Extremes of Aggregated Random Variables and Spatial Processes

Jordan Richards, M.Sci.(Hons.), M.Res

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

July 2021

STOR-i
excellence with impact

Abstract

This thesis aims to develop on literature for modelling the extremal behaviour of aggregates of random variables or spatial processes, where here the aggregate refers to the arithmetic mean, or sum, of a collection of random variables, or the integral of a stochastic process. The tail behaviour of aggregates is of interest to practitioners in industries such as financial trading, where extreme returns or losses of a portfolio, i.e., a weighted aggregate of financial derivatives, are of interest. Another area where the literature is applicable is in risk management for river flooding, which typically occur with heavy rain- or snowfall over a catchment area; this problem can be formulated as an extreme value analysis of the total volume of rain or snow that falls within a specified spatio-temporal region.

Aggregation acts as a smoothing operation, meaning that all information about the underlying process that feeds the aggregate variable is lost; this can potentially lead to unreliable inference when only the sample aggregate data are available for modelling. However, given that data for the underlying process are available, we can exploit the relationship between the statistical properties of this process and the extremal behaviour of the aggregate to improve on inference; we provide some approaches for establishing such a relationship.

We derive the first-order behaviour of the survival function of the weighted sum of random variables, as this aggregate variable tends to its upper-endpoint. We do this first for a bivariate sum with dependence within the set of underlying variables

modelled using two widely applied limiting characterisations of extremal dependence. We then extend these results to a $d-$variate sum for finite d, and with dependence modelled fully using certain copulae. In both cases, we establish links between the extremal behaviour of the underlying random variables and the aggregate variable.

We further detail a data-driven approach for modelling the extremes of spatial aggregates. Here we propose a fully spatial model for the extremal behaviour of the underlying process, which relies on conditional methods; we then draw replications from this model to approximate the distribution of the spatial aggregate. Whilst this approach can be applied to any spatial process, we apply it to precipitation and detail considerations that must be taken to make this feasible.

A method for accommodating spatial non-stationarity in the extremal dependence structure of data is also proposed. This relies on transformation of the original coordinate system to a new latent space where stationarity can reasonably be assumed.

Acknowledgements

First and foremost, I would like to acknowledge the substantial patience of my three academic supervisors: Jon Tawn, Jenny Wadsworth and Simon Brown; there's very little I would change about my PhD experience and I could not have asked for a more supportive and knowledgeable trio. Thank you for the time that you spent developing me as a researcher and giving me the confidence to pursue this work further. I'd like to give particular thanks (and maybe some apologies) to Jon, as we've had more conversations together over the past few months than I think I've had with any other person. I can only hope that you're not too happy to see the back of me.

I'm very grateful for the opportunities afforded to me through funding from EPSRC and the STOR-i centre for doctoral training. Speaking of STOR-i, I'm grateful for the working environment that was created by both the students and staff; I'd particularly like to thank Jen, Wendy, Kim, Nicky, Idris, Kevin and Jon for all of your hard work keeping STOR-i running; Oli for your helpful tech support and Dan for your help building my R package. Although I want to acknowledge the support from all of my STOR-i colleagues, I'd especially like to mention my cohort for the continued support and the rather unforgettable shared experiences. Special thanks go to the members of the extremes coffee sessions, for entertaining Tuesday chats and supportive discussions, and attendees of the STOR-i football sessions, for not commenting on my lack of any ability all that often.

I'd like to give thanks to my family and wider friends group for the general support

given to me throughout my life. And last, but certainly by no means least, I'd like to thank Tiegan, who has been my emotional rock for the past 4 years. You have supported me from start to finish; a constant in a world of uncertainty. The work in this thesis is equal parts yours as it is mine, as you've listened attentively through every talk practice, every stress-fuelled rant, and I imagine will be sharing in the celebrations accordingly. Cheers!

Declaration

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

The content in Chapter 3 has been submitted for publication as Richards. J., Tawn, J. A. (2021). On the tail behaviour of aggregated random variables.

The content in Chapter 5 has been submitted for publication as Richards. J., Tawn, J. A., Brown. S. (2021). Modelling extremes of spatial aggregates of precipitation using conditional methods.

The content in Chapter 7 has been accepted for publication as Richards. J., Wadsworth, J. L. (2021). Spatial deformation for nonstationary extremal dependence. *Environmetrics*, e2671. The code and data that supports the findings in this study are available in the accompanying `R` package, `sdfEXTREME`, which can be found at `https://github.com/Jbrich95/sdfExtreme`.

The word count for this thesis is approximately 78,000.

Jordan Richards

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

Quantification of the stochastic behaviour of extreme events is important in numerous applications, ranging from financial, e.g., stock market analysis and insurance pricing, to those of an environmental nature, e.g., modelling of extreme weather or climate events. In both cases, the questions that we wish to address often concern risk mitigation of some kind; can our understanding of the occurrence of extreme events be used to lessen the negative impacts associated with these events, such as financial loss, damage to infrastructure or property, and loss of life?

An intuitive example of risk mitigation for extreme events can be found within the design criteria for weather defences, which are built to withstand all events that they are likely to experience within their projected T-year lifespan. For such a defence to retain this lifespan, they must be built to withstand an event that is expected to occur, or be exceeded, at least once in a T-year period and extreme value theory is a particularly powerful tool that can be used to estimate such events. Here, conventional statistical methods are likely to perform poorly as the T-year period of interest is typically longer than the historical record for which data are observed; that is, we

may be interested in events that are more extreme than those that have ever been observed. Extreme value theory is underpinned by asymptotic or limit arguments that facilitate a framework which allows extrapolation beyond the maxima of data to estimate such events. For example, if were interested in the upper-tail of a random variable X, i.e., its most extreme values, then we could consider the distribution of exceedances $a(u)(X-u)|X > u$ as $u \to x^F$, where x^F denotes the upper-endpoint of X and $a(u)$ is a normalising function which is selected so that there is a non-degenerate limit distribution for the threshold exceedances. Extreme value theory can be used to illustrate that an appropriate distribution for such a random variable is the generalised Pareto distribution, often denoted GPD, with scale and shape parameters, $\sigma > 0$ and $\xi \in \mathbb{R}$ respectively, see Section 2.1.3.

Whilst there exists a richly studied class of statistical models that can be used for inference on the extremal behaviour of univariate random variables, these alone are not sufficient for applications where problems of interest cannot be succinctly described as being univariate, e.g., modelling the aggregate of a multivariate random vector. It is often apt to account for dependence between variables through the use of multivariate models. However, it can quickly become cumbersome to fit said models to high-dimensional data, especially if we expect dependence between variables to differ over different pairs; environmental data often suffers from both of these issues, and it is this type of data that we are most interested in studying in the context of aggregation. We must then turn to certain classes of multivariate models, termed spatial models, which are derived from stochastic processes that are indexed over space; these are appropriate for data that we expect to exhibit statistical characteristics that are affected by the location at which an observation is measured, and dependence is typically characterised as a function of distance between locations, making inference computationally easier.

A key concept that motivates this thesis is that of extremal dependence, i.e.,

the tendency for variables to exhibit extreme events simultaneously, and how this underlying structure affects the univariate aggregate over random vectors or processes. Spatial models for extremal dependence often fall into one of two classes: those that model asymptotic dependence or those that model asymptotic independence, where the two definitions correspond to a non-zero, and zero, probability of extreme events occurring together in the limit as the events become increasingly extreme, respectively. An active area of focus in the literature is the development of parsimonious spatial models that can account for both classes simultaneously.

Modern risk assessments in many areas of interest require estimation of the extremal behaviour of sums or averages of random variables. Such areas include financial risk management, where one may be interested in extreme losses or profits associated with a financial portfolio; this can be formulated as the weighted average of returns from a number of securities. Another key area is fluvial flood risk management, as river flooding is typically caused by prolonged heavy rainfall over a catchment area, which can be quantified as the total volume of rainfall over a spatial region and temporal period.

We consider there to be two main strategies for studying the extremal behaviour of aggregates of random variables: the first is analytical in nature, where we can derive the theoretical behaviour of the upper-tails (or lower-tails) of aggregates of random variables using asymptotic arguments: the second is a data-driven approach and relies on simulation from fitted statistical models. The first approach is a very natural extension of the literature for extreme value theory, as much of this is underpinned by limiting arguments that characterise the first-order behaviour of the extremes of random variables. However, when relying on asymptotics, we only gain insight into the behaviour of the most extreme values of the aggregate, and we might find that this is completely dominated by only a subset of the marginal variables. To learn about the extremes of the aggregate at a sub-asymptotic level, we propose a second

strategy: a model is proposed and fitted to the data that we wish to aggregate and then Monte-Carlo methods can be used to approximate the distribution of the extremes of the aggregate. Both approaches have their advantages and disadvantages, which we discuss.

We begin by deriving the first-order upper-tail behaviour of the weighted sum of bivariate random variables under weak assumptions on their marginal distributions and their copula. The extremal behaviour of the marginal variables is characterised by the generalised Pareto distribution, and for their dependence, we rely on subclasses of the two most general characterisations of joint tail dependence. These representations were first proposed by Ledford and Tawn (1997) and Heffernan and Tawn (2004), and describe both components being jointly extreme, and the behaviour of one component conditional on the other component being large, respectively. We find that the upper-tail behaviour of the aggregate is driven by different factors dependent on the signs of the marginal shape parameters, and the strength of extremal dependence; these relationships are quantified and succinctly presented in four theorems. We also derive the upper-tail behaviour of the aggregate for some well-known copulae which reveals further insight into the tail structure when the copula falls outside the conditions for the subclasses of the limiting dependence representations.

Inference on the extremal behaviour of spatial aggregates of precipitation is important for quantifying river flood risk. There are two classes of previous approach, with one failing to ensure self-consistency in inference across different regions of aggregation and the other requiring highly inflexible marginal and spatial dependence structure assumptions. To overcome these issues, we propose a model for the extremes of high-resolution precipitation data, from which we can simulate realistic fields and explore the extremal behaviour of spatial aggregates. Recent developments in spatial extremes literature have seen promising progress with spatial extensions of the Heffernan and Tawn (2004) model for conditional multivariate extremes, which can handle a

wide range of dependence structures, and so we use an extension of this model for the extremal behaviour of the high-resolution fields. Our contribution is two-fold: new parametric forms for the dependence parameters of the spatial conditional extremes model; and a novel framework for deriving aggregates addressing edge effects and sub-regions without rain. By simulating from our model, we are able to approximate the distribution of aggregates over different spatial regions and can illustrate that this approach provides comparatively more reliable inference on the extremes of spatial aggregates than previous approaches.

Modelling the extremal dependence structure of spatial data is considerably easier if that structure is stationary; that is, pairwise dependence of data observed at different locations is simply a function of the distance beween them. However, for data observed over large or complicated domains, non-stationarity will often prevail. Current methods for modelling non-stationarity in extremal dependence rely on models that are either computationally difficult to fit or require prior knowledge of covariates. Sampson and Guttorp (1992) proposed a simple technique for handling non-stationarity in spatial dependence by smoothly mapping the sampling locations of the process from the original geographical space to a latent space where stationarity can be reasonably assumed. This methodology is designed for modelling dependence in the bulk of the distribution, and it is possible that the strength of dependence changes as we move into the tails, making models for dependence in the bulk no longer appropriate. We thus adapt this methodology to make it appropriate for modelling extremal dependence in a spatial framework, which we achieve by considering least squares minimisation of pairwise theoretical and empirical extremal dependence measures.

1.2 Thesis outline

This thesis aims to develop on methods for modelling the extremal behaviour of aggregates of random variables or spatial processes. The models for aggregates that we introduce in Chapters 3, 4 and 5 rely on weak stationarity assumptions on the dependence structure of the underlying process; when this assumption is not satisfied in an application, a novel methodology for addressing this issue is presented in Chapter 7. The rest of this section follows with an outline of the thesis.

Chapter 2 provides an overview of existing methodology for modelling extreme values. We begin by conducting a thorough review of univariate methods before detailing multivariate extensions of these approaches. The concept of extremal dependence and measures for its quantification are discussed before we provide extensions of multivariate extreme value models to a spatial setting. We conclude with an introduction to the Heffernan and Tawn (2004) conditional approach to modelling multivariate extremes and the spatial extension of this model proposed by Wadsworth and Tawn (2019); these methods provide the foundation for the content of Chapter 5.

Chapters 3 and 4 concern the tail behaviour of aggregates of random vectors. These studies are conducted analytically by deriving the first order upper-tail behaviour of the survival function of the aggregate as the aggregate tends to its upper-endpoint. A wide range of dependence structures and marginal tail behaviour are considered for the components of the random vector and we discuss their effect on the upper-tail of the aggregate. In both chapters, marginal components X_i are assumed to be GPD random variables with scale and shape parameters, $\sigma_i > 0$ and $\xi_i \in \mathbb{R}$ respectively, for $i = 1, \ldots, d$ and for $d \in \mathbb{N}$; note that we do not have necessarily have equality in the marginal parameters. The value of d and the approach to modelling dependence differs between the two chapters, and is to be discussed.

In Chapter 3, we consider the sum of bivariate random variables X_1 and X_2, where their extremal dependence is modelled using the limiting dependence characterisations

of Ledford and Tawn (1996) and Heffernan and Tawn (2004). An interesting collection of results are presented that show that, in some cases, the tail behaviour of the aggregate X_1+X_2 can be linked to the coefficient of asymptotic independence (Ledford and Tawn, 1996). We further illustrate that the derived results can provide good approximations in practice by conducting inference on the upper-tail behaviour of aggregates of gridded precipitation and temperature data. In Chapter 4, we consider instead the $d-$dimensional sum of a random vector $(X_1, \ldots, X_d)$, where dependence is modelled fully using copulas, rather than limiting dependence models. A comparison of the results derived using the two different approaches to modelling dependence is detailed in Chapter 3.

Chapter 5 presents a method for inference on the tail behaviour of spatial aggregates of spatial processes. We first fit a model to high-resolution data; this model is an extension of the spatial conditional extremes model proposed by Wadsworth and Tawn (2019). We then use Monte-Carlo methods to approximate the distribution of the upper-tail of aggregates over different spatial regions. Although this approach is particularly versatile and can be used for inference on the tails of spatial aggregates of any environmental data, we detail specific methodology that allow us to model extreme precipitation. We apply our modelling approach to gridded East-Anglia, UK precipitation data from a convection permitting climate model.

Chapter 6 extends the modelling approach presented in Chapter 5. We utilise an algorithm developed at the Met Office Hadley Centre, UK, to cluster observed precipitation fields from the UK climate predictions (2018) into one of two classes: data produced from a convective storm cell and otherwise. We then fit extensions of the model detailed in Chapter 5 to the two clusters, separately, and contrast their estimated extremal dependence structures. Monte-Carlo methods can then be used to draw realisations from both fitted models, which can be combined into a single sample and used for approximating the upper-tail behaviour of spatial aggregates; we

compare this modelling approach to that detailed in Chapter 5 and find that improved inference on the tails of spatial aggregates is made.

In Chapter 7, we present an extension of the deformation method proposed by Sampson and Guttorp (1992) to a spatial extremes framework. Along with some practical advice on applying these deformations, we provide a detailed simulation study in which we propose three spatial processes with varying degrees of non-stationarity in their extremal and central dependence structures. The methodology is applied to Australian summer temperature extremes and UK precipitation to illustrate its efficacy compared to a naive modelling approach where non-stationarity is ignored.

Chapter 8 concludes with a summary of the contribution of this thesis and a discussion of potential opportunities for further work.

2

Literature review

2.1 Univariate extreme value theory

2.1.1 Overview

Classical univariate extreme value theory is underpinned by the properties of max-stability and threshold-stability which give rise to uniquely defined distributions for the limiting characteristics of random variables. We present the two most widely used approaches for modelling univariate extreme values: the block maxima approach using the generalised extreme value distribution, and the peaks-over-threshold approach, which models exceedances above a threshold with the generalised Pareto distribution. We further present an alternative characterisation of both approaches using point processes. A comphrehensive overview of these methods is given in Coles (2001).

2.1.2 Limiting distribution of maxima

Fisher and Tippett (1928) provide a class of limiting distributions for the maxima of univariate random variables via the extremal types theorem. Let $X_1, \ldots, X_n$ be independent random variables with common distribution function F and consider the maximum $M_n = \max\{X_1, \ldots, X_n\}$.

Theorem 2.1.1 (Extremal types theorem)**.** *If there exist sequences of normalising constants* $\{a_n > 0\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$*, such that*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \quad as \;\; n \to \infty \tag{2.1.1}$$

for $z \in \mathbb{R}$ *and for non-degenerate distribution function* G*, then* G *belongs to one of three extreme value classes: Fréchet, negative Weibull or Gumbel.*

The extreme value classes of distributions are defined as:

- **Gumbel**: $G(z) = \exp\left[-\exp\left\{-\left(\frac{z-b}{a}\right)\right\}\right], \quad z \in \mathbb{R};$
- **Negative Weibull**: $G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \geq b; \end{cases}$
- **Fréchet**: $G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{\alpha}\right\}, & z > b, \end{cases}$

for $a > 0,\ b \in \mathbb{R}$ and $\alpha > 0$, and, if (2.1.1) holds, we say that F is in the maximum domain of attraction (MDA) of G. It can be shown that the above distributions can be represented by special cases of a single family. The generalised extreme value (GEV) distribution has distribution function

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}, \quad z \in \mathbb{R}, \tag{2.1.2}$$

with $\{y\}_{+} = \max\{0, y\}$ and location and scale parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, respectively. The shape parameter $\xi \in \mathbb{R}$ determines the extreme value class: for $\xi > 0$ and $\xi < 0$, we have Fréchet and negative Weibull distributions, respectively, and we interpret $\xi = 0$ as the limit $\xi \to 0$, which leads to the Gumbel family. This implies that the lower- and upper-endpoints, z_G and z^G, respectively, of G are dependent on the value of ξ. If $\xi < 0$, then $z^G = \mu - \sigma/\xi$ and if $\xi > 0$, we have $z_G = \mu - \sigma/\xi$.

To motivate the importance of the normalising sequences, consider the distribution function of M_n, which is

$$\Pr(M_n \leq x) = \Pr(X_1 \leq x, \ldots, X_n \leq x) = \{\Pr(X \leq x)\}^n = F^n(x).$$

In practice, $F(x)$ is unknown and $F^n(x)$ is degenerate as $n \to \infty$. This follows as, for all $x < x^F$ where x^F is the upper end-point of F, we have that $F^n(x) \to 0$ as $n \to \infty$. Conversely, $F^n(x) \to 1$ as $n \to \infty$ for all $x \geq x^F$.

The GEV family is the only class of distributions which satisfy the max-stability property. A distribution G is said to be max-stable if there exist constants $\alpha_t > 0$ and β_t such that

$$\{G(\alpha_t z + \beta_t)\}^t = G(z),$$

for all $t \in \mathbb{N}$ and $z \in \mathbb{R}$. That is, max-stability is the property satisfied by distributions for which the operation of taking sample maxima of independent copies leads to an identical distribution, albeit with a change of location and scale. If we assume equality in (2.1.1) holds for large n, then

$$\Pr\{M_n \leq z\} = G\left(\frac{z - b_n}{a_n}\right) = \bar{G}(z),$$

where $\bar{G}$ is a GEV distribution with a different location and scale parameter to that of G. This result allows us to model maxima in practice, as data can be blocked into sections of equal length m and the maxima of each block treated as a realisation from $\bar{G}$. This is termed the block maxima approach, which we illustrate in Figure 2.1.1.

Inverting equation (2.1.2) provides a method for estimating the quantiles z_p for

the GEV distribution. The solution to $G(z_p) = 1 - p$ is

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

We term the quantile z_p as the return level associated with a return period of $1/p$ "blocks" for a GEV fitted to block maxima of length m, i.e., the probability of exceeding z_p in a given period of length m is p. For example, if $p = 0.01$ and each of the m blocks corresponds to a year of observations, then $z_{0.01}$ is the return level associated with a return period of $1/0.01 = 100$ years.

Coles (2001) detail methods for obtaining parameter and uncertainty estimates of the GEV parameters (μ, σ, ξ) via maximum likelihood estimation. However, we note that the standard asymptotic normality properties of maximum likelihood estimators do not necessarily apply in the case of the GEV distribution when the shape parameter $\xi < -1/2$, as the upper-endpoint of the distribution is a function of the parameters; Smith (1985) studies this particular problem and highlight that if the true shape $\xi < -1$, then maximum likelihood estimators may not exist and alternative means of parameter estimation must be used.

We further note that if the variable of interest is minima, rather than maxima, then this can be studied under the same framework by considering

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

2.1.3 Threshold exceedance modelling

An alternative approach to modelling univariate extreme events is to consider exceedances above a high threshold u. This approach has the immediate advantage over the former that typically more data is available for model fitting. We illustrate this

in Figure 2.1.1 by comparing the two approaches, applied to a common simulated dataset.

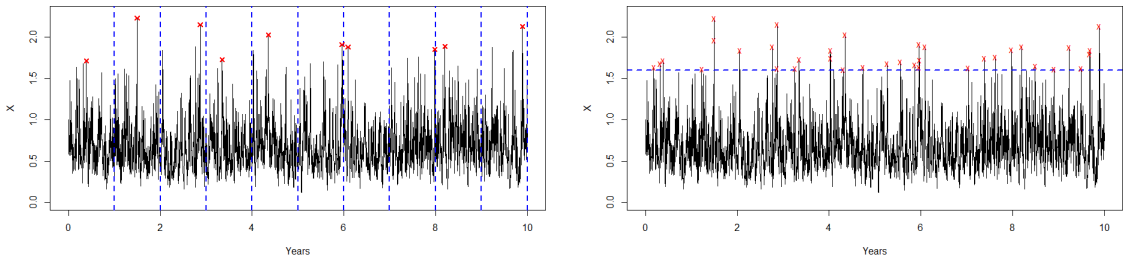

Figure 2.1.1: Illustration of block maxima (left) and threshold exceedance (right) modelling for univariate extremes. Both approaches are applied to the same simulated data with the red points denoting those used for inference.

With justifcation from Pickands (1975), exceedances above a fixed high threshold u are modelled using the generalised Pareto distribution (GPD). Consider $X \sim F$, where F is in the MDA of a GEV(μ, σ, ξ) distribution. Then the distribution function of $(X - u)|X > u$ is approximated by

$$H(y) = \left(1 + \frac{\xi}{\sigma_u} y\right)_+^{-1/\xi}, \quad (y > 0),$$

where $\sigma_u = \sigma + \xi(u - \mu) > 0$. The scale parameter σ_u is determined by the choice of threshold u and the shape parameter ξ is equal to that of the associated GEV distribution. Davison and Smith (1990) define the characterising property of the GPD as that of threshold stability, which is unique to this class of distributions. Suppose that $(X - u_0)|X > u_0$ is GPD(σ_{u_0}, ξ) above threshold u_0. By definition of a GPD, the distribution of $(X - u)|X > u$ for $u > u_0$ is also GPD, with a scale change to $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$.

This model provides information on the upper-tail behaviour of X, i.e., $X|X > u$, only and so an appropriate model must be chosen for $X < u$; we refer to this as the bulk of the distribution of X. A typical non-parametric model for the bulk is the empirical distribution $\tilde{F}(x)$ of observations less than u, see Coles (2001). The

resulting distribution function of X is

$$F(x) = \begin{cases} \tilde{F}(x) & \text{if } x \leq u \\ 1 - \lambda_u \left[1 + \frac{\xi(x-u)}{\sigma_u}\right]_+^{-1/\xi} & \text{if } x > u, \end{cases}$$

where $\lambda_u = 1 - \tilde{F}(u)$. Parametric models for the bulk of X with GPD lower and/or upper tails, have been proposed in the literature, e.g., Gamma-GPD and Weibull-GPD mixtures (Behrens et al., 2004) and Normal-GPD mixture (de Melo Mendes and Lopes, 2004); these models rely on continuity constraints to ensure that the density of X is smooth across the exceedance threshold.

The choice of u is particularly important when conducting inference using a GPD; if u is too small, then it is unlikely that the asymptotic arguments which give rise to the GPD will hold in practice. Choosing u too high reduces the number of data available for inference, and so parameter estimates will have higher standard errors. There are numerous heuristic techniques for selecting the threshold and one of the most commonly used is parameter stability plots. For this technique, estimates of the shape parameter are observed for a range of u and a suitable u chosen such that the parameter value lies in a "stable" region, i.e., estimates appear to be somewhat constant over a neighbourhood of u values. A review of alternative techniques is provided by Scarrott and MacDonald (2012). To avoid the trade-off entirely, there has been recent developments in parametric mixture models for the entire distribution of X, but do not require threshold selection, see Papastathopoulos and Tawn (2013) and Naveau et al. (2016).

2.1.4 Point process characterisation

Pickands (1971) provides a framework for characterising the modelling approaches detailed in Sections 2.1.2 and 2.1.3 using point processes. Coles (2001) formulates the

result in the following theorem:

Theorem 2.1.2. *Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables and suppose there exists sequences of normalising constants $\{a_n > 0\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ such that* (2.1.1) *holds, where G takes the form given in* (2.1.2) *with lower- and upper-endpoints, z_G and z^G, respectively. Then for any $u > z_G$, the sequence of point processes*

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : \ i = 1, \ldots, n \right\},$$

converges on regions of the form $(0,1) \times [u, \infty)$ as $n \to \infty$, to a Poisson process with intensity measure

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi}, \tag{2.1.3}$$

where $A = [t_1, t_2] \times [z, z^G)$ and $0 \leq t_1 < t_2 \leq 1$.

