5. We plot simulated excursions of EVAR(4) such that the excursion

maximum significant wave height takes on values between 11.5m and 12.5m (left) and we visually compare these with observed excursions for the same interval of excursion maxima (middle). On the right, we summarize simulated and observed excursions in terms of the median, the 10% and 90% percentiles. Finally, in the bottom panel we plot the survival probability for an excursion relative to the time of the excursion maximum, conditional on the excursion maximum taking a value between 11.5m and 12.5m, i.e., we plot

$$\mathbb{P}\left(\min\{H_{S,i}^L: \ i = \min(0,\tau),\ldots,\max(0,\tau)\} > u \ \Big| \ H_{S,0} \in [11.5, 12.5]\right) \qquad (5.4.3)$$

for $\tau \in \mathbb{Z}$. In Appendix F, we produce analogous plots for each of the 18 models considered and HM. We observe that EVAR(4) characterizes the period of the peak, and also the pre-peak and post-peak periods of the excursion well. Moreover, EVAR(4) also reproduces the observed excursion survival probability.

## 5.5   Conclusion

In this paper, we provide models for extreme excursions of multivariate time-series. Excursions are characterized by a three-stage modelling procedure for the period of the peak, the pre-peak and the post-peak periods. We model the period of the peak using the conditional extremes framework (Heffernan and Tawn, 2004), and for the pre-peak and post-peak periods, we define two classes of time-series models: MMEM, motivated by the Markov extremal model of Winter and Tawn (2017); and EVAR, an extreme-value extension of a vector autoregressive model. We compare these excursion models with a baseline historical-matching method, motivated by Feld et al. (2015). We find that the excursion models are at least competitive with historical-matching and often outperform it in the estimation of the tail of a range of notional structure response variables for a met-ocean application in the northern North Sea.

   Statistical modelling of extreme excursions of multivariate time-series is difficult as it requires the estimation of complex model forms. MMEM requires the estimation of the conditional distribution of high-dimensional residual random variables and
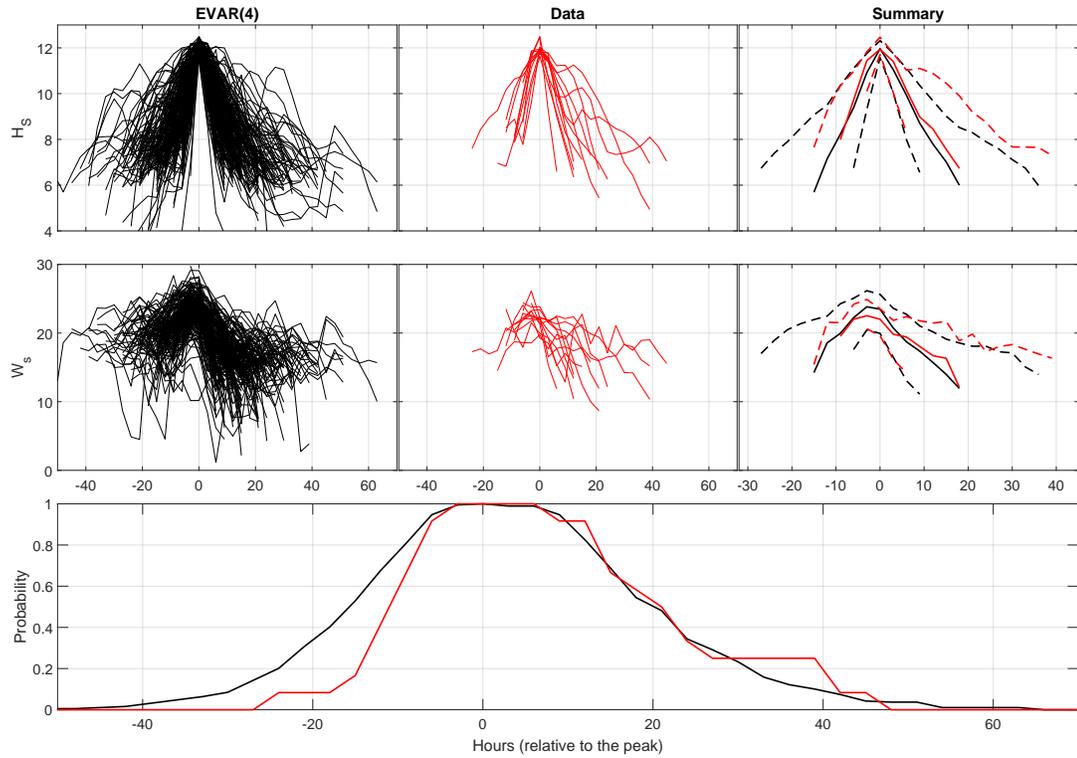
Figure 5.4.5: Excursions of $H_S$ and $W_s$ from EVAR(4) model (left; black), and data (middle; right) on original margins such that storm peak significant wave height is in $[11.5, 12.5]$; (right) summaries of the data (black) and EVAR(4) (red) excursions: median (solid), and the 10% and 90% quantiles (dashed). In the bottom panel, we plot survival probabilities for observed (black) and EVAR(4) (red) excursions relative to the time of the excursion maximum, see equation (5.4.3).

EVAR is highly parameterized. Nevertheless, for realistically sized directional samples of significant wave height and wind speed time-series, we found that MMEM(3) and EVAR(4) perform well. Even when the empirical historical-matching procedure is competitive, adoption of an excursion model is advantageous because it allows for rigorous uncertainty quantification. We expect that our excursion models are applicable more generally, e.g., for the modelling of higher-dimensional met-ocean data and spatial fields.

# Chapter 6

# Conclusions and further work

In this chapter, we look back on the conclusions we draw from this thesis and summarize how our work fits into the broader scope of the field of met-ocean inference as described in the Introduction. Moreover, we also point out directions for further work.

We recall that in the Introduction we specified a five-step statistical procedure for estimating the distribution of extreme responses to offshore structures corresponding to extreme weather events. Step 1 involved estimating the distribution of the long-term met-ocean environment, and steps two to five convert this information into distributions on responses of offshore structures. The overall goal of the thesis was to answer questions that arise for the modelling procedures in step 1.

In Chapter 3, we propose an extension to the conditional extremes model (Heffernan and Tawn, 2004) such that it is appropriate to be applied to bivariate mixture distributions. This extension was motivated by a very specific element of step 1: the understanding of the dependence of the joint distribution of significant wave height and wave period for large wave periods. Modelling this part of the distribution could be relevant to applications where resonance frequencies occur at rare large wave periods.

For the bivariate mixtures extension, we describe two methods of inference. The first one is a novel approach for inference relying on estimating model parameters

using quantile regression. The problem of this method, however, is that it cannot be extended to higher dimensions in the absence of a theory on multivariate quantile regression. On the other hand, the second method for performing inference - which is derived by using likelihoods - does extend naturally to higher dimensions. Either way both of these methods work reasonably well when considering the two-dimensional application.

There are a variety of ways how one can extend this work into further research. For example, one can consider generic multivariate applications for mixture distributions. For example a model for wind speed and significant wave height conditional on wave period requires a three-dimensional mixture model with at least two mixture components. Some of the many challenges that need to be taken into account when the multivariate extension of this model is applied:

- Model misspecification can play a more significant role.

- Subasymptotic models might be necessary to use when the number of mixture components is large and, thus, smaller thresholds need to be used to keep enough data per mixture component for reliable inference.

- Data sparsity of higher dimensional applications needs to be addressed: what are the practical limits of this model in higher dimensions?

- The proposed procedure to select the number of mixture components would need to be reassessed: can we comfortably estimate the number of modes of a $(d-1)$-dimensional residual distribution? Do these estimates actually correspond to the actual number of mixture components in applications where we use synthetic data?

A different but also useful way to extend on the work is by exploring the subasymptotic variant of the mixture model as discussed in Section 3.4, in which the mixture probability is a function of the conditioning variable. In particular, from a theoretical point of view it is interesting to derive bounds on the $\beta$-parameters for the mixture components that are associated with a mixture probability function that tends to 0

when the conditioning variable gets closer to its upper end point. On the other hand, from a practical point of view, we remark that the quantile regression based inference methodology is probably not applicable for estimating these subasymptotic variants - and if they are applicable, i.e., when parametric assumptions on the residual distribution and the mixture probability function are made, then the quantile regression methodology gives up its beneficial non-parametric nature whilst keeping its overly complicated form. We believe that a suitable adjustment to the likelihood approach can be made for estimating simultaneously a parametric mixture probability function and the mixture model parameters; for example by using penalized likelihoods. However, further work would be needed to assess what works here, both in theory and in practice.

In Chapter 4, we contribute to the field of multivariate extreme value theory by providing a mathematical toolset for calculating extremal properties of conditional models. This work is important for increasing the understanding of a researcher or engineer on the extremal assumptions that (conditional) models make. In particular, this understanding is valuable for models which are currently being used for modelling step 1, e.g., Haver and Winterstein (2009). We remark here that our theoretical work does not just limit itself to conditional models but is generally applicable to calculate the rate of convergence of analytically intractable integrals. So, other extreme value applications could potentially be found, but we didn't search for those in this work.

There are many possibilities for extending the current work. The most obvious one is to add a corollary to Proposition 4.2.2 which would be more specific to calculating extremal properties of conditional models. We recall that we use a very general $g_n$ in the theorem, whereas we only apply the theorem to a survival function multiplied with a density function. So, it is natural to search for a more specific result in terms of the survival and density function. A second possibility is to include different extremal characteristics of conditional models like $\tau_C$ or $\lambda$, see Section 2.4.2.

Another possibility for extending the work is to explore subasymptotic extreme value characteristics of conditional models. For example, the curve $\eta(p)$ for $p$ close to 1 contains more information about the distribution than the limiting quantity $\eta$, and

thus it would be interesting to consider the difference of two models: for example, one that agrees with $\eta(p)$ for $p \in [0, 1 - 10^{-6}]$, and another one that agrees only with the limit $\eta$. From a different point of view, can we develop conditional models that agree with empirical estimations of $\eta(p)$? We remark that theory from Chapter 4, is only applicable to calculate limiting values, and thus this would mean that either subasymptotic models need to be derived or one would instead need to focus on numerical methods.

Finally, the current work lends itself to be extended to higher dimensions relatively well. However, we do remark that creating generic theory for applications with dimension $d > 2$ becomes increasingly harder because of the vast variety of dependence models to consider: e.g., multivariate Heffernan-Tawn, graphical models with conditional independence (Engelke and Hitz, 2020), vine copula dependence models (Simpson et al., 2021), and more.

In Chapter 5, we address one of the main open problems in met-ocean variables: there is a need for a parsimonious multivariate temporal model that has a foundation in extreme value theory to describe the temporal evolution of a storm. So, we extend the Heffernan-Tawn model to obtain a temporal model that does exactly that.

There are many further research opportunities that build on the research of this chapter. The most apparent one is to extend the case study to include additional oceanographic variables at the same location, include multiple variables at different locations, or redefine a cluster by requiring, for example, that significant wave height exceeds some threshold at least one of the considered locations.

When extending the models spatially, it might be more realistic to make smoothing assumptions - in space - on the parameters of the models. Moreover, assumptions on the dependence structure of higher dimensional residual distributions need to be assessed. Can we use kernel density estimates or do we need to assume parametric marginals and/or parametric copula models?

A different extension to this work is to find asymptotic arguments for or against the extremal vector autoregression (EVAR) model introduced in Chapter 5. Currently EVAR does not have a strong theoretical foundation but is a merger of two widely

used models: vector autoregression - widely used in time-series applications - combined with the Heffernan-Tawn model - widely used in multivariate extremes applications.

Another extension would merge the mixture modelling ideas of Chapter 3 with the temporal extremes model such that a statistical model can be developed for the evolution of significant wave height and wave period during extreme events. This leads to some interesting questions: for example, at a given time step and some historical observations, can we estimate the probability that the next observation is part of a different mixture component than the previous?

The three novel methodological chapters in this thesis merely answer some of the statistical questions about met-ocean variables that arise in step 1. With this respect, many more questions can still be asked and answered. However, we do believe that a suitable extension of Chapter 5 yields a solution to solving step 1 of the procedure for modelling large return values induced by extreme met-ocean conditions.

We do note that such an extension is still far away from being accomplished since models for many features would still need to be developed, assessed, and/or combined with each other. We list some of these:

- a model for the period of down-time relating to observations below a threshold and the size of such observations;

- a seasonal model that allows storms to be simulated at different times of the year;

- a temporal model for the multivariate extreme events with dimension $d > 2$;

- a non-stationary model for the time-series that incorporates climate change.

Moreover, the ideal model would additionally be able to use efficient methods for estimation of large return periods. Finally, the models in steps 2 to 5 need to be combined with the ideal model for step 1 such that uncertainty in estimating large return periods and their joint dependencies is propagated properly to yield sensible conclusions.

# Appendix A

# Appendix to Chapter 3

## A.1   Proof of Theorem 3.4.1

Let $X, Y \sim \text{Laplace}(1)$. We consider the following simplified version of model (3.4.2)

$$Y|X \sim \alpha_k X, \text{ with probability } p_k(X) \tag{A.1.1}$$

with $0 \leq \alpha_k < \alpha_{k'}$ for all $1 \leq k \leq k' \leq K$, and where $p_k : \mathbb{R} \to [0, 1]$ are functions such that for all $x$, $\sum_{k=1}^{K} p_k(x) = 1$. We will find a necessary condition on $\alpha_k$ 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 the simplified model (A.1.1), we simplify this probability as follows

$$
\begin{aligned}
\mathbb{P}(Y > y, X > u) &= \sum_{k=1}^{K} \mathbb{P}(X > \max\{y/\alpha_k, u\}, \ J = k) \\
&= \sum_{k=1}^{K} \int_{\max\{y/\alpha_k, u\}}^{\infty} \mathbb{P}(J = k | X = x) f_X(x) \, \mathrm{d}x \\
&= \frac{1}{2} \sum_{k=1}^{K} \int_{\max\{y/\alpha_k, u\}}^{\infty} p_k(x) e^{-x} \, \mathrm{d}x \\
&= \frac{1}{2} e^{-\max(y/\alpha_1, u)} + \frac{1}{2} \sum_{k=2}^{K} \int_{\max\{y/\alpha_k, u\}}^{\max\{y/\alpha_{k-1}, u\}} \left( \sum_{i=k}^{K} p_i(x) \right) e^{-x} \, \mathrm{d}x, \tag{A.1.2}
\end{aligned}
$$

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 $\alpha_1 < \alpha_2 < \cdots < \alpha_j \leq 1 < \alpha_{j+1} < \cdots < \alpha_K$.

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

$$\alpha_{i_0} = \liminf_{x \to \infty} -\frac{\log p_{i_0}(x)}{x} + 1 + \varepsilon \tag{A.1.3}$$

for some $\varepsilon > 0$. Note that this expression excludes $i_0 \leq j$ as $\liminf_{x \to \infty} -\frac{\log p_{i_0}(x)}{x} + 1 + \varepsilon > 1$. Hence, we deduce that $i_0 > j$. Now, define $A_{i_0}(x) = p_{i_0}(x)e^{(\alpha_{i_0}-1)x}$. Equation (A.1.3) implies

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

Hence, there exists an $x'$ such that for all $x > x'$, $A_{i_0}(x) > \alpha_{i_0}$. For $y > a_K \max\{x', u\}$, we get that equation (A.1.2) simplifies as follows

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

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, $\liminf_{x \to \infty} A_{i_0}(x) > \alpha_{i_0}$ contradicts with the marginal distribution of $Y$. We conclude that for all $i \geq j+1$, we need to have $\liminf_{x \to \infty} p_i(x)e^{(\alpha_i-1)x} \leq \alpha_i$. So,

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

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

# Appendix B

# Supplementary Information to Chapter 3

This document contains details corresponding to the main paper and is organised as follows. In Section B.1, we present a pseudo code algorithm for fitting the HT mixture model. In Section B.2, we describe how one can estimate probability measures using the two methods, $\mathrm{HT}(K)$ and $\mathrm{QR}(K)$, introduced in the main paper. In particular, one can find how one simulates observations from both models. Section B.3.1 contains details on preprocessing of oceanographical data to gain a sample without temporal dependence and directional dependence. Section B.3.2 shows trace plots of MCMC chains of the models fitted to the data. Calculations with respect to the example in Section 2.4 from the main paper are shown in Section B.4. In Section B.5, we discuss additional details that are highlighted throughout the main paper. Finally, we give a summary of the simulation study in Section B.6. The raw simulation study results are found in Section B.7.

## B.1  Pseudo code for the HT mixture model

We present the pseudo-code for the pseudo-Bayesian inference procedure as described in Section 3.1 of the main paper in Algorithm 1.

Here, $B$ represents the length of the MCMC chain, and $A$ is the minimal amount

of burn-in before we start the adaptive MCMC procedure, which updates the stepsize for more efficient steps in the MCMC, see Roberts and Rosenthal (2009). Moreover, instead of recalculating $\sigma_\theta^2 := \sigma_{\theta;b}^2 = \mathrm{Var}\left(\theta_{1:b-1}\right)$, we update this value in an online fashion by using the following equations

$$\mu_{\theta;b} = \mu_{\theta;b-1} + \frac{1}{b}(\theta_{b-1} - \mu_{\theta;b}) \ \ \text{and} \ \ \sigma_{\theta;b}^2 = \frac{b-1}{b}\sigma_{\theta;b-1}^2 + \frac{b-1}{b^2}(\theta_{b-1} - \mu_{\theta;b-1})^2.$$

We then follow Roberts and Rosenthal (2009) and simulate $\theta_b$ from a linear combination of two normals. The shape of the first normal will be in line with the real shape of the posterior distribution. The second normal ensures mixing. Finally, we allocate data to a mixture component $k$ at random using the probabilities that are calculated using equation (10).

**Data:** $(x_i, y_i)$, $i$ from 1 to $n$

Initialization;

Transform data to Laplace margins;

**for** $b = 1 : B$ **do**

    **if** $b > A$ **then**

        **for** $k = 1 : K$, $\theta = [\alpha_k, \beta_k, \mu_k, \sigma_k]$ **do**

            Update $\sigma_\theta^2 = \mathrm{Var}\,(\theta_{1:b-1})$;

            Simulate $\theta_b$ from a linear combination of a

            $\mathcal{N}\left(\theta_{b-1}, 2.38^2/(4K) \cdot \sigma_p^2\right)$ with weight 0.95 and $\mathcal{N}(\theta_{b-1}, 0.05^2)$

            with weight 0.05.

        **end**

    **else**

        Simulate $\theta_b$ from a $\mathcal{N}(\theta_{b-1}, 0.05^2)$.

    **end**

    **for** $k = 1 : K$ **do**

        **for** $\theta = [\alpha_k, \beta_k, \mu_k, \sigma_k]$ **do**

            Allocate data to a mixture component $k$ using equation (10) where

              the model parameters are given by the previously accepted

              parameters and the current proposed $\theta$;

            Compute the augmented log-llikelihood of component $k$;

            Accept parameter $\theta$ using the Metropolis-Hastings acceptance

              ratio;

        **end**

    **end**

**end**

    **Algorithm 1:** HT mixture model implementation in pseudo-code.

## B.2 Estimating probability measures of extreme sets

### B.2.1 Using the Heffernan-Tawn mixture model

Let $S \subseteq \mathbb{R}^2$ be a measurable set that is extreme in $X$ and for which we want to estimate the probability $\mathbb{P}((X, Y) \in S)$. We define the following key properties of the set $S$: (i) the lower bound of $S$ in the $x$-dimension $x_l^S := \inf\{x \in \mathbb{R} : \exists y \in \mathbb{R} : (x, y) \in S\}$, which is assumed to be large in the Laplace domain; (ii) the upper bound of $S$ in the $x$-dimension $x_u^S := \sup\{x \in \mathbb{R} : \exists y \in \mathbb{R} : (x, y) \in S\}$ with $0 < x_l^S < x_u^S \leq \infty$; and (iii) the slice of the set $S$ given the $x$-component $Y^S(x) := \{y \in \mathbb{R} : (x, y) \in S\}$ for $x_l^S < x < x_u^S$. Using the HTM model formulation, we estimate $\mathbb{P}((X, Y) \in S \mid \boldsymbol{\theta})$ conditional on the HTM model parameters $\boldsymbol{\theta}$, by conditioning on $X = x$ and integrating over $[x_l^S, x_u^S]$, i.e., with

$$\sum_{k=1}^{K} \hat{p}_k \int_{x_l^S}^{x_u^S} \left( \frac{1}{n_k} \sum_{i=1}^{n} \mathbb{1}\left\{ \hat{\gamma}_k + \hat{\alpha}_k x + x^{\hat{\beta}_k} \hat{z}_{ki} \in Y^S(x), j_i = k \right\} \right) f_X(x) \, dx, \qquad \text{(B.2.1)}$$

where $f_X(x) = \exp(-|x|)/2$ is the density of a standard Laplace distribution. The contribution per component is calculated by integrating out the conditioning random variable. In this formula, we sum over the $K$ mixture components weighting the contribution of each component $k$ by their respective estimated mixture probabilities $\hat{p}_k$. Finally, we use the empirical estimator

$$\hat{\mathbb{P}}(Y \in Y^S(x), J = k | X = x, \boldsymbol{\theta}) := \frac{1}{n_k} \sum_{i=1}^{n} \mathbb{1}\left\{ \hat{\gamma}_k + \hat{\alpha}_k x + x^{\hat{\beta}_k} \hat{z}_{ki} \in Y^S(x), j_i = k \right\}.$$

Equation (B.2.1) suggests two estimators: one that uses numerical integration and one that uses stochastic integration, both of which induce negligible errors compared to the model estimation uncertainty. We use stochastic integration since it is computationally cheaper than its numerical counterpart.

We simulate a suitably large number $M$ of observations of $(X, Y)|(x_l^S < X < x_u^S)$. The proportion of this sample that fall into the set $S$ is our estimator of $\mathbb{P}((X, Y) \in S \mid x_l^S < X < x_u^S)$. Using the binomial distribution, we can estimate the variance of this estimator with respect to the size of $M$. We choose $M$ large enough such that the estimator is suitably precise. This algorithm has a relatively short computational

time as the value of $M$ does not need to be excessively large since we can directly get a realisation of the conditional random variable $X|(x_l^S < X < x_u^S)$.

To simulate $M$ observations from the model, we first sample a mixture component $1 \leq k \leq K$ with probability $\hat{p}_k$. Next, we simulate a realisation $U$ of a standard uniform distribution, and calculate

$$\tilde{x} = x_l^S - \log\left(1 - u(1 - \exp(x_l^S - x_u^S))\right) \tag{B.2.2}$$

to simulate from a standard Exponential random variable, truncated between $x_l^S$ and $x_u^S$. Finally, we draw $z^{(k)}$ independently of $\tilde{x}$ from the residual distribution of component $k$ by resampling residuals, and calculate

$$\tilde{y} = \hat{\gamma}_k + \hat{\alpha}_k \tilde{x} + \tilde{x}^{\hat{\beta}_k} z^{(k)}. \tag{B.2.3}$$

Using this simulation procedure, we generate a sample $\{(\tilde{x}_i, \tilde{y}_i) : i = 1, \ldots, M\}$ from the model conditional on $x_l^S < X < x_u^S$. This simulation methodology is summarised in pseudo code in Algorithm 2. The estimation of the probability measure of the set $S$ conditional on $x_l^S < X < x_u^S$ is now given by

$$\hat{\mathbb{P}}((X,Y) \in S | \boldsymbol{\theta},\ x_l^S < X < x_u^S) := \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}\left\{(\tilde{x}_j, \tilde{y}_j) \in S\right\}.$$

This equation implies the following unconditional estimator

$$\hat{\mathbb{P}}((X,Y) \in S | \boldsymbol{\theta}) = \hat{\mathbb{P}}((X,Y) \in S | \boldsymbol{\theta},\ x_l^S < X < x_u^S) \cdot \frac{1}{2}\left(e^{-x_l^S} - e^{-x_u^S}\right), \tag{B.2.4}$$

as the marginal distribution of $X$ is standard Laplace.

**Data:** $(x_i, y_i)$, $i$ from 1 to $n$

Initialisation;

Simulate $N$ realisations $\hat{x}_i$, $i = 1, \ldots, N$, from the marginal distribution of $X$,

   or from $X|(x_l < X < x_u)$ for some $x_l < x_u$, see equation (B.2.2);

**for** $i = 1 : N$ **do**

      Simulate a random index of the MCMC chain after burn-in;

      Extract the parameters at this index;

      Calculate residuals, allocate residuals randomly using equation (10), and

        empirically calculate allocation probabilities given the allocated residuals;

      Simulate which component to draw from and empirically sample from the

        residuals allocated to this component;

      Calculate the associated value $\hat{y}_i$ using equation (B.2.3);

**end**

**Algorithm 2:** Pseudo code for simulating $N$ observations from the Heffernan-Tawn mixture model with $K$ components.

Note that using the generated MCMC chain, we have a sample $\{\boldsymbol{\theta}^b : b = B_0, \ldots, B\}$ from the posterior distribution for $\boldsymbol{\theta}$, where $B$ is the number of iterations of the MCMC chain and $B_0 < B$ is the burn-in of the adaptive MCMC algorithm. Now, we obtain a sample from the posterior distribution of the probability measure of the set $S$ by

$$\left\{ \hat{\mathbb{P}}((X, Y) \in S | \boldsymbol{\theta}^b, x_l^S < X < x_u^S) : b = B_0, \ldots, B \right\}.$$

Randomness throughout the MCMC chain in the simulation of $\hat{\mathbb{P}}$ can be avoided through appropriate fixing of the random seed.

## B.2.2  Using the quantile-regression model

We focus on estimating $\mathbb{P}((X, Y) \in S \mid \boldsymbol{\omega})$ using the quantile-regression model fit where $S$ is as in Section B.2.1. We discuss two methods: one based on numerical calculation and one based on a combination of numerical calculation and simulation.

It is not straightforward to numerically estimate $\mathbb{P}((X, Y) \in S \mid \boldsymbol{\omega})$ within the quantile-regression formulation since $S$ can have a complicated shape. However, it is

straightforward to estimate probabilities of sets of the form $\{(x,y) \in \mathbb{R}^2 : \ l < x < u, \ \hat{q}_\tau(x) < y < \hat{q}_{\tau'}(x)\}$ for $l < u$ and $0 \le \tau < \tau' \le 1$. For $l > 0$

$$\mathbb{P}(l < X < u, \ \hat{q}_\tau(X) < Y < \hat{q}_{\tau'}(X)) \approx \frac{\tau' - \tau}{2} \cdot (\exp(-l) - \exp(-u)), \qquad \text{(B.2.5)}$$

where this approximation would be exact if we used the true value of $q_\tau(x)$ and $q_{\tau'}(x)$ instead of their corresponding estimators. We use the smoothed estimator $\hat{q}_\tau(x)$ as in Section B.5.5. The key of the numerical approach is to approximate the set $S$ as a disjoint union of sets of the form (B.2.5). Let $\varepsilon > 0$ and $N \in \mathbb{N}$ and define

$$\tilde{S} := \bigcup_{i=1}^{N} \{(x,y) \in \mathbb{R}^2 : \ a_i < x < b_i, \ \hat{q}_{c_i}(X) < Y < \hat{q}_{d_i}(X)\}.$$

for a real-valued sequences $(a_i, b_i, c_i, d_i)_{i=1}^{N}$. We can always find a real-valued sequence $(a_i, b_i, c_i, d_i)_{i=1}^{N} = (a_i(\boldsymbol{\omega}), b_i(\boldsymbol{\omega}), c_i(\boldsymbol{\omega}), d_i(\boldsymbol{\omega}))_{i=1}^{N}$ such that $\tilde{S}$ as defined above is the union of disjoint sets and

$$\mathbb{P}\left((X,Y) \in S \triangle \tilde{S} \mid \boldsymbol{\omega}\right) < \varepsilon,$$

where $A \triangle B := (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference between the sets $A$ and $B$. This means we could approximate the set $S$ with the disjoint union of sets for which we can estimate the measures using equation (B.2.5) to any level of desired precision relative to the probability measure. In particular, if $S$ when transformed to uniform margins is a convex set, then we can choose $b_i = a_{i+1}$ and $d_i = c_{i+1}$ for all $i$. For any desired level of numerical precision, we estimate

$$\hat{\mathbb{P}}((X,Y) \in S \mid \boldsymbol{\omega}) := \sum_{i=1}^{N} \frac{d_i - c_i}{2} \cdot (\exp(-a_i) - \exp(-b_i)). \qquad \text{(B.2.6)}$$

It is unclear how uncertainty propagates in equation (B.2.5) and there is no rule of thumb to inform grid spacing to ensure a given level of precision. Our unreported simulation studies show that using a relatively fine grid, i.e., $N \approx 1000$, does not require long computation times and that the numerical error is overshadowed by the model uncertainty.

