Assessing and Modelling Extremal Dependence in Spatial Extremes

Mónika Kereszturi, B.Sc.(Hons.), M.Res.

Submitted for the degree of Doctor of Philosophy

at Lancaster University.

December 2016

STOR-i

excellence with impact

Abstract

Offshore structures, such as oil platforms and vessels, must be built such that they can withstand extreme environmental conditions (e.g., high waves and strong winds) that may occur during their lifetime. This means that it is essential to quantify probabilities of the occurrence of such extreme events. However, a difficulty arises in that there are very limited data available at these levels. The statistical field of extreme value theory provides asymptotically motivated models for extreme events, hence allowing extrapolation to very rare events.

In addition to the risk to a single site, we are also interested in the joint risk of multiple offshore platforms being affected by the same extreme event. In order to understand joint extremal behaviour for two or more locations, the spatial dependence between the different locations must be considered. Extremal dependence between two locations can be of two types: asymptotic independence (AI) when the extremes at the two sites are unlikely to occur together, and asymptotic dependence (AD) when it is possible for both sites to be affected simultaneously. For finite samples it is often difficult to determine which type of dependence the data are more consistent with. In a large ocean basin it is reasonable to expect both of these features to be present, with some close by locations AD, with the dependence decreasing with distance, and some far apart locations AI. In this thesis we develop new diagnostic tools for distinguishing between AD and AI and illustrate these on North Sea wave height data. We also investigate how extremal dependence changes with direction and find evidence for spatial anisotropy in our data set.

The most widely used spatial models assume asymptotic dependence or perfect independence between sites, which is often unrealistic in practice. Models that attempt to capture both AD and AI exist, but they are difficult to implement in practice due to their complexity and they are restricted in the forms of AD and AI they can model. In this thesis we introduce a family of bivariate distributions that exhibits all the required features of short, medium and long range extremal dependence required for pairwise dependence modelling in spatial applications.

Acknowledgements

First of all, I would like to thank Jonathan Tawn for his guidance and encouragement over the past four years. Jon's continual patience, support and expertise proved invaluable and this thesis would not exist without him. I could not have asked for a better PhD supervisor.

I would also like to thank Philip Jonathan for his insights and enthusiasm that lead to countless interesting discussions, and for giving me a taste of what research in statistics looks like in an industry setting. Thanks also to David Randell, Mirrelijn van Nee and Emma Ross of the Shell statistics team for many useful discussions.

My PhD was funded by EPSRC through the STOR-i Centre for Doctoral Training at Lancaster University and Shell Research Limited, for which I am very grateful. STOR-i provided a wonderful environment and community, and thanks is owed to all students and staff involved. I would especially like to thank Christian, Ivar, Kaylea, Lawrence and Lisa, whose company, both in the office and outside of it, made my time in Lancaster memorable and enjoyable.

Thank you also to my family for raising me in an environment where learning was held in high regard, and for the sacrifices they took that made my studies possible.

Finally, the person I would like to thank most is my husband, Phil. You have been by my side throughout this whole journey and I will be forever grateful for your unwavering love and support.

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.

Chapter 3 has been published as M. Kereszturi, J.A. Tawn, and P. Jonathan. (2016). Assessing extremal dependence of North Sea storm severity. *Ocean Engineering*, 118:242–259.

Chapter 5 has been accepted for publication as M. Kereszturi and J.A. Tawn. (2017). Properties of extremal dependence models built on bivariate max-linearity. *Journal of Multivariate Analysis*.

All analysis in this thesis has been undertaken in the statistical computing language R.

Mónika Kereszturi

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Inadequate design of offshore structures can have wide ranging negative consequences. Weather induced structural damage to oil platforms and vessels can lead to lost revenue, danger to operating staff and environmental pollution. In order to avoid such consequences, design codes stipulate that all offshore structures must be built such that they can withstand extreme environmental events (e.g., high waves and strong winds) with a low probability of failure. This means that it is essential to quantify probabilities of the occurrence of such extreme events, with values sought for events as rare as once in a 100 (or even 10000) years. Hence, interest lies in environmental phenomena that are very rare with limited data available that exceed these levels. An additional challenge is that often we need to estimate probabilities of events that are more extreme than what has been observed previously.

In order to design reliable offshore structures, it is critical to gain an understanding of extreme ocean environments. Significant wave height is one of the most commonly used variables to characterise the ocean environment. This represents an average height of the highest one-third of the waves in a given time period, or in a specific wave or storm system. There are several factors that influence the formation of ocean waves, such as wind direction

and speed, distance of open water that the wind has blown over (i.e., fetch), width of area affected by fetch, distance from shore, time duration and water depth. These factors all work together to determine the size of ocean waves. Traditionally, physical models have been used to model ocean environments and to assess the reliability of offshore structures, but increasingly, the benefits that statistical models can bring to this field are being recognised. One particular area of statistics that is well suited for modelling extreme events is extreme value theory. Unlike standard statistical methods that are driven by mean values, extreme value theory provides asymptotically motivated models for the tails of a distribution, hence allowing extrapolation to very rare events.

An additional benefit of using statistical models is that this allows us to quantify not only the risk of extreme events at a single location, but also the joint risk of several locations being affected by the same extreme event. The joint risk of multiple offshore platforms being affected by the same storm is important information for insurance purposes. In general, neighbouring locations are likely to be affected by the same physical phenomenon, whereas locations further apart are unlikely to be affected simultaneously. In the case of ocean storms, some characteristics of the storm, such as the direction of the storm path, might also have an impact on the risk of multiple sites being affected.

In order to understand joint extremal behaviour for two or more locations, the spatial dependence between the different locations must be considered. However, since the interest lies in the extreme values, the dependence considered here only concerns the tail of the distribution and not the body. This is termed tail dependence or extremal dependence, and can be of two types: asymptotic independence (AI) and asymptotic dependence (AD). In simple terms, we say that two locations are AI if they are unlikely to be affected by extreme events simultaneously. On the other hand, if two locations are AD it is implied that if one location is affected by an extreme event then it is possible for the other location to also be affected. In a relatively large ocean basin, such as the North Sea, it is reasonable to expect both of these features to be present, with some close by locations AD, with the dependence

decreasing with distance, and some far apart locations AI.

Broadly speaking there exist three classes of statistical models for spatial extremes; (i) models that assume AD, (ii) models that assume AI, and (iii) models that attempt to model both AI and AD. This means that we either need to know whether our data is AI or AD in advance and then fit a model of class (i) or (ii) accordingly, or we need to fit a model of class (iii). This may sound simple but there are two problems. Firstly, it is very difficult to determine with reasonable certainty whether a finite data set is consistent with being AD or AI. Secondly, existing models of class (iii) are difficult to implement in practice due to their complexity and they are restricted in the forms of AD and AI they can model.

In this thesis we aim to address both of these problems. The first major contribution of this thesis is the development of new diagnostic tools for distinguishing between AI and AD that can improve confidence in model selection between classes (i) and (ii). However, as suggested previously, it is desirable to have a model that can capture a wide range of extremal dependence structures, ranging from strong AD, weak AD, all the way to AI. As a second major contribution of this thesis we introduce a family of bivariate distributions, with simple multivariate extensions, that exhibits all the required features of short, medium and long range extremal dependence for pairwise dependence modelling in spatial applications. This family is shown to capture all possible bivariate distributions with these properties. We propose novel bivariate characterisations of the extremal dependence structure that reveal structure of this family of distributions that standard measures of extremal dependence fail to identify. Additionally, we investigate how extremal dependence changes with direction and find evidence for spatial anisotropy in extremal dependence in North Sea storm peak significant wave height data.

1.2 Thesis outline

The aim of this thesis is to provide novel statistical methodology for the modelling of extreme ocean environments. The thesis is divided into the following chapters.

Chapter 2 gives a brief overview of the basic results of extreme value theory that are core to the methods developed in the thesis. The overview starts with univariate methods. The two main approaches in univariate extremes arise from different definitions of the concept of being "extreme". The first approach divides the data into blocks (e.g., months or years) and models the block maxima, while the second approach sets a high threshold and models the exceedances of this threshold. These two approaches have extensions in multivariate extremes that we also review in this chapter, with particular focus given to the notion of extremal dependence. We conclude our review by describing the most widely studied class of models for spatial extremes, max-stable processes. Then in the last part of this chapter, we present our work on two extensions to current approaches for modelling extremal dependence. The first extension seeks to use more information when estimating tail dependence by including information about extreme observations in lower dimensional cases. In the second extension we test a recently proposed bias correction method for estimating tail dependence.

As suggested above, it is vital to have good diagnostics to identify the appropriate extremal dependence class. If variables are AI, fitting an AD model can lead to overestimation of the joint risk of extreme events, and hence to higher than necessary design costs of offshore structures. In Chapter 3, we develop improved diagnostics for differentiating between AD and AI dependence classes, which leads to increased confidence in model selection. Application to samples of North Sea sea-state and storm-peak significant wave height suggest that for sites located close by AD is a reasonable assumption, but for sites further apart AI seems to be more appropriate. Our results also suggest that tail dependence changes with direction and distance between spatial locations and with the level of overall dependence in the data.

In Chapter 4 we further investigate the effects of direction on extremal dependence in samples of North Sea storm-peak significant wave height. It is well known that in the North Sea the biggest storms come from a northerly or north-westerly direction. Here we want to see whether this translates to an effect in the extremal dependence between various locations in the data set. We use composite likelihood methods to fit various AI and AD models to data at locations along strips with different orientations. Simplistically, we can imagine a storm propagating along a straight line in time as a half-plane in space. Symmetry then suggests that the extremal dependence spatially would exhibit limiting forms (i) along the storm trajectory and (ii) perpendicular to the storm trajectory (i.e., along the edge of the half plane representing the storm front). In the North Sea, large storms mostly travel in an approximately north to south direction and we observe that two locations X_1 and X_2 located on a strip with a west-east orientation show stronger dependence than two locations Y_1 and Y_2 located on a strip with a north-south orientation (when the distance between X_1 and X_2 and Y_1 and Y_2 is kept constant). This indicates that extremal dependence is stronger across the storm front than in the storm direction. We also find that extremal dependence varies smoothly with the direction of the strips, suggesting that direction needs to be considered when modelling extremal dependence.

In Chapter 5 we introduce a family of bivariate distributions that contains both AD and AI components and has the flexibility to capture all dependence forms within very broad classes in each case. Bivariate max-linear models provide a core building block for characterising bivariate max-stable distributions. The limiting distribution of marginally normalised componentwise maxima of bivariate max-linear models can be dependent (asymptotically dependent) or independent (asymptotically independent). However, for modelling bivariate extremes they have weaknesses in that they are exactly max-stable with no penultimate form of convergence to asymptotic dependence, and asymptotic independence arises if and only if the bivariate max-linear model is independent. In this chapter we present more realistic structures for describing bivariate extremes. We show that these models are built on

bivariate max-linearity but are much more general. In particular, we present models that are dependent but asymptotically independent and others that are asymptotically dependent but have penultimate forms. We characterise the limiting behaviour of these models using two new different angular measures in a radial-angular representation that reveal more structure than existing measures. The bivariate distributions studied here have simple multivariate extensions that can be used in spatial applications with higher dimensions.

In Chapter 6 we show how to do conditional simulation for the models introduced in Chapter 5. Conditional simulation is useful when we want to estimate probabilities of an extreme event at a particular location, given the characteristics of the same event at one or more other locations. In an oceanographic application, for example, we might be interested in the probability of wave heights exceeding a certain level at a site, given the wave heights being large at some other sites.

In Chapter 7 we discuss the outcomes of the thesis and present some ideas for further work.

Chapter 2

Introduction to Extreme Value Theory

In oceanography and other environmental applications it is often the case that extreme events (e.g. extreme wave heights and wind speeds) are of interest, as these are the ones that can cause the most damage. Some of these events can be as rare as once in 100 or 500 years, with very limited data available at these levels, or in some cases none at all. In statistical terms, the interest lies in the tails rather than the bulk of the distribution. By definition, observations in the tails are scarce, as most data points are concentrated towards the centre of the distribution. In standard statistical methods, parameter estimates and assessment of model fit are strongly driven by these central values, and different models that fit the body of the data well can have very different extrapolations in the tail. Further difficulty arises from the fact that estimation of the probability of events beyond the sample maxima or minima is often required.

The ideas described above gave rise to the statistical field of Extreme Value Theory (EVT). The problem requires an extrapolation from observed levels to unobserved levels, and EVT provides asymptotic models to enable such extrapolation (Coles, 2001).

In this chapter we aim to give an overview of some of the most important results in EVT. In Section 2.1 we focus on approaches for modelling the tail distribution of a single variable of interest. In Section 2.2 we present models for multivariate extreme values, with particular focus on the idea of extremal dependence. In Section 2.3 we give an overview of models for spatial extremes. Finally, in Section 2.4 we present some extensions to the described methods.

2.1 Univariate extremes

Within univariate extremes there are two common approaches within the EVT literature; models for block maxima and models for exceedances of a high threshold. In the following we outline the main theoretical results for both of these approaches, along with advantages and disadvantages of using them.

2.1.1 Models for block maxima

Consider an independent identically distributed (IID) sample $X_1, \ldots, X_n$ with common distribution function F. Let $M_{X,n} = \max(X_1, \ldots, X_n)$ be the sample maximum. When it is clear the subscript X will be dropped and we will write M_n to denote the sample maximum. Here we will only consider maxima, i.e. the upper tail of the distribution, but it is easy to see that the theory for maxima can be also applied to minima:

$$\begin{aligned} m_{X,n} &= \min(X_1, \ldots, X_n) \\ &= -\max(-X_1, \ldots, -X_n) \\ &= -M_{-X,n}. \end{aligned}$$

The distribution of M_n can be obtained in the following way:

$$\begin{aligned}\Pr(M_n \leqslant x) &= \Pr(X_1 \leqslant x, \ldots, X_n \leqslant x) \\ &= \Pr(X_1 \leqslant x) \cdots \Pr(X_n \leqslant x) \\ &= \prod_{i=1}^{n} \Pr(X_i \leqslant x) \\ &= \{F(x)\}^n.\end{aligned}$$

However, as generally F is unknown, this formula is not useful. As we are often interested in the maximum of a large number of variables it would be convenient to approach modelling M_n using an asymptotic argument. However, $M_n \to x^F$ in probability as $n \to \infty$, where x^F is the upper end point of F. Hence, the asymptotic distribution of M_n is degenerate. Limit arguments similar to the Central Limit Theorem suggest that some kind of scaling is required.

Theorem 1. *The Extremal Types Theorem (ETT) states that if there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leqslant x\right) \to G(x), \tag{2.1.1}$$

for some non-degenerate distribution G, then G is of the same type as one of the following distributions:

$$\underline{\textit{Gumbel}}\textit{:}\ G(x) = \exp[-\exp(-x)]\ \textit{for}\ -\infty < x < \infty$$

$$\underline{\textit{Fréchet}}\textit{:}\ G(x) = \begin{cases} 0 & \textit{for } x \leqslant 0, \\ \exp(-x^{-\alpha}) & \textit{for } x > 0, \alpha > 0 \end{cases}$$

$$\underline{\textit{Negative Weibull}}\textit{:}\ G(x) = \begin{cases} \exp[-(-x)^{\alpha}] & \textit{for } x < 0, \alpha > 0, \\ 1 & \textit{for } x \geqslant 0. \end{cases}$$

In words, the ETT states that the rescaled sample maxima converge in distribution to a variable having a distribution within one of the above families of distributions, as suggested by Fisher and Tippett (1928). The Gumbel, Fréchet and Negative Weibull families of distributions are the only possible limits for the distributions of the rescaled sample maxima, regardless of the distribution F of the population.

It is inconvenient to work with three separate classes of limiting distributions, so it is preferable to use a parametrisation that unifies these distributions. Von Mises (1936) and Jenkinson (1955) derived the GEV(μ, σ, ξ) distribution with distribution function

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} \tag{2.1.2}$$

where $x_+ = \max(x, 0)$ and μ, $\sigma > 0$, and ξ are the location, scale and shape parameters, respectively.

The shape parameter ξ describes the behaviour of the distribution in the tail and determines the family of the limit distribution.

- $\xi > 0$ corresponds to the Fréchet distribution and a heavy upper tail,
- $\xi = 0$ corresponds to the Gumbel distribution and an exponential upper tail,
- $\xi < 0$ corresponds to the Negative Weibull distribution and a finite upper tail.

If $\xi \to 0$ the GEV takes the form

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}.$$

Theorem 2. *The Unified Extremal Types Theorem (UETT) states that in equation (2.1.1), G is a member of the GEV family, i.e. it is of the same type (up to location and scale parameter) as*

$$G(x) = \exp[-(1+\xi x)_+^{-1/\xi}]$$

for some value of ξ.

The UETT does not guarantee the existence of a non-degenerate limit or say which type will arise. It also does not tell us how to pick a_n and b_n. In order to justify the UETT (or equivalently the ETT) we need to introduce the notion of max-stability. For a full proof the reader is referred to Leadbetter et al. (1983).

Definition. *A distribution* G *is said to be max-stable if, for every integer* $n > 0$ *there are constants* $\alpha_n > 0$ *and* β_n *such that*

$$G^n(\alpha_n x + \beta_n) = G(x). \tag{2.1.3}$$

In simple terms, equation (2.1.3) states that the operation of taking sample maxima leads to an identical distribution to the distribution of the original sample, except a change in the location and scale. It can be proved that a distribution is max-stable if, and only if, it is a GEV distribution (Resnick (1987) Proposition 5.9).

In most practical applications interest lies in the estimation of the so called return period and return level. For $1/p$ return period let z_p be the $1-p$ quantile of the GEV distribution for $0 < p < 1$. Then z_p is the return level that is expected to be exceeded on average once every $1/p$ years. An estimate of z_p can be obtained by inverting equation (2.1.2) and substituting the maximum likelihood estimates of the GEV parameters, giving

$$\widehat{z}_p = \begin{cases} \widehat{\mu} - \frac{\widehat{\sigma}}{\widehat{\xi}}\left[1 - \{-\log(1-p)\}^{-\widehat{\xi}}\right] & \text{for } \widehat{\xi} \neq 0, \\ \widehat{\mu} - \widehat{\sigma}\log\{-\log(1-p)\} & \text{for } \widehat{\xi} = 0. \end{cases}$$

The block maxima approach is sensible when we only have access to data in the form of weekly, monthly or yearly maxima. However, if an entire dataset of, say, hourly or daily observations is available, then blocking the data and using only the maximum value in each block is a wasteful approach. Figures 2.1.1 and 2.1.2 illustrate this idea. We can make

better use of the data by avoiding the procedure of blocking. This leads to the methods described in the following section on threshold models.

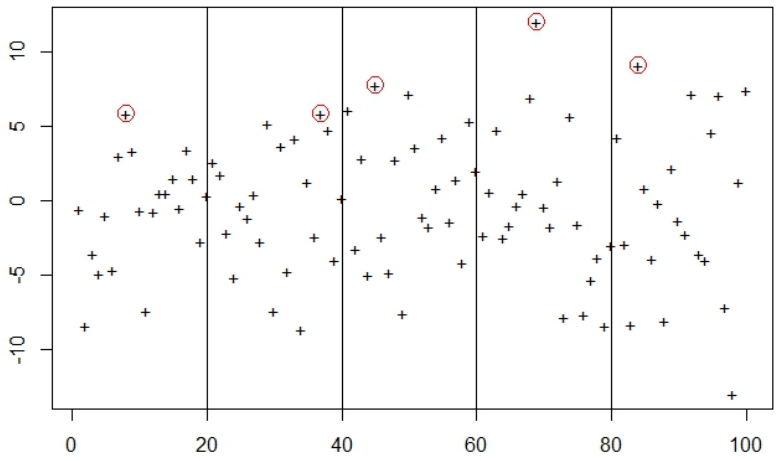


Figure 2.1.1: Illustration of the block maxima approach

2.1.2 Threshold models

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common marginal distribution F, and let us consider those of the X_i that exceed some high threshold u to be extreme. We will focus on the distribution of the exceedances of this high threshold, i.e. $X|X > u$, which can be written as

$$\Pr\{X > u + y | X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \quad y > 0. \tag{2.1.4}$$

It can be shown that the limit distribution of scaled exceedances as u tends to the upper endpoint of F is a generalized Pareto distribution, first derived by Pickands (1975), with distribution function

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-1/\xi}, \tag{2.1.5}$$

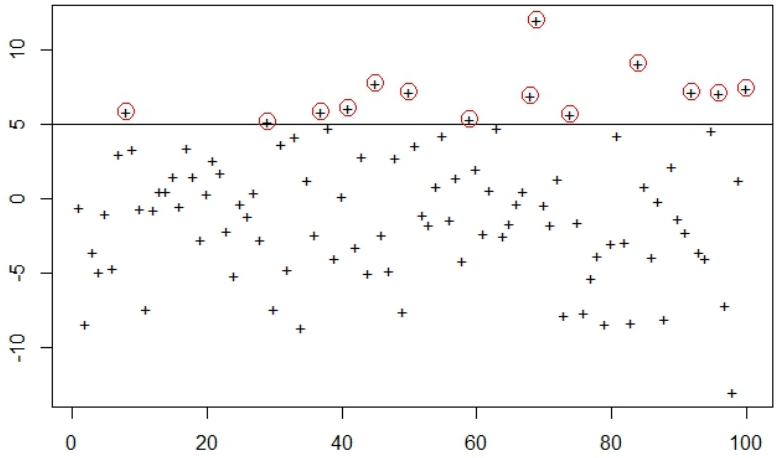


Figure 2.1.2: Illustration of the threshold exceedances approach

where σ_u is the scale parameter dependent on some high threshold u, and ξ is the shape parameter. If block maxima have limiting distribution G from the GEV family with shape parameter ξ, then threshold excesses have a corresponding limiting distribution within the generalised Pareto (GP) family with the same shape parameter. Leadbetter et al. (1983) gives a full proof of this. We can define the threshold excesses as $Y_u = \max(X - u, 0)$, and then $Y_u | Y_u > 0 \sim \text{GP}(\sigma_u, \xi)$.

One of the difficulties of implementing the threshold method is choice of threshold. On the one hand we want to use as low a threshold as possible to maximise the amount of data used. On the other hand we want to use as high a threshold as possible to make sure that the asymptotic arguments hold and the GP distribution fits well. This choice is equivalent to choosing the block size in the block maxima approach. Usually graphical methods are used to make decisions about threshold selection, with the most common being mean residual life plots and parameter stability plots. The threshold stability property states that if Y_u satisfies $Y_u | Y_u > 0 \sim \text{GP}(\sigma_u, \xi)$ for some high threshold u, then for any higher threshold $v \geqslant u$, $Y_v | Y_v > 0 \sim \text{GP}(\sigma_u + \xi(v - u), \xi)$. So ξ is constant with threshold, but $\sigma_v = \sigma_u + \xi(v - u)$ is not. More recent advances in threshold selection methods generally

build on and seek to improve the traditional diagnostic methods (e.g. Wadsworth and Tawn (2012b), Northrop and Coleman (2014) and Wadsworth (2016)). Scarrott and MacDonald (2012) give an overview of traditional and more recent methods of extreme value threshold selection.

Return levels can be derived similarly to block maxima, but since here we model data conditionally on exceeding a high threshold u, we must first undo this conditioning by multiplying by the rate of exceedance $\lambda_u = \Pr(X > u)$. Hence, the unconditional survivor function is

$$\Pr(X > x) = \lambda_u \left[1 + \xi \left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, \quad x > u.$$

This can be inverted to give the return level that will be exceeded once every N years:

$$\widehat{x}_N = \begin{cases} u + \frac{\widehat{\sigma}_u}{\widehat{\xi}} \left[(N n_y \lambda_u)^{\widehat{\xi}} - 1\right], & \text{for } \widehat{\xi} \neq 0, \\ u + \widehat{\sigma}_u \log(N n_y \lambda_u), & \text{for } \widehat{\xi} = 0, \end{cases}$$

where $\widehat{\sigma}_u$ and $\widehat{\xi}$ are the maximum likelihood estimates of the generalised Pareto scale and shape parameters and n_y is the number of observation in a year.

The GEV and the GP results can both be derived from a Poisson point process representation (Pickands, 1971). Let $P_1, P_2, \ldots$ be a sequence of point processes on $[0,1] \times \mathbb{R}$ defined by

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right); i = 1, \ldots, n \right\}, \tag{2.1.6}$$

where a_n and b_n are norming constants. Then, on the set $[0,1] \times (b_l, \infty)$, where $b_l = \lim_{n\to\infty}(x_F - b_n)/a_n$, with x_F being the lower endpoint of the support of the underlying distribution of the X_i, we have that $P_n \to P$ as $n \to \infty$, where P is a non-homogeneous Poisson process with intensity $\lambda(x) = (1 + \xi x)_+^{-1-1/\xi}$. See Smith (1989) for the statistical implementation of this asymptotic result.

2.2 Multivariate extremes

The methods for univariate extremes described above assume that the variables are independently and identically distributed (IID). In practice this is often not a realistic assumption to make. Dependence between variables can occur, for example, when the different processes are influenced by a common physical phenomenon; in the context of oceanography an obvious example would be wind speed and significant wave height that are both due to the same storm event. Another way dependence can arise is when a single process is observed at different spatial locations or points in time. Neighbouring locations are likely to be affected by the same physical phenomenon, and similarly, data points observed close together in time are likely to be impacted by the same storm event. Hence, when modelling multivariate extremes of environmental data, this dependence between the different variables must be taken into account.

In this section we will introduce two approaches commonly used in multivariate extremes; componentwise block maxima, which is an extension of the block maxima approach, and methods for multivariate threshold exceedances that are extensions of univariate threshold models.

2.2.1 Componentwise block maxima

Following Coles et al. (1999), let $(X_{i,1}, \ldots, X_{i,d})$, for $i = 1, \ldots, n$, be a collection of independent and identically distributed d-dimensional random vectors with standard Fréchet margins, i.e. $\Pr(X_{ij} \leqslant x) = \exp(-1/x)$ for $x > 0$ for all i, j. Then we define each of the componentwise maxima in the following way:

$$M_{n,k} = \max\{X_{1,k}, \ldots, X_{n,k}\}, \quad k = 1, \ldots, d,$$

and hence define the vector of componentwise maxima by $\mathbf{M}_n = \{M_{n,1}, \ldots, M_{n,d}\}$.

Theorem 3. *If*

$$\Pr(M_{n,1}/n \leqslant x_1, \ldots, M_{n,d}/n \leqslant x_d) = \{F(nx_1, \ldots, nx_d)\}^n \to G(x_1, \ldots, x_d),$$

as $n \to \infty$*, with* $G(x_1, \ldots, x_d)$ *non-degenerate in each margin then* G *is a multivariate extreme value distribution function, defined as* $G(x_1, \ldots, x_d) = \exp\{-V(x_1, \ldots, x_d)\}$*, where*

$$V(x_1, \ldots, x_d) = d \int_{\mathcal{D}} \max_{1 \leqslant j \leqslant d} (w_j / x_j) H(d\mathbf{w}), \tag{2.2.1}$$

where H *is a distribution function on the* $(d-1)$*-dimensional unit simplex* $\mathcal{D} = \{(w_1, \ldots, w_d) : \sum_{j=1}^{d} w_j = 1, w_j \geqslant 0, j = 1, \ldots, d\}$ *satisfying the following condition:*

$$\int_{\mathcal{D}} w_j H(d\mathbf{w}) = 1/d, \;\; j = 1, \ldots, d. \tag{2.2.2}$$

Note that the exponent measure V is homogeneous or order -1, i.e.

$$V(nx_1, \ldots, nx_d) = V(x_1, \ldots, x_d)/n. \tag{2.2.3}$$

For simplicity, we will now consider the bivariate case. For a pair of random variables X_1 and X_2 on Fréchet margins, the bivariate distribution function G is $G(x_1, x_2)) = \exp\{-V(x_1, x_2)\}$, where

$$V(x_1, x_2) = \int_0^1 \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) 2dH(w), \tag{2.2.4}$$

for some distribution function H on the interval $[0, 1]$ satisfying the moment constraint

$$\int_0^1 w dH(w) = 1/2.$$

For example, the bivariate logistic extreme value distribution is defined as

$$F(x_1, x_2) = \exp\left\{-\left(x_1^{-1/\gamma} + x_2^{-1/\gamma}\right)^{\gamma}\right\}, \quad 0 < \gamma \leqslant 1, \tag{2.2.5}$$

and is a bivariate member of the family of multivariate extreme value distributions defined in Theorem 3. For this distribution V and H are defined as

$$V(x_1, x_2) = \left(x_1^{-1/\gamma} + x_2^{-1/\gamma}\right)^{\gamma},$$

and

$$H(w) = \frac{1}{2}\left[\{w^{(1-\gamma)/\gamma} - (1-w)^{(1-\gamma)/\gamma}\}\{w^{1/\gamma} + (1-w)^{1/\gamma}\}^{\gamma-1} + 1\right].$$

Here $\gamma = 1$ corresponds to independence and $\gamma \to 0$ to perfect dependence.

2.2.2 Multivariate threshold models

As in the univariate case, we can gain improvements in efficiency if we have entire series of data available, not just block maxima. Multivariate threshold models provide a more flexible approach for modelling the joint tail than componentwise maxima. A key consideration in multivariate extremes is the extremal dependence structure, hence we start by introducing some measures for extremal dependence. Then we will present three different tail models that are widely used in the multivariate extremes literature; these are the Ledford and Tawn joint tail model, the Ramos and Ledford model and the Heffernan and Tawn conditional model. For ease of notation we will present the bivariate form of these models but they can be extended to more dimensions in a straightforward fashion.

Extremal dependence measures

In standard statistical analysis the correlation (or an equivalent measure) can be used to determine the dependence between two variables. When dealing with extremes the dependence can no longer be measured using correlation, as the dependence in the body can be quite different from the tail dependence. There are several measures for extremal dependence in the extreme value theory literature; here we will introduce some of the most commonly used ones. In Chapter 3 we will see that in applications we can sometimes gain additional insight if we supplement these extremal dependence measure with a dependence measure for the body of the data (such as Spearman's rank correlation coefficient ρ or Kendall's τ).

A common technique in extreme value theory is to remove the effect of the marginals by transforming the variables onto common margins. To achieve this, we need to introduce the copula function (Nelsen, 2006). Subject to continuity conditions, there is a unique function $C(\cdot, \cdot)$ with domain $\mathcal{A} = [0, 1] \times [0, 1]$, such that

$$F(x, y) = C\{F_X(x), F_Y(y)\}, \tag{2.2.6}$$

where $F_X(x) = F(x, \infty)$ and $F_Y(y) = F(\infty, y)$ are the marginal distribution functions. The copula contains complete information about the dependence between X and Y, in a form that is invariant to marginal monotone transformation.

An intuitive measure of the tail dependence for two identically distributed variables, X and Y, is

$$\chi = \lim_{z \to z^*} \Pr(Y > z | X > z),$$

where z^* is the upper limit of the support of the common marginal distribution. In Chapter 3 we will present a version of this measure that holds for (X, Y) having any marginal distributions. Hence, χ is essentially the probability of one variable being extreme given that the other is extreme.

In practice, data could lead to estimates of $\chi = 0$, suggesting asymptotic independence, i.e. the most extreme events do not occur simultaneously. Hence, for the class of asymptotically independent variables χ does not provide any information about dependence at finite levels. To deal with this issue, Coles et al. (1999) introduce the measure $\bar{\chi}$. We can define the joint survivor function $\Pr(X > x, Y > y)$ by

$$\begin{aligned}\bar{F}(x, y) &= 1 - F_X(x) - F_Y(y) + F(x, y) \\ &= \bar{C}\{F_X(x), F_Y(y)\},\end{aligned}$$

where $\bar{C}(u, v) = 1 - u - v + C(u, v)$, $u, v \in [0, 1]$, and C is the copula function. Then, following Coles et al. (1999), we have

$$\bar{\chi}(u) = \frac{2 \log \Pr(U > u)}{\log \Pr(U > u, V > u)} - 1 = \frac{2 \log(1 - u)}{\log \bar{C}(u, u)} - 1 \text{ for } 0 \leqslant u \leqslant 1, \tag{2.2.7}$$

and

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u), \text{ for which } -1 < \bar{\chi} \leqslant 1.$$

The two measures, χ and $\bar{\chi}$, together give a complete measure of dependence for both asymptotically dependent and asymptotically independent variables; $(\chi > 0, \bar{\chi} = 1)$ signifies asymptotic dependence, for which the value of χ gives a measure of strength of dependence; and $(\chi = 0, \bar{\chi} < 1)$ signifies asymptotic independence, for which the value of $\bar{\chi}$ gives the strength of dependence. Table 2.2.1 presents a summary of these dependence measures.

Measure of Dependence	Scale	Asymptotic Dependence	Asymptotic Independence
χ	$[0, 1]$	$\chi \in (0, 1]$	$\chi = 0$
$\bar{\chi}$	$[-1, 1]$	$\bar{\chi} = 1$	$\bar{\chi} \in [-1, 1)$

Table 2.2.1: Summary of dependence measures χ and $\bar{\chi}$

In simple terms, if we have asymptotic independence, extreme events for both variables (X, Y) are very unlikely to occur simultaneously, whereas for asymptotic dependence, if X

is extreme it is also possible for Y to be simultaneously extreme. Most statistical methods for multivariate extremes assume asymptotic dependence, leading to overestimation of extremes if in reality the data are asymptotically independent. On the other hand, if asymptotic independence is assumed when the data are in fact asymptotically dependent, the extremes will be underestimated. Hence it is very important to be able to differentiate between the two cases when fitting a model.

The extremal coefficient is also a commonly used measure of extremal dependence. This measure is based on the joint probability function. In the case of the bivariate extreme value distribution, for (X, Y) on standard Fréchet margins, this is given as,

$$\Pr(X < z, Y < z) = \exp\{-V(1,1)/z\} = \exp(-\theta/z), \quad z > 0, \tag{2.2.8}$$

where $V(x, y)$ is the exponent measure of the joint distribution of (X, Y), and θ is the extremal coefficient of (X, Y). Here θ takes values between $\theta = 1$ when the observations are fully dependent, and $\theta = 2$ when they are independent. The dependence measure χ defined above can also be expressed in terms of the extremal coefficient: $\chi = 2 - \theta$. The idea and definition of the extremal coefficient extends to the multivariate case (with d dimensions) in the obvious way:

$$\Pr(X_1 \leqslant x, \ldots, X_d \leqslant x) = \exp\{-V(1, \ldots, 1)/x\} = \exp(-\theta_d/x), x > 0.$$

Point process limit

Similarly to the univariate case, there is a point process approach that links the componentwise block maxima and multivariate threshold approaches. Following Coles and Tawn (1991), let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of IID random vectors on $\mathbb{R}^d_+$ with Fréchet margins. Let us consider a point process P_n on $\mathbb{R}^d_+$ where $P_n = \{n^{-1}\mathbf{X}_i : i = 1, \ldots, n\}$. Then $P_n \xrightarrow{d} P$ as $n \to \infty$, where P is a non-homogeneous Poisson process on $\mathbb{R}^d_+ \setminus \{0\}$. Pseudoradial and

angular measures can be defined as $r_i = \sum_{j=1}^{d} X_{i,j}/n$ and $w_{i,j} = X_{i,j}/nr$, for $i = 1, \ldots, n$ and $j = 1, \ldots, d$, where $X_{i,j}$ is the jth component of $\mathbf{X}_i$. The intensity measure μ of the limiting process P then satisfies

$$\mu(dr \times d\mathbf{w}) = 2\frac{dr}{r^2} dH(\mathbf{w}), \tag{2.2.9}$$

where $\mathbf{w} = \{w_1, \ldots, w_d\}$ and H is the distribution function introduced in Theorem 3 in Section 2.2.1.

2.2.3 Ledford and Tawn joint tail model

Ledford and Tawn (1996) prove that, under broad conditions, the joint survivor function of (X, Y), with common Fréchet margins, satisfies the asymptotic condition

$$\Pr(X > z, Y > z) \sim \mathcal{L}(z)\{\Pr(X > z)\}^{1/\eta} \text{ as } z \to \infty, \tag{2.2.10}$$

where $\mathcal{L}(z)$ is a slowly varying function at infinity (i.e. $\mathcal{L}(tz)/\mathcal{L}(z) \to 1$ as $z \to \infty$ for all fixed $t > 0$), and $\eta \in (0, 1]$ is the coefficient of the tail dependence. If $\eta = 1$ and $\mathcal{L}(z) \to c$ as $z \to \infty$, with $0 < c \leqslant 1$, then $\chi = c$ and $\bar{\chi} = 1$, and the variables are asymptotically dependent of degree c. If $\eta < 1$, then $\chi = 0$ and $\bar{\chi} = 2\eta - 1$, and X and Y are asymptotically independent. If $0.5 < \eta < 1$ the variables are positively associated; $\eta = 0.5$ occurs under near independence, and $0 < \eta < 0.5$ suggests that the variables are negatively associated.

Inference

The easiest way to estimate η is to transform the problem to a one-dimensional form by defining $T = \min(X, Y)$, where X and Y have Fréchet margins. If $\mathcal{L}(t)$ is approximated as

a constant c for $t > u$, then

$$\Pr(T > t) = \Pr(X > t, Y > t) \sim \frac{c}{t^{1/\eta}} \text{ for } t > u, \tag{2.2.11}$$

by the joint tail model given in (2.2.10). Other non-constant forms for $\mathcal{L}(t)$ will be explored in Chapter 3, where we show that these give similar η estimates. Hence, we find that the approximation of $\mathcal{L}(t)$ as a constant c is not very restrictive. Note that in expression (2.2.11), η is in fact the shape parameter of a generalised Pareto tail fitted to T. Hence, we can use univariate methods by fitting a generalized Pareto distribution to T and estimating η as the shape parameter.

Another method to estimate η is by using the approximation to the joint tail probability given in (2.2.11), but instead of using the generalised Pareto fit, we can construct a likelihood and use this for estimation. Using a 'censored' likelihood approach we obtain the likelihood as

$$\begin{aligned} L(\eta, c) &= \left\{ \prod_{i=1}^{n_u} f_T(t_i) \right\} \Pr(T \leqslant u)^{n-n_u}, \\ &= \left\{ \prod_{i=1}^{n_u} \frac{c}{\eta t_i^{1/\eta+1}} \right\} \left(1 - \frac{c}{u^{1/\eta}}\right)^{n-n_u}, \end{aligned} \tag{2.2.12}$$

where n_u is the number of points above the threshold u, and $t_1, \ldots, t_{n_u}$ are the exceedances of this threshold. Estimates of c and η can be obtained numerically or analytically by maximizing the likelihood. Analytically, we obtain the following estimates:

$$\begin{aligned} \widehat{\eta} &= \min\left(\frac{1}{n_u} \sum_{i=1}^{n_u} \log\left(\frac{t_i}{u}\right), 1 \right), \\ \widehat{c} &= \frac{n_u}{n} u^{1/\widehat{\eta}}. \end{aligned} \tag{2.2.13}$$

Note that $\widehat{\eta}$ coincides with the Hill estimator (Hill, 1975). For numerical maximisation of the likelihood in (2.2.12), the constraint $0 \leqslant c \leqslant u^{1/\eta}$ must be observed. Further parameter

constraints and extensions for the Ledford and Tawn model will be given in Section 2.4. We will also present and compare other estimators for the coefficient of tail dependence η in Chapter 3.

2.2.4 Ramos and Ledford model

Ramos and Ledford (2009) obtain an asymptotically motivated bivariate model for (X, Y) when both marginal variables are simultaneously large. They derive a general form for the joint survivor model for (X, Y) with Fréchet margins, an example of which is the following:

$$\bar{F}_{XY}(x,y) = \frac{\lambda u^{1/\eta}}{N_\rho}\left[(\rho x)^{-1/\eta} + \left(\frac{y}{\rho}\right)^{-1/\eta} - \left\{(\rho x)^{-1/\alpha} + \left(\frac{y}{\rho}\right)^{-1/\alpha}\right\}^{\alpha/\eta}\right], \qquad (2.2.14)$$

for $(x, y) \in [u, \infty)\times[u, \infty)$, where u is a high threshold and λ is the joint threshold exceedance probability $\Pr(X > u, Y > u)$, and $N_\rho = \rho^{-1/\eta}+\rho^{1/\eta}-(\rho^{-1/\alpha}+\rho^{1/\alpha})^{\alpha/\eta}$ with $0 < \alpha, \rho, \eta \leqslant 1$.

This model has four key parameters, namely η, ρ, α and λ. As before, η is the coefficient of tail dependence, with (X, Y) asymptotically dependent for $\eta = 1$, and asymptotically independent if $\eta < 1$. The parameter ρ is the so-called asymmetry parameter, with $\rho = 1$ corresponding to the symmetric case. The relative size of the α parameter to η defines the dependence between the angular $X/(X+Y)$ and radial $(X+Y)$ re-parametrisation of (X,Y). Lastly, the parameter λ is the joint threshold exceedance probability. For a fixed threshold u, λ can be estimated empirically by $\widehat{\lambda} = \Pr(X > u, Y > u)$, independently of the other three parameters.

2.2.5 Heffernan and Tawn conditional model

The conditional tail model of Heffernan and Tawn (2004) is a semi-parametric approach that allows for both asymptotic dependence and asymptotic independence. Most other multivariate models focus on the case when all variables are extreme, whereas the conditional

model is appropriate also when only a subset of the variables are extreme. The model can be applied to problems of higher dimensions, but here we will only consider the bivariate case.

We will initially follow the derivation of the model from Heffernan and Tawn (2004). Consider random variables (X, Y) with Gumbel marginal distributions and examine the limiting behaviour of the conditional distribution $\Pr(Y \leqslant y|X = x)$, with X being the conditioning margin as $x \to \infty$. We assume that there exist normalising functions $a(x)$ and $b(x) > 0$, which can be chosen such that for all fixed z and for any sequence of x-values such that $x \to \infty$,

$$\lim_{x\to\infty} \Pr(Z \leqslant z|X = x) = G(z), \tag{2.2.15}$$

for

$$Z = \frac{Y - a(x)}{b(x)}, \tag{2.2.16}$$

where the limit distribution G is non-degenerate.

Under assumption (2.2.15), we have that, conditionally on $X > u$, as $u \to \infty$ the variables $X - u$ and Z are independent in the limit with their limiting marginals being exponential and $G(z)$, respectively.

The class of limit distributions is unique up to type, and the normalising functions $a(x)$ and $b(x)$ can be identified up to the constants A and B in $a^*(x) = a(x) + Ab(x)$ and $b^*(x) = Bb(x)$, where if $a(x)$ and $b(x)$ give a non-degenerate distribution $G(z)$, then the normalising functions $a^*(x)$ and $b^*(x)$ give the non-degenerate limit $G(Bz + A)$.

Keef et al. (2013) find that this form of the conditional model runs into difficulties when modelling variables with some components positively associated and some negatively associated. This is due to the choice of the Gumbel distribution for the marginal distribution in which to apply the model. To overcome this problem, Keef et al. (2013) suggest the use of the Laplace marginal distributions instead. So X and Y both need to be transformed to

have marginal distributions of the form

$$\Pr(X < x) = \begin{cases} \exp(x)/2 & \text{if } x < 0, \\ 1 - \exp(-x)/2 & \text{if } x \geqslant 0, \end{cases}$$

which ensures that both the lower and the upper tails of X and Y are exponentially distributed. Now assume that there exist vector-valued normalising functions, $a(x)$ and $b(x) > 0$, such that for $x > 0$

$$\Pr\left(X - u > x, \frac{Y - a(X)}{b(X)} \leqslant z \mid X > u\right) \to \exp(-x)G(z) \quad \text{as} \quad u \to \infty, \tag{2.2.17}$$

where G is a non-degenerate distribution function. Heffernan and Resnick (2007) show that the norming functions are regularly varying functions, and Heffernan and Tawn (2004) derive the following specific finite form:

$$a(x) = \alpha x \quad \text{and} \quad b(x) = x^{\beta},$$

with $(\alpha, \beta) \in [-1, 1] \times [-\infty, 1]$, which they show holds widely. This representation covers both positive and negative association, with $0 < \alpha \leqslant 1$ corresponding to positive dependence and $-1 \leqslant \alpha < 0$ corresponding to negative dependence. The parameters α and β take different values for different dependence structures. Asymptotic independence is implied when $\alpha = \beta = 0$ and asymptotic dependence (with positive association) when $\alpha = 1$ and $\beta = 0$. Assuming that the limit in (2.2.17) holds for finite u, the model takes the form:

$$Y = \alpha X + X^{\beta} Z, \;\; X > u,$$

where Z is a random variable with distribution function G and is independent of X.

Details about inference and simulation from the Heffernan and Tawn model will be discussed in Chapter 3.

2.3 Spatial extremes

Environmental applications are often spatial in nature so benefits can be gained by using spatial methods. For example, we can improve inference for a single location by using data from the spatial neighbourhood around the location. Spatial extremes also allows us to quantify the joint risk of multiple locations being affected by a single extreme event.

Most spatial extremes models can capture only one of the two classes of extremal dependence described in Section 2.2. The most widely studied and used spatial extremes models are in the family of max-stable processes, which can model asymptotic dependence or perfect independence. Asymptotically independent models include Gaussian processes and inverted max-stable processes (Wadsworth and Tawn, 2012a), but these are not widely adopted in practice due to a general preference for conservativeness (i.e. overestimation of extremes is considered safer than underestimation). The hybrid model of Wadsworth and Tawn (2012a) is capable of modelling both asymptotic dependence and independence, but its complexity makes it difficult to use in practice. We will touch upon all of these models in the subsequent chapters of this thesis (Chapters 3 and 4 in particular), but since max-stable processes are the most widely used, we will focus our attention on them for the rest of this section.

For a more comprehensive review of methods for spatial extremes, the reader is referred to Davison et al. (2012).

2.3.1 Max-stable processes

The random process $Z(t)$ is called max-stable on Ω, if for each $k = 1, 2, \ldots$, there exist continuous functions $a_k(t) > 0$ and $b_k(t)$ such that for any function $z(t)$,

$$\Pr\{Z(t) \leqslant a_k(t)z(t) + b_k(t), t \in \Omega\}^k = \Pr\{Z(t) \leqslant z(t), t \in \Omega\},$$

i.e. $Z(t)$ and the maximum of k independent copies of $\{Z(t) - b_k(t)\}/a_k(t)$ have the same distribution. A consequence of this property is that $\{Z(t_1), \ldots, Z(t_m)\}$ follows a multivariate extreme value distribution for all $t_1, \ldots, t_m \in \Omega$ and all m.

All max-stable processes are asymptotically dependent over all distances, i.e. $\eta = 1$, unless they are perfectly independent. Here we will show this for the bivariate extreme value (BEV) distribution. The BEV distribution function takes the form:

$$G(x, y) = \exp\{-V(x, y)\}. \tag{2.3.1}$$

Hence it follows that,

$$\begin{aligned}
\Pr(X > x, Y > x) &= 1 - 2\exp\left\{-\frac{1}{x}\right\} + \exp\{-V(x, x)\}, \\
&= 1 - 2\exp\left\{-\frac{1}{x}\right\} + \exp\left\{-\frac{V(1,1)}{x}\right\}, \\
&\sim \frac{2 - V(1,1)}{x}, \quad \text{as } x \to \infty,
\end{aligned}$$

where the second line follows due to the homogeneity property of V as shown in (2.2.3). By comparison with (2.2.11), $\eta = 1$, unless $V(1,1) = 2$, which only occurs when X and Y are perfectly independent. Hence, indeed the bivariate extreme value distribution is asymptotically dependent.

In the following, we will introduce four commonly used classes of max-stable processes. For the sake of simplicity of notation, here we will only present joint distributions for two-dimensional max-stable models. Higher order joint distributions have been studied by Wadsworth and Tawn (2014) and Genton et al. (2015).

Smith process

Following Smith (1990), consider a stochastic process $\{Z(t)\}$, with $t \in \Omega$ for some arbitrary index set Ω. Without loss of generality we may assume that $Z(t)$ has standard Fréchet

margins:

$$\Pr(Z(t) \leqslant z) = e^{-1/z}, \forall t \in \Omega.$$

Smith defines $\{(\xi_i, s_i), i \geqslant 1\}$ as the points of a Poisson process on $(0, \infty) \times S$, with intensity measure $\xi^{-2} d\xi \times \nu(ds)$, where S is an arbitrary measurable set and ν is a positive measure on S. Furthermore, let $\{f(s,t), s \in S, t \in \Omega\}$ denote a non-negative function for which

$$\int_S f(s,t)\nu(ds) = 1, \forall t \in \Omega.$$

Then Smith defines $Z(t)$ as:

$$Z(t) = \max_i \{\xi_i f(s_i, t)\}, t \in \Omega,$$

where $f(s,t)$ is a multivariate normal density with mean s and covariance matrix Σ,

$$f(s,t) = f_0(s-t) = (2\pi)^{-d/2} |\Sigma|^{-1} \exp\left\{-\frac{1}{2}(s-t)^T \Sigma^{-1} (s-t)\right\}, \tag{2.3.2}$$

and takes $\nu(ds) = ds$.

To motivate this representation from a practical point of view, we can think of s_i as storm centres in Ω, with magnitude ξ_i, distributed over space according to a Poisson process, so centres are uniformly distributed with intensity ds and sizes decay with a density $1/\xi^2$. The function f represents the 'shape' of the storm. Hence, $\xi_i f(s_i, t)$ is the size of the storm at position t from a storm of size ξ_i centred at location s_i. Figure 2.3.1 shows three realisations of the process in one dimension. The underlying events are shown in black and the pointwise maximum of these is taken to obtain a one-dimensional realisation from the Smith process (shown in red).

It can be shown that this choice of $Z(t)$ is max-stable and that it has unit Fréchet margins for any $t \in \Omega$ (see Smith (1990)). Figure 2.3.2 shows two simulations from the two-dimensional Smith model with different covariance matrices. The realisation shown on the left panel is

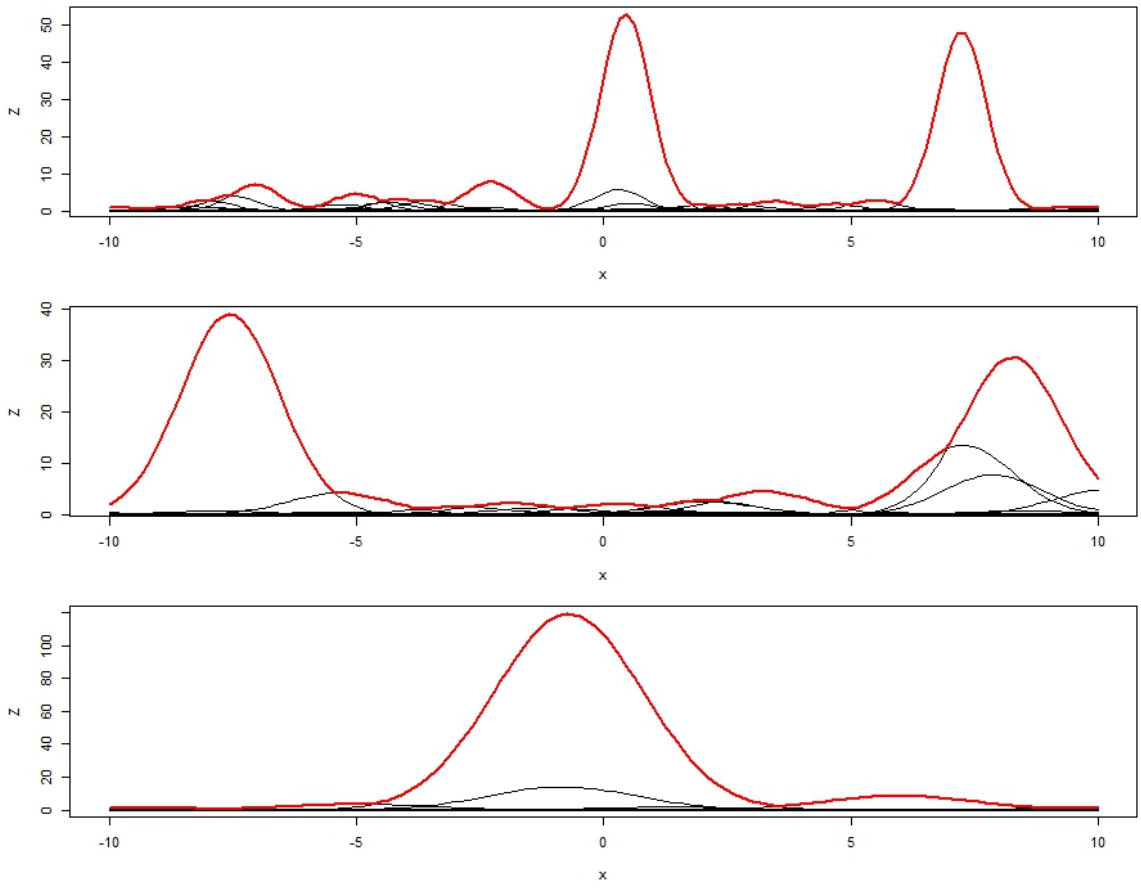


Figure 2.3.1: Three one-dimensional realisations from the Smith process (top: $\sigma = 0.5$, middle: $\sigma = 1$, bottom: $\sigma = 1.5$), with the red line being the pointwise maximum, and the black lines the underlying events.

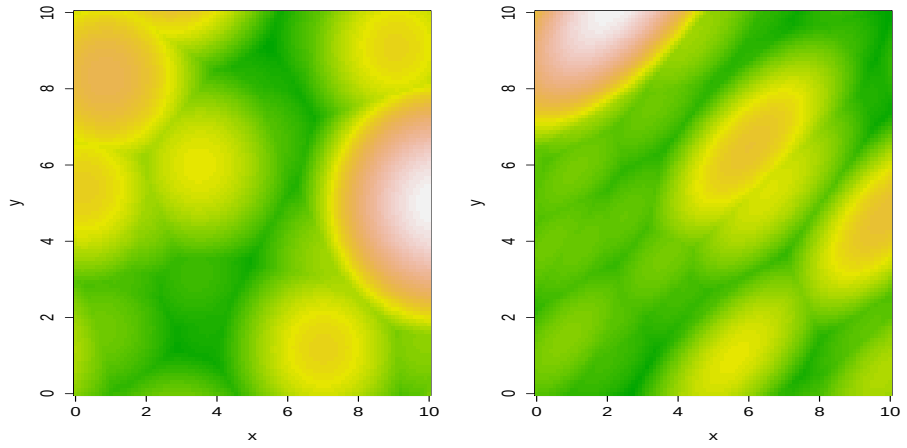


Figure 2.3.2: Two simulations from the Smith model with different covariance matrices on a 100×100 grid. Left panel: $\sigma_{11} = \sigma_{22} = 1.5$ and $\sigma_{12} = 0$; right panel: $\sigma_{11} = \sigma_{22} = 1.5$ and $\sigma_{12} = 1$.

isotropic whereas the one on the right is not. Both realisations are very smooth due to the underlying Gaussian densities.

For two sites ($d = 2$), say the origin $\mathbf{o}$ and a location $\mathbf{h}$, the joint distribution is given by

$$\Pr(Z(\mathbf{o}) \leqslant z_1, Z(\mathbf{h}) \leqslant z_2) = \exp\{-V_{\mathbf{h}}(z_1, z_2)\}$$

where $V_{\mathbf{h}}$ is the exponent measure defined as

$$V_{\mathbf{h}}(z_1, z_2) = z_1^{-1}\Phi\left\{\frac{a(\mathbf{h})}{2} + a^{-1}(\mathbf{h})\log\left(\frac{z_2}{z_1}\right)\right\} + z_2^{-1}\Phi\left\{\frac{a(\mathbf{h})}{2} + a^{-1}(\mathbf{h})\log\left(\frac{z_1}{z_2}\right)\right\}, \quad (2.3.3)$$

where Φ is the standard normal distribution function, and $a(\mathbf{h}) = \sqrt{\mathbf{h}^T\Sigma^{-1}\mathbf{h}}$ is the Mahalanobis distance between $\mathbf{h}$ and the origin. Higher order joint distributions become increasingly complicated.

The extremal coefficient for the Smith model is $\theta(\mathbf{h}) = 2\Phi\{a(\mathbf{h})/2\}$, for which $\theta(\mathbf{h}) \to 2$ as $\|\mathbf{h}\| \to \infty$, and $\theta(\mathbf{h}) \to 1$ as $\|\mathbf{h}\| \to 0$, spanning the range of possible asymptotic dependencies.