To use the point process framework for modelling the extremes of an observed sample $x_1, \ldots, x_n$, we select a high exceedance threshold u. We consider $A = (0,1) \times [u, \infty)$ and re-label the $N(A)$ points observed in A as $\{(t_1, x_1), \ldots, (t_{N(A)}, x_{N(A)})\}$. Note that if data are observed in k blocks of length m, and we are interested in the distribution of block-maxima, rather than $k-$block maxima, we must make a slight adjustment to the intensity measure in (2.1.3), as the parameters in this measure currently correspond to a GEV for $k-$block maxima. For example, if we observe 20 years of data, we may be interested in the distribution of yearly maxima; moreover, we are unlikely to be interested in the distribution of 20$-$year maxima as only a single observation will be available for inference. Thus to make this approach applicable for modelling block maxima of length m, we simply replace the intensity measure in (2.1.3) with

$$\Lambda(A) = k(t_2 - t_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\xi}.$$

As all data larger than u are used for estimating the GEV parameters, rather than just the sample of block maxima, this approach can result in more accurate inference compared to that described in Section 2.1.2.

We can now establish a link to the approaches detailed in Sections 2.1.2 and 2.1.3. Let $N_n(A)$ be the number of points in A, where $N_n(A) \to N(A) \sim \text{Poi}\{\Lambda(A)\}$, as $n \to \infty$. The event $\{(M_n - b_n)/a_n \leq z\}$ is equivalent to $N_n(A_z) = 0$ for $A_z = (0,1) \times (z, \infty)$. It follows that

$$\begin{aligned}\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \Pr\{N_n(a_z) = 0\} \\ &\to \Pr\{N(A_z) = 0\} = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},\end{aligned}$$

as $n \to \infty$, and so equality with (2.1.1) is achieved.

A similar approach can be used to show equivalence with the threshold exceedance approach; first, we factorise $\Lambda(A_z)$ into $\Lambda_1([t_1, t_2]) \times \Lambda_2([z, \infty))$ where $\Lambda_1([t_1, t_2]) = (t_2 - t_1)$ and $\Lambda_2([z, \infty)) = \left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}$. It follows that

$$\begin{aligned}\Pr\left\{\left(\frac{X_i - b_n}{a_n} \geq z\right) \middle| \left(\frac{X_i - b_n}{a_n} > u\right)\right\} \to \frac{\Lambda_2([z,\infty))}{\Lambda_2([u,\infty))} &= \frac{[1 + \xi(z-\mu)/\sigma]^{-1/\xi}}{[1 + \xi(u-\mu)/\sigma]^{-1/\xi}} \\ &= \left[1 + \xi\left(\frac{z-u}{\tilde{\sigma}}\right)\right]^{-1/\xi},\end{aligned}$$

as $n \to \infty$ and for $\tilde{\sigma} = \sigma + \xi(u - \mu)$. Although equivalence is achieved between the models themselves, there are differences in how inference is conducted. For the approach described in Section 2.1.3, the probability of exceeding the threshold u and the two model parameters are estimated separately; for the point process representation given above, estimation of the threshold exceedance probability is incorporated into the inference for the three model parameters. The latter approach is often favourable over the former as its parameter values are independent of the threshold u; this is particularly advantageous when incorporating covariate effects into the parameters.

2.2 Multivariate extreme value theory

2.2.1 Multivariate ordering

Unlike in the univariate case, there exists no natural ordering of variables to define extremes in a multivariate setting. Barnett (1976) suggest several methods for defining multivariate extremes; we focus on the two most commonly studied.

A natural extension of the univariate block-maxima approach is component-wise maxima. We let $\mathbf{X}_i = \{X_{i,1}, \ldots, X_{i,d}\}$ for $i = 1, \ldots, n$ be a sequence of independent d-dimensional random vectors with common distribution function F, then

$$\mathbf{M}_n = \left(\max_{1 \leq i \leq n} X_{i,1}, \ldots, \max_{1 \leq i \leq n} X_{i,d} \right) \tag{2.2.1}$$

is the vector of component-wise maxima. Note that the indices for which the components attain their respective maxima need not be equal, i.e., $\arg\max_{1 \leq i \leq n} X_{i,j}$ is not necessarily equal to $\arg\max_{1 \leq i \leq n} X_{i,k}$ where $j \neq k$. This implies that the vector $\mathbf{M}_n$ does not necessarily correspond to an observed vector of the original data, which we illustrate in Figure 2.2.1. We discuss the limiting distribution of (2.2.1) as $n \to \infty$ in Section 2.2.3.

Alternative characterisations for multivariate extremes can be conducted by considering the distribution of $\mathbf{X}$ conditioned on some event defined through a risk functional $l : \mathbb{R}^d \to \mathbb{R}$. The univariate random variable $l(\mathbf{X})$ summarises characteristics of the random vector $\mathbf{X}$, and so $\mathbf{X}|(l(\mathbf{X}) > v)$, for some $v \in \mathbb{R}$, is the variable of interest. The form of $l(\cdot)$ is usually specific to the application; Coles and Tawn (1994) provide some natural examples of $l(\mathbf{X})$ which correspond to combinations of variables that cause structural failure in certain applications, e.g., offshore structures, river-bank flood defences. A pertinent choice for $l(\cdot)$ is a weighted sum of the components of $\mathbf{X}$, which has application in extreme precipitation and flooding models, as $l(\mathbf{X})$ can

denote the total volume of rain over a river catchment area; models for this particular case are discussed in Sections 2.3.6 and 2.3.5. Another example is $l(X) = \min_i X_i$, equivalent to considering $\mathbf{X}|\cup_i (X_i > v)$, which can be seen to be a natural extension of the peaks-over-threshold modelling approach; this scenario is considered in Section 2.2.5.

Illustration of two approaches to defining multivariate extreme events is illustrated in Figure 2.2.1. Before detailing models for these cases, we proceed by describing a general approach to modelling dependence in multivariate random vectors, which is done using copula models. Note that as the $j-$th component of $\mathbf{X}$ is a sequence of i.i.d. univariate random variables, the results discussed in Section 2.1 are applicable to their marginal behaviour.

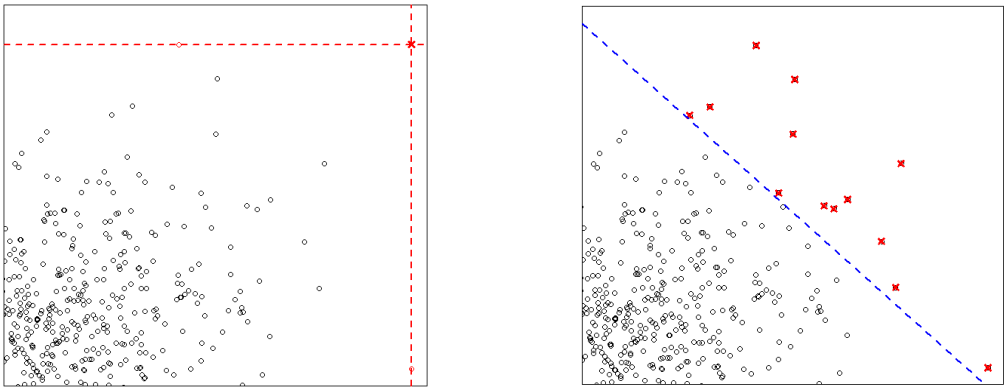

Figure 2.2.1: Examples of classifying bivariate extremes. The left panel gives componentwise maxima, where the red cross corresponds to the vector $\mathbf{M_n}$. The right panel illustrates the use of a structure variable, where the red crosses correspond to the vector $\{(X_1, X_2) : X_1 + X_2 \geq m\}$ and the blue line is $X_1 + X_2 = m$.

2.2.2 Copula based modelling

A valid and oft used approach for modelling multivariate extremes utilises a copula to describe dependence between univariate variables (Joe, 1997; Nelsen, 2006). Marginal

modelling and dependence modelling is conducted separately, which can lead to faster and more intuitive inference. All dependence is encapsulated in the copula, which is established by Sklar's Theorem (Sklar, 1959).

Theorem 2.2.1 (Sklar's Theorem)**.** *If* $\mathbf{X} = (X_1, \ldots, X_d)$ *has joint distribution* F*, and* $X_i \sim F_i$ *for* $i = 1, \ldots, d$ *and each* F_i *are continuous, then there exists a unique copula* C *such that*

$$F(\mathbf{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}.$$

The copula C is a multivariate distribution function $C : [0,1]^d \to [0,1]$ on standard uniform margins. A copula can incorporate any marginal distributions and preserves the dependence in $\mathbf{X}$ if the margins are transformed; this can be achieved using the probability integral transform. That is, if X_i is a continuous random variable with distribution function F_i, and its inverse distribution function F_i^{-1} exists, then $U = F_i(X) \sim \text{Unif}(0,1)$ and $F_i^{-1}(U) \sim F_i$. For example, if $C_U = C$ is a copula on uniform margins, a copula C_F on standard Frechét margins is simply

$$F(\mathbf{x}) = C_F\{-1/\log F_1(x_1), \ldots, -1/\log F_d(x_d)\}.$$

Similarly, a copula on standard exponential margins would be

$$F(\mathbf{x}) = C_E\{-\log[1 - F_1(x_1)], \ldots, -\log[1 - F_d(x_d)]\},$$

and further examples can be found using the corresponding inverse distribution functions for the desired margins.

Two examples of widely used parametric copulae are the logistic, or Gumbel, copula (Gumbel, 1960), given by

$$C_F(\mathbf{x}) = \exp\left\{-\left[\sum_{j=1}^{d} x_j^{-1/\alpha}\right]^{\alpha}\right\} \tag{2.2.2}$$

for $\alpha \in (0, 1]$, and the Gaussian copula,

$$C_U(\mathbf{x}) = \int_{-\infty}^{\Phi^{-1}(x_1)} \cdots \int_{-\infty}^{\Phi^{-1}(x_d)} \phi_d(\mathbf{y}; \Sigma) \mathrm{d}\mathbf{y}, \tag{2.2.3}$$

where $\phi_d(\cdot; \Sigma)$ denotes the standard $d-$dimensional Gaussian density with correlation matrix Σ, which determines the dependence. Dependence in (2.2.2) is determined by α: if $\alpha = 1$, we have independence and as $\alpha \to 0$, we have perfect dependence.

Results in multivariate extreme value theory are often given in standardised margins; for example, we consider the multivariate max-stable distribution in Section 2.2.3 which is given on standard Fréchet margins, and Heffernan and Tawn (2004) use standard Gumbel margins for their conditional extremes approach, which is discussed in Section 2.4.2. Different margins are often used to accentuate different properties of X; Fréchet margins accentuate the dependence in the largest values, whilst Gaussian margins accentuate dependence in the body. This is illustrated in Figure 2.2.2, where we plot transformations of simulated data from a standard bivariate Gaussian distribution.

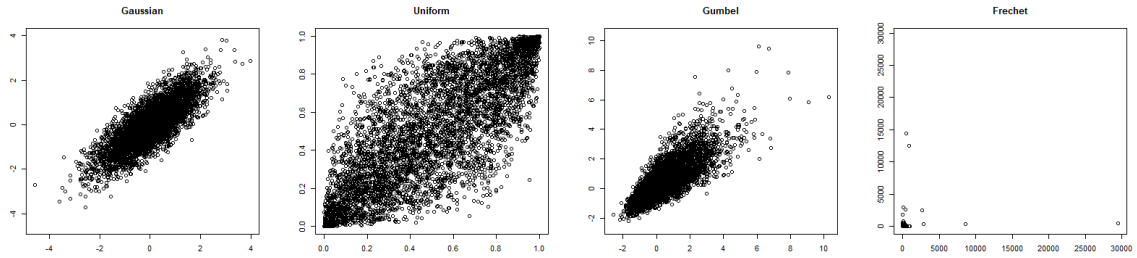

Figure 2.2.2: Illustration of marginally transformed data. The parent distribution is the left panel, which is standard Gaussian with correlation 0.8.

2.2.3 Multivariate extreme value distribution

We now consider componentwise maxima as defined in (2.2.1). The probability distribution for $\mathbf{M}_n$ can be derived exactly by considering an i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$

with common distribution $F(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$. Then

$$\Pr(\mathbf{M}_n \leq \mathbf{x}) = \Pr(\mathbf{X}_1 \leq \mathbf{x}, \ldots, \mathbf{X}_n \leq \mathbf{x}) = F^n(\mathbf{x}),$$

with vector operations taken componentwise above and thereafter. As in the univariate case discussed in Section 2.1.2, the distribution F^n is not typically used in practice as F is unknown and F^n is potentially degenerate for $n \to \infty$. We instead detail a normalisation of $\mathbf{M}_n$ for which the limiting distribution, as $n \to \infty$, is the multivariate extreme value distribution; the multivariate analogue of (2.1.2), which is reviewed in (Beirlant et al., 2006, Chapter 8).

Assume that there exist vectors $\mathbf{a}_n = (a_{n1}, \ldots, a_{nd}) > \mathbf{0}$ and $\mathbf{b}_n$, such that

$$\Pr\left\{\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right\} \to G(\mathbf{x}), \quad \text{as } n \to \infty, \tag{2.2.4}$$

for some d-dimensional distribution G, which is non-degenerate in each marginal. Similarly to the univariate case, we say that F is in the MDA of G if the limit in (2.2.4) holds. Here G is termed the multivariate extreme value (MEV) distribution, and can be considered as a copula, i.e.,

$$G(x_1, \ldots, x_d) = C^{MEV}(G_1(x_1), \ldots, G_d(x_d)), \tag{2.2.5}$$

where G_j, for $j = 1, \ldots, d$, is the distribution function of a GEV(μ_j, σ_j, ξ_j) random variable. Following the discussion in Section 2.2.2, we standardise the margins of this copula; without loss of generality, we rewrite (2.2.5) as $G(\mathbf{x}) = C_F^{MEV}(\mathbf{x}) = \exp\{-V(\mathbf{x})\}$, which has standard Fréchet margins, i.e,

$$G(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty) = \Pr\{X_j \leq x_j\} = \exp\left\{-\frac{1}{x_j}\right\}$$

for $x_j > 0$ and all $j = 1, \ldots, d$. We now discuss properties of the exponent measure

V, which is used to model dependence in $\mathbf{X}$.

Analogously to the univariate case, we require that G must satisfy max-stability, which implies that the function V must satisfy certain criteria. A multivariate distribution is said to be max-stable if, for all $N \in \mathbb{N}$, there exists vectors $\mathbf{A}_N > \mathbf{0} = (0, \ldots, 0)$ and $\mathbf{B}_N$, such that

$$G^N(\mathbf{x}) = G(\mathbf{A}_N\mathbf{x} + \mathbf{B}_N). \tag{2.2.6}$$

A random vector follows an MEV distribution if and only if its distribution function satisfies max-stability, which implies that V is an homogeneous function of order -1, i.e., $V(sx_1, \ldots, sx_d) = s^{-1}V(\mathbf{x})$ for all $s > 0$. Furthermore, the copula $C_F^{MEV} : \mathbb{R}^d \to [0, 1]$ must satisfy all properties of a valid probability distribution function. Pickands (1981) illustrates that these two conditions are met if and only if

$$V(\mathbf{x}) = d \int_{\mathcal{S}_{d-1}} \max_{i=1,\ldots,d} \left\{ \frac{w_i}{x_i} \right\} \mathrm{d}H(\mathbf{w}), \tag{2.2.7}$$

where $\mathcal{S}_{d-1} = \{\mathbf{w} \in [0, 1]^d : \sum_{i=1}^d w_i = 1\}$ is a $(d-1)$-dimensional unit simplex, and H is termed the spectral measure and satisfies

$$\int_{\mathcal{S}_{d-1}} \mathrm{d}H(\mathbf{w}) = 1, \quad \text{and} \quad \int_{\mathcal{S}_{d-1}} w_i \mathrm{d}H(\mathbf{w}) = 1/d \ \text{ for } \ i = 1, \ldots, d, \tag{2.2.8}$$

i.e., H is a valid probability distribution, or measure, on $\mathcal{S}_{d-1}$. Special cases of dependence in $\mathbf{X}$ arise when H is a discrete measure, e.g., if H puts equal mass $1/d$ on the boundaries of $\mathcal{S}_{d-1}$, this leads to $V(x_1, \ldots, x_d) = x_1^{-1} + \cdots + x_d^{-1}$, i.e., independence between all components of $\mathbf{X}$. Alternatively, if H place all mass along the diagonal of $\mathcal{S}_{d-1}$, i.e., $H(\{1/d\}, \ldots, \{1/d\}) = 1$, this leads to $V(\mathbf{x}) = \max\{x_1^{-1}, \ldots, x_d^{-1}\}$,which corresponds to perfect dependence between all components of $\mathbf{X}$.

If H is a differentiable distribution function with valid density h, then Coles and

Tawn (1991) provide an alternate link between V and H, i.e.,

$$\frac{\partial V}{\partial x_1 \dots \partial x_d} = -\frac{1}{\left(\sum_i x_i\right)^{d+1}} h\left(\frac{\mathbf{x}}{\sum_i x_i}\right).$$

Using this link, it is possible to derive classes of parametric forms for V through careful specification of h. One notable example is the Gumbel, or logistic copula, given in (2.2.2), which follows by specifying

$$h(\mathbf{w}) = \frac{1}{d}\left\{\prod_{i=1}^{d-1}\left(\frac{j}{\alpha} - 1\right)\right\}\left(\prod_{i=1}^{d} w_i\right)^{-1/\alpha-1}\left(\sum_{i=1}^{d} w_i^{-1/\alpha}\right)^{\alpha-d},$$

for $\mathbf{w} \in \mathcal{S}_{d-1}$ and $\alpha \in (0, 1]$. The limiting value of α leads to the boundary cases for H detailed above; for $\alpha = 1$ and $\alpha \to 0$, we induce independence and perfect dependence, respectively. Other examples for V include the asymmetric logistic, negative logistic and Hüsler-Reiss distribution, for which forms were first proposed by Tawn (1990), Galambos (1975) and Hüsler and Reiss (1989), respectively.

2.2.4 Regular variation

Resnick (1987) details an alternative approach to characterising the extremal dependence in a random vector $\mathbf{X}$ in terms of pseudo-radial and -angular components $(R, \mathbf{W})$. Let $\mathbf{X} \in \mathbb{R}_+^d$ have common marginals satisfying $\Pr\{X_j > x\} \sim cx^{-1}$, as $x \to \infty$, for $j = 1, \dots, d$ and for some $c > 0$. With $\|\cdot\|$ the L_1 norm, we define $(R, \mathbf{W})$ as

$$R = \|\mathbf{X}\| \quad \text{and} \quad \mathbf{W} = \frac{\mathbf{X}}{\|\mathbf{X}\|},$$

and it follows that $R > 0$ and $\mathbf{W} \in \mathcal{S}_{d-1} = \{\mathbf{w} \in [0, 1]^d : \sum_{i=1}^d w_i = 1\}$; the $(d-1)$-dimensional unit simplex. The vector $\mathbf{X}$ is said to be regularly varying if, for $r \geq 1$,

$$\lim_{t\to\infty} \Pr\{R > tr, \mathbf{W} \in \mathcal{B} | R > t\} = H(\mathcal{B})r^{-1}, \tag{2.2.9}$$

where $\mathcal{B} \subset \mathcal{S}_{d-1}$ is a measurable set and the spectral measure H satisfies the constraints detailed in (2.2.8). We note that (2.2.9) states that under the assumption of regular variation, that we have independence of R and $\mathbf{W}$ in the limit; and so for the largest magnitude events, i.e., large R, the distribution of mass on $\mathcal{S}_{d-1}$ controls the extremal dependence of $\mathbf{X}$. de Haan et al. (1984) illustrate that the distribution of $\mathbf{X}$ is in the MDA of the multivariate extreme value distribution if $\mathbf{X}$ is regularly varying.

2.2.5 Multivariate peaks-over-threshold

In Section 2.2.1, we briefly discussed the use of risk functionals for modelling the distribution of $\mathbf{X}$, given that some univariate function $l(\mathbf{X})$ exceeds a threshold; this can be regarded as a multivariate analogue of the univariate peaks-over-threshold approach to modelling extreme events, described in Section 2.1.3. Another analogue is the multivariate GPD distribution proposed by Rootzén and Tajvidi (2006) with further work by Rootzén et al. (2018b), Rootzén et al. (2018a) and Kiriliouk et al. (2019).

Using the notation in Section 2.2.3 and assuming that the limit in (2.2.4) holds, we define $\mathbf{x}_G \in (-\infty, 0]^d$ as the vector of lower endpoints of the marginal distributions $G_1, \ldots, G_d$. Rootzén et al. (2018b) show that, as $n \to \infty$,

$$\max\left\{\frac{\mathbf{X} - \mathbf{b}_n}{\mathbf{a}_n}, \mathbf{x}_G\right\} \Bigg| (\mathbf{X} \not\leqslant \mathbf{b}_n)$$

converges in distribution to the random vector $\mathbf{Y}$, which follows a multivariate generalised Pareto distribution. We denote the CDF of $\mathbf{Y}$ by $G^*(y_1, \ldots, y_d)$ and its marginals $G_i^*(y_i)$ for $i = 1, \ldots, d$; we can link G^* and the corresponding G in (2.2.4) through the expression

$$G^*(\mathbf{y}) = \frac{\log G(\min\{y_1, 0\}, \ldots, \min\{y_d, 0\}) - \log G(\mathbf{y})}{\log G(\mathbf{0})}, \tag{2.2.10}$$

which implies that the dependence in G^* is determined by that of G. Whilst we do not detail models for the dependence structure in G^* here, we do discuss some interesting properties that it holds.

We begin by noting that the marginals, $G_1^*, \ldots, G_d^*$, need not be univariate GPD, as the conditioning of the events $\{X_j > b_{n,j}\}$ and $\{\mathbf{X} \not\leqslant \mathbf{b}_n\}$ are not equivalent. However, conditioning on the marginal distributions being strictly positive does yield GPD margins; that is,

$$\Pr\{Y_j > y | Y_j > 0\} = 1 - \left(1 + \xi_j \frac{y}{\sigma_j}\right)_+^{-1/\xi_j},$$

for $\sigma_j > 0$ and for all $j = 1, \ldots, d$.

Furthermore, G^* satisfies a multivariate analogue of threshold stability (see Section 2.1.3). Formally, if $\mathbf{Y} \sim G^*$ and we have $\mathbf{u} \geq \mathbf{0}$, $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{u} > \mathbf{0}$ and $G^*(\mathbf{u}) < 1$, then the distribution of $(\mathbf{Y} - \mathbf{u})|(\mathbf{Y} \not\leqslant \mathbf{u})$ is also multivariate GPD with the same marginal shape parameters $\boldsymbol{\xi}$ as G^*, but with translated scale parameters $\boldsymbol{\sigma} + \boldsymbol{\xi}\mathbf{u}$.

2.2.6 Extremal dependence

A particularly important consideration for modelling multivariate, or spatial, extremes is the notation of extremal dependence, i.e., the tendency of variables to concurrently experience extreme events. Extremal dependence within a random vector is often described by one of two classes: asymptotic dependence, for which there is a non-zero probability of the most extreme events occurring together, and asymptotic independence; here our variables of interest may exhibit positive association, but their most extreme values do not occur simultaneously. Limiting sub-classes of these two dependence types include perfect dependence and independence, respectively.

We formally define asymptotic dependence and asymptotic independence using some commonly used measures in the literature; Coles et al. (1999) detail these for

bivariate data. The extremal dependence models discussed in Section 2.2.3 and 2.2.5 are designed for modelling asymptotic dependence only. The copula models described in Section 2.2.2 can be used for modelling asymptotic independence and this is discussed in a spatial setting in Section 2.3.3.

For a bivariate vector (X_1, X_2) with arbitrary margins, asymptotic dependence can be quantified through the upper tail index, $\chi \in [0, 1]$ (Joe, 1997), where

$$\chi := \lim_{q \uparrow 1} \chi(q), \quad \text{with} \quad \chi(q) = \Pr\{X_1 > F_1^{-1}(q) | X_2 > F_2^{-1}(q)\}, \tag{2.2.11}$$

and F_1 and F_2 the distribution functions of X_1 and X_2, respectively. In practice, χ cannot be estimated in the limit as $q \uparrow 1$; instead, estimates are provided by fixing high $q < 1$ and approximating χ via $\chi(q)$. Assessing the choice of q can be achieved through the use of threshold stability plots, similar to those described in Section 2.1.3.

Estimates of $\chi > 0$ suggest asymptotic dependence in (X_1, X_2), with strengthening dependence as $\chi \to 1$. For $\chi = 1$ and $\chi = 0$, we have perfect dependence and asymptotic independence, respectively, in (X_1, X_2). Theoretical values of χ for distributions can often be derived; consider a bivariate extreme value distribution for (X_1, X_2) with distribution function $G(x_1, x_2) = \exp\{-V(x_1, x_2)\}$, with V defined in (2.2.7). It can be shown that $\chi = 2 - V(1, 1)$, e.g., the logistic copula in (2.2.2) has $\chi = 2 - 2^{\alpha}$ for $\alpha \in (0, 1]$. Furthermore, it can be shown that $V(1, 1) < 2$, except in the special case of independence; it follows that $\chi > 0$ for all $(X_1, X_2) \sim G$ that exhibit positive association and thus all other bivariate extreme value distributions are asymptotically dependent. We further note that $\chi = 0$ for the bivariate Gaussian distribution with $\rho < 1$, implying that data of this type are asymptotically independent.