We can also use a simulation based method as opposed to the numerical method described above. Using the same notation as before, we define $x_l^S$ and $x_u^S$, and we let

$M$ be a large number. Then for any $1 \leq j \leq M$, we simulate $\tilde{x}_j$ from a standard exponential truncated between $x_l^S$ and $x_u^S$ as described in Section B.2.1. Finally, sample $u$ from a standard uniform distribution and calculate $\tilde{y}_j = \hat{q}_u(\tilde{x}_j \mid \boldsymbol{\omega})$ using the quantile calibration method from Section B.5.5. This simulation methodology is summarised in pseudo code in Algorithm 3. We note that this calculation is quick as it only involves optimising a function once in a one-dimensional space. The estimation of the probability measure of the set $S$ conditional on $x_l^S < X < x_u^S$ is now given by

$$\hat{\mathbb{P}}((X, Y) \in S \mid x_l^S < X < x_u^S, \boldsymbol{\omega}) := \frac{1}{M} \sum_{j=1}^{M} \mathbb{1} \left\{ (\tilde{x}_j, \tilde{y}_j) \in S \right\}.$$

Similar to before, this implies an unconditional estimator $\hat{\mathbb{P}}((X, Y) \in S \mid \boldsymbol{\omega})$.

> **Data:** $(x_i, y_i)$, $i$ from 1 to $n$
>
> Initialisation;
>
> Perform quantile calibration as detailed in Section B.5.5;
>
> Simulate $N$ realisations $\hat{x}_i$, $i = 1, \ldots, N$, from the marginal distribution of $X$,
>   or from $X|(x_l < X < x_u)$ for some $x_l < x_u$, see equation (B.2.2);
>
> **for** $i = 1 : N$ **do**
> > Simulate $u \in (0, 1)$ uniformly;
> >
> > Use the calibrated quantile functions to find $\hat{q}_u(\cdot \mid \boldsymbol{\omega})$;
> >
> > Calculate $\hat{y}_i = \hat{q}_u(\hat{x}_i \mid \boldsymbol{\omega})$.
>
> **end**

**Algorithm 3:** Pseudo code for simulating $N$ observations from the quantile-regression model with $K$ components with parameters $\boldsymbol{\omega}$.

## B.3   Data analysis

### B.3.1   Preprocessing

**Removing temporal dependence**

The peak-picking method consists of three steps: (i) separate the observations into temporal clusters, such that for any observation within the cluster there exists at least one observation, recorded within 24 hours, with an $H_S$ value larger than the

empirical 70% quantile of the marginal distribution of $H_S$; (ii) collect per cluster the maximal value $H_{S,peak}$ of $H_S$, the associated wave period $T_{2,ass}$ and wave direction $D_{ass}$ corresponding to the same occurence time as $H_{S,peak}$; (iii) delete the observations for which direction $D_{ass}$ is outside $[182 \deg, 266 \deg]$. This procedure reduces the $176,765$ observations recorded over the period $1957 - 2018$ to $1597$. See Figure 1 in the main paper for a scatter plot of the data and the storm peak sample.

**Removing directional dependence**

Wave directionality has a significant influence on the marginal behaviour of $H_S$ and $T_2$. Eg, in the northern North Sea, extreme sea states rarely occur for directions that are associated with the land shadows of Norway, the British Isles or mainland Europe. At our location, the land shadows of Norway and the open Atlantic Ocean are the main directional covariates for explaining the directional variability in the data. For our application, we do not aim to model this directional variability. We split up the data into two directional sectors corresponding to the two main directional covariates. For our applications, we only consider storm peak values with associated wave directions corresponding to the Atlantic Ocean as they generally correspond to more extreme sea states.

## B.3.2   Heffernan-Tawn mixture model fits

To show that the MCMC chains converged and mix well, we show trace plots of the MCMC chains for the Heffernan-Tawn mixture models fitted to the application, see Figures B.3.1-B.3.2. We show these in response to one of the referees.

# B.4   Calculations for example from Section 4.

Recall the following set-up for $\lambda > 1$, $0 < t < 1$.

$$
B \sim \text{Bernoulli}(p), \quad X \sim \begin{cases} \text{Exp}(1) & \text{if } B = 0, \\ \text{Exp}(\lambda) & \text{if } B = 1, \end{cases} \quad Y \sim \begin{cases} tX & \text{if } B = 0, \\ \lambda X & \text{if } B = 1. \end{cases} \tag{B.4.1}
$$
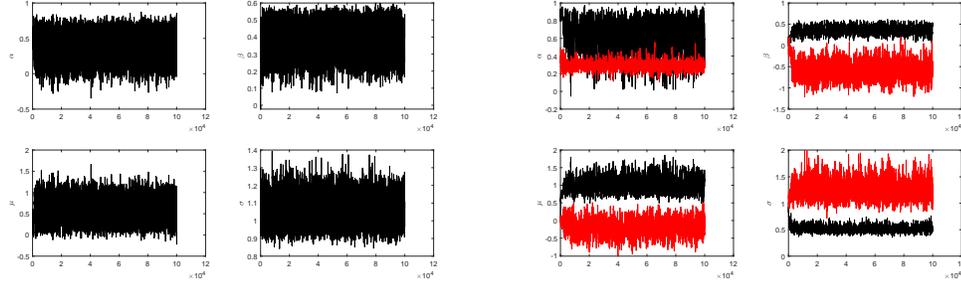
Figure B.3.1: Trace plots of the HT(1) model fit (left) and the HT(2) model fit (right) to data from the application in the main paper. In the figure on the right, the black trace plot corresponds to the sea waves component of the HT(2) model, and the red trace plot corresponds to the swell waves component of the HT(2) model.
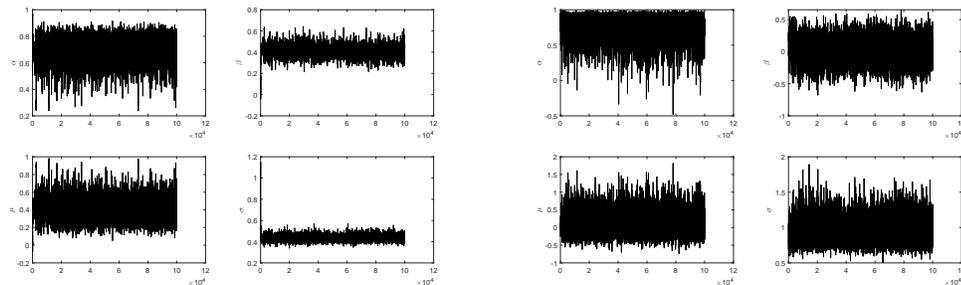


Figure B.3.2: Trace plots of the Part-HT(1) model fit to data from the application in the main paper. (left) sea waves component, (right) swell waves component.

We transform $X$ and $Y$ to Laplace scale using the probability integral transform, i.e., $X_L = F_L^{-1}(F_X(X))$ (and similarly for $Y$), where $F_L$ is the cumulative distribution function of a standard Laplace distribution, and $F_X$ is the distribution function of $X$. We investigate the conditional distribution of $Y_L|X_L = x_L$ for large values of $x_L$.

We first calculate the distribution functions of $B \mid X = x$, $X$ and $Y$ evaluated at large values. We get as $x \to \infty$

$$\mathbb{P}(B = 1|X = x) = \frac{\mathbb{P}(X = x \mid B = 1)\mathbb{P}(B = 1)}{\mathbb{P}(X = x \mid B = 1)\mathbb{P}(B = 1) + \mathbb{P}(X = x \mid B = 0)\mathbb{P}(B = 0)}$$

$$= \frac{p\lambda e^{-\lambda x}}{p\lambda e^{-\lambda x} + (1 - p)e^{-x}} \sim \frac{p\lambda}{1 - p} \cdot e^{-(\lambda - 1)x},$$

$\mathbb{P}(X > x) \sim (1 - p)e^{-x}$ as $x \to \infty$ and

$$\mathbb{P}(Y > y) = \mathbb{P}(B = 0)\mathbb{P}(tX > y|B = 0) + \mathbb{P}(B = 1)\mathbb{P}(\lambda X > y|B = 1)$$

$$= (1 - p)e^{-y/t} + pe^{-y} \sim pe^{-y}$$

as $y \to \infty$. Hence, as $X \to \infty$ we get that

$$X_L = F_L^{-1}(F_X(X)) = -\text{sign}(F_X(X) - 1/2)\log(1 - 2|F_X(X) - 1/2|)$$

$$\sim -\log(2 - 2F_X(X)) \sim -\log(2(1 - p)e^{-X}) = X - \log(2(1 - p)).$$

Similarly, $Y_L \sim Y - \log(2p)$ as $Y \to \infty$. Moreover,

$$\mathbb{P}(B = 1 \mid X_L = x_L) \sim \frac{p\lambda}{1 - p} \cdot e^{-(\lambda-1)(x_L + \log(2(1-p)))} = \frac{p\lambda}{2^{\lambda-1}(1 - p)^\lambda} \cdot e^{-(\lambda-1)x_L}.$$

Finally, we get that for large $X_L$

$$Y_L \mid X_L \sim \begin{cases} tX_L + [t\log(2(1 - p)) - \log(2p)] & \text{with prob. } 1 - \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)X_L} \\ \lambda X_L + [\lambda\log(2(1 - p)) - \log(2p)] & \text{with prob. } \frac{p\lambda}{2^{\lambda-1}(1-p)^\lambda} \cdot e^{-(\lambda-1)X_L}. \end{cases}$$

# B.5 Further details on the models

## B.5.1 The Metropolis-acceptance probability from Section 3.2.

The Metropolis-Hastings acceptance ratio $\alpha_{MH}$ given proposal $\boldsymbol{\theta}^{(t+1)} = (\gamma_1^{(t+1)}, \alpha_1^{(t+1)}, \beta_1^{(t+1)}, \mu_1^{(t+1)}, \sigma_1^{(t+1)})$ is given by

$$
\alpha_{MH}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)}) = 
\begin{cases}
\frac{L(\boldsymbol{\theta}^{(t+1)};\mathbf{x})}{L(\boldsymbol{\theta}^{(t)};\mathbf{x})} & \text{if } (\alpha_1^{(t)} \neq 1, \alpha_1^{(t+1)} \neq 1) \\[2em]
& \text{or } (\alpha_1^{(t)} = 1, \alpha_1^{(t+1)} = 1) \\[2em]
\frac{L(\boldsymbol{\theta}^{(t+1)};\mathbf{x})}{L(\boldsymbol{\theta}^{(t)};\mathbf{x})} \cdot \frac{(1-\omega)/2}{\omega} \cdot \frac{1-\Phi\left([1-\alpha_1^{(t+1)}]/h\right)}{\varphi\left([\alpha_1^{(t+1)}-1]/h\right)/h} & \text{if } \alpha_1^{(t)} = 1, \alpha_1^{(t+1)} \neq 1 \\[2em]
\frac{L(\boldsymbol{\theta}^{(t+1)};\mathbf{x})}{L(\boldsymbol{\theta}^{(t)};\mathbf{x})} \cdot \frac{\omega}{(1-\omega)/2} \cdot \frac{\varphi\left([\alpha_1^{(t)}-1]/h\right)/h}{1-\Phi\left([1-\alpha_1^{(t)}]/h\right)} & \text{if } \alpha_1^{(t)} \neq 1, \alpha_1^{(t+1)} = 1
\end{cases}
$$

with candidate acceptance probability $p = \min\{1, \alpha_{MH}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)})\}$. This follows straightforwardly from the definition of $g(\cdot|\alpha_1^{(t)})$.

## B.5.2 Model (7) when $\alpha_k = \alpha_{k'}$ for $k \neq k'$

If we allow $\alpha_k = \alpha_{k'}$ for $k \neq k'$ in model (7), then some key details change. We list two of them here concisely. (i) For model identifiability, we impose the constraint that $\alpha_k = \alpha_{k'}$ for $k < k'$ implies $\beta_k < \beta_{k'}$. Note that if $\beta_k = \beta_{k'}$, then the model is equivalent to a model with a smaller value for $K$ and the respective residual distribution is a mixture of $Z_k$ and $Z_{k'}$ with $Z_k$ arising with probability $p_k/(p_k + p_k')$. (ii) The distribution $G_k$ as defined earlier similarly puts mass $p_k$ on $\mathbb{R} \setminus \{0\}$. However, it puts a different amount of mass at $\{+\infty\}$, $\{-\infty\}$ and 0 as before. More precisely, let $I = \{i \neq k : \alpha_i = \alpha_k\} = \{i_1, \ldots, i_j\}$ and write $q_k = \mathbb{P}(\mu_k + \sigma_k \tilde{Z}_k \leq 0) = 1 - \bar{q}_k$. Then, $G_k(z)$ puts weight: $p_{k+1}\bar{q}_{k+1} + \cdots + p_{i_j}\bar{q}_{i_j} + p_{i_j+1} + \cdots + p_K$ on $\{+\infty\}$; $p_{i_1} + \cdots + p_{k-1}$ on $\{0\}$; and $p_1 + \cdots + p_{i_1-1} + p_{k+1}q_{k+1} + \cdots + p_{i_j}q_{i_j}$ on $\{-\infty\}$.

### B.5.3 Quantile-regression model where $K > 2$

The quantile-regression model extends naturally to $K > 2$ mixture components. Though, a naive extension of the method from Section 3.3.3. requires optimising over a space of size $\mathcal{O}(m^{K-1})$, which becomes infeasible for relatively large values of $K$. However, through an adapted implementation, the computational complexity can be reduced to $\mathcal{O}(m^2)$ as long as we do not impose a non-crossing contraint. This reduction is achieved by estimation for all subsequences $\boldsymbol{\omega}_{\tau_i:\tau_j} = (\boldsymbol{\omega}_{\tau_i}, \boldsymbol{\omega}_{\tau_{i+1}}, \ldots, \boldsymbol{\omega}_{\tau_j})$ where $1 \leq i \leq j \leq m$ using $K = 1$. In any of these sub-problems, we optimise over a $(4 + j - i)$ dimensional space. We save the results of these sub problems in the $m \times m$ matrix $M = (M_{i:j})_{i,j=1}^m$, where $M_{i:j}$ contains all information that is related to the fit of the one-component HTM model where the sequence of quantile levels is given by $(\tau_i, \ldots, \tau_j)$, e.g., the parameters of the fit are stored in this location, and where $M_{i:j} = \varnothing$ for $j < i$.

A model fit with $K$ mixture components now consists of a path of length $K$ starting in the top row of $M$ and ending in the rightmost column. The restriction of the path through the matrix is that you have to move from column $j$ to row $j+1$ and you are not allowed to visit elements of the matrix that are equal to the empty set. As an example, a general path might look like

$$M_{i_1+1:i_2} \to M_{i_2+1:i_3} \to M_{i_3+1:i_4} \to \cdots \to M_{i_K+1:i_{K+1}}$$

where $i_1 = 0$, $i_{K+1} = m$, and $i_{j+1} > i_j$ for all $j$. The best fit can now be found by finding the path that minimizes equation (12). To that end, it is sufficiently quick to evaluate all possible sequences.

In practice, it is a sensible assumption to only allow subsequences in the above such that $i_{j+1} \geq i_j + l$ for all $j$ and some $l > 1$, i.e., when we expect the mixture probability $p_k$ for $k = 1, 2, \ldots, K$ to be at least of size $\min\{\tau_{i+l-1} - \tau_i : i = 1, \ldots, m + 1 - l\}$. This assumption requires less computation time as more elements of the matrix structure above are considered empty. Additionally, this assumption makes the inference procedure less sensitive to boundary problems.

## B.5.4 Estimating the mixture probabilities using the quantile-regression model

Here, we discuss how to estimate $p_1$ in Section 3.3.3. The estimator $\hat{p}_1$ for $p_1$ is derived from the property that the quantile function $q_{p_1}(x)$, which is at the boundary between the two mixture components, should equally likely belong to either component of the model fit. With this in mind, we define the estimator $\hat{p}_1$ for the mixture probability $p_1$ as

$$\hat{p}_1 := \underset{p \in [\tau_{\hat{m}_0}, \tau_{\hat{m}_0+1}]}{\text{solve}} \left\{ \sum_{i=1}^{n} \rho_p(y_i - q(x_i|\hat{\gamma}_1, \hat{\alpha}_1, \hat{\beta}_1, \hat{\zeta}_p^1)) = \sum_{i=1}^{n} \rho_p(y_i - q(x_i \mid \hat{\gamma}_2, \hat{\alpha}_2, \hat{\beta}_2, \hat{\zeta}_p^2)) \right\}$$
(B.5.1)

where $\hat{\zeta}_p^k$ for $k = 1, 2$ is defined as

$$\hat{\zeta}_p^k = \text{argmin}_{\zeta \in \mathbb{R}} \left\{ \sum_{i=1}^{n} \rho_p(y_i - q(x_i|\hat{\gamma}_k, \hat{\alpha}_k, \hat{\beta}_k, \zeta)) \right\}.$$

## B.5.5 Quantile calibration

The conditional quantile function $q_\tau(x)$ of the underlying model (7) is a continuous function of $\tau$ on $(0, 1)$ when $K = 1$. The estimator $\hat{q}_\tau(x)$ for $\tau \in (0, 1)$ obtained from the grid $\{\tau_1, \ldots, \tau_m\}$ using equation (13), however, does not satisfy this property. Here, we describe modified smooth estimators for $q_\tau(x)$.

We demonstrate the discontinuous nature of $\hat{q}_\tau(x)$ in the following simplified case. Let $x_1 < x_2 < x_3$ be three observations and let $q_\tau(x) := q_\tau$. The estimator $\hat{q}_\tau(x)$ as defined above is now not continuous as a function of $\tau$ and is given by

$$\hat{q}_\tau(x) = x_1 \mathbb{1}\{\tau \in (0, 1/3)\} + x_2 \mathbb{1}\{\tau \in (1/3, 2/3)\} + x_3 \mathbb{1}\{\tau \in (2/3, 1)\}.$$

We note that for $\tau \in \{1/3, 2/3\}$, the estimator is not well-defined as the objective function in equation (11) takes on the same value for any $q_\tau$ in $[x_1, x_2], [x_2, x_3]$, respectively. Although less obvious, it can be verified using simulations that this phenomenon also holds when $q_\tau(x)$ takes on different parametric forms.

Suppose that we have used the following non-exceedance probabilties $\tau_j = 0.05j$ for $j = 1, \ldots, 19$ to fit model (13). In practice, we then estimate $\hat{q}_\tau(x)$ on a finer grid

$\tau \in \mathcal{G}$, e.g., $\mathcal{G} := \{0.001, 0.002, \ldots, 0.999\}$ using equation (13), and linearly interpolate $\hat{q}_\tau(x)$ for $\tau \notin \mathcal{G}$. We now produce smooth estimators for $q_\tau(x)$ and $\tau \in (0,1)$ by applying the following two techniques: one to the body of the residual distribution, i.e., $\tau \in [q_l, q_u]$, and the other to the tails. For $\tau \in [q_l, q_u]$, we smooth $\hat{q}_\tau(x)$ using a third order increasing spline from the SLM-shape language modelling toolbox (D'Errico, 2020). For values of $\tau$ corresponding to the upper $[q_u, 1)$ quantiles of the residual distribution, we fit a generalised Pareto distribution (Pickands, 1975) that minimises the Anderson-Darling test statistic

$$\int_{[q_u, 1)} \frac{(\tilde{F}(z) - \hat{F}_Z(z))^2}{\hat{F}_Z(z)(1 - \hat{F}_Z(z))} \, d\hat{F}_Z(z),$$

where $\tilde{F}(z) = 1 - (1 + \xi(z - \mu_Z)/\sigma_Z)_+^{-1/\xi}$ for $z > \mu_Z$ is the distribution function of a generalised Pareto and $\hat{F}_Z(z) = \hat{\zeta}_{\hat{F}_Z(z)}$ is estimated using quantile-regression. A similar approach is applied to the lower tail. We use this extrapolation technique to both reduce variability in estimates for $\tau$ near 0 and 1 and allow for extrapolations beyond the data that do not contradict standard univariate extreme value assumptions.

Parameter estimation via minimisation of the Anderson-Darling test statistic is appropriate since likelihood inference would require a data sample which, in this case, is not available. The commonly used Kolmogorov-Smirnov and Cramer-Rao test statistics were felt to be unsuitable since both put an emphasis on the body of the distribution at the expense of the tail whereas the Anderson-Darling test statistic does not have that property.

## B.6 Simulation Study

### B.6.1 Set-up

We test the model performance of the Heffernan-Tawn mixture and quantile-regression models on data generated from six different distributions, by comparing estimated probabilities of extreme sets. We consider the following bivariate distributions on Laplace margins, see Table B.6.1: (A) the bivariate extreme value distribution with a logistic dependence structure, i.e., a special case of the distribution from equation (5);

(B-D) the bivariate extreme value distribution with an asymmetric logistic dependence structure, i.e., the distribution from equation (5); (E) a bivariate Gaussian copula; (F) a model which is a maxima-mixture of distributions (A) and (C) as in Simpson et al. (2020). Distributions (A) and (E) are described well using the vanilla HT model since the residual distributions do not put weight on $\{\pm\infty\}$ and there are no mixtures present. Distributions (B-D) and (F) on the other hand fall into the HTM model framework with $K = 2$ components.

For the simulations, we draw 5000 observations from the distributions, and use the 90% empirical quantile as the HT modelling threshold $u$, giving 500 observations with which to make inference for the conditional extremal models. We fit the two models for a number of $K \geq 1$ and replicate this procedure 500 times. For the method based on quantile-regression, we choose $\tau_i = 0.05i$ for $i = 1, \ldots, 19$ which was was found to give stable results without requiring long computational times.

We set $\gamma_k = 0$ for all $k = 1, \ldots, K$ in the Heffernan-Tawn mixture model. Theoretically, performance should be improved upon letting $\gamma_k \neq 0$. However, the correlation between the HT parameters in the MCMC chain is increased significantly by adding the parameter $\gamma_k$. In particular, the HT parameters $\alpha_k$, $\mu_k$ and $\gamma_k$ become near to unidentifiable. This problem can be solved by reparameterising the regression to for example $\gamma_k + \alpha_k(X - \bar{X}) + \mu_k(X^\beta - (\bar{X})^\beta) + \sigma_k X^\beta \tilde{Z}_i$. In the quantile-regression model, we impose that the median of the residual distribution is 0. We could in a similar manner impose here that $\mu_k = 0$ to avoid a similar problem. There is hardly any practical difference in assuming $\mu_k = 0$ compared to $\gamma_k = 0$.

We test model performance by studying the sets in the simulation study of Heffernan and Tawn (2004). The difference is that they estimate the sets given the probability measure and we estimate the probaility measure given the sets. More precisely, we test the performance by comparing estimates of $\mathbb{P}((X, Y) \in S_i^{\mathcal{D}})$ for $i = 1, 2$ where $S_i^{\mathcal{D}}$ is defined as follows

$$S_1^{\mathcal{D}}(p) := (v_{\mathcal{D}}(p), \infty) \times (v_{\mathcal{D}}(p), \infty) \quad \text{and} \quad S_2^{\mathcal{D}}(p, q) := (r_{\mathcal{D}}(p, q), \infty) \times (-\infty, w_{\mathcal{D}}(p, q)),$$

| I.D. | Name | Model parameters | Heffernan-Tawn parameters |
|------|------|------------------|---------------------------|
| (A) | Logistic | $\eta = 0.5,\ \theta = (0,0)$ | $K = 1,\ \alpha = 1,\ \beta = 0$ |
| (B) | Asymmetric Logistic | $\eta = 0.5,\ \theta = (0.5, 0.5)$ | $K = 2,\ \alpha = (0,1),\ \beta = (0,0)$ |
| (C) | Asymmetric Logistic | $\eta = 0.5,\ \theta = (0.25, 0.5)$ | $K = 2,\ \alpha = (0,1),\ \beta = (0,0)$ |
| (D) | Asymmetric Logistic | $\eta = 0.5,\ \theta = (0.75, 0.5)$ | $K = 2,\ \alpha = (0,1),\ \beta = (0,0)$ |
| (E) | Bivariate Gaussian | $\rho = 0.5$ | $K = 1,\ \alpha = 0.25,\ \beta = 0.5$ |
| (F) | Mixture of E and A $(p = 0.5)$ | $\rho = 0.5,\ \xi = 0.5$ | $K = 2,\ \alpha = (0.25, 1),\ \beta = (0.5, 0)$ |

Table B.6.1: List of distributions used in the simulation study. The marginals of all distributions are standard Laplace.

where $v_{\mathcal{D}}(p) \in \mathbb{R}$ is defined such that

$$\mathbb{P}(X > v_{\mathcal{D}}(p),\ Y > v_{\mathcal{D}}(p)) = p \tag{B.6.1}$$

and $r_{\mathcal{D}}(p, q), w_{\mathcal{D}}(p, q) \in \mathbb{R}$ are defined such that

$$\mathbb{P}(X > r_{\mathcal{D}}(p, q)) = \frac{p}{q} \quad \text{and} \quad \mathbb{P}(X > r_{\mathcal{D}}(p, q),\ Y < w_{\mathcal{D}}(r, q)) = p. \tag{B.6.2}$$

Here $\mathcal{D}$ denotes the distribution of $(X, Y)$. The set of rectangles under investigation for distribution $\mathcal{D} \in \{(A), \ldots, (F)\}$ is given by

$$\mathcal{S}^{\mathcal{D}} = \left\{ S \subseteq \mathbb{R}^2 :\ S = S_1^{\mathcal{D}}(p) \text{ for } p \in \mathcal{P} \text{ or } S = S_2^{\mathcal{D}}(p, q) \text{ for } (p, q) \in \mathcal{P} \times \mathcal{Q} \right\},$$

where $\mathcal{P} = \{10^{-4}, 10^{-6}, 10^{-8}\}$ and $\mathcal{Q} = \{0.2, 0.5, 0.8\}$. This gives 12 rectangles per distribution. The HTM model returns a posterior distribution and the quantile-regression model gives a point estimate. To make the two methods comparable, we use only the posterior medians of the probability measures from the HTM model.

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 write upper case $K$ for the choice of the number of components $K$ in the inference procedures, and we introduce $K_0$ as the true number of components.