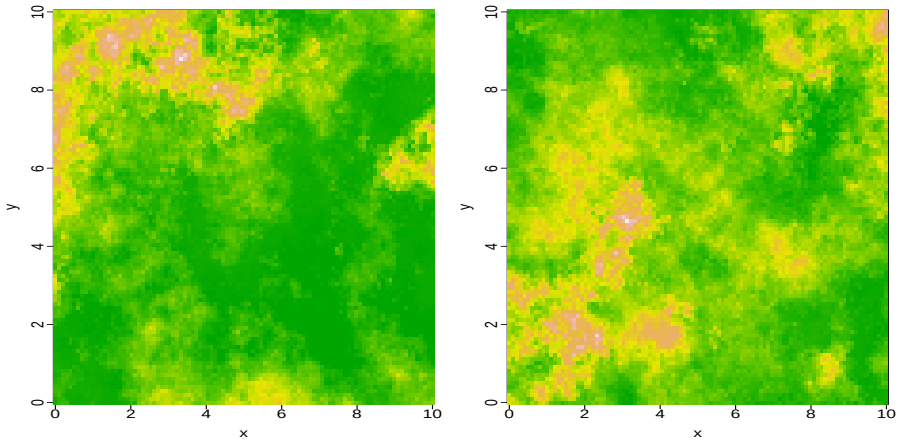


Figure 2.3.3: Two simulations from the Schlather model with different range parameters on a 100×100 grid. Left panel: range = 10; right panel: range = 3.

Schlather process

Following Schlather (2002), let $\{S_j\}_{j=1}^{\infty}$ be the points of a Poisson process on $\mathbb{R}_+$ with intensity ds/s^2. Let $\{W_j(x)\}_{j=1}^{\infty}$ be independent replicates of a stationary Gaussian process $W(x)$ on $\mathbb{R}^d$ satisfying $\mathbb{E}[\max\{0, W_j(o)\}] = 1$, where o denotes the origin. Then Schlather defines

$$Z(x) = \max_j S_j \max\{0, W_j(x)\}. \tag{2.3.4}$$

and proves that $Z(x)$ is a stationary max-stable process on $\mathbb{R}^d$ with unit Fréchet marginals.

The exponent measure for this model in the bivariate case is given by

$$V_{\mathbf{h}}(z_1, z_2) = \frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right) \times \left(1 + \left[1 - 2\frac{(\rho(\mathbf{h}) + 1)z_1 z_2}{(z_1 + z_2)^2}\right]^{1/2}\right),$$

where ρ is a valid correlation function. The most commonly used correlation functions are isotropic, i.e. $\rho(\mathbf{h}) = \rho(\|\mathbf{h}\|)$, and include the Whittle-Matern, Cauchy and powered exponential correlation functions. As for the Smith model, higher order forms are difficult to express analytically. Figure 2.3.3 shows two realisations of the two-dimensional Schlather process with powered exponential correlation functions with different range parameters. Note that the Schlather model realisations are less smooth than the Smith realisations.

The extremal coefficient is $\theta(\mathbf{h}) = 1 + \{[1 - \rho(\mathbf{h})]/2\}^{1/2}$. Because of the requirement that $\rho(\mathbf{h})$ be a positive definite function for $\mathbf{h} \in \mathbb{R}^2$, and $W_j(x)$ be stationary and isotropic, $\theta(\mathbf{h}) < 1.838$. So this means that the model cannot account for extremes that become independent when $\|\mathbf{h}\| \to \infty$.

Brown-Resnick process

Let $\epsilon(x)$ be an isotropic fractional Brownian process with semivariogram $\gamma(h) \propto \|\mathbf{h}\|^{\alpha}$, $0 < \alpha \leqslant 2$ and $\epsilon(0) = 0$ almost surely. Then $W(x)$ in (2.3.4) can be taken as:

$$W(x) = \exp\{\epsilon(x) - \gamma(x)\}.$$

This process was introduced by Brown and Resnick (1977). When ϵ is a Brownian process and $\alpha = 2$, $W(x)$ corresponds to the Smith model. The bivariate extremal coefficient is $\theta(\mathbf{h}) = 2\Phi\{a(\|\mathbf{h}\|)/2\}$, as for the Smith process, but a takes a different value ($a = \{2\gamma(\|h\|)\}^{1/2}$). Therefore, $\theta(\|\mathbf{h}\|) \to 2$ as $\|\mathbf{h}\| \to \infty$, so the process captures complete independence for large distances.

Extremal-t process

The extremal-t process was first proposed by Demarta and McNeil (2005) and it assumes the following representation of $W(x)$ in (2.3.4):

$$W(x) = \sqrt{\pi} 2^{-\nu/2+1} \Gamma\left(\frac{\nu+1}{2}\right)^{-1} \max\{\epsilon(x), 0\}^{\nu},$$

where $\mu \geqslant 1$, Γ is the Gamma function, and $\epsilon(x)$ is a Gaussian random field with mean zero and correlation function $\rho(\mathbf{h})$. The case when $\nu = 1$ corresponds to the Schlather process. The bivariate extremal coefficient is $\theta(\mathbf{h}) = 2T_{\nu+1}(\sqrt{(\nu+1)[1-\rho(\mathbf{h})]/[1+\rho(\mathbf{h})]})$, where T_{ν} denotes the cumulative distribution function of a student-t random variable with

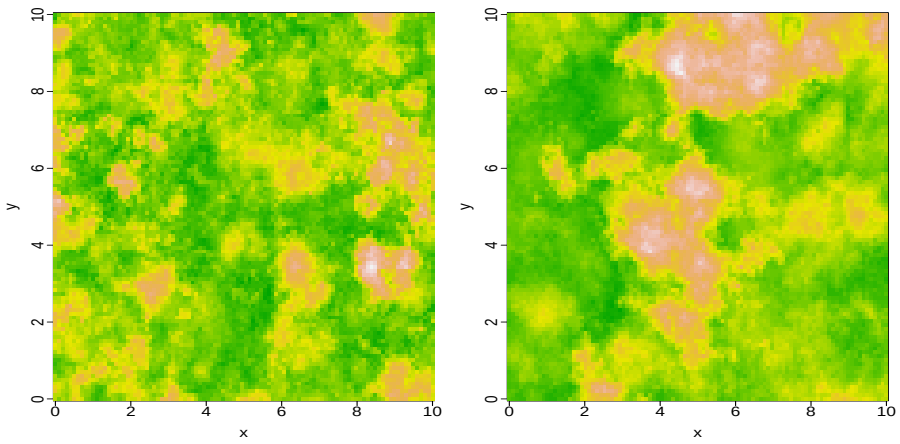


Figure 2.3.4: Two simulations from the extremal-t model with different range parameters on a 100×100 grid. Left panel: range = 1; right panel: range = 3.

ν degrees of freedom, and correlation function $\rho(\mathbf{h})$. Figure 2.3.4 shows two realisations of the extremal-t process with different dependence structures. These realisations appear similar in roughness to the simulations from the Schlather proces.

2.4 Extensions

In this section we will present two extensions to the multivariate approaches presented in Section 2.2. The first one builds on the idea of including marginal information in the Ledford and Tawn joint tail model introduced in Winter (2015). The second one seeks to use the bias correction method introduced in Fougères et al. (2015) to reduce the bias in estimating the coefficient of tail dependence η.

2.4.1 Including marginal information in the Ledford and Tawn model

The censored likelihood in (2.2.12) only uses information about the points that are above the threshold in both margins, i.e. in the region $\{X > u, Y > u\}$. Winter (2015) shows that some efficiency can be gained in the bivariate case by incorporating marginal information about points that are above the threshold in one margin but not the other, i.e. in the regions

$\{X > u, Y \leqslant u\}$ and $\{X \leqslant u, Y > u\}$.

The two examples we will use in this section are the bivariate logistic extreme value distribution (2.2.5) and the inverted bivariate logistic extreme value distribution, which has the following distribution function:

$$F(x,y) = e^{-1/x} + e^{-1/y} - 1 + \exp\left\{-\left[\{-\log(1-e^{-1/x})\}^{1/\gamma} + \{-\log(1-e^{-1/y})\}^{1/\gamma}\right]^{\gamma}\right\}, \tag{2.4.1}$$

where $\gamma \in (0, 1]$.

Figure 2.4.1 shows the main results of the simulation study in Winter (2015). It is clear that there is a considerable reduction in the root mean square error of the χ estimate when including marginal information in the Ledford and Tawn bivariate joint tail model. Here we extend this approach to three dimensions.

In the bivariate case, the following constraints must be observed. As $0 \leqslant \Pr(T \leqslant u) \leqslant 1$, where $T = \min(X, Y)$ and has survivor function given in expression (2.2.11), it follows that $0 \leqslant c \leqslant u^{1/\eta}$. In addition, the conditional probability of $Y > u$ given that $X > u$ must also be in $[0, 1]$;

$$\begin{aligned}\Pr(Y > u | X > u) &= \frac{\Pr(Y > u, X > u)}{\Pr(X > u)} \\ &= \frac{\Pr(T > u)}{\Pr(X > u)} \\ &= \frac{c}{u^{1/\eta - 1}}.\end{aligned}$$

Hence, we obtain the following constraint on the parameter space:

$$0 < c \leqslant u^{1/\eta - 1}. \tag{2.4.2}$$

As this constraint is stronger, it is sufficient to enforce this instead of $0 \leqslant c \leqslant u^{1/\eta}$.

Now in the three-dimensional case, consider the random variables X_1, X_2 and X_3 on common

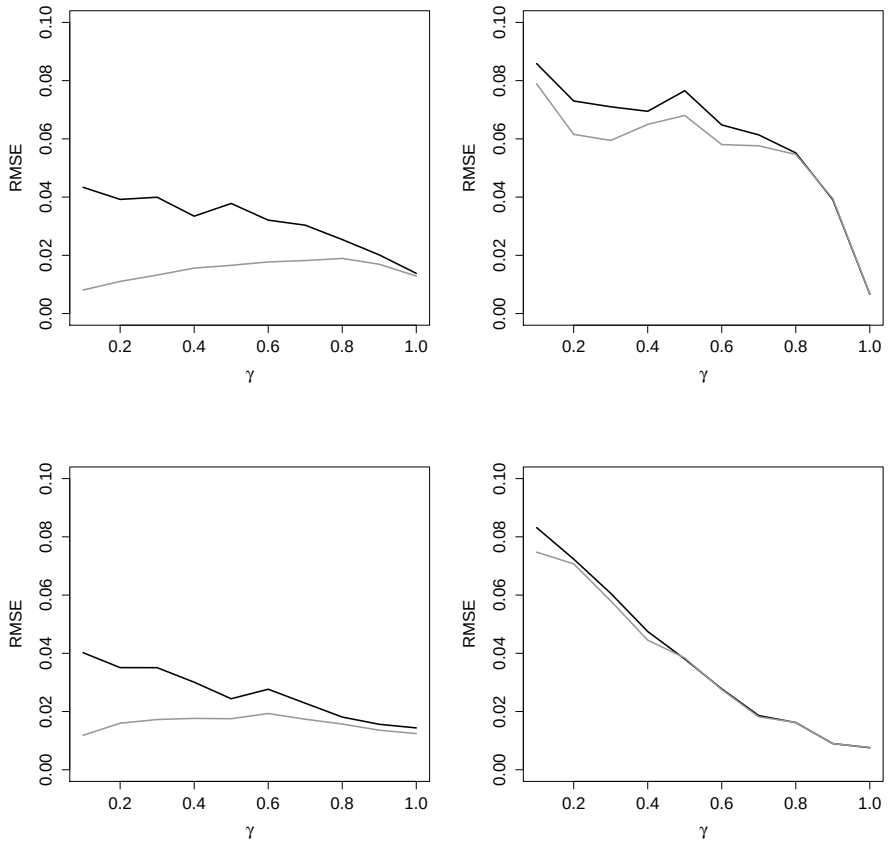


Figure 2.4.1: Estimates of the RMSE of $\widehat{\chi}(v)$ for the Ledford and Tawn approach (black) and the Ledford and Tawn approach with marginal information (grey). Simulated samples are from the bivariate logistic extreme value distribution (top row) and the inverted bivariate logistic extreme value distribution (bottom row). The critical level v is set at the 90% quantile (left) and the 99% quantile (right).

Pareto margins, i.e. $\Pr(X_i > t) = 1/t$ for $i = 1, 2, 3$. Define the following regions for some high threshold u:

$$R_{000} = \{X_1 \leqslant u, X_2 \leqslant u, X_3 \leqslant u\} \qquad R_{011} = \{X_1 \leqslant u, X_2 > u, X_3 > u\}$$
$$R_{001} = \{X_1 \leqslant u, X_2 \leqslant u, X_3 > u\} \qquad R_{101} = \{X_1 > u, X_2 \leqslant u, X_3 > u\}$$
$$R_{010} = \{X_1 \leqslant u, X_2 > u, X_3 \leqslant u\} \qquad R_{110} = \{X_1 > u, X_2 > u, X_3 \leqslant u\}$$
$$R_{100} = \{X_1 > u, X_2 \leqslant u, X_3 \leqslant u\} \qquad R_{111} = \{X_1 > u, X_2 > u, X_3 > u\}$$

such that the whole three-dimensional space is divided into these nine regions. Then, the probability of falling into region R_{111} is given by the expression in (2.2.11) as before, except now $T = \min(X_1, X_2, X_3)$. The probability of falling in the other regions can be derived by the use of inclusion-exclusion arguments. For example,

$$\begin{aligned}\Pr(X_1 \leqslant u, X_2 \leqslant u, X_3 > u) &= \Pr(X_3 > u) - \Pr(X_1 > u, X_3 > u) \\ &\quad - \Pr(X_2 > u, X_3 > u) + \Pr(X_1 > u, X_2 > u, X_3 > u), \\ &= \frac{1}{u} - \frac{c_{12}}{u^{1/\eta_{12}}} - \frac{c_{23}}{u^{1/\eta_{23}}} + \frac{c}{u^{1/\eta}},\end{aligned}$$

where c_{ij} and η_{ij}, $i, j = 1, 2, 3$, $i \neq j$, represent the parameters of the bivariate models, and c and η the parameters of the three-dimensional model. The probabilities of falling in the other regions can be derived similarly. Let us denote the number of points falling in the region R_{000} by n_{000}, the number of points falling in the region R_{001} by n_{001} and so on. Then

the we can write down the likelihood as follows,

$$
\begin{aligned}
L(\boldsymbol{\theta}) = & \left(1 - \frac{3}{u} + \frac{c_{12}}{u^{1/\eta_{12}}} + \frac{c_{13}}{u^{1/\eta_{13}}} + \frac{c_{23}}{u^{1/\eta_{23}}} - \frac{c}{u^{1/\eta}}\right)^{n_{000}} \\
& \times \left(\frac{1}{u} - \frac{c_{13}}{u^{1/\eta_{13}}} - \frac{c_{23}}{u^{1/\eta_{23}}} + \frac{c}{u^{1/\eta}}\right)^{n_{001}} \\
& \times \left(\frac{1}{u} - \frac{c_{12}}{u^{1/\eta_{12}}} - \frac{c_{23}}{u^{1/\eta_{23}}} + \frac{c}{u^{1/\eta}}\right)^{n_{001}} \\
& \times \left(\frac{1}{u} - \frac{c_{12}}{u^{1/\eta_{12}}} - \frac{c_{13}}{u^{1/\eta_{13}}} + \frac{c}{u^{1/\eta}}\right)^{n_{100}} \\
& \times \left(\frac{c_{23}}{u^{1/\eta_{23}}} - \frac{c}{u^{1/\eta}}\right)^{n_{011}} \\
& \times \left(\frac{c_{13}}{u^{1/\eta_{13}}} - \frac{c}{u^{1/\eta}}\right)^{n_{101}} \\
& \times \left(\frac{c_{12}}{u^{1/\eta_{12}}} - \frac{c}{u^{1/\eta}}\right)^{n_{110}} \\
& \times \prod_{i=1}^{n_{111}} \left(\frac{c}{\eta t_i^{1+1/\eta}}\right),
\end{aligned}
\tag{2.4.3}
$$

where $\boldsymbol{\theta} = (c_{12}, c_{23}, c_{13}, c, \eta_{12}, \eta_{23}, \eta_{13}, \eta)$. The following parameter constraints must be observed, $\forall i, j \in \{1, 2, 3\},\ i \neq j$:

(i) $0 \leqslant c_{ij} \leqslant u^{1/\eta_{ij}-1}$ and $0 \leqslant c \leqslant u^{1/\eta-1}$,

(ii) $c/u^{1/\eta} \leqslant c_{ij}/u^{1/\eta_{ij}}$,

(iii) $1/u - c_{13}/u^{1/\eta_{13}} - c_{23}/u^{1/\eta_{23}} + c/u^{1/\eta} \geqslant 0$,

(iv) $1/u - c_{12}/u^{1/\eta_{12}} - c_{23}/u^{1/\eta_{23}} + c/u^{1/\eta} \geqslant 0$,

(v) $1/u - c_{12}/u^{1/\eta_{12}} - c_{13}/u^{1/\eta_{13}} + c/u^{1/\eta} \geqslant 0$,

(vi) $0 < \eta \leqslant \eta_{ij} \leqslant 1$.

Constraint (i) is the three-dimensional equivalent of the constraint (2.4.2). Constraints (ii) and (iii)-(v) ensure that the terms inside the round brackets in lines 5-7 and lines 2-4, respectively, of the expression in (2.4.3) are positive. Finally, constraint (vi) ensures that $\Pr(\min(X_1, X_2, X_3) > x) \leqslant \Pr(\min(X_i, X_j) > x)$ for all $x > 0$.

To obtain parameter estimates for $\boldsymbol{\theta}$ we can maximise the likelihood given in (2.4.3). We can either do this in one step and estimate all eight parameters at once, or we can take a two-step approach by fitting three bivariate models first to estimate c_{ij} and η_{ij} for $i, j = 1, 2, 3$, $i \neq j$,

and then maximising the likelihood with these parameters fixed to obtain the maximum likelihood estimates of c and η.

Next we will examine whether this approach gives any additional benefits over the approach with no marginal information. To do this we conduct a simulation study similar to the one carried out in Winter (2015). We simulate 200 replicate samples of size 5000 from a three-dimensional extreme value distribution with a logistic dependence structure (MEVL) and from the inverted version of this distribution (IMEVL). For each of the samples we estimate c and η using (i) the approach with no marginal information, (ii) the two-step approach including marginal information, and (iii) the one-step approach including marginal information. Then we can estimate $\chi(v) = \Pr(X_1 > v, X_2 > v \mid X_3 > v)$ as

$$\widehat{\chi}(v) = \widehat{c}/v^{1/\widehat{\eta}-1}.$$

Note that $\Pr(X_1 > v, X_2 > v \mid X_3 > v) = \Pr(X_1 > v, X_3 > v \mid X_2 > v) = \Pr(X_2 > v, X_3 > v \mid X_1 > v)$. To compare the methods with and without marginal information we calculate the root mean squared error (RMSE) of the estimated $\chi(v)$ values for each method for various critical values v:

$$\text{RMSE}[\chi(v)] = \sqrt{\text{E}\{[\widehat{\chi}(v) - \chi_{true}(v)]^2\}}$$

For the MEVL the true $\chi(v)$ can be derived to first order for large v as

$$\chi_{true}(v) \sim 3 - 3(2^{\gamma}) + 3^{\gamma},$$

and for the IMEVL as

$$\chi_{true}(v) \sim v^{1-3^{\gamma}},$$

where γ is the dependence parameter of the MEVL and IMEVL. Figure 2.4.2 shows the RMSE estimates for both models at two different critical levels. It seems that the approach

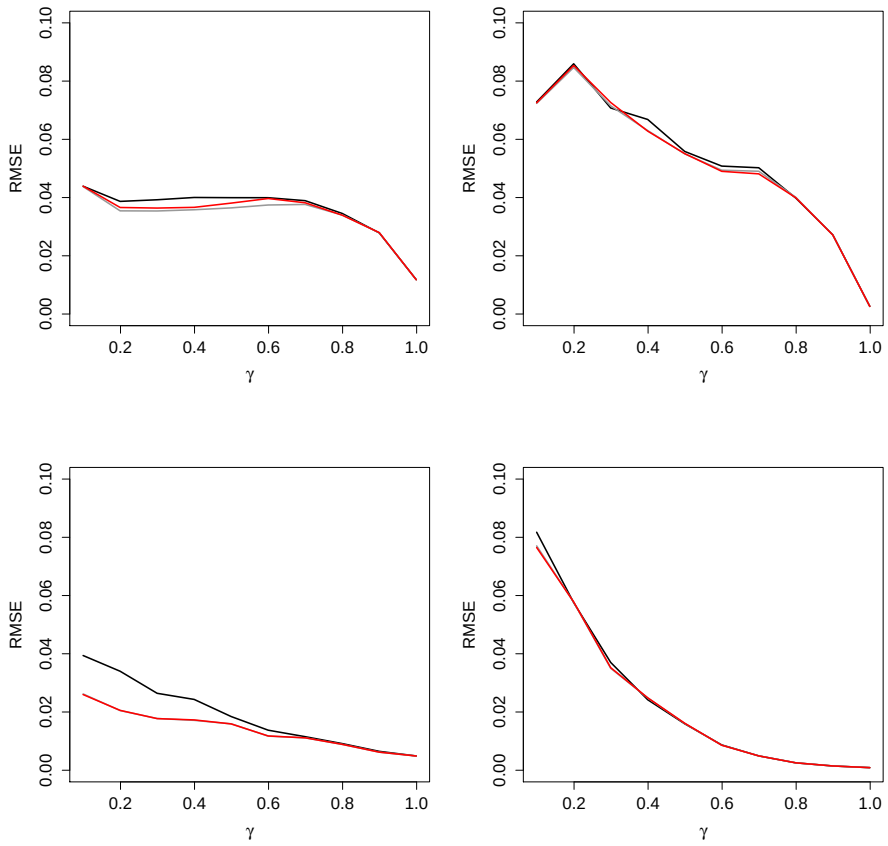


Figure 2.4.2: Estimates of the RMSE of $\widehat{\chi}(v)$ for the Ledford and Tawn approach (black) and the Ledford and Tawn approach with marginal information (grey for two-step estimation, red for one-step estimation). Simulated samples are from the MEVL distribution (top row) and the IMEVL distribution (bottom row). The critical level v is set at the 90% quantile (left) and the 99% quantile (right).

including marginal information reduces the RMSE slightly, especially for the lower critical value of $v = 90\%$. There is negligible difference between the one-step and two-step approaches.

In the three-dimensional case we could also define $\chi(v)$ as the probability of one of (X_1, X_2, X_3) being greater than v conditionally on the other two being above v, e.g. $\Pr(X_1 > v \mid X_2 > v, X_3 > v)$. We have also tested estimating this $\chi(v)$ using marginal information, but since the RMSE estimates were only very slightly reduced compared to not including marginal information, we do not include these results here.

We conclude that although including marginal information provides some efficiency gains in the bivariate case (as shown in Winter (2015) and on Figure 2.4.1), the benefits decrease in the three-dimensional case. Hence, we think that on balance the additional complexity of the model is not worthwhile beyond the bivariate case.

2.4.2 Bias correction using the approach of Fougères et al. (2015)

Estimation of the extremal dependence structure is biased, with the bias increasing with the number of points used for the estimation. Fougères et al. (2015) suggest a method for correcting the bias in estimating the stable tail dependence function, which in the bivariate case is

$$L(x_1, x_2) = \lim_{t\to\infty} t \Pr\{1 - F_1(X^{(1)}) \leqslant t^{-1}x_1 \text{ or } 1 - F_2(X^{(2)}) \leqslant t^{-1}x_2\}, \tag{2.4.4}$$

where $(X^{(1)}, X^{(2)})$ is a bivariate vector with continuous marginal distributions F_1 and F_2. The bivariate stable tail dependence function $L(x_1, x_2)$ is equivalent to $V(1/x_1, 1/x_2)$ where V is the exponent measure defined in (2.2.4). An empirical estimator of $L(\mathbf{x})$ (where $\mathbf{x} = (x_1, x_2)$) is then obtained as

$$\widehat{L}_k(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left\{X_i^{(1)} \geqslant X_{n-[kx_1]+1,n}^{(1)} \text{ or } X_i^{(2)} \geqslant X_{n-[kx_2]+1,n}^{(2)}\right\}, \tag{2.4.5}$$

where $X_{k,n}^{(j)}$ is the kth order statistics among n realisations of the margin $X^{(j)}$, with $k/n \to 0$ for $n \to \infty$. For large n, $\widehat{L}_k(\mathbf{x})$ is a biased estimator of the stable tail dependence function $L(\mathbf{x})$.

Fougères et al. (2015) suggest the following bias correction method to give an unbiased estimator. Let us denote

$$\widehat{L}_{k,a}(\mathbf{x}) := a^{-1}\widehat{L}_k(a\mathbf{x}), \;\; 0 < a < 1, \tag{2.4.6}$$

and

$$\Delta_{k,a}(\mathbf{x}) := \widehat{L}_{k,a}(\mathbf{x}) - \widehat{L}_k(\mathbf{x}), \tag{2.4.7}$$

where a is a positive scale parameter which allows the contraction or dilation of the observed data points. Then, the following asymptotically unbiased estimator of $L(\mathbf{x})$ can be derived:

$$\mathring{L}_{k,1,k_\rho}(\mathbf{x}) := \widehat{L}_k(\mathbf{x}) - \Delta_{k,2^{-1/\widehat{\rho}}}(\mathbf{x}), \tag{2.4.8}$$

and, more generally,

$$\mathring{L}_{k,a,k_\rho}(\mathbf{x}) := \widehat{L}_{k,a}(\mathbf{x}) - \Delta_{k,(a^{-\widehat{\rho}}+1)^{-1/\widehat{\rho}}}(\mathbf{x}), \tag{2.4.9}$$

with a fixed as $a^{-\rho} = 2$, where ρ is a second order parameter estimated as $\widehat{\rho}$ and k_ρ is an intermediate sequence that represents the number of order statistics used in the estimator $\widehat{\rho}$. The authors derive the following estimator for ρ:

$$\widehat{\rho}_{k,a,r}(\mathbf{x}) := \left(1 - \frac{1}{\log r} \log\left\{\frac{\Delta_{k,a}(r\mathbf{x})}{\Delta_{k,a}(\mathbf{x})}\right\}\right), \tag{2.4.10}$$

and they suggest fixing $a = r = 0.4$. The a and k_ρ parameters in (2.4.9) have also been tuned and can be fixed to $a = 0.4$ and $k_\rho = 990$.

It is possible to avoid estimating ρ by using a combination of estimators. Fougères et al. (2015) derive the following alternative estimator:

$$\tilde{L}_{k,a,k_\rho}(\mathbf{x}) := \frac{\widehat{L}_k(\mathbf{x})\Delta_{k_\rho,a}(a\mathbf{x}) - \widehat{L}_k(a\mathbf{x})\Delta_{k_\rho,a}(\mathbf{x})}{\Delta_{k_\rho,a}(a\mathbf{x}) - a\Delta_{k_\rho,a}(\mathbf{x})}. \tag{2.4.11}$$

We now test these bias correction methods on the bivariate logistic extreme value (BVEL) distributions which has stable tail dependence function

$$L(x_1, x_2) = \left(x_1^{1/\gamma} + x_2^{1/\gamma}\right)^\gamma \quad 0 < \gamma \leqslant 1.$$

Fougères et al. (2015) have shown that the estimators in expressions (2.4.9) and (2.4.11) give good results in estimating the stable tail dependence function for a number of theoretical examples, including the BVEL. Our interest lies in estimating the second order tail dependence structure, represented by the coefficient of tail dependence η, introduced in (2.2.10). The stable tail dependence function can be used to estimate η in the following way. We can write the bivariate stable tail dependence function as

$$\begin{aligned} L(tx,tx) &= t\Pr(X > tx \text{ or } Y > tx), \\ &= t[\Pr(X > tx) + \Pr(Y > tx) - \Pr(X > tx, Y > tx)], \\ &= \frac{2}{x} - \frac{c}{x^{1/\eta} t^{1/\eta - 1}}. \end{aligned} \tag{2.4.12}$$

Then, setting $x = 1$ in (2.4.12) and rearranging we have

$$2 - L(t,t) = \frac{c}{t^{1/\eta - 1}}. \tag{2.4.13}$$

Similarly, setting $x = y$, for some $y > 0$, in (2.4.12) and rearranging we have

$$2 - L(ty,ty) = \frac{c}{y^{1/\eta} t^{1/\eta - 1}}. \tag{2.4.14}$$

Then dividing (2.4.14) by (2.4.13) and rearranging we obtain the following estimator for η:

$$\widehat{\eta} = \log(y) \Big/ \log\left(\frac{2 - L(t,t)}{2/y - L(ty,ty)}\right). \tag{2.4.15}$$

We will now test the performance of this estimator on samples from the bivariate logistic extreme value distribution (BEVL) and its inverted version (IBEVL), i.e. bivariate versions of the MEVL and and IMEVL distributions we used in Section 2.4.1. For L in expression (2.4.15) we use the empirical estimator $\widehat{L}$ given in (2.4.5), the unbiased estimator $\mathring{L}$ in expression (2.4.9) and the alternative unbiased estimator $\tilde{L}$ in (2.4.11). We compare the performance of these estimators against the Hill estimator given in (2.2.13). For asymptotic

independence, Fougères et al. (2015) derive that $\eta = 1/(1-\rho)$ so we also include the estimator

$$\widehat{\eta}_\rho = 1/(1-\widehat{\rho}), \tag{2.4.16}$$

where $\widehat{\rho}$ was estimated using (2.4.10).

Figure 2.4.3 shows the performance of this estimator on 100 BEVL (top row) and 100 IBEVL (bottom row) samples of size 1000 with $\gamma = 0.2$ (i.e. relatively strong dependence). We set y equal to a value near 1, and set t such that on the left hand side plots 50 points (out of 1000) were used for estimation, whereas on the right hand side 100 points. Note that the true η for the BEVL is 1 and for the IBEVL is $2^{-\gamma}$, which is less than 1 for all values $\gamma > 0$, and this is captured by the Hill estimator well. The estimator $\widehat{\eta}_\rho$ performs reasonably well for the IBEVL samples, but poorly for the BEVL samples, as expected. Even for IBEVL the estimator $\widehat{\eta}_\rho$ is much more variable than the Hill estimator, and so is inefficient in terms of RMSE relative to $\widehat{\eta}_{Hill}$. The three estimators based on the stable tail dependence function all vastly underestimate η and perform significantly worse than $\widehat{\eta}_{Hill}$. Hence, the bias correction introduced in Fougères et al. (2015) does not seem to capture the second order structure of the tail dependence of the data well.

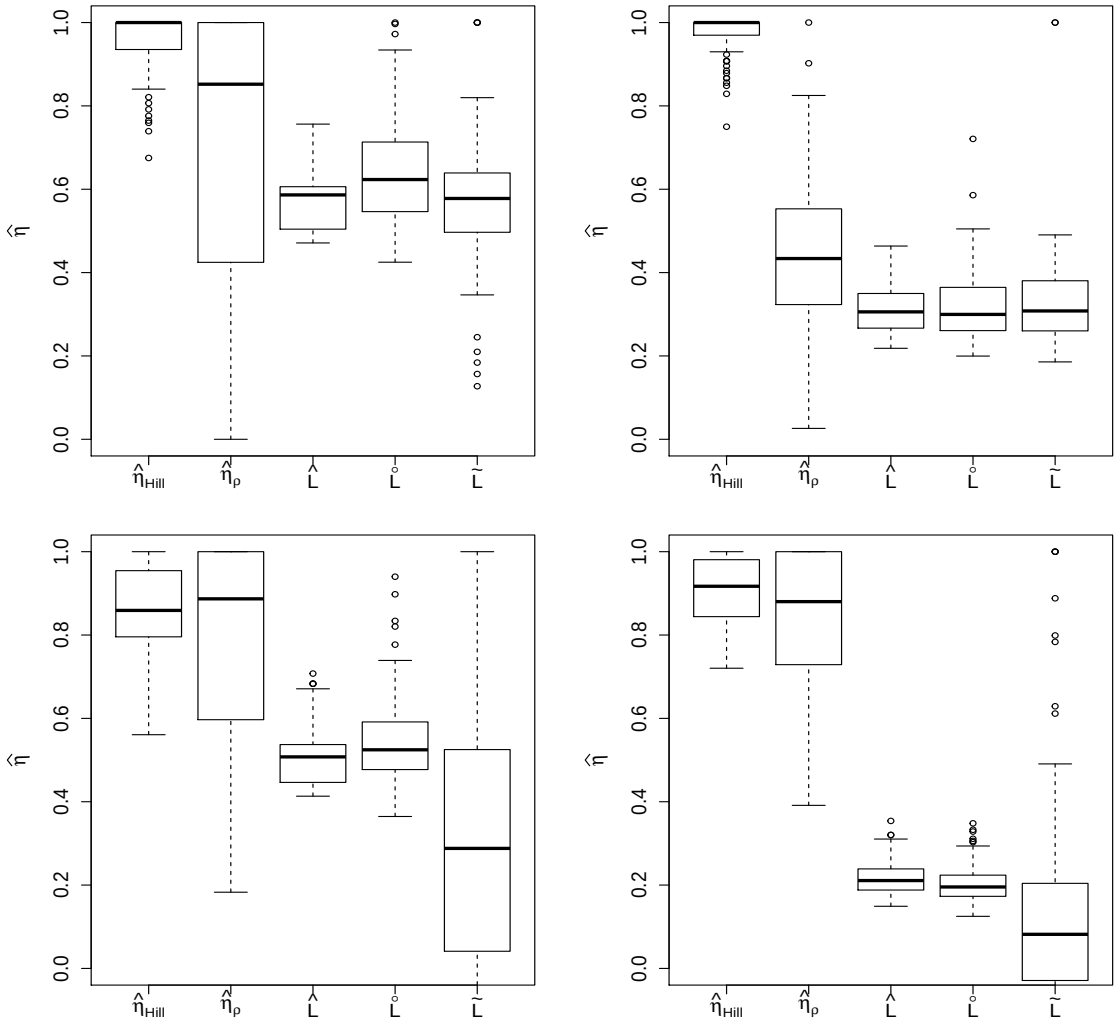


Figure 2.4.3: Comparison of estimators for η for 100 BEVL (top row) and 100 IBEVL (bottom row) samples of size 1000 with $\gamma = 0.2$. On the left hand side 50 points (out of 1000) were used for estimation, whereas on the right hand side 100 points. The estimators used are: $\widehat{\eta}_{Hill}$ in (2.2.13), $\widehat{\eta}_{\rho}$ in (2.4.16), and expression (2.4.15) with L estimated by the empirical estimator $\widehat{L}$ in (2.4.5), the unbiased estimator $\mathring{L}$ in (2.4.9) and the alternative unbiased estimator $\tilde{L}$ in (2.4.11).

Chapter 3

Assessing Extremal Dependence of North Sea Storm Severity

3.1 Introduction

Metocean and coastal engineers are generally interested in the estimation of design conditions for a specific spatial location. Of primary concern is the estimation of return values for wind-, wave- and current-related variables corresponding to a long return period. However, there are instances when the engineer is concerned with characterising joint occurrences of rare events at different locations or multiple hazard occurrence at a single location. Examples of the former type might include (i) risk to multiple offshore facilities from a single cyclone or hurricane event, (ii) impact of a single storm surge event on multiple coastal locations, and (iii) insurance risk to a moving vessel traversing an ocean basin. Examples of the latter type include the combined risk of wind-, wave- and current-induced forces on a fixed structure, or motions of a floating structure. In such circumstances, a description of the joint structure of extreme events is necessary. This chapter develops diagnostic methods for assessing the appropriate form of dependence structure to model metocean data sets and explores the implications of getting this decision right and wrong in application to significant

wave height at neighbouring locations.

To help formulate the issues, consider the simplified example of two identically distributed random variables X and Y. For example, the random variables X and Y could represent significant wave height at neighbouring locations. We are typically interested in cases where both variables are large together, i.e., we need to find the probability $\Pr\{X > x, Y > x\}$ for large x, where x could correspond to the 50-year return level of each variable marginally. For large x, this joint probability will be small, so for studying dependence between extremes, it is typical to focus on the conditional probability $\Pr\{Y > x | X > x\}$, recognising that $\Pr\{X > x, Y > x\}$ can be recovered from this conditional probability by multiplying by the marginal probability $\Pr\{X > x\}$. Two special cases of dependence structure between (X, Y) correspond to the variables being independent (which we term perfect independence) or perfectly dependent (i.e., $X = Y$ always). Then,

$$\Pr\{Y > x | X > x\} = \begin{cases} \Pr\{Y > x\} & \text{if perfect independence} \\ 1 & \text{if perfect dependence.} \end{cases} \tag{3.1.1}$$

As x tends to the upper endpoint x^* of the common marginal distribution of X and Y, the limiting probability χ then will satisfy

$$\chi = \lim_{x \to x^*} \Pr\{Y > x | X > x\} = \begin{cases} 0 & \text{if perfect independence} \\ 1 & \text{if perfect dependence.} \end{cases} \tag{3.1.2}$$

The chance of joint occurrence of extremes is completely different in these cases. We cannot assume observations of variables, such as significant wave height at different locations, to be perfectly independent nor perfectly dependent of one another. An occurrence of an extreme event at one location influences the chance of an extreme event at a location in its neighbourhood, but even locations very close by can experience different conditions. The most well known dependence model is the multivariate normal distribution, or equivalently

the Gaussian copula, which is widely used across fields of empirical modelling, including spatial interpolation using kriging or Gaussian process regression. Consideration of the properties of the bivariate normal distribution with correlation parameter ρ_P (where we use the subscript P to denote the Pearson correlation coefficient) suggests an alternative form for $\Pr\{Y > x|X > x\}$ for large x, which is intermediate to perfect dependence and perfect independence for all $0 < \rho_P < 1$, but has $\chi = 0$.

Hence, to model tail dependence reliably we need models which will account for the different possibilities for how the conditional probability $\Pr\{Y > x|X > x\}$, and its extensions for our metocean examples, vary for large x. Clearly, models which capture intermediate forms between perfect independence and perfect dependence are required, and we will explain that these correspond to two broad classes, with either $0 < \chi < 1$ or $\chi = 0$, termed asymptotic dependence and asymptotic independence, respectively. For asymptotic independence, it is impossible for the most extreme events for both variables (X, Y) to occur simultaneously. For asymptotic dependence, if X is extreme it is also possible for Y to be simultaneously extreme. In theory, there are also cases where $\Pr\{Y > x|X > x\}$ decays to 0 faster than $\Pr\{Y > x\}$ as $x \to x^*$, corresponding to negative dependence between (X, Y), but for our metocean examples this typically does not occur. Hence, the key focus in the chapter will be on positively dependent variables.

As we are interested in joint probabilities for large events, it is natural to look for asymptotic arguments that provide guidance regarding the possible form that these probabilities can take. Extreme value theory provides this framework for one variable. For example, although the distribution of a continuous random variable can take any form, subject to weak convergence conditions, above a sufficiently high threshold, peaks over threshold of this variable follow the generalised Pareto distribution approximately (Pickands, 1975). This motivates the adoption of a specific tail functional form which can be estimated, and from which extrapolations to long return period events can be obtained. For metocean data this has been found to provide a reliable method for return level estimation of a single variable.

Furthermore, the parameters of this distribution may be allowed to vary with one or more covariates, such as direction or season (e.g., Jonathan et al. 2008).

Unfortunately, analogous specific asymptotic results are not available from extreme value theory for the joint distribution of the extremes of two or more variables. For this reason, a number of different approaches, none of which is completely satisfactory, have emerged for modelling multivariate and spatial extremes. In the applied statistics literature popular methodologies are motivated by consideration of componentwise maxima and regular variation (e.g., Coles and Tawn 1994; Cooley et al. 2012; Davis et al. 2013; Davison et al. 2012; de Haan and de Ronde 1998). These lead to tractable models for joint extremes. However, they require the assumption of asymptotic dependence between extremes of pairs of random variables, i.e, $\chi > 0$ (as defined in (3.1.2)).

Suppose that X and Y represent significant wave height at two different locations $\boldsymbol{r}_X$ and $\boldsymbol{r}_Y$, respectively. Then if the distance $d_{XY} = |\boldsymbol{r}_Y - \boldsymbol{r}_X|$ between these locations is small relative to the spatial extent of an ocean storm, the assumption of $\chi > 0$ may well be valid. However, as d_{XY} increases, the assumption appears increasingly untenable, since the same physical cause (i.e., a low-pressure field causing wind causing waves) is increasingly unlikely to produce simultaneous extreme values of significant wave height at both locations despite there being some dependence between observed values of X and Y. Thus, it seems reasonable for these d_{XY} that $\chi = 0$ corresponding to asymptotic independence. If we continue to increase d_{XY}, eventually the locations are so distant that the characteristics of X and Y cannot possibly be influenced by the same physical cause. In this limit, we can say that X and Y exhibit perfect independence of extremes.

The difference between joint extremes estimated under asymptotic dependence and asymptotic independence assumptions can be very large. For example, for the storm-peak significant wave height from a typical location in the North Sea application that we present in Section 3.5, marginal extreme value analysis suggests that the 100-year return level is approximately 15.5m and that this does not change much over neighbouring sites. If a 100 year

event occurs at one location, the median of the conditional distribution at a neighbouring location will also be approximately 15.5m, under strong asymptotic dependence. However, under the assumption of an asymptotically independent Gaussian process model with correlation parameter 0.8, the median of the conditional distribution at the neighbouring location reduces to approximately 13.5m (see Section 3.5.3 for details). In fact all quantiles of the conditional distribution reduce by approximately 2m; a reduction of 2m in the conditional distribution is of considerable practical importance.

The majority of applicable models for multivariate and spatial extremes assume asymptotic dependence, leading potentially to over-estimation of the risk of joint occurrences of extremes of X and Y if asymptotic dependence is incorrectly assumed. On the other hand, if asymptotic independence is assumed when the data are in fact asymptotically dependent, the risk of the extremes will be underestimated. Conservatism in design is desirable, provided that the extent of conservatism is itself reasonable and well bounded. For extreme dependence modelling this conservatism may be unbounded. For instance, if the variables are falsely assumed to be asymptotically dependent (with $\chi > 0$) when they were actually perfectly independent, then $\Pr\{X > x, Y > x\}$ can be over-estimated by the factor $\chi / \Pr\{X > x\} \to \infty$ as $x \to x^*$. This is an issue, as over-conservatism in some aspects of design can lead to increased risks in other aspects, particularly when resources are constrained. If asymptotic dependence is not a reasonable assumption for a particular application, alternative approaches admitting asymptotic independence must be used. The conditional extremes model of Heffernan and Tawn (2004) and the hybrid model of Wadsworth and Tawn (2012a) are possibilities, but the field of modelling asymptotically independent spatial extremes is considerably less well developed in general.

These risks of over- and under-design illustrate that we must justify the specific assumptions made concerning extremal dependence in any metocean application. Therefore reliable diagnostic methods for identifying between the asymptotic independence and asymptotic dependence classes of extremal dependence based on a sample of data are required. Tawn

and co-workers (e.g., Ledford and Tawn 1997; Heffernan and Tawn 2004; Wadsworth and Tawn 2013b) have developed the description of asymptotic independence, and have offered different characterisations. A number of diagnostic statistics for the form of extremal dependence have been proposed (see e.g., Coles et al. 1999) and applied in metocean settings (see e.g., Eastoe et al. 2013). In ideal situations, these provide good guidance regarding the nature of extremal dependence in play in a particular application. However, for a typical metocean application, where sample size might be small, the distributions of X and Y vary with respect to multiple covariates, and are observed only with error (either from a physical hindcast model or from observation), such diagnostics tend to be inconclusive.

This chapter attempts to refine the way diagnostics for extremal dependence are employed in practice, to improve their interpretability. We develop improved diagnostics for differentiating between asymptotic independence and dependence, which leads to increased assurance in model selection. In turn, we hope this facilitates better understanding of the dependence of extremes of ocean environmental variables, so that joint design conditions may be estimated with improved confidence. A key component of our method is the novel idea to supplement measures of extremal dependence with a measure of the dependence for the body of the data. This enables the assessment of extremal dependence in the light of the general dependence and we find that it noticeably increases diagnostic performance.

The layout of the chapter is as follows: in Section 3.2 we summarise measures of asymptotic independence and illustrate their value for a range of the widely used extremal models; in Section 3.3 we explore a range of inference methods to aid the diagnostic process; in Section 3.4 we conduct a simulation study to illustrate the performance of the diagnostic methods for a range of distributions and samples sizes; and in Section 3.5 we apply the methods to North Sea significant wave height data; show that the tail dependence changes with wave direction and distance between spatial locations; and infer that in many cases the assumption of asymptotic independence is more plausible, and in all cases that the presence of asymptotic independence cannot be dismissed. In Section 3.6 we conclude with

a discussion.

3.2 Extremal dependence

3.2.1 Different dependence measures for typical and extreme value data

In standard statistical analysis the Pearson correlation coefficient ρ_P is often used to determine the dependence between two variables. A weakness of this measure is that it is not invariant to non-linear monotone increasing transformations of the marginal variables. For oceanographic applications this means we have different correlations between wave heights at two locations and their associated wave powers. To overcome this issue we use Spearman's rank correlation coefficient ρ (Spearman, 1904), but we could have also used the broadly equivalent Kendal τ. In the context of oceanographic variables it is reasonable to interpret a value of ρ close to 0 (and 1) as indicating near independence (and near perfect dependence) respectively.

Spearmans rank correlation measures dependence for the body of the distribution, i.e., typical values, as all observations are given equal weighting in the assessment of dependence. When dealing with extremes, the dependence can no longer be measured using any form of correlation estimate, as the dependence in the body can be quite different from the extremal or tail dependence. This property is illustrated in Sections 3.4 and 3.5. A nice illustration of this feature arises in time series: a Gaussian autoregressive process has temporal dependence as measured by the autocorrelation function but no tail dependence as measured by χ (Sibuya, 1960); whereas the reverse holds for the ARCH and GARCH processes (de Haan and Resnick (1989), Laurini and Tawn (2012)).

Therefore in Section 3.2.2 we introduce some measures of dependence structure specific to extreme values and illustrate these on different probabilistic models for spatial processes. As there is limited information in the extreme values, there is typically not sufficient evidence

from these extremal dependence measures for them to be effective diagnostic tools when identifying the form of the extremal dependence, e.g., whether the processes the data are observed on are asymptotically dependent or asymptotically independent. Therefore, the novel approach we take in this chapter is to supplement these measures of extremal dependence with ρ, measuring dependence of typical values. The combined information from these two dependence measures improves diagnostic performance for extremal dependence, as ρ provides a context in which to assess estimated values of extremal dependence.

3.2.2 Measures of Extremal Dependence

There exist several summary statistics for extremal dependence, see Coles et al. (1999), Beirlant et al. (2004) and Wadsworth and Tawn (2013b). Here we focus on four measures which identify different characteristics of the joint tail for asymptotically independent and asymptotically dependent distributions.

In Section 3.1 we introduced χ, with $0 \leq \chi \leq 1$, as a measure of extremal dependence for variables which are identically distributed. This measure holds more generally when X and Y have marginal distribution functions F_X and F_Y, and associated inverses F_X^{-1} and F_Y^{-1}, respectively, with

$$\chi = \lim_{q \to 1} \Pr\{Y > F_Y^{-1}(q) | X > F_X^{-1}(q)\} = \lim_{q \to 1} \Pr\{F_Y(Y) > q | F_X(X) > q\}. \tag{3.2.1}$$

The interpretation of χ is as in Section 3.1. From an oceanographic context this corresponds to the the limit of the probability that one variable exceeds its T year return level given that the other variable exceeds its T year return level as $T \to \infty$, thus there is a greater risk for metocean designs the larger the value of χ.

All measures of extremal dependence can be specified on general and different marginal forms, like for χ, but it is often more simple mathematically to present them for identical marginal distributions; compare expressions (3.1.2) and (3.2.1). The concept of copulas

(Nelsen, 2006), shows there is no loss of generality in presenting extremal measures for a specific marginal distribution choice. Specifically when presenting the measure for variables $\tilde{X}$ and $\tilde{Y}$ with common marginal distribution choice F, then we need to transform (X, Y) to $(\tilde{X}, \tilde{Y})$, by

$$\tilde{X} = F^{-1}\{F_X(X)\} \text{ and } \tilde{Y} = F^{-1}\{F_Y(Y)\}. \tag{3.2.2}$$

To apply this transformation in practice, we estimate F_X and F_Y using the empirical distribution functions. Alternatively, we could replace the empirical distributions in the upper tail with a generalised Pareto distribution (Coles and Tawn, 1991). Various measures are more simple to present mathematically for a specific marginal choice, which then dictate the choice of common distribution function F. Specifically in this chapter we use Pareto margins, $F(x) = 1 - x^{-1}$ for $x \geq 1$ with $(\tilde{X}, \tilde{Y})$ then denoted by (X_P, Y_P), Fréchet margins, $F(x) = \exp(-x^{-1})$ for $x > 0$ with $(\tilde{X}, \tilde{Y})$ then denoted by (X_F, Y_F), and Laplace margins, $F(x) = \exp(x)/2$ for $x \leqslant 0$ and $1 - \exp(-x)/2$ for $x > 0$ with $(\tilde{X}, \tilde{Y})$ then denoted by (X_L, Y_L). Note that $\Pr\{X_F > x\} = 1 - \exp(-x^{-1}) = x^{-1} + x^{-2} + O(x^{-3}) \approx x^{-1} = \Pr\{X_P > x\}$ for large x. Hence, the marginal tails of the Pareto and Fréchet distributions are approximately equivalent and so these two margins can be used inter-changeably when focusing on the upper tail.

The measure χ has limitations: all asymptotically independent variables have $\chi = 0$; it does not provide information about the dependence at levels of interest, e.g., at the 50 year return level; and it is hard to estimate as it involves an extrapolation. In particular, there could be stronger dependence at the 50 year level for asymptotically independent variables than for asymptotically dependent variables. Therefore we need a more refined measure than χ. The second extremal dependence measure $\chi(x)$ is presented for Pareto marginal variables, with

$$\chi(x) = \Pr\{Y_P > x | X_P > x\} \text{ for } x \geqslant 1 \tag{3.2.3}$$

and $\chi = \lim_{x\to\infty} \chi(x)$. On this scale x corresponds to the return period on the marginal

scale, so $x = 50$ is the one in every 50 observations return level. It follows from Section 3.1 that for positively dependent variables $x^{-1} \leqslant \chi(x) \leqslant 1$, the bigger the value of $\chi(x)$, for large x, the stronger the level of extremal dependence.

A complication with $\chi(x)$ is that it is a function over $x > 1$ and so not parsimonious. Ideally what is required is a formulation for how $\chi(x)$ behaves for large x. Ledford and Tawn (1996) prove that under weak conditions

$$\Pr\{X_P > x, Y_P > x\} \sim \mathcal{L}(x)x^{-1/\eta} \text{ for } x \geqslant 1, \tag{3.2.4}$$

where $\eta \in (0, 1]$ is termed the coefficient of the tail dependence, and $\mathcal{L}$ is any positive slowly varying function at infinity, i.e.,

$$\lim_{t\to\infty} \mathcal{L}(tx)/\mathcal{L}(t) = 1, \tag{3.2.5}$$

provided the derivative with respect to x of the right hand side of expression (3.2.4) is non-negative for all $x \geqslant 1$. This formulation gives that

$$\chi(x) = \mathcal{L}(x)x^{-(1-\eta)/\eta} \text{ for } x \geqslant 1. \tag{3.2.6}$$

From expression (3.2.6) the relevance of η is apparent. When the variables (X_P, Y_P) are perfectly independent (perfectly dependent) expression (3.1.1) shows that $\eta = \frac{1}{2}$ and $\mathcal{L}(x) = 1$ ($\eta = 1$ and $\mathcal{L}(x) = 1$), respectively. More generally, when $\eta = 1$ and $\mathcal{L}(x) \to c$ with $0 < c \leq 1$, it follows that $\chi = c$ and the variables are asymptotically dependent. Furthermore, $\chi(x)$ converges to c slower than any power of x for large x, and hence, for practical cases, can be viewed as constant for large x, i.e., $\chi(x) = c$ for $x > u$ for some large u. When $0 < \eta < 1$ then $\chi(x) \to 0$ as $x \to \infty$, so $\chi = 0$ and the variables are asymptotically independent. However, the value of η provides additional information about the level of asymptotic independence as η describes the rate of convergence to 0, and therefore is key in determining the difference

between $\chi(x)$ and χ. Specifically $\frac{1}{2} < \eta < 1$ corresponds to positive extremal dependence and $0 < \eta < \frac{1}{2}$ to negative extremal dependence. An example when $0 < \eta < 1$ occurs is for the Gaussian dependence structure, with $\eta = (1 + \rho_P)/2$, where ρ_P is the Pearson correlation coefficient, so $0 < \rho_P < 1$ implies $\frac{1}{2} < \eta < 1$, i.e., all non-perfectly dependent Gaussian dependence structure variables are asymptotically independent.

Although the characterisation of Ledford and Tawn (1996) provides a more concrete formulation for the structure of asymptotic independence, it is restricted to situations where both variables are large. A more general formulation, based on slightly different assumptions, has been proposed by Heffernan and Tawn (2004) and found to have substantial uses in metocean applications (see e.g., Jonathan et al. 2013 and Eastoe et al. 2013). The formulation is most apparent in Laplace margins, i.e., we have (X_L, Y_L), with the connection for large values of these variables with the Pareto variable case being that $X_L \approx \log(X_P/2)$ and $Y_L \approx \log(Y_P/2)$. Under weak assumptions, a combination of Heffernan and Tawn (2004) and Heffernan and Resnick (2007) show that there exist values α and β, with $-1 \leqslant \alpha \leqslant 1$, $0 \leqslant \beta < 1$, for $x > 0$ and $-\infty < z < \infty$, such that

$$\lim_{u\to\infty} \Pr\left\{ \frac{Y_L - \alpha X_L}{X_L^{\beta}} < z, X_L - u > x \middle| X_L > u \right\} = \exp(-x)G(z), \tag{3.2.7}$$

where G is a non-degenerate distribution function. It follows that α, β and G jointly determine the form of extremal dependence. Here, if $\alpha = 1$ and $\beta = 0$ the variables are asymptotically dependent with $\chi = \int_0^{\infty} [1 - G(-t)] \exp(-t)dt > 0$, but otherwise $\chi = 0$. Specifically, the form of asymptotic independence is given by

$$\begin{aligned} \chi(x) &= \Pr\{Y_L > x_L | X_L > x_L\} \\ &= \int_0^{\infty} \left[1 - G\left(\frac{x_L - \alpha(x_L + t)}{(x_L + t)^{\beta}}\right)\right] \exp(-t)dt, \end{aligned} \tag{3.2.8}$$

where $x_L = \log(x/2)$. For statistical purposes, Heffernan and Tawn (2004) assume that limit (3.2.7) holds exactly for a large finite u. As a consequence, conditionally on $X_L > u$,

$X_L - u$ and $(Y_L - \alpha X_L)/X_L^{\beta}$ are independent random variables, and conditionally on X_L, with $X_L > u$, we have

$$Y_L = \alpha X_L + X_L^{\beta} Z, \tag{3.2.9}$$

where Z has mean μ and variance σ^2, with distribution function G. It follows that $E[Y_L \mid X_L] = \alpha X_L + X_L^{\beta}\mu$ and $\text{Var}(Y_L|X_L) = X_L^{2\beta}\sigma^2$ for $X_L > u$. Then maximum likelihood methods can be used, under a working assumption of normality for G, to estimate $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\mu}$ and $\widehat{\sigma}$. Rearranging (3.2.9) gives

$$\widehat{Z} = \frac{Y_L - \widehat{\alpha} X_L}{X_L^{\widehat{\beta}}}, \text{for } X_L > u,$$

and using the empirical distribution of such $\widehat{Z}$'s gives an estimate of the distribution G, i.e., we estimate the distribution of Z using all pairs (X_L, Y_L) with $X_L > u$.