The measure χ gives a simple summary measure of extremal dependence within the class of asymptotically dependent distributions, but it fails to provide any measure of discrimination for asymptotically independent variables; thus we require another measure alongside χ. Coles et al. (1999) define the coefficient of asymptotic indepen-

dence as the measure $\bar{\chi} \in [-1, 1]$, with $\bar{\chi} := \lim_{q \uparrow 1} \bar{\chi}(q)$ where

$$\bar{\chi}(q) = \frac{2 \log \Pr\{X_1 > F_1^{-1}(q)\}}{\log \Pr\{X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q)\}} - 1.$$

This measure has the following properties: if (X_1, X_2) are asymptotically dependent, then $\bar{\chi} = 1$ and the strength of the extremal dependence between X_1 and X_2 decreases with $\bar{\chi}$; a value of $\bar{\chi} = 0$ corresponds to near extremal independence. Tail dependence can be completely summarised using the pair of measures $(\chi, \bar{\chi})$. For asymptotic dependence, we have $(\chi > 0, \bar{\chi} = 1)$ where χ increases with the level of asymptotic dependence. Conversely, we have $(\chi = 0, \bar{\chi} \leq 1)$ for the class of asymptotically independent variables; and $\bar{\chi}$ increasing with the strength of extremal dependence.

An alternative measure for characterising asymptotic independence, provided by Ledford and Tawn (1996), is the coefficient of tail dependence $0 < \eta \leq 1$. This measure is defined through the assumption on the joint tail distribution that

$$\Pr\left\{F_1(X_1) > 1 - \frac{1}{u}, F_2(X_2) > 1 - \frac{1}{u}\right\} = L(u)u^{1/\eta}, \text{ as } u \to \infty, \tag{2.2.12}$$

where L is a slowly-varying function[1]. It can be shown that $\bar{\chi} = 2\eta - 1$, but we detail both as η plays a particularly vital role in Chapters 3 and 4.

We now detail measures that characterise extremal dependence in higher dimensional settings, and so consider now $\mathbf{X} = (X_1, \ldots, X_d)$. Schlather and Tawn (2003) propose a measure for quantifying multivariate dependence using the extremal coefficient $\theta_d \in [1, d]$; in the bivariate case, the subscript d is often dropped. Assuming that all marginals of $\mathbf{X}$ are unit Fréchet, they illustrate that

$$\Pr\{X_1 \leq z, \ldots, X_d \leq z\} = \exp\left(-\frac{V(1, \ldots, 1)}{z}\right) = \exp\left(-\frac{\theta_d}{z}\right) = \left[\exp\left(-\frac{1}{z}\right)\right]^{\theta_d},$$

[1] $L(x)$ satisfies $L(cx)/L(x) \to 1$ as $x \to \infty$ for any fixed $c > 0$.

which follows as V is homogeneous of order -1. The measure can be interpreted as the effective number of independent marginal variables in $\mathbf{X}$; for $\theta_d = 1$ and $\theta_d = d$ we have complete dependence and independence, respectively, between all d marginals.

Given $\mathcal{D}$ as the set of all possible subsets of $\{1, \ldots, d\}$ of cardinality at least equal to two, Wadsworth and Tawn (2013) provide a multivariate extension of χ to d dimensions, which we denote $\chi(D)$ for $D \in \mathcal{D}$. Asserting that $\mathbf{X}$ has arbitrary marginals with distribution functions $F_1, \ldots, F_d$, then the d-dimensional joint tail dependence for marginals of $\mathbf{X}$ indexed by D is

$$\chi(D) = \lim_{q \uparrow 1} \Pr\{X_j > F_j^{-1}(q), \forall j \in D | X_i > F_i^{-1}(q), i \in D\}.$$

Wadsworth and Tawn (2013) refer to the cases $\chi(D) > 0$ and $\chi(D) = 0$ as strong joint tail dependence and weak joint tail dependence, respectively, between variables $\{X_j : j \in D\}$.

The d-dimensional extension of η, denoted $\eta_D \in (0, 1]$, is defined by Eastoe and Tawn (2012) and describes extremal dependence amongst a subset of components of $\mathbf{X}$; they make the assumption that

$$\Pr\left(\min_{i \in D}\{F_i(X_i)\} > 1 - \frac{1}{u}\right) = L_D(u) u^{-1/\eta_D}, \text{ as } u \to \infty, \tag{2.2.13}$$

for a slowly-varying function L_D. If there is asymptotic dependence between all components, i.e., $X(D) > 0$ for all $D \in \mathcal{D}$, then $\eta_D = 1$. The cases $1/|D| < \eta_D < 1$, $\eta_D = 1/|D|$ and $0 < \eta_D < 1/|D|$ correspond to positive extremal association, near extremal independence and negative extremal association, respectively.

2.3 Spatial extremes

2.3.1 Overview of spatial statistics

Before detailing models for the extremes of spatial processes, we provide a brief overview of some fundamental concepts of spatial statistics; this material is covered by Diggle and Ribeiro (2007).

We define a stochastic process $\{X(s) : s \in \mathcal{S}\}$ for some index set $\mathcal{S} \subseteq \mathbb{R}^p$ for $p \in \mathbb{N}$. When considering spatial processes, it is often appropriate to take $p = 2$ as this corresponds to (x, y) or (lon, lat) coordinates, and so we proceed as such. Typically, data that are available for inference are realisations $\mathbf{X}_t = (X_t(s_1), \ldots, X_t(s_d))$ for times $t = 1, \ldots, n$ and are treated as observations of said process $\{X(s)\}$ at sampling locations $(s_1, \ldots, s_d)$ for $d \in \mathbb{N}$. Inference on $\mathbf{X}_t$ can be made by specification of some distribution F to describe the characteristics of $\{X(s)\}$, such that for a finite collection of sites $(s_1, \ldots, s_d)$, we have $(X(s_1), \ldots, X(s_d)) \sim F$. A common specification for distribution F is a multivariate Gaussian, which gives rise to a Gaussian process.

Properties of F are often specified to be a function of $s \in \mathcal{S}$; that is, the statistical characteristics of a spatial process are dependent on location. For example, if $\{X(s)\}$ is a Gaussian process, then F is determined by a mean component $\mu(s) = \mathbb{E}[X(s)]$, a variance component $\sigma^2(s) = \mathrm{Var}\,(X(s))$, and a correlation function $\rho(s, s')$ which determines the dependence between $X(s)$ and $X(s')$ for $s, s' \in \mathcal{S}$.

A practical assumption that is often made about spatial processes is one of stationarity. If a Gaussian process $\{X(s)\}$ is second-order stationary, then its marginal characteristics are constant over $\mathcal{S}$ and correlation is a function of displacement $h = s - s'$; we can then rewrite the correlation function as $\rho(h)$ where $\rho(0) = 1$. A further assumption to make is that the process is isotropic, i.e., invariant to rotation or direction. In this case, we can replace displacement $s - s'$ with $\|s - s'\|$, where $\|\cdot\|$ denotes the Euclidean norm; this implies that correlation is a function of distance only, and

not dependent on the orientation of sites. Potential parametric forms for the correlation function include the Matérn, or powered exponential, families (Diggle and Ribeiro, 2007). The assumption of stationarity further implies that the variogram, denoted $\gamma(s, s') = \text{var}\left[X(s) - X(s')\right]/2$ for $s, s' \in \mathcal{S}$, can be written as a function of h; we mention the variogram as some dependence models are specified through its characteristics, rather than those of $\rho(\cdot)$.

Although Gaussian processes are well studied and full inference with them is computationally feasible, they may not be well suited for modelling the extreme characteristics of some spatial processes as they do not necessarily capture tail decay appropriately; Gaussian processes are inherently asymptotically independent, i.e., for any $s, s' \in \mathcal{S}$ we have $\chi = 0$ for $(X(s), X(s'))$, regardless of the distance $\|s - s'\|$. We follow with an overview of methods for statistical modelling of the extremes of spatial processes; a review of these methods is given by Davison et al. (2012) with recent advances detailed by Huser and Wadsworth (2020).

2.3.2 Max-stable processes

Max-stable processes concern the behaviour of spatial maxima and are widely applied in the literature, as they provide a convenient analogue of the multivariate extreme value distribution discussed in Section 2.2.3; a review of max-stable processes is provided by Ribatet (2013). Let $\{X_t(s) : s \in \mathcal{S}\}_{t=1}^{n}$ for index set $\mathcal{S} \subseteq \mathbb{R}^p,\ p \in \mathbb{N}$ be a sequence of n independent replications of a continuous stochastic process $\{X(\cdot)\}$. For suitable scaling functions $a_n(s) > 0$ and $b_n(s) \in \mathbb{R}$, which are continuous over all $s \in \mathcal{S}$, we define the spatial process of maxima as

$$\{Z_n(s) : s \in \mathcal{S}\} = \left\{ \max_{1 \leq t \leq n} \frac{X_t(s) - b_n(s)}{a_n(s)} : s \in \mathcal{S} \right\}. \tag{2.3.1}$$

This can be thought of as the point maxima of n independent and identically distributed copies of $\{X(s) : s \in \mathcal{S}\}$. Figure 2.3.1 illustrates an example of spatial maxima for a 1-dimensional process, i.e., $p = 1$.

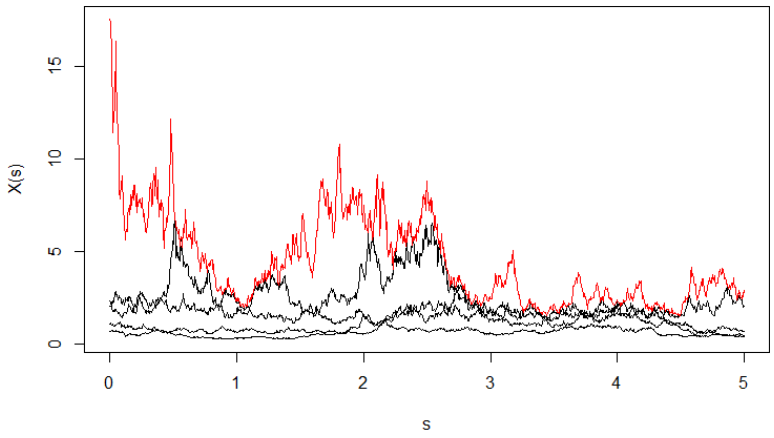

Figure 2.3.1: Illustration of spatial maxima for a 1-dimensional process. The red line gives the maxima over $n = 5$ replications. The underlying process is Gaussian with powered exponential correlation, with margins transformed to standard exponential.

Our interest lies in the limiting process $\{Z_n(s) : s \in \mathcal{S}\}$ as $n \to \infty$: if it exists and is non-degenerate in all marginals, then $\{Z(s)\} = \lim_{n\to\infty}\{Z_n(s)\}$ is a stationary max-stable process with GEV marginals. Similarly to Section 2.2, it is convenient to assert unit Fréchet margins for $\{Z(s)\}$ and so we do so throughout the remainder of this section. The joint distribution of $(Z(s_1), \ldots, Z(s_d))$ at a finite collection of sites $\mathbf{s} = (s_1, \ldots, s_d)$ is an MEV distribution, and so the distribution function is given by

$$\Pr\{Z(s_1) \leq z_1, \ldots, Z(s_d) \leq z_d\} = \exp\left(-V(z_1, \ldots, z_d)\right), \tag{2.3.2}$$

with exponent V, defined in (2.2.7); if $\{Z(s) : s \in \mathcal{S}\}$ is stationary, then V is a function of the pairwise distances between sites $\mathbf{s}$. Models for V arise through point process constructions of max-stable processes; we detail two here and describe some of the forms for V that they generate.

Smith (1990) propose the following construction for a max-stable process. Let $\{(W_i, R_i)\}_{i=1}^{\infty}$ be points of a Poisson process P on the space $\mathcal{S} \times \mathbb{R}_+$ for $\mathcal{S} \subseteq \mathbb{R}^p$, with intensity given by $\lambda(w, r) = \mathrm{d}w \times r^{-2}\mathrm{d}r$. With f a non-negative function on $\mathcal{S}$ such that $\int_{\mathcal{S}} f(s)\mathrm{d}s = 1$, then

$$\{Z(s) : s \in \mathcal{S}\} = \left\{ \sup_{(W,R)\in P} Rf(s - W) : s \in \mathcal{S} \right\} \tag{2.3.3}$$

is a stationary max-stable process with standard Fréchet margins. Smith (1990) provide a physical interpretation of (2.3.3) by suggesting that the function f defines the shape of an event, centred at W, and R describes its magnitude. Specification of f can lead to parametric models for $\{Z(s)\}$; Smith (1990) develops a model, usually termed the Smith process, by letting f be a $p-$variate Gaussian density with covariance matrix Σ. This leads to closed parametric form for the joint distribution (2.3.2) with exponent

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left\{\frac{a(h)}{2} + a(h)^{-1}\log\left(\frac{z_2}{z_1}\right)\right\} + \frac{1}{z_2}\Phi\left\{\frac{a(h)}{2} + a(h)^{-1}\log\left(\frac{z_1}{z_2}\right)\right\}, \tag{2.3.4}$$

where $a^2(h) = h^T\Sigma^{-1}h$ with $h = \|s_1 - s_2\|$, and Φ is the standard normal distribution function. Figure 2.3.2 illustrates a simulated Smith process; these are typically too smooth to provide realistic models for environmental data.

Schlather (2002) provides another characterisation of $\{Z(s)\}$ which gives rise to other commonly used models, and is detailed in Davison et al. (2012). Let $\{R_i\}_{i=1}^{\infty}$ be points of a Poisson process on $\mathbb{R}_+$ with intensity measure given by $r^{-2}\mathrm{d}r$ and let $\{W_i(s)\}_{i=1}^{\infty}$ be independent replicates of a stationary stochastic process $\{W(s) : s \in \mathcal{S}\}$ for $\mathcal{S} \subseteq \mathbb{R}^p$, that satisfies $\mathbb{E}[\max\{0, W_i(s^*)\}] = 1$ for any $s^* \in \mathcal{S}$. Then

$$\left\{Z(s) : s \in \mathcal{S}\right\} = \left\{\max_i R_i \max\{0, W_i(s)\} : s \in \mathcal{S}\right\}. \tag{2.3.5}$$

Similarly to (2.3.3), the stochastic process $\{W(s)\}$ describes the shape of an event with R being its magnitude. Specification of $W(s)$ leads to parametric forms for the exponent V in (2.3.2). For example, the Schlather process (Schlather, 2002) is derived by taking $W(s)$ to be a stationary standard Gaussian process with correlation function $\rho(h)$, scaled such that $\mathbb{E}[\max\{0, W_j(s^*)\}] = 1$. This gives rise to the exponent

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

An example realisation of a Schlather process can be found in Figure 2.3.2.

Other popular models that can arise from the representation given by (2.3.5) are the Brown-Resnick (Davis and Resnick, 1984; Kabluchko et al., 2009), and the extremal-t (Opitz, 2013), processes. The former takes $W(s) = \exp[\varepsilon_1(s) - \gamma(s)]$, where $\varepsilon_1(s)$ is an intrinsically Gaussian process with variogram $\gamma(h)$ and with $\varepsilon_1(o) = 0$ almost surely; here o denotes the origin. This is a generalisation of the Smith process, as its exponent is equal to (2.3.4) except with $a^2(h) = 2\gamma(h)$; note that if $\gamma(h) \propto h^2$, then the two processes are equivalent. Furthermore, if $\{Z(s)\}$ is a Brown-Resnick process, then $(Z(s_1), \ldots, Z(s_d))$ follows the Hüsler-Reiss distribution mentioned in Section 2.2.3. The extremal-t process generalises the Schlather process by replacing $\max\{0, W_i(s)\}$ in (2.3.5) with $c_\nu \max\{0, \varepsilon_2(s)\}^\nu$, where $c_\nu = \sqrt{\pi} 2^{-(\nu-2)/2} \Gamma\left((\nu+1)/2\right)^{-1}$ for $\nu \geq 1$ and Γ denotes the Gamma function; here ε_2 is a standard Gaussian process with correlation function ρ. The exponent function for this model is

$$V(z_1, z_2) = \frac{1}{z_1} T_{\nu+1}\left\{-\frac{\rho(s_2 - s_1)}{b} + b^{-1}\left(\frac{z_2}{z_1}\right)^{1/\nu}\right\} + \frac{1}{z_2} T_{\nu+1}\left\{-\frac{\rho(s_2 - s_1)}{b} + b^{-1}\left(\frac{z_1}{z_2}\right)^{1/\nu}\right\},$$

where T_ν denotes the CDF of a Student's t random variable with ν degrees of freedom and $b^2 = \{1 - \rho(s_2 - s_1)^2\}/(\nu + 1)$.

The extremal dependence measures in Section 2.2.6 can be naturally extended to a spatial setting. For example, replacing X_1 and X_2 in (2.2.11) with $Z(s_1)$ and

$Z(s_2)$, we can rewrite χ as $\chi(s_1, s_2)$, i.e., a function of $s_1, s_2 \in \mathcal{S}$. For stationary Z, we expect these measures to be a function of distance $h = \|s_1 - s_2\|$. Following the properties of V discussed in Section 2.2.6, max-stable processes exhibit asymptotic dependence, or independence, at all distances. That is, if $\{Z(s) : s \in \mathcal{S}\}$ exhibits positive spatial association, then $\chi(s_1, s_2) > 0$ for all $s_1, s_2 \in \mathcal{S}$, regardless of the value of h. This makes them an inappropriate choice for modelling environmental data that exhibits asymptotic independence. Wadsworth and Tawn (2012) describe an inverted max-stable process that exhibits asymptotic independence, detailed in Section 2.3.3.

2.3.3 Inverted max-stable processes

Following the approach of Wadsworth and Tawn (2012), let $\{Z(s) : s \in \mathcal{S}\}$ be a max-stable process with unit Fréchet margins, as defined in (2.3.5). Then $Y(s) = 1/Z(s)$ is the corresponding inverted max-stable process on standard exponential margins. The joint survivor function for $Y(s)$ is

$$\Pr\{Y(s_1) \geq y_1, \ldots, Y(s_d) \geq y_d\} = \exp\{-V(1/y_1, \ldots, 1/y_d)\}, \tag{2.3.6}$$

where V is the exponent for $Z(s)$, defined in (2.3.2). From (2.3.6), it can be shown that $\eta(s_1, s_2) = 1/V(1, 1)$ for all $s_1, s_2 \in \mathcal{S}$, i.e., unless Z is perfectly dependent, $Y(s)$ is asymptotically independent at all distances. The intuition behind this follows by noting that the reciprocal is a monotonically decreasing transformation, which means that the copula for $Z(\cdot)$ is inverted; the lower tails of the max-stable process become the upper tails of the inverted max-stable process, and vice-versa. As the lower tails of $Z(s)$ are asymptotically independent, so too are the upper tails of $Y(s)$.

Figure 2.3.2 illustrates two max-stable processes and their respective inverted max-stable counterparts; we observe that tail dependence is reversed by taking the inverse.

Any of the max-stable models in Section 2.3.2 can be transformed in this way, and will exhibit asymptotic independence.

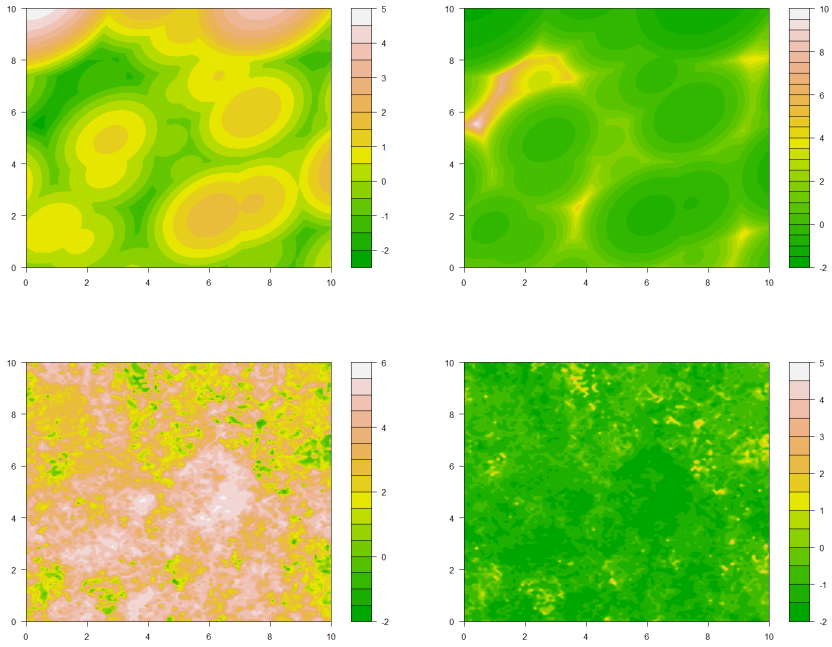

Figure 2.3.2: Simulated max-stable (left panels), and inverted max-stable (right panels), processes. Top row: Smith, with covariance matrix $\Sigma \in$ (2.3.4) with $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\sigma_{12} = 0.3$. Bottom row: Schlather process with powered exponential $\rho(h)$ with scale and shape parameters, 1 and 0.5, respectively. All processes are on Gumbel margins.

2.3.4 Alternatives to max-stable processes

Recent developments in the literature propose extensions of max-stable, and inverted max-stable, processes that are able to capture a much wider range of extremal dependence structures. We follow Huser and Wadsworth (2020) and give a brief overview of these models.

Wadsworth and Tawn (2012) propose a hybrid process, often termed a max-mixture, that takes point-wise maxima of a weighted mixture of an asymptotically dependent process, say $\{Z_1(s) : s \in \mathcal{S}\}$, and an asymptotically independent process $\{Z_2(s) : s \in \mathcal{S}\}$; both of which have common Fréchet margins. They define the

max-mixture process by $Z(s) = \max\{\alpha Z_1(s), (1-\alpha)Z_2(s)\}$ for $\alpha \in [0,1]$. The resulting process exhibits a mixture of the dependence in Z_1 and in Z_2, with the mixing proportion controlled by the weight α. We briefly consider a model of this type in Chapter 7.

Another proposed set of models for spatial extremes are random scale mixtures, which are of the form $X(s) = RW(s)$ for general $R > 0$ and W; processes of this type have been almost fully characterised in a bivariate setting by Engelke et al. (2019b). Models constructed from random scale mixtures can be particularly flexible as they can be constructed to be able to capture either class of extremal dependence. The heaviness of the tails of R relative to those of W is the driving factor of the dependence exhibited by the resulting $X(s)$ (Huser and Wadsworth, 2020). For example, if W is a standard Gaussian process and R has Pareto upper-tails, i.e., $\Pr\{R > r\} \sim Kr^{-\gamma}$ as $r \to \infty$ for $K > 0$ and $\gamma > 0$, then $X(s)$ is asymptotically dependent; for the same W and R with Weibull upper-tails, i.e., $\Pr\{R > r\} \sim Kr^{\alpha}\exp(-\theta r^{\beta})$ as $r \to \infty$ for constants $K, \theta, \beta > 0$ and $\alpha \in \mathbb{R}$, then $X(s)$ exhibits asymptotic independence. Thus, through careful model specifciation for R which allows for both heavy and light tail decay as sub-classes, a model can be constructed that is appropriate for data exhibiting either class of extremal dependence; an approach first considered by Huser et al. (2017). An alternative approach is proposed by Huser and Wadsworth (2019), where they let $X^*(s) = R^{\delta}W(s)^{1-\delta}$ for $\delta \in [0,1]$ with $R \geq 1$ a Pareto random variable and $W(s)$ an asymptotically independent process with unit Pareto margins. The relative heaviness of the tails of R and of $W(s)$ in this mixture is controlled by the parameter δ. Huser and Wadsworth (2019) show that if $\delta \leq 0.5$ then $X(s)$ is asymptotically independent, and asymptotically dependent otherwise.

The final class of models we discuss are max-infinitely divisible (max-id) processes, which were first considered in the context of modelling block-maxima by Padoan (2013). Whilst we omit full details for these processes, we note that they satisfy a

similar property to max-stability, namely max-infinite divisiblity. A d-dimensional distribution G is max-id if G^t is a valid distribution for all $t \in \mathbb{R}_+$. Every max-id distribution is max-stable, which can be seen by setting $t = N \in \mathbb{N}$ and considering (2.2.6); however, the reverse is not true. Thus, max-id processes can theoretically capture a wider range of dependence structures than max-stable processes, which includes asymptotic independence. Whilst the max-id model of Padoan (2013) does not allow for max-stable processes as a sub-class, recent advances by Bopp et al. (2020) and Huser et al. (2020) have developed models that accommodate this feature.