## B.6.2 Results and conclusions

We present a summary of the raw results in Table B.6.2, and we provide the entire list of raw results in the Section B.7. To explain the results in the table, we introduce the following notation. Denote $r = 500$ as the number of replications and let for $1 \leq i \leq r$, $\hat{p}_{S,\mathcal{D},i}$ denote the $i$th estimate of the probability measure of $S$ given the data that are generated using distribution $\mathcal{D}$. Moreover, denote $\hat{F}_{p_{S,\mathcal{D}}}$ as the empirical distribution function of the sample $\{\hat{p}_{S,\mathcal{D},i} : 1 \leq i \leq r\}$. Now, we define the two summary performance measures $N^{\mathcal{D}}$ and $R^{\mathcal{D}}$, which are reported in the table. The first

$$N^{\mathcal{D}} := \#\{S \in \mathcal{S}^{\mathcal{D}} : \hat{F}_{p_{S,\mathcal{D}}}^{-1}(0.025) \leq \mathbb{P}(S) \leq \hat{F}_{p_{S,\mathcal{D}}}^{-1}(0.975)\} \qquad (\text{B.6.3})$$

is an integer between 0 and 12 and represents how often the the true value lies between the 2.5% and 97.5% percentiles of the estimates. The second number

$$R^{\mathcal{D}} := \frac{1}{12} \sum_{S \in \mathcal{S}^{\mathcal{D}}} \sqrt{\sum_{i=1}^{r} \left( \frac{\hat{p}_{S,\mathcal{D},i} - \mathbb{P}(\mathbf{X} \in S | \mathbf{X} \sim \mathcal{D}))}{\mathbb{P}(\mathbf{X} \in S \mid \mathbf{X} \sim \mathcal{D})} \right)^2} \qquad (\text{B.6.4})$$

is the average relative root mean square of the methods, averaged over $\mathcal{S}^{\mathcal{D}}$. As an illustrative example, we also provide boxplots of the ensemble of estimates for distribution (F) and rectangle $R_2^{(\text{F})}(10^{-8}, 0.5)$, see Figure B.6.1. Because there is no reason to assume normality of the ensemble, the upper and lower whiskers are given by the 2.5% and 97.5% empirical ensemble quantiles.

From the tables, we note that the models with $K \geq 2$ outperform the model with $K = 1$ significantly, where $K_0 \geq 2$. The HT($K$) and QR($K$) models with $K \geq 2$ perform comparably across method and $K$, but HT($K$) has slightly poorer performance overall. In particular, for increasing $K \geq 2$ there is a slight decrease in level of performance of the HT($K$) model. This is not the case for the QR($K$) models. The main difference between QR($K$) and the HT($K$) is the width of the confidence intervals. The confidence intervals for QR($K$) tend to be slightly larger than for HT($K$) such that often the true value is contained within this interval, which is not always the case for HT($K$). We conclude that the quantile-regression model is performing well in all scenarios so long as the number of components is not

| Distribution | Method | $K$ | | |
|:---:|:---:|:---:|:---:|:---:|
| | | 1 | 2 | 3 |
| (A) | QR | (12; 0.41) | (12; 0.36) | (12; 0.36) |
| | HT | (1; 0.67) | (0; 0.65) | (0; 0.68) |
| (B) | QR | (3; 0.52) | (12; 0.35) | (12; 0.33) |
| | HT | (3; 0.54) | (9; 0.35) | (6; 0.36) |
| (C) | QR | (0; 0.64) | (9; 0.43) | (12; 0.39) |
| | HT | (0; 0.71) | (3; 0.44) | (3; 0.44) |
| (D) | QR | (8; 0.56) | (12; 0.43) | (12; 0.40) |
| | HT | (4; 0.55) | (6; 0.51) | (8; 0.37) |
| (E) | QR | (12; 0.46) | (12; 0.71) | (12; 0.80) |
| | HT | (12; 0.44) | (12; 0.49) | (12; 0.53) |
| (F) | QR | (3; 0.58) | (12; 0.35) | (12; 0.31) |
| | HT | (1; 0.63) | (6; 0.38) | (6; 0.35) |

Table B.6.2: Summary performance of the simulation studies. Each element in the table corresponds to a method, HT or QR, with $k$ components applied to distribution $\mathcal{D} \in \{(A), \ldots, (F)\}$. The value in each cell is given by $(N^{\mathcal{D}}; \ R^{\mathcal{D}})$, see equations (B.6.3) and (B.6.4).
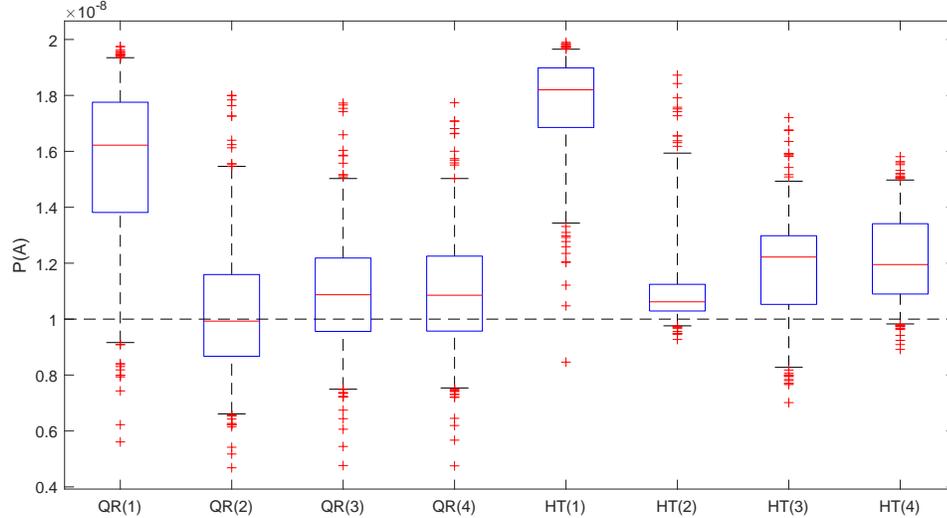
Figure B.6.1: Boxplots of the estimates for the probability corresponding to the rectangle $S_2^{(\mathrm{F})}(10^{-8}, 0.5) = (17.0, \infty) \times (-\infty, 13.2)$ of distribution (F). $\mathrm{QR}(K)$ and $\mathrm{HT}(K)$ are abbreviations for the estimates generated by the quantile-regression model and the Heffernan-Tawn mixture model with $K$ components.

underestimated.

In the case of distributions (A), (B), (C), (D) and (F), the true HTM model is an asymptotically dependent model, i.e., $\alpha_{K_0} = 1$ and the parameters lie thus on the boundary of the parameter space, see Table B.6.1. It appears that for these types of distributions the Heffernan-Tawn mixture model consistently underestimates the probability measures of sets of the form $R_1(p)$ because a uniform prior over the parameter space puts 0 weight on the asymptotically dependent case. The quantile-regression model, however, does have a chance of classifying asymptotically dependent models but since it theoretically cannot over-estimate, the model also consistently underestimates albeit to a smaller extent.

If there is a reason to believe that the asymptotic dependent model is the true model, then we can incorporate this information into the prior, as in Section 3.2. We have tested the adjusted version of the method where we allow for positive prior mass on asymptotic dependence on distributions (A) and (E), capturing a model exhibiting asymptotic dependence and asymptotic independence, respectively. Initial

simulations indicate that this method improves significantly on the original method for Distribution (A), i.e., where $\alpha_1$ is equal to one. Moreover, the new method performs only slightly worse for Distribution (E), i.e., where $\alpha_1$ is not equal to one.

In more detail, let $K = 1$ and set $\omega_1 = 0$ and $\omega_2 = 0.5$, i.e., we set the prior probability $\mathbb{P}(\alpha_1 = 1)$ equal to 0 and 0.5, respectively. In our simulations, we simulate and compute the posterior of $\alpha_1$. We repeat this 1000 times to get a good estimate for average performance. From the results, we found that the average MSE of $\hat{\alpha}_1$ using $\omega_1$ for distribution (A) is estimated to be $0.008(\pm 0.0003)$ and with $\omega_2$ it is estimated to be $0.002(\pm 0.0002)$. For distribution (E), the average MSE using $\omega_1$ is estimated to be around $0.04(\pm 0.0009)$, whereas the average MSE under the model with $\omega_2$ is equal to $0.07(\pm 0.004)$. Hence, in this case we improve our results with a factor of 5 when the data exhibits asymptotic dependence and the results worsen with a factor of 2 when the data does not exhibit asymptotic dependence.

These results show that the choice of prior can positively influence estimates. We assert that using a uniform prior over the parameter space is a good choice as long as asymptotic dependence is not appropriate. However, if one is uncertain about the asymptotic nature of the data, then we suggest the use of a prior which puts mass on the asymptotic dependence model as well as a uniformly distributed mass on the parameter space corresponding to asymptotic independence. The question arises of how to choose an optimal balance here, i.e., how do we choose the prior probability $\omega$ of asymptotic dependence. Further research is needed to provide a sensible answer but we believe that in practice, an optimal value can be found by applying the methods for different $\omega$'s.

## B.7 Raw simulation study results

We present the raw results of each distribution and give a concise summary of the results. The tables used within this section all use the same notation. In the tables, we give estimates of the measures of $S_1^{\mathcal{D}}(p)$, expressed as a fraction of $p$, for different values of $p$ and distributions $\mathcal{D} \in \{(A), (B), (C), (D), (E), (F)\}$. Here $\text{QR}(K)$ and $\text{HT}(K)$ are abbreviations for the estimates generated by the quantile-regression model and the Heffernan-Tawn mixture model with $K$ components. Boldface indicates whether or not the true value, which is always 1, is captured within the error bounds. Boldface with an asterisk is used to indicate the best-performing model for the corresponding $p$, i.e., for which the root mean squared error is minimal.

### B.7.1 Distribution (A)

For this distribution, the 1 component model is the true limit model. However, it turns out that the 2 component quantile-regression model is dominant in the results although it is similar to the 1 component quantile-regression model. We note that the $\text{HT}(K)$ models for $K \geq 1$ consistently underestimate the probabilities in the first table and overestimate in the second table. This bias is there because the true HT parameters lie on the boundary of the parameter space.

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (A) | QR(1) | **0.85** | **(0.59, 1.09)** | **0.72** | **(0.35, 1.08)** | **0.59** | **(0.21, 1.08)** |
| | QR(2) | **0.89*** | **(0.63, 1.08)*** | **0.80*** | **(0.41, 1.07)*** | **0.72*** | **(0.24, 1.06)*** |
| | QR(3) | **0.89** | **(0.62, 1.10)** | **0.81** | **(0.38, 1.10)** | **0.73** | **(0.23, 1.09)** |
| | HT(1) | 0.68 | (0.42, 0.92) | 0.46 | (0.18, 0.81) | 0.29 | (0.07, 0.71) |
| | HT(2) | 0.66 | (0.50, 0.85) | 0.49 | (0.30, 0.71) | 0.37 | (0.17, 0.59) |
| | HT(3) | 0.64 | (0.49, 0.80) | 0.46 | (0.29, 0.65) | 0.34 | (0.18, 0.52) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (A) | QR(1) | 0.2 | **1.18** | **(0.70, 1.89)** | **1.55** | **(0.70, 2.83)** | **1.95** | **(0.70, 3.59)** |
| | QR(1) | 0.5 | **1.16** | **(0.91, 1.44)** | **1.31** | **(0.92, 1.67)** | **1.44** | **(0.92, 1.81)** |
| | QR(1) | 0.8 | **1.06** | **(0.98, 1.14)** | **1.10** | **(0.98, 1.19)** | **1.13** | **(0.98, 1.22)** |
| | QR(2) | 0.2 | **1.26*** | **(0.75, 1.86)*** | **1.57** | **(0.79, 2.63)** | **1.81*** | **(0.83, 3.31)*** |
| | QR(2) | 0.5 | **1.12*** | **(0.92, 1.40)*** | **1.21*** | **(0.93, 1.61)*** | **1.30*** | **(0.93, 1.78)*** |
| | QR(2) | 0.8 | **1.04*** | **(0.98, 1.13)*** | **1.07*** | **(0.98, 1.18)*** | **1.09*** | **(0.98, 1.21)*** |
| | QR(3) | 0.2 | **1.27** | **(0.74, 1.85)** | **1.59*** | **(0.78, 2.59)*** | **1.84** | **(0.80, 3.42)** |
| | QR(3) | 0.5 | **1.11** | **(0.88, 1.42)** | **1.20** | **(0.89, 1.65)** | **1.29** | **(0.90, 1.79)** |
| | QR(3) | 0.8 | **1.04** | **(0.97, 1.14)** | **1.06** | **(0.97, 1.20)** | **1.08** | **(0.97, 1.22)** |
| | HT(1) | 0.2 | **1.52** | **(0.96, 2.38)** | 2.31 | (1.26, 3.64) | 3.05 | (1.53, 4.40) |
| | HT(1) | 0.5 | 1.35 | (1.10, 1.60) | 1.57 | (1.22, 1.83) | 1.73 | (1.33, 1.94) |
| | HT(1) | 0.8 | 1.12 | (1.04, 1.18) | 1.17 | (1.08, 1.22) | 1.20 | (1.11, 1.24) |
| | HT(2) | 0.2 | 1.68 | (1.12, 2.33) | 2.43 | (1.60, 3.13) | 2.91 | (2.07, 3.80) |
| | HT(2) | 0.5 | 1.36 | (1.17, 1.53) | 1.54 | (1.32, 1.73) | 1.66 | (1.45, 1.85) |
| | HT(2) | 0.8 | 1.12 | (1.07, 1.17) | 1.16 | (1.11, 1.21) | 1.19 | (1.14, 1.22) |
| | HT(3) | 0.2 | 1.80 | (1.30, 2.27) | 2.52 | (1.86, 3.15) | 3.08 | (2.30, 3.76) |
| | HT(3) | 0.5 | 1.39 | (1.22, 1.54) | 1.57 | (1.38, 1.73) | 1.69 | (1.51, 1.84) |
| | HT(3) | 0.8 | 1.13 | (1.08, 1.17) | 1.17 | (1.12, 1.20) | 1.19 | (1.15, 1.22) |

## B.7.2 Distribution (B)

For this distribution, the 2 component model is the true limit model. We note that the 2 and 3 component models perform similarly. Peculiarly, the 1 component models do a good job in estimating $S_2^{\mathcal{D}}(p, 0.5)$. This is the case because the mixture probability is exactly 0.5. Hence, for $q < 0.5$, $S_2^{\mathcal{D}}(p, q)$ is over-estimated, and for $q > 0.5$, it is under-estimated using these models. In general, this clearly shows that the multi-component models are the clear winners here. However, both methods do struggle with the extreme sets as the true HT modelling parameters lie on the boundary of the parameter space.

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 0.26 | (0.09, 0.56) | 0.07 | (0.01, 0.28) | 0.02 | (0.00, 0.13) |
| | QR(2) | **0.73** | **(0.32, 1.19)** | **0.46** | **(0.10, 1.19)** | **0.27** | **(0.03, 1.17)** |
| | QR(3) | **0.79** | **(0.36, 1.18)** | **0.57** | **(0.12, 1.15)** | **0.40** | **(0.04, 1.15)** |
| | HT(1) | 0.18 | (0.06, 0.39) | 0.03 | (0.01, 0.14) | 0.01 | (0.00, 0.04) |
| | HT(2) | 0.76 | (0.38, 0.94) | **0.58*** | **(0.15, 0.82)*** | 0.43 | (0.05, 0.71) |
| | HT(3) | 0.64 | (0.35, 0.82) | 0.44 | (0.12, 0.66) | 0.29 | (0.04, 0.54) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 0.2 | 0.39 | (0.12, 0.57) | 0.23 | (0.02, 0.52) | 0.13 | (0.00, 0.47) |
| | QR(1) | 0.5 | **1.03** | **(0.65, 1.25)** | **1.09** | **(0.54, 1.43)** | **1.13** | **(0.45, 1.58)** |
| | QR(1) | 0.8 | 1.19 | (1.09, 1.23) | 1.23 | (1.16, 1.25) | 1.24 | (1.19, 1.25) |
| | QR(2) | 0.2 | **0.48** | **(0.27, 1.47)** | **0.43** | **(0.14, 1.68)** | **0.41** | **(0.06, 1.83)** |
| | QR(2) | 0.5 | **0.93** | **(0.68, 1.21)** | **0.96** | **(0.71, 1.23)** | **0.97** | **(0.73, 1.25)** |
| | QR(2) | 0.8 | **1.03*** | **(0.88, 1.19)*** | **1.10** | **(0.93, 1.23)** | **1.15** | **(0.94, 1.24)** |
| | QR(3) | 0.2 | **0.59** | **(0.26, 1.54)** | **0.49** | **(0.03, 1.82)** | **0.45** | **(0.00, 2.02)** |
| | QR(3) | 0.5 | **0.93** | **(0.65, 1.21)** | **0.95** | **(0.66, 1.25)** | **0.97** | **(0.67, 1.30)** |
| | QR(3) | 0.8 | **1.04** | **(0.87, 1.19)** | **1.09** | **(0.90, 1.23)** | **1.14** | **(0.92, 1.24)** |
| | HT(1) | 0.2 | 0.48 | (0.18, 0.98) | 0.27 | (0.04, 0.86) | 0.16 | (0.01, 0.78) |
| | HT(1) | 0.5 | **1.13** | **(0.80, 1.39)** | **1.23** | **(0.73, 1.62)** | **1.31** | **(0.64, 1.77)** |
| | HT(1) | 0.8 | 1.22 | (1.18, 1.24) | 1.24 | (1.23, 1.25) | 1.25 | (1.24, 1.25) |
| | HT(2) | 0.2 | **1.22** | **(0.39, 2.09)** | **1.47** | **(0.15, 2.45)** | **1.68** | **(0.05, 2.55)** |
| | HT(2) | 0.5 | **1.01*** | **(0.93, 1.09)*** | **1.01*** | **(0.95, 1.07)*** | **1.01*** | **(0.95, 1.08)*** |
| | HT(2) | 0.8 | 1.09 | (1.03, 1.19) | 1.13 | (1.07, 1.23) | 1.17 | (1.09, 1.24) |
| | HT(3) | 0.2 | **1.13** | **(0.41, 1.82)** | **1.29** | **(0.25, 2.08)** | **1.44** | **(0.17, 2.11)** |
| | HT(3) | 0.5 | **1.14** | **(0.85, 1.31)** | **1.19** | **(0.78, 1.33)** | **1.22** | **(0.73, 1.33)** |
| | HT(3) | 0.8 | 1.12 | (1.07, 1.20) | 1.17 | (1.11, 1.23) | 1.20 | (1.14, 1.24) |

## B.7.3   Distribution (C)

For this distribution, the 2 component model is the true model. We see that the quantile-regression model with either 2 or 3 components dominate the results. The $HT(K)$ models again underperform because the true parameters lie on the boundary. The same holds true for the quantile-regression model albeit less apparent. The only sets where the Heffernan-Tawn mixture model has good results is for sets of the form $R_2(p, 0.2)$, which is not much affected by the parameters that are on the boundary.

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (C) | QR(1) | 0.48 | (0.20, 0.87) | 0.21 | (0.05, 0.63) | 0.09 | (0.02, 0.46) |
| | QR(2) | **0.71**$^*$ | **(0.36, 1.10)**$^*$ | **0.46**$^*$ | **(0.13, 1.04)**$^*$ | 0.30 | (0.05, 0.98) |
| | QR(3) | **0.77** | **(0.37, 1.12)** | **0.57** | **(0.13, 1.07)** | **0.41**$^*$ | **(0.04, 1.07)**$^*$ |
| | HT(1) | 0.23 | (0.09, 0.50) | 0.06 | (0.01, 0.25) | 0.02 | (0.01, 0.12) |
| | HT(2) | 0.61 | (0.41, 0.74) | 0.46 | (0.21, 0.62) | 0.36 | (0.11, 0.54) |
| | HT(3) | 0.58 | (0.36, 0.77) | 0.39 | (0.17, 0.62) | 0.27 | (0.08, 0.49) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (C) | QR(1) | 0.2 | 0.14 | (0.03, 0.45) | 0.02 | (0.00, 0.17) | 0.00 | (0.00, 0.08) |
| | QR(1) | 0.5 | 1.47 | (1.09, 1.77) | 1.75 | (1.30, 1.94) | 1.89 | (1.48, 1.98) |
| | QR(1) | 0.8 | 1.16 | (1.05, 1.22) | 1.21 | (1.11, 1.24) | 1.23 | (1.15, 1.25) |
| | QR(2) | 0.2 | **0.59**\* | **(0.22, 1.21)**\* | **0.35**\* | **(0.04, 1.13)**\* | 0.25 | (0.01, 1.00) |
| | QR(2) | 0.5 | **1.19** | **(0.89, 1.53)** | **1.44** | **(0.93, 1.81)** | **1.63**\* | **(0.98, 1.91)**\* |
| | QR(2) | 0.8 | **1.07**\* | **(0.96, 1.17)**\* | **1.13**\* | **(0.99, 1.22)**\* | 1.17 | (1.01, 1.24) |
| | QR(3) | 0.2 | **0.66** | **(0.22, 1.39)** | **0.43** | **(0.02, 1.32)** | **0.30**\* | **(0.00, 1.31)**\* |
| | QR(3) | 0.5 | **1.16**\* | **(0.88, 1.52)**\* | **1.34**\* | **(0.93, 1.80)**\* | **1.49** | **(0.96, 1.93)** |
| | QR(3) | 0.8 | **1.05** | **(0.95, 1.18)** | **1.11** | **(0.96, 1.22)** | **1.15**\* | **(0.97, 1.24)**\* |
| | HT(1) | 0.2 | 0.27 | (0.07, 0.76) | 0.07 | (0.00, 0.45) | 0.02 | (0.00, 0.29) |
| | HT(1) | 0.5 | 1.73 | (1.43, 1.90) | 1.93 | (1.71, 1.98) | 1.98 | (1.86, 2.00) |
| | HT(1) | 0.8 | 1.21 | (1.15, 1.24) | 1.24 | (1.20, 1.25) | 1.25 | (1.23, 1.25) |
| | HT(2) | 0.2 | **0.78** | **(0.18, 1.70)** | **0.50** | **(0.02, 1.93)** | **0.33** | **(0.00, 2.07)** |
| | HT(2) | 0.5 | 1.33 | (1.22, 1.51) | 1.45 | (1.32, 1.71) | 1.56 | (1.39, 1.84) |
| | HT(2) | 0.8 | 1.12 | (1.08, 1.18) | 1.16 | (1.11, 1.21) | 1.18 | (1.13, 1.23) |
| | HT(3) | 0.2 | **1.00** | **(0.31, 1.58)** | **0.93** | **(0.09, 1.62)** | **0.90** | **(0.02, 1.64)** |
| | HT(3) | 0.5 | 1.36 | (1.17, 1.55) | 1.53 | (1.32, 1.77) | 1.65 | (1.45, 1.89) |
| | HT(3) | 0.8 | 1.13 | (1.08, 1.19) | 1.17 | (1.12, 1.22) | 1.20 | (1.15, 1.24) |

## B.7.4   Distribution (D)

For this distribution we see the same performance as for distribution (C). The difference that the HTM model also performs well now for sets of the form $R_2(p, 0.5)$. This is because of the different parameter choice in the distribution such that the estimates for these sets are also less affected by the boundary problem.

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (D) | QR(1) | 0.28 | (0.04, 0.57) | 0.05 | (0.00, 0.21) | 0.01 | (0.00, 0.07) |
| | QR(2) | **0.84** | **(0.28, 1.40)** | **0.54** | **(0.05, 1.38)** | **0.30** | **(0.01, 1.38)** |
| | QR(3) | **0.87** | **(0.33, 1.37)** | **0.65**\* | **(0.07, 1.35)**\* | **0.44**\* | **(0.02, 1.35)**\* |
| | HT(1) | 0.23 | (0.06, 0.48) | 0.03 | (0.00, 0.14) | 0.01 | (0.00, 0.04) |
| | HT(2) | 0.39 | (0.11, 0.98) | 0.09 | (0.01, 0.70) | 0.02 | (0.00, 0.45) |
| | HT(3) | **0.82**\* | **(0.20, 1.15)**\* | 0.51 | (0.04, 0.98) | 0.28 | (0.00, 0.82) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| (D) | QR(1) | 0.2 | **0.38** | **(0.08, 1.67)** | **0.42** | **(0.24, 1.95)** | **0.46** | **(0.26, 2.15)** |
| | QR(1) | 0.5 | **0.45** | **(0.34, 1.10)** | **0.42** | **(0.22, 1.14)** | **0.39** | **(0.16, 1.17)** |
| | QR(1) | 0.8 | **1.06** | **(0.90, 1.23)** | **1.14**[*] | **(0.97, 1.24)**[*] | 1.17 | (1.01, 1.25) |
| | QR(2) | 0.2 | **0.73** | **(0.11, 1.71)** | **0.65** | **(0.12, 2.30)** | **0.60** | **(0.09, 2.73)** |
| | QR(2) | 0.5 | **0.74** | **(0.40, 1.23)** | **0.64** | **(0.25, 1.37)** | **0.58** | **(0.11, 1.44)** |
| | QR(2) | 0.8 | **1.01** | **(0.72, 1.20)** | **1.09** | **(0.77, 1.24)** | **1.15** | **(0.81, 1.25)** |
| | QR(3) | 0.2 | **0.82** | **(0.23, 1.67)** | **0.72** | **(0.03, 2.35)** | **0.65** | **(0.00, 2.92)** |
| | QR(3) | 0.5 | **0.87** | **(0.44, 1.29)** | **0.78** | **(0.17, 1.37)** | **0.71** | **(0.04, 1.45)** |
| | QR(3) | 0.8 | **1.01** | **(0.78, 1.17)** | **1.06** | **(0.83, 1.23)** | **1.11**[*] | **(0.85, 1.25)**[*] |
| | HT(1) | 0.2 | **0.78**[*] | **(0.39, 1.44)**[*] | **0.60**[*] | **(0.20, 1.55)**[*] | **0.48**[*] | **(0.10, 1.70)**[*] |
| | HT(1) | 0.5 | 0.56 | (0.28, 0.99) | 0.39 | (0.12, 1.00) | **0.29** | **(0.07, 1.02)** |
| | HT(1) | 0.8 | 1.22 | (1.16, 1.25) | 1.24 | (1.23, 1.25) | 1.25 | (1.24, 1.25) |
| | HT(2) | 0.2 | **1.14** | **(0.32, 2.21)** | **1.62** | **(0.20, 2.62)** | **2.07** | **(0.12, 2.66)** |
| | HT(2) | 0.5 | **1.00**[*] | **(0.65, 1.22)**[*] | **1.00**[*] | **(0.44, 1.20)**[*] | **1.00**[*] | **(0.32, 1.20)**[*] |
| | HT(2) | 0.8 | **1.18** | **(1.02, 1.24)** | 1.23 | (1.09, 1.25) | 1.25 | (1.14, 1.25) |
| | HT(3) | 0.2 | **1.10** | **(0.46, 1.75)** | **1.34** | **(0.32, 2.18)** | **1.53** | **(0.25, 2.35)** |
| | HT(3) | 0.5 | **0.90** | **(0.58, 1.24)** | **0.79** | **(0.49, 1.27)** | **0.73** | **(0.41, 1.30)** |
| | HT(3) | 0.8 | **1.06**[*] | **(0.97, 1.22)**[*] | 1.13 | (1.01, 1.24) | 1.18 | (1.04, 1.25) |

## B.7.5  Distribution (E)

For this distribution, the 1 component model is the truth and the parameters lie in the interior of the parameter space. However, we know that the convergence of this model to the Heffernan-Tawn model is relatively slow. However, this doesn't seem to be a problem within the results. The 1 component models dominate the results and all models have an acceptable performance.