3.2.3 Extremal Dependence Models

Introduction

We will consider three classes of spatial processes: max-stable process, Gaussian processes and inverted max-stable processes. Max-stable processes can be asymptotically dependent or perfectly independent only. In contrast, Gaussian processes and inverted max-stable processes are asymptotically independent always. The max-stable family is the most studied asymptotically dependent model, but it has no finite parametrisation. To illustrate some of the max-stable process features we will present a range of parametric sub-models that are widely used in applications. For each of the three classes of processes we will present the dependence measures χ, $\chi(x)$ for large x, η and α and β to reveal different features of their dependence structures. These processes are also used in Section 3.4 in a simulation study for assessing the performance of the estimation methods of Section 3.3.

Max-stable Processes

Max-stable processes arise as the class of marginally non-degenerate limit distributions for componentwise maxima of independent and identically distributed replicates of a spatial process. For such processes all finite dimensional distributions are multivariate extreme value distributions (Beirlant et al. (2004), Ch. 8). Max-stable fields are typically presented via the following constructive approach (see e.g., de Haan (1984); Schlather (2002), de Haan and Ferreira (2006), Ch. 9).

For a max-stable process on the space S, let $\{r_i\}$, $i = 1, \ldots$, be the points of a Poisson process on $(0, \infty)$, with intensity dr, and define the spectral function to be $\{W(s) \geq 0; s \in S\}$, a continuous random function on S, satisfying $E\{W(s)\} = 1$ for all $s \in S$. If the processes $W_i(\cdot)$ for $i = 1, \ldots$ are independent and identically distributed copies of $W(\cdot)$, then

$$X_F(s) = \max_i W_i(s)/r_i \text{ for } s \in S$$

is a max-stable random field with unit Fréchet margins and d-dimensional distribution function

$$\Pr\{X_F(s_1) \leq x_1, \ldots, X_F(s_d) \leq x_d\} = \exp\left(E\left[\max\left\{\frac{W(s_1)}{x_1}, \ldots, \frac{W(s_d)}{x_d}\right\}\right]\right)$$

for $x_i > 0$, $i = 1, \ldots, d$.

To motivate this representation from a practical point of view, we can view the ith event having magnitude $1/r_i$, and event profile/shape over space S of $W_i(\cdot)$, so the ith event for the process is $W_i(s)/r_i$ for $s \in S$. Thus, $X_F(s)$ is the componentwise maxima of these events at location $s \in S$. The dependence structure of the max-stable process is determined by the form of the exponent function, V,

$$V(\mathbf{x}) = E\left[\max\left\{\frac{W(s_1)}{x_1}, \ldots, \frac{W(s_d)}{x_d}\right\}\right],$$

with $\mathbf{x} = (x_1, \ldots, x_d)$ for all $\mathbf{x}$ and all d. Here $1 \leq V(\mathbf{x}) \leq d$, with the lower and upper bounds achieved when the variables are perfectly dependent and perfectly independent respectively. Key to the properties of max-stable processes is homogeneity of order -1 of V, i.e., $V(n\mathbf{x}) = n^{-1}V(\mathbf{x})$ for all n and $\mathbf{x}$. Thus

$$\Pr\{X_F(s)/n \leq x(s), s \in S\}^n = \Pr\{X_F(s) \leq x(s), s \in S\},$$

i.e., $X_F(\cdot)$ and the maximum of n independent copies of $X_F(\cdot)/n$ have the same distribution.

All max-stable processes are pairwise asymptotically dependent or perfectly independent. Dropping the index notation for sites, we have the joint survivor function of $(X_F, Y_F) = (X_F(s_1), X_F(s_2))$ as

$$\begin{aligned}\Pr\{X_F > x, Y_F > x\} &= 1 - 2\exp(-1/x) + \exp\{-V(x,x)\} \\ &= 1 - 2\exp(-1/x) + \exp\{-V(1,1)/x\} \\ &= \frac{2 - V(1,1)}{x} + \left[\frac{V(1,1)^2}{2} - 1\right]/x^2 + O(x^{-3}), \text{ as } x \to \infty. \quad (3.2.10)\end{aligned}$$

It follows that

$$\chi(x) = 2 - V(1,1) + \left[\frac{V(1,1)^2}{2} - 1\right]/x + O(x^{-2}), \text{ as } x \to \infty.$$

Hence, $\chi = 2 - V(1,1) > 0$ if $V(1,1) < 2$, i.e. the variables are asymptotically dependent when they are not perfectly independent, and $\chi = 0$ when $V(1,1) = 2$, i.e., the variables are only asymptotically independent when they are perfectly independent. For the bivariate extreme value distribution with $V(1,1) < 2$, $\eta = 1$, $\alpha = 1$ and $\beta = 0$, with $G(z) = -V_1(1, e^z)$, where $V_1(\cdot,\cdot)$ is the first derivative of $V(\cdot,\cdot)$ (see Heffernan and Tawn (2004), Section 8.4).

Typically, it is difficult to derive V explicitly, and this is the limiting factor in finding tractable max-stable models. There exists a slowly growing set of models for which the bivariate distributions are available; see, e.g., Smith (1990), Schlather (2002), Kabluchko

et al. (2009), Davison et al. (2012) and Wadsworth and Tawn (2012a). We list the details of the most widely used below:

Smith process: Smith (1990) takes $W_i(s) = \phi_d(s - t_i, \Sigma)$, where $d = \dim(S)$, $\phi_d(\cdot, \Sigma)$ is the d-dimensional multivariate normal density function where the vector mean is $\mathbf{0}$ and variance is Σ, and where t_i have measure dt over $\mathbb{R}^d$. For this model the pairwise exponent measure for locations $\mathbf{o}$, the origin, and $\mathbf{h}$ is

$$V_{\mathbf{h}}(x, y) = x^{-1}\Phi\left\{\frac{\psi(\mathbf{h})}{2} + \frac{1}{\psi(\mathbf{h})}\log\left(\frac{y}{x}\right)\right\} + y^{-1}\Phi\left\{\frac{\psi(\mathbf{h})}{2} + \frac{1}{\psi(\mathbf{h})}\log\left(\frac{x}{y}\right)\right\}, \quad (3.2.11)$$

where $x > 0$ and $y > 0$, Φ is the standard normal distribution function, and $\psi(\mathbf{h}) = \sqrt{\mathbf{h}^T\Sigma^{-1}\mathbf{h}}$ is the Mahalanobis distance between $\mathbf{h}$ and $\mathbf{o}$. It follows that the level of asymptotic dependence at separation $\|\mathbf{h}\|$ is $\chi_{\mathbf{h}} = 2[1-\Phi\{\psi(\mathbf{h})/2\}]$ for which $\chi_{\mathbf{h}} \to 0$ as $\|\mathbf{h}\| \to \infty$, and $\chi_{\mathbf{h}} \to 1$ as $\|\mathbf{h}\| \to 0$, spanning the range of possible asymptotic dependencies. At all finite separations of the locations we have asymptotic dependence.

Schlather process: Schlather (2002) took $W_i(s) = \max\{0, Z_i(s)\}$ where the $Z_i(s)$ are independent replicates of a stationary process $Z(s)$ on $\mathbb{R}^d$, satisfying $\mathbb{E}[\max\{0, Z(\mathbf{o})\}] = 1$, where $\mathbf{o}$ denotes the origin. When $\sqrt{2\pi}Z(s)$ is a stationary Gaussian process with mean zero, variance 2π and correlation function $\rho(\mathbf{h})$, then the pairwise exponent measure for locations $\mathbf{o}$, the origin, and $\mathbf{h}$ is

$$V_{\mathbf{h}}(x, y) = \frac{1}{2}\left(\frac{1}{x} + \frac{1}{y}\right) \times \left(1 + \left[1 - 2\frac{(\rho(\mathbf{h}) + 1)xy}{(x+y)^2}\right]^{1/2}\right),$$

where $x > 0$ and $y > 0$. Here $\chi_{\mathbf{h}} = 1 - \{[1 - \rho(\mathbf{h})]/2\}^{1/2}$. When $d = 2$, because $\rho(\mathbf{h})$ is a positive definite function, $\chi_{\mathbf{h}} \geq 0.162$. This means that the model cannot account for processes with low asymptotic dependence at any separation.

Brown-Resnick process: Brown and Resnick (1977) let $W(s) = \exp\{\epsilon(s) - \gamma(s)\}$, where $\epsilon(s)$ is a fractional Brownian process with semivariogram $\gamma(h) = (h/c)^\alpha$, $0 < \alpha \leqslant 2$, $c > 0$,

$h = \|\mathbf{h}\| > 0$, and $\epsilon(0) = 0$ almost surely. Here $\chi_{\mathbf{h}} = 2[1 - \Phi\{\{2\gamma(h)\}^{1/2}/2\}]$. A special case is the Smith model when $\alpha = 2$.

Extremal-t process: Demarta and McNeil (2005) assume that

$$W(s) = \sqrt{\pi} 2^{-\nu/2+1} [\Gamma\{(\nu+1)/2\}]^{-1} \max\{\epsilon(s), 0\}^{\nu},$$

where $\nu \geqslant 1$, Γ is the Gamma function, and $\epsilon(x)$ is a stationary Gaussian random field with mean zero and correlation function $\rho(\mathbf{h})$. The case when $\nu = 1$ corresponds to the Schlather process. The level of asymptotic dependence between locations $\mathbf{o}$ and $\mathbf{h}$ is $\chi_{\mathbf{h}} = 2[1 - T_{\nu+1}(\sqrt{(\nu+1)[1-\rho(\mathbf{h})]/[1+\rho(\mathbf{h})]})]$, where T_ν denotes the cumulative distribution function of a student-t random variable with ν degrees of freedom.

Gaussian Processes

Gaussian processes are the simplest and most well-known asymptotically independent processes. Let $\{Z(s), s \in S\}$ be a stationary Gaussian process with $E\{Z(s)\} = 0$, $\mathrm{Var}\{Z(s)\} = 1$ and $\mathrm{Corr}\{Z(\mathbf{o}), Z(\mathbf{h})\} = \rho(\mathbf{h})$. Define

$$X_F(s) = -1/\log \Phi(Z(s)),$$

then $X_F(s)$ has Fréchet margins and a Gaussian process copula. For this process, $\chi_{\mathbf{h}} = 0$ for all $\|\mathbf{h}\| \neq o$, and $\eta_{\mathbf{h}} = \{1 + \rho(\mathbf{h})\}/2$, where $\eta_{\mathbf{h}}$ is the measure in (3.2.4) for the process at the pair of sites $\mathbf{o}$ and $\mathbf{h}$. For large x we have

$$\chi_{\mathbf{h}}(x) = \mathcal{L}_{\mathbf{h}}(x)/x^{1/\eta_{\mathbf{h}}},$$

with $\mathcal{L}_{\mathbf{h}}$ a slowly varying function. Furthermore, in the Heffernan and Tawn formulation, the dependence parameters (α, β) for the process at the pair of sites $\mathbf{o}$ and $\mathbf{h}$ are

$$\alpha_{\mathbf{h}} = (\rho(\mathbf{h}))^2 \operatorname{sign}(\rho(\mathbf{h})) \text{ and } \beta_{\mathbf{h}} = 1/2.$$

The limiting conditional distribution $G_{\mathbf{h}}$ (corresponding to G in equation (3.2.7)) is Gaussian with variance $2\rho(\mathbf{h})^2[1-\rho(\mathbf{h})^2]$. This shows that simply using a Gaussian copula for asymptotically independent cases is not a sufficiently general approach as both $\beta_{\mathbf{h}}$ and $G_{\mathbf{h}}$ have restrictive forms relative to the more general Heffernan and Tawn class.

Inverted Max-Stable Processes

Wadsworth and Tawn (2012a) introduce an alternative class of asymptotically independent processes - the inverted max-stable processes. Essentially, these processes have the same copula as max-stable processes, but with the copula inverted, so that lower tail dependence translates to upper tail dependence, and vice versa. Formally, if $X_{MS}(\cdot)$ and $X_{IMS}(\cdot)$ are a max-stable process and an inverted max-stable process, respectively, both with Fréchet margins, then

$$X_{IMS}(s) = -1/\log[1-\exp\{-1/X_{MS}(s)\}], \text{ for } s \in S. \tag{3.2.12}$$

This process has $\chi_{\mathbf{h}} = 0$ for all $\mathbf{h} \neq \mathbf{0}$, and

$$\chi_{\mathbf{h}}(x) = x^{-1/V_{\mathbf{h}}(1,1)} + O(x^{-1-1/V_{\mathbf{h}}(1,1)}) \text{ for large } x,$$

so $\eta_{\mathbf{h}} = 1/V_{\mathbf{h}}(1,1)$, where $V_{\mathbf{h}}$ is the exponent measure for $X_{MS}(\mathbf{o})$ and $X_{MS}(\mathbf{h})$. The form of the conditional extremes parameters $(\alpha_{\mathbf{h}}, \beta_{\mathbf{h}})$ is complicated in general, see Papastathopoulos and Tawn (2016). We focus in particular on the example where $V(x,y)$ corresponds

to the logistic model (Tawn, 1988)

$$V(x,y) = \left(x^{-1/\gamma} + y^{-1/\gamma}\right)^{\gamma}$$

where $x > 0$, $y > 0$ and $\gamma \in (0, 1]$, with $\gamma \to 1$ implying independence and $\gamma \to 0$ implying perfect dependence. Then the joint distribution function for the inverted bivariate logistic extreme value distribution is

$$\begin{aligned} F(x,y) &= \exp(-1/x) + \exp(-1/y) - 1 \\ &\quad + \exp\left\{-\left[\left(\frac{-1}{\log(1-e^{-1/x})}\right)^{-1/\gamma} + \left(\frac{-1}{\log(1-e^{-1/y})}\right)^{-1/\gamma}\right]^{\gamma}\right\}. \end{aligned} \tag{3.2.13}$$

For this distribution $\eta = 2^{-\gamma}$, $\alpha = 0$ and $\beta = 1 - \gamma$.

3.3 Estimation

In this section we introduce several methods for estimating the dependence measures described in Section 3.2. The performance of these methods will be explored in Section 3.4. Our particular focus is on the estimation of η and $\chi(x)$; the former as it gives a single number summary indicating how close the joint tail is to asymptotic dependence, and the latter as it provides a clear picture of the implications of the nature of the extremal dependence on actual joint occurrences. We base our estimation of η on methods of Ledford and Tawn (1996), and propose new models and constraints. Equations (3.2.6) and (3.2.8) present different formulations for $\chi(x)$; we present inference based on both of these.

3.3.1 Estimating η

To estimate the tail dependence given in representation (3.2.4) it is useful to define $T = \min(X_P, Y_P)$ and note that

$$\begin{aligned} \Pr\{T > x\} &= \Pr\{\min(X_P, Y_P) > x\} \\ &= \Pr\{X_P > x, Y_P > x\} \\ &= \mathcal{L}(x)x^{-1/\eta} \text{ for } x \geqslant 1. \end{aligned} \tag{3.3.1}$$

Thus, $0 < \eta \leq 1$ is the tail index of the variable T. This means that, even though dependence is a bivariate feature, univariate techniques can be used to estimate η simply by using data on T constructed from (X_P, Y_P). To estimate η using the tail form (3.3.1), we need to make some modelling assumptions about the positive slowly varying function $\mathcal{L}(x)$ above a high threshold u, in addition to the property (3.2.5) and expression (3.3.1) representing a valid survivor function. In previous studies when parametric models have been specified for $\mathcal{L}(x)$ for $x \geq u$, the only constraint imposed is that $0 < \Pr\{T > u\} \leq 1$, so $0 < \mathcal{L}(u) \leq u^{1/\eta}$. However, from conditional probability (3.2.6) we have the stronger condition $0 < \mathcal{L}(u) \leq u^{(1-\eta)/\eta}$. For the tail of T to have a density $f_T(x) > 0$ for $x > u$, we also require that

$$\mathcal{L}(x) > \eta x \mathcal{L}'(x) \text{ for all } x \geq u.$$

Possible choices for $\mathcal{L}(x)$, for $x > u$, that have been suggested in the past (see, Ledford and Tawn (1996), Ledford and Tawn (2003) and Pickands (1975), respectively, for models

(3.3.2)-(3.3.4)) are:

$$\mathcal{L}_1(x) = c \tag{3.3.2}$$

$$\mathcal{L}_2(x) = c + dx^{1/\eta - 2} \tag{3.3.3}$$

$$\mathcal{L}_3(x) = \lambda \left[x^{-1} \left(1 + \frac{\eta(x-u)}{\sigma} \right) \right]^{-1/\eta}. \tag{3.3.4}$$

Model (3.3.2) leads to the Hill estimator for η and model (3.3.4) corresponds to a generalised Pareto tail with the shape $\eta > 0$, scale $\sigma > 0$ and threshold exceedance rate $0 < \lambda < 1$. In contrast, the two-term model (3.3.3) attempts to capture a natural second order decay rate with the tail $cx^{-1/\eta} + dx^{-2}$, so for this to be the second term, we additionally require $\eta > 1/2$, corresponding to positive extremal dependence. Our constraints impose that $0 \leqslant c \leqslant u^{(1-\eta)/\eta}$ for model (3.3.2), and $cu^{-(1-\eta)/\eta} + du^{-1} \leqslant 1$ and if $d < 0$ that $c > 2\eta|d|u^{-2+1/\eta}$ for model (3.3.3). Additionally, we imposed that η must be in the range $(0, 1]$.

The models given above can be fitted using a censored likelihood approach. We take the likelihood to be

$$L(\boldsymbol{\theta}) = \left\{ \prod_{i=1}^{n_u} f_T(t_i) \right\} \Pr\{T \leqslant u\}^{n-n_u}, \tag{3.3.5}$$

where $\boldsymbol{\theta}$ is the vector parameter for the tail of T, $T = \min(X_P, Y_P)$, n_u is the number of points above the threshold u, and $t_1, \ldots, t_{n_u}$ is an enumeration of the values of $T > u$. We use constrained maximum likelihood to impose our restrictions on the parameter space when estimating $\boldsymbol{\theta} = (c, \eta)$, $\boldsymbol{\theta} = (c, d, \eta)$ or $\boldsymbol{\theta} = (\lambda, \sigma, \eta)$ for models (3.3.2), (3.3.3) and (3.3.4), respectively. For model (3.3.2) analytical estimates can be obtained

$$\widehat{\boldsymbol{\theta}} = (\widehat{c}, \widehat{\eta}) = \left(\frac{n_u}{n} u^{1/\widehat{\eta}}, \min \left\{ \frac{1}{n_u} \sum_{i=1}^{n_u} \log \left(\frac{t_i}{u} \right), 1 \right\} \right).$$

The other two models require constrained numerical maximisation of the likelihood to obtain parameter estimates as no tractable solutions exist.

Simply using the likelihood function does not exploit knowledge about the processes in-

volved, e.g., they are positively dependent in the extremes, nor our caution in the inference about being conservative from the outset, i.e., expecting the extremal dependence structure to be more like asymptotic dependent than asymptotic independence prior to seeing data. To incorporate this information it is natural to adopt a Bayesian framework to the inference. Although much work (see e.g., Lee 2012) has been undertaken on selecting priors to represent our lack of knowledge (i.e., non-informative priors), less work exists for guiding the incorporation of informative prior knowledge. A key development in the latter has been through the approach of Simpson et al. (2015). They introduce the penalised complexity (PC) prior as a way of constructing prior distributions for situations where there is a natural nested structure in the model components; i.e., there is a more flexible model that is an extension of a base model. PC priors penalise the complexity induced by deviating from the simpler base model. For our context, a natural base model is asymptotic dependence and the more flexible model covers both asymptotic independence and dependence parametrised by η. Focusing on exceedances of the threshold u by T and adopting model (3.3.2) for $\mathcal{L}(x)$, the base model is

$$g(x) = \frac{u}{x^2}, \quad x > u,$$

and the more flexible model is

$$f(x) = \frac{u^{1/\eta}}{\eta x^{1/\eta+1}}, \quad x > u,\ 0 < \eta \leqslant 1,\ u > 0.$$

So, $f(x) = g(x)$ when $\eta = 1$. The penalised complexity prior for η (derived in Appendix 3.A) is then

$$\pi(\eta; \lambda) = \frac{\lambda(1-\eta)\exp\left(-\lambda\sqrt{-2\log\eta + 2(\eta-1)}\right)}{\eta\sqrt{-2\log\eta + 2(\eta-1)}}, \text{ for } 0 < \eta \leq 1,$$

where the parameter $\lambda > 0$ controls the shape of the prior. The value of λ determines the strength of the prior belief that a model close to an asymptotically dependent model is appropriate to the data. In the studies in this chapter we chose λ such that $\pi(\eta > 0.5; \lambda) = 0.99$, which essentially ensures positive extremal dependence. From this it follows

that $\lambda = 7.409$. We use this prior with the likelihood given in (3.3.5) to obtain a maximum a posteriori (MAP) estimate of η.

3.3.2 Estimating $\chi(x)$

Though it is possible to estimate $\chi(x)$, defined by probability (3.2.3), empirically, this restricts estimates to x in the range of the data. To estimate beyond the data we need to use a model; here we use asymptotic forms (3.2.6) and (3.2.8).

Using the Ledford and Tawn model (3.2.6), with estimates $\widehat{\eta}$ and $\widehat{\mathcal{L}}(x)$ for each of the models (3.3.2)-(3.3.4), substituting the maximum likelihood estimates of the parameters, gives

$$\widehat{\chi}_i(x) \approx \frac{\widehat{\mathcal{L}}_i(x)}{x^{1/\widehat{\eta}-1}}, \text{ for } x \geqslant u, i = 1, 2, 3. \tag{3.3.6}$$

Now consider inference for $\chi(x)$ using the conditional extremes model (3.2.7), with $\chi(x)$ given by the integral expression (3.2.8). The simplest way to evaluate this integral is through Monte Carlo methods using estimated values $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{G}$ of α, β and G, as shown in the algorithm below.

Algorithm 1 Algorithm for Monte Carlo evaluation of $\chi(x)$ using conditional extremes model

Simulate X_L above $x_L = \log(x/2)$ as $X_L = x_L + v$, where v is a realisation of an Exponential(1) random variable;
Simulate Z from $\widehat{G}$ independently of X_L;
Set $Y_L = \widehat{\alpha} X_L + (X_L)^{\widehat{\beta}} Z$;
Repeat steps 1-3 m-times;
Set $\widehat{\chi}_{HT}(x)$ as the proportion of the m simulated pairs (X_L, Y_L) with $Y_L > x_L$.

3.4 Simulation study

Here we compare different methods for the estimation of η and $\chi(x)$ presented in Section 3.3 for samples from asymptotically dependent and asymptotically independent models given in Section 3.2.3. We further investigate how sample size and threshold choice affect estimation.

3.4.1 Design

Simulation processes

We simulate samples from the max-stable Smith, Schlather, Brown-Resnick and extremal-t processes, the Gaussian process, and from the inverted logistic bivariate extreme value distribution.

Simulating from a max-stable process is relatively straightforward; for details, see Schlather (2002). In this chapter we used the *SpatialExtremes* package in *R* (Ribatet 2008), which gives samples on Fréchet margins that can then be transformed to the desired margins using the probability integral transform. We simulated from the processes over 10 equally spaced locations on a straight line. For the Smith process, we set $\Sigma = \{\sigma_{ij}\}$ with $\sigma_{11} = \sigma_{22} = 2$ and $\sigma_{12} = \sigma_{21} = 0$, in order to give a good range of Spearman's ρ on (0,1) between pairs of locations. For similar reasons, we simulated the Schlather process with correlation function $\rho_P(h) = \exp(-h/4)$ and the Brown-Resnick process with semivariogram $\gamma(h) = \exp(-h/0.8)$. For the extremal-t simulation, the correlation function used was $\rho_P(h) = \exp(-h/3)$ with $\nu = 5$ degrees of freedom.

The Gaussian process samples were simulated over 15 equally spaced locations on a line, with a distance of one unit between the closest sites. The process has correlation function $\rho_P(h) = e^{-\lambda h}$, where h is the distance between two locations, and λ was chosen such that $\rho_P(1) = 0.6$. Hence, $\rho_P(mh) = \rho_P(h)^m$, for $m = 1, \ldots, 15$, leading to a process at the observed sites being a first order Markov process. We generated the samples using the *mvtnorm* package in *R* (Genz et al. 2014). The chosen combination of number of sites and first order dependence parameter λ ensures a good coverage of the range (0,1) for Spearman's ρ.

Samples from the inverted logistic distribution were obtained by simulating from the bivariate extreme value distribution using the *evd* package in *R* (Stephenson 2002), and then

inverting the lower tail of this using (3.2.12). We simulated over 10 equally spaced locations on a line, and the dependence parameter γ was chosen such that a good range of Spearman's ρ is obtained over (0,1).

General behaviour of η and $\chi(x)$ with ρ

Throughout the simulation study, we simulate samples from each of the processes described above over 10 (or 15) equally spaced locations on a straight line, and we estimate ρ, $\chi(x)$ and η for pairs of locations with different separation lags along this line. In general, we expect pairs of locations with smaller lag to be more correlated (and hence have higher ρ estimates) than pairs of locations further apart. Hence, ρ corresponds to the lag between locations to some extent. Since the estimation of ρ is very precise relative to the estimation of η and $\chi(x)$, we only show confidence intervals for η and $\chi(x)$ throughout.

First we explore how values of η and $\chi(x)$ link to Spearman's ρ for the different processes we have introduced above. Here we evaluate each of these properties using estimates obtained from replicates of very large samples.

Figure 3.4.1 illustrates the relationship between median estimates for η and $\chi(x)$ and values for ρ from 1000 realisations of samples of size 10^6 from each of the processes, with η estimated using model (3.3.4) with threshold probability $q = \Pr\{T \leqslant u\} = 0.99$, and $\chi(x)$ estimated empirically using equation (3.2.3) with $x = 100$. The point-wise $(2.5\%, 97.5\%)$ uncertainty band is also shown for η and $\chi(x)$. Here $\widehat{\rho}$ denotes the median estimate of ρ over the 1000 sample realisations. The width of the 95% confidence interval for $\widehat{\rho}$ is at most 0.01 for this large sample size, so the ρ estimates are very precise. The relationship between $\widehat{\eta}$ and $\widehat{\rho}$ is very similar for all asymptotically dependent max-stable processes (shown in magenta, red, blue and green), as is the relationship between $\widehat{\chi}(x)$ and $\widehat{\rho}$. The behaviour of the asymptotically independent Gaussian process (shown in black) and inverted logistic distribution (shown in cyan) is clearly different for both $\widehat{\eta}$ with $\widehat{\rho}$ and $\widehat{\chi}(x)$ with $\widehat{\rho}$. However,

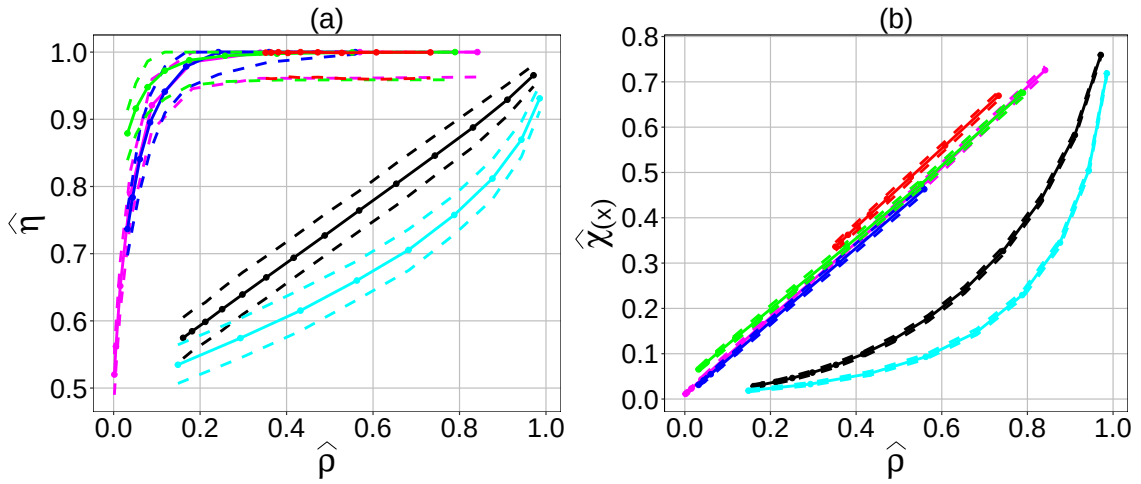


Figure 3.4.1: Estimates for (a) η and (b) $\chi(x)$ against estimates for Spearman's ρ for simulated samples of size $n = 10^6$ from each of the models introduced in Section 3.2.3; Smith (magenta), Schlather (red), Brown-Resnick (blue), extremal-t (green) and Gaussian (black) processes, and the inverted logistic distribution (cyan). Estimation methods use model (3.3.4) for η with $q = 0.99$, and the empirical estimate for $\chi(x)$ with $x = 100$. Solid lines are median estimates (of η and $\chi(x)$) from 1000 sample replications, dashed lines give 2.5% and 97.5% quantiles. The estimation of Spearman's ρ is very precise relative to the estimation of η and $\chi(x)$, so $\widehat{\rho}$ here is the median of the 1000 sample estimates for each model with no confidence intervals shown.

for asymptotic dependence, the known limiting value of unity for η is only achieved (approximately) for max-stable processes with $\widehat{\rho}$ exceeding 0.2 (see Figure 3.4.1(a)). Similarly, for asymptotic independence, the known limiting value of zero for $\chi(x)$ is never achieved for either the Gaussian process or inverted logistic distribution for our choices of x (see Figure 3.4.1(b)). This suggests that relying purely on sample estimates for η and $\chi(x)$ approximating limiting values, even for large samples, is not a useful diagnostic for extremal dependence. However, the obvious differences in behaviour between estimates of η with ρ, and $\chi(x)$ with ρ suggest that distinction between asymptotic dependence and asymptotic independence is possible.

We confirmed by simulation that the functional relationships between η and ρ, and $\chi(x)$ and ρ, are not strongly dependent on the values of model parameters for a specific model. The results we present in the remainder of this section are therefore effectively independent of the model parameter specification.

Estimators, sample sizes and thresholds

We investigate the performance of the set of estimators for η and $\chi(x)$ for samples from different processes, with different sample sizes and threshold choices. Since the characteristics of all max-stable models (for given sample size and threshold) are found to be very similar, we choose here to illustrate their behaviour using results for the Smith process only; corresponding results for other max-stable processes are given in Appendix 3.B.

In Sections 3.4.2 and 3.4.3, we examine the performance of different estimators for η and $\chi(x)$, respectively, for samples from the Smith max-stable process and the Gaussian process. We chose these two processes, so that the behaviour of η and $\chi(x)$ can be inspected for both an AI and an AD process. In Section 3.4.4, we assess the effect of sample size on estimates for η and $\chi(x)$, for samples from the Smith and Gaussian processes, and the inverted logistic distribution. We also consider in Section 3.4.4 the effect of the threshold probability q and level x on estimates for η and $\chi(x)$, respectively, for these samples.

We consider three sample sizes, motivated by the North Sea application discussed in Section 3.5: a large sample with $n = 10^6$ observations (probably unrealistically large for typical metocean applications, but useful to assess large sample performance), and two more realistic sample sizes of $n = 58585$ (medium size, equivalent to the size of the sea-state significant wave height sample in Section 3.5), and $n = 916$ (small size, equivalent to the size of the storm-peak significant wave height sample in Section 3.5). For each sample size, we adjust the threshold probability q and the level x to achieve acceptable numbers of observations for estimation of η and $\chi(x)$, respectively. For the large sample, high threshold probabilities and levels are chosen, whereas for the more realistically-sized medium and small samples we are forced to choose lower threshold probabilities.

The key feature of figures shown in this section is the degree of agreement between sample estimates for η and $\chi(x)$ and their known limiting behaviour given the underlying process (or distribution) used to simulate the sample. For asymptotically dependent samples, we

expect estimates for η close to unity, and estimates for $\chi(x)$ bounded away from zero even for large x. For asymptotically independent samples, we expect estimates for η less than unity, and estimates for $\chi(x)$ approaching zero for large x. We expect that agreement between sample estimate and limiting behaviour further depends on the overall extent of dependence in the full sample, hence we plot these estimated extremal dependence features against the Spearman's rank correlation ρ.

3.4.2 Comparison of estimation methods for η

We denote the estimators of η given by the censored likelihood approach (3.3.5) for the models (3.3.2)-(3.3.4) by $\widehat{\eta}_1$, $\widehat{\eta}_2$ and $\widehat{\eta}_3$ respectively. We also estimated models (3.3.2) and (3.3.3) with the Bayesian penalised complexity (PC) prior described in Section 3.3.1. However, we found that the estimates were effectively unchanged unless very strong priors were used. Hence, we only show these results for the smallest sample size, where the prior is not dominated by the sample likelihood.

We compare estimators for the following combinations (n, q) of sample size n and threshold probability q: $(10^6, 0.99)$, $(58585, 0.90)$ and $(916, 0.90)$. For each sample size, 1000 replicate samples are generated from each of the Smith and Gaussian processes to allow estimation of 2.5% and 97.5% quantiles of the sampling distribution of the estimator of η. Since ρ is relatively precise to estimate, with the width of the 95% confidence interval around $\widehat{\rho}$ at most 0.01 for the large sample and at most 0.1 for the small sample, we use the median of $\widehat{\rho}$ over the 1000 sample replications. Figure 3.4.2 shows estimates for η using the different methods, represented by lines of different colour, for the large sample from the Smith (see Figure 3.4.2(a)) and Gaussian processes (see Figure 3.4.2(b)), plotted against Spearman's ρ.

The Gaussian process is asymptotically independent, with $\eta = (1+\rho_P)/2 < 1$. The estimates in Figure 3.4.2(b) are consistent with this. There is no material difference between the means

of the sampling distributions for the different methods for estimating η here. However, note that the variance of the sampling distributions is slightly larger for model (3.3.3). This is likely due to the additional term in the tail model.

The Smith model is asymptotically dependent over all distances, so the true value of η is unity for all pairs. Whilst there is some difference between the different estimation methods, all estimators underestimate η for asymptotically dependent samples when dependence is low and the sample size is relatively small (see Figure 3.4.2(a)). All η estimators are close to unity for large $\widehat{\rho}$, but $\widehat{\eta}_2$ and $\widehat{\eta}_3$ seem to perform better than $\widehat{\eta}_1$ for pairs with lower $\widehat{\rho}$, as these better capture the higher order features of the tail decay.

The following argument helps understand why η estimates start to perform poorly for the Smith process with weak dependence. For the Smith process, using expression (3.2.10), we know that

$$\begin{aligned}\bar{F}_{T,h}(x) &= \Pr(\min(X(s), X(s+h)) > x) \\ &= \Pr(X(s) > x, X(s+h) > x) \\ &= \frac{(2 - 2\Phi\{a(h)/2\})}{x} + \frac{2\Phi^2\{a(h)/2\} - 1}{x^2} + O(x^{-3}), \text{ for large } x, \end{aligned} \tag{3.4.1}$$

where Φ is the standard normal distribution function and $a(h)$ is the Mahalanobis distance between h and the origin o. For two locations with Spearmans ρ approximately equal to 0.1, the upper tail in equation (3.4.1) becomes $\bar{F}_{T,h}(x) \approx cx^{-1} + dx^{-2} + O(x^{-3})$, with $c = 0.004$ and $d = 0.991$. In order for the first term to dominate we would need $cx^{-1} \gg dx^{-2}$, but this can only be achieved if x is very large. Even to just have $cx^{-1} > dx^{-2}$, for $x > u$, we need $u \approx 250$ on the Fréchet scale, which means a threshold quantile of approximately 99.995% for T. Hence, for practical sample sizes and threshold choices, the estimator $\widehat{\eta}_1$ will be biased for pairs with low Spearmans ρ.

Figures 3.4.3 and 3.4.4 present analogous estimates of η for the medium and small sample sizes, respectively. Note that the smaller the sample size, the η estimates for the Smith

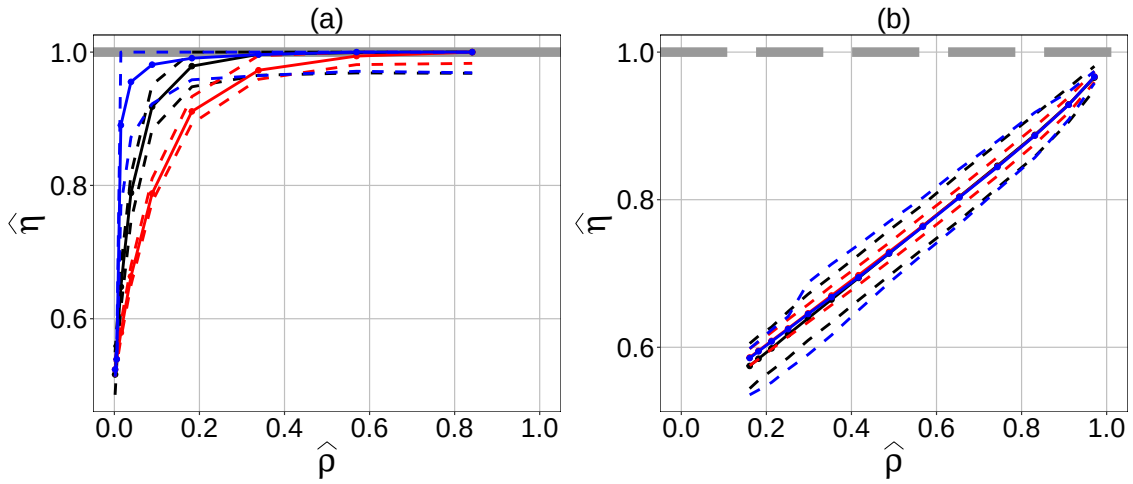


Figure 3.4.2: Estimates for η against Spearman's ρ for sample size $n = 10^6$ with $q = 0.99$, from (a) Smith process and (b) Gaussian process. Estimation methods are model (3.3.2) (red), model (3.3.3) (blue), and model (3.3.4) (black). Solid lines give the median from 1000 sample replications, dashed lines give 2.5% and 97.5% quantiles. The solid grey line shows a correct limiting value for η. The dashed grey line shows a limiting value that η should not take.

model have more bias for pairs with low dependence, and show more uncertainty (see Figures 3.4.3(a) and 3.4.4(a)). However, when the sample size is small, the prior is more important and its inclusion gives better estimates, drawing $\widehat{\eta}$ towards unity (see Figure 3.4.4(a)). For the Gaussian process the estimates are consistent with the true η value, but here too there is more uncertainty in the estimates due to the small sample sizes. The increased uncertainty means that unity lies within the confidence bounds for pairs with $\rho > 0.6$ (see Figure 3.4.4(b)), so in these cases asymptotic dependence might be selected incorrectly.

3.4.3 Comparison of estimation methods for $\chi(x)$

For samples of size $n = 58585$ from the Smith and Gaussian processes, we compare different methods for estimating $\chi(x)$ outlined in Section 3.3.2. Specifically, model-based estimates $\widehat{\chi}_1(x)$, $\widehat{\chi}_3(x)$ and $\widehat{\chi}_{HT}(x)$ are compared with empirical estimate $\widehat{\chi}_{EMP}(x)$. We estimate $\widehat{\chi}_{EMP}(x)$ using the conditional probability given in (3.2.3), by simply counting the points in the region $\{Y_P > x, X_P > x\}$ and dividing by the number of points in the region $\{Y_P > x\}$. Estimates for $\widehat{\chi}_2(x)$ are omitted in figures for clarity due to their close similarity to those for

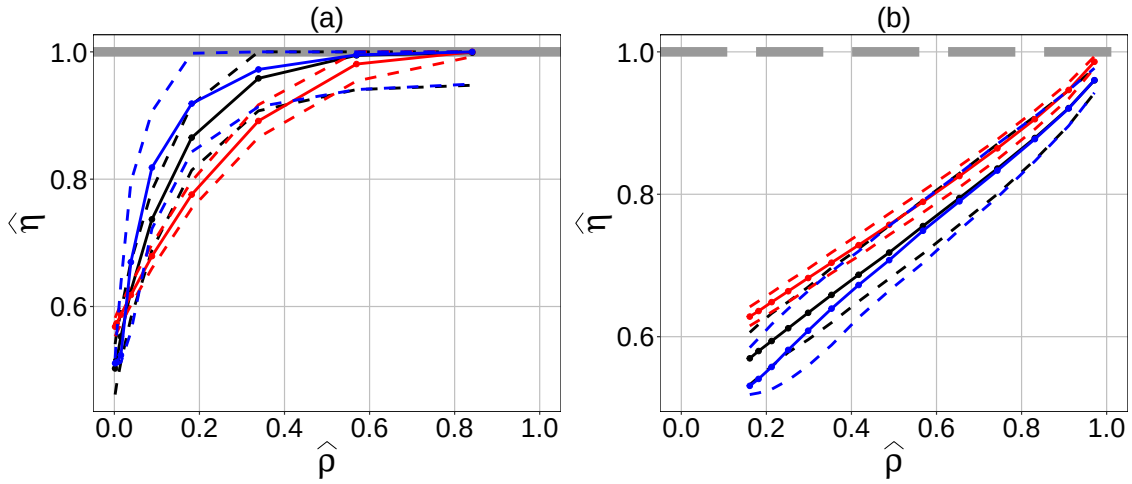


Figure 3.4.3: Estimates for η against Spearman's ρ for sample size $n = 58585$ with $q = 0.90$, from (a) Smith process and (b) Gaussian process. Lines are as described in Figure 3.4.2.

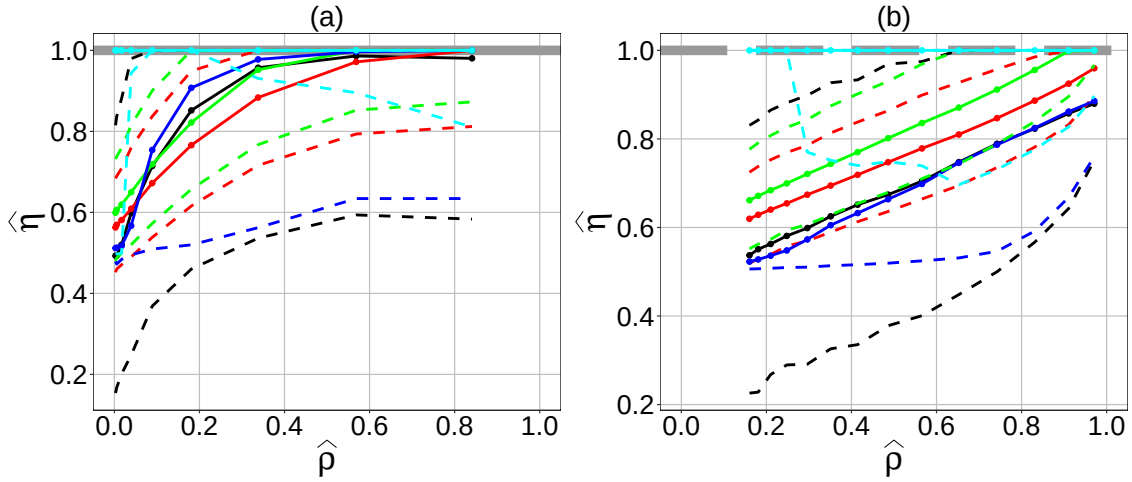


Figure 3.4.4: Estimates for η against Spearman's ρ for sample size $n = 916$ with $q = 0.90$ from (a) Smith process and (b) Gaussian process. Lines are as described in Figure 3.4.2. Additionally, estimation methods using model (3.3.2) with PC prior (green) and model (3.3.3) with PC prior (cyan) are shown.

$\widehat{\chi}_1(x)$. For the model-based estimates $\widehat{\chi}_1(x)$ and $\widehat{\chi}_3(x)$ a modelling threshold u was chosen such that $\Pr\{T \leqslant u\} = 0.90$. For the third model-based estimate $\widehat{\chi}_{HT}(x)$ we estimated the model parameters α, β, μ, and σ above a 90th percentile threshold (see Section 3.2.2). We estimate $\chi(x)$ at the level $x = 10000$. This level was chosen, since it is a very high quantile of the sample, with only 5 data points above this level, so the empirical estimate is expected to be poor. Hence, we can better compare the model-based estimates. We have also compared the estimators at the level $x = 100$, which is within the data, but at this level all the estimators were performing very similarly, so these results are not shown here.

The different estimates are shown against estimates of Spearman's ρ in Figure 3.4.5. The empirical estimate is not reliable, due to the small number of points above this level. The model-based estimates perform similarly for the Gaussian sample (see Figure 3.4.5(b)). However, for the Smith model, $\widehat{\chi}_3(x)$ gives higher mean values and wider uncertainty bands than the other model-based estimators. This might be due to the tail model (3.3.4) that this estimator is based on. Note the different pattern of the relationship between $\widehat{\chi}(x)$ and $\widehat{\rho}$ for the two processes, as already noted on Figure 3.4.1; the asymptotically dependent Smith process shows higher levels of tail dependence than the asymptotically independent Gaussian process across all levels of Spearman's $\widehat{\rho}$. At the high level $x = 10000$, $\widehat{\chi}(x)$ approaches 0 for the Gaussian process sample, which is the limiting value as $x \to \infty$.

In summary, it appears that the different estimation methods lead to fairly similar estimates for both processes, when x is set to a level that is within the data. For this reason, in Section 3.4.4, we adopt the empirical estimate only due to ease of computation. We note however that model-based estimates offer the additional benefit of extrapolation to rare levels (beyond the sample) when necessary.

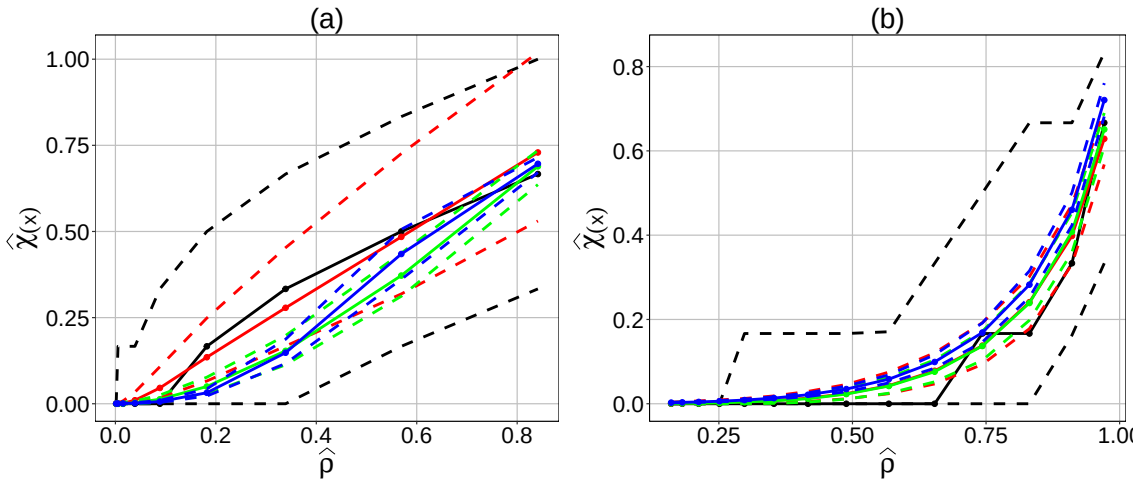


Figure 3.4.5: Estimates for $\chi(x)$ against Spearman's ρ for sample size $n = 58585$ with $x = 10000$ from (a) Smith process and (b) Gaussian process. Estimation methods are empirical $\chi_{EMP}(x)$ (black), $\chi_1(x)$ (blue), $\chi_3(x)$ (red) and $\chi_{HT}(x)$ (green). Solid lines give the median from 1000 sample replications, dashed lines give 2.5% and 97.5% quantiles.

3.4.4 Effect of sample size and threshold effect

Here we assess the effect of sample size, threshold probability and level x on selected estimates for η and $\chi(x)$ for samples from the Smith and Gaussian processes, and the inverted logistic distribution. The estimator $\widehat{\eta}_3$ as given by model (3.3.4) is used for η estimation, and the empirical estimate $\widehat{\chi}_{EMP}(x)$ for $\chi(x)$. For estimating η, the following pairs (n, q) of sample size n and threshold probability q are considered: $(10^6, 0.99)$, $(10^6, 0.999)$, $(10^6, 0.9999)$, $(58585, 0.80)$, $(58585, 0.90)$ and $(58585, 0.99)$. For estimating $\chi(x)$, the level x was set so that the same number of observations would be used for estimation as for the corresponding η estimation. This gives the following pairs (n, x) of sample size n and level x: $(10^6, 100)$, $(10^6, 1000)$, $(10^6, 10000)$, $(58585, 5)$, $(58585, 10)$ and $(58585, 100)$.

For asymptotically dependent samples, we expect η estimates close to unity, and $\chi(x)$ estimates bounded away from zero for all x. Results for the Smith process clearly show that $\widehat{\chi}(x) > 0$ even for large x and small ρ (see Figure 3.4.6(b) and 3.4.6(d)). Further, $\eta \approx 1$ for pairs with high dependence in the body (as measured by Spearman's ρ), however, η is underestimated for pairs with weak dependence in the body, particularly for smaller sample sizes

and lower thresholds. Figure 3.4.6(a) shows that the η estimate is good for our large sample, but this sample size and threshold probability are not typically achievable in practice. For realistic sample sizes and threshold probabilities, η is underestimated for low levels of dependence (see Figure 3.4.6(c)). This is the case for other asymptotically dependent models as well. We conducted the same analysis for the other max-stable processes described in Section 3.2.3 with similar results (see Appendix 3.B).

For asymptotically independent samples, the limiting value of η is less than 1 for all pairs, and so $\chi = 0$. For both the Gaussian and the inverted logistic samples, $\widehat{\eta} < 1$ for all pairs (see Figures 3.4.7(a,c) and 3.4.8(a,c)). However, $\widehat{\chi}(x)$ is approximately 0 only for pairs with low dependence in the body of the data for all levels of x considered here. At high levels x, estimates approach the limiting value of χ (see Figures 3.4.7(b,d) and 3.4.8(b,d)). Face-value interpretation of $\chi(x)$ estimates suggests asymptotic dependence; and we know this is not the case. This shows the critical importance η plays in distinguishing asymptotic independence from asymptotic dependence.

We note also that $\widehat{\eta}$ and $\widehat{\chi}(x)$ appear to decrease faster with decreasing $\widehat{\rho}$ for the inverted logistic sample than for the Gaussian sample (compare Figures 3.4.7 and 3.4.8). This behaviour is not affected by the distributional parameters chosen for sample simulation from these processes; these only affect the range of ρ, not the relationship between ρ and η and $\chi(x)$. These features therefore provide a valuable tool for model diagnostics for asymptotically independent processes, and will be used in Section 3.5.

3.5 Application

3.5.1 Data

Characterisation of marginal and spatial dependence for extremes of significant wave height H_S is essential to reliable design and operation of marine and coastal installations. Sea-state

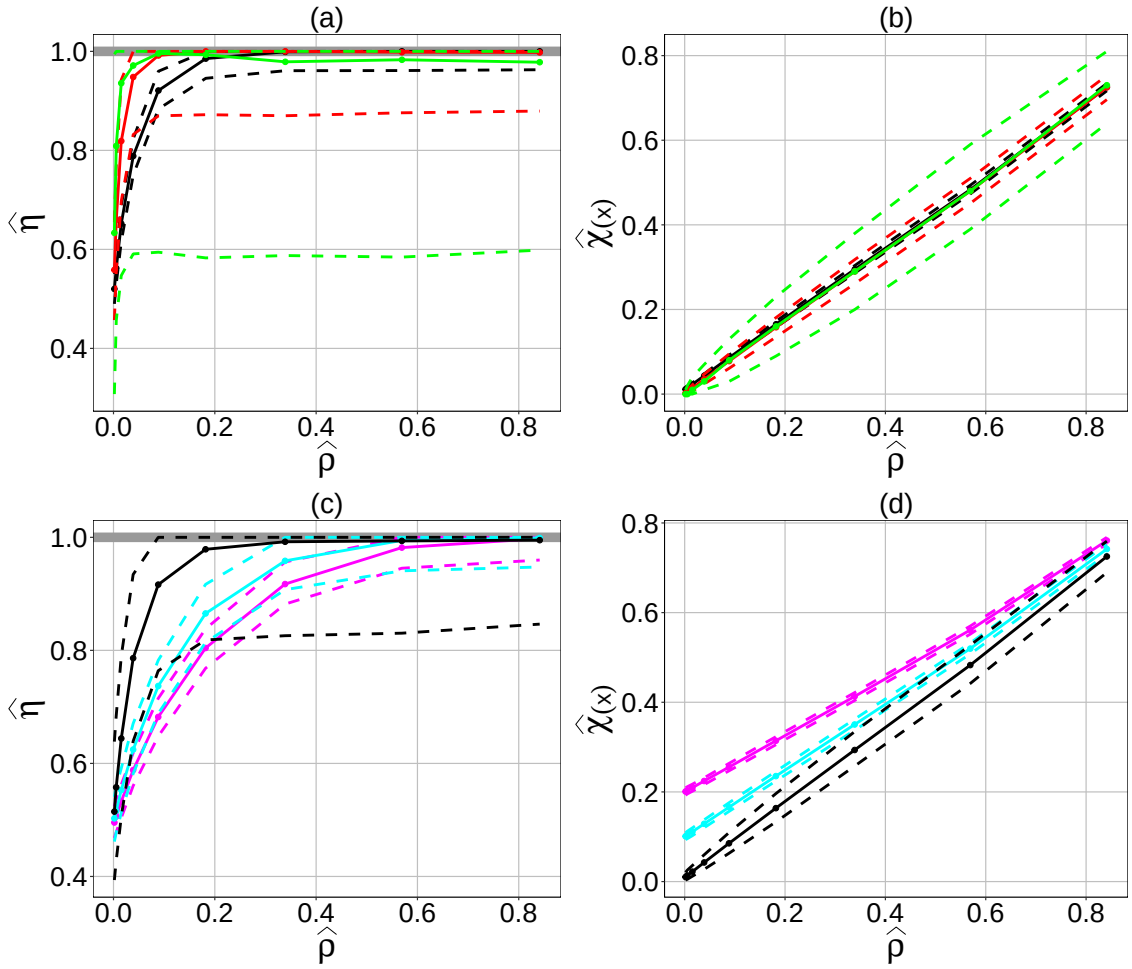


Figure 3.4.6: Estimates of η and $\chi(x)$ plotted against Spearman's ρ for simulated data from the Smith model. Sample size is $n = 10^6$ (top row) and $n = 58585$ (bottom row). Threshold probabilities for the η estimation are $q = 0.9999$ (green line), $q = 0.999$ (red line), $q = 0.99$ (black line), $q = 0.90$ (cyan line) and $q = 0.80$ (magenta line). Corresponding levels for the $\chi(x)$ estimation are $x = 10000$ (green line), $x = 1000$ (red line), $x = 100$ (black line), $x = 10$ (cyan line) and $x = 5$ (magenta line). Solid lines give the median from a 1000 sample replications, dashed lines give the 2.5% and 97.5% quantiles. The solid grey line shows a correct limiting value for η.