2.3.5 Pareto processes

Whilst max-stable processes can be considered to be the spatial extension of MEV random variables, the natural spatial analogue of multivariate generalised Pareto random variables is the generalised Pareto process. First considered by Buishand et al. (2008) with further details provided by Ferreira and de Haan (2014), these processes arise as the limiting distribution of

$$\left[1 + \xi(s)\left\{\frac{X(s) - b_n(s)}{a_n(s)}\right\}\right]_+^{1/\xi(s)} \,\middle|\, \left(\sup_{s\in\mathcal{S}} \tilde{X}_n(s) > 0\right), \quad \text{as} \quad n \to \infty, \tag{2.3.7}$$

where $\{\tilde{X}_n(s)\} = \{(X(s) - b_n(s))/a_n(s)\}$ for sequences $a_n(s) > 0$ and $b_n(s)$ as described in (2.3.1), and $\xi(s)$ is the shape parameter of the GEV distribution at site $s \in \mathcal{S}$. Although we do not discuss models for (2.3.7), we note that inference for Pareto processes is conducted in a finite-dimensional setting and so the modelling techniques for the multivariate Pareto distribution described in Section 2.2.5 can be applied here, i.e., the relationship described in (2.2.10) can be exploited. It follows that the Pareto process is asymptotically dependent at all distances; note that if the underlying process $X(s)$ is asymptotically independent, then the corresponding Pareto process is degenerate.

Recent extensions of generalised Pareto processes, termed $l-$Pareto processes by Dombry and Ribatet (2015) and $r-$Pareto processes by de Fondeville and Davison (2018), generalise the formulation in (2.3.7) by replacing the conditioning event $\sup_{s\in\mathcal{S}}\{\tilde{X}(s)\} > 0$ with the exceedance of a risk-functional l, i.e., $l(\{\tilde{X}(s)\}) > v$ for some $s \in \mathcal{S}$ and fixed $v \in \mathbb{R}$. Valid functionals must be homogeneous of order 1 when $\{\tilde{X}(s)\}$ has standardised Pareto-type marginals, and so possible examples include $\int_{\mathcal{A}} \tilde{X}(s)\mathrm{d}s$, $\inf_{s\in\mathcal{A}}\{\tilde{X}(s)\}$ or $\sup_{s\in\mathcal{A}}\{\tilde{X}(s)\}$ for $\mathcal{A} \subseteq \mathcal{S}$. Equivalent functionals can be derived for a finite dimensional domain $\mathcal{S}$, and extensions to accommodate other types of functionals have been recently proposed by de Fondeville and Davison (2020).

2.3.6 Aggregation of spatial processes

We now detail some of the approaches that have been developed for modelling the extremal behaviour of the aggregate of a spatial process over a region $\mathcal{A} \subset \mathcal{S}$, i.e., the upper tail behaviour of

$$R_i(\mathcal{A}) = \frac{1}{|\mathcal{A}|}\int_{\mathcal{A}} X_i(s)\mathrm{d}s,$$

where $|\mathcal{A}|$ denotes the area of $\mathcal{A}$.

Following a recharacterisation of the max-stable process construction in (2.3.3) by Coles (1993), Coles and Tawn (1996) derive results for the distribution of the maxima of the aggregate $\max_{1\leq i\leq n}\{R_i(\mathcal{A})\}$. They begin with the assertion that $\max_{1\leq i\leq n}\{X_i(s)\} \sim GEV\{\mu(s), \sigma(s), \xi(s)\}$ for each $s \in \mathcal{A}$ and then derive the CDF of $\max_{1\leq i\leq n}\{R_i(\mathcal{A})\}$ as a function of the marginal GEV parameters and the components of (2.3.3). We omit this full derivation here as in the general case it has no analytical solution; however, if the shape parameter is homogeneous over $\mathcal{A}$, i.e., $\xi(s) = \xi$ for all

$s \in \mathcal{A}$, then Coles and Tawn (1996) illustrate that

$$\Pr\left(\max_{i=1,\ldots,n}\{R_i(\mathcal{A})\} \leq r\right) = \begin{cases} \exp[-n\theta_{\mathcal{A}}\{1+\xi(r-\bar{\mu})/\bar{\sigma}\}_+^{-1/\xi}], & \text{if } \xi \neq 0, \\ \exp[-n\theta_{\mathcal{A}}\exp\{-(r-\bar{\mu})/\bar{\sigma}\}], & \text{if } \xi = 0, \end{cases} \tag{2.3.8}$$

where $\bar{\mu} = |\mathcal{A}|^{-1}\int_{\mathcal{A}}\mu(s)\mathrm{d}s$ and $\bar{\sigma} = |\mathcal{A}|^{-1}\int_{\mathcal{A}}\sigma(s)\mathrm{d}s$. The scaling factor $\theta_{\mathcal{A}}$ is termed the areal coefficient, and summarises the effect of extremal dependence over $\mathcal{A}$; following the notation in (2.3.3), it is defined as

$$\theta_{\mathcal{A}} = \begin{cases} (\bar{\sigma}|\mathcal{A}|)^{-1/\xi}\int_{\mathcal{A}}\left\{\int_{\mathcal{A}} f(s-w)^{\xi}\sigma(s)\mathrm{d}s\right\}^{1/\xi}\mathrm{d}w), & \text{if } \xi \neq 0, \\ \int_{\mathcal{A}}\exp\left\{(\bar{\sigma}|\mathcal{A}|)^{-1}\int_{\mathcal{A}}\log f(s-w)\sigma(s)\mathrm{d}s\right\}\mathrm{d}w, & \text{if } \xi = 0. \end{cases}$$

Further formalisation of this work is provided by Ferreira et al. (2012).

Extensions of this framework are given by Engelke et al. (2019a) who derive a similar coefficient to $\theta_{\mathcal{A}}$, which they denote θ^l. This approach differs slightly from that of Coles and Tawn (1996) as it links the distribution of $X(s^*)$ for any $s^* \in \mathcal{S}$ to the distribution of a general risk functional $l(\cdot)$ of the process; examples of these were given in Section 2.3.5. Engelke et al. (2019a) illustrate that, for any $s^* \in \mathcal{S}$ and $x \in \mathbb{R}$, then

$$\Pr\left[\frac{l(\{X(s)\}) - l(b_n(s))}{l(a_n(s))} > x\right] \approx \theta^l \Pr\left\{\frac{X(s^*) - b_n(s^*)}{a_n(s^*)} > x\right\}, \tag{2.3.9}$$

for sufficiently large n, and where the sequences $a_n(s) > 0$ and $b_n(s)$ are as described in (2.3.1). While we omit the form for θ^l, we note that equality with $\theta_{\mathcal{A}}$ can be achieved by considering $l(\{X(s)\}) = |\mathcal{A}|^{-1}\int_{\mathcal{A}} X(s)\mathrm{d}s$.

Alternative modelling approaches for $R_{\mathcal{A}}$ utilise the functional Pareto processes described in Section 2.3.5, see de Fondeville and Davison (2020) for one example. These, and the earlier approaches described, rely on models constructed from asymptotically

dependent processes, which means that they may be inappropriate for application to data that exhibits asymptotic independence. We detail an approach for modelling $R_{\mathcal{A}}$ in Chapter 5 which can accommodate the underlying marginal process $\{X(s)\}$ being asymptotically independent; this is based on the conditional extremes framework.

2.4 Conditional extremes

2.4.1 Overview

We now introduce a wholly different approach to modelling multivariate and spatial extremes. First proposed by Heffernan and Tawn (2004) with generalistions by Heffernan and Resnick (2007), the conditional extremes approach differs from the methods described in Sections 2.2 and 2.3 as it does not focus on the joint extremal behaviour of a random vector or process; rather, Heffernan and Tawn (2004) study the behaviour of a random vector by conditioning on one component being extreme and then modelling the vector with the conditioning variable removed. Inference for this model and its extensions are typically less computationally expensive than other models for multivariate extremes, and they prosper from the benefit of being able to model both asymptotic dependence and asymptotic independence in a parsimonious manner.

2.4.2 Multivariate conditional extremes

Suppose we have a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with exponential upper-tails for each margin, i.e., $\Pr\{X_j \geq x\} \sim C\exp(-x)$ as $x \to \infty$ for all $j = 1, \ldots, d$ and constant $C > 0$. We denote one of its components as a conditioning variable X_i for any $i = 1, \ldots, d$, and define $\mathbf{X}_{-i}$ as the vector $\mathbf{X}$ with the $i-$th component removed. Heffernan and Tawn (2004) define the vector of standardised residuals by $\mathbf{Z}_{|i} = [\mathbf{X}_{-i} - \mathbf{a}_{|i}(X_i)]/\mathbf{b}_{|i}(X_i)$, for normalising functions $\mathbf{a}_{|i}, \mathbf{b}_{|i} : \mathbb{R} \to \mathbb{R}^{d-1}$, and

assume that

$$\Pr(\mathbf{Z}_{|i} \leq \mathbf{z}_{|i} | X_i = x) \to G_{|i}(\mathbf{z}_{|i}), \tag{2.4.1}$$

as $x \to \infty$, where $G_{|i}$ is non-degenerate in each marginal. From (2.4.1), and for fixed $\mathbf{Z}_{|i}$ and $x > 0$, they show that

$$\Pr(\mathbf{Z}_{|i} \leq \mathbf{z}_{|i}, X_i - u > x | X_i > u) \to G_{|i}(\mathbf{z}_{|i}) \exp(-x), \tag{2.4.2}$$

as $u \to \infty$; that is, $\mathbf{Z}_{|i}$ and $X_i - u | X_i > u$ are independent in the limit as $u \to \infty$.

Modelling using (2.4.2) can be conducted by assuming that the limit holds for large fixed u, and with parametric forms given for the normalising functions and for $G_{|i}$; the former are typically simplified to location and scale parameter vectors given by $\mathbf{a}_{|i}(x) = \boldsymbol{\alpha}_{|i} x$ for $\boldsymbol{\alpha}_{|i} \in (0, 1]$, under positive dependence, and $\mathbf{b}_{|i}(x) = x^{\boldsymbol{\beta}_{|i}}$ for $\boldsymbol{\beta}_{|i} \leq 1$, and where operations are taken component-wise. Tail dependence between X_i and components of $\mathbf{X}_{-i}$ is characterised through the values of $\boldsymbol{\alpha}_{|i}$ and $\boldsymbol{\beta}_{|i}$. For example, asymptotic dependence between X_j and X_i is implied by values $\alpha_{j|i} = 1$ and $\beta_{j|i} = 0$. Within the class of asymptotically independent variables $\alpha_{j|i} < 1$, with $\alpha_{j|i} = \beta_{j|i} = 0$ giving near extremal independence. Usually $G_{|i}$ is modelled using a $(d-1)$-dimensional Gaussian copula with some chosen margins; Heffernan and Tawn (2004) use Gaussian, but recent extensions of this methodology to a spatial context use more flexible alternatives.

2.4.3 Spatial conditional extremes

A natural extension of the Heffernan and Tawn (2004) multivariate conditional extremes model to a spatial setting is given by Wadsworth and Tawn (2019); the general proposal replaces the random vectors described in Section 2.4.2 with corresponding spatial processes, and represents $\boldsymbol{a}_{|i}$ and $\boldsymbol{b}_{|i}$ as functions of space. They begin by defining $\{X(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}$ as a stationary spatial process with standard exponen-

tial upper-tailed marginals and condition that it is extreme at a specified site $s_O \in \mathcal{S}$. Under the assumption that there exists normalising function $\{a : (\mathbb{R} \times \mathbb{R}^2) \to \mathbb{R}\}$, with $a(x, 0) = x$, and $\{b : (\mathbb{R} \times \mathbb{R}^2) \to (0, \infty)\}$, such that as $u \to \infty$, Wadsworth and Tawn (2019) assume that for each $s_O \in \mathcal{S}$,

$$\left(\left\{ \frac{X(s) - a\{X(s_O), s - s_O\}}{b\{X(s_O), s - s_O\}} : s \in \mathcal{S} \right\}, X(s_O) - u \right) \Bigg| \Big(X(s_O) > u \Big) \xrightarrow{d} \left(\Big\{ Z(s|s_O) : s \in \mathcal{S} \Big\}, E \right), \tag{2.4.3}$$

where convergence is in the sense of finite-dimensional distributions, and E is a standard exponential variable that is independent of the process $\{Z(s|s_O)\}$, assumed to be non-degenerate for all $s \in \mathcal{S}$ where $s \neq s_O$. That is, there is convergence in distribution of the normalised process to $\{Z(s|s_O) : s \in \mathcal{S}\}$, termed the residual process, which is independent of E, and $Z(s_O|s_O) = 0$ almost surely. A discussion of modelling choices for the normalising functions a and b, and the residual process $Z(s|s_O)$ is given in Chapter 5.

The strength and class of extremal dependence between $X(s)$ and $X(s_O)$ can be determined by the corresponding values of a and b. Equivalence with previously discussed spatial processes can arise if a, b and $Z(s|s_O)$ take certain forms:

- **l-Pareto process**: If $a(x, s - s_O) = x$ and $b(x, s - s_O) = 1$ for $s, s_O \in \mathcal{S}$, then $X(s)$ and $X(s_O)$ are asymptotically dependent. If this is achieved for all $s_O \in \mathcal{S}$, then the formulation (2.4.3) is equivalent to an l-Pareto process with valid risk functional $l\{X(s_O)\} = X(s_O)$ (see Section 2.3.5).

- **Gaussian process**: If $X(s)$ is a Gaussian process with correlation function $\rho(\cdot) \geq 0$ and $a(x, s - s_O) = \rho(s - s_O)^2 x$ and $b(x, s - s_O) = 1 + \rho(s - s_O)x^{1/2}$, then $Z(s|s_O)$ is a zero mean Gaussian process, subject to the condition that $Z(s_O|s_O) = 0$. As discussed previously, this process is asymptotically indepen-

dent at all distances.

- **Mixture process**: If $b(x, s - s_O) = 1$ for all $s, s_O \in \mathcal{S}$, and $a(x, s - s_O) = x$ for $\|s - s_O\| \leq \tau$ and $0 \leq \alpha(x, s - s_O) < x$ otherwise, then the process is asymptotically dependent up to distance τ from the conditioning site s_O, and asymptotically independent thereafter.

3

On the tail behaviour of sums of random variables

3.1 Introduction

The extremal behaviour of aggregated data is of importance in two key areas of risk management; financial portfolio optimisation and fluvial flooding. In financial risk management, it is standard practice to aggregate over returns from several assets in a portfolio in an attempt to mitigate investment risk. It is important that the uncertainty surrounding the tail behaviour of the aggregate is assessed so that the risk of large negative cumulative returns can be quantified (Hauksson et al., 2001; Chen et al., 2012; Embrechts et al., 2015; Kole et al., 2017; Bernard et al., 2018). For flood risk management, consider that fluvial floods are typically caused by prolonged extreme precipitation over a catchment area; more succinctly, precipitation aggregated both spatially and temporally (Coles and Tawn, 1996; Sangati and Borga, 2009; Spekkers et al., 2013; Eggert et al., 2015; Morbidelli et al., 2018). In both cases, the assumption of independence within the multivariate variable of interest is unlikely to hold; we derive the first order behaviour of the upper-tail of a weighted sum of a bivari-

ate random vector with different marginal tail behaviours and extremal dependence structures and demonstrate that both factors have a significant effect on the extremal behaviour of the aggregate variable.

We define the aggregate R as a weighted sum of the components of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$, with marginal distribution functions F_i for $i = 1, \ldots, d$, as

$$R = \sum_{i=1}^{d} \omega_i X_i, \tag{3.1.1}$$

with weights $\boldsymbol{\omega} = \{\omega_i; 0 \leq \omega_i \leq 1, \sum_{i=1}^{d} \omega_i = 1\}$, and where components of $\mathbf{X}$ are all positive and not necessarily independent and identically distributed and $\mathbf{X}$ has a joint density. Dependence between components can be described using copulae, see Sklar's theorem (Nelsen, 2006). The joint distribution function of $\mathbf{X}$ can be uniquely written as

$$F(\mathbf{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \quad \mathbf{x} \in \mathbb{R}^d$$

where C is the copula, i.e., some multivariate distribution function $C : [0,1]^d \to [0,1]$ on uniform margins. Our interest lies in the tail behaviour of R, which we quantify by considering $\Pr\{R \geq r\}$ as $r \to r^F$, where $r^F \leq \infty$ is the upper-endpoint of R, and how this behaviour is driven by the marginal tails and dependence structure of $\mathbf{X}$. Modelling the marginal tails of a random vector $\mathbf{X}$ has been widely studied, see Pickands (1975); Davison and Smith (1990) and Coles (2001). The typical approach is to assume that there exists a threshold u_i for each X_i, such that the distribution of $(X_i - u_i)|(X_i > u_i)$ can be characterised by a generalised Pareto distribution, denoted $\mathrm{GPD}(\sigma_i, \xi_i)$, which has distribution function

$$F_i(x) = \begin{cases} 1 - \left(1 + \frac{\xi_i x}{\sigma_i}\right)_+^{-1/\xi_i}, & \xi_i \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma_i}\right), & \xi_i = 0, \end{cases} \tag{3.1.2}$$

for $x > 0$, scale parameter $\sigma_i > 0$, shape parameter $\xi_i \in \mathbb{R}$ and where $z_+ = \max\{0, z\}$. The operator z_+ forces X_i to have upper-endpoint $x_i^F = -\sigma_i/\xi_i$ if and only if $\xi_i \leq 0$ and the shape parameter ξ_i controls the heaviness of the upper tails of X_i: for $\xi_i > 0$, $\xi_i = 0$ and $\xi_i < 0$, we have that X_i has heavy, exponential and bounded, upper tails, respectively. It is important to make the distinction between these three cases as we show that the sign of the marginal shape parameters, ξ_i, has a large effect on the tail behaviour of R. To analytically quantify the effect of the marginal tail behaviour and dependence on the sum, we focus on the bivariate sum $R = X_1 + X_2$, where $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ and $X_i > 0$ for $i = 1, 2$, and with some specified joint distribution on (X_1, X_2). We assume that $u_i = 0$ for $i = 1, 2$ and discuss the implications of this choice in Section 3.2.

It remains to specify the dependence structure between X_1 and X_2 which leads to large R. The dependence between extreme values of variables is often classified into one of two classes: asymptotic dependence or asymptotic independence with respective measures of dependence: χ the coefficient of asymptotic dependence and $\bar{\chi}$ the coefficient of asymptotic independence (Coles et al., 1999). The former is defined

$$\chi = \lim_{q \uparrow 1} \Pr\{F_1(X_1) > q | F_2(X_2) > q\}, \tag{3.1.3}$$

where the value of χ determines the class and stength of extremal dependence between X_1 and X_2; for $\chi = 0$ and $\chi > 0$, we have asymptotic independence and asymptotic dependence, respectively, between X_1 and X_2, with χ increasing with strength of extremal dependence. Conversely, Ledford and Tawn (1996) characterise asymptotic independence between X_1 and X_2 through the assumption that the joint survivor function has the property

$$\Pr\left\{F_1(X_1) > 1 - \frac{1}{u}, F_2(X_2) > 1 - \frac{1}{u}\right\} = \mathcal{L}(u)u^{-1/\eta}, \quad \text{as } u \to \infty, \tag{3.1.4}$$

where $0 < \eta \leq 1$, $\mathcal{L}(\cdot)$ is slowly varying and $\bar{\chi} = 2\eta - 1$, so $-1 < \bar{\chi} \leq 1$. In particular, if $\bar{\chi} = 1$ and $\mathcal{L}(u)$ tends to a positive constant as $u \to \infty$, we have asymptotic dependence, and for $\bar{\chi} \in [0, 1)$ we have asymptotic independence with decreasing strength of dependence as $\bar{\chi} \to 0$. We consider two special cases of these extremal dependence classes, namely perfect positive dependence and independence. For the former, we have $\chi = 1$ in (3.1.3) and $\eta = 1$ in (3.1.4), and for the latter, $\chi = 0$ and $\eta = 1/2$. In both cases, $\mathcal{L}(u) = 1$ for $u > 1$.

Previous studies on the tail behaviour of aggregated random variables focus on the effects of the marginal distributions, with limited cases of the dependence structure being considered. Numerous studies on the sum of independent ($\chi = 0$, $\bar{\chi} = 0$) Pareto random variables, corresponding to GPD random variables with $\xi = 1$, have been conducted, see Zaliapin et al. (2005); Ramsay (2006, 2008); Nguyen and Robert (2015). Goovaerts et al. (2005) study the tail behaviour of weighted sums of Pareto random variables, where the weights are random and exhibit dependence which is modelled using elliptical distributions. Opitz (2016) describes the relationship between marginal exceedance probabilities for both an exponential-tailed Laplace random vector and its sum. Nadarajah (2008) give the exact distribution of independent exponential random variables with nonhomogenous, i.e., different, marginal scale parameters, and Nadarajah and Kotz (2008); Nadarajah et al. (2018) extend this framework to independent GPD margins. Nadarajah and Espejo (2006) further derive the distribution of R with GPD margins and a Clayton copula ($\chi > 0$, $\bar{\chi} = 1$), see Ghosh and Banks (2020). Under a general assumption that $\chi > 0$ and that the shape parameters are equal, studies that focus on the extremal behaviour of R include Coles and Tawn (1994) and Klüppelberg and Resnick (2008) and where R is an integral of a stochastic process by Coles and Tawn (1996) and Engelke et al. (2019a). The extension to asymptotically independent structures has been made by Engelke et al. (2019b), who study the relationship between the relative tail decay rates of the bivariate sum R and

random vector $(X_1/R, X_2/R)$, and the corresponding values of χ and η for (X_1, X_2); however, these are general results and do not link the marginal shapes to the tail decay rate of R. Other general results for the tail behaviour of sums include extensions of Breiman's lemma (Breiman, 1965), which link the decay rate of a multivariate regularly varying random vector (see Resnick (1987)) to the decay rate of the sum of its components, see Fougeres and Mercadier (2012) and Li (2018); whilst we omit the details of these results, we note that they apply to cases where $\min\{\xi_1, \xi_2\} > 0$. Therefore there are important gaps in the literature for the tail behaviour of R relating unequal marginal shape parameters and copulae with $\chi = 0$ and $\bar{\chi} < 1$. The case where $\bar{\chi} < 0$ implies negative dependence between X_1 and X_2; this case is also absent from the literature, but we constrain our focus to $\bar{\chi} \geq 0$.

The paper is structured as follows. Section 3.2 follows with a numerical study that motivates our use of the limiting dependence models of Ledford and Tawn (1997) and Heffernan and Tawn (2004). Section 3.3 introduces preliminary model set-up and the results that follow by modelling dependence in (X_1, X_2) using the limit models given above; these are easily interpretable and give a strong insight into the tail behaviour of the aggregate. In Section 3.4, we provide examples of our results for widely used copulae and give further insight into the tail behaviour of R when the dependence in (X_1, X_2) does not satisfy the conditions detailed in Section 3.3.2. We apply our results to UK precipitation and temperature data in Section 3.5. Appendix A provides the proofs of the results in Section 3.3.2.

3.2 Motivation

We explore the upper-tail of R numerically using Monte-Carlo methods for copulas with a range of χ and $\bar{\chi}$ values; this is to motivate the form in which we present the results in Section 3.3.2 and our choice of the frameworks of Ledford and Tawn (1997)

and Heffernan and Tawn (2004) for modelling. We consider two copulas based on the bivariate extreme value copula, see Tawn (1988) and Gudendorf and Segers (2010). An example of a bivariate extreme value copula is the logistic model,

$$C_L(u,v) = \exp\left\{-\left[(-\log u)^{1/\gamma} + (-\log v)^{1/\gamma}\right]^{\gamma}\right\}, \quad u,v \in [0,1], \tag{3.2.1}$$

where $\gamma \in [0,1)$; where here we avoid the case $\gamma = 1$ which is the independence copula, but allow $\gamma = 0$, taken as the limit in (3.2.1) as $\gamma \to 0$. From (3.1.3) and (3.1.4), this copula gives values $\chi = 2 - 2^{\gamma} > 0$ and $\bar{\chi} = 1$, and the variables are asymptotically dependent with the strength of asymptotic dependence decreases with γ increasing. Inverting this copula gives the inverted-logistic copula which is asymptotically independent, see Wadsworth and Tawn (2012). This is defined through its survival copula,

$$\bar{C}_{IL}(u,v) = \exp\left\{-\left[(-\log(1-u))^{1/\gamma} + (-\log(1-v))^{1/\gamma}\right]^{\gamma}\right\}, \quad u,v \in [0,1], \tag{3.2.2}$$

where $\gamma \in (0,1]$. In contrast to the logistic copula, we have $\chi = 0$ and $\bar{\chi} = 2^{1-\gamma} - 1$, with strength of asymptotic independence increasing as γ decreases.

In Section 3.3.2, we present our results for $\Pr\{R \geq r\}$ in the form

$$\Pr\{R \geq r\} \sim \begin{cases} K_1 r^{-1/\xi_R}, & \text{if } \xi_R > 0, \\ K_2 \exp\left\{-\frac{r}{\sigma_R}\right\}, & \text{if } \xi_R = 0, \\ K_3 \left\{1 - \frac{r}{r^F}\right\}^{-1/\xi_R}, & \text{if } \xi_R < 0, \end{cases} \tag{3.2.3}$$

as r tends to r^F, the upper-endpoint of R, which is infinite if $\xi_R \geq 0$ and is finite when $\xi_R < 0$. Here $\sigma_R > 0$ and $K_1, K_2, K_3 > 0$ are proportionality constants. In Supplementary Material A.2, we show how expression (3.2.3) links to the GPD tail formulation which is typically required for modelling using (3.1.2). We use this

formulation, instead of the GPD form, to avoid introducing an arbitrary threshold u_R. Expression (3.2.3) highlights that the tail of R is predominantly determined by ξ_R, with σ_R important when $\xi_R = 0$, and r^F when $\xi_R < 0$. Note that in general $r^F \leq x_1^F + x_2^F$, where x_i^F is the upper-endpoint of X_i for $i = 1, 2$, but for the copulas considered in this section the equality holds.