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 1.01 | (0.33, 1.68) | 1.03 | (0.07, 2.79) | 1.10 | (0.00, 4.73) |
| | QR(2) | 1.05 | (0.08, 2.22) | 1.07 | (0.00, 5.10) | 1.16 | (0.00, 10.78) |
| | QR(3) | 1.06 | (0.03, 2.29) | 1.07 | (0.00, 5.67) | 1.15 | (0.00, 17.99) |
| | HT(1) | 0.86* | (0.28, 1.49)* | 0.73* | (0.01, 2.16)* | 0.72* | (0.00, 3.51)* |
| | HT(2) | 0.94 | (0.26, 2.01) | 0.75 | (0.01, 3.30) | 0.59 | (0.00, 5.34) |
| | HT(3) | 1.02 | (0.27, 2.04) | 0.98 | (0.03, 4.02) | 0.82 | (0.00, 7.57) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 0.2 | 0.70* | (0.25, 1.42)* | 0.48* | (0.08, 1.58)* | 0.33* | (0.03, 1.73)* |
| | QR(1) | 0.5 | 0.95 | (0.57, 1.32) | 0.89 | (0.32, 1.45) | 0.82 | (0.15, 1.53) |
| | QR(1) | 0.8 | 1.00 | (0.84, 1.14) | 0.99* | (0.75, 1.18)* | 0.98 | (0.65, 1.20) |
| | QR(2) | 0.2 | 0.88 | (0.29, 1.58) | 0.75 | (0.07, 1.91) | 0.65 | (0.01, 2.17) |
| | QR(2) | 0.5 | 0.96 | (0.59, 1.37) | 0.93 | (0.42, 1.51) | 0.90 | (0.28, 1.60) |
| | QR(2) | 0.8 | 0.98 | (0.76, 1.17) | 0.95 | (0.64, 1.21) | 0.92 | (0.56, 1.23) |
| | QR(3) | 0.2 | 0.92 | (0.28, 1.60) | 0.87 | (0.03, 1.93) | 0.83 | (0.00, 2.14) |
| | QR(3) | 0.5 | 0.95 | (0.62, 1.38) | 0.92 | (0.44, 1.52) | 0.89 | (0.32, 1.60) |
| | QR(3) | 0.8 | 0.97 | (0.76, 1.18) | 0.95 | (0.67, 1.22) | 0.91 | (0.60, 1.23) |
| | HT(1) | 0.2 | 0.61 | (0.18, 1.45) | 0.37 | (0.04, 1.64) | 0.24 | (0.01, 1.87) |
| | HT(1) | 0.5 | 0.94 | (0.52, 1.32) | 0.87 | (0.22, 1.46) | 0.78 | (0.07, 1.57) |
| | HT(1) | 0.8 | 1.02* | (0.87, 1.16)* | 1.02 | (0.75, 1.20) | 1.02 | (0.64, 1.21) |
| | HT(2) | 0.2 | 0.97 | (0.34, 1.89) | 1.04 | (0.15, 2.48) | 1.12 | (0.07, 2.60) |
| | HT(2) | 0.5 | 1.01 | (0.65, 1.28) | 1.01 | (0.49, 1.30) | 1.01 | (0.39, 1.33) |
| | HT(2) | 0.8 | 0.94 | (0.74, 1.14) | 0.86 | (0.65, 1.16) | 0.79 | (0.62, 1.17) |
| | HT(3) | 0.2 | 1.20 | (0.43, 1.84) | 1.47 | (0.32, 2.11) | 1.61 | (0.23, 2.32) |
| | HT(3) | 0.5 | 0.94* | (0.69, 1.30)* | 0.90* | (0.64, 1.34)* | 0.87* | (0.59, 1.34)* |
| | HT(3) | 0.8 | 0.93 | (0.76, 1.10) | 0.86 | (0.67, 1.10) | 0.84* | (0.54, 1.08)* |

## B.7.6   Distribution (F)

For this distribution, the two component model is the true model. This distribution is relatively hard to model given there is a lot of overlap at the finite levels that we used. However, QR(2) does a very good job and provides estimates close to the true value. The HT(2) underestimates

| Distribution | Method | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 0.42 | (0.18, 0.80) | 0.14 | (0.04, 0.44) | 0.05 | (0.01, 0.23) |
| | QR(2) | **0.76** | **(0.37, 1.26)** | **0.45** | **(0.11, 1.24)** | **0.26** | **(0.03, 1.17)** |
| | QR(3) | **0.82** | **(0.37, 1.27)** | **0.57** | **(0.09, 1.28)** | **0.36** | **(0.02, 1.26)** |
| | HT(1) | 0.30 | (0.12, 0.54) | 0.08 | (0.02, 0.22) | 0.02 | (0.01, 0.09) |
| | HT(2) | 0.69 | (0.31, 0.95) | 0.44 | (0.08, 0.80) | 0.26 | (0.02, 0.64) |
| | HT(3) | 0.67 | (0.37, 0.94) | 0.42 | (0.13, 0.72) | 0.26 | (0.04, 0.55) |

| Distribution | Method | q | Median, (2.5 and 97.5 percentiles) for different values of $p$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $p = 10^{-4}$ | | $p = 10^{-6}$ | | $p = 10^{-8}$ | |
| | QR(1) | 0.2 | 0.26 | (0.05, 0.85) | 0.07 | (0.00, 0.56) | 0.02 | (0.00, 0.38) |
| | QR(1) | 0.5 | **1.25** | **(0.86, 1.62)** | **1.46** | **(0.88, 1.85)** | **1.62** | **(0.92, 1.93)** |
| | QR(1) | 0.8 | 1.18 | (1.09, 1.22) | 1.22 | (1.16, 1.24) | 1.24 | (1.21, 1.25) |
| | QR(2) | 0.2 | **0.74** | **(0.27, 1.39)** | **0.55** | **(0.09, 1.56)** | **0.44** | **(0.04, 1.69)** |
| | QR(2) | 0.5 | **0.98** | **(0.73, 1.32)** | **0.99** | **(0.69, 1.42)** | **0.99** | **(0.66, 1.55)** |
| | QR(2) | 0.8 | **1.09** | **(0.93, 1.19)** | **1.16** | **(0.95, 1.23)** | **1.20** | **(0.98, 1.24)** |
| | QR(3) | 0.2 | **0.86** | **(0.31, 1.52)** | **0.75** | **(0.12, 1.69)** | **0.69** | **(0.03, 1.83)** |
| | QR(3) | 0.5 | **1.03** | **(0.77, 1.33)** | **1.07** | **(0.76, 1.44)** | **1.09** | **(0.75, 1.50)** |
| | QR(3) | 0.8 | **1.07\*** | **(0.92, 1.19)\*** | **1.14** | **(0.92, 1.24)** | **1.18** | **(0.93, 1.25)** |
| | HT(1) | 0.2 | **0.35** | **(0.08, 1.03)** | 0.13 | (0.01, 0.94) | 0.05 | (0.00, 0.81) |
| | HT(1) | 0.5 | 1.40 | (1.06, 1.69) | 1.67 | (1.21, 1.91) | 1.82 | (1.34, 1.97) |
| | HT(1) | 0.8 | 1.20 | (1.15, 1.23) | 1.24 | (1.21, 1.25) | 1.24 | (1.23, 1.25) |
| | HT(2) | 0.2 | **1.12** | **(0.39, 2.06)** | **1.27** | **(0.20, 2.40)** | **1.43** | **(0.08, 2.47)** |
| | HT(2) | 0.5 | **1.11** | **(0.99, 1.35)** | **1.08\*** | **(0.98, 1.49)\*** | **1.06\*** | **(0.98, 1.59)\*** |
| | HT(2) | 0.8 | **1.11** | **(1.03, 1.20)** | 1.17 | (1.07, 1.24) | 1.20 | (1.12, 1.25) |
| | HT(3) | 0.2 | **1.14** | **(0.51, 1.74)** | **1.30** | **(0.30, 1.80)** | **1.42** | **(0.19, 1.83)** |
| | HT(3) | 0.5 | **1.08** | **(0.86, 1.36)** | **1.16** | **(0.84, 1.45)** | **1.22** | **(0.83, 1.49)** |
| | HT(3) | 0.8 | 1.12 | (1.05, 1.20) | 1.17 | (1.10, 1.23) | 1.20 | (1.14, 1.24) |

# Appendix C

# Appendix to Chapter 4

## C.1  Proofs

*Proof of Proposition 4.2.2.*  We prove that for sufficiently large $n$, there exists a constant $C_1 > 0$ such that

$$\mathscr{I}_n := \int_I e^{g_n(x) - g_n(x_n^*)} \, \mathrm{d}x \cdot \left(-g_n^{(k_0)}(x_n^*)\right)^{\frac{1}{k_0}} \geq C_1.$$

To bound $\mathscr{I}_n$ from below, we first simplify its expression by applying the variable transformation $y = t_n(x) := (x - x_n^*) \left(-g_n^{(k_0)}(x_n^*)\right)^{1/k_0}$ and defining

$$h_n(y) := g_n \left(x_n^* + y \left(-g_n^{(k_0)}(x_n^*)\right)^{-\frac{1}{k_0}}\right), \quad \text{for } y \in I_n' := \{t_n(x): \ x \in I\}.$$

Then, the integral $\mathscr{I}_n$ becomes

$$\mathscr{I}_n = \int_{I_n'} e^{h_n(y) - h_n(0)} \, \mathrm{d}y.$$

We note that for all $n \in \mathbb{N}$, we have $0 \in I_n'$, $h_n \in C^{k_0}(I_n')$, and $h_n(0) > h_n(y)$ for all $y \in I_n' \setminus \{0\}$. Moreover, we have for $y \in I_n'$, $i = 1, \dots, k_0$,

$$h_n^{(i)}(y) = g_n^{(i)} \left(x_n^* + y \left(-g_n^{(k_0)}(x_n^*)\right)^{-1/k}\right) \cdot \left(-g_n^{(k_0)}(x_n^*)\right)^{-i/k_0}.$$

Hence, $h_n^{(k_0)}(0) = -1$ and $\lim_{n \to \infty} h_n^{(i)}(0) = 0$ for all $1 \leq i < k_0$. Using Taylor's theorem, there exists a function $\xi(y)$ taking on a value between $0$ and $y$ such that

$$h_n(y) - h_n(0) = \sum_{i=1}^{k_0 - 1} \frac{y^i}{i!} h_n^{(i)}(0) + \frac{y^{k_0}}{k_0!} h_n^{(k_0)}(\xi(y)).$$

Let $\tilde{\varepsilon} > 0$. Because $\lim_{n \to \infty} h_n^{(i)}(0) = 0$ for all $i < k_0$, we can find an $N_0 \equiv N_0(\tilde{\epsilon}) \in \mathbb{N}$ such that for all $n > N_0$, we have $\max_{i=1,\dots,k_0-1} |h_n^{(i)}(0)| < \tilde{\varepsilon}$. Moreover, from the assumptions of the proposition, we can find a $\delta > 0$ and associated $\varepsilon > 0$ and $N_1 \equiv N_1(\delta) \in \mathbb{N}$ such that for all $n > N_1$, $h_n^{(k_0)}(y) > -(1 + \varepsilon)$ for $y \in (-\delta, \delta) \cap I_n'$. For $n > \max\{N_0, N_1\}$,

$$h_n(y) - h_n(0) > -|y|\tilde{\varepsilon} - \frac{|y|^2}{2!}\tilde{\varepsilon} - \cdots - \frac{|y|^{k_0-1}}{(k_0-1)!}\tilde{\varepsilon} - \frac{(1+\varepsilon)|y|^{k_0}}{k_0!} > -\tilde{\varepsilon}e^\delta - \frac{(1+\varepsilon)|y|^{k_0}}{k_0!}$$

for $y \in (-\delta, \delta) \cap I_n'$. Hence, we derive a lower bound

$$\mathscr{I}_n \geq e^{-\tilde{\varepsilon}e^\delta} \int_{I_n' \cap (-\delta, \delta)} e^{-\frac{(1+\varepsilon)|y|^{k_0}}{k_0!}} \, \mathrm{d}y =: C_1.$$

From the connectedness of $I$ and $0 \in I_n'$, we conclude that $I_n' \cap (-\delta, \delta)$ has positive mass under the Lebesgue measure. Hence, $C_1 \in (0, \infty)$.                     $\square$

*Proof of Proposition 4.4.1.* Let $(X, Y)$ be a random vector such that $X$ and $Y$ both have standard Laplace margins. Moreover, assume that there exist $-1 \leq \alpha \leq 1$, $0 \leq \beta < 1$ and $u > 0$ such that for $x > u$

$$\mathbb{P}(Y > y \mid X = x) = \overline{H}\left(\frac{y - \alpha x}{x^\beta}\right)$$

holds for all $y \in \mathbb{R}$ with

$$\overline{H}(z) = \exp(-\gamma z^\delta)\mathbb{1}\{z > 0\} + \mathbb{1}\{z \leq 0\},$$

where $\gamma, \delta > 0$. Now, let $Z$ be a random variable that is independent of $X$ and have survival function $\overline{H}$. We derive that $\delta \geq (1 - \beta)^{-1}$ must hold. Since $Y$ is distributed as a standard Laplace, we have for $y > 0$

$$\frac{\exp(-y)}{2} = \mathbb{P}(\alpha X + X^\beta Z \geq y, \ X \geq u) + \mathbb{P}(Y \geq y, \ X < u)$$

$$\geq \mathbb{P}(\alpha X + X^\beta Z \geq y, \ X \geq u) \geq \mathbb{P}(X^\beta Z \geq y, \ X \geq u)$$

$$= \int_u^\infty \mathbb{P}\left(Z \geq \frac{y}{x^\beta}\right) f_X(x) \, \mathrm{d}x = \frac{1}{2}\int_u^\infty \exp\left(-\frac{\gamma y^\delta}{x^{\beta\delta}} - x\right) \mathrm{d}x =: \tilde{\mathscr{I}}_y.$$

We will show that $2\exp(y)\tilde{\mathscr{I}}_y > 1$ for sufficiently large $y$, if $\delta < (1 - \beta)^{-1}$, which thus would contradict with the marginal distribution of $Y$. This result holds trivially

for $\beta = 0$. So, for now, we let $\beta > 0$. We will prove this asymptotic inequality by applying Proposition 4.2.2, with $k_0 = 2$, to bound $\tilde{\mathscr{I}}_y$ from below.

First define $I := [u, \infty)$ as the integration domain, and

$$g_y(x) := \exp\left(-\frac{\gamma y^\delta}{x^{\beta\delta}} - x\right) \mathbb{1}\{x \in I\}, \quad \text{and} \quad h_y(x) := \left(-\frac{\gamma y^\delta}{x^{\beta\delta}} - x\right) \mathbb{1}\{x \in I\}.$$

Next we find the mode $x_y^*$ of $g_y(x)$. We assume that $x_y^*$ lies in the interior of $I$ such that $h_y'(x_y^*) = 0$, which implies that $\beta\delta\gamma y^\delta (x_y^*)^{-\beta\delta-1} = 1$. So, $x_y^* = (\beta\delta\gamma)^{\frac{1}{\beta\delta+1}} y^{\frac{\delta}{\beta\delta+1}}$, which lies in the interior of $I$ for sufficiently large $y$. We now compute

$$g_y(x_y^*) = \exp\left(-\frac{\gamma y^\delta}{(x_y^*)^{\beta\delta}} - x_y^*\right) = \exp\left(-Ay^{\frac{\delta}{\beta\delta+1}}\right)$$

with $A := \gamma (\beta\delta\gamma)^{-\frac{\beta\delta}{\beta\delta+1}} + (\beta\delta\gamma)^{\frac{1}{\beta\delta+1}}$. Secondly,

$$h_y''(x_y^*) = -\beta\delta(\beta\delta+1)(x_y^*)^{-\beta\delta-2}\gamma y^\delta = -(\beta\delta+1)(\beta\delta\gamma)^{-\frac{1}{\beta\delta+1}} y^{-\frac{\delta}{\beta\delta+1}}.$$

Using these expressions, we check that the assumptions from Proposition 4.2.2 with $k_0 = 2$ are satisfied. First, note that $h_y'(x_y^*)(-h_y''(x_y^*))^{-1/2} = 0$. Next let $C > 0$ and $|x| \leq C$, then

$$\lim_{y\to\infty} \frac{h_y''\left(x_y^* + \frac{x}{\sqrt{-h_y''(x_y^*)}}\right)}{h_y''(x_y^*)}$$

$$= \lim_{y\to\infty} \frac{-\beta\delta(\beta\delta+1)\left((\beta\delta\gamma)^{\frac{1}{\beta\delta+1}} y^{\frac{\delta}{\beta\delta+1}} + \frac{x}{\sqrt{(\beta\delta+1)(\beta\delta\gamma)^{-\frac{1}{\beta\delta+1}} y^{-\frac{\delta}{\beta\delta+1}}}}\right)^{-\beta\delta-2} \gamma y^\delta}{-(\beta\delta+1)(\beta\delta\gamma)^{-\frac{1}{\beta\delta+1}} y^{-\frac{\delta}{\beta\delta+1}}}$$

$$= \lim_{y\to\infty} \frac{\left(y^{\frac{\delta}{\beta\delta+1}} + \frac{x}{\sqrt{(\beta\delta+1)(\beta\delta\gamma)^{\frac{1}{\beta\delta+1}} y^{-\frac{\delta}{\beta\delta+1}}}}\right)^{-\beta\delta-2} y^\delta}{y^{-\frac{\delta}{\beta\delta+1}}}$$

$$= \lim_{y\to\infty} \left(1 + \frac{x}{\sqrt{(\beta\delta+1)(\beta\delta\gamma)^{\frac{1}{\beta\delta+1}} y^{\frac{\delta}{\beta\delta+1}}}}\right)^{-\beta\delta-2}$$

$$= 1,$$

which is sufficient to show that for each $\tilde{x}$, Proposition 4.2.2 is applicable with $k_0 = 2$ on interval $I_{\tilde{x}} := \left[x_y^* - \frac{\tilde{x}}{\sqrt{-h_y''(x_y^*)}}, x_y^* + \frac{\tilde{x}}{\sqrt{-h_y''(x_y^*)}}\right]$. Hence for each $\tilde{x}$, there exists a

constant $C_1(\tilde{x}) > 0$ such that for sufficiently large $y$,

$$y^{-\frac{\delta/2}{\beta\delta+1}} \exp\left(Ay^{\frac{\delta}{\beta\delta+1}}\right) \cdot \tilde{\mathscr{I}}_y \geq y^{-\frac{\delta/2}{\beta\delta+1}} \exp\left(Ay^{\frac{\delta}{\beta\delta+1}}\right) \cdot \int_{I_{\tilde{x}}} g_y(x)\,\mathrm{d}x$$
$$= \frac{C_1(\tilde{x})\,(\beta\delta\gamma)^{\frac{1}{2(\beta\delta+1)}}}{\sqrt{\beta\delta+1}}.$$

Using the inequality $2\exp(y)\tilde{\mathscr{I}}_y \leq 1$, we must have

$$y^{-\frac{\delta/2}{\beta\delta+1}} \exp\left(Ay^{\frac{\delta}{\beta\delta+1}}\right) \cdot \frac{1}{2}\exp(-y) \geq \frac{C_1(\tilde{x})\,(\beta\delta\gamma)^{\frac{1}{2(\beta\delta+1)}}}{\sqrt{\beta\delta+1}} \qquad \text{(C.1.1)}$$

for sufficiently large $y$. Since $0 \leq \beta < 1$, we note that if $\delta < (1-\beta)^{-1}$ then inequality (C.1.1) does not hold. So, we derive that $\delta \geq (1-\beta)^{-1}$. $\qquad \square$

# Appendix D

# Supplementary Information to Chapter 4

## D.1 Introduction

We give an overview of the content in the Supplementary Material. In Section D.2.1, we give parameter estimates of the Haver-Winterstein (HW) distribution as referred to in Section 4.3. In Sections D.3-D.3.4, we give the details of the calculations that support the arguments in Section 4.3. Finally, in Section D.4 one can find the mathematical derivations of the results stated in of Section 4.4.

## D.2 Supplementary Material

### D.2.1 HW model parameters

See Table D.2.1.

## D.3 Details on calculations for the HW distribution

Let $(X, Y)$ follow the HW model, see Section 4.3, with $0 < \mu_2 < 0.5$ and $2\mu_2 < k$. The goal is to calculate the asymptotic behaviour of joint probabilities $\mathbb{P}(X > F_X^{-1}(p), Y >$

| Parameter | $\alpha$ | $\theta$ | $u$ | $k$ | $\lambda$ | |
|-----------|-------|-------|-------|-------|-------|-------|
| | 0.573 | 0.893 | 3.803 | 1.550 | 2.908 | |
| Parameter | $\mu_0$ | $\mu_1$ | $\mu_2$ | $\sigma_0$ | $\sigma_1$ | $\sigma_2$ |
| | 1.134 | 0.892 | 0.225 | 0.005 | 0.120 | 0.455 |

Table D.2.1: Parameter estimates of the joint probability density function of significant wave height $H_S$ (m) and wave period $T_p$ (s) for the Northern North Sea (Haver and Winterstein, 2009).

$F_Y^{-1}(p))$ when $p$ tends to 1 where $F_X$ and $F_Y$ denote the distribution functions of the random variables $X$ and $Y$, respectively. First, we evaluate the distribution function of $Y$ at large values such that we can calculate $F_Y^{-1}(p)$. After, we compute joint probabilities, like $\mathbb{P}(X_E > u, Y_E > u)$, where $X_E$ and $Y_E$ denote $X$ and $Y$, respectively, on exponential margins, i.e., $X_E = -\log(1 - F_X(X))$ and $Y_E = -\log(1 - F_Y(Y))$.

We write down an analytical expression for the survival function $\overline{F}_Y$ of $Y$

$$\overline{F}_Y(y) := \mathbb{P}(Y > y) = \int_0^\infty \overline{\Phi}\left(\frac{\log y - \mu(x)}{\sigma(x)}\right) f_X(x)\, dx \qquad \text{(D.3.1)}$$

where $\mu(x)$ and $\sigma(x)$ are defined in the main paper. We remark that we need to evaluate $\overline{F}_Y$ at large $y$. To that end, we denote $p_y(x) := (\log y - \mu(x))/\sigma(x)$, and the integrand

$$g_y(x) := \overline{\Phi}\left(p_y(x)\right) f_X(x). \qquad \text{(D.3.2)}$$

As seen in Figure 4.3.1, the integrand $g_y$ has two local maxima for $y$ large enough. Hence, Proposition 4.2.2 is not directly applicable. However, we can use the proposition to indirectly prove a lower bound for the intergal (D.3.1). Next, it is straightforward to find an upper bound for the integral with the same rate of decay as the proven lower bound.

We adopt the following set of steps: (a) we prove that there exist (at least) two local maxima $x_y^*$ and $x_y^{**}$, and find expressions for both. If there are more then $x_y^*$ is the one with the smallest argument, and $x_y^{**}$ is the one with the second smallest argument; (b) we show that we can apply part of Proposition 4.2.2 to the smaller of the two local maxima, which gives a lower bound for the integral; (c) we define an

upperbound $\tilde{g}_y$ for the integrand $g_y$, compute the integral of $\tilde{g}_y$, and show that this integral has the same rate of decay as the lower bound; (d) finally, we combine the results to get an asymptotic expression for $\overline{F}_Y(y)$ as $y \to \infty$.

We need to start by working out the expressions for the local maxima. We do this by considering all possible options, which yields five (types of) local extrema $0 < x_0 < x_1 < x_2 < x_3 < x_4 < \infty$ that satisfy the following: (i) as $y \to \infty$, $p_y(x_0) \to \infty$ holds and $x_0 \to 0$; (ii) as $y \to \infty$, $p_y(x_1) \to \infty$ holds and $x_1 \to c \in (0, \infty)$; (iii) as $y \to \infty$, $p_y(x_2) \to \infty$ holds and $x_2 \to \infty$; (iv) as $y \to \infty$, $p_y(x_3) \to c \in \mathbb{R}$ holds and $x_3 \to \infty$; (v) as $y \to \infty$, $p_y(x_4) \to -\infty \in \mathbb{R}$ holds and $x_4 \to \infty$. It is straightforward to show that $x_3$ and $x_4$ cannot exist. However, this argument is unnecessary for the purpose of this section.

Finally, we calculate $\overline{F}_Y(y)$ using Proposition 4.2.2. In particular, we will get a lower bound by applying Proposition 4.2.2 around the local maximum $x_0$ and we derive an upper bound directly.

## D.3.1 Finding local extrema

We consider cases (i), (ii) and (iii). These cases have in common that $p_y(x_*) \to \infty$ for $x_* \in \{x_0, x_1, x_2\}$. We will write $x_*$ rather than either $x_0, x_1, x_2$ to not distinguish between arguments that are applicable to all three cases. To find an expression for $x_*$ in closed form, we define $h_y(x) := \log g_y(x)$ and we solve $h_y'(x_*) = 0$. First, we calculate $h_y'(x)$,

$$h_y'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \left( \log \overline{\Phi}\left( p_y(x) \right) + \log f_X(x) \right)$$
$$= \frac{-\varphi\left( p_y(x) \right)}{\overline{\Phi}\left( p_y(x) \right)} \cdot p_y'(x) + \frac{f_X'(x)}{f_X(x)}.$$

Since $p_y(x_*) \to \infty$, we can simplify this expression by using Mills' ratio, which says that

$$\frac{\overline{\Phi}(x)}{\varphi(x)} = \frac{1}{x} - \frac{1}{x^3} + O(x^{-5})$$

as $x \to \infty$, which implies $\varphi(x)/\overline{\Phi}(x) = x + x^{-1} + O(x^{-3})$ as $x \to \infty$. Moreover, we can write

$$
\begin{aligned}
p_y'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\log y - \mu(x)}{\sigma(x)}\right) \\
&= -(\log y - \mu(x)) \cdot \frac{\sigma'(x)}{\sigma(x)^2} - \frac{\mu'(x)}{\sigma(x)} \\
&= -p_y(x) \cdot \frac{\sigma'(x)}{\sigma(x)} - \frac{\mu'(x)}{\sigma(x)}.
\end{aligned}
$$

So,

$$
\begin{aligned}
h_y'(x_*) &= -\left(p_y(x_*) + \frac{1}{p_y(x_*)} + O\left(p_y(x_*)^{-3}\right)\right) \cdot \left(-p_y(x_*) \cdot \frac{\sigma'(x_*)}{\sigma(x_*)} - \frac{\mu'(x_*)}{\sigma(x_*)}\right) \\
&\quad + \frac{f_X'(x_*)}{f_X(x_*)} \\
&= p_y(x_*)^2 \cdot \frac{\sigma'(x_*)}{\sigma(x_*)} + p_y(x_*) \cdot \frac{\mu'(x_*)}{\sigma(x_*)} + \frac{\sigma'(x_*)}{\sigma(x_*)} + \frac{\mu'(x_*)}{p_y(x_*)\sigma(x_*)} \\
&\quad + O\left(\frac{\sigma'(x_*)}{p_y(x_*)^2\sigma(x_*)} + \frac{\mu'(x_*)}{p_y(x_*)^3\sigma(x_*)}\right) + \frac{f_X'(x_*)}{f_X(x_*)}
\end{aligned}
$$

as $y \to \infty$. We now fill in the parametric forms for $\mu$ and $\sigma$. We can then simplify this expression to

$$
\begin{aligned}
h_y'(x_*) &= \frac{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^2 \cdot -\frac{1}{2}\sigma_1\sigma_2 \exp\left(-\sigma_2 x_*\right)\left(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*)\right)^{-1/2}}{(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^{3/2}} \\
&\quad + \frac{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})\mu_1\mu_2 x_*^{\mu_2-1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*)} \\
&\quad - \frac{\frac{1}{2}\sigma_1\sigma_2 \exp(-\sigma_2 x_*)(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^{-1/2}}{(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^{1/2}} \\
&\quad + \frac{\mu_1\mu_2 x_*^{\mu_2-1}}{\log y - \mu_0 - \mu_1 x_*^{\mu_2}} \\
&\quad + O\left(\frac{x_*^{\mu_2-1}}{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^3} + \frac{\exp(-\sigma_2 x_*)}{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^2}\right) + \frac{f_X'(x_*)}{f_X(x_*)} \\
&= -\frac{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^2 \cdot \sigma_1\sigma_2 \exp\left(-\sigma_2 x_*\right)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^2} \\
&\quad + \frac{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})\mu_1\mu_2 x_*^{\mu_2-1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*)} - \frac{\sigma_1\sigma_2 \exp(-\sigma_2 x_*)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))} \\
&\quad + \frac{\mu_1\mu_2 x_*^{\mu_2-1}}{\log y - \mu_0 - \mu_1 x_*^{\mu_2}} + O\left(\frac{x_*^{\mu_2-1}}{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^3} + \frac{\exp(-\sigma_2 x_*)}{(\log y - \mu_0 - \mu_1 x_*^{\mu_2})^2}\right) \\
&\quad + \frac{f_X'(x_*)}{f_X(x_*)}.
\end{aligned}
$$

Since, $h'_y(x_*) = 0$ for all $y$, we can let $y \to \infty$, to further simplify

$$
\begin{aligned}
0 = {} & \lim_{y \to \infty} h'_y(x_*) \\
= {} & \lim_{y \to \infty} \Bigg( - \log^2 y \cdot \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_*)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^2} \\
& + \log y \cdot \left( \frac{(\mu_0 + \mu_1 x_*^{\mu_2}) \sigma_1 \sigma_2 \exp(-\sigma_2 x_*)}{(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^2} + \frac{\mu_1 \mu_2 x_*^{\mu_2 - 1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*)} \right) \\
& - \frac{(\mu_0 + \mu_1 x_*^{\mu_2}) \mu_1 \mu_2 x_*^{\mu_2 - 1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*)} - \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_*)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))} \\
& - \frac{(\mu_0 + \mu_1 x_*^{\mu_2})^2 \cdot \sigma_1 \sigma_2 \exp(-\sigma_2 x_*)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))^2} + \frac{\mu_1 \mu_2 x_*^{\mu_2 - 1}}{\log y - \mu_0 - \mu_1 x_*^{\mu_2}} + \frac{f'_X(x_*)}{f_X(x_*)} \Bigg). \\
& \hspace{11cm} \text{(D.3.3)}
\end{aligned}
$$

We now split up the analysis into the three cases: (i) $x_* = x_0 \to 0$; (ii) $x_* = x_1 \to c \in (0, \infty)$; (iii) $x_* = x_2 \to \infty$.