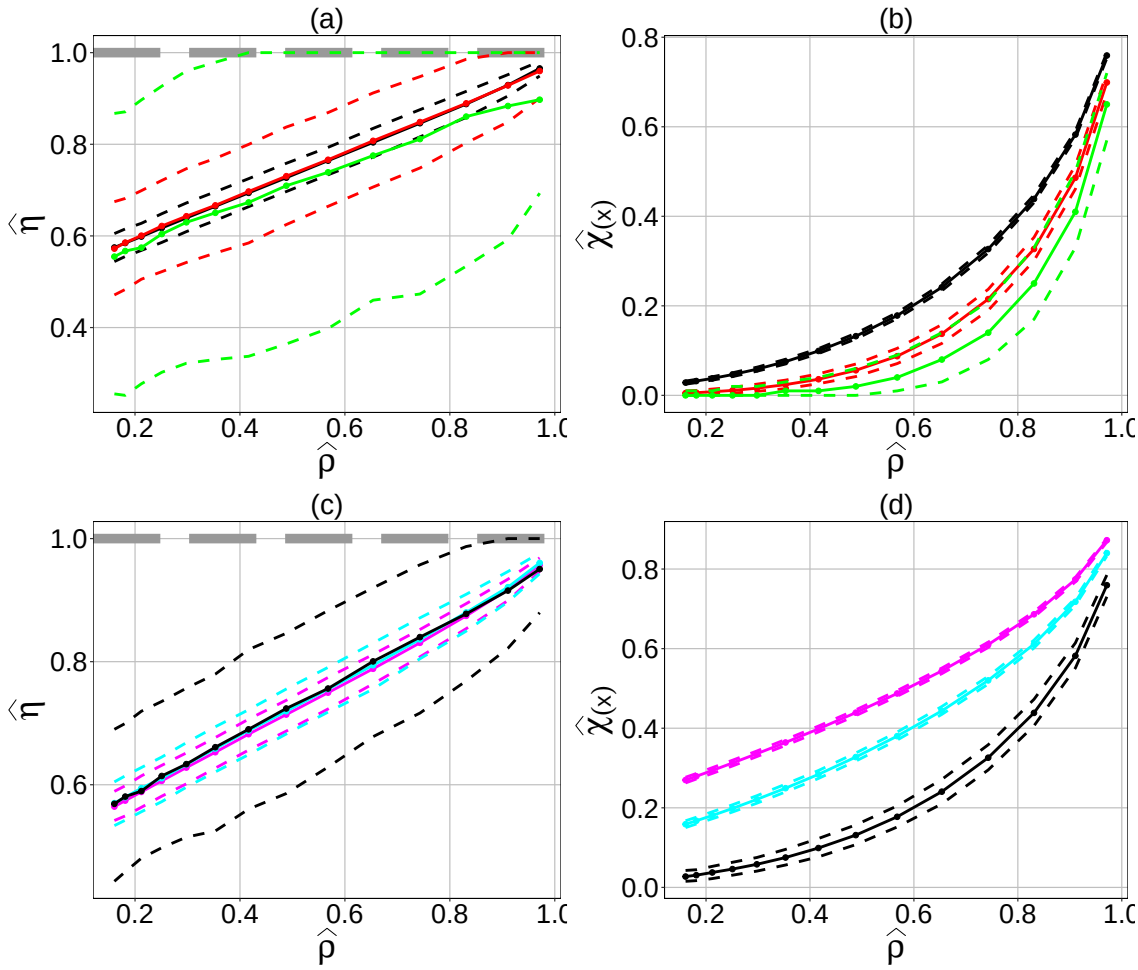


Figure 3.4.7: Estimates of η and $\chi(x)$ plotted against Spearman's ρ for simulated data from the Gaussian process. Sample size is $n = 10^6$ (top row) and $n = 58585$ (bottom row). Lines are as described on Figure 3.4.6. The dashed grey line shows a limiting value for η that the estimates should not take.

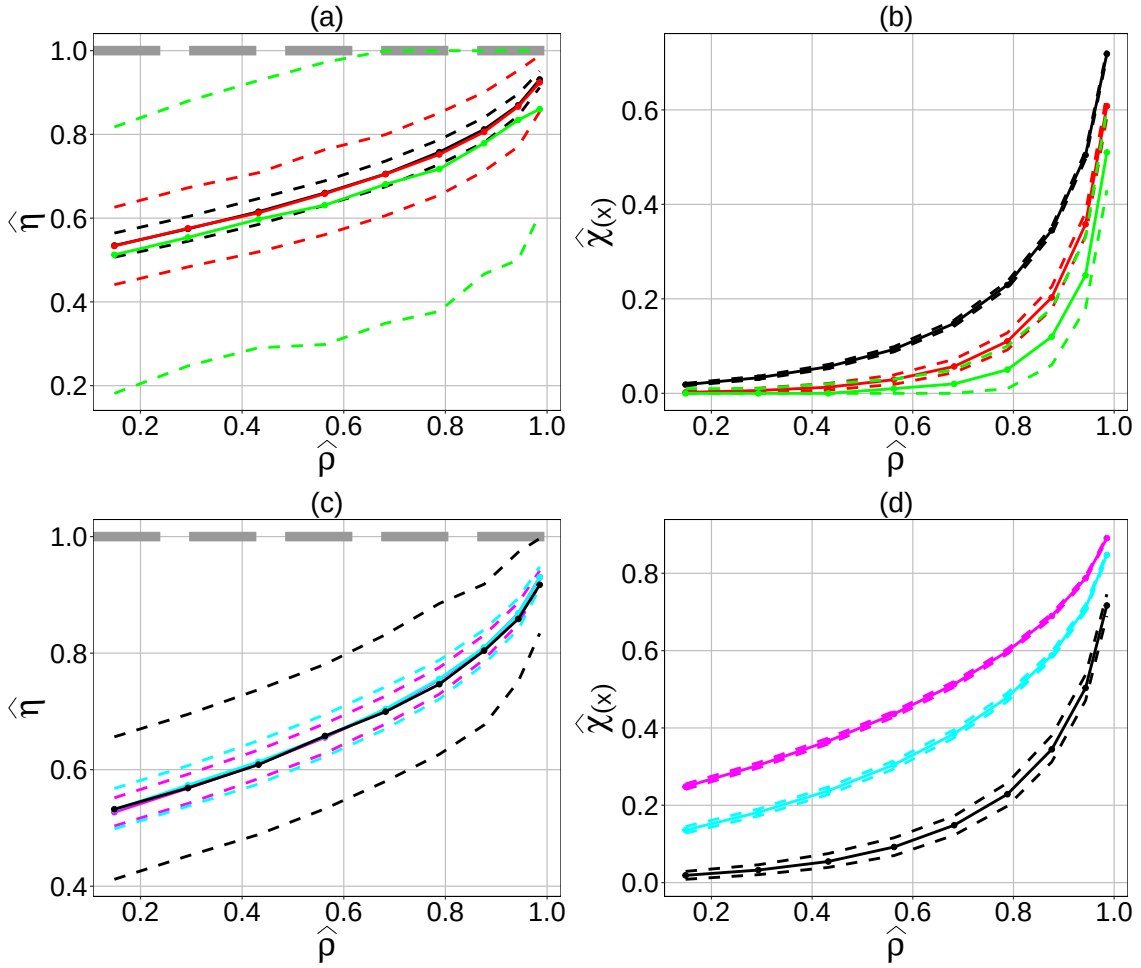


Figure 3.4.8: Estimates of η and $\chi(x)$ plotted against Spearman's ρ for simulated data from the inverted logistic model. Sample size is $n = 10^6$ (top row) and $n = 58585$ (bottom row). Lines are as described on Figure 3.4.6. The dashed grey line shows a value for η that the estimates should not take.

H_S is defined as four times the standard deviation of the ocean surface elevation in a given time period. The data examined here consists of time-series for H_S for three hour sea-states (from the hindcast of Reistad et al. 2011) for the period September 1957 to December 2012 on a grid of 150 northern North Sea locations, covering an area of approximately 5° longitude by 5° latitude, with an approximate grid spacing of 0.4°; see Figure 3.5.1. The univariate extreme value characteristics of H_S from the hindcast have been examined by Aarnes et al. (2012) and Breivik et al. (2013). We analyse this sample of sea-state H_S and a further sample of 916 independent storm-peak H_S values extracted from the sea-state data using the procedure of Ewans and Jonathan (2014): briefly, contiguous intervals of sea-state H_S above a low peak-picking threshold are identified, each interval corresponding to a storm event. The maximum of H_S during the interval is taken as the storm-peak H_S for the storm. Consecutive values of sea-state H_S at a location typically exhibit temporal dependence, whereas consecutive storm-peak values do not.

The distribution of sea-state H_S varies with wave direction θ due to the combination of prevailing wind field and land shadow effects. Figure 3.5.2 shows (θ, H_S) for one location. Note that θ represents the direction from which waves emanate, measured clockwise from North. The density estimates for $\theta|H_S > v$ for $v = 0, 4, 8$m and 10m are also shown. It can be seen that as the threshold v increases, the density is increasingly concentrated in the [320, 10] and [120, 200] directional sectors. This suggests that for this particular location the biggest waves come from these directions, corresponding to storms emanating from the North North-West and South South-East approximately, respectively, roughly the orientation of the blue strip in Figure 3.5.1. We therefore might expect directional effects to influence spatial dependence, and that ignoring θ in assessing spatial dependence might not be justified. Focusing on strips of locations with different orientations helps us assess and illustrate directional influences on spatial dependence in a more straightforward way. Hence, we proceed to examine the extremal dependence between pairs of locations on the four straight-line strips, shown on Figure 3.5.1. The blue strip has approximately North-

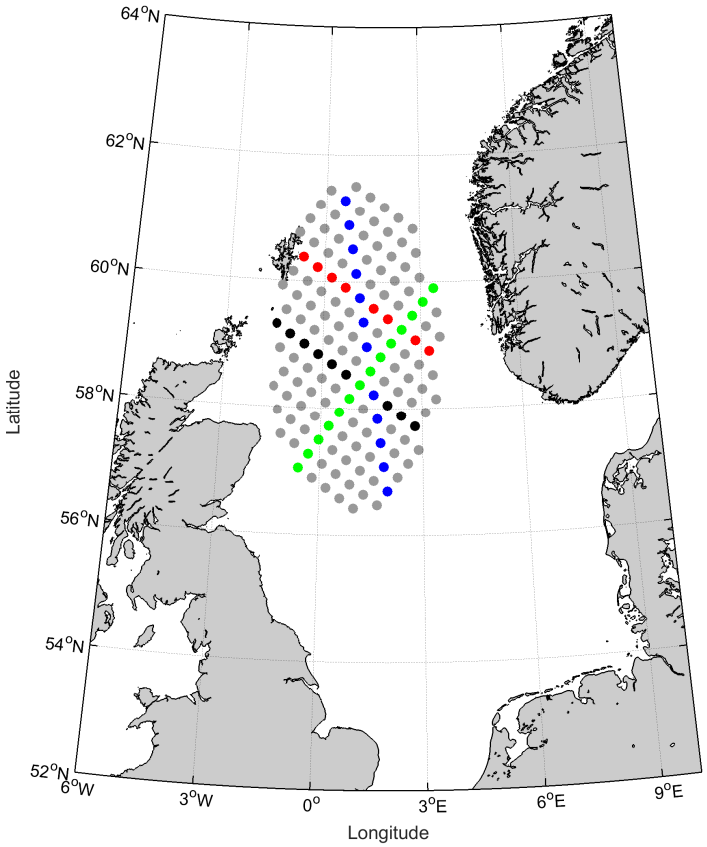


Figure 3.5.1: Map of location of data, showing four colour-coded sets of locations lying on straight lines with particular orientations, referred to in the text as "strips".

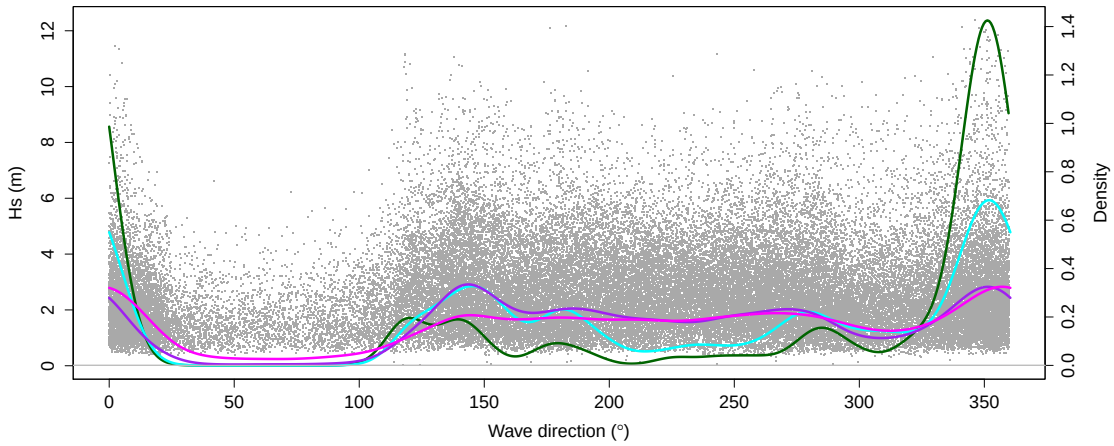


Figure 3.5.2: Sea-state H_S against wave direction θ for a central location, and corresponding density estimates for $\theta|H_S > v$ for $v = 0$m (magenta), $v = 4$m (purple), $v = 8$m (cyan) and $v = 10$m (dark green).

South orientation with long along-strip fetches in both directions. The red and black strips are both aligned North-West – South-East. The westerly end of the red strip is in the land shadow of the Shetland islands, but the westerly end of the black strip is located between the Shetland and Orkney islands, so has a long fetch from the North-West Atlantic Ocean. The green strip has a South-West – North-East orientation with along-strip fetch limited by the land masses of the Britain Isles and Norway.

3.5.2 Identification of extremal dependence

Figure 3.5.3 illustrates that the dependence between locations on each of the strips varies with inter-location separation. In Figure 3.5.3(a), plots of convex hulls for sea-state H_S from pairs of locations on the black strip corresponding to inter-location separations of 1, 5 and 10 are shown (or separations of 0.5°, 2.3° and 4.5° respectively). Figure 3.5.3(a), and corresponding figures for other strips (not shown), suggests that dependence reduces with increasing separation as expected. Because of the orientation of the location grid used for the hindcast, inter-location separation along the red and black strips is approximately the same; however, inter-location separations along the green and blue strips are different.

Figure 3.5.3(b) shows convex hulls of sea-state H_S for pairs from all strips with inter-location separation of approximately 2.4°, achieved by taking an inter-location separation of 5 for the red and black strips, 6 for the blue and both 6 (approximately 2.2°) and 7 (approximately 2.6°) for the green strip. From Figure 3.5.3(b) it is difficult to identify differences in dependence between strips for pairs at a separation of approximately 2.4°; however there is some suggestion that the dependence between sea-state H_S for pairs of locations on the red and black strips is similar. Blue and green strips also appear to be relatively similar. The corresponding plots for the storm-peak sample are given in Figure 3.5.3(c, d), with very similar characteristics.

We next estimate ρ, η and $\chi(x)$ for sea-state and storm-peak H_S for pairs of locations on each of the four strips. Each of η, $\chi(x)$ and ρ were estimated for sea-state H_S at pairs of locations on the four strips shown on Figure 3.5.1. We used the estimator $\widehat{\eta}_3$ for η estimation, and $\widehat{\chi}_{EMP}(x)$ for $\chi(x)$ estimation, as in Section 3.4.4. The estimates are shown in Figure 3.5.4. Equivalent estimates for simulated samples of the same size from the Smith and Gaussian processes and the inverted logistic distribution are also shown. This enables us to compare the patterns of relationship between η, $\chi(x)$ and ρ for the different asymptotically dependent and asymptotically independent processes that we observed in Figure 3.4.1 with the pattern the data exhibits. It appears that behaviour on the blue (North-South) and green (South-West – North-East) strips is reasonable well represented by the Gaussian model, whereas the dependence characteristics on black and red strips (North-West – South-East) are better captured by the inverted logistic model. Both the functional relationships between η and ρ and $\chi(x)$ and ρ support this fit. This is particularly clear for higher threshold probabilities q and levels x. Both these best-matching distributions are asymptotically independent. This clearly shows how our novel approach of supplementing measures of extremal dependence with ρ improves diagnostic performance.

We further estimate η, $\chi(x)$ and ρ for pairs of values of storm-peak H_S from locations on each of the four strips. Figure 3.5.5 illustrates these estimates together with the corresponding

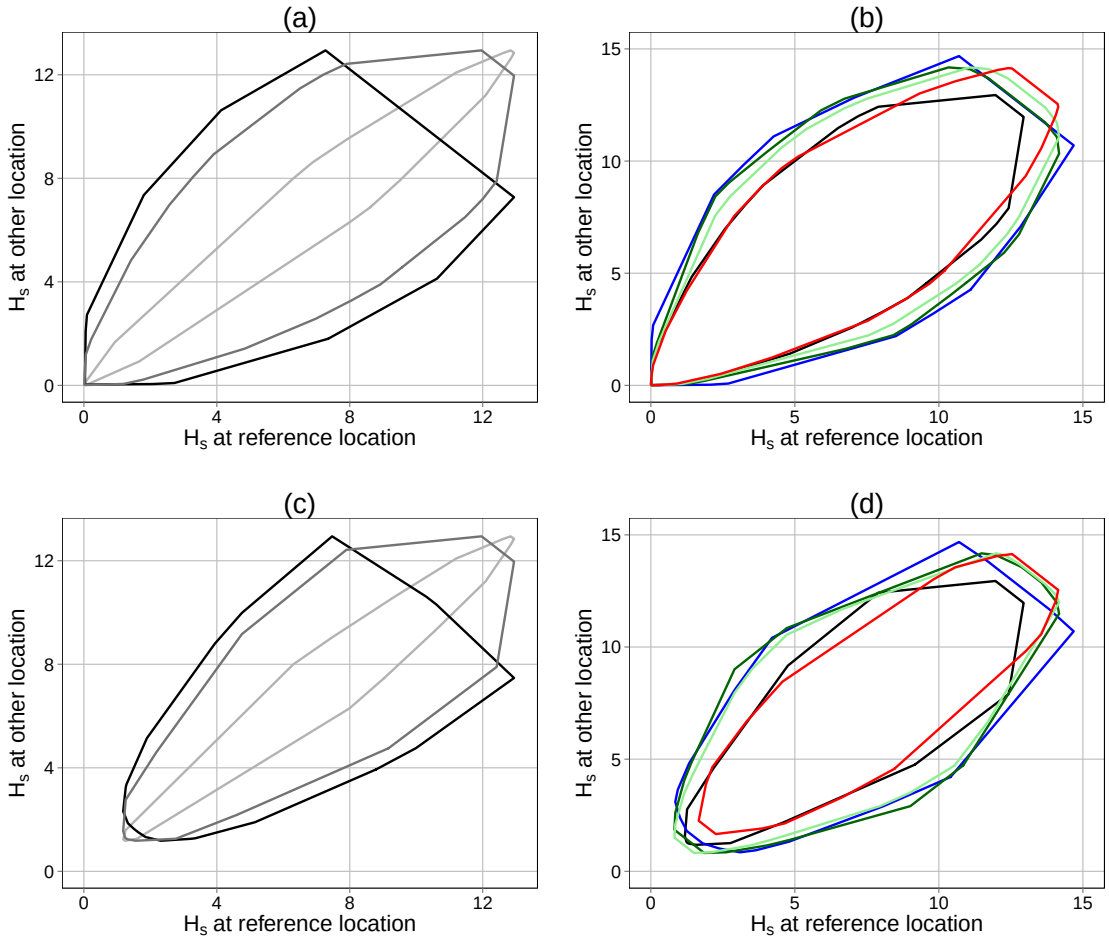


Figure 3.5.3: Convex hulls for pairs of sea-state significant wave heights H_S illustrating dependence. (a) Convex hulls for locations from the black strip (see Figure 3.5.1) are plotted, corresponding to inter-location separation of 0.5° (light grey), 2.3° (grey) and 4.5° (black); (b) convex hulls for locations with inter-location separation of approximately 2.4° are plotted, for all strips, coloured accordingly. Since there are no pairs of locations on the green strip corresponding to separation of 2.4°, those with separation 2.2° (light green) and 2.6° (dark green) are shown. The corresponding convex hulls for the storm-peak sample are shown in plots (c) and (d).

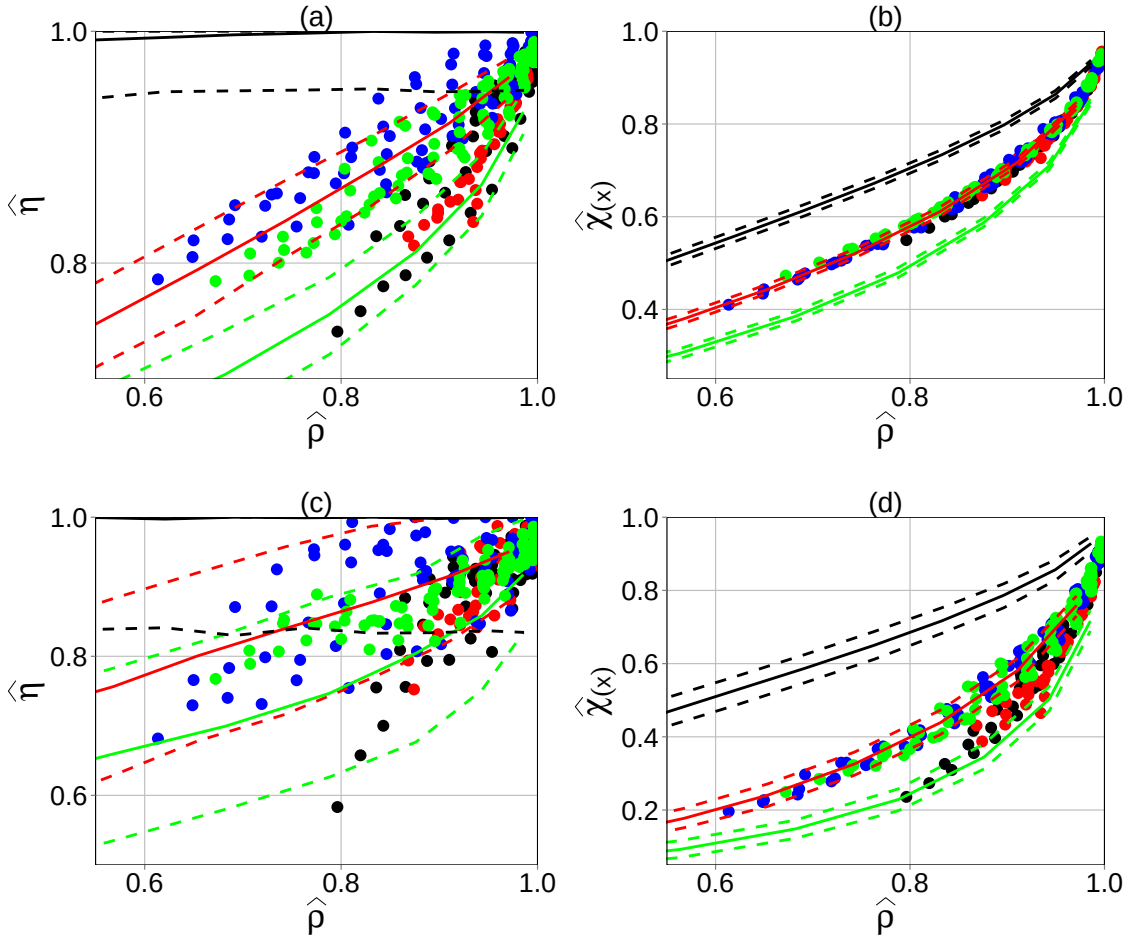


Figure 3.5.4: Estimates of η with (a) $q = 0.90$ and (c) $q = 0.99$, and $\chi(x)$ with (b) $x = 10$ and (d) $x = 100$, plotted against Spearman's ρ for sea-state H_S sample of size $n = 58585$. Coloured points identify estimates from corresponding strip. Lines identify estimates using simulated samples of same size from Smith (black) and Gaussian (red) processes, and from the inverted logistic distribution (green).

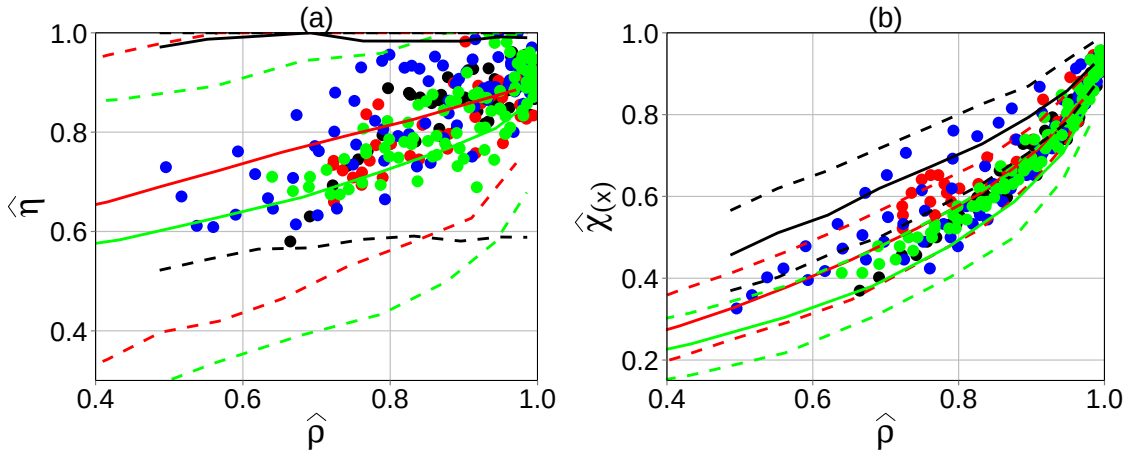


Figure 3.5.5: Estimates of (a) η with $q = 0.90$ and (b) $\chi(x)$ with $x = 10$, plotted against Spearman's ρ for storm-peak H_S sample of size $n = 916$. Points and lines as described in Figure 3.5.4.

estimates from simulated samples of the same size from the Smith and Gaussian processes and the inverted logistic distribution. It appears that storm-peak H_S dependence is more consistent with the asymptotically independent Gaussian and inverted logistic models for all strips, but due to small sample size there is a large overlap between the estimated uncertainty bands for different models. Inferring the nature of extremal dependence from samples of this size is extremely difficult.

3.5.3 Design implications

The choice of extremal dependence structure of storm-peak significant wave height values at neighbouring locations has material implications for structural design and reliability assessment. To illustrate the type of difference, consider the storm-peak significant wave heights (X, Y) at two of our North Sea locations separated along a strip by approximately 2.6°. If the risk of joint extremes at the two locations is of interest, then we would want to find the distribution of Y given that X was equal to its 100 year return level, denoted by

x_{100}, i.e. we want to find y_p, the pth conditional quantile, such that

$$P(Y < y_p | X = x_{100}) = p.$$

For simplicity of calculation we make the assumption that the marginal distribution of X and Y is identical to a typical site in our region shown in Figure 3.5.1. This is a reasonable assumption, as over this spatial scale, away from land shadows, the marginal distributions do not change rapidly. As mentioned in Section 3.1, a naive approach is to assume perfect dependence, and hence $y_p = x_{100}$ for all p. Under strong asymptotic dependence similar values for y_p arise. The question is what reduction can we get in y_p from using our best fitting model for extremal dependence, i.e., an asymptotically independent distribution, which we take here as the Gaussian dependence structure. For the storm-peak significant wave heights we take a threshold u and assume that the marginal conditional distribution of threshold exceedances is described by a generalised Pareto distribution

$$F(x) = P(X < x | X > u) = P(Y < x | Y > u) = 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}, \text{ for } x > u,$$

where ξ and $\sigma > 0$ are the shape and scale parameters respectively and $t_+ = \max(t, 0)$. The threshold u is selected using parameter stability plots (Coles 2001). Here we choose $P\{X > u\} = \phi = 0.5$, and estimate the parameters ξ and σ using maximum likelihood. For our data, $u = 6.66$m and the estimates are $(\widehat{\xi}, \widehat{\sigma}) = (-0.23, 2.48)$. Then the expected number of storm peaks above some value x in 100 years is

$$100 \times n_{py} \times \phi \times \left[1 + \widehat{\xi} \frac{x-u}{\widehat{\sigma}}\right]^{-1/\widehat{\xi}}, \tag{3.5.1}$$

where n_{py} is the average number of storm peaks per year ($n_{py} = 30.11$ for our data). Setting expression (3.5.1) to 1 when $x = x_{100}$, and rearranging, we find the marginal 100-year return value x_{100} for location X to be the marginal quantile of storm peaks with non-exceedance probability $F(x_{100}) = 1 - (100 n_{py} \phi)^{-1}$, corresponding to a value $x_{100} \approx 15.5$m. Thus under

perfect dependence $y_p = 15.5$m.

The conditional extremes model of Heffernan and Tawn (2004) is most straightforward to apply on Laplace margins. To transform to Laplace margins we use expression (3.2.2) which transforms quantiles of F to the same quantile of the Laplace distribution F_L. Hence the transformed 100-year level on Laplace scale is x^L_{100} where $F(x_{100}) = F_L(x^L_{100})$, giving $x^L_{100} = -\log[2(100 n_{py}\phi)^{-1}]$. Now suppose further that the extremal dependence between the locations corresponds approximately to the asymptotically independent Gaussian process with correlation ρ_P, which is consistent with our findings of Section 3.5.2. We had found in our analysis that two locations at separation of approximately 2.6° correspond to a Spearman's ρ value of approximately 0.8 (see Figure 3.5.5), which, in turn, corresponds to a Gaussian process with Pearson correlation coefficient of $\rho_P = 0.8$. The conditional extremes model of Heffernan and Tawn then suggests for Gaussian processes, with Laplace margins (X_L, Y_L), that

$$P\left\{\frac{Y_L - \rho_P^2 X_L}{X_L^{1/2}} < z \middle| X_L = x^L_{100}\right\} \approx \Phi\left(\frac{z}{\rho_P(1-\rho_P^2)^{1/2}}\right); \tag{3.5.2}$$

see Section 8 of Heffernan and Tawn (2004) for more details. If we assume that approximation (3.5.2) holds for x^L_{100} then the pth quantile of the distribution of $Y_L|(X_L = x^L_{100})$ is

$$y^L_p = \rho_P^2 x^L_{100} + \rho_P(1-\rho_P^2)^{1/2}(x^L_{100})^{1/2}\Phi^{-1}(p), \tag{3.5.3}$$

where Φ is the distribution function of a standard normal variable. Here y^L_p is on Laplace scale, and converting back to the original scale, quantile to quantile transformation suggests that y_p satisfies $F(y_p) = F_L(y^L_p)$. This means that the associated quantile of storm-peak significant wave height at the neighbouring location will be

$$y_p = u + \frac{\widehat{\sigma}}{\widehat{\xi}}\left(\left\{\frac{1}{2\phi}\exp\left(-y^L_p\right)\right\}^{-\widehat{\xi}} - 1\right),$$

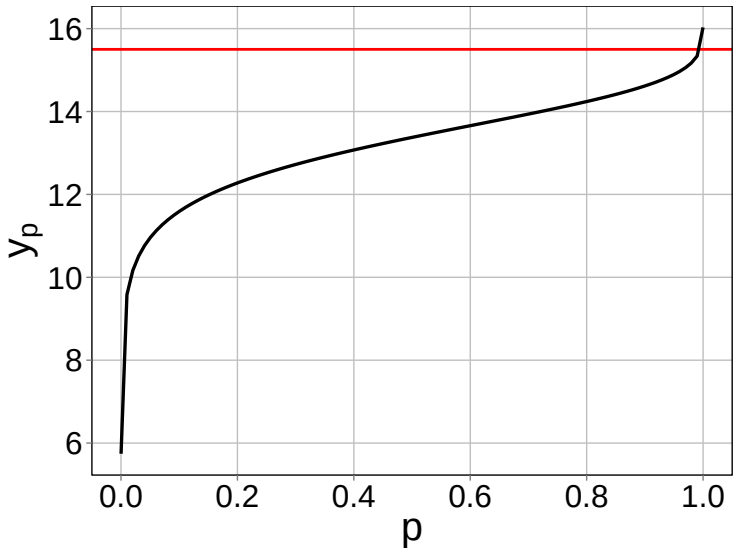


Figure 3.5.6: Values of y_p plotted against values of p, where p is a quantile of the distribution of $Y_L|(X_L = x^L_{100})$.

where y^L_p is given by expression (3.5.3). For our data the y_p values are plotted against the quantile p (see Figure 3.5.6). For almost all values of p, $y_p \ll 15.5$m, so the estimated risk is much less than under perfect dependence; for example, the median is 2m lower. Only for ρ_P very close to 1 is there essentially no practical difference between the two approaches.

3.6 Discussion and conclusions

The extremal dependence between contemporaneous observations of significant wave height at neighbouring locations determines the rate and size of occurrences of joint extremes at those locations. In this work we examined the extremal dependence of significant wave height for pairs of locations from a North Sea hindcast. We showed that the characteristics of extremal dependence vary systematically with distance between locations and with orientation of locations, given prevailing atmospheric conditions and fetch variation. Using the variation of coefficient of tail dependence η and the $\chi(x)$ statistic with (full-sample) Spearman rank correlation ρ to quantify extremal dependence, we find that a sample of sea-state significant wave height exhibits asymptotic independence consistent with either

the Gaussian process or inverted logistic models. The nature of extremal dependence for a sample of storm-peak significant wave height is very difficult to estimate, due to its small size of approximately 1000 in comparison to 58000 for the full sample. However, again the extremal dependence of the sample examined is consistent with either the Gaussian process or the inverted max-stable model. Had we found instead that our sample exhibits asymptotic dependence, a choice between various max-stable model needs to be made. Our diagnostics are not ideal for this; for more details about model selection for asymptotic dependence see Davison et al. (2012).

We examined the behaviour of different estimators for η and $\chi(x)$ for simulated samples from processes with different known extremal dependencies. We found that, in general, different estimators yield consistent estimates for η and $\chi(x)$. As might be expected, the bias of estimates reduces with increasing tail threshold, and variance of estimates reduces with increasing tail sample size. For samples with sizes of the order of 10^4 typical for metocean applications, interpretation of plots of estimated η and $\chi(x)$ as a function of estimated ρ is difficult. In particular, when estimated ρ is small, η estimated from asymptotically dependent processes is biased, and $\chi(x)$ from asymptotically independent processes only approaches the limiting value of zero for levels x that are unachievable in practice. It appears that comparison of sample-based plots of η and $\chi(x)$ with ρ for significant wave height with those for samples of the same size from bivariate processes with known extremal dependence provides more intuitive interpretation.

The extremal spatial dependence of sea-state and storm-peak significant wave height in the North Sea appears, at the very least, not to be inconsistent with asymptotic independence, implying that extreme value models admitting asymptotic independence should be sought for such applications. The great majority of models for spatial extremes which have been applied to environmental problems, motivated by max-stable assumptions, admit asymptotic dependence only; these are not ideal therefore for the current application. The conditional extremes model of Heffernan and Tawn (2004) provides a potential solution, although the

approach does not lend itself naturally to spatial modelling over continuous space. The hybrid model of Wadsworth and Tawn (2012a) is an alternative more suitable to spatial application; but considerable effort would be needed to develop a reliable engineering implementation. We found that the asymptotically independent Gaussian process and the inverted max-stable processes seem to fit well to the data. As the inverted max-stable models benefit from the recent investment of model development in max-stable processes, we foresee future spatial modelling of extremes of H_S using these models.

Inspection of Figure 3.5.5 shows that identifying the appropriate form of extremal dependence for storm-peak significant wave height is problematic for samples of size approximately 1000. Since this is critical to reliable design, and since samples of size less than 1000 are typical for metocean applications, it is appealing to seek means of improved identification of extremal dependence for storm-peak H_S. The obvious approach is to gather larger samples of storm-peak values, but this is often impossible in practice. Another approach might be to assume that the dependence between contemporaneous values of sea-state H_S (for which sample size is large, and identification of extremal dependence somewhat easier) is similar if not equal to that of contemporaneous peak-over-threshold values. However, in reality, times of storm-peak occurrences at neighbouring locations are not the same; for example, they are obviously associated with the storm trajectory in space and time across the neighbourhood. Perhaps estimates of extremal dependence for time-series of sea-state H_S, appropriately time-lagged such that storm-peak events are contemporaneous, offer an approximate solution.

In some ocean basins, pooling of samples from locations within a spatial neighbourhood is performed routinely to increase sample size for extreme value modelling (see e.g., Heideman and Mitchell 2009). Data from the different locations that are pooled are assumed to follow a common marginal extreme value distribution, so that any marginal spatial non-stationarity is neglected. Moreover, the extremal dependence of values from different locations is also typically ignored; at best, uncertainties in estimated return values are inflated to reflect de-

pendence using a suitable bootstrapping scheme. A well-informed spatial extremes model, accommodating the appropriate form of extremal dependence, offers reduced bias and improved uncertainty quantification.

The diagnostic tools explored in this work are of general relevance to any application in which understanding the characteristics of the joint occurrences of extreme events is of interest. For example, the nature of extremal dependence between H_S at two locations may differ from location to location within an ocean basin, and between ocean basins. Application of the diagnostic analysis suggested here for other locations in the North Sea, and for other ocean basins, would quantify this. We might explore the spatial behaviour of joint extremes of variables other than H_S, such as wind speed between locations (or altitudes), or current speed between locations (or water depths). There is no reason to expect that wind speed and current speed exhibit the same form of extremal dependence as H_S, and the form of dependence may differ temporally depending on prevailing conditions at the locations of interest. For example, if we are interested in understanding the loading on risers of a moored vessel, appropriate models for extreme current profiles with depth are required, incorporating appropriate forms for extremal dependence; otherwise biased estimates of riser loads and tensions may result, as explained in Section 3.5 above.

The current work is also relevant in improving understanding of engineering design practices which implicitly assume, often with little justification, a particular form of extremal dependence between variables: an estimate of return value for a particular oceanographic variable, with corresponding estimate of associated return value for another related variable, falls into this category. For example, any parametric model for the joint structure of two oceanographic variables (e.g., H_S and peak wave period T_P, or H_S and wind speed) is usually estimated by fitting the model to observations, usually not exclusively of jointly extreme events. We therefore expect that the assumed parametric form explains the bulk of observations well. But the parametric form also imposes a specific structure on the extremal dependence between variables which may or may not be justified. Estimating the type of

extremal dependence present explicitly using the current diagnostics for jointly extreme events is at least advisable if not essential in demonstrating that the parametric form used is also appropriate for joint extremes.

When there is doubt concerning the nature of extremal dependence present, conventional wisdom is that the metocean engineer should err on the side of caution and assume asymptotic dependence, since this will yield more and larger joint occurrences of large events. However, this approach too may have undesirable knock-on consequences. For example, consider the design of a multi-component system. The incorrect assumption of asymptotic dependence for variables in one component (which actually exhibit asymptotic independence) might result in greater build cost for that component. When overall build cost is constrained, this might lead to less build resources for other system components. That is, over-conservatism in one aspect of design causes lack of conservatism in others, and hence overall increased structural risk. A sensible approach in such circumstances is to balance risk in a statistically valid manner with respect to competing design requirements, requiring consistent assessment of risk throughout design. In a metocean context, specification of consistent return values for a set of oceanographic variables is challenging, because of the difficulty of quantifying the dependence between extreme values of two or more variables. The current work provides diagnostics to assist in quantifying extremal dependence rationally.

3.A Derivation of the PC prior

Simpson et al. (2015) use the Kullback-Leibler divergence (KLD) to measure the increased complexity from g to f. In other words, it is a measure of the information lost when the base model g is used to approximate the more flexible model f. This is defined by

$$\mathrm{KLD}(f||g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx.$$

For $g(x)$ and $f(x)$ as defined in Section 3.3.1, we can calculate the KLD between g and f as,

$$\text{KLD}(f||g) = -\log\eta + \eta - 1.$$

This can be transformed onto a physically interpretable distance scale:

$$\begin{aligned} d_{fg} &= \sqrt{2\text{KLD}(f||g)} \\ &= \sqrt{-2\log\eta + 2(\eta-1)} \end{aligned}$$

Simpson et al. (2015) assume that the penalisation rate is constant, which implies an exponential prior on the distance scale. Hence, $d_{fg} \sim$ Exponential(λ) and $\pi_\lambda(d_{fg}) = \lambda\exp(-\lambda d_{fg})$, $\lambda > 0$. Since d_{fg} is a function of η ($d = \sqrt{-2\log\eta + 2(\eta-1)}$), we can use Jacobian transformation to obtain a prior for η; i.e. $\pi_\lambda(\eta) = \pi_\lambda(d_{fg})\cdot|J|$, where $J = dd_{fg}/d\eta$. This gives the following prior for η,

$$\pi(\eta;\lambda) = \frac{\lambda(1-\eta)\exp\left(-\lambda\sqrt{-2\log\eta + 2(\eta-1)}\right)}{\eta\sqrt{-2\log\eta + 2(\eta-1)}}.$$

3.B Results for other processes

In this section we give equivalent results to those shown on Figure 3.4.6 for other max-stable processes. For computational reasons, here we only used 100 replications, but this does not affect the main conclusions we draw. As all max-stable processes are AD, we expect $\widehat{\eta} \approx 1$ and $\widehat{\chi} > 0$.

3.B.1 Schlather process

Figure 3.B.1 shows η and χ estimates against Spearman's ρ for simulated samples from the Schlather process. Due to the constraint mentioned in Section 3.2.3, we cannot simulate samples with low dependence. Hence, $\widehat{\eta} \approx 1$ and $\widehat{\chi} > 0$ everywhere.

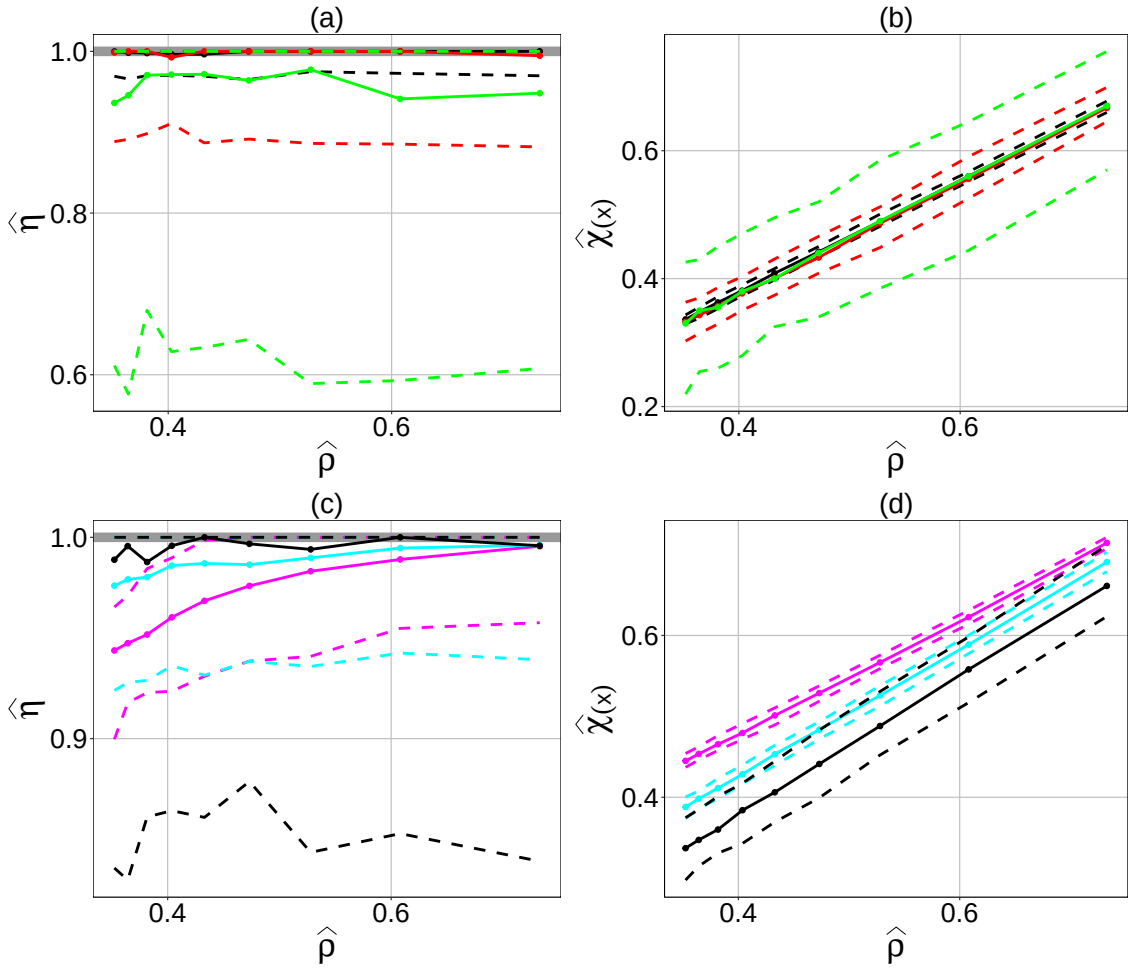


Figure 3.B.1: Estimates of η and χ plotted against Spearman's ρ for simulated data from the Schlather model; sample size: $n = 10^6$ (top row), $n = 58585$ (bottom row). Threshold probabilities for the η estimation are $q = 0.9999$ (green line), $q = 0.999$ (red line), $q = 0.99$ (black line), $q = 0.90$ (cyan line) and $q = 0.80$ (magenta line). Corresponding levels for the $\chi(x)$ estimation are $x = 10000$ (green line), $x = 1000$ (red line), $x = 100$ (black line), $x = 10$ (cyan line) and $x = 5$ (magenta line). Solid lines give the median from a 100 sample replications, dashed lines give the 2.5% and 97.5% quantiles. The solid grey line shows a correct limiting value for η.

3.B.2 Brown-Resnick process

The Brown-Resnick process behaves similarly to the Smith process in terms of its extremal properties. Figure 3.B.2 shows that $\widehat{\eta}$ is estimated as approximately 1 for higher dependence ρ, but it is underestimated for lower dependence, especially for small sample sizes and lower thresholds. The χ estimates are greater than 0, as it is expected for asymptotically dependent models.

3.B.3 Extremal-t process

Results for the extremal-t process are very similar to those for the Brown-Resnick process. Figure 3.B.3 shows that $\widehat{\eta}$ is estimated as approximately 1 when the general level of dependence in the data is stronger (as measured by ρ), but $\widehat{\eta}$ is underestimated when the dependence is weaker. The χ estimates are greater than 0 here too, as it is expected for asymptotically dependent models.

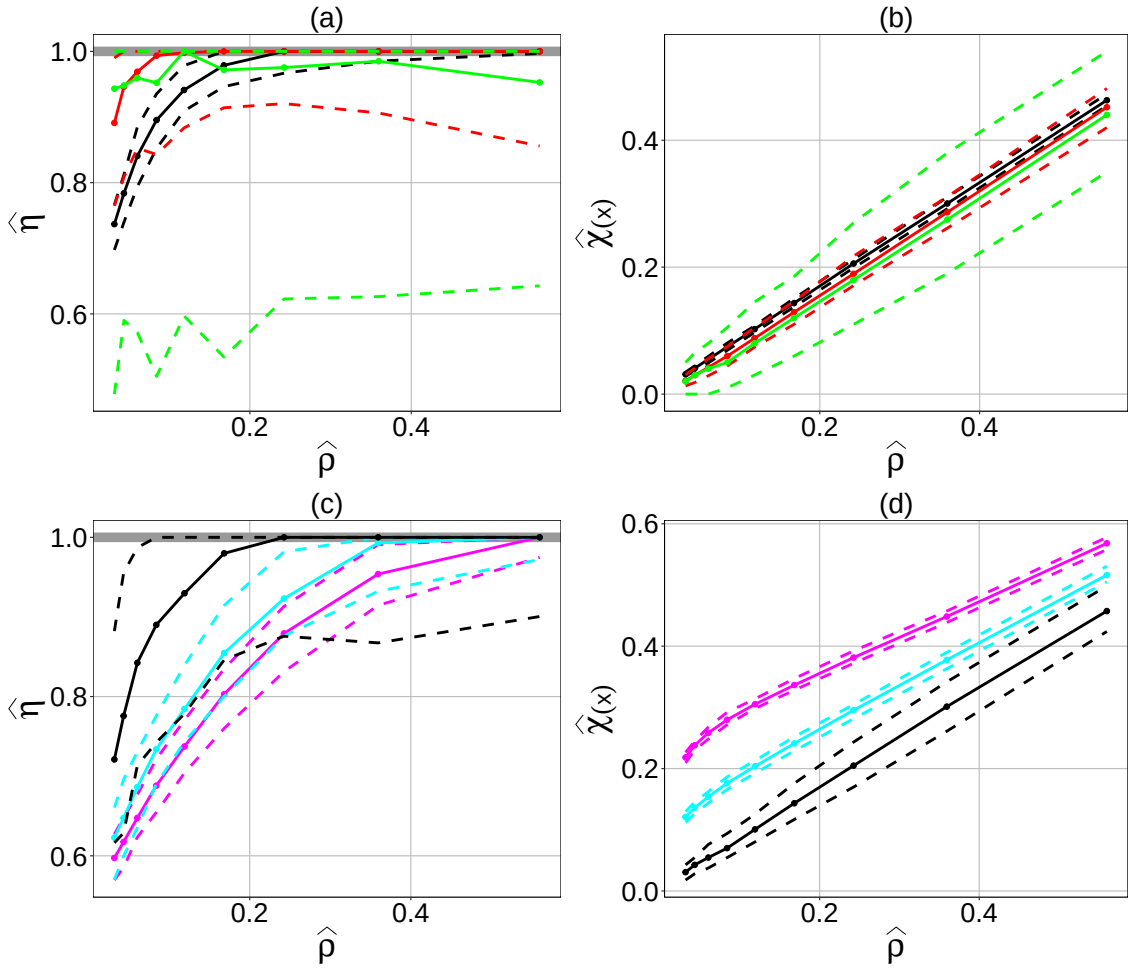


Figure 3.B.2: Estimates of η and χ plotted against Spearman's ρ for simulated data from the Brown-Resnick model; sample size: $n = 10^6$ (top row), $n = 58585$ (bottom row). Threshold probabilities for the η estimation are $q = 0.9999$ (green line), $q = 0.999$ (red line), $q = 0.99$ (black line), $q = 0.90$ (cyan line) and $q = 0.80$ (magenta line). Corresponding levels for the $\chi(x)$ estimation are $x = 10000$ (green line), $x = 1000$ (red line), $x = 100$ (black line), $x = 10$ (cyan line) and $x = 5$ (magenta line). Solid lines give the median from a 100 sample replications, dashed lines give the 2.5% and 97.5% quantiles. The solid grey line shows a correct limiting value for η.

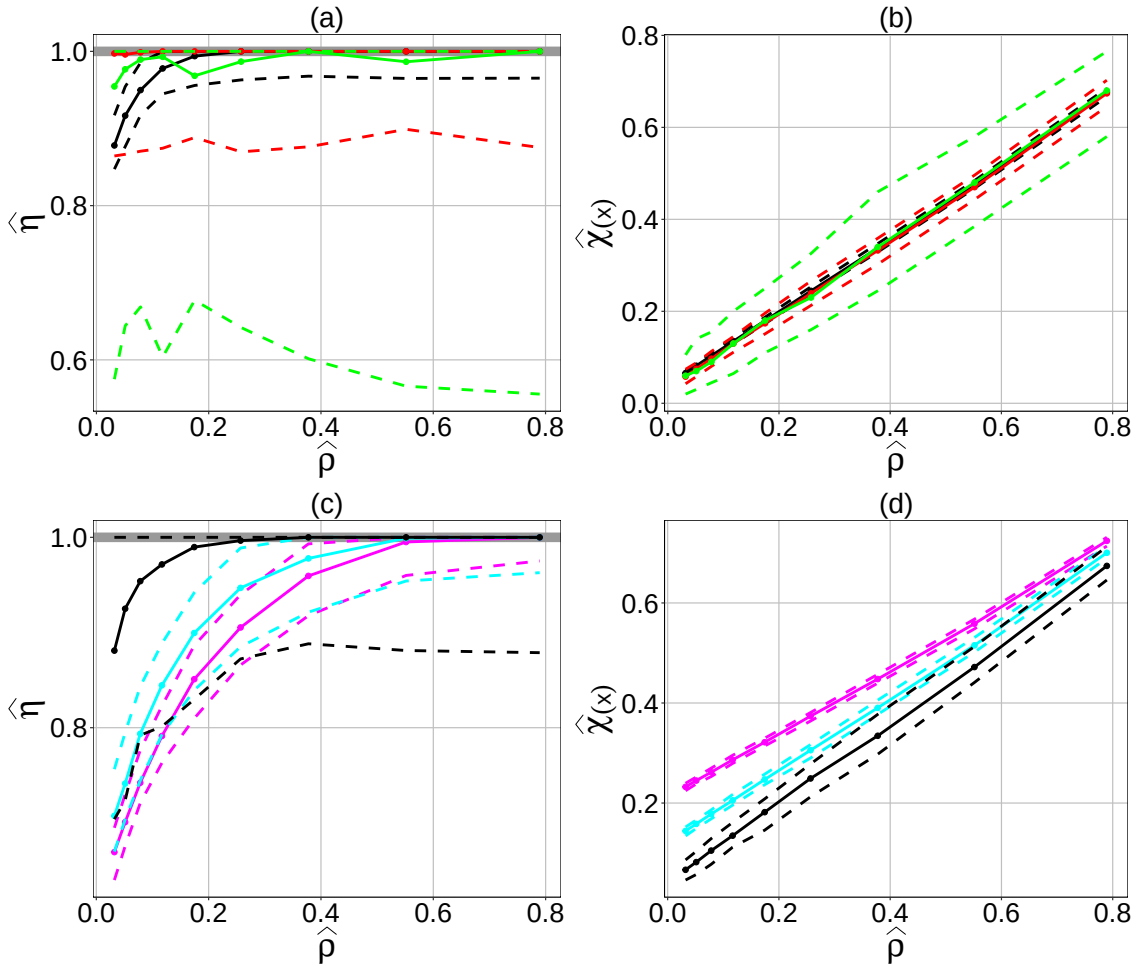


Figure 3.B.3: Estimates of η and χ plotted against Spearman's ρ for simulated data from the extremal-t model; sample size: $n = 10^6$ (top row), $n = 58585$ (bottom row). Threshold probabilities for the η estimation are $q = 0.9999$ (green line), $q = 0.999$ (red line), $q = 0.99$ (black line), $q = 0.90$ (cyan line) and $q = 0.80$ (magenta line). Corresponding levels for the $\chi(x)$ estimation are $x = 10000$ (green line), $x = 1000$ (red line), $x = 100$ (black line), $x = 10$ (cyan line) and $x = 5$ (magenta line). Solid lines give the median from a 100 sample replications, dashed lines give the 2.5% and 97.5% quantiles. The solid grey line shows a correct limiting value for η.

Chapter 4

Directional Effects in the Extremal Dependence of North Sea Storm Severity

4.1 Introduction

In Chapter 3 we have estimated various extremal dependence measures for samples of North Sea sea-state and storm-peak significant wave height. We have found that there is typically not sufficient evidence from these extremal dependence measures for them to be effective diagnostic tools when identifying the form of the extremal dependence, e.g., whether the data are asymptotically dependent (AD) or asymptotically independent (AI). However, supplementing these measures of extremal dependence with a dependence measure for the body of the data improves diagnostic performance, particularly when comparing estimates from the data to estimates for known AI and AD processes.

Application of this novel diagnostic method to the North Sea data showed that the dependence characteristics of the data are better captured by AI models. We also found that

data along locations on strips with different orientations have different extremal dependence structures. In this chapter we further investigate the effects of direction on extremal dependence in samples of North Sea storm-peak significant wave height. Since we are interested in the effects on extremal dependence and not marginal effects, the data set has been marginally "whitened". At each location a directional Gamma-generalised Pareto mixture model has been fitted independently and the data transformed to unit Fréchet margins, thus the marginal variables are all independent of direction. For more details on this procedure see Jonathan et al. (2014) and Randell et al. (2016). This transformation ensures that any directional effects we see should be due to differences in the dependence structure rather than common marginal effects across space. Similarly to Chapter 3, we examine the effects of direction by examining extremal dependence between locations on six straight-line strips, shown in Figure 4.1.1. For each of these six strips we also look at nine parallel strips (four on one side and 5 on the other side of the strips shown on Figure 4.1.1), so in total there are 60 strips, ten with each of the orientations shown on Figure 4.1.1.