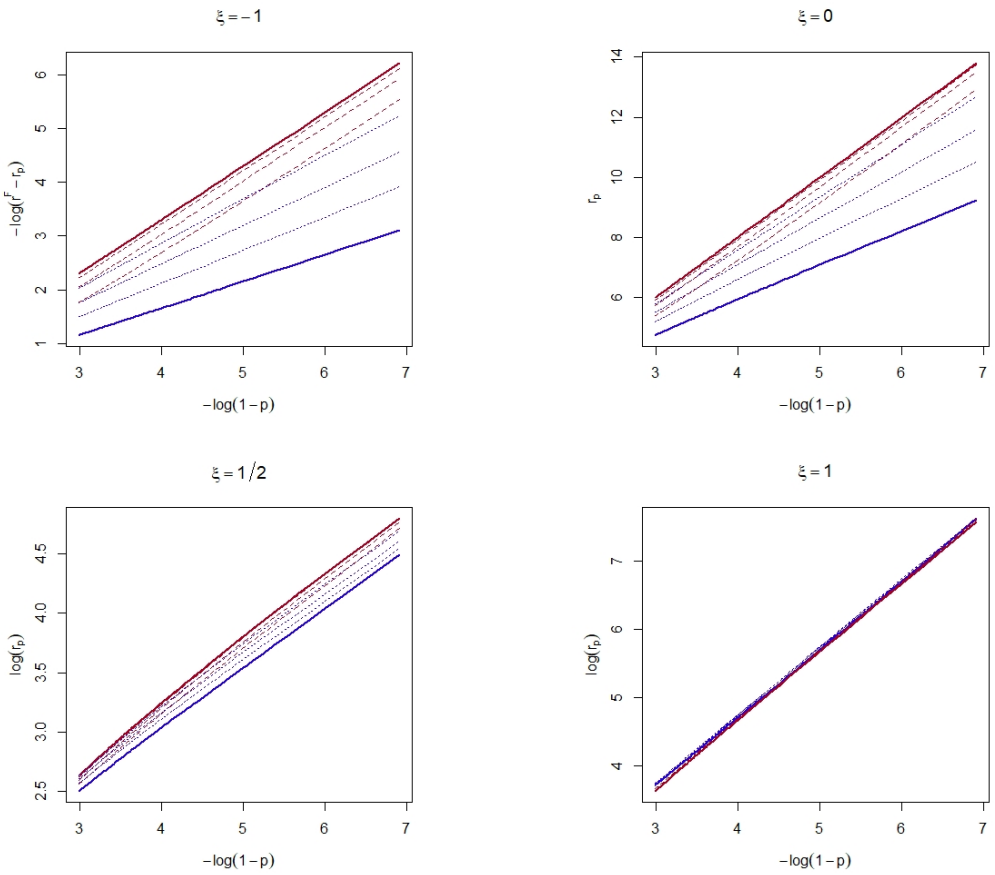

Figure 3.2.1: Quantiles r_p of R; the sum of two GPD$(1, \xi)$ random variables, with copula (3.2.1) in red and (3.2.2) in blue and for $\xi = -1, 0, 1/2, 1$ and $\gamma = 0.3, 0.5, 0.9$ and $p \in [0.95, 0.999]$. To emphasise their similarities, these are displayed on the scales $-\log(r^F - r_p)$, r_p and $\log(r_p)$ for $\xi < 0$, $\xi = 0$ and $\xi > 0$ respectively, where r^F is the upper-endpoint of R. Solid lines correspond to perfect dependence and independence, and the values on the $y-$axis decrease in each plot with increasing γ. Curves are estimated using Monte Carlo methods, with samples taken to be sufficiently large that any observed differences in the plot are statistically significant.

Figure 3.2.1 provides simulated quantiles of samples of size 5×10^6 for $R = X_1 + X_2$, where $X_1, X_2 \sim$ GPD$(1, \xi)$ with copulae (3.2.1) and (3.2.2) for selected values of ξ

and the copulae parameter γ. Quantiles r_p, where $\tilde{F}_R(r_p) = p$ for $\tilde{F}_R$ the empirical distribution of R, are given for p close to 1. We observe that growth of the quantiles of R is affected by both the underlying dependence in (X_1, X_2) and the marginal shape parameters. The scales of the axes in Figure 3.2.1 are chosen so that the gradients of the lines reveal structure about the shape parameter of R. To illustrate this, let $\Pr\{R \geq r_p\} = 1 - p$ where the survivor function of R takes the form (3.2.3). Then

$$-\log(1-p) \sim \begin{cases} \frac{1}{\xi_R}\log(r_p) - \log(K_1), & \text{if } \xi_R > 0, \\ \frac{r_p}{\sigma_R} - \log(K_2), & \text{if } \xi_R = 0, \\ -\frac{1}{\xi_R}\log(r^F - r_p) - \log(K_3) - \frac{1}{\xi_R}\log(r^F), & \text{if } \xi_R < 0, \end{cases} \tag{3.2.4}$$

as $p \to 1$. Thus, with the axes scaling used in Figure 3.2.1, we expect the gradient of each quantile curve to be $1/\xi_R$, $1/\sigma_R$ and $-1/\xi_R$ if $\xi_R > 0$, $\xi_R = 0$ and $\xi_R < 0$, respectively.

Relationship (3.2.4) and Figure 3.2.1 reveal interesting preliminary insights into the upper tail behaviour of R. For marginal positive shape parameters, we find that the gradients in Figure 3.2.1 are approximately equal; implying that the dependence structure has no significant effect on the shape parameter of R. For zero and negative marginal shape parameters, the reverse is true; for $\xi = 0$, we observe that for the asymptotically independent copulas, the scale parameter of R changes with the strength of dependence; a similar property can be observed for $\xi < 0$, albeit giving a change in the shape parameter for R. In both cases, the gradients remain approximately equal for the quantiles derived using the asymptotically dependent copula, which implies that some of the structure in the shape parameter of the survival function of R is driven by the strength of asymptotic independence, rather than the degree of asymptotic dependence.

Figure 3.2.2 motivates our choice of the regions on which we focus for charac-

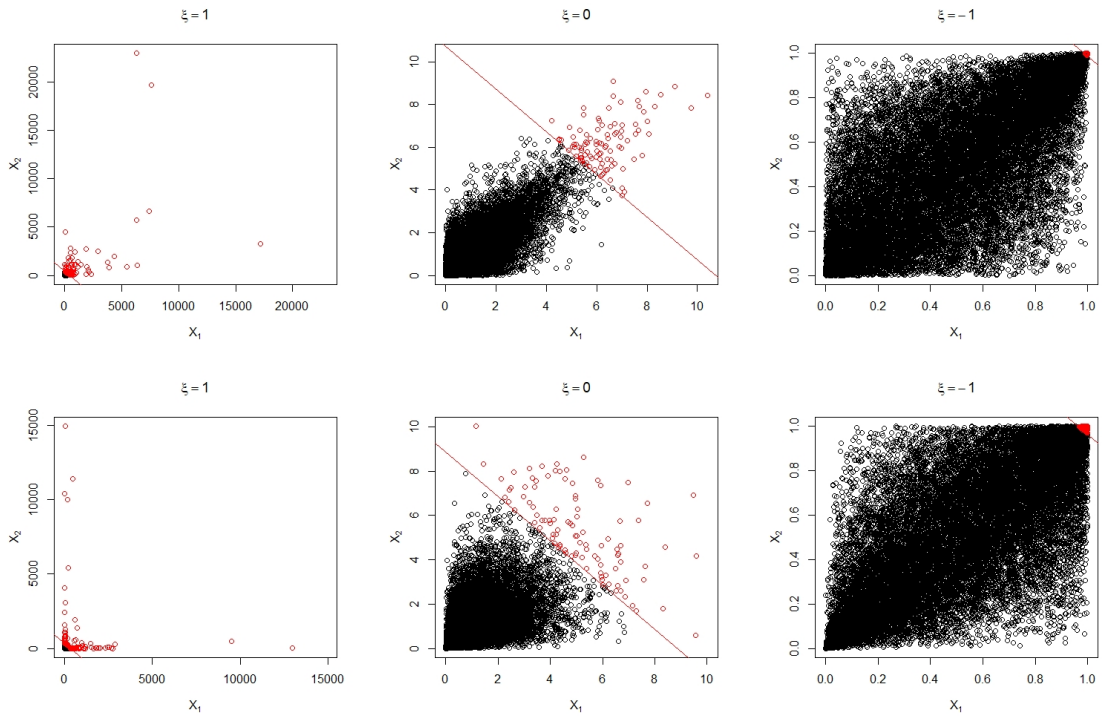

Figure 3.2.2: Scatter plots of 20000 simulated $X_1, X_2 \sim \text{GPD}(1, \xi)$ with copula (3.2.1) (top) and (3.2.2) (bottom). Both copulas take parameter value $\gamma = 0.5$ and so $(\chi, \eta) = (2 - 2^{1/2}, 1)$ and $(0, 2^{-1/2})$ in the two rows, respectively. The red points are those for which $X_1 + X_2 > r_{0.999}$, the estimated 0.999−quantile of $R = X_1 + X_2$.

terising dependence within (X_1, X_2) to derive the extremes of R. Here we plot simulated $X_1, X_2 \sim \text{GPD}(1, \xi)$ with dependence induced through the logistic and inverted-logistic copula, equations (3.2.1) and (3.2.2), respectively. The regions of (X_1, X_2) for which $R \geq r_{0.999}$ are shown, with points in these regions highlighted in red. The combinations of $(X_1, X_2)|(R > r_{0.999})$ highlight which aspects of the copula are important for studying the tail behaviour of R. These combinations are similar for different copulas, or dependence structures, but differ for different signs on the marginal shape parameter. For $\xi \leq 0$, the large values of R occur for values which are large in both marginals, which suggests that the important regions of the copula are those where both arguments are simultaneously large; Ledford and Tawn (1997, 1998) detail dependence in these regions. Conversely, Figure 3.2.2 illustrates that for $\xi > 0$, large values of R occur when (X_1, X_2) is extreme in at least one component.

We thus require a model that considers the distribution of one variable whilst the other is already extreme; which is covered by the characterisation of Heffernan and Tawn (2004). We use both approaches for describing limiting dependence of (X_1, X_2) and detail these characterisations in Section 3.3.1.

In Section 3.1, we specified that throughout we would assume that $X_i > 0$ with $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $i = 1, 2$. These assumptions are clearly highly restrictive when describing marginal behaviour, but as our interest lies in the upper tail behaviour of R, we find that the full distribution of X_i is not always relevant. For example, Figure 3.2.2 indicates that when $\max\{\xi_1, \xi_2\} < 0$, the combinations of (X_1, X_2) which give large R require both X_i variables to be in their upper tails. When $\max\{\xi_1, \xi_2\} \geq 0$ and (X_1, X_2) are positively dependent in their extremes, large values of R tend to occur when both marginal variables are in their tails. In the case where extremal dependence is weak and the marginal tails are heavy, then R is dominated by only one large marginal variable; the distribution of the values in the body of the smaller variable is not important for the characteristics of the upper tail of R.

These arguments indicate that it is predominantly the upper tail of the marginal variables that are important. The widely adopted approximation for the upper tails of arbitrary marginal variables is that, for some high quantile $u_i > 0$ of X_i, that $(X_i - u_i)|(X_i > u_i)$ follows a GPD (Pickands, 1975). Our approach is consistent with this, following the threshold stability property (Coles, 2001) of the GPD: that for all $0 < u_i < x_i^F$ we have $(X_i - u_i)|(X_i > u_i) \sim \text{GPD}(\sigma_i - \xi_i u_i, \xi_i)$, and so our approach is consistent with the usual tail model without any loss of generality. Thus, our modelling of the marginal distribution has the following properties: it avoids the arbitrary choice of u_i; it determines the shape parameter of the tail of R for all ξ_i; when $\xi_i = 0$ it uniquely determines the scale parameter of the tail through σ_i; and if the marginal variables are not lower bounded by zero, then similar results are obtained by location shifting the X_i, where X_i has a finite lower bound.

3.3 Limit results

3.3.1 Background and model set-up

We now describe the extremal dependence characteristics that we assume for (X_1, X_2). Figure 3.2.2 illustrates that we require two characterisations; one for when X_1 and X_2 are joint large, and one where at least one is large. For the former, Ledford and Tawn (1997, 1998) present an extension of (3.1.4); this model was further extended by Ramos and Ledford (2009). Presented here for general marginals F_1 and F_2, they characterise the joint survival function as

$$\Pr\left\{F_1(X_1) > 1 - \frac{1}{x_1}, F_2(X_2) > 1 - \frac{1}{x_2}\right\} = \frac{\mathcal{L}(x_1 + x_2)}{(x_1 x_2)^{\frac{1}{2\eta}}} g\left(\frac{x_1}{x_1 + x_2}\right), \qquad (3.3.1)$$

for any $x_1 \to \infty,\ x_2 \to \infty$ such that $x_1/(x_1 + x_2) \to w$ for $0 < w < 1$, and where $\mathcal{L}(\cdot)$ is slowly-varying and the continuous function $g : (0, 1) \to \mathbb{R}_+$. Ledford and Tawn (1997) have different powers of x_1 and x_2 which then requires that g satisfies a property they term quasi-symmetry; however, Ramos and Ledford (2009) use equal powers of x_1 and x_2 in the denominator which removes the need for this property. Ledford and Tawn (1997) provide examples of g for certain copulae, e.g., for the logistic copula, they illustrate that $g(w) = \{w(1 - w)\}^{-1/2}[1 - V((1 - w)^{-1}, w^{-1})]$ where $V(x, y) = (x^{-1/\gamma} + y^{-1/\gamma})^{\gamma}$ for γ defined in (3.2.1), and for the inverted logistic copula defined in (3.2.2), they show that $g(w) \to 1$ for all $w \in [0, 1]$ as $r \to r^F$.

Heffernan and Tawn (2004) and Keef et al. (2013) quantify extremal dependence between variables by conditioning on one variable being extreme; whilst their characterisations can accommodate negative extremal dependence, we focus on non-negative association only. To model extremal dependence in (X_1, X_2), they consider the transformed variables $Y_1 = -\log\{1 - F_1(X_1)\}$ and $Y_2 = -\log\{1 - F_2(X_2)\}$, such that Y_1, Y_2 are standard exponential random variables. Under the assumption that there exists

normalising functions $a : \mathbb{R} \to \mathbb{R}, b : \mathbb{R} \to \mathbb{R}_+$, then for any fixed $z \in \mathbb{R}, y \in \mathbb{R}_+$ and for any sequence $u \to \infty$, we have

$$\Pr\left\{\frac{Y_2 - a(Y_1)}{b(Y_1)} < z, Y_1 - u > y \middle| Y_1 > u\right\} \to \exp(-y)G(z), \quad \text{as } u \to \infty, \qquad (3.3.2)$$

where $G(\cdot)$ is non-degenerate and $\lim_{z\to\infty} G(z) = 1$. Often, for the purposes of modelling non-negative extremal association, the normalising functions are simplified to location and scale parameters, i.e., $a(y) = \alpha y$ for $\alpha \in [0, 1]$ and $b(y) = y^\beta$ for $0 \leq \beta < 1$. The values of α and β determine the strength of dependence between Y_1 and Y_2, and, thus, between X_1 and X_2. For example, asymptotic dependence between the two is implied by values $\alpha = 1, \beta = 0$. Within the class of asymptotic independence, we have $\alpha < 1, \beta \geq 0$, with $\alpha = \beta = 0$ giving near perfect independence; we further require $G(\cdot)$ to be standard exponential if (X_1, X_2) are independent.

3.3.2 Results

We now present the results for the tail behaviour of $R = X_1 + X_2$ derived by using the limiting structures described in (3.3.1) and (3.3.2) to model dependence in (X_1, X_2). Recall in (3.1.1) we define R as a weighted sum, i.e., $R = \omega_1 X_1 + \omega_2 X_2$ with $0 < \omega_1, \omega_2 < 1$ and $\omega_1 + \omega_2 = 1$. By setting $Y_i = \omega_i X_i$ where $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ it follows from (3.1.2) that $Y_i \sim \text{GPD}(\omega_i\sigma_i, \xi_i)$ and $R = Y_1 + Y_2$, and so we present results for $R = X_1 + X_2$ without loss of generality. We begin with Theorems 3.3.1 and 3.3.2, which detail the cases where the marginal shape parameters are equal and non-zero, and zero, respectively. Theorems 3.3.3 and 3.3.4 provides results for the cases where the marginal shapes are unequal; Theorem 3.3.3 covers those cases where both shapes are strictly negative and the other cases are covered by Theorem 3.3.4. The proofs for all theorems are provided in Appendix A.

Throughout Theorems 3.3.1-3.3.3, we make the assumption that $X_1 \sim \text{GPD}(\sigma_1, \xi_1)$

and $X_2 \sim \text{GPD}(\sigma_2, \xi_2)$, with distribution functions defined in (3.1.2), and that the extremal dependence in (X_1, X_2) satisfies the regularity conditions for model (3.3.1); we further assume that there exists a fixed $v > 0$ such that, for all $y > v$, we have $\mathcal{L}(y)$ acts like a positive constant which is absorbed by the function g. We assume that model (3.3.1) holds in equality for $x_1 + x_2 \geq \max\{c, u^*\}$ for a fixed constant $0 < c < r^F$, and where $u^* = \max\{x_1^F, x_2^F\}$ if $\max\{\xi_1, \xi_2\} < 0$, and $u^* = 0$ otherwise; that is, we require that model (3.3.1) holds for large R. We make the assumption that the first- and second-order derivatives of g exists; further assumptions on g are made for specific cases. If $\min\{\xi_1, \xi_2\} = \xi > 0$, we require an additional assumption that the limit in (3.3.2) holds in equality for fixed $u > 0$ and that the residual distribution G is differentiable. For the theorems that require specific assumptions for g, we consider g satisfying different conditions:

Condition 1 There exists a fixed $v^* > 0$, such that for $r = x_1 + x_2 > v^*$, we have $g(\omega_x) = 1$ for all $\omega_x = \exp(x_1/\sigma_1)/[\exp(x_1/\sigma_1) + \exp(x_2/\sigma_2)] \in [0, 1]$; or equivalently, X_1 and X_2 are independent when $R = X_1 + X_2 > v^*$, and where $X_i \sim \text{Exp}(1/\sigma_i)$ for $\sigma_i > 0$ and $i = 1, 2$.

Condition 2 The tails of g satisfy $g(w) \sim K_g w^\kappa$ as $w \to 0$ and $g(w) \sim K_g(1-w)^\kappa$ as $w \to 1$ for constant $K_g > 0$ and fixed $0 \leq \kappa < 1/(2\eta)$.

Condition 3 As $w \to 0$ or $w \to 1$, we have that

$$g(w) \sim w^{-1/(2\eta)}(1-w)^{-1/(2\eta)}[1 - H((1-w)^{-1}, w^{-1})],$$

where the bivariate function H is homogeneous of order -1 and its first and second-order partial derivatives exist and are continuous, and $H(\infty, t) = H(t, \infty) = t^{-1}$ for $t > 0$. We denote H_1 and H_2 as the first-order partial derivatives of H with respect to the first and second arguments respectively, and H_{12} the second-order

partial derivative with respect to both arguments, and present two sub-conditions of Condition 3: Condition 3a, $H_{12}(1,z) \sim -K_{H_1} z^{c_1}$ as $z \to 0$ for constants $K_{H_1} > 0$ and $c_1 > -1$ and $H_{12}(1,z) \sim -K_{H_2} z^{c_2}$ as $z \to \infty$ for constants $K_{H_2} > 0$ and $c_2 < -2$; Condition 3b, $H_1(1,\infty) < \infty$ and $H_1(1,z) \to 0$ as $z \to 0$.

Although Conditions 3a and 3b appear quite restrictive, they are both satisfied by the bivariate extreme value distribution with $\kappa = 1/2$; this is a widely-applied copula for modelling asymptotic dependence, with the logistic copula as a special case.

Theorem 3.3.1. *If $\xi_1 = \xi_2 = \xi \neq 0$, then*

$$\Pr\{R \geq r\} \sim \begin{cases} K\left(1 + \frac{\xi r}{\sigma_1 + \sigma_2}\right)^{-\frac{1}{\eta\xi}}, & \text{if } \xi < 0, \\ K^* r^{-1/\xi}, & \text{if } \xi > 0, \text{ and } ^*, \end{cases}$$

as $r \to r^F$, where $r^F = \infty$ for $\xi > 0$ and $r^F = -(\sigma_1 + \sigma_2)/\xi$ for $\xi < 0$, and for constants K and K^* defined in (A.3.3) *and* (A.3.15)*, respectively. *Condition 2 holds or Condition 3 with $\eta = 1$ holds*

Theorem 3.3.2. *If $\xi_1 = \xi_2 = 0$, then*

$$\Pr\{R \geq r\} \sim \begin{cases} \frac{\sigma_{max}}{\sigma_{max} - \sigma_{min}} \exp\left(-\frac{r}{2\eta\sigma_{max}}\right), & \text{if } \sigma_1 \neq \sigma_2, \text{Condition 1 holds}, \\ \frac{r}{2\eta\sigma} \exp\left(-\frac{r}{2\eta\sigma}\right), & \text{if } \sigma_1 = \sigma_2 = \sigma, \text{Condition 1 holds}, \\ K \exp\left(-\frac{r}{(\sigma_1 + \sigma_2)}\right), & \text{if Condition 3a holds with } \eta = 1, \end{cases}$$

as $r \to \infty$ and for constant K defined (A.4.3)*, where $\sigma_{max} = \max\{\sigma_1, \sigma_2\}$ and $\sigma_{min} = \min\{\sigma_1, \sigma_2\}$.*

Note that there is a power term in the second case for $\Pr\{R \geq r\}$ given by Theorem 3.3.2 that is not covered by the general form given by (3.2.3). However, this is in the domain of attraction of a GPD with shape and scale parameters zero and $2\eta\sigma$, respectively.

Theorem 3.3.3. *If* $\xi_1 \neq \xi_2$ *and* $\max\{\xi_1, \xi_2\} < 0$*, then*

$$\Pr\{R \geq r\} \sim \begin{cases} K_1 \left(1 + \frac{\xi_1 \xi_2 r}{\sigma_1 \xi_2 + \sigma_2 \xi_1}\right)^{-\frac{1}{\xi_{max}}\left(\frac{1}{2\eta}+\kappa\right)-\frac{1}{\xi_{min}}\left(\frac{1}{2\eta}-\kappa\right)}, & \text{if Condition 2 holds,} \\ K_2 \left(1 + \frac{\xi_1 \xi_2 r}{\sigma_1 \xi_2 + \sigma_2 \xi_1}\right)^{-\frac{1}{\eta \xi_{max}}}, & \text{if Condition 3b holds,} \end{cases}$$

as $r \to r^F = -(\sigma_1/\xi_1 + \sigma_2/\xi_2)$*, and for constants* $K_1 > 0$ *and* $K_2 > 0$ *defined in* (A.5.3) *and* (A.5.6)*, respectively, and* $\xi_{max} = \max\{\xi_1, \xi_2\}, \xi_{min} = \min\{\xi_1, \xi_2\}$.

The set conditions on the dependence between X_1 and X_2 described above for Theorems 3.3.1-3.3.3 are not necessary for Theorem 3.3.4; instead, this theorem applies for any non-negative association between X_1 and X_2.

Theorem 3.3.4. *If* $\xi_1 \neq \xi_2$ *and* $\max\{\xi_1, \xi_2\} \geq 0$*, then*

$$\Pr\{R \geq r\} \sim \begin{cases} \left(\frac{\xi_{max}}{\sigma_{max}}\right)^{-1/\xi_{max}} r^{-1/\xi_{max}}, & \text{if } \max\{\xi_1, \xi_2\} > 0, \\ C \exp\left(-\frac{r}{\sigma_{max}}\right), & \text{if } \max\{\xi_1, \xi_2\} = 0, \end{cases}$$

as $r \to \infty$ *and where* $\xi_{max} = \max\{\xi_1, \xi_2\}$ *and* $\sigma_{max} = \{\sigma_i; i \text{ is s.t. } \xi_i = \xi_{max}\}$ *and for constant* $C \in [C_1, C_2]$ *for* $C_1 > 0$ *and* $C_2 > 0$ *defined in* (A.6.1) *and* (A.6.2) *respectively.*

To illustrate that $\xi_R = \max\{\xi_1, \xi_2\}$ for ξ_R defined in (3.2.3), Koutsoyiannis (2020) provide a similar result to the case in Theorem 3.3.4 where $\min\{\xi_1, \xi_2\} > 0$ using a different approach. We further note that the cases where $\min\{\xi_1, \xi_2\} > 0$ in Theorems 3.3.1 and 3.3.3 agree with Breiman's Lemma (Breiman, 1965), as we have $\xi_R = \max\{\xi_1, \xi_2\}$.