**Case (i):** $x_* = x_0 \to 0$

In this case, there must exist a $y' > 0$ such that for all $y > y'$, $x_0(y) < u$. So, let $y > y'$, then

$$
\frac{f'_X(x_0)}{f_X(x_0)} = -\frac{\log x_0 - \theta}{x_0 \alpha^2} - \frac{1}{x_0}.
$$

Filling in $x_* = x_0$ simplifies equation (D.3.3) to

$$
\begin{aligned}
\lim_{y \to \infty} \Bigg( & - \log^2 y \cdot \frac{\sigma_1 \sigma_2}{2(\sigma_0 + \sigma_1)^2} + \log y \cdot \left( \frac{\mu_0 \sigma_1 \sigma_2}{(\sigma_0 + \sigma_1)^2} + \frac{\mu_1 \mu_2 x_0^{\mu_2 - 1}}{\sigma_0 + \sigma_1} \right) - \frac{\mu_0 \mu_1 \mu_2 x_0^{\mu_2 - 1}}{\sigma_0 + \sigma_1} \\
& - \frac{\sigma_1 \sigma_2}{2(\sigma_0 + \sigma_1)} - \frac{\mu_0^2 \sigma_1 \sigma_2}{2(\sigma_0 + \sigma_1)^2} + \frac{\mu_1 \mu_2 x_0^{\mu_2 - 1}}{\log y} - \frac{\log x_0 - \theta}{x_0 \alpha^2} - \frac{1}{x_0} \Bigg) = 0.
\end{aligned}
$$
$$
\hspace{12cm} \text{(D.3.4)}
$$

Because $0 < \mu_2 < 0.5$, the dominating terms within this limit are of the order $\log^2(y)$ and $\log y \cdot x_0^{\mu_2 - 1}$. Indeed, $(\log x_0)/x_0$ is dominated by both of these terms since, we must eventually have $x_0^{2\mu_2 - 2} > (\log x_0)/x_0$. So $x_0$ must satisfy as $y \to \infty$

$$
-\log y \cdot \frac{\sigma_1 \sigma_2}{2(\sigma_0 + \sigma_1)} + \log y \cdot \mu_1 \mu_2 \cdot x_0^{\mu_2 - 1} = O\left( \frac{\log x_0}{x_0 \log y} \right).
$$

Finally, we derive the following asymptotic expression for $x_0$ as $y \to \infty$

$$
x_0 = \left( \frac{\sigma_1 \sigma_2}{2 \mu_1 \mu_2 (\sigma_0 + \sigma_1)} \right)^{-\frac{1}{1 - \mu_2}} \cdot (\log y)^{-\frac{1}{1 - \mu_2}} + O\left( \log^{-2}(y) \right). \hspace{1cm} \text{(D.3.5)}
$$

We will later show that $h_y''(x_0) < 0$. So, indeed $x_0$ corresponds to a local maximum.

**Case (ii):** $x_* = x_1 \to c \in (0, \infty)$

In this case, equation (D.3.3) is equivalent to

$$\lim_{y \to \infty} -c_1 \log^2 y + c_2 \log y - c_3 = 0$$

where

$$0 < c_1 = \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 c)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 c))^2}$$

$$0 < c_2 = \frac{(\mu_0 + \mu_1 c^{\mu_2})\sigma_1 \sigma_2 \exp(-\sigma_2 c)}{(\sigma_0 + \sigma_1 \exp(-\sigma_2 c))^2} + \frac{\mu_1 \mu_2 c^{\mu_2 - 1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 c)}$$

$$0 < c_3 = \frac{(\mu_0 + \mu_1 c^{\mu_2})\mu_1 \mu_2 c^{\mu_2 - 1}}{\sigma_0 + \sigma_1 \exp(-\sigma_2 c)} + \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_*)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 x_*))}$$
$$+ \frac{(\mu_0 + \mu_1 c^{\mu_2})^2 \cdot \sigma_1 \sigma_2 \exp(-\sigma_2 c)}{2(\sigma_0 + \sigma_1 \exp(-\sigma_2 c))^2} + \frac{\mu_1 \mu_2 c^{\mu_2 - 1}}{\log y - \mu_0 - \mu_1 c^{\mu_2}} - \frac{f_X'(c)}{f_X(c)}.$$

We can now clearly see that equation (D.3.3) cannot be valid under this assumption. We conclude that $x_1$ does not exist.

**Case (iii):** $x_* = x_2 \to \infty$

Finally, let $x_* = x_2 \to \infty$. In this case, there must exist a $y'' > 0$ such that for all $y > y''$, $x_0(y) > u$. So, let $y > y''$, then

$$\frac{f_X'(x_0)}{f_X(x_0)} = \frac{k-1}{x_*} - \frac{k x_*^{k-1}}{\lambda^k}.$$

Now, equation (D.3.3) is equivalent to

$$\lim_{y \to \infty} \left( -\log^2 y \cdot \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_2)}{2\sigma_0^2} \right.$$
$$+ \log y \cdot \left( \frac{(\mu_0 + \mu_1 x_2^{\mu_2})\sigma_1 \sigma_2 \exp(-\sigma_2 x_2)}{\sigma_0^2} + \frac{\mu_1 \mu_2 x_2^{\mu_2 - 1}}{\sigma_0} \right)$$
$$- \frac{(\mu_0 + \mu_1 x_2^{\mu_2})\mu_1 \mu_2 x_2^{\mu_2 - 1}}{\sigma_0} - \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_2)}{2\sigma_0} \qquad \text{(D.3.6)}$$
$$- \frac{(\mu_0 + \mu_1 x_2^{\mu_2})^2 \cdot \sigma_1 \sigma_2 \exp(-\sigma_2 x_2)}{2\sigma_0^2}$$
$$\left. + \frac{\mu_1 \mu_2 x_2^{\mu_2 - 1}}{\log y - \mu_0 - \mu_1 x_2^{\mu_2}} + \frac{k-1}{x_2} - \frac{k x_2^{k-1}}{\lambda^k} \right) = 0.$$

The dominating terms in equation (D.3.6) are of the order $\log^2 y$, $\log y \cdot x_2^{\mu_2 - 1}$ and $x_2^{k-1}$. So, we can simplify equation (D.3.6) to

$$\lim_{y \to \infty} -\log^2 y \cdot \frac{\sigma_1 \sigma_2 \exp(-\sigma_2 x_2)}{2\sigma_0^2} + \log y \cdot \frac{\mu_1 \mu_2 x_2^{\mu_2 - 1}}{\sigma_0} - \frac{k x_2^{k-1}}{\lambda^k} = 0. \qquad \text{(D.3.7)}$$

We note that the first and third terms have a negative sign, and the second has a positive sign. We note that we cannot simplify this further without considering the following two options as $y \to \infty$: (a) $\exp(-\sigma_2 x_2) \log^2 y \gg x_2^{k-1}$; (b) $\exp(-\sigma_2 x_2) \log^2 y \ll x_2^{k-1}$. Both of these cases will yield a solution to equation (D.3.7) which we call $x_{2a}$ and $x_{2b}$ respectively.

**Case (iii-a):** $x_* = x_{2a} \to \infty$ **and** $\exp(-\sigma_2 x_{2a}) \log^2 y \gg x_{2a}^{k-1}$

We derive from equation (D.3.7) that $x_{2a}$ must satisfy as $y \to \infty$

$$-\log y \cdot \frac{\sigma_1 \sigma_2}{2\sigma_0} \exp(-\sigma_2 x_{2a}) + \mu_1 \mu_2 x_{2a}^{\mu_2 - 1} = O\left(\frac{x_{2a}^{k-1}}{\log y}\right).$$

So, $x_{2a}$ must satisfy as $y \to \infty$

$$x_{2a}^{\mu_2 - 1} \exp(\sigma_2 x_{2a}) = \log y \cdot \left(\frac{\sigma_1 \sigma_2}{2\sigma_0 \mu_1 \mu_2} + O\left(\frac{x_{2a}^{k-1}}{\exp(-\sigma_2 x_{2a}) \log^2 y}\right)\right).$$

Finally, we derive the following asymptotic expression for $x_{2a}$ as $y \to \infty$

$$x_{2a} = \frac{\log \log y}{\sigma_2} + O(\log \log \log y). \qquad \text{(D.3.8)}$$

**Case (iii-b):** $x_* = x_{2b} \to \infty$ **and** $\exp(-\sigma_2 x_{2a}) \log^2 y \ll x_{2a}^{k-1}$

We derive from equation (D.3.7) that $x_{2b}$ must satisfy as $y \to \infty$

$$\log y \cdot \frac{\mu_1 \mu_2}{\sigma_0} - \frac{k x_2^{k - \mu_2}}{\lambda^k} = O\left(\log^2 y \exp(-\sigma_2 x_2) x_2^{1 - \mu_2}\right).$$

So, $x_{2b}$ must satisfy as $y \to \infty$

$$x_{2b} = \left(\frac{\lambda^k \mu_1 \mu_2}{k \sigma_0}\right)^{\frac{1}{k - \mu_2}} \cdot (\log y)^{\frac{1}{k - \mu_2}} + O\left((\log y)^{\frac{1}{k - \mu_2} - \frac{k - 2\mu_2}{k - \mu_2}}\right). \qquad \text{(D.3.9)}$$

## D.3.2   Identifiying local maxima and local minima

In the previous section, we have found expressions for local extrema, see equations (D.3.5), (D.3.8) and (D.3.9). In this section, we will show by using the second derivative $h_y''$ that $x_0$ and $x_{2b}$ correspond to local maxima and that $x_{2a}$ corresponds to a local minimum.

We calculated before

$$h_y'(x) = \frac{-\varphi\left(p_y(x)\right)}{\overline{\Phi}\left(p_y(x)\right)} \cdot p_y'(x) + \frac{f_X'(x)}{f_X(x)}.$$

So,

$$h_y''(x) = -\frac{\varphi\left(p_y(x)\right)^2 p_y'(x)^2}{\overline{\Phi}\left(p_y(x)\right)^2} - \frac{\varphi'\left(p_y(x)\right) p_y'(x)^2}{\overline{\Phi}\left(p_y(x)\right)} - \frac{\varphi\left(p_y(x)\right) p_y''(x)}{\overline{\Phi}\left(p_y(x)\right)} - \frac{f_X'(x)^2}{f_X(x)^2} + \frac{f_X''(x)}{f_X(x)}.$$

We can simplify $h_y''(x_*)$ for $x_* \in \{x_0, x_{2a}, x_{2b}\}$ as $y \to \infty$ by using the identities $\varphi(p_y(x_*))/\overline{\Phi}(p_y(x_*)) \sim p_y(x_*) + p_y(x_*)^{-1}$ as $y \to \infty$ and $\varphi'(x) = -x\varphi(x)$. We get as $y \to \infty$

$$h_y''(x_*) \sim -\left(p_y(x_*) + \frac{1}{p_y(x_*)}\right)^2 p_y'(x_*)^2 + \left(p_y(x_*) + \frac{1}{p_y(x_*)}\right) p_y(x_*)p_y'(x_*)^2$$

$$- \left(p_y(x_*) + \frac{1}{p_y(x_*)}\right) p_y''(x_*) - \frac{f_X'(x_*)^2}{f_X(x_*)^2} + \frac{f_X''(x_*)}{f_X(x_*)}$$

$$\sim -p_y'(x_*)^2 - \frac{p_y'(x_*)^2}{p_y(x_*)^2} - p_y(x_*)p_y''(x_*) - \frac{p_y''(x_*)}{p_y(x_*)} - \frac{f_X'(x_*)^2}{f_X(x_*)^2} + \frac{f_X''(x_*)}{f_X(x_*)}$$

$$\sim -p_y'(x_*)^2 - p_y(x_*)p_y''(x_*) - \frac{f_X'(x_*)^2}{f_X(x_*)^2} + \frac{f_X''(x_*)}{f_X(x_*)}.$$

We work out $p_y'(x)^2$ and $p_y''(x)$ in terms of $p_y(x)$

$$p_y'(x)^2 = \left(-p_y(x) \cdot \frac{\sigma'(x)}{\sigma(x)} - \frac{\mu'(x)}{\sigma(x)}\right)^2 = p_y(x)^2 \cdot \frac{\sigma'(x)^2}{\sigma(x)^2} + p_y(x) \cdot \frac{2\sigma'(x)\mu'(x)}{\sigma(x)^2} + \frac{\mu'(x)^2}{\sigma(x)^2}$$

and

$$p_y''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\log y - \mu(x)}{\sigma(x)}\right)$$

$$= -\frac{\mu''(x)}{\sigma(x)} + 2 \cdot \frac{\mu'(x)\sigma'(x)}{\sigma(x)^2} + (\log y - \mu(x)) \cdot \left(\frac{2\sigma'(x)^2}{\sigma(x)^3} - \frac{\sigma''(x)}{\sigma(x)^2}\right)$$

$$= p_y(x) \cdot \left(\frac{2\sigma'(x)^2}{\sigma(x)^2} - \frac{\sigma''(x)}{\sigma(x)}\right) + 2 \cdot \frac{\mu'(x)\sigma'(x)}{\sigma(x)^2} - \frac{\mu''(x)}{\sigma(x)}.$$

So, as $y \to \infty$

$$h_y''(x_*) \sim -p_y(x_*)^2 \cdot \frac{\sigma'(x_*)^2}{\sigma(x_*)^2} - p_y(x_*) \cdot \frac{2\sigma'(x_*)\mu'(x_*)}{\sigma(x_*)^2} - \frac{\mu'(x_*)^2}{\sigma(x_*)^2}$$

$$- p_y(x_*) \left( p_y(x_*) \cdot \left( \frac{2\sigma'(x_*)^2}{\sigma(x_*)^2} - \frac{\sigma''(x_*)}{\sigma(x_*)} \right) + 2 \cdot \frac{\mu'(x_*)\sigma'(x_*)}{\sigma(x_*)^2} - \frac{\mu''(x_*)}{\sigma(x_*)} \right)$$

$$- \frac{f_X'(x_*)^2}{f_X(x_*)^2} + \frac{f_X''(x_*)}{f_X(x_*)}$$

$$\sim -p_y(x_*)^2 \cdot \left( \frac{3\sigma'(x_*)^2}{\sigma(x_*)^2} - \frac{\sigma''(x_*)}{\sigma(x_*)} \right) - p_y(x_*) \left( \frac{4\mu'(x_*)\sigma'(x_*)}{\sigma(x_*)^2} - \frac{\mu''(x_*)}{\sigma(x_*)} \right)$$

$$- \frac{\mu'(x_*)^2}{\sigma(x_*)^2} - \frac{f_X'(x_*)^2}{f_X(x_*)^2} + \frac{f_X''(x_*)}{f_X(x_*)}.$$

For $x = x_0$, we have

$$\mu'(x_0) = \mu_1\mu_2 x_0^{\mu_2-1},$$

$$\mu''(x_0) = -\mu_1\mu_2(1-\mu_2)x_0^{\mu_2-2},$$

$$\sigma(x_0) \sim \sqrt{\sigma_0 + \sigma_1},$$

$$\sigma'(x_0) \sim -\sigma_1\sigma_2/(2\sqrt{\sigma_0 + \sigma_1}),$$

$$\sigma''(x_0) \sim \sigma_1^2\sigma_2^2/(4(\sigma_0 + \sigma_1)^{3/2}), \text{ and}$$

$$p_y(x_0) \sim \log y/\sqrt{\sigma_0 + \sigma_1}.$$

So,

$$h_y''(x_0) \sim -\frac{\log^2 y}{\sigma_0 + \sigma_1} \cdot \left( \frac{3\sigma_1^2\sigma_2^2}{4(\sigma_0 + \sigma_1)^2} - \frac{\sigma_1^2\sigma_2^2}{4(\sigma_0 + \sigma_1)^2} \right)$$

$$+ \frac{\log y}{\sqrt{\sigma_0 + \sigma_1}} \left( \frac{2\mu_1\mu_2 x_0^{\mu_2-1} \cdot \sigma_1\sigma_2}{(\sigma_0 + \sigma_1)^{3/2}} - \frac{\mu_1\mu_2(1-\mu_2)x_0^{\mu_2-2}}{\sqrt{\sigma_0 + \sigma_1}} \right)$$

$$- \frac{\mu_1^2\mu_2^2 x_0^{2\mu_2-2}}{\sigma_0 + \sigma_1} + \frac{1}{x_0^2} + \frac{\log x_0 - \theta}{x_0^2\alpha^2} - \frac{1}{x_0^2\alpha^2}$$

$$\sim -\frac{\sigma_1^2\sigma_2^2}{2(\sigma_0 + \sigma_1)^3} \cdot \log^2 y - \frac{\mu_1\mu_2(1-\mu_2)}{\sigma_0 + \sigma_1} \cdot \log y \cdot x_0^{\mu_2-2}$$

$$- \frac{\mu_1^2\mu_2^2}{\sigma_0 + \sigma_1} \cdot x_0^{2\mu_2-2} + \frac{1}{x_0^2} + \frac{\log x_0 - \theta}{x_0^2\alpha^2} - \frac{1}{x_0^2\alpha^2}.$$

We combine this result with equation (D.3.5), to get

$$h_y''(x_0) \sim -\frac{\mu_1\mu_2(1-\mu_2)}{\sigma_0 + \sigma_1} \cdot \log y \cdot x_0^{\mu_2-2} \sim -C \left(\log y\right)^{2+\frac{1}{1-\mu_2}}$$

with

$$C = \frac{\mu_1\mu_2(1-\mu_2)}{\sigma_0 + \sigma_1} \cdot \left( \frac{\sigma_1\sigma_2}{2\mu_1\mu_2(\sigma_0 + \sigma_1)} \right)^{1+\frac{1}{1-\mu_2}}.$$

We conclude that $h_y''(x_0) < 0$ and that indeed $x_0$ corresponds to a local maximum.

For $x = x_{2*}$ with $* = a, b$, we have

$$\mu'(x_{2*}) = \mu_1\mu_2 x_{2*}^{\mu_2-1},$$

$$\mu''(x_{2*}) = -\mu_1\mu_2(1-\mu_2)x_{2*}^{\mu_2-2},$$

$$\sigma(x_{2*}) \sim \sqrt{\sigma_0},$$

$$\sigma'(x_{2*}) \sim -\sigma_1\sigma_2/(2\sqrt{\sigma_0}) \cdot \exp(-\sigma_2 x_{2*}),$$

$$\sigma''(x_{2*}) \sim \sigma_1\sigma_2^2/(2\sqrt{\sigma_0}) \cdot \exp(-\sigma_2 x_{2*}), \text{ and}$$

$$p_y(x_{2b}) \sim \log y/\sqrt{\sigma_0}.$$

So,

$$h_y''(x_{2*}) \sim -\frac{\log^2 y}{\sigma_0} \cdot \left( \frac{3\sigma_1^2\sigma_2^2 \cdot \exp(-2\sigma_2 x_{2*})}{4\sigma_0^2} - \frac{\sigma_1\sigma_2^2 \cdot \exp(-\sigma_2 x_{2*})}{2\sigma_0} \right)$$
$$+ \frac{\log y}{\sqrt{\sigma_0}} \left( \frac{2\mu_1\mu_2 x_{2*}^{\mu_2-1} \cdot \sigma_1\sigma_2 \exp(-\sigma_2 x_{2*})}{\sigma_0^{3/2}} - \frac{\mu_1\mu_2(1-\mu_2)x_{2*}^{\mu_2-2}}{\sqrt{\sigma_0}} \right)$$
$$- \frac{\mu_1^2\mu_2^2 x_{2*}^{2\mu_2-2}}{\sigma_0} - \frac{k-1}{x_{2*}^2} - \frac{k(k-1)x_{2*}^{k-2}}{\lambda^k}.$$

For $x_{2*} = x_{2a}$, we simplify

$$h_y''(x_{2a}) \sim \frac{\sigma_1\sigma_2^2}{2\sigma_0^2} \cdot \log^2 y \cdot \exp(-\sigma_2 x_{2a})$$

which confirms that $x_{2a}$ corresponds to a local minimum. Finally, for $x_{2*} = x_{2b}$, we simplify

$$h_y''(x_{2b}) \sim -\frac{\mu_1\mu_2(1-\mu_2)x_{2b}^{\mu_2-2}}{\sigma_0} \cdot \log y - \frac{k(k-1)x_{2b}^{k-2}}{\lambda^k}$$
$$\sim -\left( \frac{\mu_1\mu_2(1-\mu_2)\left(\frac{\lambda^k\mu_1\mu_2}{k\sigma_0}\right)^{\frac{\mu_2-2}{k-\mu_2}}}{\sigma_0} + \frac{k(k-1)\left(\frac{\lambda^k\mu_1\mu_2}{k\sigma_0}\right)^{\frac{k-2}{k-\mu_2}}}{\lambda^k} \right)(\log y)^{\frac{k-2}{k-\mu_2}}$$
$$\sim -\left( \frac{\mu_1\mu_2(1-\mu_2)}{\sigma_0}\frac{k\sigma_0}{\lambda^k\mu_1\mu_2} + \frac{k(k-1)}{\lambda^k} \right) \cdot \left( \frac{\lambda^k\mu_1\mu_2}{k\sigma_0} \right)^{\frac{k-2}{k-\mu_2}}(\log y)^{\frac{k-2}{k-\mu_2}}$$
$$\sim -\frac{k}{\lambda^k}(k-\mu_2) \cdot \left( \frac{\lambda^k\mu_1\mu_2}{k\sigma_0} \right)^{\frac{k-2}{k-\mu_2}}(\log y)^{\frac{k-2}{k-\mu_2}}.$$

Finally, it is clear to see that $h_y''(x_{2b}) < 0$ which confirms that $x_{2b}$ is a local maximum.

### D.3.3  Calculating the survival function of $Y$

We will apply Proposition 4.2.2 to $g_y$ from equation (D.3.2), where we find that $k_0 = 2$ and $x_y^* = x_0(y)$. This gives us a lower bound for $\overline{F}_Y(y)$ as $y \to \infty$. We start with evaluating $g_y(x_0)$ and after, we check the smoothness assumption of the proposition for $k_0 = 2$. Finally, we derive an upper bound that is of the same order as the lower bound. Hence, we can combine the lower and upper bound to get an estimate for the rate of convergence to 0 of $\overline{F}_Y(y)$.