Simplistically, we can imagine a storm propagating along a straight line in time as a half-plane in space. Symmetry then suggests that the extremal dependence spatially would exhibit limiting forms (i) along the storm trajectory and (ii) perpendicular to the storm trajectory (i.e., along the edge of the half plane representing the storm front). In the North Sea, storms that cause large waves mostly travel in an approximately north to south direction, hence, we would expect extremal dependence to be the strongest either along strips with an approximately north to south orientation (in line with (i)) or along strips with an approximately west to east orientation (in line with (ii)). To determine if this is indeed the case, we fit various AI and AD models to data along strips with different orientations using composite likelihood methods. The models used are four of the processes presented in Chapter 3: two AD processes (Smith and Schlather max-stable processes) and two AI processes (Gaussian and inverted Smith processes). In Chapter 3 we used an inverted logistic extreme value distribution, which is a simple form of an inverted max-stable

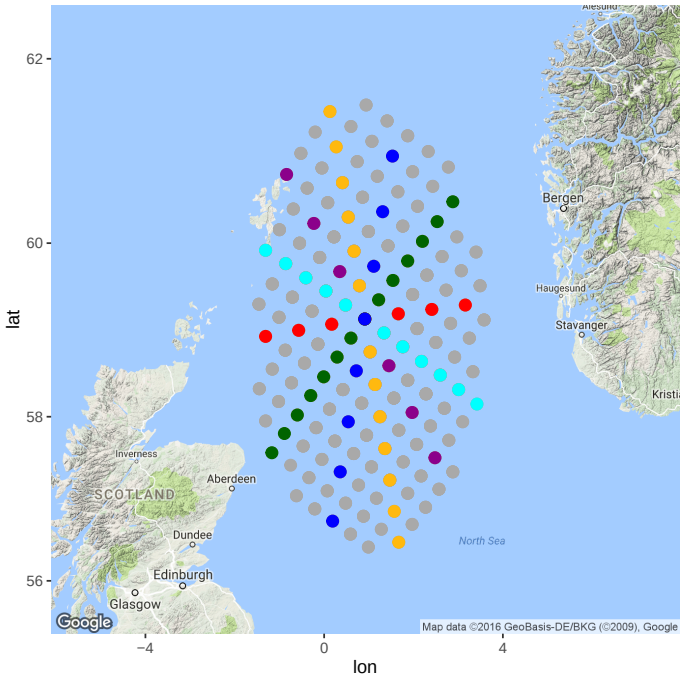


Figure 4.1.1: Map of location of data, showing six colour-coded sets of locations lying on straight lines with particular orientations, referred to in the text as "strips".

distribution. For the purposes of the analysis in Chapter 3 this model was suitable as the analysis was pairwise, but since here we are fitting models to spatial data it makes more sense to use an inverted Smith max-stable process as this already has an in-built spatial structure.

This chapter is structured as follows. In Section 4.2 we introduce the composite and censored likelihood methods we use for inference in this chapter. In Section 4.3 we derive the necessary results for the four models listed above. Then in Section 4.4 we fit these models to the North Sea storm-peak wave height data set and present the results. We close with a discussion in Section 4.5.

4.2 Inference

Density functions for max-stable processes are intractable in higher dimensions, so standard likelihood methods are difficult to implement. There is some recent work that tries to overcome this problem. Wadsworth and Tawn (2013a) exploit the structure of a specific class of max-stable processes and perform full likelihood inference for this class if the occurrence times of spatial componentwise maxima are available. More generally, Thibaud et al. (2016) use Bayesian MCMC methods to conduct inference for the Brown-Resnick max-stable process. Padoan et al. (2010) introduced what is regarded to be the standard approach to practical methods for fitting max-stable processes to spatial data using composite likelihood-based methods. Using this latter approach the pairwise composite log-likelihood is

$$l(\boldsymbol{\psi}; \mathbf{z}) = \sum_{n=1}^{N} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \log f_{i,j}(z_{n,i}, z_{n,j}; \boldsymbol{\psi}), \tag{4.2.1}$$

where $\{z_{n,k}\}$ denote the maximum of m samples over $n = 1, \ldots, N$ blocks and $k = 1, \ldots, K$ locations, $f_{i,j}$ is a marginal bivariate density function based on data at locations i and j, and $\boldsymbol{\psi}$ is the parameter vector.

As max-stable processes arise as the limit distribution of componentwise maxima, when applied to event data they only provide a valid model when two events are simultaneously extreme at both locations. Hence, we adopt a censored likelihood approach (see, e.g., Coles (2001) and Huser and Davison (2012)). Let us assume that u is a sufficiently high threshold for some bivariate distribution $F(z_1, z_2)$ to be a valid model for $z_1 > u$ and $z_2 > u$. Then

the likelihood contribution $L_u(z_1, z_2)$ of the pair (z_1, z_2) is

$$L_u(z_1, z_2) = \begin{cases} \frac{\partial^2}{\partial z_1 \partial z_2} F(z_1, z_2), & z_1 > u, z_2 > u; \\ \frac{\partial}{\partial z_1} F(z_1, u), & z_1 > u, z_2 \leqslant u; \\ \frac{\partial}{\partial z_2} F(u, z_2), & z_1 \leqslant u, z_2 > u; \\ F(u, u), & z_1 \leqslant u, z_2 \leqslant u. \end{cases}$$

In the following section we derive these first and second derivatives for each of the models we use.

4.3 Models

In this section we derive the necessary results for fitting each model to spatial data. We also present the extremal coefficient (or the sub-asymptotic extremal coefficient) for each model, as this gives a more natural way of comparing the models than the model parameters themselves.

4.3.1 Smith process

The bivariate distribution function for the Smith max-stable process can be written as

$$F_{\mathbf{h}}(z_1, z_2) = \exp\left(-\frac{\Phi(w(\mathbf{h}))}{z_1} - \frac{\Phi(v(\mathbf{h}))}{z_2}\right),$$

where $z_1 > 0$ and $z_2 > 0$, and Φ is the standard Gaussian distribution function, $\mathbf{h} = (\mathbf{t}_2 - \mathbf{t}_1)^T$ for locations $\mathbf{t}_1$ and $\mathbf{t}_2$. The terms $w(\mathbf{h})$ and $v(\mathbf{h})$ are defined as

$$w(\mathbf{h}) = a(\mathbf{h})/2 + \log(z_1/z_2)/a(\mathbf{h}),$$

and

$$v(\mathbf{h}) = a(\mathbf{h}) - w(\mathbf{h}),$$

with $a(\mathbf{h}) = (\mathbf{h}^T\Sigma^{-1}\mathbf{h})^{1/2}$ the Mahalanobis distance, where Σ is a covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \tag{4.3.1}$$

Hence the parameter set of the Smith process is $\boldsymbol{\psi} = (\sigma_1, \sigma_2, \sigma_{12})$.

Following the derivation in Padoan et al. (2010), the first derivatives are

$$\frac{\partial}{\partial z_1} F_{\mathbf{h}}(z_1, z_2) = \left(\frac{\Phi(w(\mathbf{h}))}{z_1^2} + \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h}) z_1^2} - \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h}) z_1 z_2} \right) \exp\left(-\frac{\Phi(w(\mathbf{h}))}{z_1} - \frac{\Phi(v(\mathbf{h}))}{z_2} \right),$$

and

$$\frac{\partial}{\partial z_2} F_{\mathbf{h}}(z_1, z_2) = \left(\frac{\Phi(v(\mathbf{h}))}{z_2^2} + \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h}) z_2^2} - \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h}) z_1 z_2} \right) \exp\left(-\frac{\Phi(w(\mathbf{h}))}{z_1} - \frac{\Phi(v(\mathbf{h}))}{z_2} \right),$$

where φ is the standard Gaussian density function. The second derivative gives the bivariate density function

$$\frac{\partial^2}{\partial z_1 \partial z_2} F_{\mathbf{h}}(z_1, z_2) = f_{\mathbf{h}}(z_1, z_2),$$

which for the Smith process is

$$\begin{aligned} f_{\mathbf{h}}(z_1, z_2) = \exp\left(-\frac{\Phi(w(\mathbf{h}))}{z_1} - \frac{\Phi(v(\mathbf{h}))}{z_2} \right) \Bigg\{ \left(\frac{\Phi(w(\mathbf{h}))}{z_1^2} + \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h}) z_1^2} - \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h}) z_1 z_2} \right) \times \\ \left(\frac{\Phi(v(\mathbf{h}))}{z_2^2} + \frac{\varphi(v(\mathbf{h}))}{a(\mathbf{h}) z_2^2} - \frac{\varphi(w(\mathbf{h}))}{a(\mathbf{h}) z_1 z_2} \right) + \left(\frac{v(\mathbf{h})\varphi(w(\mathbf{h}))}{a(\mathbf{h})^2 z_1^2 z_2} + \frac{w(\mathbf{h})\varphi(v(\mathbf{h}))}{a(\mathbf{h})^2 z_1 z_2^2} \right) \Bigg\}. \end{aligned}$$

The extremal coefficient for the Smith process is $\theta(\mathbf{h}) = 2\Phi\{a(\mathbf{h})/2\}$.

4.3.2 Schlather process

From Schlather (2002), the bivariate distribution function for a Schlather max-stable process with latent correlation function $\rho(\mathbf{h})$ is

$$F(z_1, z_2) = \exp\left\{-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2(\rho(\mathbf{h}) + 1)\frac{z_1 z_2}{(z_1 + z_2)^2}}\right)\right\}.$$

Some commonly used correlation functions are listed in Chapter 2. Here we use a powered exponential correlation function of the form $\rho(\mathbf{h}) = \exp\{-(\mathbf{h}^T\Sigma^{-1}\mathbf{h})/2\}$, where Σ is a covariance matrix. This allows a parametrisation similar to the one used for the Smith process, with parameters $(\sigma_1, \sigma_2, \sigma_{12})$.

To obtain the first and second derivatives we need for the censored likelihood, first let $F^* = \log(F)$. Then, dropping arguments, the bivariate density can be derived as,

$$f(z_1, z_2) = \frac{\partial^2}{\partial z_1 \partial z_2} F(z_1, z_2) = \left(\frac{\partial F^*}{\partial z_1}\frac{\partial F^*}{\partial z_2} + \frac{\partial^2 F^*}{\partial z_1 \partial z_2}\right) F.$$

and the first partial derivatives as

$$\frac{\partial}{\partial z_1} F(z_1, z_2) = \frac{\partial F^*}{\partial z_1} F,$$
$$\frac{\partial}{\partial z_2} F(z_1, z_2) = \frac{\partial F^*}{\partial z_2} F.$$

Let us define the following

$$W = 1 - 2(\rho(\mathbf{h}) + 1)\frac{z_1 z_2}{(z_1 + z_2)^2},$$
$$A = W^{-1/2}\frac{\rho(\mathbf{h}) + 1}{2(z_1 + z_2)^3}.$$

Then the first partial derivatives of F^* are

$$\frac{\partial F^*}{\partial z_1} = \frac{1}{2z_1^2}(1+\sqrt{W}) + A\frac{z_2^2 - z_1^2}{z_1};$$
$$\frac{\partial F^*}{\partial z_2} = \frac{1}{2z_2^2}(1+\sqrt{W}) + A\frac{z_1^2 - z_2^2}{z_2},$$

with the second partial derivative being

$$\frac{\partial^2 F^*}{\partial z_1 \partial z_2} = A\left[2 - W^{-1}(\rho(\mathbf{h})+1)\left(\frac{z_1 - z_2}{z_1 + z_2}\right)^2\right].$$

The extremal coefficient for the Schlather process is $\theta(\mathbf{h}) = 1 + \sqrt{\{1-\rho(\mathbf{h})\}/2}$.

4.3.3 Gaussian process

Here we will consider Gaussian processes with univariate marginals transformed to Fréchet margins. First consider a stationary Gaussian process $\{X_t^*\}$ at location t, for t in some suitable set. Assume the Gaussian process has standard normal margins and correlation $\rho(\mathbf{h}) = \exp(-\mathbf{h}^T\Sigma^{-1}\mathbf{h})$, where $\mathbf{h} = (\mathbf{t}_1 - \mathbf{t}_2)^T$, for locations $\mathbf{t}_1$ and $\mathbf{t}_2$. Now consider the bivariate distribution of $(X_{t_1}^*, X_{t_2}^*)$ and denote these variables, for simplicity, as (X_1, X_2). Note that the transformations $Z_1 = -1/\log\Phi(X_1)$ and $Z_2 = -1/\log\Phi(X_2)$ give (Z_1, Z_2) on Fréchet margins. Then, on Fréchet margins,

$$\begin{aligned}
F_{\mathbf{h}}(z_1, z_2) &= \Pr(Z_1 \leqslant z_1, Z_2 \leqslant z_2), \\
&= \Pr\left(-1/\log\Phi(X_1) \leqslant z_1, -1/\log\Phi(X_2) \leqslant z_2\right), \\
&= \Pr\left(X_1 \leqslant \Phi^{-1}(e^{-1/z_1}), X_2 \leqslant \Phi^{-1}(e^{-1/z_2})\right), \\
&= \Phi_2\left(\Phi^{-1}(e^{-1/z_1}), \Phi^{-1}(e^{-1/z_2}); \rho(\mathbf{h})\right), \\
&= \Phi(x_1, x_2; \rho(\mathbf{h})),
\end{aligned}$$

where $\Phi_2(x_1, x_2; \rho(\mathbf{h}))$ denotes the bivariate Gaussian distribution function with correlation $\rho_{\mathbf{h}}$, and $(x_1, x_2) = \left(\Phi^{-1}(e^{-1/z_1}), \Phi^{-1}(e^{-1/z_2})\right)$. Thus, this model corresponds to a bivariate normal copula with Fréchet margins. Note that, for $i = 1, 2$,

$$\frac{dx_i}{dz_i} = \frac{1}{z_i^2 e^{1/z_i} \varphi\left[\Phi^{-1}(e^{1/z_i})\right]}.$$

Then, for $z_1 > u$ and $z_2 > u$,

$$\frac{\partial^2 F_{\mathbf{h}}(z_1, z_2)}{\partial z_1 \partial z_2} = \frac{\partial^2 \Phi(x_1, x_2; \rho(\mathbf{h}))}{\partial z_1 \partial z_2} = \varphi(x_1, x_2; \rho(\mathbf{h})) \frac{\partial x_1}{\partial z_1} \frac{\partial x_2}{\partial z_2},$$

where $\varphi(x_1, x_2; \rho(\mathbf{h}))$ is the standard form bivariate Gaussian density function that can be written as

$$\varphi(x_1, x_2; \rho(\mathbf{h})) = \frac{1}{2\pi\sqrt{1 - \rho^2(\mathbf{h})}} \exp\left(-\frac{x_1^2 - 2\rho^2(\mathbf{h}) x_1 x_2 + x_2^2}{2(1 - \rho^2(\mathbf{h}))}\right).$$

The first partial derivative then is

$$\begin{aligned}
\frac{\partial F_{\mathbf{h}}(z_1, z_2)}{\partial z_1} = \frac{\partial \Phi(x_1, x_2; \rho(\mathbf{h}))}{\partial z_1} &= \frac{\partial x_1}{\partial z_1} \int_{-\infty}^{x_2} \varphi(x_1, t) dt, \\
&= \frac{\partial x_1}{\partial z_1} \int_{-\infty}^{x_2} \varphi(z_1) \varphi(t \mid z_1) dt, \\
&= \frac{\partial x_1}{\partial z_1} \varphi(x_1) \Phi\left(\frac{x_2 - \rho(\mathbf{h}) x_1}{\sqrt{1 - \rho^2(\mathbf{h})}}\right), \qquad (4.3.2)
\end{aligned}$$

where the last line follows since $X_2 \mid X_1 \sim \mathcal{N}(\rho(\mathbf{h}) z_1, 1 - \rho^2(\mathbf{h}))$. By symmetry, the second partial derivative is

$$\frac{\partial F_{\mathbf{h}}(z_1, z_2)}{\partial z_2} = \frac{\partial \Phi(x_1, x_2; \rho(\mathbf{h}))}{\partial z_2} = \frac{\partial x_2}{\partial z_2} \varphi(x_2) \Phi\left(\frac{x_1 - \rho(\mathbf{h}) x_2}{\sqrt{1 - \rho^2(\mathbf{h})}}\right).$$

The sub-asymptotic extremal coefficient (equivalent to the sub-asymptotic extremal index

introduced in Ledford and Tawn (2003)) can be derived as

$$\theta_z(\mathbf{h}) = \frac{\log F_{\mathbf{h}}(z, z; \rho(\mathbf{h}))}{\log F(z)}.$$

This is equivalent to

$$\begin{aligned}
\theta_z(\mathbf{h}) &= \frac{\log \Phi(x, x; \rho(\mathbf{h}))}{\log \Phi(x)}, \text{ where } x = \Phi^{-1}(e^{-1/z}) \\
&\approx \frac{1 - \Phi(x, x; \rho(\mathbf{h}))}{1 - \Phi(x)}, \\
&= \frac{2\bar{\Phi}(x) - \bar{\Phi}(x, x; \rho(\mathbf{h}))}{\bar{\Phi}(x)}, \\
&= 2 - \frac{\bar{\Phi}(x, x; \rho(\mathbf{h}))}{\bar{\Phi}(x)}, \text{ for large } x \text{ or } z,
\end{aligned} \tag{4.3.3}$$

where $\bar{\Phi}(\cdot)$ denotes the survival function, and the second line follows from Taylor series expansions for both the numerator and the denominator. It follows from Ledford and Tawn (1996) that $\theta_z(\mathbf{h}) \to 2$ for any $\mathbf{h} \neq 0$ as $z \to \infty$, as expected for an AI model. For very large values of x, both $\bar{\Phi}(x, x; \rho(\mathbf{h}))$ and $\bar{\Phi}(x)$ tend to zero very rapidly, hence numerically calculating $\theta_x(\mathbf{h})$ runs into difficulties. In Appendix 4.A we present a numerical trick to get around this problem.

4.3.4 Inverted Smith process

The bivariate distribution function of the inverted Smith max-stable process is

$$F_{\mathbf{h}}(z_1, z_2) = e^{-1/z_1} + e^{-1/z_2} - 1 + (1 - e^{-1/z_1})^{\Phi(w(\mathbf{h}))} \times (1 - e^{-1/z_2})^{\Phi(v(\mathbf{h}))}, \tag{4.3.4}$$

where Φ, $w(\mathbf{h})$ and $v(\mathbf{h})$ are defined as in Section 4.3.1.

The first derivatives are

$$\frac{\partial F_{\mathbf{h}}(z_1,z_2)}{\partial z_1} = z_1^{-2}e^{-1/z_1} - \frac{z_1^2 e^{-1/z_1}}{1-e^{-1/z_1}}\left[(1-e^{-1/z_1})^{\Phi(w(\mathbf{h}))} + (1-e^{-1/z_2})^{\Phi(v(\mathbf{h}))}\right] \times \left[\Phi(w(\mathbf{h})) + \frac{\phi(w(\mathbf{h}))}{a(\mathbf{h})} - \frac{\log(1-e^{-1/z_2})}{\log(1-e^{-1/z_1})}\frac{\phi(v(\mathbf{h}))}{a(\mathbf{h})}\right],$$

and

$$\frac{\partial F_{\mathbf{h}}(z_1,z_2)}{\partial z_2} = z_2^{-2}e^{-1/z_2} - \frac{z_2^2 e^{-1/z_2}}{1-e^{-1/z_2}}\left[(1-e^{-1/z_1})^{\Phi(w(\mathbf{h}))} + (1-e^{-1/z_2})^{\Phi(v(\mathbf{h}))}\right] \times \left[\Phi(v(\mathbf{h})) + \frac{\phi(v(\mathbf{h}))}{a(\mathbf{h})} - \frac{\log(1-e^{-1/z_1})}{\log(1-e^{-1/z_2})}\frac{\phi(w(\mathbf{h}))}{a(\mathbf{h})}\right],$$

where φ and $a(\mathbf{h})$ are defined as in Section 4.3.1. Then the bivariate density can be derived as

$$\begin{aligned} f_{\mathbf{h}}(z_1,z_2) = {} & \frac{z_1^2 e^{-1/z_1}}{1-e^{-1/z_1}}\frac{z_2^2 e^{-1/z_2}}{1-e^{-1/z_2}}\left[(1-e^{-1/z_1})^{\Phi(w(\mathbf{h}))} + (1-e^{-1/z_2})^{\Phi(v(\mathbf{h}))}\right] \\ & \times \left\{ a^{-2}(\mathbf{h})\left[\frac{w(\mathbf{h})\phi(w(\mathbf{h}))}{\log(1-e^{-1/z_2})} + \frac{v(\mathbf{h})\phi(v(\mathbf{h}))}{\log(1-e^{-1/z_1})}\right] \right. \\ & - a^{-1}(\mathbf{h})\left[\frac{\phi(w(\mathbf{h}))}{\log(1-e^{-1/z_2})} + \frac{\phi(v(\mathbf{h}))}{\log(1-e^{-1/z_1})}\right] \\ & + \left[\Phi(w(\mathbf{h})) + \frac{\phi(w(\mathbf{h}))}{a(\mathbf{h})} - \frac{\log(1-e^{-1/z_2})}{\log(1-e^{-1/z_1})}\frac{\phi(v(\mathbf{h}))}{a(\mathbf{h})}\right] \\ & \left. \times \left[\Phi(v(\mathbf{h})) + \frac{\phi(v(\mathbf{h}))}{a(\mathbf{h})} - \frac{\log(1-e^{-1/z_1})}{\log(1-e^{-1/z_2})}\frac{\phi(w(\mathbf{h}))}{a(\mathbf{h})}\right]\right\}. \end{aligned}$$

The sub-asymptotic extremal coefficient is

$$\theta_z(\mathbf{h}) = \frac{\log F_{\mathbf{h}}(z,z)}{\log F(z)},$$

where $F_{\mathbf{h}}(z,z)$ is the inverted Smith bivariate distribution function as given in (4.3.4) and

$F(z)$ is the unit Fréchet distribution function. Hence,

$$\theta_z(\mathbf{h}) = \frac{\log\left(2e^{-1/z} - 1 + (1 - e^{-1/z})^{2\Phi\{a(\mathbf{h})/2\}}\right)}{-1/z},$$
$$\approx 2 - \left(\frac{1}{z}\right)^{2\Phi\{a(\mathbf{h})/2\}-1} \quad \text{for large } z.$$

Similarly to the Gaussian process case, here too $\theta_z(\mathbf{h}) \to 2$ for any $\mathbf{h} \neq 0$ as $z \to \infty$. This is as expected, since the inverted Smith max-stable process is AI.

4.4 Application to North Sea data

In this section we use the inference methods introduced in Section 4.2 and the results derived in Section 4.3 to fit two AD and two AI models to the whitened North Sea storm-peak significant wave height data set described in Section 4.1. We first briefly investigate how the censoring threshold choice affects the results. Then, we explore how the orientation of the strip affects the model parameters, and hence the extremal dependence. Finally, we briefly look at whether there is also a difference in extremal dependence between strips with the same orientation.

4.4.1 Threshold effects

First, we fit the Smith max-stable process along each of the six strips shown on Figure 4.1.1 with different censoring thresholds u. As all locations lie on a strip for each fit, $\mathbf{h}$ reduces to a distance h, and $\sigma_2^2 = \sigma_{12} = 0$ in Σ, so only one parameter, σ_1^2, needs to be estimated. Denote this parameter σ_ϕ^2 when the strip is in direction ϕ, measured clockwise relative to East. Figure 4.4.1 shows the estimated Smith parameter σ_ϕ^2 plotted against the censoring threshold level. First, note that the Smith parameter estimates decrease as the threshold is increased for all strips. However, the ordering of the estimates for the various strips is broadly maintained for any threshold, suggesting that any directional effect we might

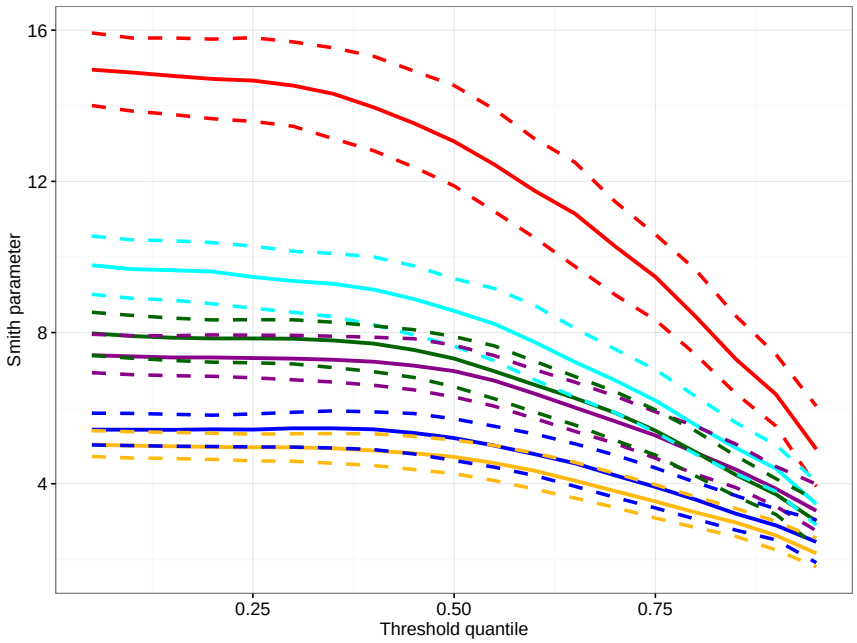


Figure 4.4.1: Estimated Smith parameter σ^2_ϕ plotted against the censoring threshold as quantile of the data. The colours correspond to the strips shown on Figure 4.1.1 and determine ϕ. Confidence intervals were obtained as the 2.5% and 97.5% quantiles from fits to 100 bootstrap samples.

see would not be affected by threshold choice. Since the models fitted are valid only for sufficiently high thresholds, for the rest of this chapter we will use the 90% quantile of the data as the threshold.

4.4.2 Directional effects

Next, we fit all four models in two ways. First, along each strip, in which case the parameter set is reduced to just one parameter σ^2_ϕ for each of the models. This allows us to see whether there is any difference between parameter estimates along strips with different orientations. Secondly, we fit each model to the full data set obtaining estimates of σ_1, σ_2 and σ_{12} for each model. Then we can use the following transformation to estimate the parameter in a given direction ϕ:

$$\widehat{\sigma}^2_\phi = \begin{bmatrix} \cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} \widehat{\sigma}^2_1 & \widehat{\sigma}_{12} \\ \widehat{\sigma}_{12} & \widehat{\sigma}^2_2 \end{bmatrix} \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}. \tag{4.4.1}$$

This allows us to compare the results along the strips with the results from fitting the models to the whole data set. Throughout, the threshold is set to the 90% quantile. Confidence intervals are obtained from 100 bootstrap samples of the data. Result are shown on Figure 4.4.2.

There is a clear pattern present across all four models, with parameter estimates for the red strip highest everywhere and those for the yellow strip the lowest. This means that the extremal dependence is strongest along the red strips and weakest along the yellow strips. The plots suggests that there is a smooth relationship between the parameters (and hence extremal dependence) with the orientation of the strips. The estimates obtained from fitting the models to the whole region (solid black line) also show a similar pattern. There is a reasonably good agreement with the estimates for the individual strips, although the fitted parameters for the whole region seem to suggest stronger dependence than the fitted parameters for the individual strips for the Smith, Schlather max-stable processes and the Gaussian process, and slightly lower dependence for the inverted Smith max-stable process.

The estimated parameters are on different scales for the different models and there is no intuitive way of interpreting the values. Hence, we use the estimated parameters to calculate the extremal coefficient $\theta(\mathbf{h})$ (for AD models) or the sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ (for AI models) for various $\mathbf{h}$ and z levels. We show results for three different levels of $\mathbf{h}$; $\mathbf{h}_1 = (\mathbf{t}_i - \mathbf{t}_j)^T$ where $\mathbf{t}_i$ and $\mathbf{t}_j$ are neighbouring location, $\mathbf{h}_2 = (\mathbf{t}_i - \mathbf{t}_j)^T$ where $\mathbf{t}_i$ and $\mathbf{t}_j$ are locations in the centre and the edge of the region, respectively, and $\mathbf{h}_3 = (\mathbf{t}_i - \mathbf{t}_j)^T$ where $\mathbf{t}_i$ and $\mathbf{t}_j$ are locations at opposite edges of the region. The extremal coefficient and the sub-asymptotic extremal coefficient range between the values 1 and 2, with a value close to 1 signifying strong dependence, and a value close to 2 implying near independence.

Figure 4.4.3 shows $\theta(\mathbf{h}_1)$, $\theta(\mathbf{h}_2)$ and $\theta(\mathbf{h}_3)$ obtained from the fitted Smith parameters. The values for $\theta(\mathbf{h}_1)$ are close to 1 suggesting strong dependence, whereas $\theta(\mathbf{h}_2)$ and $\theta(\mathbf{h}_3)$ are higher, with $\theta(\mathbf{h}_3)$ close to 2, suggesting near independence. Since $\mathbf{h}_1$ corresponds to locations close by and $\mathbf{h}_3$ to locations relatively far apart, the results confirm that extremal

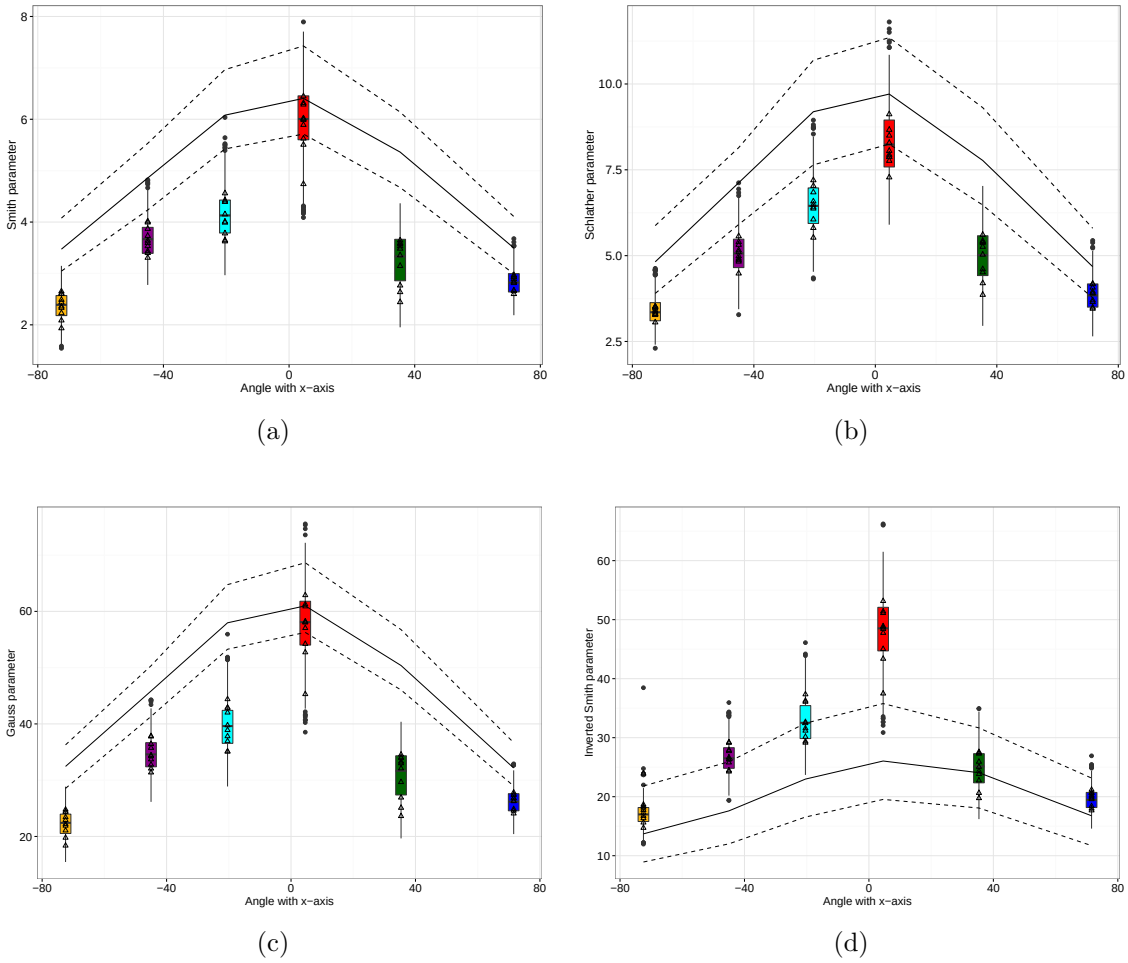


Figure 4.4.2: Triangles show the estimated parameters σ_ϕ^2 along strips with orientations as shown on Figure 4.1.1 for the (a) Smith process, (b) Schlather process, (c) Gaussian process, and (d) inverted Smith process. Boxplots show the estimates for a 100 bootstrap samples of each strip. Solid black line shows the parameter estimate obtained from fitting the model to the whole data set. Dashed black lines show the 2.5% and 97.5% quantiles obtained from 100 bootstrap samples.

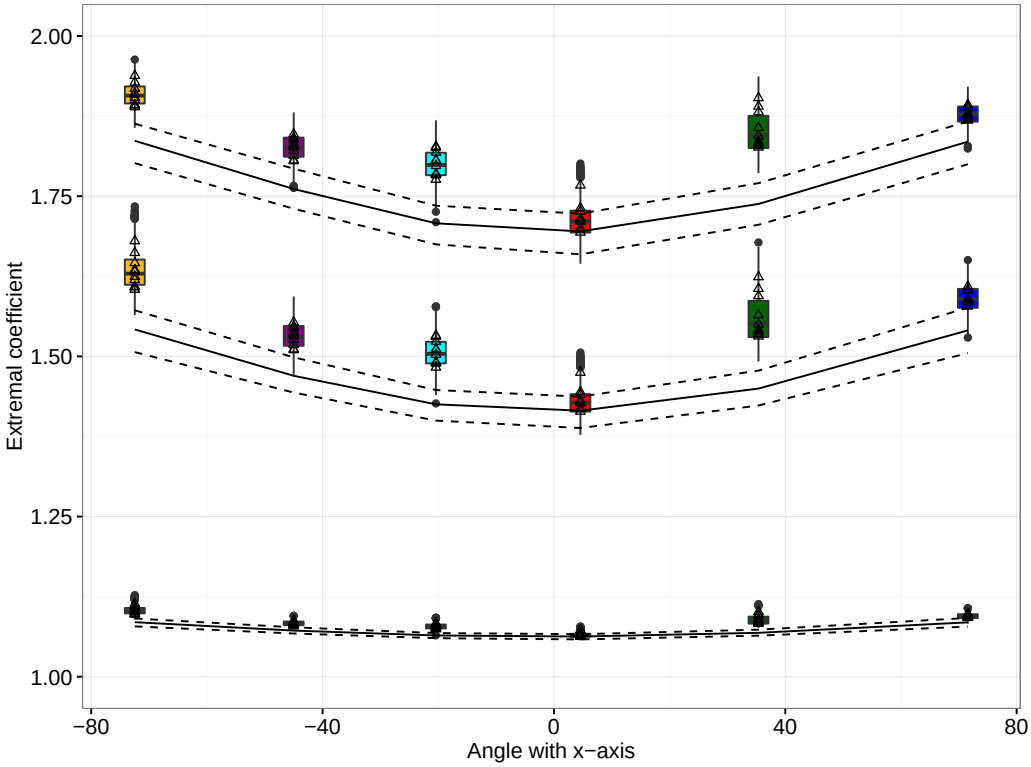


Figure 4.4.3: The extremal coefficient $\theta(\mathbf{h})$ for $\mathbf{h} = \mathbf{h}_1$ (closest to 1), $\mathbf{h} = \mathbf{h}_2$ and $\mathbf{h} = \mathbf{h}_3$ (closest to 2), obtained from the fitted Smith parameters. The interpretation of colours, lines, boxplots and symbols is the same as before.

dependence decreases with distance. Note also that the pattern we observed on Figure 4.4.2 is maintained here, with the red strips showing the strongest dependence for all three levels of $\mathbf{h}$. Equivalent results for the Schlather model are shown on Figure 4.4.4. Note that the Schlather process cannot account for extremes that become independent for large distances as the extremal coefficient is bounded $\theta(\mathbf{h}) < 1.838$ for any $\mathbf{h}$ (Schlather (2002)). This property is clearly shown in Figure 4.4.4, where $\theta(\mathbf{h})$ stays below 1.75 even for $\mathbf{h}_3$.

The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ can be evaluated at various levels z, so we first explore how $\theta_z(\mathbf{h})$ behaves for a range of levels z for $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$ for both the Gaussian and the inverted Smith processes. Since, these processes are AI, it is expected that $\theta_z(\mathbf{h}) \to 2$ as $z \to \infty$, for any $\mathbf{h} \neq 0$. Figures 4.4.5a, 4.4.5c and 4.4.5d show $\theta_z(\mathbf{h})$ calculated from the fitted Gaussian parameters along each strip for a range of z levels (with z on the log scale), for $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$, respectively. For $\mathbf{h}_2$ and $\mathbf{h}_3$, $\theta_z(\mathbf{h})$ reaches 2 for

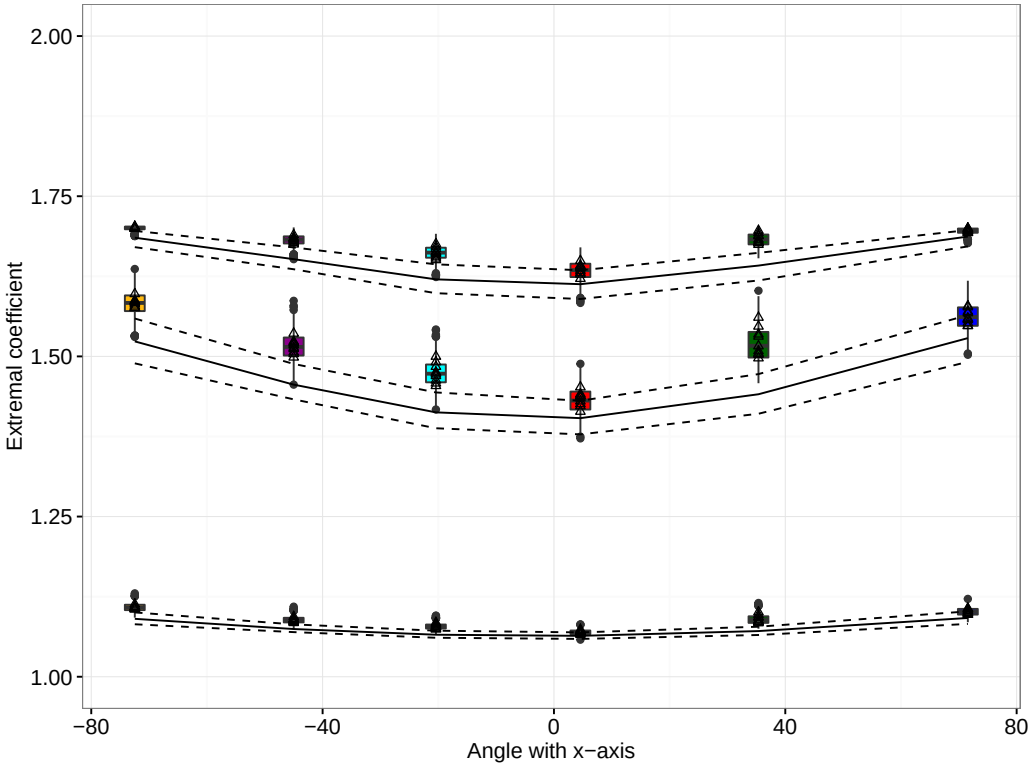
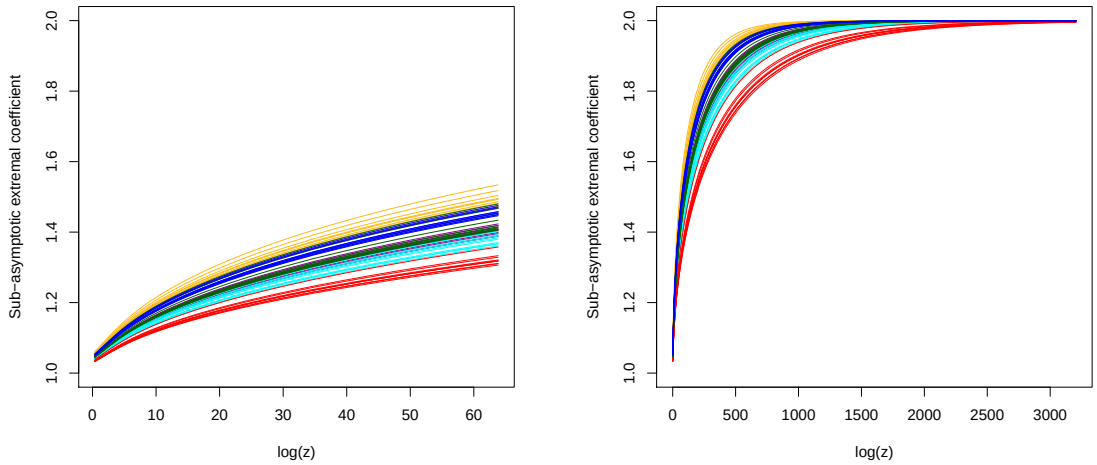


Figure 4.4.4: The extremal coefficient $\theta(\mathbf{h})$ for $\mathbf{h} = \mathbf{h}_1$ (closest to 1), $\mathbf{h} = \mathbf{h}_2$ and $\mathbf{h} = \mathbf{h}_3$ (closest to 2), obtained from the fitted Schlather parameters. The interpretation of colours, lines, boxplots and symbols is the same as before.

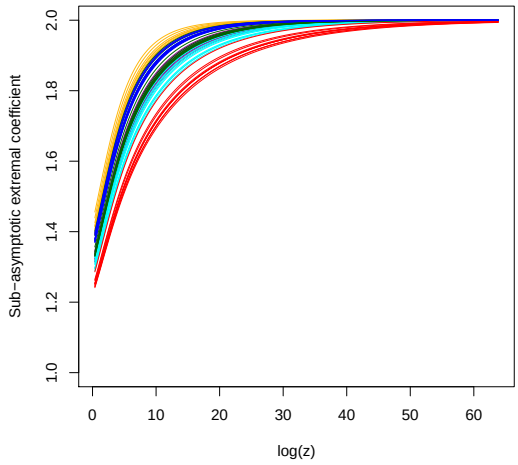
$\log(z) < 60$. However, for $\mathbf{h}_1$, $\theta_z(\mathbf{h})$ only reaches 2 for $\log(z) \approx 3000$ (see Figure 4.4.5b), corresponding to an exceedance probability of approximately e^{-3000}. Figure 4.4.6 shows equivalent results for the inverted Smith process. Similarly to the Gaussian process, $\theta_z(\mathbf{h})$ reaches 2 for relatively low z values ($\log(z) < 50$) for $\mathbf{h}_2$, $\mathbf{h}_3$, but for $\mathbf{h}_1$, $\theta_z(\mathbf{h})$ reaches 2 for all strips only around $\log(z) \approx 300$, which corresponds to an exceedance probability of approximately e^{-300}. This suggests that the inverted Smith process has weaker extremal dependence and reaches independence at lower levels z than the Gaussian process.

This is also visible on Figures 4.4.7 and 4.4.8 where we show the sub-asymptotic extremal coefficient for two values of $\log(z)$, (a) $\log(z) = 2$ and (b) $\log(z) = 15$, obtained from the fitted Gaussian process copula parameters on Figure 4.4.2c and the fitted inverted Smith process parameters on Figure 4.4.2d, respectively. For the same level $\log(z)$, $\theta_z(\mathbf{h})$ is closer to 2 (independence) for the inverted Smith max-stable process than for the Gaussian process

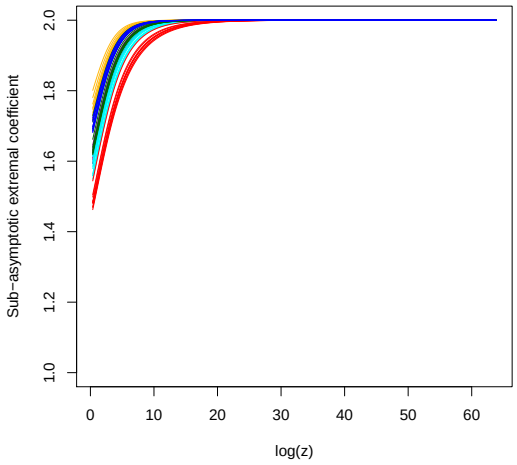


(a)

(b)

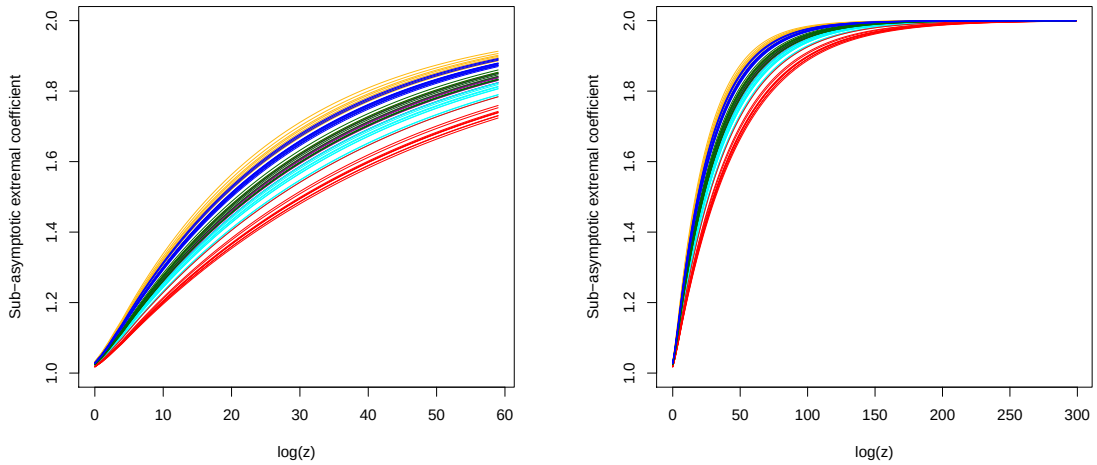


(c)



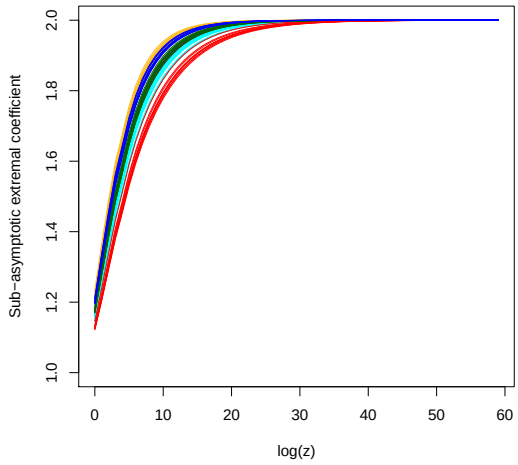
(d)

Figure 4.4.5: The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ for a range of levels z, for (a) $\mathbf{h} = \mathbf{h}_1$, (c) $\mathbf{h} = \mathbf{h}_2$ and (d) $\mathbf{h} = \mathbf{h}_3$ for $\log(z) < 60$, and (b) $\mathbf{h} = \mathbf{h}_1$ for $\log(z)$ extended to > 60, obtained from the fitted Gaussian process copula parameters for locations along the strips shown on Figure 4.1.1.

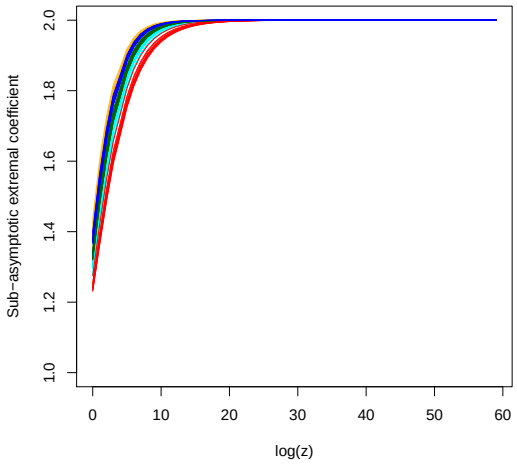


(a)

(b)

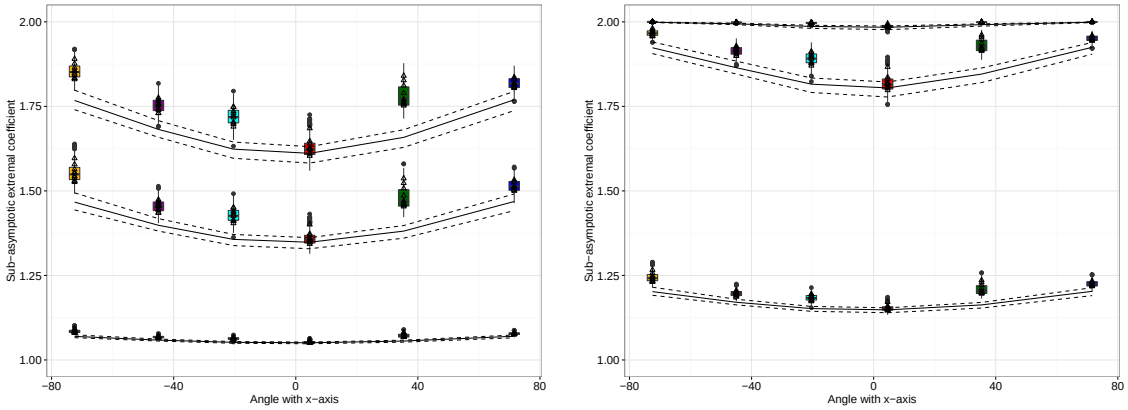


(c)



(d)

Figure 4.4.6: The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ for a range of levels z, for (a) $\mathbf{h} = \mathbf{h}_1$, (c) $\mathbf{h} = \mathbf{h}_2$ and (d) $\mathbf{h} = \mathbf{h}_3$ for $\log(z) < 60$, and (b) $\mathbf{h} = \mathbf{h}_1$ for $\log(z)$ extended to > 60, obtained from the fitted inverted Smith process parameters for locations along the strips shown on Figure 4.1.1.



(a) (b)

Figure 4.4.7: The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ for $\mathbf{h} = \mathbf{h}_1$ (closest to 1), $\mathbf{h} = \mathbf{h}_2$ and $\mathbf{h} = \mathbf{h}_3$ (closest to 2), obtained from the fitted Gaussian process copula parameters, for (a) $\log(z) = 2$, and (b) $\log(z) = 15$. The interpretation of colours, lines, boxplots and symbols is the same as before.

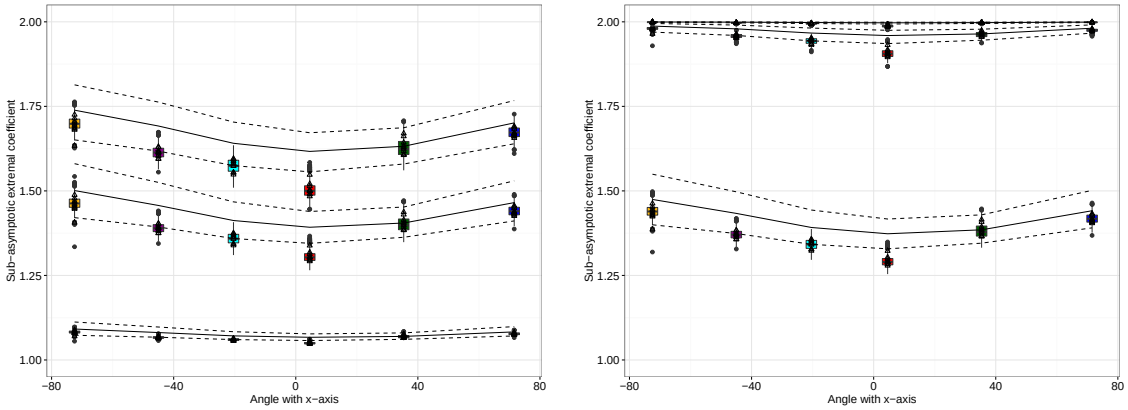
for all distances $\mathbf{h}$. The same patterns as noted in the case of the Smith and Schlather models is present here also, with extremal dependence being weaker the larger $\mathbf{h}$ is.

The sub-asymptotic extremal coefficient can also be estimated empirically from the data itself using

$$\theta_z(\mathbf{h}) = \frac{\log \Pr(Z_1 \leqslant z, Z_2 \leqslant z)}{\log \Pr(Z_1 \leqslant z)}, \tag{4.4.2}$$

where Z_1 and Z_2 are observations at two locations $\mathbf{t}_1$ and $\mathbf{t}_2$, and $\mathbf{h} = (\mathbf{t}_1 - \mathbf{t}_2)^T$. Both probabilities in expression (4.4.2) can be estimated from the data using empirical counts. Figures 4.4.9a-4.4.9c show the empirical estimates for the strips shown on Figure 4.1.1 calculated for the distances $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$ for a range of levels z. First, note that, similarly to Figures 4.4.5 and 4.4.6, the estimates for the different strips are ordered in roughly the same way with the red strip showing the strongest extremal dependence ($\theta_z(\mathbf{h})$ closest to 1) and the yellow strip the weakest. All the models we fitted capture this feature in the data well.

Also note on Figures 4.4.9a-4.4.9c that the empirical estimates of $\theta_z(\mathbf{h})$ increase with distance $\mathbf{h}$ and level z. Figure 4.4.9d, where the estimated $\theta_z(\mathbf{h})$ is shown for a range of



(a) (b)

Figure 4.4.8: The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ for $\mathbf{h} = \mathbf{h}_1$ (closest to 1), $\mathbf{h} = \mathbf{h}_2$ and $\mathbf{h} = \mathbf{h}_3$ (closest to 2), obtained from the fitted inverted Smith process parameters, for (a) $\log(z) = 2$, and (b) $\log(z) = 15$. The interpretation of colours, lines, boxplots and symbols is the same as before.

distances for three levels of z, also clearly shows these features. For all the fitted models we observed a similar increase with distance $\mathbf{h}$, signifying that dependence decreases with distance in both the data and in the fitted models. However, for the max-stable Smith and Schlather processes the extremal coefficient does not increase with z. This is only the case for the AI processes fitted (Gaussian and inverted max-stable Smith), and hence the fact that the empirical estimates do increase with z suggest that the data is more consistent with AI models.

Comparing the empirical estimates with the model estimates on Figures 4.4.5 and 4.4.6 is difficult, as for large values of z there is high uncertainty in the empirical estimates. For the relatively low level of $\log(z) = 5$ the empirical estimates are $\theta_z(\mathbf{h}_1) \approx 1.1$, $\theta_z(\mathbf{h}_2) \approx 1.6$ and $\theta_z(\mathbf{h}_3) \approx 1.8$ (see Figures 4.4.9a-4.4.9c). For $\mathbf{h}_1$ and $\mathbf{h}_2$ both the Gaussian and the inverted Smith max-stable processes reach similar values at $\log(z) = 5$. However, for the Gaussian process $\theta_z(\mathbf{h}_3) \approx 1.9$, whereas for the inverted Smith process $\theta_z(\mathbf{h}_3) \approx 1.8$, suggesting that at this distance the Gaussian process is less dependent than the data. Hence, the inverted Smith model might be more suitable for this data set. Interestingly, the Gaussian process seems to exhibit weaker dependence at long distances ($\mathbf{h}_3$), but stronger dependence for

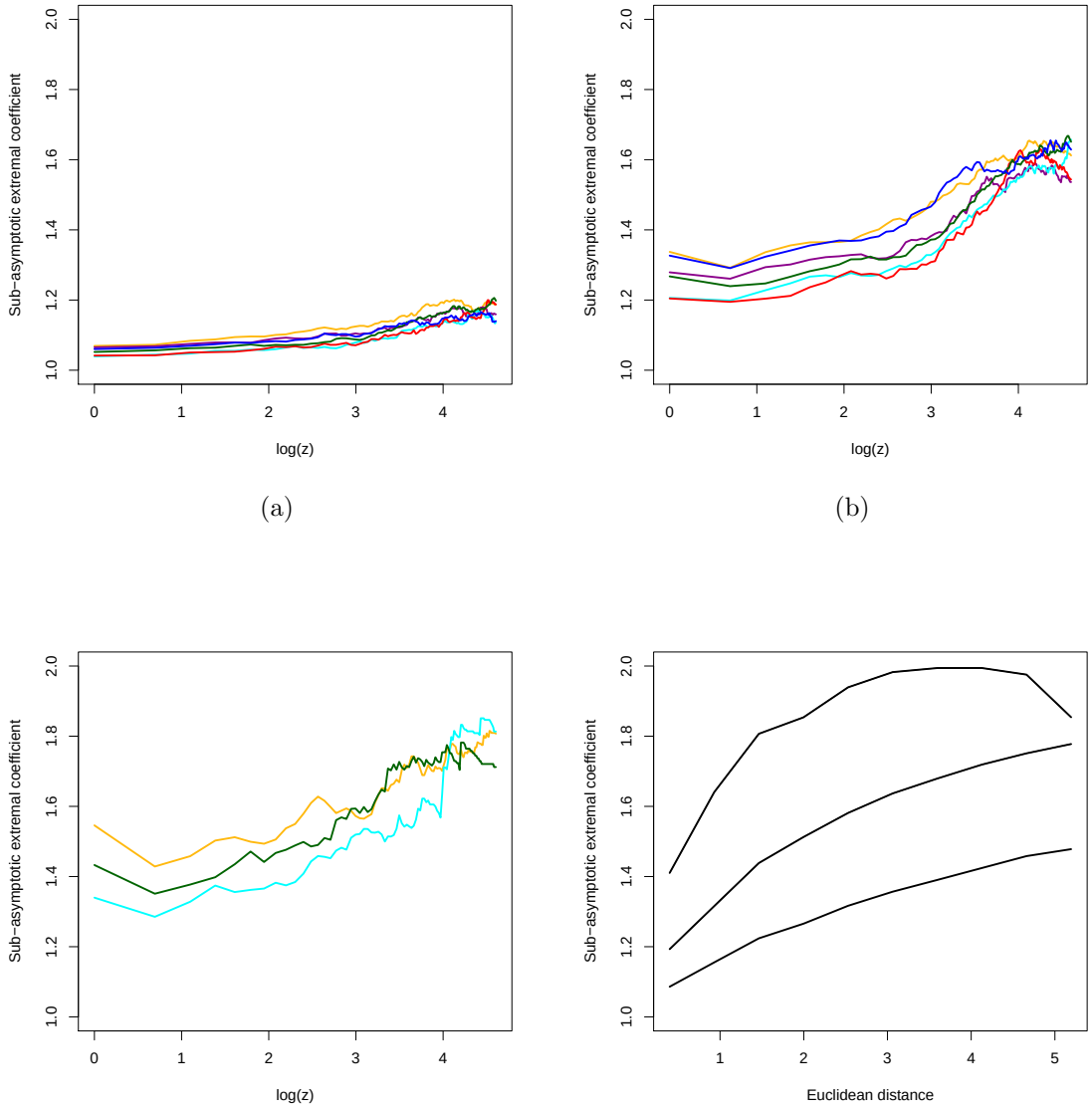
short distances ($\mathbf{h}_1$) than the inverted Smith process, highlighting the differing extremal dependence structures of the two models.