3.4 Copula examples

We now compare the limit results detailed in Section 3.3.2 with results for the upper-tail behaviour of R when dependence in (X_1, X_2) is fully modelled using copula families

and their marginal models remain the same, i.e., $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $i = 1, 2$. The assumptions we made in Section 3.3.2 hold in some cases and in these we obtain identical results to Section 3.3.2. However, where the assumptions of Section 3.3.2 are too strong, our direct derivations from the copulae, with details in Chapter 4, provide insight into the tails of R in these specific cases. We consider the extreme value copula and the inverted extreme value copula and the limiting forms of these two classes, i.e., perfect dependence and independence. We further consider a standard Gaussian copula with correlation parameter ρ $(0 < \rho < 1)$; this copula exhibits asymptotic independence, i.e., $\chi = 0$, $\bar{\chi} = \rho$.

The extreme value copula takes the form

$$C_{ev}(u, v) = \exp\left\{-V(-1/\log(u), -1/\log(v))\right\}, \tag{3.4.1}$$

where

$$V(x, y) = 2\int_0^1 \max\left\{\frac{w}{x}, \frac{1-w}{y}\right\} \mathrm{d}M(w), \tag{3.4.2}$$

is a homogeneous function of order -1 and $M(w)$ is a univariate distribution function/probability measure for $w \in [0, 1]$, which has expectation $1/2$. Note that $1 \leq V(1, 1) \leq 2$, where the boundary cases correspond to special cases of the extreme value copula, i.e., we have perfect dependence, and independence, between X_1 and X_2 when $V(1, 1) = 1$ and $V(1, 1) = 2$ respectively. This copula gives $\eta = 1$ $(\bar{\chi} = 1)$ and $\eta = 1/2$ $(\bar{\chi} = 0)$ when $V(1, 1) < 2$ and $V(1, 1) = 2$, respectively. Furthermore, Ledford and Tawn (1997) illustrate that this copula satisfies Condition 3a/3b, with $H = V$ and $\kappa = 1/2$, that is required for Theorems 3.3.2 and 3.3.3.

The inverted extreme value copula follows by inverting (3.4.1), see Wadsworth and Tawn (2012), and is defined through its survival copula

$$\bar{C}_{iev}\{u, v\} = \exp\left\{-V(-1/\log(1-u), -1/\log(1-v))\right\}, \tag{3.4.3}$$

with a similarly defined V. This, and the Gaussian copula, have $\eta = V(1,1)^{-1}$ and $\eta = (1+\rho)/2$, respectively, where $\bar{\chi} = 2\eta - 1$. The logistic and inverted logistic copulas defined in (3.2.1) and (3.2.2), respectively, are subclasses of (3.4.1) and (3.4.3), respectively. When discussing results pertaining to copulas (3.4.1) and (3.4.3), we assume that the first- and second-order partial derivatives of V exist, which corresponds to the existence of a joint density, hence this excludes perfect dependence which is derived separately.

We report the parameters that determine the behaviour of $\Pr\{R \geq r\}$ as $r \to \infty$ as given by form (3.2.3), i.e., $\xi_R \neq 0$ and σ_R, otherwise. Consider three cases: $\min\{\xi_1, \xi_2\} > 0$, $sgn(\xi_1) \neq sgn(\xi_2)$, and $\max\{\xi_1, \xi_2\} \leq 0$. In the first two cases, no further insight into the uper-tails of R is revealed when modelling dependence in $\mathbf{X}$ using copulaes, and Chapter 4 gives the same results as detailed in Section 3.3.2, i.e., $\xi_R = \max\{\xi_1, \xi_2\}$; this suggests that, for these cases, modelling dependence using the limiting models of Ledford and Tawn (1997) and Heffernan and Tawn (2004) is sufficient to derive the first-order behaviour of the upper-tail of R.

However, if we have that $\max\{\xi_1, \xi_2\} \leq 0$ and (X_1, X_2) exhibits asymptotic independence but positive association, then Chapter 4 shows that further insight into the tail behaviour of R can be gained by modelling dependence with copulas in the following two cases: $\xi_1 = \xi_2 = \xi_R = 0$ and $\max\{\xi_1, \xi_2\} < 0, \xi_1 \neq \xi_2$, see Table 3.4.1; we find no change in all other cases. Ledford and Tawn (1997) illustrate that none of Conditions 1-3 are met by either the inverted extreme value, or standard Gaussian, copulas, and so the results for these copulae, given in Chapter 4, are presented in Table 3.4.1. We observe that for these copulae, the parameters in (3.2.3) cannot be represented as the product of a function of the marginal parameters and the summary measure η; instead, the upper-tail behaviour of R is driven by a function of both the marginal parameters and dependence structure which cannot be factorised, which suggest that there is a more subtle relationship between the marginal shapes,

extremal dependence structure and tail behaviour of R.

	$\xi_1 = \xi_2 = \xi_R = 0$	$\max\{\xi_1, \xi_2\} < 0,\ \xi_1 \neq \xi_2$
Dependence Structure	σ_R	ξ_R
Theorems 3.3.2/3.3.3, $\chi > 0$	$\sigma_1 + \sigma_2$	$\max\{\xi_1, \xi_2\}$
Theorems 3.3.2/3.3.3, $0 \leq \bar{\chi} < 1$	$2\eta \max\{\sigma_1, \sigma_2\}$	$2\eta \left(\frac{1}{\xi_1} + \frac{1}{\xi_2}\right)^{-1}$
Independence $\bar{\chi} = 0$	$\max\{\sigma_1, \sigma_2\}$	$\left(\frac{1}{\xi_1} + \frac{1}{\xi_2}\right)^{-1}$
Perfect dependence $\chi = 1$	$\sigma_1 + \sigma_2$	$\max\{\xi_1, \xi_2\}$
Extreme value copula $\chi > 0$	$\sigma_1 + \sigma_2$	$\max\{\xi_1, \xi_2\}$
Inverted extreme value copula $0 \leq \bar{\chi} < 1$	$\max_{0 \leq w \leq 1} \left\{\left[V\left(\frac{\sigma_1}{w}, \frac{\sigma_2}{1-w}\right)\right]^{-1}\right\}$	$-1/V(-\xi_1, -\xi_2)$
Standard Gaussian $\bar{\chi} = \rho, \rho \in [0, 1)$	$(1-\rho^2) \max_{0 \leq w \leq 1} \{h(w)^{-1}\}$	$\frac{1-\rho^2}{\xi_1^{-1} + 2\rho(\xi_1\xi_2)^{-1/2} + \xi_2^{-1}}$

Table 3.4.1: Parameter values for $R = X_1 + X_2$ where (X_1, X_2) have GPD margins with $\max\{\xi_1, \xi_2\} \leq 0$, and $h(w) = \sigma_1 w - 2\rho\sqrt{\sigma_1\sigma_2 w(1-w)} + \sigma_2(1-w)$.

3.5 Application to aggregated environmental data

We now present an application of the results discussed in Section 3.3.2 to climate model data. We study precipitation and temperature data, which have heavy and bounded marginal upper-tails respectively. Both datasets are obtained from the UK climate projections 2018 (UKCP18) (Lowe et al., 2018) which contains values aggregated over a given time interval and a spatial grid-box. The size of these grid-boxes and the specified time interval differ between the two studies. In both cases, we investigate the marginal upper-tail for the variables observed at a configuration of grid-boxes and the spatial average of them over adjacent boxes. Note that we aggregate the data as we are interested in the extremal behaviour of the climate processes at lower resolutions; for precipitation, this is for the reasons described in Section 3.1, and for temperature, we are interested in the average extreme heat over a large spatial domain since a heatwave has societal impact owing to it affecting a spatial region not simply a single location.

Recall from Section 3.3.1 that the driving factor for the extremal behaviour of the aggregates is the GPD shape parameter, ξ. We focus on just the relationship between estimates of ξ for the marginal variables and ξ for the aggregates. To investigate this relationship, we begin with a 2 by 2 configuration of adjacent grid-boxes. For each grid-box, we fit the GPD to excesses above the sample p-th quantile using maximum likelihood methods, under the assumption that observations are independent and identically distributed (Coles, 2001). Following many spatial extreme value applications (Coles and Tawn, 1990, 1996; Fowler and Kilsby, 2003; Coelho et al., 2008; Li et al., 2019; Davison et al., 2012, 2019), we anticipate that the shape parameters for each grid-box should be identical. Therefore we also pool information across grid-boxes with a model that the distribution of excesses in grid-box i is GPD(σ_i, ξ), i.e., a common shape parameter but with the scale parameter unconstrained over grid-boxes. For each of the 4 pairs of adjacent grid-boxes, we take the spatial aggregate of the data at each separate time interval and fit a GPD to excesses of these data above its empirical p-th quantile. Quantiles are estimated separately for marginal, pooled and aggregate variables. To account for strong spatial and temporal dependence in the data, standard errors for ξ are estimated using a stationary bootstrap (Politis and Romano, 1994) with 1000 samples, with temporal block size drawn randomly from a Geometric distribution with expectation corresponding to a week of observations.

3.5.1 Precipitation

The data are precipitation flux (mm/day) from a convection permitting model on $2.2 \times 2.2km^2$ grid-boxes and hourly intervals. To account for seasonality, we use only winter, December to February, observations between the years 1980 and 2000. We study a 2×2 configuration of grid-boxes centred around $(52.18°, 0.14°)$, approximately Cambridge, UK; this is a flat area so no orographic features are important and marginal distributions are expected to be nearly homogeneous. We conduct our anal-

ysis on outputs of the model at two spatial resolutions - high using data on $(2.2)^2 km^2$ and coarse $(22)^2 km^2$ grids. The latter is produced by taking the spatial average over 10 by 10 configurations of the former data. We analyse both resolutions to investigate the effect of extremal dependence on the observed results. This is quantfied using the measure η, given in (3.3.1), which is estimated as in Coles et al. (1999). All GPD models are fit to exceedances above 99.5% quantiles.

Table 3.5.1 presents estimates and the 95% confidence intervals for the shape parameters using the three inference methods. The marginal shape parameter estimates are predominately positive which suggests that Theorem 3.3.1 is relevant, i.e., for a homogeneous marginal shape parameter $\xi > 0$, the shape parameter of the aggregate is also ξ, regardless of the dependence structure. We aim to see if this applies in the observed tail.

		Marginal	
		0.210(0.045, 0.339)	0.197(0.037, 0.350)
Marginal	0.154(-0.030, 0.286)	0.172(0.017, 0.306) 0.160(-0.006, 0.288)	0.178(0.019, 0.320) 0.172(0.020, 0.328)
	0.225(0.040, 0.344)	0.214(0.049, 0.333) 0.177(0.036, 0.316)	0.168(-0.001, 0.283) 0.184(0.041, 0.347)

Table 3.5.1: High resolution precipitation case study: shape parameter estimates and 95% confidence intervals for margins (black), pooled marginals (red) and aggregate variable (blue).

Table 3.5.1 shows the point estimates of confidence intervals for ξ using the marginal variables and the pooled analysis. As we observe similar estimates for ξ as well as substantial overlap in the confidence intervals, this suggests that it is reasonable to assume homogeneous marginal shape parameters. Using the same criteria as above, the marginal estimates also have good agreement with ξ for the aggregate variable, suggesting that the positive shape result in Theorem 3.3.1 holds well for

these data. Pairwise η estimates for Table 3.5.1 fall in the range $[0.956, 0.967]$, which suggests strong extremal dependence between the marginal variables.

		Marginal	
		0.146(-0.033, 0.277)	0.089(-0.024, 0.197)
Marginal	0.104(-0.083, 0.218)	0.108(-0.015, 0.239) 0.177(-0.095, 0.318)	0.101(0.000, 0.186) 0.011(-0.123, 0.085)
	0.068(-0.055, 0.183)	0.105(-0.119, 0.212) 0.085(-0.082, 0.182)	0.082(-0.012, 0.176) 0.061(-0.065, 0.189)

Table 3.5.2: Coarse resolution precipitation case study: shape parameter estimates and 95% confidence intervals for margins (black), pooled marginals (red) and aggregate variable (blue).

To investigate the effect of weaker dependence on the relationship between the marginal and aggregate ξ parameter, we now consider the coarse resolution data and conduct the same analyses as previously; pairwise $\hat{\eta}$ for the coarser data are in the range $[0.859, 0.895]$, which is lower than the estimates for Table 3.5.1. Table 3.5.2 suggests that it is reasonable to assume homogeneous marginal shape parameters at this coarse resolution, as we again observe good agreement between the ξ estimates for both the marginal and pooled variables. We also observe good agreement between ξ for the pooled variables and aggregate variables even with weaker extremal dependence.

3.5.2 Temperature

The data are average daily temperature (°C) from a global climate model scaled to $60 \times 60 km^2$ grid-boxes and to account for seasonality we use only summer, July to August, observations. The model is run through the years 1899 to 2099, providing 18000 observations per grid-box. We consider a 2×2 configuration of grid-boxes centred around $(53.14°, -1.70°)$, south of the Peak District, UK. As in Section 3.5.1, we conduct our analyses on outputs of the model at two spatial resolutions - high

using data on $(60)^2 km^2$ and coarse $(300)^2 km^2$ grids; the latter produced by taking the spatial average over 5 by 5 configurations of the former data. All GPD models are fit to exceedances above 98% quantiles.

Table 3.5.3 presents estimates and the 95% confidence intervals for the shape parameter for the marginal and pooled variables, which suggest that these variables have bounded upper-tails. As such, we consider the results in Theorem 3.3.1; this states that, asymptotically, the shape parameter of the aggregate should be $\eta\xi$ given that the marginal variables have equal, negative shape $\xi < 0$. To see if this result is consistent with the observed tails, Table 3.5.3 presents estimates and 95% confidence intervals for a scaling of the aggregate shape parameter by $1/\hat{\eta}$, where the estimate $\hat{\eta}$ of η is calculated for each bootstrap sample of the aggregate; if Theorem 3.3.1 holds for these data, then this should be equal to the marginal ξ.

		Marginal	
		-0.156(-0.276, -0.067)	-0.211(-0.308, -0.108)
Marginal	-0.198(-0.310, -0.106)	-0.180(-0.268, -0.106) -0.214(-0.339, -0.103)	-0.199(-0.293, -0.133) -0.201(-0.318, -0.103)
	-0.148(-0.266, -0.082)	-0.165(-0.255, -0.069) -0.166(-0.278, -0.083)	-0.161(-0.250, -0.094) -0.160(-0.297, -0.067)

Table 3.5.3: High resolution temperature case study: shape parameter estimates and 95% confidence intervals for margins (black) and pooled variable (red). Blue confidence intervals are for a scaling of the aggregated shape parameter by $1/\eta$.

Table 3.5.3 suggests that we can assume homogeneous marginal shape parameters and these estimates also have clear agreement with the scaled shape parameter for the aggregate variable, suggesting that the negative shape result in Theorem 3.3.1 holds well for these data. Pairwise η estimates for Table 3.5.3 fall in the range $[0.918, 0.981]$, which suggests strong extremal dependence between the marginal variables, and so we repeat the analyses with the coarser data to investigate the effect of weaker dependence

on the aggregate shape parameter.

		Marginal	
		-0.113(-0.277, -0.020)	-0.207(-0.298, -0.132)
Marginal	-0.183(-0.280, -0.106)	-0.145(-0.219, -0.088) -0.200(-0.356, -0.102)	-0.200(-0.272, -0.158) -0.204(-0.342, -0.129)
	-0.053(-0.317, 0.057)	-0.138(-0.258, -0.067) -0.083(-0.338, 0.011)	-0.066(-0.255, 0.010) -0.178(-0.410, -0.071)

Table 3.5.4: Coarse resolution temperature case study: shape parameter estimates and 95% confidence intervals for margins (black) and pooled variable (red). Blue confidence intervals are for a scaling of the aggregated shape parameter by $1/\eta$.

Table 3.5.4 suggests that it is still reasonable to assume homogeneous marginal shape parameters at the coarser resolution, as we again observe good agreement between the ξ estimates for both the marginal and pooled variables. We found that pairwise values of $\hat{\eta}$ for Table 3.5.4 were in the range $[0.789, 0.921]$, which suggests weaker extremal dependence than that observed for the high resolution temperature data. We also observe good agreement between these estimates and the estimates for the scaled aggregate shape parameter, confirming that the result in Theorem 3.3.1 applies well, even for weaker extremal dependence.

3.6 Discussion

In Section 3.3.2, we provide results that begin to explore the extremal behaviour of R; the bivariate aggregate of two GPD random variables, X_1 and X_2. These results focus primarily on the effect of the marginal ξ parameters and dependence within (X_1, X_2) on the shape parameter of the aggregate, or the scale parameter if we have $\xi = 0$. Through Figure 3.2.1 and Section 3.3.1, we illustrate that the value of ξ is the most important driver in the tail behaviour of the aggregate, and so when we apply

our results to data in Section 3.5, we investigate the behaviour of ξ only. However, it is important that the behaviour of the scale parameter of R is still explored.

Although the results given in Section 3.3.2 were derived by modelling the dependence in (X_1, X_2) using the limiting extremal dependence models of Ledford and Tawn (1996) and Heffernan and Tawn (2004), results using full copula dependence models are given in Section 3.4. Here we show broad agreement between results derived using the two methods, and so we conclude that the extremal behaviour of R is mostly driven by the limiting behaviour of (X_1, X_2) as $x_1 \to \infty$ and/or $x_2 \to \infty$, and that modelling the full dependence in (X_1, X_2) is not necessary to capture the first order behaviour of $\Pr\{R \geq r\}$ as $r \to \infty$.

We cannot analytically determine the theoretical scale parameter for R in application, as the results described in Section 3.3.2 follow from asymptotic arguments that remove the conditioning typically found in GPD modelling. However, this is not to say that the results in Section 3.3.2 are not useful for modelling the extremal behaviour of aggregates; on the contrary, we can utilise these results for inference on the shape parameter ξ_R of R and then estimate the scale parameter given that ξ_R is fixed. This technique is viable as the GPD shape parameter is independent of the exceedance threshold used for modelling, and so does not suffer the same issues as the scale; this follows from the threshold-stability property of the GPD, see Coles (2001).

4

Copula-based aggregation

4.1 Introduction

In Chapter 3, we derived the upper-tail behaviour of the bivariate sum of GPD random variables with dependence characterised by the limiting models of Ledford and Tawn (1997) and Heffernan and Tawn (2004). Here we conduct a similar study, but with the following differences: first, we study the upper-tail behaviour of R_d, where

$$R_d = \sum_{i=1}^{d} X_i,$$

where $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $\sigma_i > 0$ and $\xi_i \in \mathbb{R}$ and $d \in \mathbb{N}$, i.e., R_d is the sum of d positive random components, rather than two. We use an identical framework for the marginal distributions of the components of $\mathbf{X} = (X_1, \ldots, X_d)$ as was proposed in Section 3.1, i.e., that $X_i > 0$ is GPD above zero for all $i = 1, \ldots, d$. However, we stipulate that dependence in $\mathbf{X}$ is characterised fully using one of five copulae described in Section 2.2.2: these are the extreme value copula, inverted extreme value copula and the standard Gaussian copula with zero mean, unit variance and correlation matrix Σ, and the limiting forms of these classes, i.e., perfect dependence and independence.

Copulae are described by Sklar's theorem (Nelsen, 2006), which illustrates that

the joint distribution function of $\mathbf{X}$ can be written as

$$F(\mathbf{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \quad \mathbf{x} \in \mathbb{R}^d \tag{4.1.1}$$

where C is the copula. The d-dimensional extreme value copula is given by

$$C_{MEV}\{u_1, \ldots, u_d\} = \exp\left\{-V\left(-\frac{1}{\log(u_1)}, \ldots, -\frac{1}{\log(u_d)}\right)\right\}, \tag{4.1.2}$$

for $(u_1, \ldots, u_d) \in [0,1]^d$ and where

$$V(z_1, \ldots, z_d) = d\int_{\mathcal{S}_{d-1}} \max_{i=1,\ldots,d}\left\{\frac{w_i}{z_i}\right\} \mathrm{d}H(\mathbf{w}), \quad (z_1, \ldots, z_d) \in (0,\infty)^d, \tag{4.1.3}$$

where $\mathcal{S}_{d-1} = \{\mathbf{w} \in [0,1]^d : \sum_{i=1}^d w_i = 1\}$ is the $(d-1)$-dimensional unit simplex, and H is a valid multivariate distribution function or probability measure for $\mathbf{w} \in \mathcal{S}_{d-1}$ with 1-dimensional marginal expectations $1/d$. Inverting each of the marginal distributions of (4.1.2) gives the inverted extreme value copula (Wadsworth and Tawn, 2012), which is defined through its survival copula

$$\bar{C}_{iev}\{u_1, \ldots, u_d\} = \exp\left\{-V\left(-\frac{1}{\log(1-u_1)}, \ldots, -\frac{1}{\log(1-u_d)}\right)\right\}, \tag{4.1.4}$$

for $(u_1, \ldots, u_d) \in [0,1]^d$ and V as above.

In Section 3.3.2, we found that the driving behaviour of R_2 was often linked to the coefficient of tail dependence (Ledford and Tawn, 1997), denoted η, which is a measure that characterises extremal dependence between two random variables. In this chapter, we now consider R_d, which is a d-dimensional sum and so we require an extension of η to d-dimensions. Eastoe and Tawn (2012) detail such a measure,

denoted η_d, which is defined through the assumption that

$$\Pr\left(\min_{i=1,\ldots,d}\{F_i(X_i)\} > 1 - \frac{1}{u}\right) = L(u)u^{-1/\eta_d}, \text{ as } u \to \infty, \tag{4.1.5}$$

where F_i denotes the CDF of X_i and L is a slowly-varying function. Whilst $\eta_d \in (0, 1]$, we constrain our focus to $\eta_d \in (1/d, 1]$ which corresponds to positive extremal association only. The value of η_d implies asymptotic dependence between all components of $\mathbf{X}$ when $\eta_d = 1$. The cases $1/d < \eta_d < 1$ and $\eta_d = 1/d$ correspond to asymptotic independence and near extremal independence of the full collection of d variables, respectively.

Theoretical values of η_d can be derived for the copulae discussed above; the extreme value, and perfect dependence, copulae exhibit asymptotic dependence and hence give $\eta_d = 1$. For those copulae exhibiting asymptotic independence, we have $\eta_d = 1/V(1, \ldots, 1)$, $\eta_d = [\mathbb{1}_d^T \Sigma^{-1} \mathbb{1}_d]^{-1}$ and $\eta_d = 1/d$ for the inverted extreme value copula, standard Gaussian copula and independence copula, respectively. Note that $\mathbb{1}_d$ denotes a d-vector of ones.

The rest of this chapter is structured as followed: in Section 4.2, we detail the strategies and assumptions that are required to prove the results we derive and present in Section 4.3; these pertain to the form of the survival function $\Pr\{R_d \geq r\}$ as $r \to r^F$, where r^F denotes the upper-endpoint of R_d. The proofs for these results are given in Section 4.4 and Appendix B; the former contains proofs for the main results in Section 4.3 which pertain to R_d with non-zero marginal shapes; the latter contains the proofs for R_d where $\mathbf{X}$ has exponential margins, i.e., $\xi_i = 0$ for all $i = 1, \ldots, d$, and proofs of technical results required for the main proofs.

4.2 Strategy and assumptions

4.2.1 Auxiliary variables

Following the same strategy as used in Chapter 3, we begin by deriving the joint density of $(X_1, \ldots, X_d)$ and conducting some transformation to $R_d = \sum_{i=1}^{d} X_i$ and an auxiliary variable, which we typically denote by $\mathbf{W} = (W_1, \ldots, W_{d-1})$; this variable is chosen so that we can analytically integrate $\mathbf{W}$ out of the joint density of $(R_d, \mathbf{W})$ as R_d tends to its upper-endpoint. We first provide a discussion of the different forms for $\mathbf{W}$ and the transformations used to achieve them.

Pseudo-radial and -angular components

The most common transformation used is one to pseudo-radial and -angular components, namely $\mathbf{X} \to (R_d, \mathbf{W})$, where

$$\mathbf{W} = \left\{ W_j = \frac{X_j}{\sum_{i=1}^{d} X_i}; j = 1, \ldots, d-1 \right\}, \tag{4.2.1}$$

which can be rearranged to give

$$X_i = RW_i, \quad X_d = R_d \left(1 - \sum_{j=1}^{d} W_j \right),$$

for all $i = 1, \ldots, d-1$. For the sake of notation, we introduce W_d as a place holder for $1 - \sum_{j=1}^{d} W_j$. The determinant of the Jacobian of this transformation is r^{d-1}, and a proof of this is provided in Appendix B.1.1.