Before, we evaluate $g_y(x_0)$ and $h_y''(x_0)$, we first simplify $p_y(x_0)$ and $p_y(x_0)$ as $y \to \infty$. We have

$$
p_y(x_0) = \frac{\log y}{\sqrt{\sigma_0 + \sigma_1}} - \frac{\mu_0}{\sqrt{\sigma_0 + \sigma_1}} + O\left((\log y)^{-\frac{\mu_2}{1-\mu_2}}\right)
$$

and

$$
\frac{1}{p_y(x_0)} = \frac{\sqrt{\sigma_0 + \sigma_1}}{\log y} + O\left((\log y)^{-2-\frac{\mu_2}{1-\mu_2}}\right) = \frac{\sqrt{\sigma_0 + \sigma_1}}{\log y} + O\left((\log y)^{-2}\right).
$$

So,

$$
\begin{aligned}
g_y(x_0) &= \varphi\left(\frac{\log y - \mu(x_0)}{\sigma(x_0)}\right) \cdot \left(\frac{\sigma(x_0)}{\log y - \mu(x_0)} + O\left(\left(\frac{\sigma(x_0)}{\log y - \mu(x_0)}\right)^3\right)\right) f_X(x_0) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log y - \mu(x_0)}{\sigma(x_0)}\right)^2\right\} \\
&\quad \cdot \left(\frac{\sigma(x_0)}{\log y - \mu(x_0)} + O\left(\left(\frac{\sigma(x_0)}{\log y - \mu(x_0)}\right)^3\right)\right) \frac{\exp\left\{-\frac{(\log x_0 - \theta)^2}{2\alpha^2}\right\}}{\sqrt{2\pi} x_0 \alpha} \\
&= \frac{\sqrt{\sigma_0 + \sigma_1}}{2\pi\alpha} \exp\left\{-\frac{1}{2}\left(\frac{\log y}{\sqrt{\sigma_0 + \sigma_1}} - \frac{\mu_0}{\sqrt{\sigma_0 + \sigma_1}} + O\left((\log y)^{-\frac{\mu_2}{1-\mu_2}}\right)\right)^2\right\} \\
&\quad \cdot \frac{\frac{1}{\log y} + O\left((\log y)^{-2}\right)}{\left(\frac{\sigma_1\sigma_2}{2\mu_1\mu_2(\sigma_0+\sigma_1)}\right)^{-\frac{1}{1-\mu_2}} \cdot (\log y)^{-\frac{1}{1-\mu_2}} + O\left(\log^{-2}(y)\right)} \\
&\quad \cdot \exp\left\{-\frac{\left(\log\left[\left(\frac{\sigma_1\sigma_2}{2\mu_1\mu_2(\sigma_0+\sigma_1)}\right)^{-\frac{1}{1-\mu_2}} \cdot (\log y)^{-\frac{1}{1-\mu_2}} + O\left(\log^{-2}(y)\right)\right] - \theta\right)^2}{2\alpha^2}\right\} \\
&= \exp\left\{-\frac{1}{2(\sigma_0+\sigma_1)}\left(\log^2 y - 2\mu_0 \log y + O\left((\log y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right)\right\}.
\end{aligned}
$$

Next, we check the assumptions of Proposition 4.2.2.  First, we clearly have $h'_y(x_0)(-h''_y(x_0))^{-1/2} = 0$.  Secondly, we have one clearly dominating term in the second derivative of $h_y$ near $x_0$, so it is enough to show that

$$\lim_{y\to\infty} \frac{h''_y\left(x_0 + \frac{x}{\sqrt{-h''_y(x_0)}}\right)}{h''_y(x_0)}$$

$$= \lim_{x\to 0} \frac{p_y\left(x_0 + \frac{x}{\sqrt{-h''_y(x_0)}}\right)}{p_y(x_0)} \cdot \frac{\mu''\left(x_0 + \frac{x}{\sqrt{-h''_y(x_0)}}\right)}{\mu''(x_0)} \cdot \frac{\sigma(x_0)}{\sigma\left(x_0 + \frac{x}{\sqrt{-h''_y(x_0)}}\right)}$$

is equal to 1 for any fixed $x$. Since $(-h''_y(x_0))^{-1/2} \ll x_0$ as $y \to \infty$ and since $p_y$ and $\sigma$ are differentiable at $0 = \lim_{y\to\infty} x_0$, it is clear that the first and third term of the equation above tend to 1. Since $\mu''(0)$ does not exist, we would need to work out the term involving the second derivative of $\mu$ more carefully. We get

$$\frac{\mu''\left(x_0 + \frac{x}{\sqrt{-h''_y(x_0)}}\right)}{\mu''(x_0)} = \left(1 + \frac{x}{x_0\sqrt{-h''_y(x_0)}}\right)^{\mu_2-2}. \tag{D.3.10}$$

We note that $x_0\sqrt{-h''_y(x_0)}$ is asymptotically equal to a constant times $(\log y)^{\frac{1-2\mu_2}{2-2\mu_2}}$. Since $\mu_2 < 0.5$, the second term within the brackets in equation (D.3.10) tends to 0 when $y \to \infty$. This yields that the right hand side of equation (D.3.10) converges to 1 as $y \to \infty$. This is enough to show the smoothness assumption of the proposition. We get that for any fixed $\tilde{x} > 0$, there exists a constant $C_1(\tilde{x})$ such that

$$\int_0^\infty g_y(x)\,dx \geq \int_{x_0-\frac{\tilde{x}}{\sqrt{-h''_y(x_0)}}}^{x_0+\frac{\tilde{x}}{\sqrt{-h''_y(x_0)}}} g_y(x)\,dx$$

$$\geq C_1(\tilde{x})g_y(x_0) \cdot \frac{1}{\sqrt{-h''_y(x_0)}}$$

$$(\text{as } y \to \infty) = \exp\left\{-\frac{1}{2(\sigma_0+\sigma_1)}\left[\log^2 y - 2\mu_0\log y + O\left((\log y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right]\right\}. \tag{D.3.11}$$

Next, we evaluate $g_y(x_{2b})$ but first we work out

$$p_y(x_{2b}) = \frac{\log y}{\sqrt{\sigma_0}} + O\left((\log y)^{\frac{\mu_2}{k-\mu_2}}\right)$$

and

$$\frac{1}{p_y(x_0)} = \frac{\sqrt{\sigma_0}}{\log y} + O\left((\log y)^{-2+\frac{\mu_2}{k-\mu_2}}\right).$$

So,

$$
\begin{aligned}
g_y(x_{2b}) &= \varphi\left(\frac{\log y - \mu(x_{2b})}{\sigma(x_{2b})}\right)\left(\frac{\sigma(x_{2b})}{\log y - \mu(x_{2b})} + O\left(\frac{\sigma(x_{2b})^3}{(\log y - \mu(x_{2b}))^3}\right)\right) f_X(x_{2b}) \\
&= \varphi\left(\frac{\log y}{\sqrt{\sigma_0}} + O\left((\log y)^{\frac{\mu_2}{k-\mu_2}}\right)\right) \cdot \frac{\sqrt{\sigma_0}}{\log y} \cdot \left(1 + O\left((\log y)^{-1+\frac{\mu_2}{k-\mu_2}}\right)\right) \\
&\quad \cdot \frac{k}{\lambda^k}\left(\frac{\lambda^k \mu_1 \mu_2}{k\sigma_0}\right)^{\frac{k-1}{k-\mu_2}} (\log y)^{\frac{k-1}{k-\mu_2}}\left(1 + O\left((\log y)^{-\frac{k-2\mu_2}{k-\mu_2}}\right)\right) \\
&\quad \cdot \exp\left\{-\frac{\left(\frac{\sigma_1 \lambda^k \mu_1 \mu_2}{k\sigma_0^2}\right)^{\frac{k}{k-\mu_2}} (\log y)^{\frac{k}{k-\mu_2}}}{\lambda^k}\left(1 + O((\log y)^{-\frac{k-2\mu_2}{k-\mu_2}})\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_0}\log^2(y) + O\left((\log y)^{\frac{k}{k-\mu_2}}\right)\right\}.
\end{aligned}
$$

In particular, we find that $g_y(x_0) > g_y(x_{2b})$ for $y$ large enough. We have now all tools available to find an upperbound that gives the result directly,

$$
g_y(x) \le \tilde{g}_y(x) := \begin{cases} \max\{g_y(x) : \ x \in [0, x_2]\} & \text{for } 0 \le x \le x_2, \\ f_X(x) & \text{for } x > x_2. \end{cases}
$$

Since $g_y(x_{2b}) \le g_y(x_0)$ for $y$ large enough, we have derived that the maximum over the interval $[0, x_2]$ is attained at $x_0$. We here note that we do not need to show that $x_3$ and $x_4$ cannot exist as per definition, as they would clearly need to be larger than $x_2$ if they exist. So, as $y \to \infty$

$$
\begin{aligned}
\int_0^\infty g_y(x)\,\mathrm{d}x &\le \int_0^{x_{2b}} g_y(x_0)\,\mathrm{d}x + \int_{x_{2b}}^\infty f_X(x)\,\mathrm{d}x = x_{2b}g_y(x_0) + \overline{F}_X(x_{2b}) \\
&= \left(\frac{\lambda^k \mu_1 \mu_2}{k\sigma_0}\right)^{\frac{1}{k-\mu_2}}\left(1 + O\left((\log y)^{-\frac{k-2\mu_2}{k-\mu_2}}\right)\right)(\log y)^{\frac{1}{k-\mu_2}} \\
&\quad \cdot \exp\left\{-\frac{1}{2(\sigma_0 + \sigma_1)}\left[\log^2 y - 2\mu_0 \log y + O\left((\log y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right]\right\} \\
&\quad + \exp\left\{-\lambda^{-k}\left(\frac{\lambda^k \mu_1 \mu_2}{k\sigma_0}\right)^{\frac{k}{k-\mu_2}}(\log y)^{\frac{k}{k-\mu_2}}\left[1 + O\left((\log y)^{-\frac{k-2\mu_2}{k-\mu_2}}\right)\right]\right\} \\
&= \exp\left\{-\frac{1}{2(\sigma_0 + \sigma_1)}\left[\log^2 y - 2\mu_0 \log y + O\left((\log y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right]\right\}.
\end{aligned}
$$

(D.3.12)

(D.3.13)

Now, combining equation (D.3.11) and equation (D.3.13), yields as $y \to \infty$

$$\mathbb{P}(Y > y) = \int_0^\infty g_y(x)\,\mathrm{d}x = \exp\left\{-\frac{1}{2(\sigma_0 + \sigma_1)}\left[\log^2 y - 2\mu_0 \log y + O\left((\log y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right]\right\}.$$

## D.3.4   Calculating $\eta$

We use the previous work to transform $Y$ to $Y_E$ on standard exponential margins. Thus

$$Y_E = F_E^{-1}(F_Y(Y)) = -\log(1 - F_Y(Y))$$

$$= -\log\left(\exp\left\{-\frac{1}{2(\sigma_0 + \sigma_1)}\left[\log^2 Y - 2\mu_0 \log Y + O\left((\log Y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right]\right\}\right)$$

$$= \frac{1}{2(\sigma_0 + \sigma_1)}\left(\log^2 Y - 2\mu_0 \log Y + O\left((\log Y)^{\frac{1-2\mu_2}{1-\mu_2}}\right)\right).$$

So, the function $T$ that transforms $\log Y$ to $Y_E$ is given by

$$T(y) = \frac{y^2}{2(\sigma_0 + \sigma_1)} - \frac{\mu_0 y}{\sigma_0 + \sigma_1} + O\left(y^{\frac{1-2\mu_2}{1-\mu_2}}\right),$$

as $y \to \infty$. In calculating the extremal dependence measures, we need to solve $T(y) = u$ for large $y$. We get

$$T^{-1}(u) = \sqrt{2(\sigma_0 + \sigma_1)u} + O(1)$$

as $u \to \infty$. We write down a formula for $\chi = \lim_{u\to\infty} \mathbb{P}(Y_E > u \mid (X/\lambda)^k > u)$ as $u \to \infty$

$$\mathbb{P}(Y_E > u \mid (X/\lambda)^k > u) = e^u \int_{\lambda u^{1/k}}^\infty \mathbb{P}(T(\log Y) > u \mid X = x) f_X(x)\,\mathrm{d}x$$

$$= e^u \int_{\lambda u^{1/k}}^\infty \mathbb{P}(\log Y > T^{-1}(u) \mid X = x) f_X(x)\,\mathrm{d}x$$

$$= e^u \int_{\lambda u^{1/k}}^\infty \overline{\Phi}\left(\frac{T^{-1}(u) - \mu(x)}{\sigma(x)} \mid X = x\right) \cdot \frac{kx^{k-1}}{\lambda^k} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}\,\mathrm{d}x.$$

In particular, we have for $I = [\lambda u^{1/k}, \lambda(2 + \sigma_1/\sigma_0)^{1/k}]$

$$\mathbb{P}(Y_E > u \mid (X/\lambda)^k > u) \tag{D.3.14}$$

$$> e^u \int_I \overline{\Phi}\left(\frac{T^{-1}(u) - \mu(x)}{\sigma(x)} \mid X = x\right) \cdot \frac{kx^{k-1}}{\lambda^k} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}\,\mathrm{d}x. \tag{D.3.15}$$

For ease of presentation, we define $p_u(x) = [T^{-1}(u) - \mu(x)]/\sigma(x)$. Similar to the previous section, we define $g_u$ as the integrand and $h_u := \log g_u$ as the log of the integrand, both are specified only on the integration domain $I$. For $x$ in the integration domain, we have

$$h_u(x) := \log \left( \overline{\Phi}(p_u(x)) f_X(x) \right).$$

We apply Proposition 4.2.2 to bound integral (D.3.14) from below. In particular, we first need to find the mode of $h_u$ over the integration domain. Let $x_u$ be a sequence such that for each $u$, $x_u$ lies in the integration domain. So, then we can write $x = C_u u^{1/k} + o(u^{1/k})$ for some bounded set of constants $C_u \in [\lambda, \lambda(2 + \sigma_1/\sigma_0)]$. We have

$$h_u'(x) = -\frac{\varphi(p_u(x))}{\overline{\Phi}(p_u(x))} \cdot p_u'(x) - \frac{k-1}{x} - \frac{kx^{k-1}}{\lambda^k}$$

$$= \frac{\varphi(p_u(x))}{\overline{\Phi}(p_u(x))} \cdot \left( p_u(x) \cdot \frac{\sigma'(x)}{\sigma(x)} + \frac{\mu'(x)}{\sigma(x)} \right) - \frac{k-1}{x} - \frac{kx^{k-1}}{\lambda^k}.$$

Since, $p_u(x) \sim \sqrt{2(1 + \sigma_1/\sigma_0)u} \to \infty$ as $u \to \infty$, we simplify

$$h_u'(x) \sim \sqrt{2 \left( 1 + \frac{\sigma_1}{\sigma_0} \right) u} \cdot \left( \sqrt{2 \left( 1 + \frac{\sigma_1}{\sigma_0} \right) u} \cdot \frac{-\sigma_1 \sigma_2 e^{-\sigma_2 \left( C_u u^{1/k} + o(u^{1/k}) \right)}}{2\sigma_0} \right.$$

$$\left. + \frac{\mu_1 \mu_2 \left( C_u u^{1/k} + o(u^{1/k}) \right)^{\mu_2 - 1}}{\sqrt{\sigma_0}} \right)$$

$$- \frac{k-1}{C_u u^{1/k} + o(u^{1/k})} - \frac{k \left( C_u u^{1/k} + o(u^{1/k}) \right)^{k-1}}{\lambda^k}.$$

$$\sim -\frac{kC_u^{k-1} u^{1-1/k}}{\lambda^k}.$$

In particular, we derive that $h_u'(x) < 0$ as $u \to \infty$. So, the maximum of $h_u$ over the integration domain must be attained at the boundary and hence is given by $x_0 = \lambda u^{1/k}$. In particular, we get $h_u'(x_0) \sim -ku^{1-1/k}/\lambda$. We now will show that we can apply Proposition 4.2.2 with $k_0 = 1$. We have, as $u \to \infty$,

$$h_u(\lambda u^{1/k}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} p_u(\lambda u^{1/k})^2 - \log p_u(\lambda u^{1/k}) + \log f_X(\lambda u^{1/k})$$

$$= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \left( \frac{T^{-1}(u) - \mu(\lambda u^{1/k})}{\sigma(\lambda u^{1/k})} \right)^2$$

$$- \log \left( \frac{T^{-1}(u) - \mu(\lambda u^{1/k})}{\sigma(\lambda u^{1/k})} \right) + \log \left( \frac{ku^{(k-1)/k}}{\lambda} \right) - u$$

$$= - \left( 2 + \frac{\sigma_1}{\sigma_0} \right) u + O \left( u^{1/2 + \mu_2/k} \right).$$

Next, we check the smoothness assumption of Proposition 4.2.2 with $k_0 = 1$. Let $\delta > 0$ and $0 \leq x \leq \delta$. It is now enough to show that the limit of $u$ to infinity of the following expression tends to 1. We have

$$\lim_{u \to \infty} \frac{h'_u\left(\lambda u^{1/k} + \frac{x}{-h'_u(\lambda u^{1/k})}\right)}{h'_u(\lambda u^{1/k})} = \lim_{u \to \infty} \frac{\left(\lambda u^{1/k} + \frac{\lambda x}{k u^{1-1/k}}\right)^{k-1}}{u^{(k-1)/k}} = \lim_{u \to \infty} \left(\lambda + \frac{\lambda x}{k}\right)^{k-1} = 1.$$

This is enough to show the smoothness assumption of Proposition 4.2.2 with $k_0 = 1$. We conclude that for each $\tilde{x}$, there exists a constant $C_1(\tilde{x})$ such that

$$\int_{\lambda u^{1/k}}^{\infty} g_u(x)\,\mathrm{d}x \geq \int_{\lambda u^{1/k}}^{\lambda u^{1/k} + \frac{\tilde{x}}{-h'_u(x_0)}} g_u(x)\,\mathrm{d}x \geq C_1(\tilde{x})g_u(\lambda u^{1/k}) \cdot \frac{1}{-h'_u(\lambda u^{1/k})}$$

$$\overset{(\text{as } u \to \infty)}{=} e^{-\left(2+\frac{\sigma_1}{\sigma_0}\right)u + O\left(u^{\frac{1}{2}+\frac{\mu_2}{k}}\right)}. \tag{D.3.16}$$

To get an upper bound, we use the following crude upper bound $\tilde{g}_u$ for $g_u$,

$$g_u(x) \leq \tilde{g}_u(x) := \begin{cases} g_u(\lambda u^{1/k}) & \text{for } \lambda u^{1/k} \leq x \leq \lambda\left(2 + \frac{\sigma_1}{\sigma_0}\right)^{1/k} u^{1/k}, \\ f_X(x) & \text{for } x > \lambda\left(2 + \frac{\sigma_1}{\sigma_0}\right)^{1/k} u^{1/k}. \end{cases}$$

We get as $u \to \infty$,

$$\int_{\lambda u^{1/k}}^{\infty} g_u(x)\,\mathrm{d}x \leq \left(\lambda\left(2 + \frac{\sigma_1}{\sigma_0}\right)^{1/k} u^{1/k} - \lambda u^{1/k}\right) g_u(\lambda u^{1/k}) \tag{D.3.17}$$

$$+ \overline{F}_X\left(\lambda\left(2 + \frac{\sigma_1}{\sigma_0}\right)^{1/k} u^{1/k}\right)$$

$$= \exp\left(-\left(2 + \frac{\sigma_1}{\sigma_0}\right)u + O\left(u^{1/2+\mu_2/k}\right)\right) + \exp\left(-\left(2 + \frac{\sigma_1}{\sigma_0}\right)u\right)$$

$$= \exp\left(-\left(2 + \frac{\sigma_1}{\sigma_0}\right)u + O\left(u^{1/2+\mu_2/k}\right)\right). \tag{D.3.18}$$

Combining equations (D.3.16) and (D.3.18), we get

$$\mathbb{P}(Y_E > u \mid (X/\lambda)^k > u) = \int_{\lambda u^{1/k}}^{\infty} g_u(x)\,\mathrm{d}x = \exp\left(-\left(2 + \frac{\sigma_1}{\sigma_0}\right)u + O\left(u^{1/2+\mu_2/k}\right)\right)$$

as $u \to \infty$. From this expression, it is straightforward to see that $\xi = 0$ and

$$\eta^{-1} = 2 + \frac{\sigma_1}{\sigma_0}.$$

# D.4   Details on Calculations for the Exact HT model

## D.4.1   Introduction

Assume model (4.4.3) for random vector $(X, Y)$ with $\overline{H}$ as in equation (4.4.4). We recall that $(X, Y)$ is a random vector such that $X$ and $Y$ both have standard Laplace margins. Moreover, there exist $0 \leq \alpha \leq 1$, $\beta < 1$ and $u > 0$ such that for $x > u$

$$\mathbb{P}(Y > y \mid X = x) = \overline{H}\left(\frac{y - \alpha x}{x^\beta}\right),$$

holds for all $y \in \mathbb{R}$ with

$$\overline{H}(z) = \exp(-\gamma z^\delta)\mathbb{1}\{z > 0\} + \mathbb{1}\{z \leq 0\}$$

for $\gamma > 0$ and $\delta \geq (1 - \beta)^{-1}$. In this section, we work out the value for $\eta$ when $0 < \alpha < 1$, $\beta > 0$ and $\delta = (1 - \beta)^{-1}$. The other cases are significantly easier to work out and the results of these cases are stated in the main paper.

## D.4.2   Calculating $\eta$

We write

$$
\begin{aligned}
\mathbb{P}(Y > u, X > u) &= \int_u^\infty \overline{H}\left(\frac{u - \alpha x}{x^\beta}\right) f_X(x)\, \mathrm{d}x \\
&= \frac{1}{2}\int_u^{u/\alpha} \exp\left(-\gamma\left(\frac{u - \alpha x}{x^\beta}\right)^\delta - x\right) \mathrm{d}x + \frac{1}{2}\int_{u/\alpha}^\infty \exp(-x)\, \mathrm{d}x \\
&= \frac{1}{2}\int_u^{u/\alpha} \exp\left(-\gamma\left(\frac{u - \alpha x}{x^\beta}\right)^\delta - x\right) \mathrm{d}x + \frac{1}{2}\exp\left(-\frac{u}{\alpha}\right).
\end{aligned}
$$

In general, we cannot evaluate the first integral in closed form for finite $u$. However, we can bound it from below using Proposition 4.2.2. A bound from above can again be found directly. We define the integration domain $I = [u, u/\alpha]$,

$$g_u(x) := \exp\left(-\gamma\left(\frac{u - \alpha x}{x^\beta}\right)^\delta - x\right)$$

for $x \in I$ and $h_u := \log g_u$ on $I$. We now need to determine whether or not the mode $x_0 := x_0(u)$ of the integrand $g_u$ over the integration domain $I$ lies on the boundary of

$I$ or in the interior of $I$. We assume that $x_0$ lies in the interior of $I$, then we have

$$0 = h'_u(x_0) = \gamma\delta \left(\frac{u - \alpha x_0}{x_0^\beta}\right)^{\delta-1} \cdot \left(\frac{\alpha}{x_0^\beta} + \frac{(u - \alpha x_0)\beta}{x_0^{\beta+1}}\right) - 1$$

$$= \gamma\delta\beta(u - \alpha x_0)^\delta x_0^{-\beta\delta-1} + \gamma\delta\alpha(u - \alpha x_0)^{\delta-1}x_0^{-\beta\delta} - 1$$

$$= \gamma\delta\beta(u - \alpha x_0)^\delta x_0^{-\delta} + \gamma\delta\alpha(u - \alpha x_0)^{\delta-1}x_0^{-\delta+1} - 1$$

and we derive that

$$\beta(u - \alpha x_0)^\delta + \alpha(u - \alpha x_0)^{\delta-1}x_0 = \frac{1}{\gamma\delta}x_0^\delta. \tag{D.4.1}$$

Since, we work under the premise that $x_0 \in (u, u/\alpha)$, we are only interested in finding solutions that satisfy $x_0 = \tilde{c}u + o(u)$ as $u \to \infty$ for some $\tilde{c} \in [1, 1/\alpha]$, otherwise the mode of $h_u$ is found at the boundary of the integration domain at $u$. We try $x_0 = cu$ with $c \in (0, \infty)$ in equation (D.4.1), and we derive that this is an exact solution if $c$ solves

$$0 = \gamma\delta\left(\beta(1 - \alpha c)^\delta + \alpha c(1 - \alpha c)^{\delta-1}\right) - c^\delta = \gamma(1 - \alpha c)^{\delta-1}(\delta - 1 + \alpha c) - c^\delta. \tag{D.4.2}$$

Since the right hand side is a continuous function of $c$ for $c \in [0, 1/\alpha]$, we show by the intermediate value theorem that $c \in (0, 1/\alpha)$ by inserting $c = 0$ and $c = 1/\alpha$ and comparing signs of the right hand side of equation (D.4.2). Indeed, for $c = 0$, we have that

$$\gamma(1 - \alpha c)^{\delta-1}(\delta - 1 + \alpha c) - c^\delta = \gamma(\delta - 1) > 0$$

and for $c = 1/\alpha$, we have that

$$\gamma(1 - \alpha c)^{\delta-1}(\delta - 1 + \alpha c) - c^\delta = -\alpha^{-\delta} < 0.$$

We recall that we are only interested in the value for $c$ if $c \in (1, 1/\alpha)$. Hence, let $c = 1$ in the right hand side of equation (D.4.2) to give

$$\gamma\delta\left(\beta(1 - \alpha c)^\delta + \alpha c(1 - \alpha c)^{\delta-1}\right) - c^\delta = \gamma(1 - \alpha)^{\delta-1}\left((\delta - 1)(1 - \alpha) + \delta\alpha\right) - 1$$

$$= \gamma(1 - \alpha)^{\delta-1}(\delta - 1 + \alpha) - 1$$

which is negative if and only if $\gamma(1 - \alpha)^{\delta-1}(\delta - 1 + \alpha) < 1$. We conclude that $c \in (0, 1)$ if and only if $\gamma(1 - \alpha)^{\delta-1}(\delta - 1 + \alpha) < 1$ and $c \in [1, 1/\alpha)$ if and only

if $\gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha) \geq 1$. We term these cases as Case (2a) and Case (2b), respectively. In Case (2b), $x_0$ lies in the interior of the integration domain $I$ for large enough $u$, and in Case (2a), the mode over the integration domain $I$ is found at $u$ on the boundary.

We work out $g_u(x_0)$ for both Case (2a) and (2b),

$$
\begin{aligned}
g_u(x_0) &= \exp\left\{-\gamma\left(\frac{u-\alpha x_0}{x_0^\beta}\right)^\delta - x_0\right\} \\
&= \exp\left\{-\gamma\left(\frac{u-\alpha(cu+o(u))}{(cu+o(u))^\beta}\right)^\delta - cu+o(u)\right\} \\
&= \exp\left\{-\gamma u^{\delta-\beta\delta}\frac{(1-\alpha c+o(1))^\delta}{c^{\beta\delta}+o(1))} - cu+o(u)\right\} \\
&= \exp\left\{-\left(\frac{\gamma(1-\alpha c)^\delta}{c^{\beta\delta}}+c\right)u+o(u)\right\}.
\end{aligned}
$$

Next, we work out $h_u'(x_0)$ in Case (2a)

$$
\begin{aligned}
h_u'(x_0) &= \gamma\delta\beta(u-\alpha u)^\delta u^{-\beta\delta-1}+\gamma\delta\alpha(u-\alpha u)^{\delta-1}u^{-\beta\delta}-1 \\
&= \gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha)-1.
\end{aligned}
$$

By definition of Case (2a), we have that $h_u'(x_0) < 0$. Let $C > 0$ and $|x| \leq C$, then as $u \to \infty$

$$
\begin{aligned}
h_u'\left(x_0-\frac{x}{h_u'(x_0)}\right) &= \gamma\delta\beta\left(u-\alpha\left(x_0+\frac{x}{-h_u'(x_0)}\right)\right)^\delta\left(x_0+\frac{x}{-h_u'(x_0)}\right)^{-\beta\delta-1} \\
&\quad +\gamma\delta\alpha\left(u-\alpha\left(x_0+\frac{x}{-h_u'(x_0)}\right)\right)^{\delta-1}\left(x_0+\frac{x}{-h_u'(x_0)}\right)^{-\beta\delta}-1 \\
&= \gamma\delta\beta\left(u-\alpha u-\frac{x\alpha}{1-\gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha)}\right)^\delta \\
&\quad \cdot\left(u+\frac{x}{1-\gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha)}\right)^{-\beta\delta-1} \\
&\quad +\gamma\delta\alpha\left(u-\alpha u-\frac{x\alpha}{1-\gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha)}\right)^{\delta-1} \\
&\quad \cdot\left(u+\frac{x}{1-\gamma(1-\alpha)^{\delta-1}(\delta-1+\alpha)}\right)^{-\beta\delta}-1 \\
&= \gamma\delta\beta\left(u^\delta(1-\alpha)^\delta+O\left(u^{\delta-1}\right)\right)\left(u^{-\beta\delta-1}+O\left(u^{-\beta\delta-2}\right)\right) \\
&\quad +\gamma\delta\alpha\left(u^{\delta-1}(1-\alpha)^{\delta-1}+O\left(u^{\delta-2}\right)\right)\left(u^{-\beta\delta}+O\left(u^{-\beta\delta-1}\right)\right)-1 \\
&= h_u'(x_0)+O\left(u^{\delta-2-\beta\delta}\right).
\end{aligned}
$$

So,

$$\lim_{u \to \infty} \frac{h'_u \left( x_0 + \frac{x}{-h'_u(x_0)} \right)}{h'_u(x_0)} = 1,$$

which is enough to show the smoothness assumption of Proposition 4.2.2 with $k_0 = 1$.