4.4.3 Location effects

We are also interested in whether there is a difference between estimates for different strips with the same orientation. This might be the case if, for example, a strip is protected from large storms by land shadow. To see whether there is such an effect we will look at ten parallel strips for each orientation. We refer to the strips shown on Figure 4.1.1 as central strips, and calculate the distance between the central strip and each of the parallel strips with the same orientation. More northerly strips than the central strip are assigned a positive distance, whereas more southerly strips than the central strip are assigned a negative distance. Figure 4.4.10 shows the extremal coefficient $\theta(\mathbf{h}_1)$ obtained from the fitted Smith process along each strip against the distance of each strip from the central strip in each orientation. It is clear that there are some differences between different estimates for strips with the same orientation. For example, the green and yellow strips seem to show stronger extremal dependence in the middle of the region than at the edges, but further investigation is needed to ascertain these effects. Results for other $\mathbf{h}$ values and other models are similar so not shown here.

4.4.4 Simulations

Figure 4.4.11 shows four realisations of the whitened North Sea storm-peak significant wave height dataset, plotted on Gumbel scale. To see how the fitted models compare to the data, we can simulate from each model using the fitted values of σ_1^2, σ_2^2 and σ_{12}. Figure 4.4.12 shows four simulated samples from each of the fitted models. Compared to the data, simulated samples from the Smith and Gaussian processes seem to be too smooth, whereas the Schlather process appears to be too rough. Simulated samples from the inverted Smith pro-



(a) (b)

(c) (d)

Figure 4.4.9: The sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ estimated empirically from the data with (a) $\mathbf{h} = \mathbf{h}_1$, (b) $\mathbf{h} = \mathbf{h}_2$, (c) $\mathbf{h} = \mathbf{h}_3$, plotted against z on the log-scale; and (d) the sub-asymptotic extremal coefficient $\theta_z(\mathbf{h})$ estimated empirically from the data for a range of $\mathbf{h}$, with $\mathbf{h}_1 \leqslant \mathbf{h} \leqslant \mathbf{h}_3$, and three different levels of z: the 90% quantile (line closest to 1), the 99% quantile, and the 99.9% quantile (line closest to 2). On subfigure (c) results for the red, blue and magenta strips are not shown as these strip were too short to have any pairs of locations distance $\mathbf{h}_3$ apart.

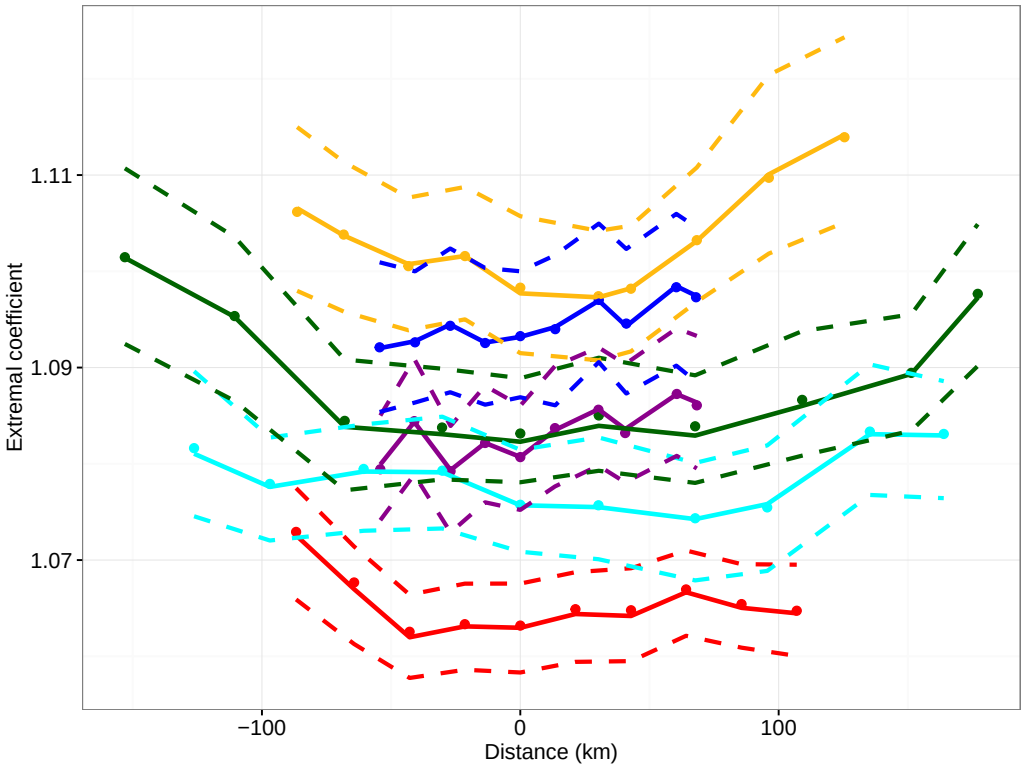


Figure 4.4.10: The extremal coefficient $\theta(\mathbf{h}_1)$, obtained from the fitted Smith process along each strip, plotted against the distance of each strip from the central strip in each orientation, with colours corresponding to the strips on on Figure 4.1.1. Positive distance is assigned if the strip crosses the y-axis higher than the central strip with the same orientation and negative distance if lower. Points represent estimates for each strip, solid lines are the median estimate from a 100 bootstrap samples, dashed lines 2.5% and 97.5% quantiles from a 100 bootstrap samples.

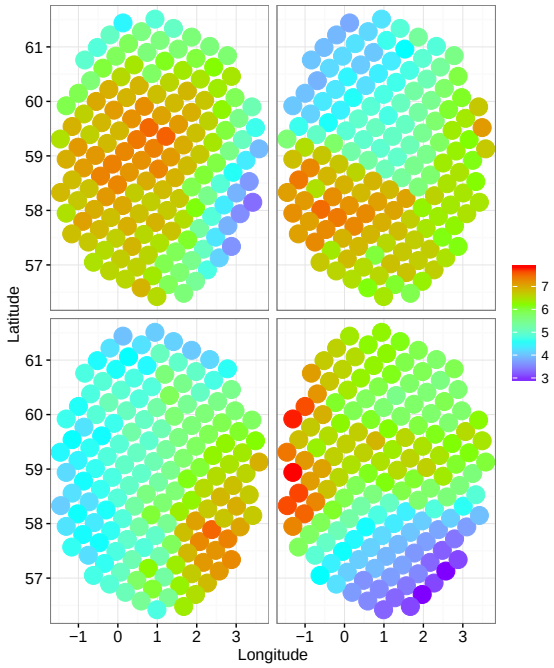


Figure 4.4.11: Four realisations of the whitened North Sea storm-peak significant wave height dataset, plotted on Gumbel scale.

cess appear somewhat smoother than the data samples, but arguably this model captures the patterns in the data the best out of these four models.

It is hard to draw definitive conclusions about directional effects from just four samples, but the north to south effect captured on Figure 4.4.11 (top and bottom right) is similar to some of the Smith or inverted Smith max-stable process realisations (Figures 4.4.12a and 4.4.12d).

Other examples in the extremes literature of higher order measures of fit for spatial models include composite likelihood information criterion (CLIC) used, for example, by Blanchet and Davison (2011) and Davison et al. (2012).

4.5 Conclusion

In this chapter we presented a study of spatial extremes that systematically tests directional features of models. As far as we are aware, this is a novel contribution to the extremes

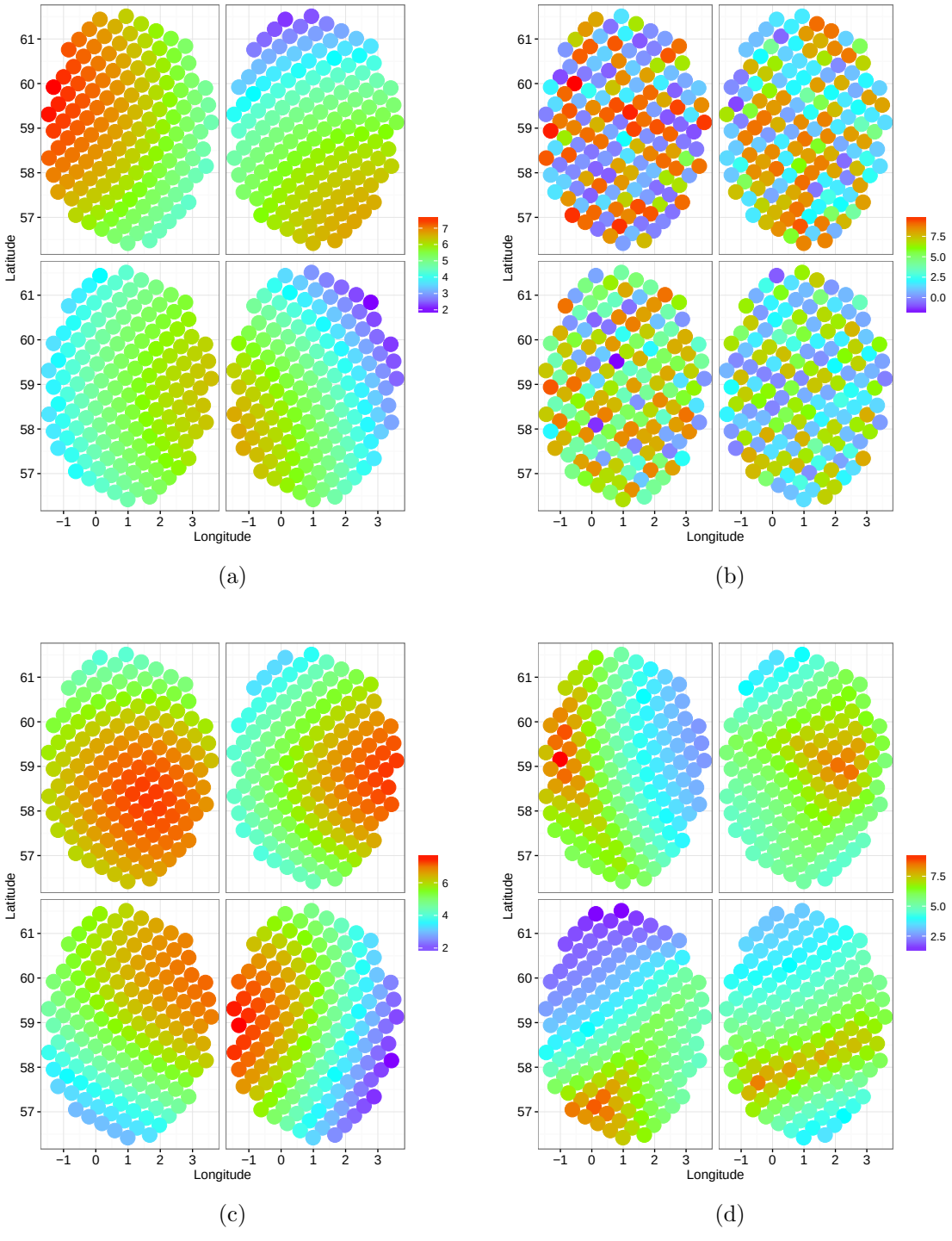


Figure 4.4.12: Four simulated samples (transformed to Gumbel margins) from each of the fitted models; (a) Smith process, (b) an isotropic version of the Schlather process, (c) Gaussian process and (d) inverted Smith process.

literature with no other equivalent studies.

The parameter estimates from the four models fitted to the North Sea storm-peak significant wave height data set suggest that extremal dependence is strongest on strips with an approximately west to east orientation. Since, the largest storms typically travel in a north to south direction in the North Sea, this suggests that the extremal dependence is stronger across the storm front than in the storm direction. We also find that extremal dependence varies smoothly with the direction of the strips, suggesting that direction needs to be considered when modelling extremal dependence. There is some suggestion in the data of a location effect, but more investigation is needed to explore these effects.

4.A Sub-asymptotic extremal coefficient for the Gaussian process

Evaluating $\theta_z(\mathbf{h})$, the expression (4.3.3) leads to numerical problems for large values of $x = \Phi^{-1}(e^{-1/z})$. To avoid this, we can use the following approximation. First, by Mill's ratio

$$\bar{\Phi}(x) \sim \frac{\varphi(x)}{x} \text{ as } x \to \infty.$$

Then, we can write the ratio in expression (4.3.3), as $x \to \infty$, as

$$\begin{aligned}
\frac{\bar{\bar{\Phi}}(x, x; \rho(\mathbf{h}))}{\bar{\Phi}(x)} &\sim \frac{\int_x^\infty \int_x^\infty \varphi(s,t) ds dt}{\varphi(x)/x}, \\
&= \frac{\int_x^\infty \int_x^\infty \varphi(t \mid s) dt \varphi(s) ds}{\varphi(x)/x}, \\
&= x \int_x^\infty \bar{\Phi}\left(\frac{x - \rho(\mathbf{h}) s}{\sqrt{1 - \rho^2(\mathbf{h})}} \right) \frac{\varphi(s)}{\varphi(x)} ds,
\end{aligned}$$

where the last line follows after a similar argument to that in (4.3.4). Using a change of variables, $s = x + y$, then gives

$$\begin{aligned}\frac{\bar{\Phi}(x, x; \rho(\mathbf{h}))}{\bar{\Phi}(x)} &= x \int_0^\infty \bar{\Phi}\left(\frac{x - \rho(\mathbf{h})(x+y)}{\sqrt{1-\rho^2(\mathbf{h})}}\right) \frac{\varphi(x+y)}{\varphi(x)} dy, \\ &= x \int_0^\infty \bar{\Phi}\left(\frac{x - \rho(\mathbf{h})(x+y)}{\sqrt{1-\rho^2(\mathbf{h})}}\right) \exp(-xy - y^2/2) dy.\end{aligned}$$

Numerical integration can be used then to evaluate this integral. This leads to results with greater numerical stability than obtained using expression (4.3.3).

Chapter 5

Properties of Extremal Dependence Models Built on Bivariate Max–Linearity

5.1 Introduction

When modelling extremes of spatial environmental processes we often care about both local dependence and long-range dependence. For example, in an oceanographic application, we would be interested in the relationship between extreme significant wave heights at two locations that might be close by or located far apart. In particular, we want to know how likely it is that both locations are affected by the same storm and have high waves simultaneously (see e.g. Jonathan et al. (2013)). Since interest lies in the extremes, the standard measures of spatial dependence are not appropriate and alternative dependence measures and models should be used. Here we introduce a family of bivariate distributions, with simple multivariate extensions, that exhibits all the required features of short, medium and long range extremal dependence for spatial applications. This family is shown to capture all possible

bivariate distributions with these properties. We propose novel bivariate characterisations of the extremal dependence structure that reveal structure of this family of distributions that standard measures of extremal dependence fail to identify.

First we identify the two core extremal dependence measures. Let X and Y be identically distributed random variables. Then, an intuitive measure of extremal dependence is the tail dependence measure χ, which is defined as the limiting probability that Y is extreme given that X is extreme,

$$\chi = \lim_{z \to z_F} \Pr(Y > z \mid X > z), \tag{5.1.1}$$

where z_F is the upper end point of the common marginal distribution. When $\chi > 0$, X and Y are said to be asymptotically dependent (AD) and the value of χ signifies the strength of asymptotic dependence. This means that X and Y can be extreme simultaneously. However, when the variables are asymptotically independent (AI), $\chi = 0$ and hence χ does not contain any information about the sub-asymptotic dependence structure. Coles et al. (1999) argue that to give a more complete summary of extremal dependence a second measure is needed to describe the rate of convergence of $\Pr(Y > z \mid X > z)$ to 0. A useful tail dependence measure can be obtained from the Ledford and Tawn (1996) joint tail dependence model, which states that

$$\Pr(X > z, Y > z) = \mathcal{L}(1/\Pr(X > z))\{\Pr(X > z)\}^{2/(\bar{\chi}+1)}, \tag{5.1.2}$$

where $\mathcal{L}$ is a slowly varying function at infinity and $\bar{\chi} \in (-1, 1]$. The exponent $2/(\bar{\chi} + 1)$ determines the decay rate of the joint probability, with smaller $\bar{\chi}$ giving more rapid convergence of χ to 0. The pair ($\chi > 0$; $\bar{\chi} = 1$) signifies AD, for which the value of χ gives a measure of strength of dependence; and ($\chi = 0$; $\bar{\chi} < 1$) signifies AI, for which the value of $\bar{\chi}$ gives the strength of dependence.

Both the dependence measures χ and $\bar{\chi}$, in expressions (5.1.1) and (5.1.2), are invariant to the marginal distribution. Of course, using the concept of copulas, all dependence measures

can be expressed independently of the marginal distributions. However, for some choices of marginal distributions extremal dependence structure properties are more simply expressed than for other marginal choices. For example, much of the traditional multivariate extreme value theory results are expressed for Fréchet marginals, as they lead to the cleanest expressions of results for componentwise maxima and multivariate regular variation (Resnick (1987)). This marginal choice is fine when the variables are AD, however for AI variables this selection leads to an identical limit form whatever the nature of the AI, i.e., whatever $\bar{\chi} < 1$. For AI variables, Heffernan and Tawn (2004), Keef et al. (2013) and Wadsworth and Tawn (2013b) all identify that non-degenerate limit distributions, under affine transformations, can be obtained using exponential margins/tails, whereas under their formulations the limits are degenerate for Fréchet margins. Furthermore, in exponential margins results for AD are also non-degenerate. The reason for this extra flexibility in exponential margins is that an affine transformation in that space is a complex non-linear transformation in Fréchet margins (see Section 2.2 of Papastathopoulos and Tawn (2016)). Therefore, we work in exponential margins to illustrate our novel extremal dependence characterisations and show that if Fréchet margins had been used, the structure we find would not have been apparent using affine transformations.

In the analysis of multivariate data, it is often difficult to make a choice between AD and AI (see e.g., Davison et al. 2013, Thibaud et al. 2013, and Kereszturi et al. 2016). By having a model that has both AD and AI components, we can avoid having to make this key decision. Wadsworth and Tawn (2012a) combine a max-stable process with an inverted max-stable process to construct a hybrid spatial dependence model. This model can capture both the AD and AI dependence structure but it is restricted in its forms of AD and AI that can be modelled. Here we use the core structure of the Wadsworth and Tawn (2012a) model as a basis for exploring bivariate extreme value modelling in a new light. Specifically, we develop a distribution that contains both AD and AI components and has the flexibility to capture all dependence forms within very broad classes in each case.

We construct our model using the multivariate max-linear model (Davis and Resnick, 1989) as the building block. This class of distributions is both mathematically elegant and was the starting point for understanding the formulation of multivariate extremes (Pickands, 1981). In the bivariate case with Fréchet marginal variables X_F and Y_F, the max-linear model takes the following form:

$$X_F = \max_{i=1,\ldots,m} \{\alpha_i Z_i\}, \qquad (5.1.3)$$
$$Y_F = \max_{i=1,\ldots,m} \{\beta_i Z_i\},$$

where $0 \leqslant \alpha_i, \beta_i \leqslant 1$ for all i, m can be finite or infinite, $\Sigma_{i=1}^m \alpha_i = 1$, $\Sigma_{i=1}^m \beta_i = 1$, and $Z_i \sim$ i.i.d. Fréchet, $i = 1, \ldots, m$, with distribution function $F_Z(z) = \exp(-1/z)$ for $z > 0$ and density denoted $f_Z(z)$. This model has joint distribution function

$$\Pr(X_F < x, Y_F < y) = \exp\left(-\sum_{i=1}^{m} \max\left(\frac{\alpha_i}{x}, \frac{\beta_i}{y}\right)\right), \text{ for } x > 0, y > 0,$$

and it is straightforward to show that this satisfies max-stability, since for any $n > 0$, $x > 0$ and $y > 0$,

$$\Pr(X_F < nx, Y_F < ny)^n = \Pr(X_F < x, Y_F < y).$$

Fundamental to our approach is that Deheuvels (1983) shows that every multivariate extreme value distribution for minima, with exponential marginals (i.e., with variables (X_F^{-1}, Y_F^{-1})), can be arbitrarily well approximated by a multivariate max-linear model. Fougères et al. (2013) showed this property holds for (X_F, Y_F), as well as presenting a broader discussion on alternative representations of multivariate extreme value distributions.

This chapter introduces two bivariate distributions, with exponential margins, that are derived from the max-linear model (5.1.3) with Fréchet margins: these are the transformed max-linear model and the inverted max-linear model, denoted by (X_E, Y_E) and $(X_E^{(I)}, Y_E^{(I)})$

respectively. Specifically,

$$(X_E, Y_E) = (-\log(1 - \exp\{-1/X_F\}), -\log(1 - \exp\{-1/Y_F\})) \tag{5.1.4}$$

and

$$(X_E^{(I)}, Y_E^{(I)}) = (1/X_F, 1/Y_F). \tag{5.1.5}$$

Here X_E (Y_E) transforms X_F (Y_F) to the exponential margins through a monotone increasing mapping, with repeated use of the probability integral transform, whereas $X_E^{(I)}$ ($Y_E^{(I)}$) transforms X_F (Y_F) to the exponential margins through a monotone decreasing mapping. So marginally both (X_E, Y_E) and $(X_E^{(I)}, Y_E^{(I)})$ are identical, but they differ significantly in their dependence structure and, in particular, their extremal dependence properties. The models (X_F, Y_F) and (X_E, Y_E) have the same copula, while $(X_E^{(I)}, Y_E^{(I)})$ has the same copula as the joint lower tail of (X_F, Y_F). Hence, we refer to (X_E, Y_E) and $(X_E^{(I)}, Y_E^{(I)})$ as having the upper tail and the lower tail copula of (X_F, Y_F), respectively. For both models we explore their joint upper tail, and so focus on different features of the (X_F, Y_F) copula.

The copula of the joint distribution $(X_E^{(I)}, Y_E^{(I)})$ is an example of the class of inverted max-stable models first introduced in Ledford and Tawn (1996, 1997). The inverted max-stable distributions are a broad class of AI distributions, covering all values of $\bar{\chi}$ with $0 \leqslant \bar{\chi} < 1$. Heffernan and Tawn (2004) found interesting conditional extremal behaviour for a sub-family of this class, with much broader structures explored by Papastathopoulos and Tawn (2016). Furthermore, Wadsworth and Tawn (2012a) explored extensions of the representations of Ledford and Tawn (1997) through a series of multivariate regular variation conditions, and found the inverted max-stable distributions to have particular importance in modelling AI. It follows from results in Deheuvels (1983) that inverted max-linear models give an arbitrarily good approximation to inverted multivariate extreme value distributions, and so for a study of AI distributions models of the form $(X_E^{(I)}, Y_E^{(I)})$ are of core importance. Next we derive χ and $\bar{\chi}$ for the transformed max-linear and inverted max-linear models.

The joint distribution function of the transformed max-linear model (X_E, Y_E) is

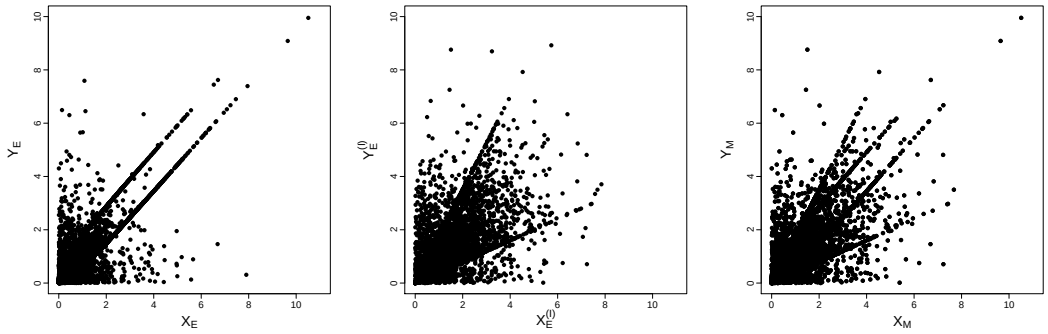
$$\Pr(X_E < x, Y_E < y) = \prod_{i=1}^{m} \min\Big((1-\exp(-x))^{\alpha_i}, (1-\exp(-y))^{\beta_i} \Big), \text{ for } x > 0, y > 0. \tag{5.1.6}$$

Unlike (X_F, Y_F), this is not max-stable, but this is due to the margin choice not the copula, which remains unchanged. The limiting distribution of normalised componentwise maxima of (5.1.6) can be shown to be max-stable, so (X_E, Y_E) is in the domain of attraction of a bivariate extreme value distribution with limiting dependence. For this model, $\bar{\chi} = 1$ and $\chi = 2 - \sum_{i=1}^{m} \max(\alpha_i, \beta_i)$, so the variables are AD. On exponential margins, simulations from the max-linear model in (5.1.3) give lines of mass, parallel with $X_E = Y_E$, and points scattered around these lines, as shown on Figure 5.1.1a, where X_E and Y_E were determined by $X_F = \max(0.7Z_1, 0.2Z_2, 0.1Z_3)$ and $Y_F = \max(0.4Z_1, 0.5Z_2, 0.1Z_4)$. The number of Z_i variables in common between X_F and Y_F determines the number of lines with mass on. In the case of Figure 5.1.1a there are two Z_i variables, Z_1 and Z_2, in common between X_F and Y_F, hence there is mass on two lines. The independent scatter of points around the lines is due to the presence of Z_3 in X_F and Z_4 in Y_F.

The joint distribution function of the inverted max-linear model $(X_E^{(I)}, Y_E^{(I)})$ is

$$\Pr(X_E^{(I)} < x, Y_E^{(I)} < y) = 1 - \exp(-x) - \exp(-y) + \exp\left(-\sum_{i=1}^{m} \max(\alpha_i x, \beta_i y) \right), \tag{5.1.7}$$

for $x > 0$, $y > 0$. For this model, that $\chi = 0$ and $\bar{\chi} = (2/\sum_{i=1}^{m} \max(\alpha_i, \beta_i)) - 1$, so the variables are AI. Figure 5.1.1b shows a random sample from $(X_E^{(I)}, Y_E^{(I)})$ derived from the same max-linear model (X_F, Y_F) as used to illustrate (X_E, Y_E) above. This model gives points on rays and points scattered around these rays. Similarly to (X_E, Y_E), the number of rays is determined by the number of Z_i variables that are common between X_F and Y_F, which in our example is two. Note, that in the inverted max-linear model the point masses are no longer on parallel lines, but on rays ($y = hx$ for $0 < h < \infty$) that meet at the origin.



(a) (b) (c)

Figure 5.1.1: Bivariate simulations derived from the max-linear model in (5.1.3) with $X_F = \max(0.7Z_1, 0.2Z_2, 0.1Z_3)$ and $Y_F = \max(0.4Z_1, 0.5Z_2, 0.1Z_4)$; (a) transformed max-linear model (X_E, Y_E), (b) inverted max-linear model $(X_E^{(I)}, Y_E^{(I)})$, and (c) mixture model (X_M, Y_M) with $\delta = 0.5$ in (5.5.1).

If there exists at least one $i = 1, \ldots, m$ such that $\alpha_i = \beta_i$, then there is a ray with gradient $h = 1$, but despite this the variables are AI.

Combining these two models provides a flexible approach to modelling extremal dependence that can capture both AI and AD. Figure 5.1.1c shows an example of a model (X_M, Y_M) that has both AI and AD components. Note, that there is a mass both on parallel lines and on rays in exponential margins, and hence, both AD and AI behaviours are represented. We are interested in the tail behaviour of these models, where this feature is most apparent.

Wadsworth and Tawn (2012a) present a statistical analysis which shows the benefit of this mixture type of model, incorporating AD and AI, over established dependence models. As illustrated in Figure 5.1.1, our models put mass on rays and lines, which is inconsistent with most data applications where an assumption of a joint density everywhere is reasonable. Consequently, if these models are fitted using likelihood/Bayesian-based inference they would need almost as many parameters as data points to get a reasonable fit as each line of mass can only explain one data point. Therefore, as currently set up, these are not parsimonious models for likelihood inference but can be used as building blocks for future parsimonious model development. Alternatively, such models can be fitted using other inference criteria which do not depend on the mass on rays/lines. That though is not the focus

of this chapter. The aim of this chapter is to study mathematically the extremal structure of this class of models, the novel tools we use for this are introduced next.

To explore the tail behaviour of bivariate distributions with identical marginals the established approach is to adopt a so-called radial-angular representation. We want the radial component, R, to represent how far we are from the origin, and the angular component, W, to represent some form of measure of angle relative to the coordinate axes. This is common practice in multivariate extremes (see e.g. de Haan and Ferreira (2006) and Resnick (1987)). For Fréchet marginals, (X_F, Y_F), these correspond to $R_F = X_F + Y_F$ and $W_F = X_F/(X_F + Y_F)$, although other norms can be used to define these. Then in the limit as $r \to \infty$ the distribution $W_F \mid (R_F > r)$ is non-degenerate if (X_F, Y_F) are AD, but not perfectly dependent, but collapses to mass on $\{0\}$ and $\{1\}$ if the variables are AI. Here the extreme events being considered are those with $R_F > r$.

The key departures to this standard radial-angular approach in our work is that we focus on exponential margins, different combinations of the variables are considered to be extreme, and we use a different dependence variable than W_F. We consider the following radial-angular variables for general bivariate variables (X, Y) on exponential margins:

$$R = X + Y, \quad W_D = Y - X, \quad W_I = X/(X + Y). \tag{5.1.8}$$

Here two different angular variables W_D and W_I are considered. Also the radial variable R differs from R_F as X and Y are on exponential scale. We will explore these radial-angular variables for the transformed max-linear model (X_E, Y_E) in (5.1.4) and the inverted max-linear model $(X_E^{(I)}, Y_E^{(I)})$ in (5.1.5).

To help understand the difference in our new radial-angular variables first consider the connection between W_D and W_F. For large X_E we have that $X_E \approx \log(X_F)$, similarly for

Y_E, and so

$$W_F \approx \exp(X_E)/(\exp(X_E) + \exp(Y_E)) = 1/(1 + \exp(Y_E - X_E)).$$

Hence, for large X_E and Y_E, W_D is simply a function of W_F, and at first sight it would appear that this choice of radial and angular variable should not reveal any new structure. But conditioning on $R > r$ leads to the selection of different extreme events than $R_F > s$, for any choice of r and s, so different results can arise. Specifically, we will show that the radial and angular representation (R, W_D) gives a non-trivial limit for the distribution of $W_D \mid (R > r)$ as $r \to \infty$ for the transformed max-linear model, but for the inverted max-linear model it gives only mass at $\{-\infty\}$, $\{0\}$ or $\{\infty\}$ depending on the (α_i, β_i), $i = 1, \ldots, m$, values. The latter limit is at odds with results for W_F, as there the associated mass at $\{0\}$, corresponding to $W_F = 1/2$, does not arise. For the radial and angular formulation (R, W_I) the limit distribution of $W_I \mid (R_I > r)$, as $r \to \infty$, for the transformed max-linear model is degenerate, with limit $W_I = 1/2$, and is a non-trivial limit for the inverted max-linear model.

The layout of the chapter is as follows. In Section 5.2 we introduce a simple case of the max-linear model given in (5.1.3), called the Marshall-Olkin model, and we will use this to derive some of the key tail dependence properties of the model. The mathematical techniques used throughout are based on the techniques shown in this section. Then in Section 5.3 we derive properties for the general case for both the transformed max-linear and inverted max-linear models. In Section 5.4 we examine the asymptotic behaviour of the upper tail for both of these models. In Section 5.5 we combine the two models together and study the extremal properties of this formulation. Proofs of the results are given in Sections 5.6. We close with a discussion in Section 5.7 that discusses multivariate and spatial models extending our bivariate models.

5.2 Marshall-Olkin model

Let us consider a simple case of model (5.1.3). This corresponds to the Marshall and Olkin (1967) model, and has the following form:

$$X_F = \max\{\alpha Z_1, (1-\alpha)Z_2\},$$
$$Y_F = \max\{\beta Z_1, (1-\beta)Z_3\},$$

where Z_i, $i = 1, 2, 3$, are defined as in (5.1.3), and $0 \leqslant \alpha, \beta \leqslant 1$ are known constants. As there is only Z_1 in common between X_F and Y_F, a similar simulation to that shown on Figure 5.1.1a would give point mass on a single line, with the rest of the points scattered above and below the line. The variables X_F and Y_F are independent only in the cases when $\alpha = 1$ and $\beta = 0$ or $\alpha = 0$ and $\beta = 1$, otherwise they are dependent.

In order to characterise this model it is useful to define the following three cases: (i) on the line $Y_F = \frac{\beta}{\alpha}X_F$, (ii) below the line with $Y_F < \frac{\beta}{\alpha}X_F$, and (iii) above the line with $Y_F > \frac{\beta}{\alpha}X_F$. In each of these cases there are certain combinations of Z_i's that can lead to them. To have points on the line we need $(X_F, Y_F) = (\alpha Z_1, \beta Z_1)$, which requires $Z_2 \leqslant \alpha/(1-\alpha)Z_1$ and $Z_3 \leqslant \beta/(1-\beta)Z_1$. Below the line we need $(X_F, Y_F) = ((1-\alpha)Z_2, \beta Z_1)$ or $(X_F, Y_F) = ((1-\alpha)Z_2, (1-\beta)Z_3)$ with $(1-\alpha)Z_2/\alpha > (1-\beta)Z_3/\beta$, and above the line $(X_F, Y_F) = (\alpha Z_1, (1-\beta)Z_3)$ or $(X_F, Y_F) = ((1-\alpha)Z_2, (1-\beta)Z_3)$ with $(1-\alpha)Z_2/\alpha < (1-\beta)Z_3/\beta$.

In each case we can derive the probability of being in that case and the density conditional on being in each region. Here we will illustrate the calculations for case (i) when $Y_F = \frac{\beta}{\alpha}X_F$; i.e., we want to work out the probability that $X_F < x$ for some $x > 0$ given that $Y_F = \beta X_F/\alpha$. We use conditional probability:

$$\Pr\left\{X_F < x \middle| Y_F = \frac{\beta}{\alpha}X_F\right\} = \frac{\Pr\left\{X_F < x, Y_F = \frac{\beta}{\alpha}X_F\right\}}{\Pr\left\{Y_F = \frac{\beta}{\alpha}X_F\right\}}. \tag{5.2.1}$$

The joint probability in the numerator is

$$\begin{aligned}\Pr\left\{X_F < x, Y_F = \frac{\beta}{\alpha}X_F\right\} &= \Pr\left\{\alpha Z_1 < x, \alpha Z_1 > (1-\alpha)Z_2, \beta Z_1 > (1-\beta)Z_3\right\} \\ &= \Pr\left\{Z_1 < \frac{x}{\alpha}, Z_2 < \frac{\alpha Z_1}{1-\alpha}, Z_3 < \frac{\beta Z_1}{1-\beta}\right\}.\end{aligned}$$

To calculate this, we can condition on one of the Z's, in this case Z_1, and integrate over the range $Z_1 < x/\alpha$, which gives

$$\begin{aligned}\Pr\left\{X_F < x, Y_F = \frac{\beta}{\alpha}X_F\right\} &= \int_0^{x/\alpha} \Pr\left\{Z_2 < \frac{\alpha z}{1-\alpha}, Z_3 < \frac{\beta z}{1-\beta}\middle| Z_1 = z\right\} f_Z(z)dz, \\ &= \int_0^{x/\alpha} e^{-(1-\alpha)/(\alpha z)} e^{-(1-\beta)/(\beta z)} \frac{1}{z^2} e^{-1/z} dz, \\ &= \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \exp\left(-\frac{\alpha+\beta-\alpha\beta}{\beta x}\right),\end{aligned}$$

where the second equality holds as Z_2 and Z_3 are independent Fréchet random variables. It follows that

$$\Pr\left\{Y_F = \frac{\beta}{\alpha}X_F\right\} = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta},$$

and hence we have obtained the conditional distribution in (5.2.1) as $\exp\{-(\alpha+\beta-\alpha\beta)/(\beta x)\}$ for $x > 0$. To obtain the one-dimensional density of the points on the line we can differentiate this distribution function, which gives

$$f_{X_F}\left(x\middle| Y_F = \frac{\beta}{\alpha}X_F\right) = \frac{\alpha+\beta-\alpha\beta}{\beta x^2} \exp\left(-\frac{\alpha+\beta-\alpha\beta}{\beta x}\right), \text{ for } x > 0,$$

or, equivalently,

$$f_{Y_F}\left(y\middle| Y_F = \frac{\beta}{\alpha}X_F\right) = \frac{\alpha+\beta-\alpha\beta}{\alpha y^2} \exp\left(-\frac{\alpha+\beta-\alpha\beta}{\alpha y}\right), \text{ for } y > 0.$$

See Appendix 5.A for similar calculations for the other two cases using this first principles approach. Deriving the densities as described above is laborious, involving many complex

integrals, which makes the calculations hard to extend to the more general case. As the densities seem to have much simpler forms than the distribution functions it seems sensible to work with densities directly. For example, above the line $Y_F > (\beta/\alpha)X_F$ the probability element is

$$\Pr\left\{X_F \in dx, Y_F \in dy \middle| Y_F > \frac{\beta}{\alpha}X_F\right\} = \frac{\Pr\left\{X_F \in dx, Y_F \in dy, Y_F > \frac{\beta}{\alpha}X_F\right\}}{\Pr\left\{Y_F > \frac{\beta}{\alpha}X_F\right\}}, \qquad (5.2.2)$$

where $X \in dx$ denotes $X \in (x, x+\delta x)$. Then, there are two possible combinations that lead to this case, $(X_F, Y_F) = (\alpha Z_1, (1-\beta)Z_3)$ and $(X_F, Y_F) = ((1-\alpha)Z_2, (1-\beta)Z_3)$, so the joint probability in the numerator of expression (5.2.2) can be broken down into the sum of two probabilities, P_1 and P_2, where

$$P_1 = \Pr\left\{X_F = \alpha Z_1 \in dx, Y_F = (1-\beta)Z_3 \in dy, Y_F > \frac{\beta}{\alpha}X_F\right\},$$
$$P_2 = \Pr\left\{X_F = (1-\alpha)Z_2 \in dx, Y_F = (1-\beta)Z_3 \in dy, Y_F > \frac{\beta}{\alpha}X_F\right\}.$$

Then, it follows that the probability P_1 is equivalent to the joint probability $\Pr\{\alpha Z_1 \in dx, (1-\beta)Z_3 \in dy, Z_2 < x/(1-\alpha)\}$ given that $y > (\beta x)/\alpha$. Hence, using that the Z_i's are independent Fréchet random variables,

$$\begin{aligned} P_1 &= \Pr\left(Z_1 \in \frac{dx}{\alpha}\right)\Pr\left(Z_3 \in \frac{dy}{1-\beta}\right)\Pr\left(Z_2 < \frac{x}{1-\alpha}\right)\mathbb{I}\left(y > \frac{\beta}{\alpha}x\right) \\ &\sim \left(\frac{\alpha}{x^2}e^{-\alpha/x}\right)\left(\frac{1-\beta}{y^2}e^{-(1-\beta)/y}\right)\left(e^{-(1-\alpha)/x}\right)\mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\delta x\delta y \\ &= \frac{\alpha(1-\beta)}{x^2y^2}e^{-1/x}e^{-(1-\beta)/y}\ \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\delta x\delta y, \end{aligned}$$

as $\delta x \to 0$ and $\delta y \to 0$. Similarly, as $\delta x \to 0$ and $\delta y \to 0$,

$$P_2 \sim \frac{(1-\alpha)(1-\beta)}{x^2y^2}e^{-1/x}e^{-(1-\beta)/y}\ \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\delta x\delta y.$$

Hence, by summing P_1 and P_2, as $\delta x \to 0$ and $\delta y \to 0$,

$$\Pr\left\{X_F \in dx, Y_F \in dy, Y_F > \frac{\beta}{\alpha}X_F\right\} \sim \frac{(1-\beta)}{x^2y^2}e^{-1/x}e^{-(1-\beta)/y}\ \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\delta x\delta y.$$

This can be integrated in the region $y > (\beta x)/\alpha$ to obtain the probability

$$\Pr\left\{Y_F > \frac{\beta}{\alpha}X_F\right\} = \frac{\alpha(1-\beta)}{\alpha+\beta-\alpha\beta}.$$

Hence, we obtain the density, conditionally on being above the line $Y_F > (\beta/\alpha)X_F$, as

$$f_{(X_F,Y_F)}\left(x, y\middle|Y_F > \frac{\beta}{\alpha}X_F\right) = \frac{\alpha+\beta-\alpha\beta}{\alpha x^2y^2}e^{-1/x}e^{-(1-\beta)/y}\ \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right).$$

Similar calculations can be performed to obtain densities for cases (i) and (ii).

5.3 General max-linear models

5.3.1 Set up and densities on Fréchet margins

Our work in this section has considerable parallels with the hitting scenarios and the conditional probability results for max-linear models developed by Wang and Stoev (2011). Here, we go beyond the scope of this chapter by calculating conditional densities.

Let us consider the general max-linear model given in expression (5.1.3). Without loss of generality, let us assume that the $\alpha_i Z_i$ and $\beta_i Z_i$ terms are ordered such that

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_{k+l}, 0, \ldots, 0),$$
$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k, 0, \ldots, 0, \beta_{k+l+1}, \ldots, \beta_m),$$

i.e., for $i = 1, \ldots, k$, $\alpha_i \neq 0$ and $\beta_i \neq 0$, for $i = k+1, \ldots, k+l$, $\alpha_i \neq 0$ and $\beta_i = 0$, and for $i = k+l+1, \ldots, m$, $\alpha_i = 0$ and $\beta_i \neq 0$, with $\sum_{i=1}^{k+l}\alpha_i = 1$ and $\sum_{i=1}^{k}\beta_i + \sum_{h=k+l+1}^{m}\beta_h = 1$.

We also assume that $\omega_i := \beta_i/\alpha_i$ are unique for $i = 1, \ldots, k$. In this general case there are k common Z_i variables between X_F and Y_F, hence there is mass on k lines, each with equation $Y_F = (\beta_i/\alpha_i)X_F$, $i = 1, \ldots, k$. If $k = 0$ then X_F and Y_F are independent. Furthermore, without loss of generality, let us assume the following ordering for the first k terms,

$$\frac{\beta_1}{\alpha_1} < \frac{\beta_2}{\alpha_2} < \ldots < \frac{\beta_k}{\alpha_k} \Leftrightarrow 0 < \omega_1 < \omega_2 < \ldots < \omega_k < \infty.$$

This notation ensures that the line with mass that has the least steep gradient is $Y_F = \omega_1 X_F$, followed by $Y_F = \omega_2 X_F$, and so on until $Y_F = \omega_k X_F$. Let us also define the following sums,

$$\alpha_{sum} = \alpha_{k+1} + \ldots + \alpha_{k+l}, \qquad \alpha_{sum}^{(j)} = \alpha_{sum} + \Sigma_{i=1}^{j}\alpha_i, \text{ for } 0 \leqslant j \leqslant k,$$
$$\beta_{sum} = \beta_{k+l+1} + \ldots + \beta_m, \qquad \beta_{sum}^{(h)} = \beta_{sum} + \Sigma_{i=h}^{k}\beta_i, \text{ for } 1 \leqslant h \leqslant k+1,$$

where we define $\sum_{i=1}^{0} x_i = 0$ and $\sum_{i=k+1}^{k} x_i = 0$, which leads to $\alpha_{sum}^{(0)} = \alpha_{sum}$ and $\beta_{sum}^{(k+1)} = \beta_{sum}$.

In this more general set up it is useful to define four types of 'regions': (i) above the line $Y_F = \omega_k X_F$, (ii) on the line $Y_F = \omega_j X_F$, $j = 1, \ldots, k$, (iii) between the two lines $Y_F = \omega_j X_F$ and $Y_F = \omega_{j+1} X_F$, $j = 1, \ldots, k-1$, and (iv) below the line $Y_F = \omega_1 X_F$. There is one region of type (i) and (iv) each, k regions of type (ii) since there are k lines, and $k-1$ regions of type (iii), since k lines define $k-1$ between-line regions.

The strategy for the derivation of the densities for each of these regions is as in Section 5.2, with full derivations given in Appendix 5.B. Here we will give the conditional density forms for each of the four region types. The density conditional on being in the region above the line $Y_F = \omega_k X_F$ is

$$f_{(X_F,Y_F)}\left(x, y \middle| \frac{Y_F}{X_F} > \omega_k\right) = \frac{\alpha_k \beta_{sum} + \beta_k}{\alpha_k x^2 y^2} \exp\left(-\frac{1}{x} - \frac{\beta_{sum}}{y}\right),$$

for $y > \omega_k x$. On the line $Y_F = \omega_j X_F$, $j = 1, \ldots, k$, the density for $x > 0$ is

$$f_{(X_F,Y_F)}\left(x, \omega_j x \middle| \frac{Y_F}{X_F} = \omega_j\right) = \frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j x^2} \exp\left(-\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j x}\right).$$

Between two lines $Y_F = \omega_j X_F$ and $Y_F = \omega_{j+1} X_F$, $j = 1, \ldots, k-1$, the conditional density is

$$f_{(X_F,Y_F)}\left(x, y \middle| \omega_j < \frac{Y_F}{X_F} < \omega_{j+1}\right) = \frac{c_j}{(\alpha_j \beta_{j+1} - \beta_j \alpha_{j+1}) x^2 y^2} \exp\left(-\frac{\alpha_{sum}^{(j)}}{x} - \frac{\beta_{sum}^{(j+1)}}{y}\right),$$

for $\omega_j x < y < \omega_{j+1} x$ where $c_j = (\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})(\alpha_{j+1} \beta_{sum}^{(j+1)} + \beta_{j+1} \alpha_{sum}^{(j)})$. Finally, in the region below the line $Y_F = \omega_1 X_F$, the conditional density is

$$f_{(X_F,Y_F)}\left(x, y \middle| \frac{Y_F}{X_F} < \omega_1\right) = \frac{\alpha_1 + \beta_1 \alpha_{sum}}{\beta_1 x^2 y^2} \exp\left(-\frac{\alpha_{sum}}{x} - \frac{1}{y}\right),$$

for $y < \omega_1 x$.

5.3.2 Densities on exponential margins

Transformed max-linear model

In Section 5.3.1 we gave densities conditional on being on each line and in the regions defined by the lines on Fréchet margins. Since it is more straightforward to expose the difference between AI and AD on exponential margins, we want to obtain the densities for exponential margins. Hence for each case (i)-(iv) defined in Section 5.3.1 we will identify the corresponding case on exponential margins and then transform to obtain the densities on the new margins.

On exponential margins for the transformed max-linear model (X_E, Y_E), the line $Y_F = \omega_j X_F$ becomes the curve $Y_E = -\log(1 - (1 - e^{-X_E})^{1/\omega_j})$ for all $j \in \{1, \ldots, k\}$. For ease of notation, let us define $g_j(X_E) = -\log(1 - (1 - e^{-X_E})^{1/\omega_j})$ for $j = 1, \ldots, k$. Note that

$g_j(X_E) \approx X_E + \log(\omega_j)$ for large X_E; this asymptotic linearity is useful when we explore the limiting behaviour of the model. The cases defined in Section 5.3.1 become (i) above the curve $Y_E = g_k(X_E)$, (ii) on the curve $Y_E = g_j(X_E)$, $j = 1, \ldots, k$, (iii) between the two curves $Y_E = g_j(X_E)$ and $Y_E = g_{j+1}(X_E)$, $j = 1, \ldots, k-1$, and (iv) below the curve $Y_E = g_1(X_E)$. Note that the transformation to exponential margins means that the lines with mass on are now curves. Furthermore, even asymptotically they are no longer rays that meet at the origin, but parallel lines each with gradient equal to one with intercepts $\log(\omega_j)$, $j = 1, \ldots, k$. For each region we transform the conditional densities given in Section 5.3.1 to exponential margins.

Hence, the conditional density in the region above the curve $Y_E = g_k(X_E)$ is:

$$f_{(X_E,Y_E)}(x, y|Y_E > g_k(X_E)) = \left(\frac{\alpha_k\beta_{sum} + \beta_k}{\alpha_k}\right) e^{-x}e^{-y}(1 - e^{-y})^{\beta_{sum}-1}, \tag{5.3.1}$$

for $x > 0$ and $y > g_k(x)$. On the curve $Y_E = g_j(X_E)$, $j = 1, \ldots, k$, the conditional density for $x > 0$ is:

$$f_{(X_E,Y_E)}(x, g_j(x)|Y_E = g_j(X_E)) = \left(\frac{\alpha_j\beta_{sum}^{(j+1)} + \beta_j\alpha_{sum}^{(j)}}{\beta_j}\right) e^{-x}(1-e^{-x})^{(\alpha_j\beta_{sum}^{(j+1)}+\beta_j\alpha_{sum}^{(j)})/\beta_j-1}.$$

The conditional density between the curves $Y_E = g_j(X_E)$ and $Y_E = g_{j+1}(X_E)$, for $j = 1, \ldots, k-1$, is:

$$f_{(X_E,Y_E)}(x, y|g_j(X_E) < Y_E < g_{j+1}(X_E)) = \frac{(\alpha_j\beta_{sum}^{(j+1)} + \beta_j\alpha_{sum}^{(j)})(\alpha_{j+1}\beta_{sum}^{(j+1)} + \beta_{j+1}\alpha_{sum}^{(j)})}{(\alpha_j\beta_{j+1} - \beta_j\alpha_{j+1})} \times e^{-x}e^{-y}(1 - e^{-x})^{\alpha_{sum}^{(j)}-1}(1 - e^{-y})^{\beta_{sum}^{(j+1)}-1},$$

for $x > 0$ and $g_j(x) < y < g_{j+1}(x)$, $j = 1, \ldots, k-1$. The conditional density below the curve $Y_E = g_1(X_E)$ is

$$f_{(X_E,Y_E)}(x, y|Y_E < g_1(X_E)) = \left(\frac{\alpha_1 + \beta_1\alpha_{sum}}{\beta_1}\right) e^{-x}(1 - e^{-x})^{\alpha_{sum}-1}e^{-y},$$

for $x > 0$ and $y < g_1(x)$.

Inverted max-linear model

Now we turn our attention to the lower tail of the max-linear model (5.1.3), i.e. the upper tail of the inverted max-linear model. Similarly to Section 5.3.2, the densities given in Section 5.3.1 can be transformed to inverted exponential margins $(X_E^{(I)}, Y_E^{(I)})$.

To invert the lower tail of X_E, set $U = 1 - e^{-X_E}$. Then the inversion of U is $U^{(I)} = 1 - U = e^{-X_E}$. Also, $U^{(I)} = 1 - e^{-X_E^{(I)}}$, which leads to $e^{-X_E} = 1 - e^{-X_E^{(I)}}$. Hence, $X_E^{(I)} = -\log(1 - e^{-X_E})$ and by substituting in X_E from expression (5.1.4) we get $X_E^{(I)} = 1/X_F$. Similarly, $Y_E^{(I)} = 1/Y_F$.

On the new inverted exponential margins, the line $Y_F = \omega_j X_F$ becomes $Y_E^{(I)} = X_E^{(I)}/\omega_j$ for $j = 1, \ldots, k$. Hence the cases defined in Section 5.3.1 become (i) below the line $Y_E^{(I)} = X_E^{(I)}/\omega_k$, (ii) on the line $Y_E^{(I)} = X_E^{(I)}/\omega_j$, $j = 1, \ldots, k$, (iii) between the two lines $Y_E^{(I)} = X_E^{(I)}/\omega_{j+1}$ and $Y_E^{(I)} = X_E^{(I)}/\omega_j$, $j = 1, \ldots, k-1$, and (iv) above the line $Y_E^{(I)} = X_E^{(I)}/\omega_j$. Note that the transformation flips the order of the lines, with the line $Y_E^{(I)} = X_E^{(I)}/\omega_1$ having the steepest gradient and $Y_E^{(I)} = X_E^{(I)}/\omega_k$ the least steep.

The conditional density below the line $Y_E^{(I)} = X_E^{(I)}/\omega_k$ has the following form:

$$f_{(X_E^{(I)}, Y_E^{(I)})}\left(x, y \mid Y_E^{(I)} < X_E^{(I)}/\omega_k\right) = \frac{\alpha_k \beta_{sum} + \beta_k}{\alpha_k} e^{-x} e^{-\beta_{sum} y},\ x > 0,\ y < x/\omega_k.$$

On the line $Y_E^{(I)} = X_E^{(I)}/\omega_j$, $j = 1, \ldots, k$, the conditional density for $x > 0$ takes the form:

$$f_{(X_E^{(I)}, Y_E^{(I)})}\left(x, x/\omega_j \mid Y_E^{(I)} = X_E^{(I)}/\omega_j\right) = \frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j} \exp\left\{-\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j} x\right\}.$$

Between the two lines $Y_E^{(I)} = X_E^{(I)}/\omega_{j+1}$ and $Y_E^{(I)} = X_E^{(I)}/\omega_j$, $j = 1, \ldots, k-1$, the conditional

density is:

$$f_{(X_E^{(I)},Y_E^{(I)})}\left(x,y \mid X_E^{(I)}/\omega_{j+1} < Y_E^{(I)} < X_E^{(I)}/\omega_j\right) = \frac{c_j}{(\alpha_j\beta_{j+1} - \beta_j\alpha_{j+1})}\, e^{-\alpha_{sum}^{(j)}x} e^{-\beta_{sum}^{(j+1)}y}.$$

where $c_j = (\alpha_j\beta_{sum}^{(j+1)} + \beta_j\alpha_{sum}^{(j)})(\alpha_{j+1}\beta_{sum}^{(j+1)} + \beta_{j+1}\alpha_{sum}^{(j)})$, $x > 0$ and $x/\omega_j < y < x/\omega_{j+1}$. Finally, above the line $Y_E^{(I)} = X_E^{(I)}/\omega_1$ the conditional density is:

$$f_{(X_E^{(I)},Y_E^{(I)})}\left(x,y \mid Y_E^{(I)} > X_E^{(I)}/\omega_1\right) = \frac{\alpha_1 + \beta_1\alpha_{sum}}{\beta_1} e^{-\alpha_{sum}x} e^{-y},\ x > 0,\ y > x/\omega_1.$$

5.4 Angular representation and limiting behaviour

In this section we explore the asymptotic behaviour of the upper tails of the transformed max-linear model (5.1.6) and the inverted max-linear model (5.1.7). As discussed in Section 5.1, we use a radial-angular representation (R, W) to explore the limiting properties of the models. For general exponential marginal variables (X, Y) we define the radial component to be of the form $R = X + Y$. For the angular component we use two different forms: $W_D = Y - X$ and $W_I = X/(X + Y)$ for the reasons given in Section 5.1. Our aim is to determine the tail behaviour of the models in the region $\{R > r\}$ as $r \to \infty$. So for each type of region $\mathcal{J}$ (identified in the previous sections), and for both forms of W, we will also calculate the joint density of R and W given that $R > r$ to give the conditional probability

$$\Pr(W > w, R > r + t \mid R > r, W \in \mathcal{J}) = \frac{\Pr(W > w, R > r + t \mid W \in \mathcal{J})}{\Pr(R > r \mid W \in \mathcal{J})},\ \ t > 0. \quad (5.4.1)$$

Then we can use these results to obtain the conditional probability of being in each region $\mathcal{J}$, given $R > r$ as $r \to \infty$, as

$$\Pr(W \in \mathcal{J} \mid R > r) = \frac{\Pr(R > r \mid W \in \mathcal{J}) \Pr(W \in \mathcal{J})}{\Pr(R > r)},\ \text{for all } \mathcal{J}. \quad (5.4.2)$$

5.4.1 Transformed max-linear model

First, we explore the asymptotic behaviour of the upper tail of the transformed max-linear model (5.1.6). We use the densities in Section 5.3.2, to obtain the densities in each region on (R, W_D) margins.