Scaled pseudo-angular components

In cases where the marginal shape parameters are negative, i.e., $\xi_i < 0$ for all $i = 1, \ldots, d$, we use scaled and shifted pseudo-angular components. This is to ensure that the range of $\mathbf{W}$ is independent of R_d, given that R_d is greater than some fixed threshold

to be defined. We first use the transformation $\mathbf{X} \to \mathbf{U}$, where for all $i = 1, \ldots, d$,

$$U_i = 1 + \xi_i X_i / \sigma_i, \tag{4.2.2}$$

and where $\prod_{i=1}^{d} \frac{\sigma_i}{\xi_i}$ is the determinant of the Jacobian for this transformation. A second transformation, $\mathbf{U} \to (R_d, \mathbf{W})$, is then applied, where

$$R_d = \sum_{i=1}^{d} X_i = \sum_{i=1}^{d} \frac{\sigma_i (U_i - 1)}{\xi_i} \quad \text{and} \quad W_j = \frac{\sigma_j U_j / \xi_j}{\sum_{i=1}^{d} \sigma_i U_i / \xi_i}, \tag{4.2.3}$$

for all $j = 1, \ldots, d-1$. This can then be rearranged to give, for all $j = 1, \ldots, d-1$,

$$U_j = \frac{\xi_j (\sum_{i=1}^{d} \frac{\sigma_i}{\xi_i} + R_d) W_j}{\sigma_j} \quad \text{and} \quad U_d = \frac{\xi_d (\sum_{i=1}^{d} \frac{\sigma_i}{\xi_i} + R_d) W_d}{\sigma_d}, \tag{4.2.4}$$

where $W_d = 1 - \sum_{j=1}^{d} W_j$. The determinant of the Jacobian for this transformation is

$$\left(\sum_{i=1}^{d} \frac{\sigma_i}{\xi_i} + r \right)^{d-1} \prod_{i=1}^{d} \frac{\xi_i}{\sigma_i}.$$

Combining both of these transformations, it can be shown that the support of $\mathbf{W}$ is $[0, 1]^{d-1}$, independent of the value R_d given that $R_d > t$, where

$$t = -\min_{1 \le j \le d} \left\{ \sum_{i \ne j}^{d} \frac{\sigma_i}{\xi_i} \right\} = \max_{1 \le j \le d} \left\{ \frac{\sigma_j}{\xi_j} \right\} - \sum_{i=1}^{d} \frac{\sigma_i}{\xi_i} = -\min_{1 \le j \le d} \left\{ -\frac{\sigma_j}{\xi_j} \right\} - \sum_{i=1}^{d} \frac{\sigma_i}{\xi_i}. \tag{4.2.5}$$

In Section A.3.1, we proved this property for the case where $d = 2$. The proof for arbitrary $d \in \mathbb{N}$ is given in Appendix B.1.2.

Pseudo-angular to radial power

In certain cases, we find that it is helpful to re-write the auxiliary variable as some power of R_d. In those cases where R_d has no upper-endpoint, we use the transforma-

tion $(R_d, \mathbf{W}) \to (R_d, \mathbf{Z})$ for $\mathbf{Z} \in (0, \infty)^{d-1}$, where

$$W_i = R_d^{-Z_i} \tag{4.2.6}$$

for all $i = 1, \ldots, d-1$. This transformation has Jacobian with determinant $r^{-z}\{\log(r)\}^{d-1}$.

In cases where R_d is bounded above by finite $r^F > 0$, we instead consider the transformation $(R_d, \mathbf{W}) \to (V, \mathbf{Z})$ for $V \in [0, 1]$ and $\mathbf{Z} \in (0, \infty)^{d-1}$, where

$$V = (r^F - R_d)/r^F, \text{ and } W_i = V^{Z_i}, \tag{4.2.7}$$

for all $i = 1, \ldots\ d-1$. The Jacobian of this transformation has determinant $1/r^F \times \{-\log(v)\}^{d-1} \prod_{j=1}^{d-1} v^{z_j}$, where $1/r^F$ corresponds to the transformation $R_d \to V$ and the other terms correspond to the transformation $\mathbf{W} \to \mathbf{Z}$.

4.2.2 Assumptions on the behaviour of the extreme value copula

The densities for the extreme value, and inverted extreme value, copulae are particularly complex; the full d-dimensional density for both copulae is the sum of the $d-$th Bell number of terms, i.e., the number of possible partitions for the set $\{1, \ldots, d\}$. Thus, use of the full density in an analytical framework is infeasible for a general $d \in \mathbb{N}$. For the extreme value copula, we are able to make reasonable assumptions on how the density acts for large R_d which allows us to identify only the important terms in the density that are necessary for investigating the first-order behaviour of the upper-tail of R_d.

To derive the results for R_d where $\mathbf{X}$ has an extreme value copula, we make assumptions on how the density of $\mathbf{X}$ acts for large R_d. We begin by making the assumption that the measure/distribution H defined in (4.1.3) places some mass/density on the

centre of the simplex $\mathcal{S}_{d-1}$; equivalently, we have

$$\Pr\left\{\min_{i\neq j}\{F_i(X_i)\} > 1-\frac{1}{u}\middle| F_j(X_j) > 1-\frac{1}{u}\right\} > 0,$$

as $u \to \infty$ and for all $j = 1, \ldots, d$. We now make the assumption that as $R \to r^F$ that all components of $\mathbf{X}$ tends to their respective upper-endpoints with associated rates determined by the following; we assume that $\bar{F}_1(x_1)/\bar{F}_j(x_j) \sim c_j$ as $\bar{F}_1(x_1) \to 0$ for constants $c_j > 0$ and for $j = 1, \ldots, d$, i.e., each component of $\mathbf{X}$ tends to its respective upper-endpoint at a similar rate, and that $R \to r^F \Rightarrow \bar{F}_1(x_1) \to 0$. Now consider the distribution function for the multivariate extreme value copula, which is given by combining (4.1.1) and (4.1.2). Then the density of $\mathbf{X}$ is

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \left\{\prod_{i=1}^{d}\frac{f_i(x_i)}{F_i(x_i)(\log(F_i(x_i)))^2}\right\}\exp\{-V(-\{\log(F_1(x_1))\}^{-1},\ldots,-\{\log(F_d(x_d))\}^{-1})\} \\
&\quad\times\sum_{\pi\in\mathcal{P}}\prod_{\pi_I\in\pi}\{-V_{\pi_I}(-\{\log(F_1(x_1))\}^{-1},\ldots,-\{\log(F_d(x_d))\}^{-1})\} \\
&\sim \left\{\prod_{i=1}^{d}\frac{f_i(x_i)}{\bar{F}_i^2(x_i)}\right\}\exp\{-V([\bar{F}_1(x_1)]^{-1},\ldots,[\bar{F}_d(x_d)]^{-1})\} \\
&\quad\times\sum_{\pi\in\mathcal{P}}\prod_{\pi_I\in\pi}\{-V_{\pi_I}([\bar{F}_1(x_1)]^{-1},\ldots,[\bar{F}_d(x_d)]^{-1})\} \\
&\sim \left\{\prod_{i=1}^{d}\frac{f_i(x_i)}{\bar{F}_i^2(x_i)}\right\}\sum_{\pi\in\mathcal{P}}\prod_{\pi_I\in\pi}\{-V_{\pi_I}([\bar{F}_1(x_1)]^{-1},\ldots,[\bar{F}_d(x_d)]^{-1})\}, \qquad (4.2.8)
\end{aligned}$$

as $r \to r^F$ and where $\mathcal{P}$ is the set of all partitions of $\{1, \ldots, d\}$ and $\pi = \{\pi_1, \ldots, \pi_k\} \in \mathcal{P}$ is one of these partitions and V_{π_I} is the partial derivative of V with respect to all indices in $\{\pi_I\}$; the second line follows as

$$-\{\log(F_i(x_i))\}^{-1} = -\{\log(1-\bar{F}_i(x_i))\}^{-1} \sim [\bar{F}_i(x_i)]^{-1},$$

as $\bar{F}_j(x_j) \to 0$ and $F_j(x_j) \sim 1$ for all $j = 1, \ldots, d$ and the last line follows as

$$\exp\{-V(\bar{F}_1^{-1}(x_1), \ldots, \bar{F}_d^{-1}(x_d))\} \sim 1 \quad \text{as } (\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)) \to \mathbf{0}^d.$$

We then have

$$f_{\mathbf{X}}(\mathbf{x}) \sim \left\{ \prod_{i=1}^{d} \frac{f_i(x_i)}{\bar{F}_i^2(x_i)} \right\} \sum_{\pi \in \mathcal{P}} [\bar{F}_1(x_1)]^{k_\pi} \prod_{\pi_I \in \pi} \{-V_{\pi_I}(1, c_2, \ldots, c_d)\}, \tag{4.2.9}$$

as $\bar{F}_1(x_1) \to 0$; note that k_π is equal to the negative of the order of homogeneity for the expression $\prod_{\pi_I \in \pi} -V_{\pi_I}([\bar{F}_1(x_1)]^{-1}, \ldots, [\bar{F}_d(x_d)]^{-1})$. As V is a homogeneous function of order -1, it follows that its n-th order partial derivatives are homogeneous order $-(n+1)$. From the properties of homogeneous functions, we have that the product of a function of order $-a$ and a function of order $-b$ gives a function of order $-(a+b)$ and so it follows that the term in the summation in (4.2.9) with the highest order of homogeneity is V_{π_I}, where $\pi_I = \{1, \ldots, d\}$ is the partition with all indices in $1, \ldots, d$; this term is non-zero if H has positive mass/density at the centre of $\mathcal{S}_{d-1}$. This corresponds to the d-th order partial derivative of V with respect to all of its components, which from now we denote $V_{\mathbf{x}}$, and has order of homogeneity $-(d+1)$. We then have

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &\sim \left\{ \prod_{i=1}^{d} \frac{f_i(x_i)}{\bar{F}_i^2(x_i)} \right\} [\bar{F}_1(x_1)]^{-(d+1)} \{-V_{\mathbf{x}}(1, c_2, \ldots, c_d)\} \\ &\sim \left\{ \prod_{i=1}^{d} \frac{f_i(x_i)}{\bar{F}_i^2(x_i)} \right\} \{-V_{\mathbf{x}}([\bar{F}_1(x_1)]^{-1}, \ldots, [\bar{F}_d(x_d)]^{-1})\}, \end{aligned}$$

as $r \to r^F$. We assume that form (4.2.9) holds for the proofs in Section 4.4.3 and Appendix B.3.3.

In the case where $\max\{\xi_1, \ldots, \xi_d\} < 0$, we require a further assumption on $V_{\mathbf{x}}$ and its lower order derivatives. For $\mathcal{D} \subset \{1, \ldots, d\}$, let $V_{\mathbf{x}_{-\mathcal{D}}}$ denote the $(d - |\mathcal{D}|)$−th order

partial derivative of V with respect to the components indexed by $\{1,\ldots,d\}\setminus\mathcal{D}$. Now let $\mathbf{z}_I$ denote the components of $\mathbf{z}=(z_1,\ldots,z_d)\in(0,\infty]^d$ indexed by the set I, and consider the function

$$l(\mathbf{z})=V_{\mathbf{x}_{-\mathcal{D}}}(z_1,\ldots,z_d).$$

Under the assumption that each component of $\mathbf{z}_{\{1,\ldots,d\}\setminus\mathcal{D}}$ is fixed at a non-zero, finite value, then we assume that the following properties hold for any $\mathcal{D}$: firstly, we have that $0<l(\mathbf{z}^*)<\infty$ for $\mathbf{z}^*=\{\mathbf{z}:\mathbf{z}_{\mathcal{D}}=(\infty,\ldots,\infty)\}$; secondly, we assume that $l(\mathbf{z})\to 0$ if $z_i\to 0$ for any $i\in\mathcal{D}$. Whilst we omit discussion for general V, we note that these two properties are satisfied by the logistic copula, see (2.2.2), and so may be applicable to other extreme value copulae.

4.2.3 Laplace's method

Some of the proofs in Section 4.4 require solutions to integrals of the form

$$\int_\Omega h(\mathbf{w},r)\exp\{-L(r)g(\mathbf{w},r)\}\mathrm{d}\mathbf{w},$$

where $\mathbf{w}=(w_1,\ldots,w_{d-1})\in\Omega$. Here h is a positive function, the Hessian for the function g with respect to $\mathbf{w}$, which we denote $G(\mathbf{w},r)$, exists and is negative definite and the function $L(r)\to\infty$ as $r\to r^F$. We further require that the global minima of $L(r)g(\mathbf{w},r)+\log(h(\mathbf{w},r))$ with respect to $\mathbf{w}$ exists and is unique. For integrals of this form, Laplace (1986) gives the approximate solution as

$$\int_\Omega h(\mathbf{w},r)\exp\{-L(r)g(\mathbf{w},r)\}\mathrm{d}\mathbf{w}\sim\left(\frac{2\pi}{L(r)}\right)^{(d-1)/2}\frac{h(\mathbf{w}^*,r)}{|-G(\mathbf{w}^*,r)|^{1/2}}\exp\{-L(r)g(\mathbf{w}^*,r)\},$$

as $L(r)\to\infty$ and where $\mathbf{w}^*=\arg\min_{\mathbf{w}\in\Omega}\{L(r)g(\mathbf{w},r)+\log(h(\mathbf{w},r))\}$ and $|\cdot|$ denotes the determinant of a matrix. For derivations that require Laplace's method, we have assumed that $\mathbf{w}^*$ without explicitly statement.

4.3 Results

The results for $\Pr\{R_d \geq r\}$ as $r \to r^F$ are presented in a similar form to that given in (3.2.3); that is, we have

$$\Pr\{R_d \geq r\} \sim \begin{cases} \left(\frac{\xi_R}{\sigma_R}\right)^{-1/\xi_R} L_1(r) r^{-1/\xi_R}, & \text{if } \xi_R > 0, \\ K_1 r^{\alpha_R} \exp\left(-\frac{r}{\sigma_R}\right), & \text{if } \xi_R = 0, \\ K_2 L_2(r) \left(1 - \frac{r}{r^F}\right)^{-1/\xi_R}, & \text{if } \xi_R < 0, \end{cases} \tag{4.3.1}$$

as $r \to r^F$ which is infinite if $\xi_R \geq 0$ and finite when $\xi_R < 0$. Here $\sigma_R > 0$ and $\alpha_R \geq 0$ and $K_1, K_2 > 0$ are proportionality constants. The form (4.3.1) given here differs from the form (3.2.3) proposed in Chapter 3 by the introduction of slowly varying functions $L_1(r), L_2(r)$ and the term r^{α_R}. Note that these terms do not change the dominant behaviour of $\Pr\{R_d \geq r\}$ as $r \to r^F$; the power term r^{α_R} varies slower than the exponential term $\exp(-r/\sigma_R)$, and both L_1 and L_2 vary slower than the corresponding power terms. We further note that for the considered copulae, that $r^F = \sum_{i=1}^d x_i^F$, where x_i^F is the upper-endpoint of X_i for $i = 1, \ldots, d$; if $x_i^F = \infty$ for any $i \in \{1, \ldots, d\}$, then $r^F = \infty$.

In Tables 4.3.1-4.3.5, we tabulate the values of the parameters ξ_R, σ_R and α_R given in (4.3.1) that arise from modelling dependence in $\mathbf{X}$ with each of the five copulae presented in Section 4.1 and for a number of different cases for the marginal GPD scale and shape parameters, ξ_i and σ_i, respectively. The five cases we consider are:

- $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$,
- $\xi_i = \xi < 0$ for all $i = 1, \ldots, d$,
- $\xi_i < 0$ for all $i = 1, \ldots, d$,
- $\xi_i = 0, \sigma_i = \sigma$ for all $i = 1, \ldots, d$,

- $\xi_i = 0$ for all $i = 1, \ldots, d$.

In the cases where $\max_{1 \le i \le d} \{\xi_i\} < 0$ and the copula is asymptotically independent, but not independent, we give the results for $d = 2$ only. For those cases where $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$, and the copula is asymptotically independent, the proofs have been conducted for $d = 2$ only; however, we postulate that this behaviour can be extended to any finite $d \in \mathbb{N}$, and so present the results as such. We further recall that all results for the extreme value copula hold only if the assumptions described in Section 4.2.2 are met.

$\xi_i = \xi > 0$ for all $i = 1, \ldots, d$

Copula	σ_R	ξ_R
Independence	$\left(\sum_{i=1}^{d} \sigma_i^{1/\xi}\right)^{\xi}$	$\vdots$
Perfect dependence	$\sum_{i=1}^{d} \sigma_i$	$\vdots$
Extreme value copula	K^{ξ}	ξ
Inverted extreme value copula	$\left(\sum_{i=1}^{d} \sigma_i^{1/\xi}\right)^{\xi}$	$\vdots$
Gaussian $(\mathbf{0}, \Sigma)$	$\left(\sum_{i=1}^{d} \sigma_i^{1/\xi}\right)^{\xi}$	$\vdots$

Table 4.3.1: Theoretical parameter values for the first-order behaviour of $\Pr\{R_d > r\}$ as $r \to \infty$ if $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$. The constant K is defined in (4.4.11).

$\xi_i = \xi < 0$ for all $i = 1, \ldots, d$

Copula	ξ_R
Independence	ξ/d
Perfect dependence	ξ
Extreme value copula	ξ
Inverted extreme value copula ($d = 2$)	$\xi/V(1,1)$
Gaussian $(\mathbf{0}, \Sigma)$ ($d = 2$)	$\xi \left\{ \mathbb{1}_2 \Sigma^{-1} \mathbb{1}_2 \right\}^{-1}$

Table 4.3.2: Theoretical parameter values for the first-order behaviour of $\Pr\{R_d > r\}$ as $r \to \infty$ if $\xi_i = \xi < 0$ for all $i = 1, \ldots, d$. The exponent V is defined in (4.1.3) and $\mathbb{1}_2$ denotes a 2-vector of ones.

We note that the results in Table 4.3.2 can be summarised for all copulas as $\xi_R = \eta_d \xi$, where η_d is defined in (4.1.5).

$\xi_i < 0$ for all $i = 1, \ldots, d$

Copula	ξ_R
Independence	$\left(\sum_{i=1}^{d} \frac{1}{\xi_i} \right)^{-1}$
Perfect dependence	$\max_{i=1,\ldots,d} \{\xi_i\}$
Extreme value copula*	$\max_{i=1,\ldots,d} \{\xi_i\}$
Inverted extreme value copula ($d = 2$)	$-1/V(-\xi_1, -\xi_2)$
Gaussian $(\mathbf{0}, \Sigma)$ ($d = 2$)	$-\{A^T \Sigma^{-1} A\}^{-1}$

Table 4.3.3: Theoretical parameter values for the first-order behaviour of $\Pr\{R_d > r\}$ as $r \to \infty$ if $\xi_i < 0$ for all $i = 1, \ldots, d$. The exponent V is defined in (4.1.3) and $A = (\sqrt{-1/\xi_1}, \sqrt{-1/\xi_2})^T$. For the extreme value copula, we have $|\arg\max_{i=1,\ldots,d} \xi_i| = 1$.

We note that if $\xi_i = \xi$ for all $i = 1, \ldots, d$, then each of the results given in Table 4.3.3 is equal to the results in Table 4.3.2.

$\xi_i = 0$ and $\sigma_i = \sigma$ for all $i = 1, \ldots, d$

Copula	α_R	σ_R
Independence	$d - 1$	σ
Perfect dependence	0	$d\sigma$
Extreme value copula	0	$d\sigma$
Inverted extreme value copula (symmetric V)	$(d-1)/2$	$d\sigma/V(1, \ldots, 1)$
Gaussian $(\mathbf{0}, \Sigma)$ ($\Sigma_{i,j} = \rho$ for all $i \neq j$)	$(\mathbb{1}^T\Sigma^{-1}\mathbb{1})^{-1}/2$	$d\sigma(\mathbb{1}^T\Sigma^{-1}\mathbb{1})^{-1}$

Table 4.3.4: Theoretical parameter values for the first-order behaviour of $\Pr\{R_d > r\}$ as $r \to \infty$ if $\xi_i = 0$ and $\sigma_i = \sigma$ for all $i = 1, \ldots, d$. The exponent V is defined in (4.1.3) and must be a symmetric function, and $\mathbb{1}_d$ denotes a d-vector of ones and $\rho \in [0, 1)$.

Note that, similarly to Table 4.3.2, the scale parameters presented in Table 4.3.4 can be written in terms of η_d, namely $\sigma_R = d\eta_d\sigma$. However, this property does not extend to the case where the marginal scale parameters are heterogeneous; these results are provided below. In the case of the inverted extreme value, and Gaussian, copulas, we have further constraints; for the former, we require that V is a symmetric function, and for the latter, we require that the correlation parameters are homogeneous, i.e., $\Sigma_{i,j} = \rho$ for all $i \neq j$. For cases where these constraints are not met, the corresponding σ_R is presented in Table 4.3.5.

$\xi_i = 0$ for all $i = 1, \ldots, d$

Copula	α_R	σ_R
Independence	$\lvert\{\arg\max_{j=1,\ldots,d}\{\sigma_j\}\}\rvert - 1$	$\max_{i=1,\ldots,d}\{\sigma_i\}$
Perfect dependence	0	$\sum_{i=1}^{d}\sigma_i$
Extreme value copula	0	$\sum_{i=1}^{d}\sigma_i$
Inverted extreme value copula	$(d-1)/2$	$\min_{\mathbf{w}} V\left(\frac{\sigma_1}{w_1}, \ldots, \frac{\sigma_d}{w_d}\right)$
Gaussian $(\mathbf{0}, \Sigma)$	α_G	σ_G

Table 4.3.5: Theoretical parameter values for the first-order behaviour of $\Pr\{R_d > r\}$ as $r \to \infty$ if $\xi_i = 0$ for all $i = 1, \ldots, d$ and where $\mathbf{w} \in \mathcal{S}_{d-1}$. The exponent V and $\mathcal{S}_{d-1}$ are defined in (4.1.3), and V must satisfy the property that $\arg\min_{\mathbf{w}\in\mathcal{S}_{d-1}} V\left(\frac{\sigma_1}{w_1}, \ldots, \frac{\sigma_d}{w_d}\right)$ is unique. The Gaussian parameters α_G and σ_G are given in (B.3.11) in Appendix B.3.5.

We note that if $\sigma_i = \sigma$ for all $i = 1, \ldots, d$ (and, for the Gaussian copula, $\Sigma_{i,j} = \rho$ for all $i \neq j$), then the results in Table 4.3.5 are equal to those presented in Table 4.3.4.

The property of which the parameters ξ_R and σ_R in Tables 4.3.2 and 4.3.4, respectively, are functions of η_d only occurs when the tails of the marginal variables decay at an equal rate; that is, if $\xi_R < 0$ we require that $\xi_i = \xi < 0$ for all $i = 1, \ldots, d$; similarly, if $\xi_R = 0$, we require $\sigma_i = \sigma$ for all $i = 1, \ldots, d$.

The proofs for the results in Tables 4.3.1, 4.3.2 and 4.3.3 are provided in Section 4.4. The proofs for the results in Tables 4.3.4 and 4.3.5 are provided in Appendix B.3.

4.4 Proofs

This section details some of the proofs of the results detailed in Section 4.3. The proofs are presented in the following manner: each subsection considers a different copula, and then within each subsection we consider two cases separately; these cases are $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$ and $\xi_i < 0$ for all $i = 1, \ldots, d$. In the case where the

marginal shape parameters are positive, we consider equality for all $i = 1, \ldots, d$ only.

4.4.1 Perfect dependence

Strictly positive marginal shapes

Perfect dependence in $\mathbf{X}$ is induced by letting $X_1 \sim \text{GPD}\{\sigma_1, \xi\}$ and then setting $X_i = \sigma_i X_i/\sigma_1$ for all $i = 1, \ldots, d$. Then if $R_d = \sum_{i=1}^{d} X_i = \sum_{i=1}^{d} \frac{\sigma_i}{\sigma_1} X_1$, it follows that

$$\Pr\{R_d \geq s\} = \Pr\left\{\sum_{i=1}^{d} \frac{\sigma_i}{\sigma_1} X_1 \geq s\right\} = \Pr\left\{X_1 \geq s\sigma_1 \left[\sum_{i=1}^{d} \sigma_i\right]^{-1}\right\} \tag{4.4.1}$$

$$= \left(1 + \xi s \left[\sum_{i=1}^{d} \sigma_i\right]^{-1}\right)^{-1/\xi} \sim \left(\frac{\xi}{\sum_{i=1}^{d} \sigma_i}\right)^{-1/\xi} s^{-1/\xi}, \tag{4.4.2}$$

as $s \to \infty$ and as needed.