We get that for any fixed $\tilde{x} > 0$ there exist a $C_1(\tilde{x}) > 0$ such that

$$
\begin{aligned}
\mathbb{P}(X > u,\ Y > u) &= \int_u^{u/\alpha} g_u(x)\, \mathrm{d}x + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&\geq \frac{1}{2} \int_{x_0 - \frac{\tilde{x}}{-h'_u(x_0)}}^{x_0 + \frac{\tilde{x}}{-h'_u(x_0)}} g_u(x)\, \mathrm{d}x + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&\geq \frac{1}{2} C_1(\tilde{x}) g_u(x_0) \cdot \frac{1}{-h'_u(x_0)} + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&\geq \frac{1}{2} C_1(\tilde{x}) \exp \left\{ - \left( \gamma(1-\alpha)^\delta + 1 \right) u + o(u) \right\} \\
&\qquad \cdot \frac{1}{1 - \gamma(1-\alpha)^{\delta-1} (\delta - 1 + \alpha)} + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&= \exp \left\{ - \left( \gamma(1-\alpha)^\delta + 1 \right) u + o(u) \right\}.
\end{aligned}
$$

In the last step we used that $\gamma(1-\alpha)^\delta + 1 < 1/\alpha$ holds, which can be directly derived from the assumptions corresponding to Case (2b). Similarly to before, we can find an upper bound rather straightforwardly using the following upperbound for $g_u(x)$

$$
g_u(x) \leq \tilde{g}_u(x) := \begin{cases} g_u(x_0) & \text{for } u \leq x \leq u/\alpha, \\ f_X(x) & \text{for } x > u/\alpha. \end{cases}
$$

So,

$$
\begin{aligned}
\mathbb{P}(X > u,\ Y > u) &= \int_u^{u/\alpha} g_u(x)\, \mathrm{d}x + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&\leq u \left( \frac{1}{\alpha} - 1 \right) g_u(x_0) + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&= u \left( \frac{1}{\alpha} - 1 \right) \exp \left\{ - \left( \gamma(1-\alpha)^\delta + 1 \right) u + o(u) \right\} + \frac{1}{2} \exp \left\{ -\frac{u}{\alpha} \right\} \\
&= \exp \left\{ - \left( \gamma(1-\alpha)^\delta + 1 \right) u + o(u) \right\}.
\end{aligned}
$$

We conclude that

$$\mathbb{P}(X > u,\ Y > u) = \exp \left\{ - \left( \gamma(1-\alpha)^\delta + 1 \right) u + o(u) \right\},$$

$\chi = 0$ and

$$\eta = \left(\gamma(1-\alpha)^\delta + 1\right)^{-1}.$$

For Case (2a), we work out $h_u''(x_0)$ as $u \to \infty$

$$h_u''(x_0) = -\alpha^2\gamma\delta(\delta-1)(u-\alpha x_0)^{\delta-2}x_0^{-\beta\delta} - 2\alpha\beta\gamma\delta^2(u-\alpha x_0)^{\delta-1}x_0^{-\beta\delta-1}$$

$$- \beta\delta(\beta\delta+1)\gamma(u-\alpha x_0)^\delta x_0^{-\beta\delta-2}$$

$$= -\alpha^2\gamma\delta(\delta-1)(u-\alpha(cu+o(u)))^{\delta-2}(cu+o(u))^{-\beta\delta}$$

$$- 2\alpha\beta\gamma\delta^2(u-\alpha(cu+o(u)))^{\delta-1}(cu+o(u))^{-\beta\delta-1}$$

$$- \beta\delta(\beta\delta+1)\gamma(u-\alpha(cu+o(u)))^\delta(cu+o(u))^{-\beta\delta-2}$$

$$= -\left[\alpha^2\gamma\delta(\delta-1)(1-\alpha c)^{\delta-2}c^{-\beta\delta} + 2\alpha\beta\gamma\delta^2(1-\alpha c)^{\delta-1}c^{-\beta\delta-1}\right.$$

$$\left. + \beta\delta(\beta\delta+1)\gamma(1-\alpha c)^\delta c^{-\beta\delta-2}\right]u^{\delta-2-\beta\delta} + o\left(u^{\delta-2-\beta\delta}\right)$$

$$= -\beta\delta^2\gamma c^{-\beta\delta-2}(1-\alpha c)^{\delta-2}\left(\alpha^2 c^2 + 2\alpha(1-\alpha c)c + (1-\alpha c)^2\right)u^{-1} + o\left(u^{-1}\right)$$

$$= -\beta\delta^2\gamma c^{-\beta\delta-2}(1-\alpha c)^{\delta-2}u^{-1} + o\left(u^{-1}\right)$$

$$= -\delta(\delta-1)\gamma c^{-\delta-1}(1-\alpha c)^{\delta-2}u^{-1} + o\left(u^{-1}\right).$$

Now, let $C > 0$ and $|x| \le C$, then we have $x_0 + x(-h_u''(x_0))^{-1/2} = cu + o(u)$. So,

$$h_u''\left(x_0 + \frac{x}{\sqrt{-h_u''(x_0)}}\right) = h_u''\left(cu + o(u)\right).$$

So,

$$\lim_{u\to\infty}\frac{h_u''\left(x_0 + \frac{x}{\sqrt{-h_u''(x_0)}}\right)}{h_u''(x_0)} = \lim_{u\to\infty}\frac{-\delta(\delta-1)\gamma c^{-\delta-1}(1-\alpha c)^{\delta-2}u^{-1}(1+o(1))}{-\delta(\delta-1)\gamma c^{-\delta-1}(1-\alpha c)^{\delta-2}u^{-1}(1+o(1))} = 1,$$

which is enough to show the smoothness assumption of Proposition 4.2.2 with $k_0 = 1$.
We get that for any fixed $\tilde{x} > 0$ there exist a $C_1(\tilde{x}) > 0$ such that as $u \to \infty$

$$\mathbb{P}(X > u, Y > u) = \int_u^{u/\alpha} g_u(x)\,\mathrm{d}x + \frac{1}{2}\exp\left\{-\frac{u}{\alpha}\right\}$$

$$\ge \frac{1}{2}\int_{x_0-\frac{\tilde{x}}{-h_u'(x_0)}}^{x_0+\frac{\tilde{x}}{-h_u'(x_0)}} g_u(x)\,\mathrm{d}x$$

$$\ge \frac{1}{2}C_1(\tilde{x})g_u(x_0)\cdot\frac{1}{-h_u'(x_0)}$$

$$= \exp\left\{-\left(\frac{\gamma(1-\alpha c)^\delta}{c^{\delta-1}} + c\right)u + o(u)\right\}.$$

Similarly to before, we can find an upper bound rather straightforwardly,

$$
\begin{aligned}
\mathbb{P}(X > u,\, Y > u) &= \int_u^{u/\alpha} g_u(x)\,\mathrm{d}x + \frac{1}{2}\exp\left\{-\frac{u}{\alpha}\right\} \\
&\leq u\left(\frac{1}{\alpha} - 1\right) g_u(x_0) + \frac{1}{2}\exp\left\{-\frac{u}{\alpha}\right\} \\
&= u\left(\frac{1}{\alpha} - 1\right)\exp\left\{-\left(\frac{\gamma(1-\alpha c)^\delta}{c^{\beta\delta}} + c\right)u + o(u)\right\} + \frac{1}{2}\exp\left\{-\frac{u}{\alpha}\right\} \\
&= \exp\left\{-\left(\frac{\gamma(1-\alpha c)^\delta}{c^{\delta-1}} + c\right)u + o(u)\right\}.
\end{aligned}
$$

So,

$$
\mathbb{P}(X > u,\, Y > u) = \exp\left\{-\left(\frac{\gamma(1-\alpha c)^\delta}{c^{\beta\delta}} + c\right)u + o(u)\right\},
$$

and we conclude that $\chi = 0$ and

$$
\eta = \left(\frac{\gamma(1-\alpha c)^\delta}{c^{\delta-1}} + c\right)^{-1}.
$$

# D.5  Supplementary examples

Here, we state without proof a list of results that I derived during the Ph.D. using Proposition 4.2.2.

**Result 1:** Let $X_1$ be a random variable with a standard Laplace distribution, and let $Z_1$ be a random variable with a standard normal distribution. If we now define $Y_1 := |Z_1|^{|X_1|}$, then

$$
\log \mathbb{P}(Y_1 > y) \sim -\frac{2\log y}{\log^2(\log y)}
$$

as $y \to \infty$. Moreover, $\chi(X_1, Y_1) = 0$ and $\eta(X_1, Y_1) = 1$.

**Result 2:** Let $X_2$ be a random variable with a standard Laplace distribution, and let $Z_2$ be a random variable with a standard normal distribution. If we now define $Y_2 := X_2^b Z_2$ for some $b > 0$, then

$$
\log \mathbb{P}(Y_2 > y) \sim -\frac{1}{2} b^{-\frac{2b}{2b+1}} (1 + 2b)\, y^2 2b + 1.
$$

If we apply the Heffernan-Tawn model to $Y_2$ on Laplace margins conditional on $X_2$, we obtain $\alpha = 0$ and $\beta = \frac{2b}{2b+1}$.

**Result 3:** Let $X_3$ be a random variable with a standard Laplace distribution, and let $Z_3$ be a random variable with a standard normal distribution. If we now define $Y_3 := e^{X_3} Z_3$, then

$$\log \mathbb{P}(Y > y) \sim -\log y$$

If we apply the Heffernan-Tawn model to $Y_3$ on Laplace margins conditional on $X_3$, we obtain $\alpha = 1$ and $\beta = 0$.

# Appendix E

# Appendix to Chapter 5

## E.1   Reparameterization of EVAR

As opposed to inference for vector autoregression models, we cannot estimate the EVAR parameters by least squares due to the presence $Y_{t,1}^{\mathbf{B}}$ term. Instead, we apply the inference methodology as discussed in Section 5.3.5. Not surprisingly, the parameter estimates $\hat{\Phi}^{(i)}$ for $i = 1, \ldots, k$ are highly intercorrelated because of the linear dependence between the components of $\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-k}$. Reparameterization to reduce the correlation between parameter estimators is therefore attractive.

To reparameterize the model, we proceed as follows. First, we assume that the conditional extremes model is applicable to $Y_{t-i,j}$ conditional on $Y_{t-k,1}$ for each $i = 0, \ldots, k$ and $j = 1, \ldots, d$ apart from $(i,j) = (k,1)$, i.e., there exist parameters $\alpha_{i,j} \in [-1,1]$ and $\beta_{i,j} < 1$ such that

$$\lim_{y \to \infty} \mathbb{P}\left( \frac{Y_{t-i,j} - \alpha_{i,j}y}{y^{\beta_{i,j}}} \le x \;\middle|\; Y_{t-k,1} = y \right) = H_{i,j}(x)$$

where $H_{i,j}$ is a non-degenerate distribution function. Following the EVAR model (5.3.4), we now must have

$$Y_{t+k,1} = \Phi_{1,1}^{(1)}Y_{t+k-1,1} + \cdots + \Phi_{d,1}^{(1)}Y_{t+k-1,d} + \cdots + \Phi_{1,1}^{(k)}Y_{t,1} + \cdots + \Phi_{d,1}^{(k)}Y_{t,d} + Y_{t,1}^{B_1}\varepsilon_{t,1}$$

$$= \left( \Phi_{1,1}^{(1)}\alpha_{k-1,1} + \cdots + \Phi_{d,1}^{(1)}\alpha_{k-1,d} + \cdots + \Phi_{1,1}^{(k)} + \cdots + \Phi_{d,1}^{(k)}\alpha_{0,d} \right) Y_{t,1} + o_p(Y_{t,1})$$

conditional on $Y_{t,1} > v$ as $v$ tends to infinity. On the other hand, we have $Y_{t+k,1}|(Y_{t,1} >$

$v) = \alpha_{k,1} Y_{t,1} + o_p(Y_{t,1})$. So,

$$\alpha_{k,1} = \Phi_{1,1}^{(1)} \alpha_{k-1,1} + \cdots + \Phi_{d,1}^{(1)} \alpha_{k-1,d} + \cdots + \Phi_{1,1}^{(k)} \cdot 1 + \cdots + \Phi_{d,1}^{(k)} \alpha_{0,d},$$

which explains the collinearity of the estimators. We now propose the following reparameterization $(\mathbf{B}, \tilde{\Phi}^{(1)}, \ldots, \tilde{\Phi}^{(k)})$. For each $1 \leq l \leq d$, we acquire $\tilde{\Phi}_{j,l}^{(k-i)}$, i.e., the $(j,l)$th element of $\tilde{\Phi}^{(k-i)}$, inductively with $0 \leq i \leq k-1$, $1 \leq j \leq d$.

$$\Phi_{j,l}^{(k-i)} = \begin{cases} \hat{\alpha}_{k,l} + \tilde{\Phi}_{1,l}^{(k)}, & \text{for } i = 0, \ j = 1 \\ -\tilde{\Phi}_{j-1,l}^{(k-i)} \ \hat{\alpha}_{i,j-1}/\hat{\alpha}_{i,j} + \tilde{\Phi}_{j,l}^{(k-i)}, & \text{for } i = 0, \ldots, k-1, \ j = 2, \ldots, d, \ \text{cond. on } \tilde{\Phi}_{1,l}^{(k-i)} \\ -\tilde{\Phi}_{d,l}^{(k-i+1)} \hat{\alpha}_{i-1,d}/\hat{\alpha}_{i,1} + \tilde{\Phi}_{1,l}^{(k-i)}, & \text{for } i = 1, \ldots, k-1, \ j = 1 \ \text{conditional on } \tilde{\Phi}_{d,l}^{(k-i+1)}. \end{cases}$$

where $\hat{\alpha}_{i,j}$ is the maximum likelihood estimate for $\alpha_{i,j}$. Under this reparametrization, estimators of $\tilde{\Phi}_{j,k}^{(i)}$ are less correlated, which is validated in unreported experiments that compares the dependence of the original parameters and the reparameterized parameters using adaptive MCMC methodology (Roberts and Rosenthal, 2009).