For (R, W_D) the curve $Y_E = g_j(X_E)$, $j = 1, \ldots, k$, is $W_D = g_j(X_E) - X_E$, which is approximately $W_D = \log(\omega_j) := w_j$, with $-\infty < w_j < \infty$, for large R and hence for large X_E. So the case (i) becomes approximately the region $W_D > w_k$ for large R. For finite samples the region is $W_D > g_k(X_E) - X_E$. The joint density conditional on being in this region is obtained from the density in (5.3.1) as:

$$f_{(R,W_D)}\left(r, w | W_D > g_k(X_E) - X_E\right) = \left(\frac{\alpha_k \beta_{sum} + \beta_k}{2\alpha_k}\right) e^{-r} (1 - e^{-(r+w)/2})^{\beta_{sum}-1}, \quad (5.4.3)$$

for $r > 0$ and $w > g_k(X_E) - X_E$. We can then calculate the conditional probability in (5.4.1) as

$$\Pr(W_D > w, R > r + t \mid R > r, W_D > g_k(X_E) - X_E) \approx \left(1 + \frac{t}{r+1}\right) e^{-t} \to e^{-t},$$

as $r \to \infty$, for $t > 0$ and $w > w_k$. This shows that $\Pr(W_D > w \mid R > r) \to 1$ for all $w > w_k$. Hence Lemma 1 follows.

Lemma 1. *In the limit as $r \to \infty$, $W_D \mid \{R > r, W_D > w_k\} \to^p w_k$, and asymptotically $W_D \perp\!\!\!\perp R \mid \{R > r, W_D > w_k\}$.*

On the curve $W_D = g_j(X_E) - X_E$, $j = 1, \ldots, k$, the density is

$$f_R\left(r | W_D = g_j(X_E) - X_E\right) = c_j e^{-(r-w_j)/2} (1 - e^{-(r-w_j)/2})^{2c_j - 1}, \; r > 0,$$

where $c_j = (\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})/(2\beta_j)$. Then, the distribution of the points, conditional on

being on this curve is

$$\Pr(R > r \mid W_D = g_j(X_E) - X_E) = 1 - (1 - e^{w_j/2} e^{-r/2})^{c_j},$$
$$\sim c_j e^{w_j/2} e^{-r/2}, \quad \text{as } r \to \infty,$$

for $j = 1, \ldots, k$.

Lemma 2. *The distribution of the radial points on the line $W_D = w_j$, $j = 1, \ldots, k$, has an exponential tail.*

In the region between the curves $W_D = g_j(X_E) - X_E$ and $W_D = g_{j+1}(X_E) - X_E$, $j = 1, \ldots, k-1$, the joint density is:

$$f_{(R,W_D)}(r, w | g_j(X_E) - X_E < W_D < g_{j+1}(X_E) - X_E) = \frac{c_j}{2} e^{-r} (1 - e^{-(r-w)/2})^{\alpha_{sum}-1}$$
$$\times (1 - e^{-(r+w)/2})^{\beta_{sum}-1},$$

for $r > 0$ and $g_j(X_E) - X_E < w < g_{j+1}(X_E) - X_E$. Then the conditional probability,

$$\Pr(W_D > w, R > r + t \mid R > r, g_j(X_E) - X_E < W_D < g_{j+1}(X_E) - X_E) \to \frac{(w_{j+1} - w)e^{-t}}{w_{j+1} - w_j},$$

as $r \to \infty$, for $t > 0$ and $w_j < w < w_{j+1}$. Hence Lemma 3 follows.

Lemma 3. *The limiting angular distribution is uniform in regions between the rays $W_D = w_j$ and $W_D = w_{j+1}$, $j = 1, \ldots, k-1$, and independent of the radial variable, which follows a unit exponential distribution.*

Lastly, in the region $W_D < g_1(X_E) - X_E$ the joint density is

$$f_{(R,W_D)}(r, w | W_D < g_1(X_E) - X_E) = \left(\frac{\alpha_j + \beta_j \alpha_{sum}}{2\beta_j} \right) e^{-r} (1 - e^{-(r-w)/2})^{\alpha_{sum}-1},$$

for $r > 0$ and $w < g_1(X_E) - X_E$. Then the conditional probability,

$$\Pr(W_D > w, R > r + t \mid R > r, W_D < w_1) \sim \frac{(w_1 - w)e^{-t}}{w_1 + r + 1}, \quad \text{as } r \to \infty, \ \ t > 0, \ \ w < w_1.$$

This suggests that $\Pr(W_D > w \mid R > r, W_D < w_1) \to 0$ as $r \to \infty$ for all $w < w_1$ and the following lemma follows.

Lemma 4. *In the limit as $r \to \infty$, $W_D \mid \{R > r, W_D < w_1\} \to^p w_1$, and asymptotically $W_D \perp\!\!\!\perp R \mid \{R > r, W_D < w_1\}$.*

Now we use the results above to calculate the probability of being in each region $\mathcal{J}$, given $R > r$ as $r \to \infty$. Theorems 4 and 5 describe the asymptotic behaviour of the conditional probability (5.4.2) for angular measures $W = W_D$ and $W = W_I$, respectively. Proofs are deferred to Section 5.6.

Theorem 4. *Let $R = X_E + Y_E$, $W_D = Y_E - X_E$ and $w_j = \log(\beta_j/\alpha_j)$. Then, as $r \to \infty$,*

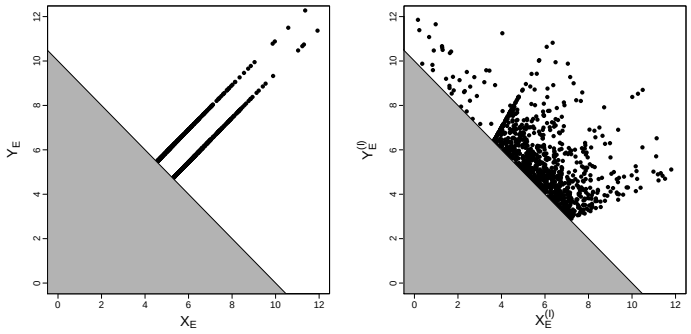
$$\Pr(W_D \in \mathcal{J} \mid R > r) \to \begin{cases} \dfrac{\alpha_j \exp(w_j/2)}{\sum_{i=1}^{k} \alpha_i \exp(w_i/2)}, & \text{for } \mathcal{J} = \{w_j\}, j = 1, \ldots, k \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 5. *Let $R = X_E + Y_E$ and $W_I = X_E/(X_E + Y_E)$. Then, as $r \to \infty$,*

$$\Pr(W_I < w \mid R > r) \to \begin{cases} 0, & w < 1/2, \\ 1, & w \geqslant 1/2, \end{cases}$$

i.e., $W_I \mid R > r \to^p 1/2$.

Thus, Theorem 4 shows that in the limit $r \to \infty$, there is only mass on the lines $W_D = w_j$, $j = 1, \ldots, k$, and not in any of the other regions for this model. Figure 5.4.1a illustrates this for the max-linear model with the same α and β parameters as in Figure 5.1.1. For the other angular form W_I, the mass collapses onto the diagonal, as shown by Theorem 5, and hence W_I is a poor angular measure for exploring the extremal dependence structure for AD variables.



(a) (b)

Figure 5.4.1: Asymptotic behaviour of the (a) transformed max-linear and (b) inverted max-linear model on exponential margins, for X_F and Y_F defined as in Figure 5.1.1.

5.4.2 Inverted max-linear model

Now we explore the asymptotic upper tail behaviour of the inverted max-linear model (5.1.7). We transform the densities given in Section 5.3.2 to obtain the densities in each region on (R, W_I) margins. On these new margins, the line $Y_E^{(I)} = X_E^{(I)}/\omega_j$ becomes the line $W_I = \omega_j/(1+\omega_j)$, $j = 1, \ldots, k$. Let us denote $w_j = \omega_j/(1+\omega_j)$ for $j = 1, \ldots, k$. Note w_j here is different than in Section 5.4.1. The lines are then ordered such that $0 < w_1 < w_2 < \ldots < w_k < 1$. Then, the region below the line $Y_E^{(I)} = X_E^{(I)}/\omega_k$ becomes the region $w_k < W_I < 1$. The conditional density in this region is:

$$f_{(R,W_I)}(r, w \mid w_k < W_I < 1) = \left(\frac{\alpha_j \beta_{sum} + \beta_j}{\alpha_j}\right) r e^{-\beta_{sum} r} e^{-wr(1-\beta_{sum})},\ r > 0,\ w_k < w < 1.$$

To determine the limiting behaviour for $r \to \infty$ we calculate the conditional probability $\Pr(W_I > w, R > r + t \mid w_k < W_I < 1, R > r)$. We obtain the joint survival function of W_I and R as

$$\Pr(W_I > w, R > r \mid w_k < W_I < 1) = \frac{\alpha_k \beta_{sum} + \beta_k}{\alpha_k(1-\beta_{sum})} \left[\frac{e^{-rc(w)}}{c(w)} - e^{-r}\right],\ \ r > 0,\ \ w_k < w < 1, \tag{5.4.4}$$

where $c(w) = \beta_{sum} + w(1 - \beta_{sum})$. Setting $w = w_k$ in (5.4.4) we get the conditional distribution for R given $\omega_k < W_I < 1$. Hence, for $t > 0$ and $w_k < w < 1$,

$$\Pr(W_I > w, R > r + t \mid R > r, w_k < W_I < 1) = \frac{[c(w)]^{-1} e^{-(r+t)c(w)} - e^{-(r+t)}}{[c(w_k)]^{-1} e^{-rc(w_k)} - e^{-r}}.$$

Now note that $c(w)$ is an increasing function for $w_k < w < 1$ and $c(1) = 1$. Hence, as $r \to \infty$,

$$\Pr(W_I > w, R > r + t \mid R > r, w_k < W_I < 1) \sim \frac{c(w_k)}{c(w)} e^{-tc(w)} e^{-r[c(w)-c(w_k)]}, \;\; w_k < w < 1$$

which as $r \to \infty$ tends to 0 for all $w \in (w_k, 1)$, and equals $e^{-tc(w_k)}$ for $w = w_k$, leading to Lemma 5.

Lemma 5. *$W_I \mid \{R > r, w_k < W_I < 1\} \to^p w_k$, as $r \to \infty$, and asymptotically W_I is independent of R, which has an exponential tail with rate $c(w_k)$.*

The conditional density of R on the line $W_I = w_j$, $j = 1, \ldots, k$, is

$$f_R(r \mid W_I = w_j) = \left(\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j} \right) \exp\left\{ -\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j} r \right\}, \text{ for } r > 0.$$

Lemma 6. *The distribution of $R \mid \{W_I = w_j\}$, is exponential with rate $(\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})/\beta_j$, $j = 1, \ldots, k$.*

For the region between the lines $W_I = w_j$ and $W_I = w_{j+1}$ the conditional density is

$$f_{(R,W_I)}(r, w \mid w_j < W_I < w_{j+1}) = \left(\frac{(\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})(\alpha_{j+1} \beta_{sum}^{(j+1)} + \beta_{j+1} \alpha_{sum}^{(j)})}{(\alpha_j \beta_{j+1} - \beta_j \alpha_{j+1})} \right) \times r e^{-\beta_{sum}^{(j+1)} r} e^{-(\alpha_{sum}^{(j)} + \beta_{sum}^{(j)}) wr},$$

for $r > 0$ and $w_j < w < w_{j+1}$. Similarly to above we calculate the conditional probability

as

$$\Pr(W_I > w, R > r + t \mid w_j < W_I < w_{j+1}, R > r) = \frac{[c(w_{j+1})]^{-1} e^{-(r+t)c(w_{j+1})} - [c(w)]^{-1} e^{-(r+t)c(w)}}{[c(w_{j+1})]^{-1} e^{-rc(w_{j+1})} - [c(w_j)]^{-1} e^{-rc(w_j)}},$$

where $t > 0$, $w_j < w < w_{j+1}$ and $c(w) = w(\alpha_{sum}^{(j)} + \beta_{sum}^{(j+1)}) + \beta_{sum}^{(j+1)}$. Here $c(w)$ is an increasing function for $w_j < w < w_{j+1}$, so $c(w_j) < c(w) < c(w_{j+1})$. Hence, as $r \to \infty$, for $w_j < w < w_{j+1}$,

$$\Pr(W_I > w, R > r + t \mid w_j < W_I < w_{j+1}, R > r) \sim \frac{c(w_j)}{c(w)} e^{-tc(w)} e^{-r[c(w)-c(w_j)]},$$

which tends to 0 for $r \to \infty$ for all $w \in (w_j, w_{j+1})$, and tends to $e^{-tc(w_j)}$ for $w = w_j$. Hence, Lemma 7 follows.

Lemma 7. *In the limit as $r \to \infty$, $W_I \mid \{R > r, w_j < W_I < w_{j+1}\} \to^p w_j$, $j = 1, \ldots, k-1$, and asymptotically $W_I \perp\!\!\!\perp R \mid \{R > r, w_j < W_I < w_{j+1}\}$.*

On (R, W_I) margins the region above the line $Y_E^{(I)} = X_E^{(I)}/\omega_1$ translates to the area represented by $W_I < w_1$. The conditional density in this region is

$$f_{(R,W_I)}(r, w \mid W_I < w_1) = \left(\frac{\alpha_1 + \beta_1 \alpha_{sum}}{\beta_1}\right) e^{-rw(\alpha_{sum}-1)} e^{-r}, \text{ for } r > 0, w < w_1.$$

Again, we can work out the conditional probability, for $t > 0$ and $w < w_1$,

$$\Pr(W_I > w, R > r + t \mid W_I < w_1, R > r) = \frac{[c(w_1)]^{-1} e^{-(r+t)c(w_1)} - [c(w)]^{-1} e^{-(r+t)c(w)}}{[c(w_1)]^{-1} e^{-rc(w_1)} - e^{-r}},$$

where $c(w) = 1 - (1 - \alpha_{sum})w$. The function $c(w)$ is in this case a decreasing function for $w \in (0, w_1)$, and $c(0) = 1$, so $0 < c(w_1) < c(w) < 1$. This means that as $r \to \infty$,

$$\Pr(W_I > w, R > r + t \mid W_I < w_1, R > r) \to e^{-tc(w)},$$

for all $w \in (0, w_1)$, suggesting that all points in this region will tend to $W_I = 0$ asymptotically as $r \to \infty$.

Lemma 8. *As $r \to \infty$, $W_I \mid \{R > r, W_I < w_1\} \to^p 0$, and asymptotically W_I is independent of R, which is unit exponential.*

Now we use the above results to calculate the conditional probability (5.4.2). Theorems 6 and 7 describe the behaviour of the conditional probability of points being in each region given that $R > r$ as $r \to \infty$, i.e., expression (5.4.2) with $W = W_I$ and $W = W_D$, respectively, where $\mathcal{J}$ denotes the different regions. Proofs are given in Section 5.6.

Theorem 6. *Let $R = X_E^{(I)} + Y_E^{(I)}$ and $W_I = X_E^{(I)}/(X_E^{(I)} + Y_E^{(I)})$. Let $\gamma_j = (\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})/(\alpha_j + \beta_j)$, $0 < \gamma_j \leqslant 1$, for $j = 1, \ldots, k$, and $\gamma_{min} = \min_{j=1,\ldots,k}(\gamma_j)$. If there is a unique γ_j value, $j = 1, \ldots, k$, equal to γ_{min}, i.e., $\gamma_{min} = \gamma_j$, then, for $t > 0$, as $r \to \infty$,*

1. $\Pr(W_I = w_j, R > r + t \mid R > r) \to (a_j/d_j)e^{-\gamma_j t}$,
2. $\Pr(w_j < W_I < w_{j+1}, R > r + t \mid R > r) \to (b_j/d_j)e^{-\gamma_j t}$,
3. $\Pr(w_{j-1} < W_I < w_j, R > r + t \mid R > r) \to (-b_{j-1}/d_j)e^{-\gamma_j t}$,
4. $\Pr(\{0 < W_I < w_{j-1}\} \cup \{w_j < W_I < 1\}, R > r + t \mid R > r) \to 0$,

where $w_0 = 0$, $w_{k+1} = 1$, $a_j = \alpha_j \beta_j/(\alpha_j + \beta_j)$, $b_j = \alpha_{sum}^{(j)} \beta_{sum}^{(j+1)}/(\alpha_{sum}^{(j)} - \beta_{sum}^{(j+1)})$ and $d_j = a_j + b_j - b_{j-1}$, $j = 1, \ldots, k$.

Theorem 7. *Let $R = X_E^{(I)} + Y_E^{(I)}$ and $W_D = Y_E^{(I)} - X_E^{(I)}$. Let a_j, b_j, d_j, γ_j and γ_{min} be defined as in Theorem 6. If there is a unique γ_j value, $j = 1, \ldots, k$, equal to γ_{min}, i.e. $\gamma_{min} = \gamma_j$, then as $r \to \infty$,*

$$W_D \mid R > r \to \begin{cases} +\infty, & \text{with probability } \mathbb{I}(\alpha_j > \beta_j) + \mathbb{I}(\alpha_j = \beta_j)(-b_{j-1}/d_j), \\ 0, & \text{with probability } \mathbb{I}(\alpha_j = \beta_j)(a_j/d_j), \\ -\infty, & \text{with probability } \mathbb{I}(\alpha_j < \beta_j) + \mathbb{I}(\alpha_j = \beta_j)(b_j/d_j). \end{cases}$$

Hence, Theorem 6 shows that asymptotically for $R > r$ and $r \to \infty$, if $\gamma_{min} = \gamma_j$, there

is an exponential density on the jth ray, and a uniform density in the regions between the $(j-1)$th and jth and the jth and $(j+1)$th rays. This is illustrated on Figure 5.4.1b for the inverted max-linear model given in Figure 5.1.1b. Note that γ_j is not necessarily unique, so it is possible that $\gamma_{min} = \gamma_j = \gamma_i$, for i and j distinct integers in $\{1, \ldots, k\}$. If this is the case then mass falls on both the ith and jth rays, and also in the regions on either side of these. For the alternative form W_D for the angular component, the mass collapses to $\{-\infty\}$, $\{0\}$ and $\{\infty\}$, as shown by Theorem 7. This is still the case, even when γ_j is not unique. Note, that even though the inverted max-linear model is AI, we find that there is mass on the diagonal $W_D = 0$ in the case when there exists $i \in \{1, \ldots, k\}$ such that $\alpha_i = \beta_i$. This is due to the fact that we defined the radial and angular components on exponential margins, which gives a different region $R > r$ than the more commonly used Fréchet margins. This illustrates one of the benefits of identifying extremal dependence structure using exponential marginal variables.

5.5 Mixture distribution

The transformed max-linear model (5.1.6) and the inverted max-linear model (5.1.7) can be combined into a mixture distribution

$$\begin{pmatrix} X_M \\ Y_M \end{pmatrix} = \begin{cases} \begin{pmatrix} X_E \\ Y_E \end{pmatrix} & \text{with probability } \delta \\ \begin{pmatrix} X_E^{(I)} \\ Y_E^{(I)} \end{pmatrix} & \text{with probability } 1-\delta \end{cases} \tag{5.5.1}$$

where $\delta \in [0, 1]$, and (X_E, Y_E) and $(X_E^{(I)}, Y_E^{(I)})$ represent a transformed max-linear model and an inverted max-linear model, respectively, on exponential margins. The statistical importance of the mixture model (5.5.1) is most easily seen by studying the sub-asymptotic

behaviour of χ defined by expression (5.1.1). Specifically let

$$\chi(z) = \Pr(Y > z \mid X > z),$$

so $\chi(z) \to \chi$ as $z \to \infty$. For the transformed max-linear model (5.1.6) it follows that $\chi = 2 - \sum_{i=1}^{m} \max(\alpha_i, \beta_i)$ and that for large z

$$\chi_E(z) \approx \chi + \frac{(2-\chi)(1-\chi)}{2} \exp(-z).$$

So here $\chi_E(z)$ converges to $\chi > 0$ at a fixed rate of decay. In contrast, for the inverted max-linear model (5.1.7) $\chi = 0$, but

$$\chi_{E^{(I)}}(z) = \{\exp(-z)\}^{\sum_{i=1}^{m} \max(\alpha_i, \beta_i) - 1}.$$

Here $\chi_{E^{(I)}}(z)$ converges to $\chi = 0$ at a rate of decay depending on the parameters of the underlying max-linear model, but there is no flexibility in the constant multiplier of this rate term. However for the mixture model (5.5.1) we have

$$\chi_M(z) \approx \delta\chi + (1-\delta)\{\exp(-z)\}^{1-\chi} + \delta\frac{(2-\chi)(1-\chi)}{2} \exp(-z),$$

where $\chi = 2 - \sum_{i=1}^{m} \max(\alpha_i, \beta_i)$. Thus, here there is AD, but also a penultimate behaviour that has flexibility in both its rate and coefficient features. Hence, although this mixture model is slightly artificial in its construction it has a sufficiently flexible form to be able to capture all natures of the leading and penultimate forms of extremal dependence.

We use results from Section 5.4 to deduce asymptotic properties of this mixture distribution. Here too, we will use the two different angular form representations W_D and W_I. Let X_M and Y_M be random variables, on exponential margins, from the mixture distribution (5.5.1), and let us define the radial and angular variables $R = X_M + Y_M$, $W_D = Y_M - X_M$, and $W_I = X_M/(X_M + Y_M)$. Then, using the angular form W_D, it follows from

Theorems 4 and 7 that, asymptotically for $R > r$ as $r \to \infty$, if there are no pairs (α_j, β_j) such that $\alpha_j = \beta_j$, $j = 1, \ldots, k$, then there is mass totalling δ on the lines $W_D = w_h$, $h = 1, \ldots, k$, and $(1 - \delta)$ mass either at $\{-\infty\}$ or at $\{+\infty\}$. If there is a pair (α_j, β_j) such that $\alpha_j = \beta_j$, then there is $(1 - \delta)(-b_{j-1}/d_j)$ mass at $\{+\infty\}$, $(1 - \delta)(-b_j/d_j)$ mass at $\{-\infty\}$, $\delta(\alpha_h \exp(w_h/2))/(\sum_{i=1}^{k} \alpha_i \exp(w_i/2))$ mass on each line $W_D = w_h$, $h \neq j$, and $\delta(\alpha_h \exp(w_h/2))/(\sum_{i=1}^{k} \alpha_i \exp(w_i/2)) + (1 - \delta)(a_j/d_j)$ mass on the diagonal $W_D = w_j = 0$. This is summarised in Theorem 8.

Theorem 8. *Let $R = X_M + Y_M$, $W_D = Y_M - X_M$, and $w_h = \log(\beta_h/\alpha_h)$, $h = 1, \ldots, k$. Then, for a_j, b_j, d_j and γ_j, $j = 1, \ldots, k$, defined as in Theorem 6, we have the following for $r \to \infty$,*

$$W_D \mid R > r \to \begin{cases} +\infty, & \text{with probability } (1 - \delta)(\mathbb{I}(\alpha_j > \beta_j) + \mathbb{I}(\alpha_j = \beta_j)(-b_{j-1}/d_j)), \\ w_h, & \text{with probability } \delta \left(\frac{\alpha_h \exp(w_h/2)}{\sum_{i=1}^{k} \alpha_i \exp(w_i/2)} \right) + (1 - \delta)\mathbb{I}(\alpha_j = \beta_j)(a_j/d_j), \\ & \qquad \text{for } h = 1, \ldots, k, \\ -\infty, & \text{with probability } (1 - \delta)(\mathbb{I}(\alpha_j < \beta_j) + \mathbb{I}(\alpha_j = \beta_j)(b_j/d_j)). \end{cases}$$

Using the second angular form W_I, it follows from Theorems 5 and 6 that asymptotically for $R > r$ as $r \to \infty$, if $\gamma_{min} = \gamma_j$ and $\alpha_j \neq \beta_j$, then there is mass totalling $(1-\delta)$ on the jth ray and the two regions adjacent to this ray, and δ mass on the diagonal ray $W_I = 1/2$. If $\alpha_j = \beta_j$, then there is $(1-\delta)a_j/d_j + \delta$ mass on the diagonal ray $W_I = 1/2$, and $(1-\delta)b_j/d_j$ and $-(1 - \delta)b_{j-1}/d_j$ mass in the two regions on either side of this ray, respectively. See Theorem 9 for details.

Theorem 9. *Let $R = X_M + Y_M$ and $W_I = X_M/(X_M + Y_M)$, and a_j, b_j, d_j and γ_j, $j = 1, \ldots, k$, defined as in Theorem 6. Then, if $\gamma_{min} = \gamma_j$, we have the following as $r \to \infty$,*

1. $\Pr(W_I = w_j \mid R > r) \to (1 - \delta)a_j/d_j + \delta\mathbb{I}(\alpha_j = \beta_j)$,
2. $\Pr(W_I = 1/2 \mid R > r) \to \delta + (1 - \delta)(a_j/d_j)\mathbb{I}(\alpha_j = \beta_j)$,

3. $\Pr(w_j < W_I < w_{j+1} \mid R > r) \to (1-\delta)b_j/d_j$,

4. $\Pr(w_{j-1} < W_I < w_j \mid R > r) \to -(1-\delta)b_{j-1}/d_j$,

5. $\Pr(\{0 < W_I < w_{j-1}\} \cup \{w_j < W_I < 1\} \mid R > r) \to 0$.

5.6 Proofs

5.6.1 Proof of Theorem 1

The probability $\Pr(R > r)$ can be written in the following way using total probability:

$$\begin{aligned}\Pr(R > r) &= \sum_{\mathcal{J}} \Pr(R > r \mid W_D \in \mathcal{J}) \Pr(W_D \in \mathcal{J}) \\ &= \Pr(R > r \mid W_D > w_k) \Pr(W_D > w_k) + \sum_{j=1}^{k} \Pr(R > r \mid W_D = w_j) \Pr(W_D = w_j) \\ &+ \sum_{j=1}^{k-1} \Pr(R > r \mid w_j < W_D < w_{j+1}) \Pr(w_j < W_D < w_{j+1}) \\ &+ \Pr(R > r \mid W_D < w_1) \Pr(W_D < w_1),\end{aligned}$$

where each of the product terms in this sum can be derived using results from Section 5.4.1. We will illustrate the derivation of the elements of the first term $\Pr(R > r \mid W_D > w_k) \Pr(W_D > w_k)$. First, note that $\Pr(W_D > w_k) \approx \Pr(Y_F > (\beta_k/\alpha_k)X_F)$ for large X_F and Y_F. Hence,

$$\Pr(W_D > w_k) \approx \frac{\alpha_k \beta_{sum}}{\alpha_k \beta_{sum} + \beta_k},$$

using results from Appendix 5.B.1. To obtain $\Pr(R > r \mid W_D > w_k)$ we first integrate expression (5.4.3) with respect to w:

$$\begin{aligned}f_R(r \mid W_D > w_k) &= \int_{w_k}^{r} f_{(R,W_D)}(r, w \mid W_D > w_k) dw, \\ &\approx \frac{\alpha_k \beta_{sum} + \beta_k}{2\alpha_k} e^{-r}(r - w_k).\end{aligned}$$

Then integrate this with respect to r to obtain the conditional probability:

$$\begin{aligned}\Pr(R > r \mid W_D > w_k) &= \int_r^\infty f_R(r \mid W_D > w_k)dr, \\ &\approx \frac{\alpha_k\beta_{sum} + \beta_k}{2\alpha_k}e^{-r}(r - w_k + 1).\end{aligned}$$

The other three product terms can be derived similarly, leading to

$$\begin{aligned}\Pr(R > r) \approx{}& \frac{\beta_{sum}}{2}(r - w_k + 1)\exp(-r) + \sum_{j=1}^{k}\frac{\alpha_j}{2}\exp\left(\frac{w_j}{2}\right)\exp\left(-\frac{r}{2}\right) \\ &+ \sum_{j=1}^{k-1}\frac{\alpha_{sum}^{(j)}\beta_{sum}^{(j+1)}}{2}(w_{j+1} - w_j)\exp(-r) + \frac{\alpha_{sum}}{2}(r + w_1 + 1)\exp(-r), \text{ for large } r.\end{aligned}$$

For $r \to \infty$ we can write

$$\Pr(R > r) \sim \exp\left(-\frac{r}{2}\right)\sum_{j=1}^{k}\frac{\alpha_j}{2}\exp\left(\frac{w_j}{2}\right),$$

since the other terms all contain $\exp(-r)$ and they go to zero faster as $r \to \infty$.

For the region $W_D > w_k$, substituting into (5.4.2), we get for large r

$$\Pr(W_D > w_k \mid R > r) \approx \frac{\beta_{sum}(r - w_k + 1)\exp(-r)}{\sum_{j=1}^{k}\alpha_j\exp(w_j/2)\exp(-r/2)}.$$

Since the numerator tends to zero faster than the denominator, $\Pr(W_D > w_k \mid R > r) \to 0$ as $r \to \infty$. This is the same for the region $W_D < w_1$ and the regions $w_j < W_D < w_{j+1}$ for all $j = 1, \ldots, k-1$, i.e., $\Pr(W_D < w_1 \mid R > r) \to 0$ and $\Pr(w_j < W_D < w_{j+1} \mid R > r) \to 0$ for all $j = 1, \ldots, k-1$ as $r \to \infty$.

For the case when $W_D = w_j$, $j = 1, \ldots, k$, both the numerator and denominator have exponent term $\exp(-r/2)$, hence,

$$\Pr(W_D = w_j \mid R > r) \to \frac{\alpha_j\exp(w_j/2)}{\sum_{i=1}^{k}\alpha_i\exp(w_i/2)}, \text{ as } r \to \infty.$$

□

5.6.2 Proof of Theorem 2

It follows from Theorem 4 that, asymptotically, W_D is on one of the k lines, i.e., on exponential margins, $Y_E = -\log(-1-(1-\exp(-X_E))^{1/\omega_j}) \approx X_E + \log(\omega_j)$, $j = 1, \ldots, k$, for large X_E. Hence, for large R,

$$W_I \approx \frac{1}{2} - \frac{\log(\omega_j)}{2R},\ j = 1, \ldots, k.$$

Hence, using Theorem 4, it follows that $W_I \mid R > r \to^p 1/2$, $j = 1, \ldots, k$, as $r \to \infty$. □

5.6.3 Proof of Theorem 3

The probability $\Pr(W_I \in \mathcal{J} \mid R > r)$ is equivalent to the expression given in (5.4.2) with $W = W_I$. The joint probability in the numerator can be calculated in each case using methods similar to those in previous sections (see e.g. equation (5.4.4) for the case when $W_I > w_k$). Then we can use total probability to calculate $\Pr(R > r)$ as shown below:

$$\begin{aligned}
\Pr(R > r) &= \sum_{\mathcal{J}} \Pr(R > r \mid W_I \in \mathcal{J}) \Pr(W_I \in \mathcal{J}) \\
&= \Pr(R > r \mid W_I > w_k) \Pr(W_I > w_k) + \sum_{j=1}^{k} \Pr(R > r \mid W_I = w_j) \Pr(W_I = w_j) \\
&+ \sum_{j=1}^{k-1} \Pr(R > r \mid w_j < W_I < w_{j+1}) \Pr(w_j < W_I < w_{j+1}) \\
&+ \Pr(R > r \mid W_I < w_1) \Pr(W_I < w_1).
\end{aligned}$$

Then,

$$\begin{aligned}\Pr(R > r) = &\frac{\beta_{sum}}{1-\beta_{sum}}\left[\gamma_k^{-1}\exp\left(-\gamma_k r\right) - \exp(-r)\right] + \sum_{j=1}^{k}\frac{\alpha_j\beta_j}{\alpha_j+\beta_j}\gamma_j^{-1}\exp\left(-\gamma_j r\right) \\ &+ \sum_{j=1}^{k-1}\frac{\alpha_{sum}^{(j)}\beta_{sum}^{(j+1)}}{\alpha_{sum}^{(j)}-\beta_{sum}^{(j+1)}}\left[\gamma_j^{-1}\exp(-\gamma_j r) - \gamma_{j+1}^{-1}\exp(-\gamma_{j+1} r)\right] \\ &+ \frac{\alpha_{sum}}{1-\alpha_{sum}}\left[\gamma_1^{-1}\exp\left(-\gamma_1 r\right) - \exp(-r)\right],\end{aligned}$$

where $\gamma_j = (\alpha_j\beta_{sum}^{(j+1)} + \beta_j\alpha_{sum}^{(j)})/(\alpha_j+\beta_j)$, $0 < \gamma_j \leqslant 1$, for $j = 1,\ldots,k$. For large r, $\Pr(R > r)$ approximately becomes

$$\begin{aligned}\Pr(R > r) \approx &\frac{\beta_{sum}}{1-\beta_{sum}}\gamma_k^{-1}\exp\left(-\gamma_k r\right) + \sum_{j=1}^{k}\frac{\alpha_j\beta_j}{\alpha_j+\beta_j}\gamma_j^{-1}\exp\left(-\gamma_j r\right) \\ &+ \sum_{j=1}^{k-1}\frac{\alpha_{sum}^{(j)}\beta_{sum}^{(j+1)}}{\alpha_{sum}^{(j)}-\beta_{sum}^{(j+1)}}\left[\gamma_j^{-1}\exp(-\gamma_j r) - \gamma_{j+1}^{-1}\exp(-\gamma_{j+1} r)\right] \\ &+ \frac{\alpha_{sum}}{1-\alpha_{sum}}\gamma_1^{-1}\exp\left(-\gamma_1 r\right), \qquad (5.6.1)\end{aligned}$$

since the terms containing $\exp(-r)$ are smaller than the terms containing the exponential terms $\exp(-\gamma_j r)$ for all $j = 1,\ldots,k-1$. Note that using the notation defined in Theorem 6, we can write (5.6.1) as

$$\Pr(R > r) \approx \sum_{j=1}^{k}(a_j + b_j - b_{j-1})\gamma_j^{-1}e^{-\gamma_j r}. \qquad (5.6.2)$$

For each region the probability in (5.4.2) will tend to zero as $r \to \infty$, unless the exponent term in the numerator is the same as the largest exponent term in expression (5.6.1). Hence, the result in Theorem 6 follows. □

5.6.4 Proof of Theorem 4

For the inverted max-linear distribution on exponential margins, $Y_E^{(I)} = X_E^{(I)}/\omega_j$, $j = 1,\ldots,k$, where $0 < \omega_j < \infty$. On (R, W_D) margins this becomes, $W_D = R\delta_j$, where $\delta_j =$

$(1-\omega_j)/(1+\omega_j)$, $j=1,\ldots,k$. Hence, as $r\to\infty$,

$$W_D \mid R>r \to \begin{cases} \infty, & \text{for } \omega_j<1, \\ 0, & \text{for } \omega_j=1, \\ -\infty, & \text{for } \omega_j>1. \end{cases}$$

To work out the mass at $\{-\infty\}$, $\{0\}$ and $\{\infty\}$ we calculate the probabilities $\Pr(W_D<0 \mid R>r)$, $\Pr(W_D=0 \mid R>r)$ and $\Pr(W_D>0 \mid R>r)$. By conditional probability,

$$\Pr(W_D<0 \mid R>r) = \frac{\Pr(R>r \mid W_D<0)\Pr(W_D<0)}{\Pr(R>r)} \tag{5.6.3}$$

where $\Pr(R>r)$ is the same as in expression (5.6.2), and $\Pr(W_D<0)$ can be easily obtained as $\Pr(W_D<0)=\Pr(X_F<Y_F)=\sum_{i=1}^{k}\mathbb{I}(\alpha_i=\beta_i)\beta_{sum}^{(j+1)}/(\alpha_{sum}^{(j)}+\beta_{sum}^{(j+1)})$. Then we calculate the conditional probability $\Pr(R>r \mid W_D<0)$ as the following sum,

$$\begin{aligned}\Pr(R>r \mid W_D<0) &= \Pr(R>r \mid W_D<0, W_D<R\delta_k)\Pr(W_D<R\delta_k \mid W_D<0) \\ &\quad + \sum_{j=i+1}^{k}\Pr(R>r \mid W_D<0, W_D=R\delta_j)\Pr(W_D=R\delta_j \mid W_D<0) \\ &\quad + \sum_{j=i}^{k-1}\Pr(R>r \mid W_D<0, R\delta_{j+1}<W_D<R\delta_j) \\ &\quad \times \Pr(R\delta_{j+1}<W_D<R\delta_j \mid W_D<0),\end{aligned}$$

where we assumed that there is an $i\in\{1,\ldots,k\}$ such that $\omega_i=1$. Using results from Section 5.3.2, we have

$$\begin{aligned}\Pr(R>r, W_D<0) &\approx b_k\gamma_k^{-1}e^{-\gamma_k r} + \sum_{j=i+1}^{k} a_j\gamma_j^{-1}e^{-\gamma_j r} + \sum_{j=i}^{k-1} b_j\left[\gamma_j^{-1}e^{-\gamma_j r} - \gamma_{j+1}^{-1}e^{-\gamma_{j+1} r}\right], \\ &= b_i\gamma_i^{-1}e^{-\gamma_i r} + \sum_{j=i+1}^{k}(a_j+b_j-b_{j-1})\gamma_j^{-1}e^{-\gamma_j r},\end{aligned}$$

where a_j, b_j and γ_j, $j = 1, \ldots, k$, are defined as in Theorem 6. The probabilities $\Pr(W_D = 0 \mid R > r)$ and $\Pr(W_D > 0 \mid R > r)$ can be calculated similarly. Substituting into expression (5.6.3) we obtain the results in Theorem 7. □

5.7 Conclusions

In this chapter we have characterised the asymptotic behaviour of models built on bivariate max-linearity, using two different angular measures defined in exponential marginal space. We found that the limiting behaviour of our three models (transformed max-linear, inverted max-linear and mixture) can be either asymptotically dependent or asymptotically independent. At finite levels, however, they feature points on rays of the form $y = hx$, $0 < h < \infty$, points on lines of the form $y = h + x$, $-\infty < h < \infty$, and independent points scattered in the regions defined by these rays and lines.

Simulation from the max-linear model (5.1.3) is straightforward by sampling Z_j, $j = 1, \ldots, m$, independently from a Fréchet distribution and simply calculating X_F and Y_F, subject to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ values being known. Then we can transform to margins (X_E, Y_E) or $(X_E^{(I)}, Y_E^{(I)})$ to obtain samples on exponential margins from the transformed max-linear or the inverted max-linear models. Assuming δ is also known, we can easily sample from the mixture distribution (5.5.1), by sampling from the transformed max-linear model with probability δ and the inverted max-linear model with probability $1 - \delta$. Simulation from the conditional distribution $Y_F \mid X_F$ is also straightforward if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are known, using methods described in Wang and Stoev (2011), so conditional simulation follows easily for our three models. For a detailed description of the simulation algorithm the reader is referred to Chapter 6.

This chapter has been restricted to bivariate models, but the formulation is straightforward to extend to multivariate cases. Specifically, consider a d-dimensional max-linear model,

with Fréchet margins, where

$$X_{F,j} = \max_{i=1,\ldots,m} (\alpha_{ij} Z_i) \text{ for } j = 1, \ldots, d,$$

where $Z_1, \ldots, Z_m$ are independent and identically distributed Fréchet variables and $\alpha_{ij} \geqslant 0$ with $\sum_{i=1}^{m} \alpha_{ij} = 1$ for all $j = 1, \ldots, d$. The multivariate transformed max-linear and inverted max-linear models follow using multivariate analogues of transformations (5.1.4) and (5.1.5), respectively. The extreme values from these joint distributions can be studied by multivariate extensions of our radial and angular transformations, in particular using

$$R = \sum_{j=1}^{d} X_j, \quad \mathbf{W}_D = (X_2 - X_1, \ldots, X_d - X_1), \quad \mathbf{W}_I = (X_2/R, \ldots, X_d/R),$$

where $(X_1, \ldots, X_d)$ are on exponential margins. We expect to obtain similar findings to the bivariate case with a range of asymptotic independence and asymptotic dependence over different subsets of the variables. Similarly, both the joint and conditional simulation algorithms can be easily extended to the multivariate case.

Our work has potential to be useful in spatial applications of extreme value theory since there is often need to model bivariate dependence for both local dependence and long-range dependence in this setting. Generally, extreme events at locations close by are expected to occur simultaneously, as they are likely to be affected by the same underlying process. Hence, it seems natural to model these as asymptotically dependent. On the other hand, extreme events at locations far apart are unlikely to occur together as the chance of both locations being affected by the same event is reduced; thus asymptotically independent models seem more appropriate in this case. In practice, it is necessary to have a model that can move through the two types and different levels of extremal dependence; e.g., to model sites close by as asymptotically dependent, with dependence decreasing with distance, and asymptotic independence for locations further apart. To achieve a smooth transition between the two types of dependence we need a model that has both components. This chapter introduced

a new model that incorporates both types of dependence, and that easily lends itself to spatial applications, since we can allow the mass on the rays and lines to vary smoothly with distance or some measure of the strength of dependence between locations. One way to do this could be to have both α_i and β_i $(i = 1, \ldots, m)$ decay with distance at different rates. The development of such models and their statistical inference is the topic for further work.

5.A Derivation of conditional densities for the Marshall-Olkin model

5.A.1 Case (ii) - Below the line with $Y_F < \frac{\beta}{\alpha} X_F$

In this case, we want to work out the distribution of the points in the region below the line. A similar approach to the one described in Section 5.2 can be taken, but here we need to consider two possible combinations of Z_i's that can give $Y_F < \frac{\beta}{\alpha} X_F$. So, by conditional probability,

$$\Pr\left\{X_F > x, Y_F > y \middle| Y_F < \frac{\beta}{\alpha} X_F\right\} = \frac{\Pr\left\{X_F > x, Y_F > y, Y_F < \frac{\beta}{\alpha} X_F\right\}}{\Pr\left\{Y_F < \frac{\beta}{\alpha} X_F\right\}}.$$

As noted before, there are two possible combinations that lead to this case: $(X_F, Y_F) = ((1-\alpha)Z_2, \beta Z_1)$, and $(X_F, Y_F) = ((1-\alpha)Z_2, (1-\beta)Z_3)$ with $(1-\alpha)Z_2/\alpha > (1-\beta)Z_3/\beta$. So, the joint probability can be broken down into the sum of the two cases, such that

$$\Pr\left\{X_F > x, Y_F > y, Y_F < \frac{\beta}{\alpha} X_F\right\} = P_1 + P_2,$$

where

$$P_1 = \Pr\{X_F > x, Y_F > y, X_F = (1-\alpha)Z_2, Y_F = \beta Z_1\},$$

$$P_2 = \Pr\{X_F > x, Y_F > y, X_F = (1-\alpha)Z_2, Y_F = (1-\beta)Z_3\}.$$

Then,

$$\begin{aligned}
P_1 &= \Pr\{(1-\alpha)Z_2 > x, \beta Z_1 > y, (1-\alpha)Z_2 > \alpha Z_1, \beta Z_1 > (1-\beta)Z_3\} \\
&= \Pr\left\{Z_2 > \frac{x}{1-\alpha}, Z_1 > \frac{y}{\beta}, Z_2 > \frac{\alpha Z_1}{1-\alpha}, Z_3 < \frac{\beta Z_1}{1-\beta}\right\} \\
&= \int_{y/\beta}^{\infty} \Pr\left\{Z_2 > \max\left(\frac{x}{1-\alpha}, \frac{\alpha z}{1-\alpha}\right), Z_3 < \frac{\beta Z_1}{1-\beta}\middle| Z_1 = z\right\} f_Z(z)dz \\
&= \mathbb{I}\left(y < \frac{\beta}{\alpha}x\right)\left[\beta - \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} - \beta e^{-1/y} - \left(\beta - \frac{\alpha\beta}{\alpha+\beta-\alpha\beta}\right)e^{-(\alpha+\beta-\alpha\beta)/(\beta x)}\right. \\
&\quad \left. + \beta e^{-(1-\alpha)/x}e^{-1/y}\right],
\end{aligned}$$

and

$$\begin{aligned}
P_2 &= \Pr\{(1-\alpha)Z_2 > x, (1-\beta)Z_3 > y, (1-\alpha)Z_2 > \alpha Z_1, (1-\beta)Z_3 > \beta Z_1, \\
&\quad \beta(1-\alpha)Z_2 > \alpha(1-\beta)Z_3\} \\
&= \mathbb{I}\left(y < \frac{\beta}{\alpha}x\right)\left[\frac{\beta(1-\alpha)(1-\beta)}{\alpha+\beta-\alpha\beta} - (1-\beta)e^{-1/y} - \frac{\beta(1-\alpha)(1-\beta)}{\alpha+\beta-\alpha\beta}e^{-(\alpha+\beta-\alpha\beta)/(\beta x)}\right. \\
&\quad \left. +(1-\beta)e^{-(1-\alpha)/x}e^{-1/y}\right],
\end{aligned}$$

where the last equalities in the derivations of P_1 and P_2 follow after extensive calculations. Summing P_1 and P_2 we get,

$$\begin{aligned}
&\Pr\left\{X_F > x, Y_F > y, Y_F < \frac{\beta}{\alpha}X_F\right\} = \\
&\quad \mathbb{I}\left(y < \frac{\beta}{\alpha}x\right)\left[\frac{(1-\alpha)\beta}{\alpha+\beta-\alpha\beta} - e^{-1/y} - \frac{(1-\alpha)\beta}{\alpha+\beta-\alpha\beta}e^{-(\alpha+\beta-\alpha\beta)/(\beta x)} + e^{-(1-\alpha)/x}e^{-1/y}\right].
\end{aligned}$$

Similarly for the denominator,

$$\Pr\left\{Y_F < \frac{\beta}{\alpha}X_F\right\} = \tilde{P}_1 + \tilde{P}_2,$$

where

$$\tilde{P}_1 = \Pr\{X_F = (1-\alpha)Z_2, Y_F = \beta Z_1\} = \beta - \frac{\alpha\beta}{\alpha+\beta-\alpha\beta},$$
$$\tilde{P}_2 = \Pr\{X_F = (1-\alpha)Z_2, Y_F = (1-\beta)Z_3\} = \frac{\beta(1-\alpha)(1-\beta)}{\alpha+\beta-\alpha\beta}.$$

Hence,

$$\Pr\left\{Y_F < \frac{\beta}{\alpha}X_F\right\} = \frac{(1-\alpha)\beta}{\alpha+\beta-\alpha\beta}.$$

Substituting into the conditional probability formula we finally obtain the joint survival function conditional on being in the region below the line, as

$$\Pr\left\{X_F > x, Y_F > y \middle| Y_F < \frac{\beta}{\alpha}X_F\right\} =$$
$$\mathbb{I}\left(y < \frac{\beta}{\alpha}x\right)\left[1 - \frac{\alpha+\beta-\alpha\beta}{(1-\alpha)\beta}e^{-1/y} - e^{-(\alpha+\beta-\alpha\beta)/(\beta x)} + \frac{\alpha+\beta-\alpha\beta}{(1-\alpha)\beta}e^{-(1-\alpha)/x}e^{-1/y}\right].$$

Differentiating, we obtain the conditional density in the region below the line as:

$$f_{(X_F,Y_F)}\left(x, y \middle| Y_F < \frac{\beta}{\alpha}X_F\right) = \frac{\alpha+\beta-\alpha\beta}{\beta x^2 y^2}e^{-(1-\alpha)/x}e^{-1/y}\ \mathbb{I}\left(y < \frac{\beta}{\alpha}x\right).$$

5.A.2 Case (iii) - Above the line with $Y_F > \frac{\beta}{\alpha}X_F$

In this case, we want to work out the distribution of the points conditional on being in the region above the line. The calculations in this case are very similar to those in Section 5.A.1

so we will give less detail. Again, by conditional probability,

$$\Pr\left\{X_F > x, Y_F > y \middle| Y_F > \frac{\beta}{\alpha}X_F\right\} = \frac{\Pr\left\{X_F > x, Y_F > y, Y_F > \frac{\beta}{\alpha}X_F\right\}}{\Pr\left\{Y_F > \frac{\beta}{\alpha}X_F\right\}},$$

and there are two possible combinations that lead to this case: $(X_F, Y_F) = (\alpha Z_1, (1-\beta)Z_3)$ and $(X_F, Y_F) = ((1-\alpha)Z_2, (1-\beta)Z_3)$ with $(1-\alpha)Z_2/\alpha < (1-\beta)Z_3/\beta$.