# Appendix F

# Supplementary Information to Chapter 5

## F.1   Introduction

In this document, we provide the interested reader with supporting figures to our case study. In these figures, we plot the diagnostics of each of the 18 considered models and the baseline historical-matching method when applied to met-ocean data. We plot simulated excursions such that the excursion maximum significant wave height takes on values between 11.5 and 12.5 (left) and we visually compare these with observed excursions (middle). On the right panels, we show summaries of the simulated and observed trajectories by plotting the median, and the 10% and 90% percentiles of the set of excursions. Finally, in the bottom panel we plot the probability that an excursion has not ended as function of hours relative to the peak conditional on the excursion maximum significant wave height taking on a value between 11.5 and 12.5.
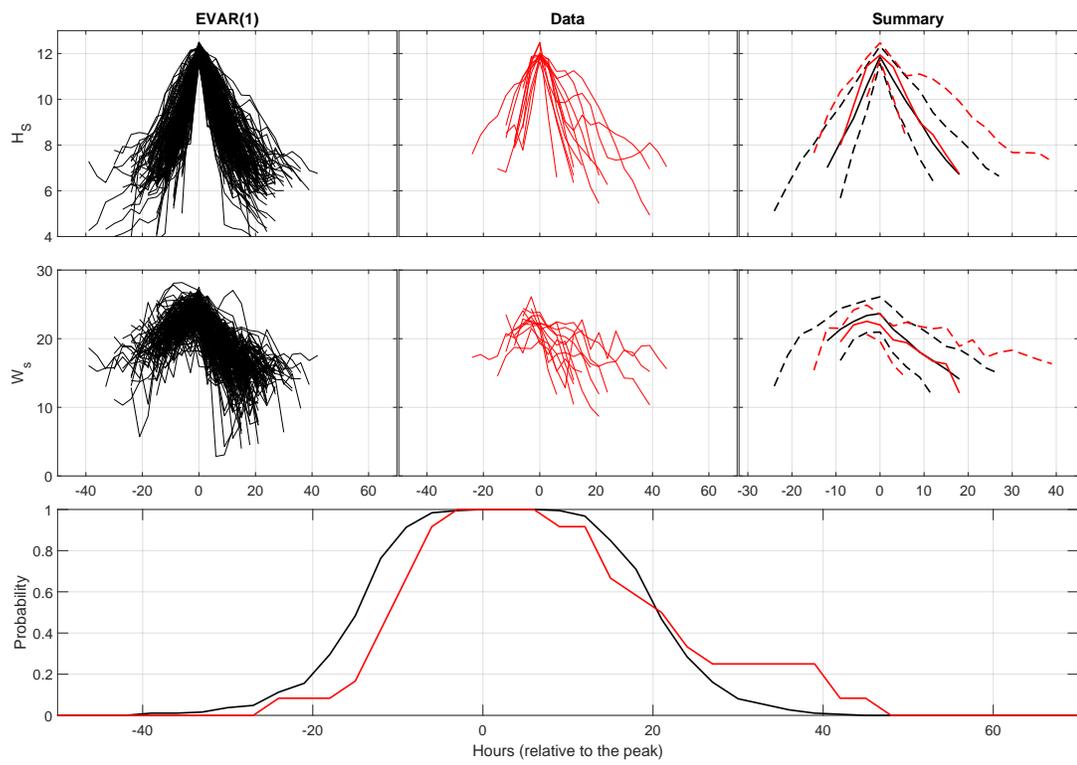
Figure F.1.1: EVAR(1)
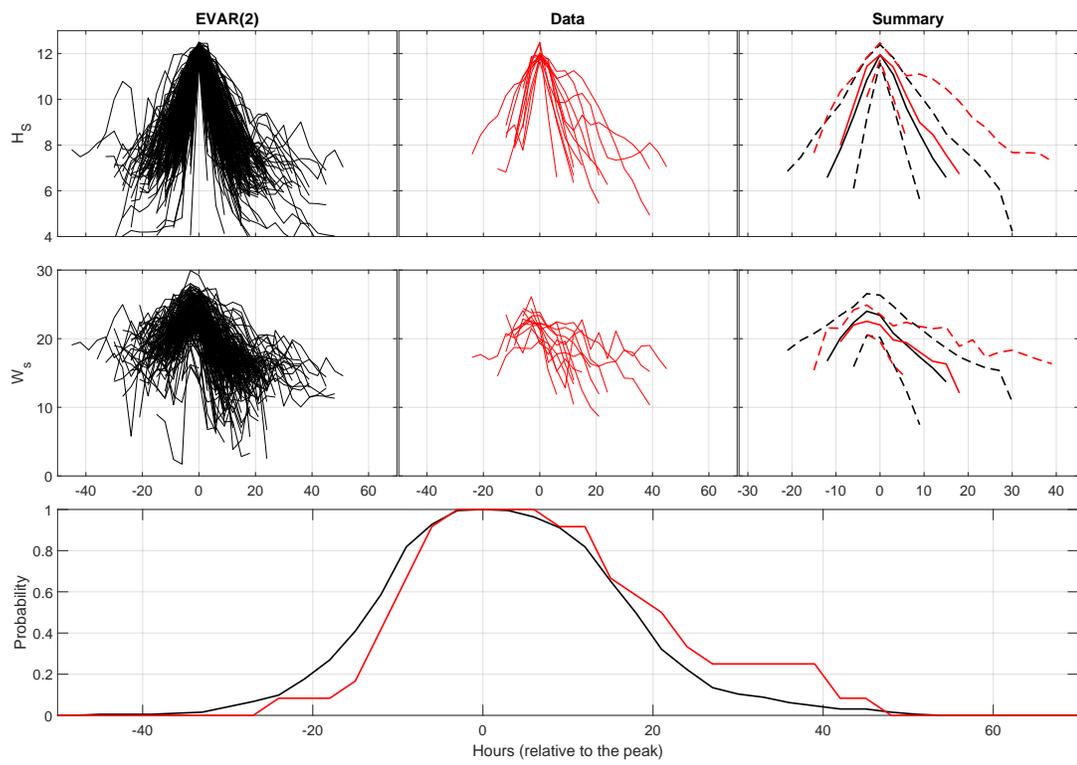
Figure F.1.2: EVAR(2)
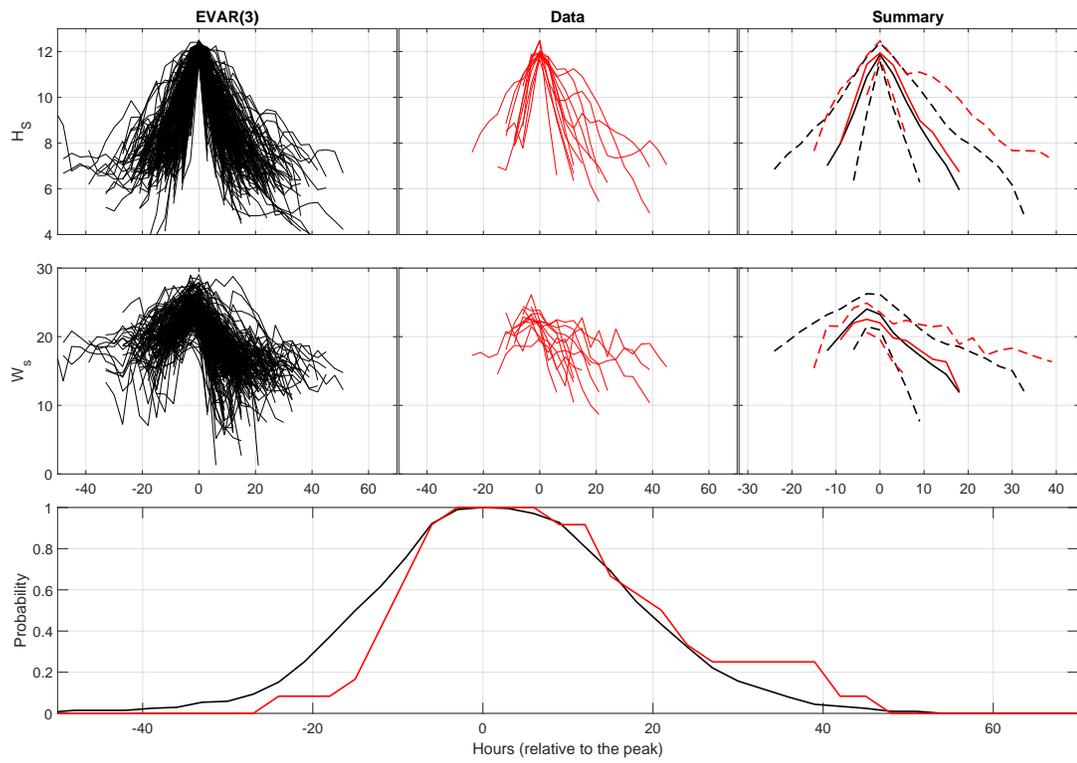
Figure F.1.3: EVAR(3)
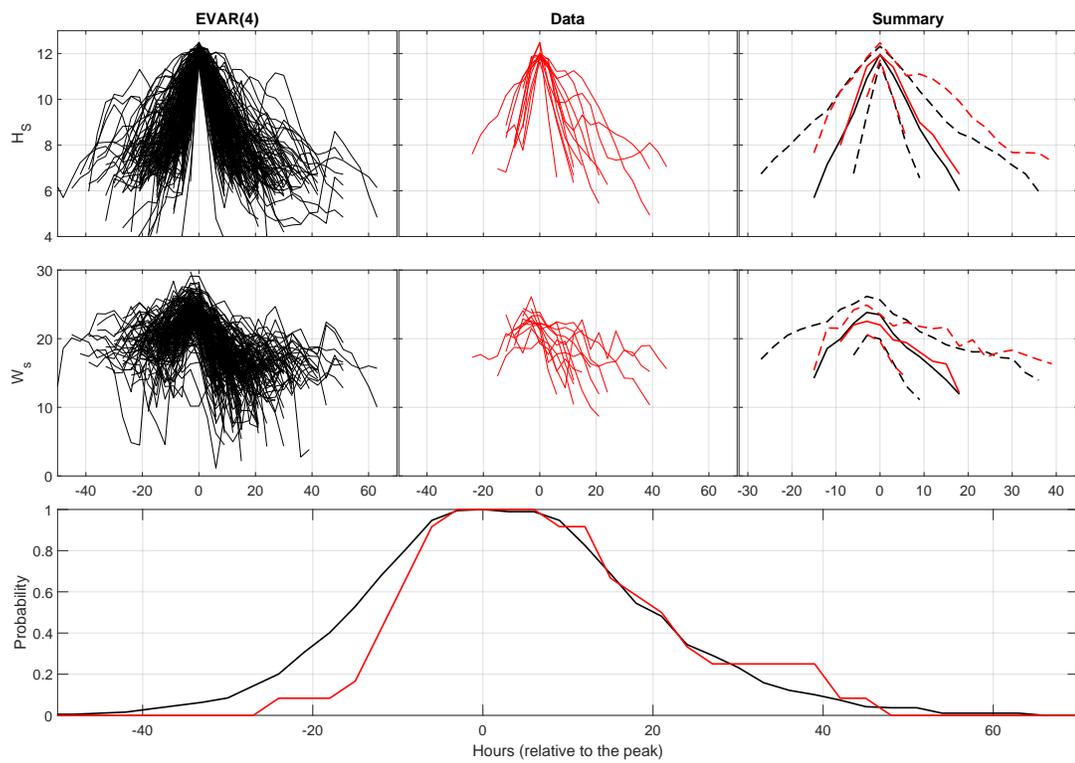
Figure F.1.4: EVAR(4)
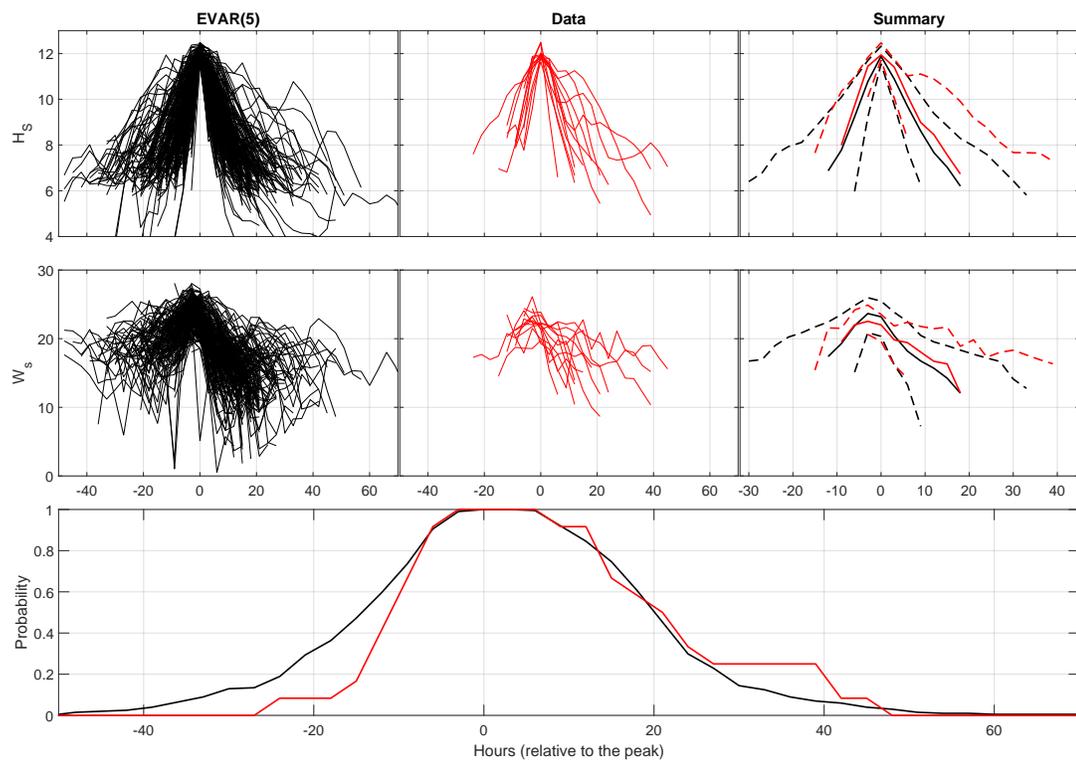
Figure F.1.5: EVAR(5)
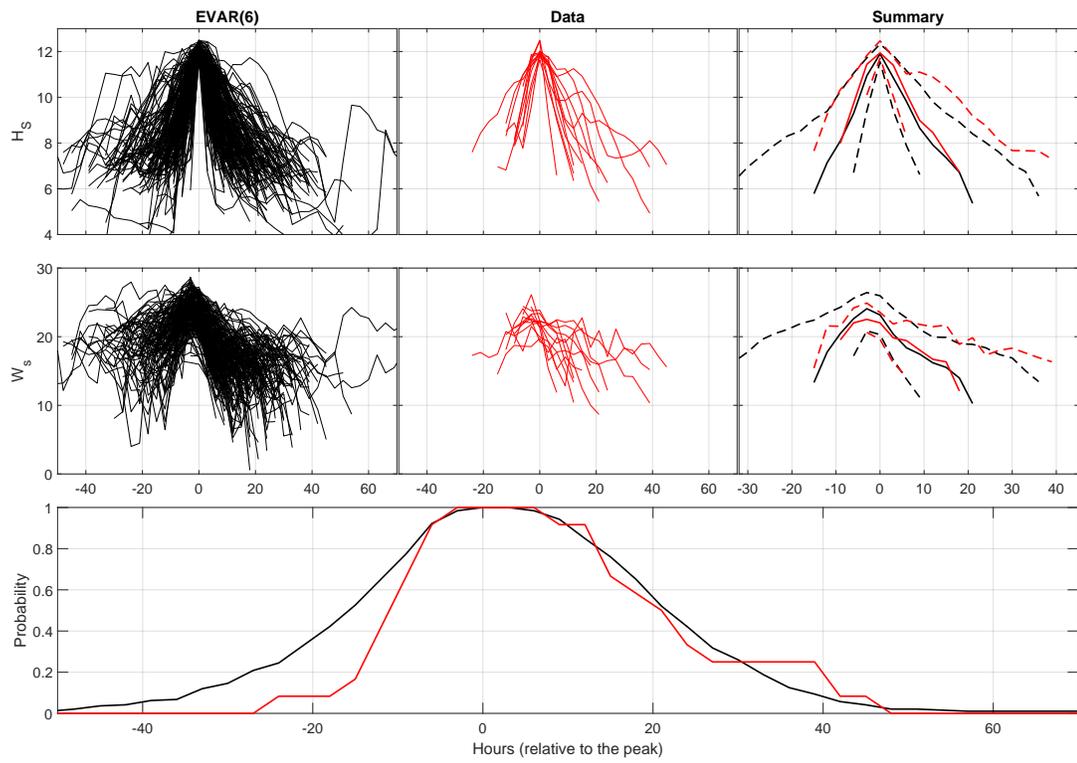
Figure F.1.6: EVAR(6)
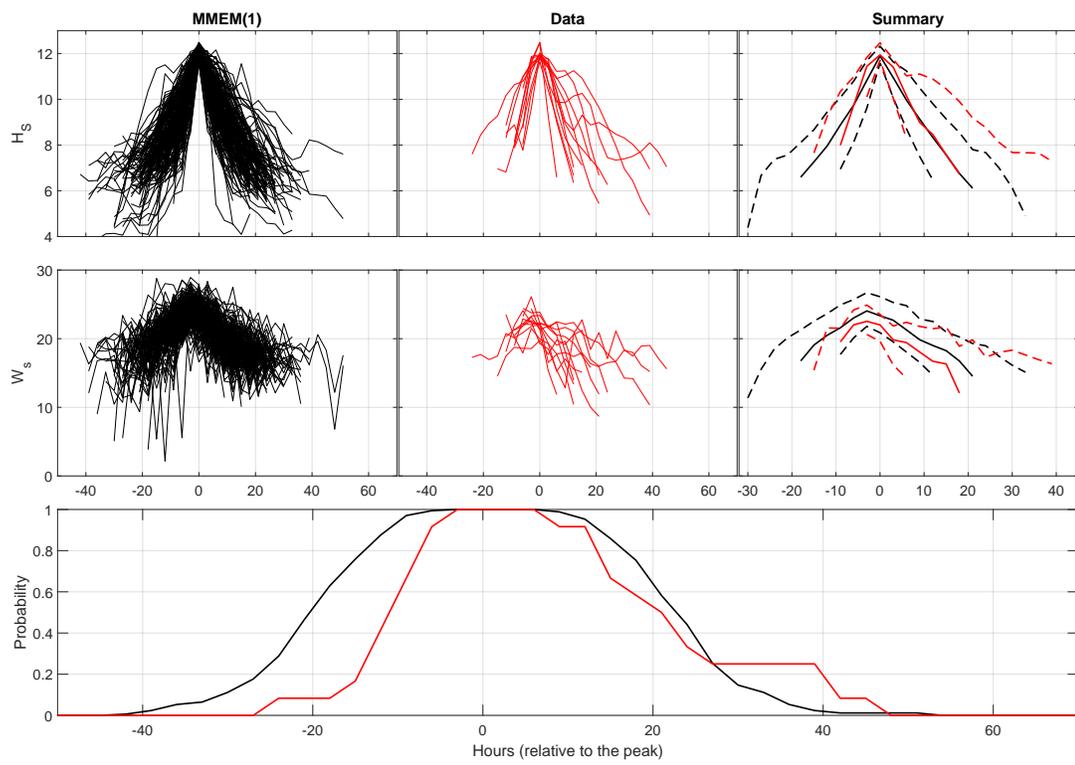
Figure F.1.7:  MMEM(1)
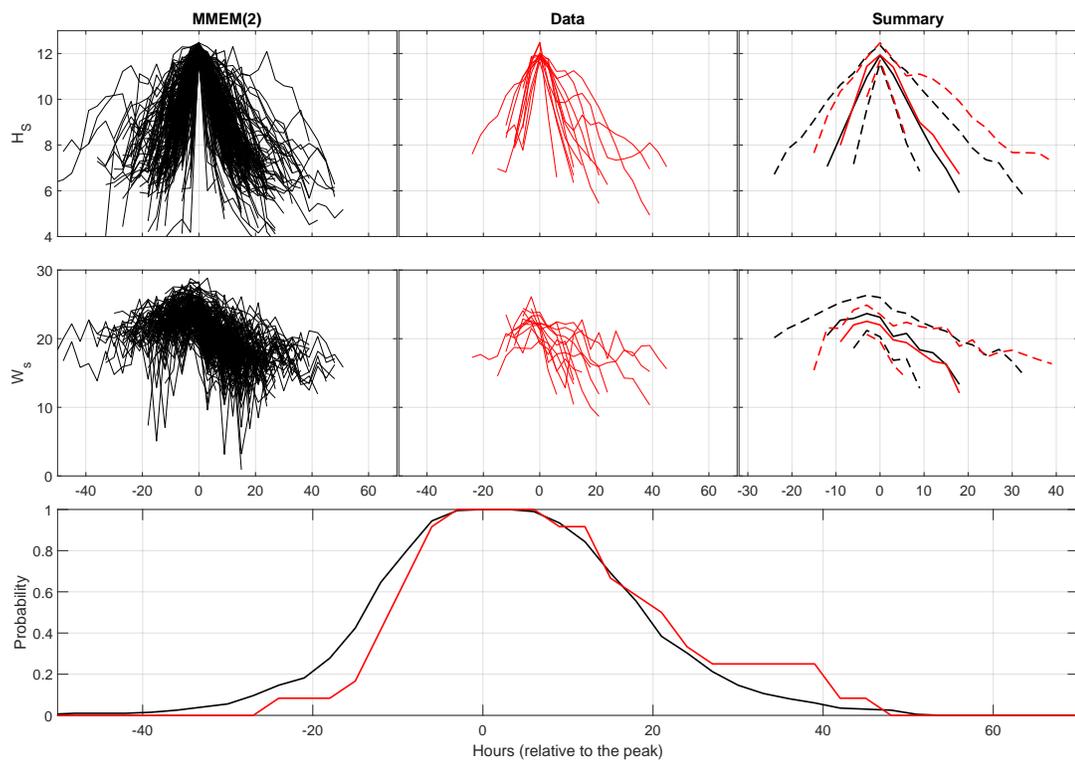
Figure F.1.8: MMEM(2)
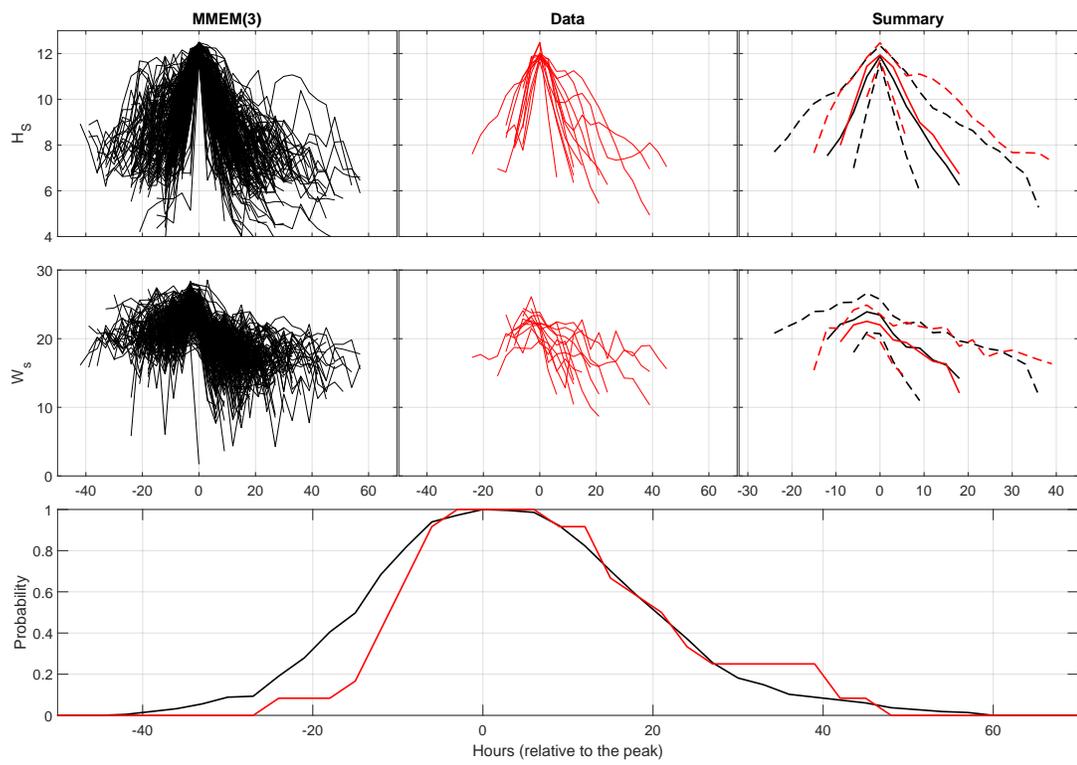
Figure F.1.9:  MMEM(3)
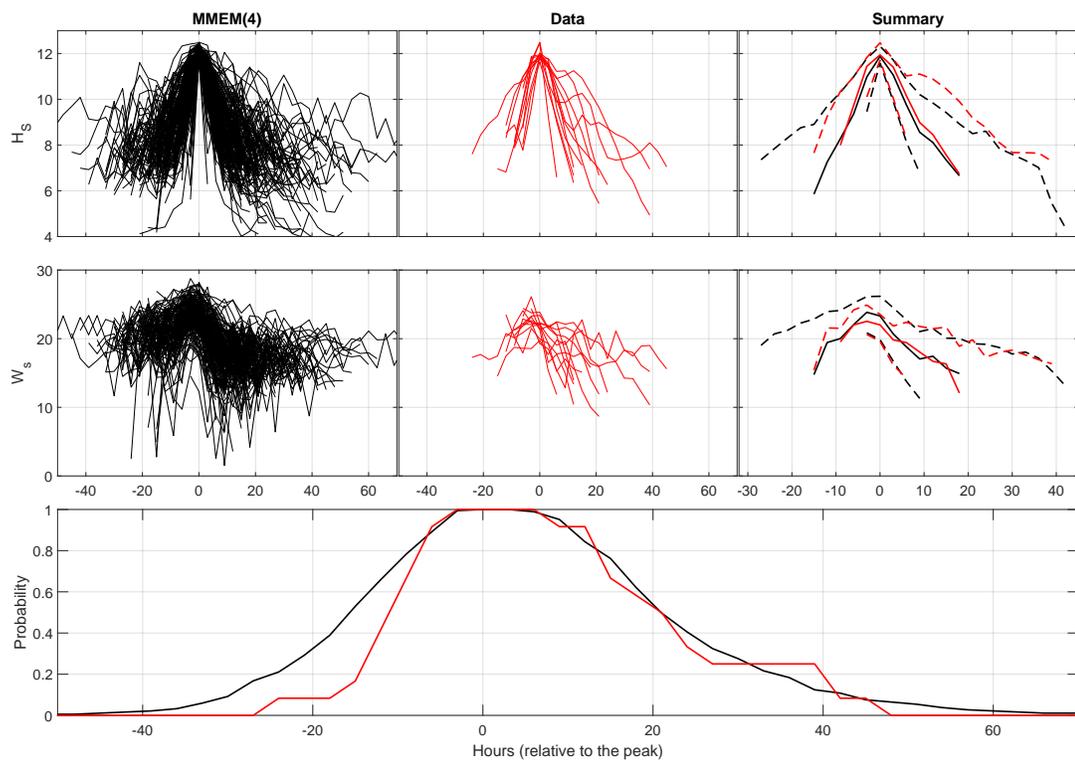
Figure F.1.10: MMEM(4)
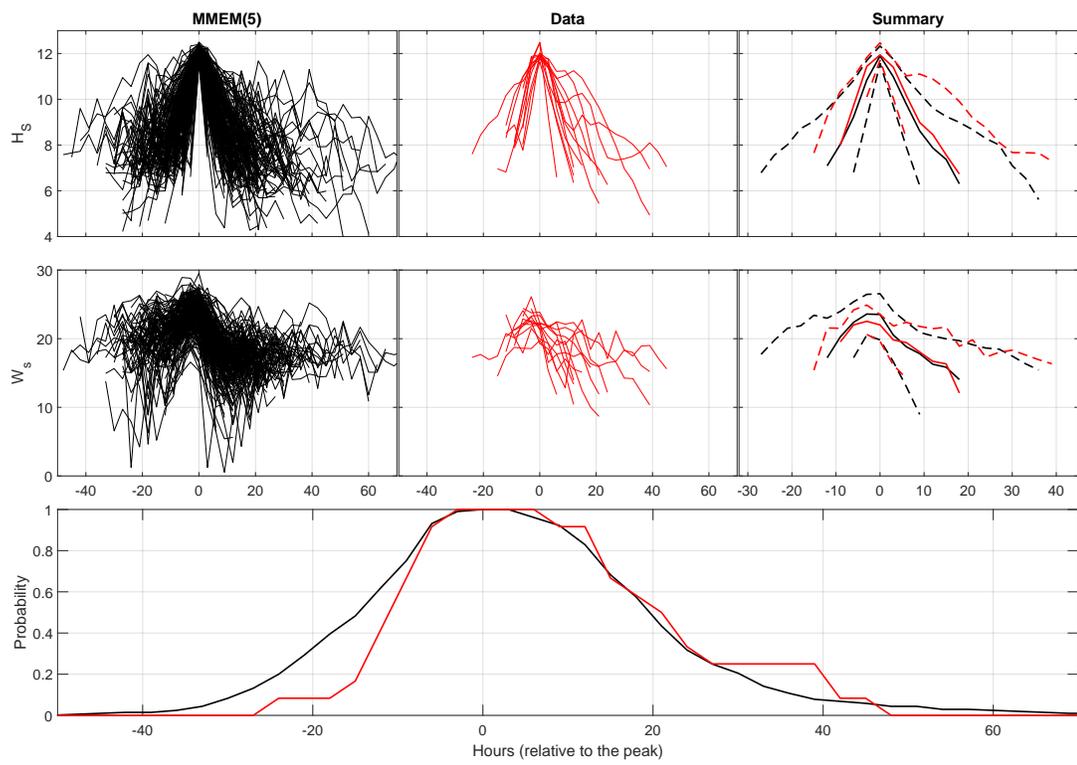
Figure F.1.11: MMEM(5)
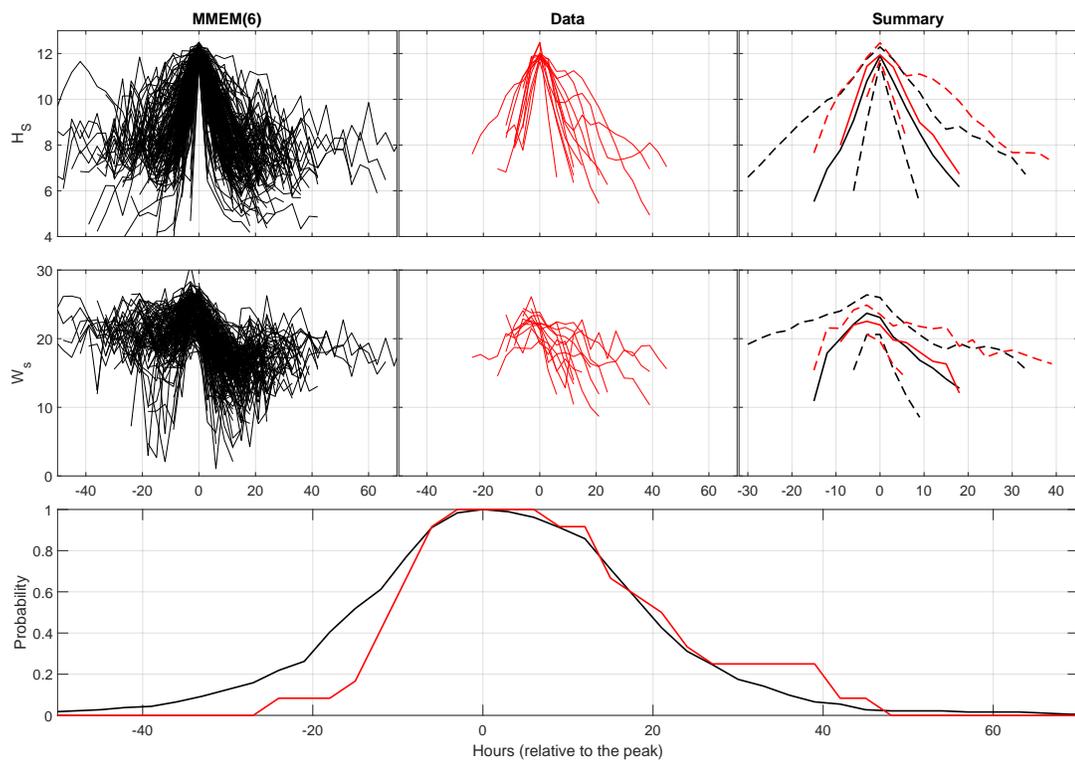
Figure F.1.12: MMEM(6)
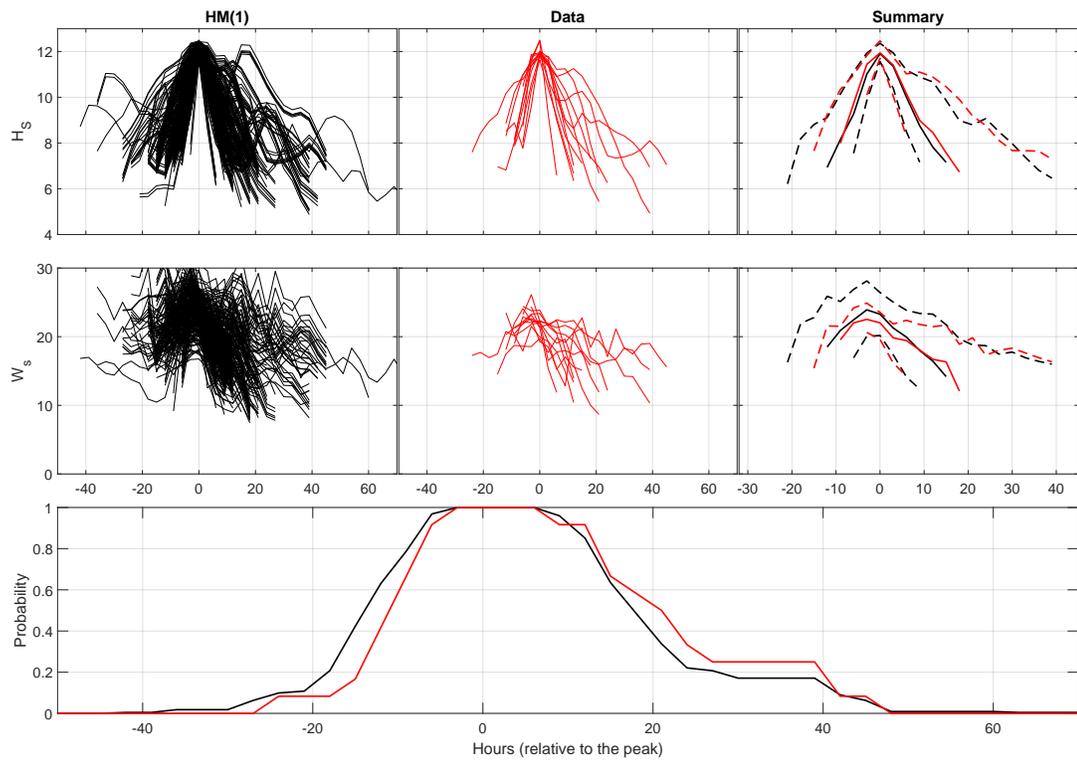
Figure F.1.13: HM

# Bibliography

Balkema, A. A. and Resnick, S. I. (1977). Max-infinite divisibility. *Journal of Applied Probability*, 14(2):309–319.

Barnett, V. (1976). The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

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.

Bortot, P. and Coles, S. (2003). Extremes of Markov chains with tail switching potential. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4):851–867.

Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867.

Bücher, A. and Segers, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125.

Cai, J.-J., de Haan, L., and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201.

Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577.

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J.-M. (2005). Extreme Value and Related Models with Applications in Engineering and Science. Hoboken, New Jersey: Wiley.

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(1):207–222.

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., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Coles, S. G., Heffernan, J. E., 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.

Cooley, D. and Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. PhD thesis, Mathematical Center Tracts, Amsterdam.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: an Introduction*. Springer Science & Business Media.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337.

D'Errico, J. (2020). Slm-shape language modeling. *(https://www.mathworks.com/matlabcentral/fileexchange/24443-slm-shape-language-modeling), MATLAB Central File Exchange.* Retrieved July 27, 2020.

Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli*, 25(3):1690–1723.

Drago, M., Giovanetti, G., and Pizzigalli, C. (2013). Assessment of significant wave height–peak period distribution considering the wave steepness limit. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 55393, page V005T06A015. American Society of Mechanical Engineers.

DuMouchel, W. H. (1983). Estimating the stable index $\alpha$ in order to measure tail thickness: A critique. *the Annals of Statistics*, 11(4):1019–1031.

Eastoe, E. F. (2019). Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics*, 30(5):e2560.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45.

Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Engelke, S. and Ivanovs, J. (2020). Sparse structures for multivariate extremes. *Annual Review of Statistics and its Application*, 8:241–270.

Feld, G., Randell, D., Wu, Y., Ewans, K., and Jonathan, P. (2015). Estimation of storm peak and intrastorm directional–seasonal design conditions in the North Sea. *Journal of Offshore Mechanics and Arctic Engineering*, 137(2):021102.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: Pwm estimators. *The Annals of statistics*, 43(1):276–298.

Ferreira, A., de Haan, L., and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*, 37(5):401–434.

Ferro, C. A. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Fougeres, A.-L. and Soulier, P. (2012). Estimation of conditional laws given an extreme component. *Extremes*, 15(1):1–34.

Galambos, J. (1994). *Extreme value theory for applications*. Springer.

Gandy, A., Jana, K., and Veraart, A. E. (2022). Scoring predictions at extreme quantiles. *AStA Advances in Statistical Analysis*, 106:527–544.

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453.

Gomes, M. I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction. *Extremes*, 4(4):331–358.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, 32(2):177–203.

Hall, P. and Weissman, I. (1997). On the estimation of extreme tail probabilities. *The Annals of Statistics*, pages 1311–1326.

Haver, S. and Winterstein, S. R. (2009). Environmental contour lines: A method for estimating long term extremes by a short term analysis. *Transactions of the Society of Naval Architects and Marine Engineers*, 116:116–127.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):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 (Statistical Methodology)*, 66(3):497–546.

Hilal, S., Poon, S.-H., and Tawn, J. A. (2011). Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX. *Journal of Banking & Finance*, 35(9):2374–2387.

Holthuijsen, L. H. (2010). *Waves in oceanic and coastal waters*. Cambridge university press.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics*, pages 2043–2071.

Hsing, T., Hüsler, J., and Leadbetter, M. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78(1):97–112.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Janßen, A. and Segers, J. (2014). Markov tail chains. *Journal of Applied Probability*, 51(4):1133–1153.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9(1):75–81.

Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, 22(1):47–64.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

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.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):521–543.

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10(1):29–52.

Liu, Y. and Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, 127:19–35.

Loretan, M. and Phillips, P. C. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of empirical finance*, 1(2):211–248.

Lugrin, T., Davison, A. C., and Tawn, J. A. (2016). Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, 19(3):491–515.

Lugrin, T., Tawn, J. A., and Davison, A. C. (2021). Sub-asymptotic motivation for new conditional multivariate extreme models. *Stat*, 10(1):e401.

Ma, L. and Swan, C. (2020). The effective prediction of wave-in-deck loads. *Journal of Fluids and Structures*, 95:102987.

Morison, J., Johnson, J., and Schaaf, S. (1950). The force exerted by surface waves on piles. *Journal of Petroleum Technology*, 2(05):149–154.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Nolde, N. and Wadsworth, J. L. (2021). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*.

Northrop, P. J. and Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2):289–303.

O'Brien, G. L. et al. (1987). Extreme values for stationary and Markov sequences. *The Annals of Probability*, 15(1):281–291.

Oorschot, J. and Zhou, C. (2020). All block maxima method for estimating the extreme value index. *arXiv preprint arXiv:2010.15950*.

Papastathopoulos, I., Strokorb, K., Tawn, J. A., and Butler, A. (2017). Extreme events of markov chains. *Advances in Applied Probability*, 49(1):134–161.

Papastathopoulos, I. and Tawn, J. A. (2020). Hidden tail chains and recurrence equations for dependence parameters associated with extremes of higher-order markov chains. *arXiv preprint arXiv:1903.04059*.

Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Advances in Applied Probability*, 29(1):138–164.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981*.

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, O., Haakenstad, H., Aarnes, O. J., and Furevik, B. R. (2009). A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea and the Barents Sea. *Norwegian Meteorological Institute*.

Reistad, M., Breivik, y., 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.

Richards, J., Tawn, J. A., and Brown, S. (2021). Modelling extremes of spatial aggregates of precipitation using conditional methods. *arXiv preprint arXiv:2102.10906*.

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.

Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains. *Advances in Applied Probability*, 20(2):371–390.

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.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical journal*, 10(1):33–60.

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.

Shooter, R., Tawn, J., Ross, E., and Jonathan, P. (2021). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24(2):241–265.

Sibuya, M. (1959). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

Simpson, E. S. (2019). *Classifying and Exploiting Structure in Multivariate Extremes*. PhD thesis, Lancaster University.

Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:100482.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2021). A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis*, 184:104736.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Smith, R. L. (1987). Approximations in extreme value theory. Technical report, North Carolina University at Chapel Hill Center for Stochastic Processes.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29(1):37–45.

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):515–528.

Tancredi, A., Anderson, C., and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9(2):87–106.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Tawn, J. A., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial statistics*, 28:39–58.

Tendijck, S., Ross, E., Randell, D., and Jonathan, P. (2019). A model for the directional evolution of severe ocean storms. *Environmetrics*, 30(1):e2541.

Tiao, G. C. and Box, G. E. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–195.

Towe, R., Tawn, J. A., Lamb, R., Sherlock, C., and Liu, Y. (2016). Improving statistical models for flood risk assessment. 7:01011.

Towe, R. P., Tawn, J. A., Lamb, R., and Sherlock, C. G. (2019). Model-based inference of conditional extreme value distributions with hydrological applications. *Environmetrics*, 30(8):e2575.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wadsworth, J. L. and Tawn, J. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

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.

Winter, H. C. and Tawn, J. A. (2016). Modelling heatwaves in Central France: a case-study in extremal dependence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):345–365.

Winter, H. C. and Tawn, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes*, 20(2):393–415.

Yun, S. (2000). The distributions of cluster functionals of extreme events in a *d*th-order Markov chain. *Journal of Applied Probability*, 37(1):29–44.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130.