Similarly to Section 5.A.1, we can calculate the following marginal and joint probabilities:

$$\Pr\left\{Y_F > \frac{\beta}{\alpha}X_F\right\} = \frac{\alpha(1-\beta)}{\alpha+\beta-\alpha\beta},$$

and

$$\Pr\left\{X_F > x, Y_F > y, Y_F > \frac{\beta}{\alpha}X_F\right\} = \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\left[\frac{\alpha(1-\beta)}{\alpha+\beta-\alpha\beta} - e^{-1/x} - \frac{\alpha(1-\beta)}{\alpha+\beta-\alpha\beta}e^{-(\alpha+\beta-\alpha\beta)/(\alpha y)} + e^{-1/x}e^{-(1-\beta)/y}\right].$$

Substituting into the conditional probability formula we obtain the conditional distribution:

$$\Pr\left\{X_F > x, Y_F > y \middle| Y_F > \frac{\beta}{\alpha}X_F\right\} = \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right)\left[1 - \frac{\alpha+\beta-\alpha\beta}{\alpha(1-\beta)}e^{-1/x} - e^{-(\alpha+\beta-\alpha\beta)/(\alpha y)} + \frac{\alpha+\beta-\alpha\beta}{\alpha(1-\beta)}e^{-1/x}e^{-(1-\beta)/y}\right].$$

Differentiating, we obtain the conditional density of the points in the region above the line as

$$f_{(X_F,Y_F)}\left(x, y \middle| Y_F > \frac{\beta}{\alpha}X_F\right) = \frac{\alpha+\beta-\alpha\beta}{\alpha x^2 y^2}e^{-1/x}e^{-(1-\beta)/y}\ \mathbb{I}\left(y > \frac{\beta}{\alpha}x\right).$$

5.B Derivation of density formulas in the general case

5.B.1 Type (i) - Above the line $Y_F = \omega_k X_F$

From the condition that $Y_F > \omega_k X_F$, it can be established that the pairs that can lead to this case are combinations of the following: $X_F = \alpha_i Z_i$ where $i = 1, \ldots, k+l$, and $Y_F = \beta_h Z_h$ where $h = k+l+1, \ldots, m$. Hence,

$$\Pr\left(X_F \in dx, Y_F \in dy, Y_F > \omega_k X_F\right) = \sum_{h=k+l+1}^{m} \sum_{i=1}^{k+l} \Pr\left(\alpha_i Z_i \in dx, \beta_h Z_h \in dy, Y_F > \omega_k X_F\right).$$

The Z_i's are independent Fréchet random variables, hence,

$$\begin{aligned}
&\Pr\left(X_F \in dx, Y_F \in dy, Y_F > \omega_k X_F\right) \\
&= \mathbb{I}\left(y > \omega_k x\right) \sum_{h=k+l+1}^{m} \sum_{i=1}^{k+l} \left[\Pr\left(Z_i \in \frac{dx}{\alpha_i}\right) \Pr\left(Z_h \in \frac{dy}{\beta_h}\right) \right. \\
&\times \prod_{p=1,\{p\neq i\}}^{k} \Pr\left(Z_p < \min\left(\frac{x}{\alpha_p}, \frac{y}{\beta_p}\right)\right) \prod_{p=k+1,\{p\neq i\}}^{k+l} \Pr\left(Z_p < \frac{x}{\alpha_p}\right) \\
&\left. \times \prod_{p=k+l+1,\{p\neq h\}}^{m} \Pr\left(Z_p < \frac{y}{\beta_p}\right) \right] dxdy \\
&= \mathbb{I}\left(y > \omega_k x\right) \sum_{h=k+l+1}^{m} \sum_{i=1}^{k+l} \left[\frac{\alpha_i}{x^2} e^{-\alpha_i/x} \frac{\beta_h}{y^2} e^{-\beta_h/y} \prod_{p=1,\{p\neq i\}}^{k+l} \left(e^{-\alpha_p/x}\right) \right. \\
&\left. \times \prod_{p=k+l+1,\{p\neq h\}}^{m} \left(e^{-\beta_p/y}\right) \right] dxdy \\
&= \mathbb{I}\left(y > \omega_k x\right) \sum_{h=k+l+1}^{m} \sum_{i=1}^{k+l} \left[\frac{\alpha_i \beta_h}{x^2 y^2} \prod_{p=1}^{k+l} \left(e^{-\alpha_p/x}\right) \prod_{p=k+l+1}^{m} \left(e^{-\beta_p/y}\right) \right] dxdy \\
&= \mathbb{I}\left(y > \omega_k x\right) \frac{\beta_{sum}}{x^2 y^2} e^{-1/x} e^{-\beta_{sum}/y} dxdy,
\end{aligned} \tag{5.B.1}$$

where the third equality follows since $x/\alpha_p < y/\beta_p$, $\forall p = 1, \ldots, k$. The marginal can be obtained simply by integrating (5.B.1) over the region

$$\begin{aligned}\Pr\left(Y_F > \omega_k X_F\right) &= \int_0^\infty \int_{\omega_k x}^\infty \left(\frac{\beta_{sum}}{x^2 y^2} e^{-1/x} e^{-\beta_{sum}/y}\right) dy dx \\ &= \frac{\alpha_k \beta_{sum}}{\beta_k + \alpha_k \beta_{sum}}.\end{aligned}$$

Hence, the conditional density is

$$\begin{aligned}f_{(X_F,Y_F)}(x, y \mid Y_F > \omega_k X_F) &= \frac{\Pr\left(X_F \in dx, Y_F \in dy, Y_F > \omega_k X_F\right)}{\Pr\left(Y_F > \omega_k X_F\right)} \\ &= \mathbb{I}\left(y > \omega_k x\right) \frac{\beta_k + \alpha_k \beta_{sum}}{\alpha_k x^2 y^2} e^{-1/x} e^{-\beta_{sum}/y}.\end{aligned}$$

5.B.2 Type (ii) - On the line $Y_F = \omega_j X_F$, $j = 1, \ldots, k$

This case only occurs if $X_F = \alpha_j Z_j$ and $Y_F = \beta_j Z_j$. Hence,

$$\Pr\left(X_F \in dx, Y_F \in dy, Y_F = \omega_j X_F\right) = \Pr\left(\alpha_j Z_j \in dx, \beta_j Z_j \in dy, Y_F = \omega_j X_F\right).$$

Similarly to the approach in Section 5.B.1, this can be written as

$$\begin{aligned}&\Pr\left(X_F \in dx, Y_F \in dy, Y_F = \omega_j X_F\right) \\ &= \mathbb{I}\left(y = \omega_j x\right) \Pr\left(Z_j \in \frac{dx}{\alpha_j}\right) \prod_{p=1}^{j-1} \Pr\left(Z_p < \frac{x}{\alpha_p}\right) \prod_{p=j+1}^{k} \Pr\left(Z_p < \frac{y}{\beta_p}\right) \\ &\times \prod_{p=k+1}^{k+l} \Pr\left(Z_p < \frac{x}{\alpha_p}\right) \prod_{p=k+l+1}^{m} \Pr\left(Z_p < \frac{y}{\beta_p}\right) dx dy \\ &= \mathbb{I}\left(y = \omega_j x\right) \frac{\alpha_j}{x^2} e^{-\alpha_j/x} \prod_{p=1}^{j-1} \left(e^{-\alpha_p/x}\right) \prod_{p=j+1}^{k} \left(e^{-\beta_p/y}\right) \prod_{p=k+1}^{k+l} \left(e^{-\alpha_p/x}\right) \\ &\times \prod_{p=k+l+1}^{m} \left(e^{-\beta_p/y}\right) dx dy \\ &= \mathbb{I}\left(y = \omega_j x\right) \frac{\alpha_j}{x^2} \exp\left\{-\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j x}\right\} dx dy.\end{aligned}$$

By integrating this from 0 to ∞ we can obtain the marginal as

$$\Pr\left(Y_F = \omega_j X_F\right) = \frac{\alpha_j \beta_j}{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}.$$

Hence, the conditional density is

$$f_{(X_F,Y_F)}(x, y \mid Y_F = \omega_j X_F) = \mathbb{I}\left(y = \omega_j x\right) \frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j x^2} \exp\left\{-\frac{\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)}}{\beta_j x}\right\}.$$

5.B.3 Type (iii) - Between the two lines $Y_F = \omega_j X_F$ and $Y_F = \omega_{j+1} X_F$, $j = 1, \ldots, k-1$

From the condition that $\omega_j X_F < Y_F < \omega_{j+1} X_F$, it follows that the pairs that lead to this case are combinations of $X_F = \alpha_i Z_i$ where $i = 1, \ldots, j, k+1, \ldots, k+l$, and $Y_F = \beta_h Z_h$ where $h = j+1, \ldots, m$. Hence,

$$\begin{aligned}&\Pr\left(X_F \in dx, Y_F \in dy, \omega_j X_F < Y_F < \omega_{j+1} X_F\right)\\ &\qquad = \sum_{h \in \mathcal{H}} \sum_{i \in \mathcal{I}} \Pr\left(\alpha_i Z_i \in dx, \beta_h Z_h \in dy, \omega_j X_F < Y_F < \omega_{j+1} X_F\right),\end{aligned}$$

where $\mathcal{I} = \{1, \ldots, j, k+1, \ldots, k+l\}$ and $\mathcal{H} = \{j+1, \ldots, m\}$. Then, due to the independence of the Z_i's this can be written as,

$$\begin{aligned}&\Pr\left(X_F \in dx, Y_F \in dy, \omega_j X_F < Y_F < \omega_{j+1} X_F\right)\\ &= \mathbb{I}\left(\omega_j x < y < \omega_{j+1} x\right) \sum_{h \in \mathcal{H}} \sum_{i \in \mathcal{I}} \Bigg[\Pr\left(Z_i \in \frac{dx}{\alpha_i}\right) \Pr\left(Z_h \in \frac{dy}{\beta_h}\right) \prod_{p=1, \{p \neq i\}}^{j} \Pr\left(Z_p < \frac{x}{\alpha_p}\right)\\ &\times \prod_{p=j+1, \{p \neq h\}}^{k} \Pr\left(Z_p < \frac{y}{\beta_p}\right) \prod_{p=k+1, \{p \neq i\}}^{k+l} \Pr\left(Z_p < \frac{x}{\alpha_p}\right)\\ &\times \prod_{p=k+l+1, \{p \neq h\}}^{m} \Pr\left(Z_p < \frac{y}{\beta_p}\right)\Bigg] dx dy.\end{aligned}$$

Then,

$$\begin{aligned}
&\Pr\left(X_F \in dx, Y_F \in dy, \omega_j X_F < Y_F < \omega_{j+1} X_F\right) \\
&= \mathbb{I}\left(\omega_j x < y < \omega_{j+1} x\right) \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}} \left[\frac{\alpha_i}{x^2} e^{-\alpha_i/x} \frac{\beta_h}{y^2} e^{-\beta_h/y} \prod_{p=1,\{p\neq i\}}^{j} \left(e^{-\alpha_p/x}\right) \right. \\
&\times \left. \prod_{p=j+1,\{p\neq h\}}^{k} \left(e^{-\beta_p/y}\right) \prod_{p=k+1,\{p\neq i\}}^{k+l} \left(e^{-\alpha_p/x}\right) \prod_{p=k+l+1,\{p\neq h\}}^{m} \left(e^{-\beta_p/y}\right) \right] dxdy \\
&= \mathbb{I}\left(\omega_j x < y < \omega_{j+1} x\right) \frac{\alpha_{sum}^{(j)} \beta_{sum}^{(j+1)}}{x^2 y^2} e^{-\alpha_{sum}^{(j)}/x} e^{-\beta_{sum}^{(j+1)}/y} dxdy.
\end{aligned}$$

Integrating this over the range $0 < x < \infty$ and $\omega_j x < y < \omega_{j+1} x$, we obtain the marginal as

$$\Pr\left(\omega_j X_F < Y_F < \omega_{j+1} X_F\right) = \frac{\alpha_{sum}^{(j)} \beta_{sum}^{(j+1)} (\alpha_j \beta_{j+1} - \alpha_{j+1} \beta_j)}{(\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})(\alpha_{j+1} \beta_{sum}^{(j+1)} + \beta_{j+1} \alpha_{sum}^{(j)})}.$$

Using conditional probability, the conditional density is,

$$\begin{aligned}
&f_{(X_F,Y_F)}(x, y \mid \omega_j X_F < Y_F < \omega_{j+1} X_F) \\
&\qquad = \mathbb{I}\left(\omega_j x < y < \omega_{j+1} x\right) \left(\frac{(\alpha_j \beta_{sum}^{(j+1)} + \beta_j \alpha_{sum}^{(j)})(\alpha_{j+1} \beta_{sum}^{(j+1)} + \beta_{j+1} \alpha_{sum}^{(j)})}{(\alpha_j \beta_{j+1} - \alpha_{j+1} \beta_j) x^2 y^2} \right) \\
&\qquad \times e^{-\alpha_{sum}^{(j)}/x} e^{-\beta_{sum}^{(j+1)}/y}.
\end{aligned}$$

5.B.4 Type (iv) - Below the line $Y_F = \omega_1 X_F$

From the condition that $Y_F < \omega_1 X_F$, it can be established that the pairs that can lead to this case are combinations of $X_F = \alpha_i Z_i$ and $Y_F = \beta_h Z_h$ where $i = k+1, \ldots, k+l$ and $h = 1, \ldots, k, k+l+1, \ldots, m$. Hence,

$$\Pr\left(X_F \in dx, Y_F \in dy, Y_F < \omega_1 X_F\right) = \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}} \Pr\left(\alpha_i Z_i \in dx, \beta_h Z_h \in dy, Y_F < \omega_1 X_F\right),$$

where $\mathcal{I} = \{k+1, \ldots, k+l\}$ and $\mathcal{H} = \{1, \ldots, k, k+l+1, \ldots, m\}$. The Z_i's are independent Fréchet random variables, hence

$$\begin{aligned}
&\Pr\left(X_F \in dx, Y_F \in dy, Y_F < \omega_1 X_F\right) \\
&\quad = \mathbb{I}\left(y < \omega_1 x\right) \sum_{h \in \mathcal{H}} \sum_{i \in \mathcal{I}} \left[\Pr\left(Z_i \in \frac{dx}{\alpha_i}\right) \Pr\left(Z_h \in \frac{dy}{\beta_h}\right) \prod_{p=1,\{p \neq h\}}^{k} \Pr\left(Z_p < \frac{y}{\beta_p}\right) \right. \\
&\quad \left. \times \prod_{p=k+1,\{p \neq i\}}^{k+l} \Pr\left(Z_p < \frac{x}{\alpha_p}\right) \prod_{p=k+l+1,\{p \neq h\}}^{m} \Pr\left(Z_p < \frac{y}{\beta_p}\right) \right] dxdy \\
&\quad = \mathbb{I}\left(y < \omega_1 x\right) \sum_{h \in \mathcal{H}} \sum_{i \in \mathcal{I}} \left[\frac{\alpha_i}{x^2} e^{-\alpha_i/x} \frac{\beta_h}{y^2} e^{-\beta_h/y} \prod_{p=1,\{p \neq h\}}^{k} \left(e^{-\beta_p/y}\right) \prod_{p=k+1,\{p \neq i\}}^{k+l} \left(e^{-\alpha_p/x}\right) \right. \\
&\quad \left. \times \prod_{p=k+l+1,\{p \neq h\}}^{m} \left(e^{-\beta_p/y}\right) \right] dxdy \\
&\quad = \mathbb{I}\left(y < \omega_1 x\right) \frac{\alpha_{sum}}{x^2 y^2} e^{-\alpha_{sum}/x} e^{-1/y} dxdy.
\end{aligned}$$

The marginal can be obtained by integrating this over the region.

$$\Pr\left(Y_F < \omega_1 X_F\right) = \frac{\alpha_{sum}\beta_1}{\alpha_{sum}\beta_1 + \alpha_1}.$$

Hence, the conditional density is

$$f_{(X_F, Y_F)}(x, y \mid Y_F < \omega_1 X_F) = \mathbb{I}\left(y < \omega_1 x\right) \frac{\alpha_{sum}\beta_1 + \alpha_1}{\beta_1 x^2 y^2} e^{-\alpha_{sum}/x} e^{-1/y}.$$

Chapter 6

Conditional Simulation of Models Built on Max–Linearity

6.1 Introduction

In applications of spatial and multivariate extremes (e.g. wave heights at different locations in an ocean basin, or wave heights and wind speed at a single location) it is often the case that we have information about one or more variables/locations and want to simulate the process for another variable/location. More specifically, we might want to estimate probabilities of an extreme event at location Y, given the characteristics of the same event at locations X_1 and X_2. In an oceanographic application, for example, we might be interested in the probability of wave heights exceeding a certain level at a site, given the wave heights being large at some other sites. Conditional simulation from the distribution of $Y \mid X_1, X_2$ is one way of estimating these probabilities of interest.

In this chapter we will show how to do conditional simulation for the max-linear and inverted max-linear models described in Chapter 5. Similarly to Chapter 5, we start with a simple example of a max-linear model, namely the Marshall-Olkin model. In Section 6.2

we describe how to simulate $Y \mid X$ for this model. Then in Section 6.3 we extend the simulation method to a general bivariate and multivariate max-linear model. In Section 6.4 we show how to do conditional simulation from an inverted max-linear model. Finally, in Section 6.5 we will combine the results of Sections 6.3 and 6.4 to show how to simulate from a mixture distribution composed of a max-linear and an inverted max-linear model. For ease of notation, in Sections 6.4 and 6.5 we will limit ourselves to the bivariate case, but the algorithms are straightforward to extend to higher dimensions.

6.2 Marshall-Olkin model

Let us consider the following simple example of a max-linear model:

$$\begin{aligned} X &= \max\{\alpha Z_1, (1-\alpha)Z_2\}, \\ Y &= \max\{\beta Z_1, (1-\beta)Z_3\}, \end{aligned} \tag{6.2.1}$$

where Z_i, $i = 1, 2, 3$, are independent Fréchet random variables and $0 < \alpha, \beta < 1$. We want to simulate samples from the conditional distribution of $Y \mid X$. To do this we need to know the probabilities of X coming from Z_1 and Z_2, respectively. Hence, we calculate the following probability:

$$\begin{aligned} \Pr(\alpha Z_1 > (1-\alpha)Z_2) &= \int_0^\infty \Pr(\alpha x > (1-\alpha)Z_2) f_{Z_1}(x)dx, \\ &= \int_0^\infty F_{Z_2}\left(\frac{\alpha x}{1-\alpha}\right) f_{Z_1}(x)dx, \\ &= \int_0^\infty e^{-(1-\alpha)/\alpha x}\frac{1}{x^2}e^{-1/x}dx, \\ &= \int_0^\infty \frac{1}{x^2}e^{-1/\alpha x}dx, \\ &= \int_0^\infty \frac{\alpha}{y^2}e^{-1/y}dy, \\ &= \alpha, \end{aligned} \tag{6.2.2}$$

where $f(\cdot)$ and $F(\cdot)$ are the Fréchet density and distribution function, respectively. It follows that $\Pr(\alpha Z_1 \leqslant (1-\alpha)Z_2) = 1-\alpha$. Then, assuming X is known, we can simulate from $Y \mid X$ as follows. With probability α, we assume that $X = \alpha Z_1$, and hence $Z_1 = X/\alpha$. Then we simulate Z_2 given that $Z_2 < \alpha/(1-\alpha)Z_1$, and draw Z_3 as an independent Fréchet variable. With probability $(1-\alpha)$, we assume that $X = (1-\alpha)Z_2$ and hence $Z_2 = X/(1-\alpha)$. Then we simulate Z_1 given that $Z_1 < (1-\alpha)/\alpha Z_2$, and draw Z_3 as an independent Fréchet variable. For details see Algorithm 2. Figure 6.2.1 shows an example of a sample simulated from $Y \mid X$ for model (6.2.1) with $\alpha = 0.3$ and $\beta = 0.8$.

Algorithm 2 Algorithm for simulating a sample of size N from $Y \mid X$ for model (6.2.1)

Input: $X_1, \ldots, X_N$; α; β;
Output: $Y_1, \ldots, Y_N$;

for $n \leftarrow 1, N$ **do**
  Draw $u \sim \text{Unif}(0,1)$.
  if $u < \alpha$ **then**
    Set $Z_1 = X_n/\alpha$
    Simulate $Z_2 \left| \left(Z_2 < \frac{\alpha}{1-\alpha} Z_1 \right) \right.$ as follows,
    Begin
      Set $p = e^{-\frac{1-\alpha}{\alpha Z_1}}$
      Draw $Z_U \sim \text{Uniform}(0, p)$
      Set $Z_2 = -1/\log(Z_U)$
    End
  else
    Set $Z_2 = X_n/(1-\alpha)$
    Simulate $Z_1 \left| \left(Z_1 < \frac{(1-\alpha)}{\alpha} Z_2 \right) \right.$ as follows,
    Begin
      Set $p = e^{-\frac{\alpha}{(1-\alpha)Z_2}}$
      Draw $Z_U \sim \text{Uniform}(0, p)$
      Set $Z_1 = -1/\log(Z_U)$
    End
  end if
  Draw $Z_3 \sim$ Fréchet(1) independently;
  Set $Y_n = \max\{\beta Z_1, (1-\beta)Z_3\}$.
end for

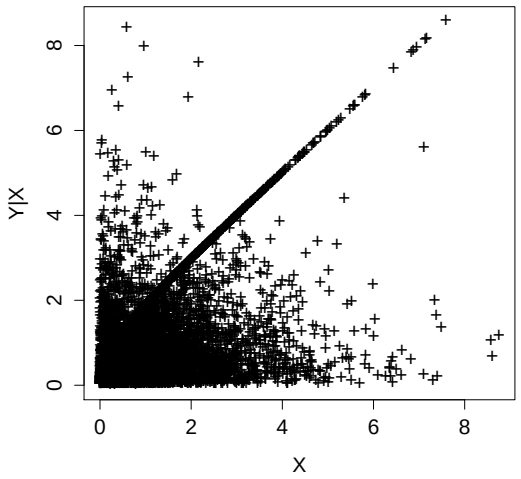


Figure 6.2.1: Simulated sample of $Y \mid X$ plotted against X for model (6.2.1) with $\alpha = 0.3$ and $\beta = 0.8$. Exponential margins are used for ease of visualisation.

6.3 Generalisation

6.3.1 Bivariate case

Consider the following bivariate max-linear model,

$$\begin{aligned} X &= \max_{i=1,\ldots,m} \{\alpha_i Z_i\}, \\ Y &= \max_{i=1,\ldots,m} \{\beta_i Z_i\}, \end{aligned} \tag{6.3.1}$$

where $0 \leqslant \alpha_i, \beta_i \leqslant 1$ for all $i = 1, \ldots, m$, $\sum_{i=1}^{m} \alpha_i = 1$, $\sum_{i=1}^{m} \beta_i = 1$ and $Z_i \sim$ i.i.d Fréchet random variables for $i = 1, \ldots, m$. Without loss of generality, let us assume that the $\alpha_i Z_i$ and $\beta_i Z_i$ terms are ordered such that,

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_{k+l}, 0, \ldots, 0), \tag{6.3.2}$$

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k, 0, \ldots, 0, \beta_{k+l+1}, \ldots, \beta_m), \tag{6.3.3}$$

i.e. for $i = 1, \ldots, k$, $\alpha_i \neq 0$ and $\beta_i \neq 0$, for $i = k+1, \ldots, k+l$, $\alpha_i \neq 0$ and $\beta_i = 0$, and for $i = k+l+1, \ldots, m$, $\alpha_i = 0$ and $\beta_i \neq 0$, with $\sum_{i=1}^{k+l} \alpha_i = 1$ and $\sum_{i=1}^{k} \beta_i + \sum_{h=k+l+1}^{m} \beta_h = 1$.

Similarly to the argument in (6.2.2), it can be shown that, for $i = 1, \ldots, k$,

$$\Pr(\alpha_i Z_i > \max_{j \in \mathcal{J}} \{\alpha_j Z_j\}) = \alpha_i, \tag{6.3.4}$$

where $\mathcal{J} = \{1, \ldots, k+l\} \setminus \{i\}$. Then, we can simulate from the conditional distribution of $Y \mid X$ as given in Algorithm 3.

Algorithm 3 Algorithm for simulating a sample of size N from $Y \mid X$ for model (6.3.1)

Input: $X_1, \ldots, X_N$; $\alpha_1, \ldots, \alpha_m$; $\beta_1, \ldots, \beta_m$;
Require: $\sum_i \alpha_i = 1$, $\sum_i \beta_i = 1$;
Output: $Y_1, \ldots, Y_N$;

for $n \leftarrow 1, N$ **do**
  Draw $u \sim \text{Unif}(0, 1)$.
  for $i \leftarrow 1, (k+l)$ **do**
    if $u < \sum_{l=1}^{i} \alpha_l$ **then**
      Set $Z_i = X_n / \alpha_i$
      for $j \in \{1, \ldots, k\} / \{i\}$ **do**
        Simulate $Z_j \left| \left(Z_j < \frac{X_n}{\alpha_j} \right) \right.$ as follows,
        Begin
          Set $p = e^{-\frac{\alpha_j}{X_n}}$
          Draw $Z_U \sim \text{Uniform}(0, p)$
          Set $Z_j = -1 / \log(Z_U)$
        End
      end for
    end if
  end for
  for $i \leftarrow (k+l+1), m$ **do**
    Draw $Z_i \sim$ Fréchet(1) independently;
  end for
  Set $Y_n = \max_{i=1,\ldots,m} \{\beta_i Z_i\}$.
end for

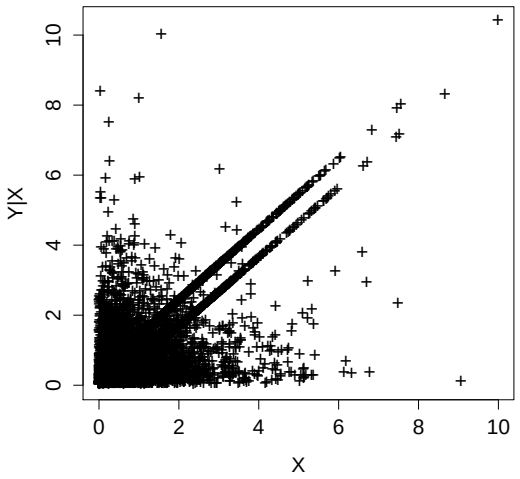


Figure 6.3.1: Simulated sample of $Y \mid X$ plotted against X for Example 6.3.1. Exponential margins are used for ease of visualisation.

Example (6.3.1)

Let $\boldsymbol{\alpha} = (0.3, 0.4, 0.3, 0)$, $\boldsymbol{\beta} = (0.2, 0.6, 0, 0.2)$, and X is a known vector of length $N = 10000$ equal to $\max_{i=1,\ldots,4}\{\alpha_i Z_i\}$, where Z_i, $i = 1, \ldots, 4$, are unknown Fréchet random variables. Hence, in the notation of Algorithm 3, $k = 2$, $l = 1$ and $m = 4$. Then, following Algorithm 3, for each $n = 1, \ldots, N$, we draw a uniform (0,1) variable u. If $u < \alpha_1 = 0.3$, we set $Z_1 = X_n/0.3$, and simulate $Z_2 \mid Z_2 < X_n/\alpha_2$ and $Z_3 \mid Z_3 < X_n/\alpha_3$. If $\alpha_1 \leqslant u < (\alpha_1 + \alpha_2) = 0.7$, we set $Z_2 = X_n/0.4$, and simulate $Z_1 \mid Z_1 < X_n/\alpha_1$ and $Z_3 \mid Z_3 < X_n/\alpha_3$. Finally, if $u > 0.7$, we set $Z_3 = X_n/0.3$, and simulate $Z_1 \mid Z_1 < X_n/\alpha_1$ and $Z_2 \mid Z_2 < X_n/\alpha_2$. Then, since $\alpha_4 = 0$, we draw Z_4 as an independent unit Fréchet variable. Finally, we calculate Y_n as $\max_{i=1,\ldots,4}\{\beta_i Z_i\}$. Figure 6.3.1 shows a simulated sample of $Y \mid X$ plotted against X on Gumbel margins.

6.3.2 Multivariate case

Let $\mathcal{A} = (\alpha_{ji})$ be a $p \times m$ matrix of constants, where $\alpha_{ji} > 0$ and $\sum_j \alpha_{ji} = 1$, for $j \in \mathcal{J}$, $i \in \mathcal{I}$, where $\mathcal{J} = \{1, \ldots, p\}$ and $\mathcal{I} = \{1, \ldots, m\}$. Let $\mathbf{Z} = \{Z_1, \ldots, Z_m\}$ be a vector of m

independent Fréchet random variables. Then, let $\mathbf{X} = \{X_1, \ldots, X_p\}$ be a vector containing p variables, where $X_j = \max\{\alpha_{j1}Z_1, \ldots, \alpha_{jm}Z_m\}$ for $j \in \mathcal{J}$, as given in (6.3.5),

$$\begin{aligned} X_1 &= \max_{i=1,\ldots,m} \{\alpha_{1i} Z_i\}, \\ &\ldots \\ X_p &= \max_{i=1,\ldots,m} \{\alpha_{pi} Z_i\}, \\ Y &= \max_{i=1,\ldots,m} \{\beta_i Z_i\}, \end{aligned} \tag{6.3.5}$$

We want to simulate from the conditional distribution of $Y \mid \mathbf{X}$, where $Y = \max_{i=1,\ldots,m}\{\beta_i Z_i\}$, with $\beta_i > 0$ for $i \in \mathcal{I}$, and $\sum_i \beta_i = 1$. Hence, in vector notation, we have that $Y = \max\{\boldsymbol{\beta}^t \mathbf{Z}\}$, where $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_m\}$.

To simulate from $Y \mid \mathbf{X}$, first we need to establish which Z_i each of the X_j, $j \in \mathcal{J}$, comes from. Let us define the following.

Definition. *For each $i \in \mathcal{I}$, let $\mathcal{J}_i$ be a subset of $\mathcal{J}$, such that for all $j \in \mathcal{J}_i$, $\alpha_{ji} \neq 0$. Let $\mathcal{J}_i^c$ denote the complement of this set, i.e., $\mathcal{J}_i^c = \mathcal{J} \setminus \mathcal{J}_i$.*

Definition. *Let $\mathcal{C} = (c_{ji})$ be a $p \times m$ matrix, where for $i \in \mathcal{I}$,*

$$c_{ji} := \begin{cases} X_j/\alpha_{ji}, & \text{for } j \in \mathcal{J}_i, \\ \infty, & \text{for } j \in \mathcal{J}_i^c. \end{cases}$$

If for all $j \in \mathcal{J}$ for some $l \in \mathcal{I}$, c_{jl} equals the same finite constant $c > 0$, that means that all X_j must come from the same Z_i variable, namely Z_l, and hence we can set $Z_l = c$. Then we can simulate all other Z_i, $i \in \mathcal{I} \setminus \{l\}$, as Fréchet random variables given that $Z_i < X_j/\alpha_{ji}$ for all j where $\alpha_{ji} \neq 0$, i.e., $Z_i < \min_{\mathcal{J}_i}\{X_j/\alpha_{ji}\}$. If $\alpha_{ji} = 0\ \forall j \in \mathcal{J}$ for some $i = 1, \ldots, m$, i.e., $\mathcal{J}_i = \emptyset$, then we draw Z_i independently from a unit Fréchet distribution.

If not all X_j come from the same Z_i then we need to look at values of c_{ji} and use the following lemmas to determine which Z_i lead to which X_j for each $j \in \mathcal{J}$.

Lemma 9. *If $c_{li} > c_{ji}$ for some $j, l \in \mathcal{J}_i$ and $i \in \mathcal{I}$, then* $\Pr(X_l = \alpha_{li} Z_i) = 0$.

Proof. From $c_{li} > c_{ji}$ it follows that $X_l/\alpha_{li} > X_j/\alpha_{ji}$, and hence

$$\max\left\{\frac{\alpha_{l1}}{\alpha_{li}}Z_1, \ldots, Z_i, \ldots, \frac{\alpha_{lm}}{\alpha_{li}}Z_m\right\} > \max\left\{\frac{\alpha_{j1}}{\alpha_{ji}}Z_1, \ldots, Z_i, \ldots, \frac{\alpha_{jm}}{\alpha_{ji}}Z_m\right\}. \tag{6.3.6}$$

It is easy to see that the left hand side of the inequality (6.3.6) cannot be equal to Z_i as this would lead to one of the following two contradictions: (i) $Z_i > Z_i$ if the right hand side of (6.3.6) is also equal to Z_i, (ii) $Z_i > \max\{\alpha_{j1}Z_1/\alpha_{ji}, \ldots, \alpha_{jm}Z_m/\alpha_{ji}\} > Z_i$ if the right hand side of (6.3.6) is greater than Z_i. Hence, it follows that $X_l/\alpha_{li} \neq Z_i$, i.e., $X_l \neq \alpha_{li}Z_i$. □

Lemma 10. *If $\alpha_{ji} = 0$, for some $j \in \mathcal{J}$ and $i \in \mathcal{I}$, then* $\Pr(Z_i = X_j/\alpha_{ji}) = 0$.

Proof. Trivial. □

It follows from Lemma 9 that $X_l \neq \alpha_{li}Z_i$, $i \in \mathcal{I}$, if $c_{li} > c_{ji}$ for some $j, l \in \mathcal{J}_i$. Also, $X_j \neq \alpha_{ji}Z_i$ for all $j \in \mathcal{J}_i^c$, $i \in \mathcal{I}$ by Lemma 10. Hence, let us define the following.

Definition. *Let $\mathcal{M} = (m_{ji})$, $i \in \mathcal{I}$ and $j \in \mathcal{J}$, be an $m \times p$ matrix, where*

$$m_{ji} = \begin{cases} 0, & \text{if } j \in \mathcal{J}_i^c, \\ 0, & \text{if } j \in \mathcal{J}_i \text{ and } c_{ji} > \min\limits_{l \in \mathcal{J}_i \setminus \{j\}} \{c_{li}\}, \\ 1, & \text{otherwise.} \end{cases}$$

This ensures that $m_{ji} = 1$ only for (i, j) pairs where there is non-negative probability of $X_j = \alpha_{ji}Z_i$ (using Lemmas 9 and 10).

Definition. *For each $j \in \mathcal{J}$, let $\mathcal{I}_j$ be a subset of $\mathcal{I}$, such that for all $i \in \mathcal{I}_j$, $m_{ji} = 1$.*

Then, it can be shown that

$$\Pr\left(\alpha_{jk}Z_k > \max_{i \in \mathcal{I}_j \setminus \{k\}} (\alpha_{ji}Z_i)\right) = \frac{\alpha_{jk}}{\sum_{i \in \mathcal{I}_j} \alpha_{ji}}, \tag{6.3.7}$$

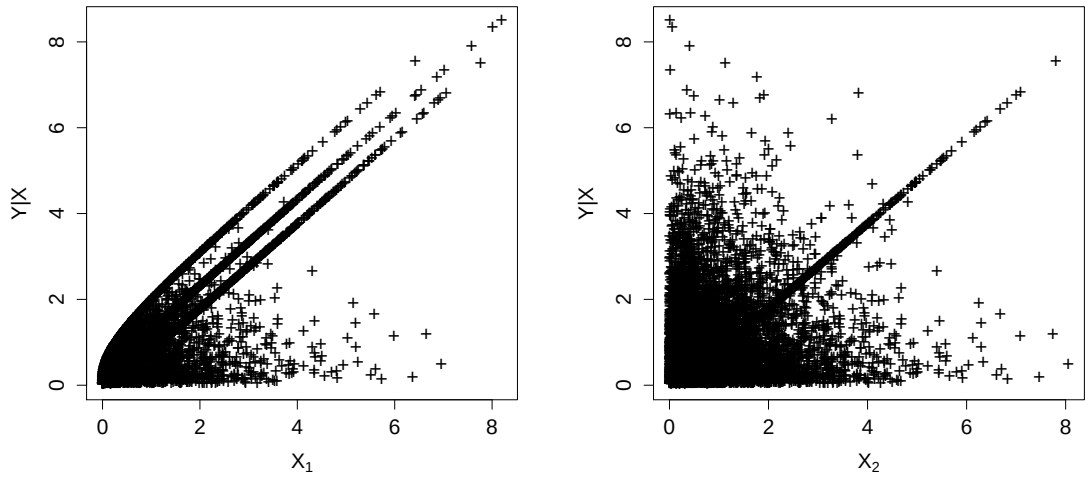


Figure 6.3.2: Simulated sample of $Y \mid \mathbf{X}$ plotted against X_1 and X_2 with $\alpha_{1\cdot} = (0.1, 0.2, 0.3, 0.4)$, $\alpha_{2\cdot} = (0.4, 0.6, 0, 0)$ and $\beta = (0.3, 0, 0.4, 0.3)$. Exponential margins are used for ease of visualisation.

for $k \in \mathcal{I}_j$ and each $j \in \mathcal{J}$. Let us denote the right hand side of equation (6.3.7) by $p_k^{(j)}$ for $k \in \mathcal{I}_j$ and $j \in \mathcal{J}$. Therefore, we can establish that the probability of $X_j = \alpha_{jk} Z_k$ is equal to $p_k^{(j)}$ for $k \in \mathcal{I}_j$ and $j \in \mathcal{J}$. So for each $j \in \mathcal{J}$, we can set $Z_k = X_j/\alpha_{jk}$ with probability $p_k^{(j)}$. Then, we can simulate the remaining Z_i as Fréchet random variables given that $Z_i < \min_{\mathcal{J}_i}\{X_j/\alpha_{ji}\}$. If $\mathcal{J}_i = \emptyset$ for some $i \in \mathcal{I}$, then we draw Z_i independently from a unit Fréchet distribution. See Algorithm 4 for more details. Figure 6.3.2 shows an example of a sample from $Y \mid \mathbf{X}$ with $\alpha_{1\cdot} = (0.1, 0.2, 0.3, 0.4)$, $\alpha_{2\cdot} = (0.4, 0.6, 0, 0)$ and $\beta = (0.3, 0, 0.4, 0.3)$.

6.4 Inverted max-linear model

Let $X_E^{(I)}$ and $Y_E^{(I)}$ represent variables from an inverted max-linear model on exponential marginal scale. Then, it can be shown that $X_E^{(I)} = 1/X$, and similarly, $Y_E^{(I)} = 1/Y$, where X and Y have Fréchet margins and are defined as in model (6.3.1). To simulate from the conditional distribution $Y_E^{(I)} \mid X_E^{(I)}$, first transform back to Fréchet margins $X = 1/X_E^{(I)}$,

Algorithm 4 Algorithm for simulating a sample of size N from $Y \mid \mathbf{X}$ for model (6.3.5)

Input: $X_{j1}, \ldots, X_{jN}$ for $j \in \mathcal{J}$;
α_{ji} for $j \in \mathcal{J}$ and $i \in \mathcal{I}$;
β_i for $i \in \mathcal{I}$;
Require: $\sum_j \alpha_{ji} = 1$, $\forall i \in \mathcal{I}$; $\sum_i \beta_i = 1$;
Output: $Y_1, \ldots, Y_N$;

Initialise $Z_i = 0$ for $i = 1, \ldots, m$, and FOUND.Z = FALSE.
for $n \leftarrow 1, N$ **do**
  for all $i \in \mathcal{I}$ **do**
    Set $\mathcal{J}_i = \{j \in \mathcal{J} : \alpha_{ji} \neq 0\}$.
    for $j \in \mathcal{J}$ **do**
      Set

$$c_{ji} := \begin{cases} X_{jn}/\alpha_{ji}, & \text{for } j \in \mathcal{J}_i, \\ \infty, & \text{for } j \notin \mathcal{J}_i. \end{cases}$$

    end for
  end for
  for all $i \in \mathcal{I}$ **do**
    if $\min_{j \in \mathcal{J}}\{c_{ji}\} = \max_{j \in \mathcal{J}}\{c_{ji}\} \neq \infty$ **then**
      Set $Z_i = X_{jn}/\alpha_{ji}$
      Set FOUND.Z = TRUE
    end if
  end for
  if FOUND.Z = FALSE **then**
    for all $i \in \mathcal{I}$ **do**
      for all $j \in \mathcal{J}$ **do**
        Set

$$m_{ji} = \begin{cases} 0, & \text{if } j \notin \mathcal{J}_i, \\ 0, & \text{if } j \in \mathcal{J}_i \text{ and } c_{ji} > \min\limits_{l \in \mathcal{J}_i \setminus \{j\}}\{c_{li}\}, \\ 1, & \text{otherwise.} \end{cases}$$

      end for
    end for
    for $j \in \mathcal{J}$ **do**
      Set $\mathcal{I}_j = \{i \in \mathcal{I} : m_{ji} = 1\}$.
      for $k \in \mathcal{I}_j$ **do**
        Set

$$p_k^{(j)} = \frac{\alpha_{jk}}{\sum_{i \in \mathcal{I}_j} \alpha_{ji}}.$$

      end for
    end for

Algorithm 4 Algorithm for simulating a sample of size N from $Y \mid \mathbf{X}$ for model (6.3.5) (continued)

for $j \in \mathcal{J}$ **do**
  Draw $u \sim \text{Unif}(0, 1)$.
  for $k \in \mathcal{I}_j$ **do**
    if $u < \sum_{l=1}^{k} p_l^{(j)}$ **then**
      Set $Z_k = X_{jn}/\alpha_{jk}$
    end if
  end for
end for
end if
if FOUND.Z = TRUE **then**
for all $i \in \mathcal{I}$ **do**
  if $Z_i = 0$ **then**
    if $\mathcal{J}_i \neq 0$ **then**
      Simulate $Z_i \,|(Z_i < \min_{\mathcal{J}_i}\{X_{jn}/\alpha_{ji}\})$ as follows,
      Begin
        Set $p = e^{-1/\min_{\mathcal{J}_i}\{X_{jn}/\alpha_{ji}\}}$
        Draw $Z_U \sim \text{Uniform}(0, p)$
        Set $Z_i = -1/\log(Z_U)$
      End
    else
      Draw $Z_i \sim$ Fréchet independently.
    end if
  end if
end for
end if
Set $Y_n = \max_{i=1,\ldots,m}\{\beta_i Z_i\}$.
end for

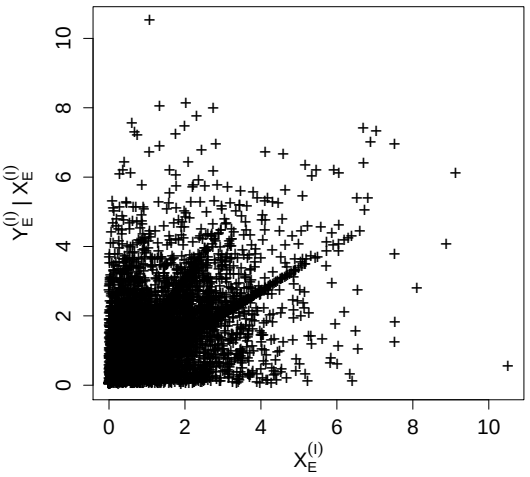


Figure 6.4.1: Simulated sample of $Y_E^{(I)} \mid X_E^{(I)}$ plotted against $X_E^{(I)}$ on Exponential margins for the inverted version of the model in Example 6.3.1.

then follow Algorithm 3 to simulate $Y \mid X$. Finally, transform back to the inverted model using $Y_E^{(I)} = 1/Y$. Figure 6.4.1 shows a simulated sample of $Y_E^{(I)} \mid X_E^{(I)}$ for the inverted version of the model in Example 6.3.1.

6.5 Mixture model

In Chapter 5 we have introduced the following bivariate mixture distribution that combines a max-linear model with an inverted max-linear model:

$$\begin{pmatrix} X_M \\ Y_M \end{pmatrix} = \begin{cases} \begin{pmatrix} X_E \\ Y_E \end{pmatrix} & \text{with probability } \delta \\ \begin{pmatrix} X_E^{(I)} \\ Y_E^{(I)} \end{pmatrix} & \text{with probability } 1-\delta \end{cases} \tag{6.5.1}$$

where $\delta \in [0,1]$, and (X_E, Y_E) and $(X_E^{(I)}, Y_E^{(I)})$ represent a max-linear model and an inverted max-linear model, respectively, on exponential margins. In this section we will show how to

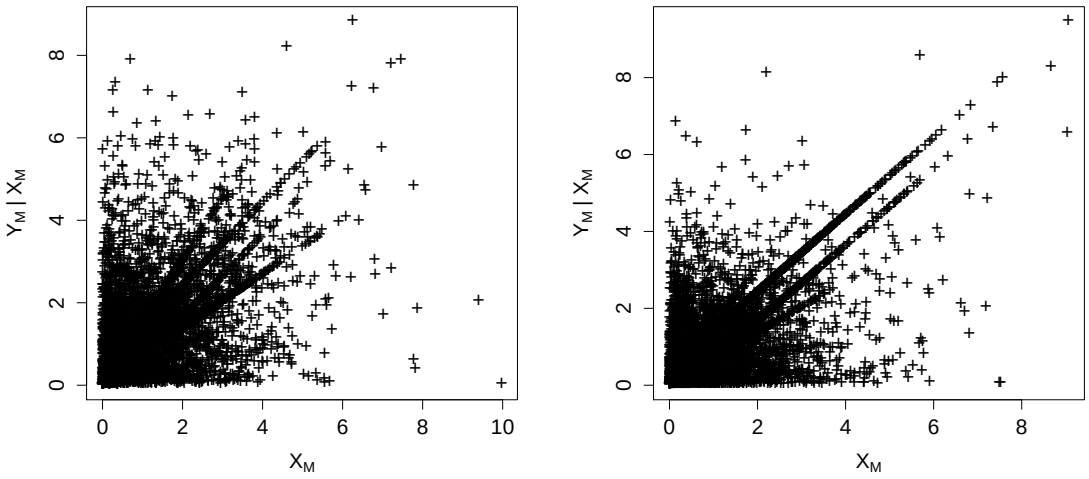


Figure 6.5.1: Simulated sample of $Y_M \mid X_M$ plotted against X_M on Exponential margins for model (6.5.1) with parameters as in example (6.3.1). Left panel: $\delta = 0.2$; right panel: $\delta = 0.8$.

simulate from the conditional distribution of $Y_M \mid X_M$.

First, we draw a random variable u from a Uniform(0,1) distribution. Since X_M comes from the max-linear part of the model with probability δ (i.e. $X_M = X_E$), if $u < \delta$, we transform X_E to Fréchet margins using $X = -1/\log(1 - \exp(-X_E))$. Then, we use Algorithm 3 to simulate $Y \mid X$, and transform to exponential margins to obtain $Y_M = Y_E = -\log(1 - e^{-1/Y})$. If $u > \delta$, then X_M comes from the inverted max-linear part of the model, hence we follow the procedure described in Section 6.4 to simulate $Y_E^{(I)} \mid X_E^{(I)}$ and hence, obtain $Y_M = Y_E^{(I)}$. Figure 6.5.1 shows simulated samples from $Y_M \mid X_M$ for different δ values, for a mixture of the max-linear model in Section 6.3.1 and its inverted version. It can be seen that for $\delta = 0.2$ the inverted max-linear part is more prominent, whereas for $\delta = 0.8$ the max-linear part is.

Chapter 7

Conclusions

In this final chapter, the main contributions of this thesis are summarised, followed by a discussion of possible directions for further work.

7.1 Outcomes of thesis

The aim of this thesis was to develop methods for assessing and modelling extremal dependence for spatial applications. In the following we summarise the main results of the thesis and outline how these results contribute to achieving the aims of the thesis.

As described in Chapter 1, it is difficult to determine with reasonable certainty whether data are asymptotically dependent (AD) or asymptotically independent (AI). This presents a problem, as most models for spatial extremes assume either AD or AI, with very few models able to capture both dependence classes. An important outcome of this thesis is the development of novel diagnostic tools that aid decision making in model selection. Results in Chapter 3 show that standard measures for extremal dependence give conflicting results for finite sample sizes. We found that supplementing these measures with a measure for dependence in the body of the data improves diagnostic performance. Applying this diagnostic method to a North Sea significant wave height data set showed that these data

are consistent with AD for shorter distances, with dependence decreasing with distance, however for sites further apart AI was found to be more appropriate. We also found that data located along straight-line strips with different orientations appear to have different extremal dependence structures.

Chapter 4 followed on from Chapter 3, further investigating the effect of direction on extremal dependence. We used composite likelihood methods to fit various spatial AI and AD models to wave height data at locations along straight-line strips with different orientations. This allowed us to systematically test directional features of the models. All model fits suggested that extremal dependence varies smoothly with the orientation of the strips and that extremal dependence is strongest along strips with an approximately west to east orientation. In the North Sea storms that produce the largest waves typically travel in a north to south direction, so our results suggest that dependence is strongest along the storm front rather than in the storm direction.

For spatial applications it is desirable to have a model that can capture both the short and long range dependence in the data. Existing models that attempt to model both AD and AI are difficult to implement in practice due to their complexity and restrictions in the forms of AD and AI they can model. In Chapter 5 we introduced a family of bivariate distributions, with simple multivariate extensions, that exhibits all the required features of short, medium and long range extremal dependence. To construct this model we used the bivariate max-linear model as the building block. We introduced two bivariate distributions that are derived from a max-linear model and an inverted max-linear model, respectively. We also show a way of combining these two models that provides a flexible approach to modelling extremal dependence that has both AI and AD components. This class of models has the potential to be useful in spatial applications, as it allows modelling of a wide range of dependence structures. In Chapter 5 we studied mathematically the extremal structure of this class of models using a new radial-angular representation. As these models put mass on rays and on lines, there are clear limitations to deriving very

specific inference from them, so in their current form these models are not immediately applicable in practice. However, in Section 7.2 we show that they make a strong building block for more practically reasonable models. Unconditional simulation from these models is straightforward using existing methods. We give algorithms for conditional simulation in Chapter 6.

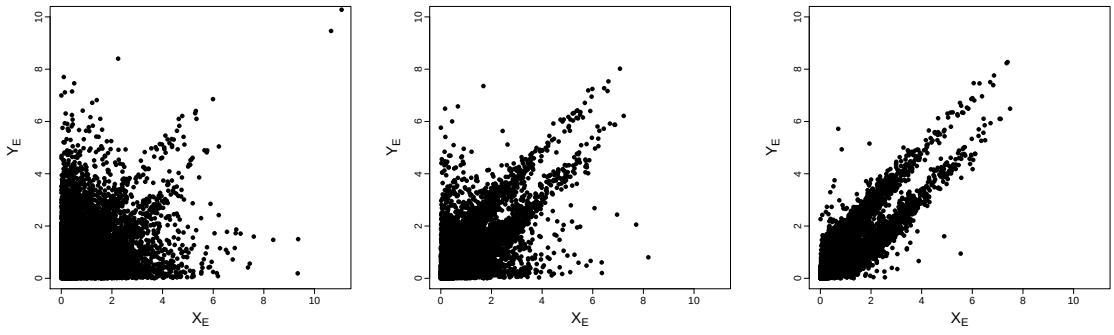
7.2 Directions for further work

The class of models introduced in Chapter 5 put mass on rays and lines, which is inconsistent with most data applications where an assumption of a joint density everywhere is reasonable. Consequently, if these models are fitted using likelihood/Bayesian-based inference they would need almost as many parameters as data points to get a reasonable fit as each line of mass can only explain one data point. Therefore, as currently set up, these are not parsimonious models for likelihood inference but can be used as building blocks for future parsimonious model development. Alternatively, such models can be fitted using other inference criteria which do not depend on the mass on rays/lines, such as moment based methods.

Here we will outline an extension of the models described in Chapter 5 that makes likelihood inference parsimonious, and also makes the model more realistic for practical applications by removing mass from being exactly on lines and rays. For simplicity, let us consider a simple case of the max-linear model given in expression (5.1.3) in Chapter 5. Let $m = 4$ and $\alpha_4 = \beta_3 = 0$, giving,

$$\begin{aligned} X_F &= \max\{\alpha_1 Z_1, \alpha_2 Z_2, \alpha_3 Z_3\}, \\ Y_F &= \max\{\beta_1 Z_1, \beta_2 Z_2, \beta_4 Z_4\}, \end{aligned} \tag{7.2.1}$$

where $0 \leqslant \alpha_i, \beta_i \leqslant 1$ for all $i = 1, \ldots, 4$, $\Sigma_{i=1}^4 \alpha_i = 1$, $\Sigma_{i=1}^4 \beta_i = 1$, and $Z_i \sim$ i.i.d. Fréchet, $i =$



(a) (b) (c)

Figure 7.2.1: Bivariate simulations from the max-linear model given in (7.2.1) with α_i and β_i, $i = 1, 2, 3$, simulated from a Dirichlet process with parameters (a) (25, 10, 100) and (10, 25, 100), (b) (25, 10, 10) and (10, 25, 10), (c) (25, 10, 1) and (10, 25, 1). The margins have been transformed to exponential.

$1, \ldots, 4$. For fixed α_i and β_i values this model gives mass on two lines for the transformed max-linear model (5.1.6), and mass on two rays for the inverted max-linear model (5.1.7). Now, let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_4)$ be independent random variables from two Dirichlet distributions of order three with parameters $\mathbf{d}_\alpha$ and $\mathbf{d}_\beta$, respectively. Hence, in this set up, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are latent parameters of the model.

To simulate from this model with known $\mathbf{d}_\alpha$ and $\mathbf{d}_\beta$ parameter, first draw $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from the distributions Dirichlet($\mathbf{d}_\alpha$) and Dirichlet($\mathbf{d}_\beta$), respectively. Then simulate $Z_1, \ldots, Z_4$ as i.i.d. Fréchet random variables, and calculate (X_F, Y_F) using (7.2.1). From (X_F, Y_F) then we can obtain samples from the transformed max-linear and the inverted max-linear models using the transformations given in Chapter 5. Some example simulations from the transformed max-linear model are given on Figure 7.2.1. First note that there is no mass exactly on lines, but points are still clustered around two lines. The underlying Dirichlet distribution parameters determine how spread out the points are. Similar simulations can be performed also from the other two models described in Chapter 5, the inverted max-linear model and the mixture model. Since there is no longer mass on lines or rays, likelihood inference is now possible for these models.

For spatial applications, it is possible to let $\mathbf{d}_\alpha$ and $\mathbf{d}_\beta$ vary with distance, and hence, the

model captures a wide range of extremal dependence structures, from AD for sites close by, with dependence decreasing with distance, to AI for sites further apart. Further work therefore involves fitting this model and its extensions with both AD and AI dependence structure over locations with a range of separation distances.

Bibliography

A. J. Aarnes, O. Breivik, and M. Reistad. Wave extremes in the North East Atlantic. *J. Climate*, 25:1529–1543, 2012.

J. Beirlant, Y. Goegebeur, J. Segers, and Teugels J. *Statistics of Extremes: Theory and Applications*. Wiley, 2004.

J. Blanchet and A.C. Davison. Spatial modelling of extreme snow depth. *Annals of Applied Statistics*, 5(3):1699–1725, 2011.

O. Breivik, O. J. Aarnes, J.-R. Bidlot, A. Carrasco, and O. Saetra. Wave extremes in the North East Atlantic from ensemble forecasts. *J. Climate*, 26:7525–7540, 2013.

B. M. Brown and S. I. Resnick. Extreme values of independent stochastic processes. *J. Appl. Prob.*, 14(4):732–739, 1977.

S. G. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *J. Roy. Statist. Soc. B*, 53(2):377–392, 1991.

S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: an application to structural design. *Appl. Statist.*, 43(1):1–48, 1994.

S. G. Coles, J. E. Heffernan, and J. A Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365, 1999.

D. Cooley, J. Cisewski, R. J. Erhardt, E. Mannshardt, B. O. Omolo, and Y. Sun. A survey of spatial extremes: measuring spatial dependence and modeling spatial effects. *Revstat*, 10(1):135–165, 2012.

R. A. Davis and S. I. Resnick. Basic properties and prediction of max-ARMA processes. *Adv. App. Prob.*, 21(4):781–803, 1989.

R. A. Davis, C. Klüppelberg, and C. Steinkohl. Statistical inference for max-stable processes in space and time. *J. Roy. Statist. Soc. B*, 75(5):791–819, 2013.

A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modeling of spatial extremes. *Statist. Sci*, 27(2):161–186, 2012.

A. C. Davison, R. Huser, and E. Thibaud. Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.*, 45(5):511–529, 2013.

L. de Haan. A spectral representation for max-stable processes. *Ann. Prob.*, 12(4):1194–1204, 1984.

L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1): 7–45, 1998.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

L. de Haan and S. I. Resnick. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.*, 32(2):213–224, 1989.

P. Deheuvels. Point processes and multivariate extreme values. *J. Mult. Anal.*, 272:257–272, 1983.

S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.

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

K. C. Ewans and P. Jonathan. Evaluating environmental joint extremes for the offshore industry. *Journal of Marine Systems*, 130:124–130, 2014.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, 24(02):180–190, 1928.

A.-L. Fougères, C. Mercadier, and J.P. Nolan. Dense classes of multivariate extreme value distributions. *J. Mult. Anal.*, 116:109–129, 2013.

A.-L. Fougères, L. de Haan, and C. Mercadier. Bias correction in multivariate extremes. *Annals of Statistics*, 43(3):903–934, 2015.

M. G. Genton, S. A. Padoan, and H. Sang. Multivariate max-stable spatial processes. *Biometrika*, pages 215–230, 2015.

A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2014. R package version 1.0-2.

J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.*, 17(2):537–571, 2007.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. Roy. Statist. Soc. B*, 66(3):497–546, 2004.

J. C. Heideman and D. A. Mitchell. Grid point pooling in extreme value analysis of hurricane hindcast data. *J. Waterway, Port, Coastal, Ocean Eng.*, 135(2):31–38, 2009.

B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 1975.

R. Huser and A. C. Davison. Space-time modelling of extreme events. *J. Roy. Statist. Soc. B*, 76(2):439–461, 2012.

A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Q. J. Roy. Meteorol. Soc.*, 81:158–172, 1955.

P. Jonathan, K. C. Ewans, and G. Z. Forristall. Statistical estimation of extreme ocean environments: the requirement for modelling directionality and other covariate effects. *Ocean Eng.*, 35:1211–1225, 2008.

P. Jonathan, K. C. Ewans, and D. Randell. Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Eng.*, 79:22–31, 2013.

P. Jonathan, Ewans K., and D. Randell. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188, 2014.

Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *Ann. Prob.*, 37(5):2042–2065, 2009.

C. Keef, I. Papastathopoulos, and J. A. Tawn. Estimation of the conditional distribution of a vector variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Mult. Anal.*, 115:396–404, 2013.

M. Kereszturi, J.A. Tawn, and P. Jonathan. Assessing extremal dependence of North Sea storm severity. *Oce. Eng.*, 118:242–259, 2016.

F. Laurini and J. A. Tawn. The extremal index for GARCH(1, 1) processes. *Extremes*, 15 (4):511–529, 2012.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. B*, 59(2):475–499, 1997.

A. W. Ledford and J. A. Tawn. Diagnostics for dependence within time series extremes. *J. Roy. Statist. Soc. B*, 65(2):521–543, 2003.

P. M. Lee. *Bayesian Statistics: An Introduction (4th Edition)*. Wiley, 2012.

A. W. Marshall and I. Olkin. A multivariate exponential distribution. *J. Amer. Statist. Assoc.*, 62(317):30–44, 1967.

R. B. Nelsen. *An Introduction to Copulas*. Springer, 2006.

P. J. Northrop and C. L. Coleman. Improved threshold diagnostic plots for extreme value analysis. *Extremes*, 17(2):289–303, 2014.

S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, 105(489):263–277, 2010.

I. Papastathopoulos and J. A. Tawn. Conditioned limit laws for inverted max-stable processes. *J. Mult. Anal.*, 150:214–228, 2016.

J. Pickands. The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, 8(4):745–756, 1971.

J. Pickands. Statistical inference using extreme order statistics. *Ann. Statist.*, 3(1):119–131, 1975.

J. Pickands. Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute: Proceedings of the 43rd Session (Buenos Aires)*, pages 859–878. Vorburg, Netherlands: ISI, 1981.

A. Ramos and A. W. Ledford. A new class of models for bivariate joint tails. *J. Roy. Statist. Soc. B*, 71(1):219–241, 2009.

D. Randell, K. Turnbull, Ewans K., and P. Jonathan. Bayesian inference for nonstationary marginal extremes. *Environmetrics*, 27:439–450, 2016.

M. Reistad, O. Breivik, H. Haakenstad, O. J. Aarnes, B. R. Furevik, and J.-R. Bidlot. A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea. *J. Geophys. Res.*, 116:1–18, 2011.

S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.

M. Ribatet. *SpatialExtremes: A R package for Modelling Spatial Extremes*, 2008. R package version 2.0-2.

C. J. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 10(1):33–59, 2012.

M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

M. Sibuya. Bivariate extreme statistics. *Ann. Inst. Stat. Math.*, 11:195–210, 1960.

D. P. Simpson, H. Rue, T. G. Martins, A. Riebler, and S.H. Sorbye. Penalising model component complexity: a principled, practical approach to constructing priors. *arXiv e-prints: 1403.4630v4*, 2015.

R. L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sci.*, 4(4):367–393, 1989.

R. L. Smith. Max-stable processes and spatial extremes. *Unpublished article, available electronically from www.stat.unc.edu/postscript/rs/spatex.pdf*, 1990.

C. Spearman. The proof and measurement of association between two things. *Am. J. Psychol.*, 15(1):72–101, 1904.

A. G. Stephenson. *evd: Extreme Value Distributions*, 2002. R package version 2.3-0.

J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3): 397–415, 1988.

E. Thibaud, R. Mutzner, and A. C. Davison. Threshold modelling of extreme spatial rainfall. *Water Resourcer Research*, 49(8):4633–4644, 2013.

E. Thibaud, J. Aalto, D. S. Cooley, A. C. Davison, and J. Heikkinen. Bayesian inference for the Brown-Resnick process, with an application to extreme low temperatures. *arXiv e-prints: 1506.07836v2*, 2016.

Richard Von Mises. La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1(1), 1936.

J. L. Wadsworth. Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126, 2016.

J. L. Wadsworth and J. A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272, 2012a.

J. L. Wadsworth and J. A. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. Roy. Statist. Soc. B*, 74(3):543–567, 2012b.

J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, 2013a.

J. L. Wadsworth and J. A. Tawn. A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714, 2013b.

J. L. Wadsworth and J. A. Tawn. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, 2014.

Y. Wang and S. A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. *Adv. Appl. Prob.*, 43:461–483, 2011.

H. Winter. *Extreme value modelling of heatwaves*. PhD thesis, Lancaster University, Lancaster, UK, 2015.