Strictly negative marginal shapes

Let $X_1 \sim \text{GPD}\{\sigma_1, \xi_1\}$, for $\xi_1 < 0$ and then let

$$X_i = \frac{\sigma_i}{\xi_i}\left(-1 + \left(1 + \frac{\xi_1}{\sigma_1} X_1\right)^{\xi_i/\xi_1}\right),$$

for $\xi_i < 0$ and for all $i = 1, \ldots, d$. It follows that $X_i \sim \text{GPD}\{\sigma_i, \xi_i\}$ with perfect dependence induced in $\mathbf{X}$. We can then write R_d as

$$R_d = \sum_{i=1}^{d} X_i = \sum_{i=1}^{d} \frac{\sigma_i}{\xi_i}\left(-1 + \left(1 + \frac{\xi_1}{\sigma_1} X_1\right)^{\xi_i/\xi_1}\right) = r^F + \sum_{i=1}^{d} \frac{\sigma_i}{\xi_i}\left(1 + \frac{\xi_1}{\sigma_1} X_1\right)^{\xi_i/\xi_1}, \tag{4.4.3}$$

where $r^F = -\sum_{i=1}^{d} \sigma_i/\xi_i$ is the upper-endpoint of R_d. We now consider $X_1 \to -\sigma_1/\xi_1$, i.e., the upper end-point of X_1. Subsequently, all other X_i, for $i = 2, \ldots, d$, approach their respective upper end-points, and thus R_d approaches r^F. Defining $J = \arg\max_{1 \leq i \leq d}\{\xi_i\}$ and $\xi_{max} = \max_{1 \leq i \leq d}\{\xi_i\}$, and considering a realisation r of R_d,

it follows that

$$r \sim r^F + \sum_{j\in J} \frac{\sigma_j}{\xi_{max}} \left(1 + \frac{\xi_1}{\sigma_1} x_1\right)^{\xi_{max}/\xi_1} \sim r^F + \frac{1}{\xi_{max}} \left(1 + \frac{\xi_1}{\sigma_1} x_1\right)^{\xi_{max}/\xi_1} \sum_{j\in J} \sigma_j,$$

as $1 + \xi_1 x_1/\sigma_1 \to 0$, and hence

$$\begin{aligned}
\Pr\{R_d \geq s\} &\sim \Pr\left\{ r^F + \frac{1}{\xi_{max}} \left(1 + \frac{\xi_1}{\sigma_1} X_1\right)^{\xi_{max}/\xi_1} \sum_{j\in J} \sigma_j \geq s \right\} \\
&\sim \Pr\left\{ X_1 \geq \frac{\sigma_1}{\xi_1} \left\{ -1 + \left[\frac{\xi_{max}}{\sum_{j\in J} \sigma_j} \left(s - r^F\right) \right]^{\xi_1/\xi_{max}} \right\} \right\} \\
&\sim \left(1 + \frac{\xi_1}{\sigma_1} \frac{\sigma_1}{\xi_1} \left\{ -1 + \left[\frac{\xi_{max}}{\sum_{j\in J} \sigma_j} \left(s - r^F\right) \right]^{\xi_1/\xi_{max}} \right\} \right)^{-1/\xi_1} \\
&\sim \left\{ \left[\frac{\xi_{max}}{\sum_{j\in J} \sigma_j} \left(s - r^F\right) \right]^{\xi_1/\xi_{max}} \right\}^{-1/\xi_1} = \left\{ \frac{\xi_{max}}{\sum_{j\in J} \sigma_j} \left(s - r^F\right) \right\}^{-1/\xi_{max}} \\
&\sim \left\{ -\frac{\xi_{max}}{\sum_{j\in J} \sigma_j} \left(r^F - s\right) \right\}^{-1/\xi_{max}} = \left\{ -\frac{\xi_{max} r^F}{\sum_{j\in J} \sigma_j} \right\}^{-1/\xi_{max}} \left(1 - \frac{s}{r^F}\right)^{-1/\xi_{max}},
\end{aligned}$$

as $s \to r^F$ and as required.

4.4.2 Independence

Strictly positive marginal shapes

For the independence copula with $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$, we have by induction that, as $s \to \infty$,

$$\Pr\{R_d \geq s\} \sim \left(\frac{\xi}{\sigma_{R_d}}\right)^{-1/\xi} s^{-1/\xi}, \tag{4.4.4}$$

where $\sigma_{R_d} = \left(\sum_{k=1}^{d} \sigma_k^{1/\xi}\right)^{\xi}$ That is, without loss of generality we can derive the result for $R_{d-1} = \sum_{k=1}^{d-1} X_i$ and then derive the result for R by considering $R_d = X_d + R_{d-1}$.

We begin with the base case, i.e., $d = 1$. For $X_1 \sim \text{GPD}(\sigma_1, \xi)$, we have

$$\Pr\{R_1 \geq s\} = \Pr\{X_1 \geq s\} = \left\{1 + \frac{\xi}{\sigma_1} s\right\}^{-1/\xi} \sim \left\{\frac{\xi}{\sigma_1} s\right\}^{-1/\xi} = \left\{\frac{\xi}{\sigma_1}\right\}^{-1/\xi} s^{-1/\xi},$$

as $s \to \infty$. Thus (4.4.4) holds and we have $\sigma_{R_1} = \sigma_1 = \left(\sum_{k=1}^{1} \sigma_1^{1/\xi}\right)^{\xi}$, as needed. As the base case holds, we assume that the result in (4.4.4) holds for R_{d-1} with and consider $R_d = R_{d-1} + X_d$. The joint density of (R_{d-1}, X_d) is

$$\begin{aligned} f_{R_{d-1},X_d}(r, x_d) &\sim \frac{1}{\sigma_{R_{d-1}}\sigma_d} \left\{\frac{\xi}{\sigma_{R_{d-1}}}\right\}^{-1/\xi} r^{-1/\xi-1} \left\{1 + \frac{\xi x_d}{\sigma_d}\right\}^{-1/\xi-1} \\ &\sim \frac{1}{\sigma_{R_{d-1}}\sigma_d} \left\{1 + \frac{\xi}{\sigma_{R_{d-1}}} r\right\}^{-1/\xi-1} \left\{1 + \frac{\xi x_d}{\sigma_d}\right\}^{-1/\xi-1}, \end{aligned} \tag{4.4.5}$$

as $r_{d-1} \to \infty$ and for any $x_d \in [0, \infty)$. We now use the transformation given by (4.2.2), i.e., $(R_{d-1}, X_d) \to (U, V)$. The joint density of (U, V) is $f_{U,V}(u, v) \sim \frac{1}{\xi^2}(uv)^{-1/\xi-1}$ as $u \to \infty$ and for $v > 1$. We then apply the second transformation $(U, V) \to (R_d, W)$ given by (4.2.3), where $R_d = R_{d-1} + X_d$ and $W = (\sigma_{R_{d-1}}/\xi + R_d)/(\sigma_{R_{d-1}}/\xi + R_d + \sigma_d/\xi + X_d)$. The joint density of (R_d, W) is

$$f_{R_d,W}(r, w) \sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} \left(\frac{\sigma_{R_{d-1}}}{\xi} + \frac{\sigma_d}{\xi} + r\right)^{-2/\xi-1} [w(1-w)]^{-1/\xi-1},$$

as $r \to \infty$ and for $w \in [t_1, 1 - t_2]$, where

$$t_1 = \frac{\sigma_{R_{d-1}}}{\xi} \left(\frac{\sigma_{R_{d-1}}}{\xi} + \frac{\sigma_d}{\xi} + r\right)^{-1}, \quad t_2 = \frac{\sigma_d}{\xi} \left(\frac{\sigma_{R_{d-1}}}{\xi} + \frac{\sigma_d}{\xi} + r\right)^{-1}.$$

The marginal density of R_d is then

$$f_{R_d}(r) = \int_{t_1}^{c_1} f_{R_d,W}(r, w)\mathrm{d}w + \int_{c_1}^{1-c_2} f_{R_d,W}(r, w)\mathrm{d}w + \int_{1-c_2}^{1-t_2} f_{R_d,W}(r, w)\mathrm{d}w \tag{4.4.6}$$

where $c_1, c_2 > 0$ are constants chosen such that $c_1 > t_1, c_2 > t_2$ and $c_1 < 1 - c_2$. We then have

$$\begin{aligned} f_{R_d}(r) &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} \left(\frac{\sigma_{R_{d-1}}}{\xi} + \frac{\sigma_d}{\xi} + r\right)^{-2/\xi-1} \int_{t_1}^{1-t_2} [w(1-w)]^{-1/\xi-1}\,\mathrm{d}w \\ &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} r^{-2/\xi-1} \left[\int_{t_1}^{c_1} [w(1-w)]^{-1/\xi-1}\mathrm{d}w + \int_{c_1}^{1-c_2} [w(1-w)]^{-1/\xi-1}\mathrm{d}w \right. \\ &\left. + \int_{1-c_2}^{1-t_2} [w(1-w)]^{-1/\xi-1}\mathrm{d}w\right] \\ &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} r^{-2/\xi-1} \left[\int_{t_1}^{c_1} w^{-1/\xi-1}\mathrm{d}w + K_1 + \int_{1-c_2}^{1-t_2} (1-w)^{-1/\xi-1}\mathrm{d}w\right], \end{aligned}$$

as $r \to \infty$ and for some constant $K_1 > 0$. The last line follows as c_1 and c_2 can be chosen small enough so that $(1 - w) \approx 1$ and $w \approx 1$ on their respective domains. Then

$$\begin{aligned} f_{R_d}(r) &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} r^{-2/\xi-1} \left[\int_{t_1}^{c_1} w^{-1/\xi-1}\mathrm{d}w + K_1 + \int_{1-c_2}^{1-t_2} (1-w)^{-1/\xi-1}\mathrm{d}w\right] \\ &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} r^{-2/\xi-1} \left[\xi(t_1^{-1/\xi} + t_2^{-1/\xi}) + K_2\right] \\ &\sim \frac{\xi^{-2/\xi-2}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} r^{-2/\xi-1} \left[\xi\left(\left\{\frac{\sigma_{R_{d-1}}}{\xi}\right\}^{-1/\xi} + \left\{\frac{\sigma_d}{\xi}\right\}^{-1/\xi}\right)\left(\frac{\sigma_{R_{d-1}}}{\xi} + \frac{\sigma_d}{\xi} + r\right)^{1/\xi} + K_2\right] \\ &\sim \frac{\xi^{-1/\xi-1}}{(\sigma_{R_{d-1}}\sigma_d)^{-1/\xi}} \left(\sigma_{R_{d-1}}^{-1/\xi} + \sigma_d^{-1/\xi}\right) r^{-1/\xi-1} \sim \xi^{-1/\xi-1} \left(\sigma_{R_{d-1}}^{1/\xi} + \sigma_d^{1/\xi}\right) r^{-1/\xi-1}, \end{aligned}$$

as $r \to \infty$ and for constant $K_2 > 0$. It follows that

$$\Pr\{R_d \geq s\} \sim \int_s^\infty \xi^{-1/\xi-1} \left(\sigma_{R_{d-1}}^{1/\xi} + \sigma_d^{1/\xi}\right) r^{-1/\xi-1}\mathrm{d}r = \left(\frac{\xi}{\left[\sigma_{R_{d-1}}^{1/\xi} + \sigma_d^{1/\xi}\right]^\xi}\right)^{-1/\xi} s^{-1/\xi},$$

as $s \to \infty$. Then as $\sigma_{R_d} = \left(\sigma_{R_{d-1}}^{1\xi} + \sigma_d^{1/\xi}\right)^\xi$, we have proven that (4.4.4) holds for any finite d by induction. Note that if $\sigma_i = \sigma$ for all $i = 1, \ldots, d$, then $\sigma_R = d^\xi\sigma$.

Strictly negative marginal shapes

If we instead have $\xi_i < 0$ for all $i = 1, \ldots, d$, then the joint density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{d} \frac{1}{\sigma_i} \left(1 + \frac{\xi_i x_i}{\sigma_i}\right)^{-1/\xi_i - 1},$$

for $0 \leq x_i \leq -\sigma_i/\xi_i$ and all $i = 1, \ldots, d$. We now conduct the two-step transformation given by (4.2.2) and (4.2.3), i.e., $\mathbf{X} \to \mathbf{U} \to (R_d, \mathbf{W})$. Recall that we use this transformation as we are able to proof that the support of $\mathbf{W} \in [0,1]^{d-1}$ is invariant of the value R_d, given that $R_d > t$ for t defined in (4.2.5). Thus, we consider $f_{R_d,\mathbf{W}}$ for $R_d > t$ and as $r \to r^F = -\sum_{i=1}^{d} \sigma_i/\xi_i$. For $\mathbf{w} = (w_1, \ldots, w_{d-1})$, we have

$$f_{R_d,\mathbf{W}}(r, \mathbf{w}) = \left(r^F - r\right)^{-\sum_i 1/\xi_i - 1} g(\mathbf{w}),$$

where

$$g(\mathbf{w}) = \left\{ \prod_{i=1}^{d-1} \frac{(-\xi_i w_i)^{-1/\xi_i - 1}}{\sigma_i^{-1/\xi_i}} \right\} \times \frac{\left(-\xi_d \left[1 - \sum_{i=1}^{d} w_i\right]\right)^{-1/\xi_d - 1}}{\sigma_d^{-1/\xi_d}}.$$

The survival function of R_d is then

$$\begin{aligned} \Pr\{R_d \geq s\} &= \int_s^{r^F} \left(r^F - r\right)^{-\sum_i 1/\xi_i - 1} \mathrm{d}r \int_{\mathbf{w}} g(\mathbf{w}) \mathrm{d}\mathbf{w} \\ &= K \int_s^{r^F} \left(r^F - r\right)^{-\sum_i 1/\xi_i - 1} \mathrm{d}r = \frac{K}{-\sum_{i=1}^{d} 1/\xi_i} \left(r^F - s\right)^{-\sum_i 1/\xi_i}, \end{aligned}$$

where $K = \int_{\mathbf{w}} g(\mathbf{w}) \mathrm{d}\mathbf{w} > 0$ is a finite constant. Note that the integral of $g(\mathbf{w})$ has an analytical solution given by the distribution function of a Dirichlet random variable.

4.4.3 Extreme value copula

Let $X_i \sim \mathrm{GPD}(\sigma_i, \xi_i)$ for $\xi_i \neq 0$ for all $i = 1, \ldots, d$ and with dependence in $\mathbf{X}$ described using the extreme value copula defined in Section 4.1. Following the assumptions made

in Section 4.2.2, the density of $\mathbf{X}$ is of the form

$$f_{\mathbf{X}}(\mathbf{x}) \sim -\left\{\prod_{k=1}^{d} \frac{\left(1+\frac{\xi}{\sigma_k}x_k\right)^{1/\xi-1}}{\sigma_k}\right\} \times V_{\mathbf{x}}\left(\left[1+\frac{\xi}{\sigma_1}x_1\right]^{1/\xi}, \ldots, \left[1+\frac{\xi}{\sigma_d}x_d\right]^{1/\xi}\right), \tag{4.4.7}$$

as each component of $\mathbf{x} = (x_1, \ldots, x_d)$ tends to its respective upper-endpoint and where $V_{\mathbf{x}}$ denotes the d-th partial derivative of V, defined in (4.1.3), with respect to all components.

Strictly positive marginal shapes

We now let $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$ and apply the transformation given by (4.2.1), namely $\mathbf{X} \to (R_d, \mathbf{W})$. The joint density of $(R_d, \mathbf{W})$ is

$$\begin{aligned} f_{R_d,\mathbf{W}}(r, \mathbf{w}) &\sim -r^{d-1}\left\{\prod_{k=1}^{d} \frac{\left(1+\frac{\xi}{\sigma_k}rw_k\right)^{1/\xi-1}}{\sigma_k}\right\} \qquad (4.4.8) \\ &\quad \times V_{\mathbf{x}}\left(\left[1+\frac{\xi}{\sigma_1}rw_1\right]^{1/\xi}, \ldots, \left[1+\frac{\xi}{\sigma_d}rw_d\right]^{1/\xi}\right) \\ &\sim -r^{d-1}\left\{\prod_{k=1}^{d} \frac{\left(\frac{\xi}{\sigma_k}rw_k\right)^{1/\xi-1}}{\sigma_k}\right\} \times V_{\mathbf{x}}\left(\left[\frac{\xi}{\sigma_1}rw_1\right]^{1/\xi}, \ldots, \left[\frac{\xi}{\sigma_d}rw_d\right]^{1/\xi}\right) \\ &\sim -r^{-1/\xi-1}\xi^{d/\xi-d}\left\{\prod_{k=1}^{d} \frac{\left(\frac{w_k}{\sigma_k}\right)^{1/\xi-1}}{\sigma_k}\right\} \times \xi^{-(d+1)/\xi}V_{\mathbf{x}}\left(\left[\frac{w_1}{\sigma_1}\right]^{1/\xi}, \ldots, \left[\frac{w_d}{\sigma_d}\right]^{1/\xi}\right) \\ &\sim r^{-1/\xi-1}\xi^{-1/\xi-1}g(\mathbf{w}) \qquad (4.4.9) \end{aligned}$$

as $r \to \infty$, and where the third line follows as $V_{\mathbf{x}}$ is a homogeneous function of order $-(d+1)$, and

$$g(\mathbf{w}) = -\xi^{-(d-1)}\left\{\prod_{k=1}^{d} \frac{w_k^{1/\xi-1}}{\sigma_k^{1/\xi}}\right\} V_{\mathbf{x}}\left(\left[\frac{w_1}{\sigma_1}\right]^{1/\xi}, \ldots, \left[\frac{w_d}{\sigma_d}\right]^{1/\xi}\right), \tag{4.4.10}$$

for $\mathbf{w} \in \mathcal{S}_{d-1}$, i.e., the $(d-1)$-dimensional unit simplex defined in (4.1.2). It follows that

$$\Pr\{R_d \geq s\} = \int_s^\infty \int_{\mathcal{S}_{d-1}} f_{R_d,\mathbf{W}}(r,\mathbf{w})\mathrm{d}\mathbf{w}\mathrm{d}r \sim \int_s^\infty K\xi^{-1/\xi-1}r^{-1/\xi-1}\mathrm{d}r \sim \left\{\frac{\xi}{K^\xi}\right\}^{-1/\xi} s^{-1/\xi},$$

as $s \to \infty$, and where

$$K = \int_{\mathbf{w}} g(\mathbf{w})\mathrm{d}\mathbf{w} < \infty. \tag{4.4.11}$$

We now give a brief discussion on the scaling constant $K > 0$. If H defined in (4.1.3) is a valid distribution function with related density h, then Coles and Tawn (1991) show that $V_{\mathbf{x}}$ can be redefined in terms of h. We do this now, and rewrite $g(\mathbf{w})$ in (4.4.10) as

$$\frac{\xi^{-(d-1)}}{\left[\sum_{i=1}^d (w_i/\sigma_i)^{1/\xi}\right]^{d+1}} \left\{\prod_{k=1}^d \frac{w_k^{1/\xi-1}}{\sigma_k^{1/\xi}}\right\} h\left(\frac{w_1^{1/\xi}}{\sigma_1^{1/\xi}\sum_{i=1}^d (w_i/\sigma_i)^{1/\xi}}, \ldots, \frac{w_{d-1}^{1/\xi}}{\sigma_{d-1}^{1/\xi}\sum_{i=1}^d (w_i/\sigma_i)^{1/\xi}}\right), \tag{4.4.12}$$

for $\mathbf{w} \in \int_{d-1}$. Using the transformation

$$\omega_j = \frac{w_1^{1/\xi}}{\sigma_1^{1/\xi}\sum_{i=1}^d (w_i/\sigma_i)^{1/\xi}},$$

for all $j = 1, \ldots, d-1$, it can be shown that K in (4.4.11) can be rewritten as

$$K = \int_\Omega \frac{1}{\sum_{i=1}^d f_i(\boldsymbol{\omega})} \times h^*\left(\omega_1, \ldots, \omega_{d-1}, \omega_d\right) \mathrm{d}\boldsymbol{\omega}, \tag{4.4.13}$$

where

$$\omega_d = 1 - \sum_{i=1}^{d-1} \left\{\sum_{j=1}^{d-1} \frac{\sigma_i}{\sigma_j}\left(\frac{\omega_i}{\omega_j}\right)^\xi + \frac{\sigma_d}{\sigma_i}\left(\frac{1-\sum_{k=1}^{d-1}\omega_k}{\omega_i}\right)^\xi\right\}^{-1}, \tag{4.4.14}$$

and

$$f_i(\boldsymbol{\omega}) = \begin{cases} \left\{ \sum_{j=1}^{d-1} \frac{\sigma_i}{\sigma_j} \left(\frac{\omega_i}{\omega_j} \right)^{\xi} + \frac{\sigma_d}{\sigma_i} \left(\frac{1-\sum_{k=1}^{d-1} \omega_k}{\omega_i} \right)^{\xi} \right\}^{-1/\xi} \sigma_i^{-1/\xi} & \text{if, } \; i \neq d, \\ (w_d/\sigma_d)^{1/\xi}, \quad \text{if } \; i = d, \end{cases}$$

and

$$\Omega = \left\{ \boldsymbol{\omega} : \omega_d + \sum_{i=1}^{d-1} \left\{ \sum_{j=1}^{d-1} \frac{\sigma_i}{\sigma_j} \left(\frac{\omega_i}{\omega_j} \right)^{\xi} + \frac{\sigma_d}{\sigma_i} \left(\frac{1 - \sum_{k=1}^{d-1} \omega_k}{\omega_i} \right)^{\xi} \right\}^{-1} = 1 \right\}.$$

We use h^* in (4.4.13) to denote the same density as given in (4.4.12), except with a different constraint on the variables. With abuse of notation, we rewrite h as $h(z_1, \ldots, z_{d-1}) = h(z_1, \ldots, z_{d-1}, z_d)$, where the last variable is constrained so that $z_d = 1 - \sum_{j=1}^{d-1} z_j$, i.e., $\mathbf{z}$ lies in $\mathcal{S}_{d-1}$. The function $h^*(\cdot)$ has the same form as h, except with the simplex constraint replaced with the constraint detailed in (4.4.14). A proof of the transformation of (4.4.11) to (4.4.13) is given in Appendix B.2.

Note that if we have homogeneous scale parameters, i.e., $\sigma_i = \sigma$ for all $i = 1, \ldots, d$, and that the marginal shape parameter is $\xi = 1$, we can use a linear transformation to reclaim the original $(d-1)$-dimensional simplex. This gives

$$K = \int_{\Omega} \frac{1}{\sum_{i=1}^{d} f_i(\boldsymbol{\omega})} \times h^* \left(\omega_1, \ldots, \omega_{d-1}, \omega_d \right) \mathrm{d}\boldsymbol{\omega} = \sigma \int_{\mathcal{S}_{d-1}} h \left(\omega_1, \ldots, \omega_{d-1} \right) \mathrm{d}\boldsymbol{\omega} = d\sigma.$$

Strictly negative marginal shapes

For $\xi_i < 0$ for all $i = 1, \ldots, d$, the joint density of $\mathbf{X}$ is given by (4.4.7). We conduct the two-step transformation given by (4.2.2) and (4.2.3), i.e., $\mathbf{X} \to \mathbf{U} \to (R_d, \mathbf{W})$. Recall that we use this transformation as we are able to illustrate that the support of $\mathbf{W} \in [0,1]^{d-1}$ is independent of the value of R_d, given that $R_d > t$ for $t = r^F + \max_{1 \leq j \leq d} \{\sigma_i / \xi_i\}$ defined in (4.2.5). Thus, we consider $f_{R_d, \mathbf{W}}$ for $R_d > t$ and as

$r \to r^F = -\sum_{i=1}^{d} \sigma_i/\xi_i$. For $\mathbf{w} = (w_1, \ldots, w_{d-1})$, we have

$$f_{R_d,\mathbf{W}}(r, \mathbf{w}) \sim K_1 \left(r^F - r\right)^{\left\{\sum_i 1/\xi_i\right\}-1} \left\{\prod_{k=1}^{d} w_k^{1/\xi_k - 1}\right\} \times \left\{-V_{\mathbf{x}}\left(\left[\frac{(-\xi_1)}{\sigma_1}\left(r^F - r\right) w_1\right]^{1/\xi_1}, \ldots, \left[\frac{(-\xi_d)}{\sigma_d}\left(r^F - r\right) w_d\right]^{1/\xi_d}\right)\right\} \tag{4.4.15}$$

as $r \to r^F$ and for constant $K_1 > 0$ and where $w_d = 1 - \sum_{j=1}^{d-1} w_j$. We now conduct the transformation $R_d \to V$, for V defined in (4.2.7), and consider $v < r_F - t$. The joint density of $(V, \mathbf{W})$ is

$$\begin{aligned} f_{V,\mathbf{W}}(v, \mathbf{w}) &\sim K_2 v^{\left\{\sum_i 1/\xi_i\right\}-1} \left\{\prod_{i=1}^{d} w_i^{1/\xi_i - 1}\right\} \\ &\quad \times \left\{-V_{\mathbf{x}}\left(\left[\frac{-\xi_1}{\sigma_1} r^F v w_1\right]^{1/\xi_1}, \ldots, \left[\frac{-\xi_d}{\sigma_d} r^F v w_d\right]^{1/\xi_d}\right)\right\} \\ &\sim K_3 v^{\left\{\sum_i 1/\xi_i\right\}-1} \left\{\prod_{i=1}^{d} w_i^{1/\xi_i - 1}\right\} (v w_d)^{-(d+1)} \left\{-V_{\mathbf{x}}\left(g_1(v, \mathbf{w}), \ldots, g_d(v, \mathbf{w})\right)\right\}, \end{aligned} \tag{4.4.16}$$

as $v \downarrow 0$ and $\mathbf{w} \in [0, 1]^{d-1}$ and for constants $K_2, K_3 > 0$, where

$$g_i(v, \mathbf{w}) = v^{1/\xi_i - 1/\xi_d} \left[\frac{-\xi_i r^F}{\sigma_i}\right]^{1/\xi_i} \left[\frac{-\xi_d r^F}{\sigma_d}\right]^{-1/\xi_d} \frac{w_i^{1/\xi_i}}{w_d^{1/\xi_d}},$$

for $i = 1, \ldots, d$. The last line of (4.4.16) follows as $V_{\mathbf{x}}$ is a homogeneous function of order $-(d+1)$.

We now set $\xi_d = \max_{i=1,\ldots,d} \xi_i$ and assume that $\xi_d > \xi_j$ for all $j = 1, \ldots, d-1$. To understand the behaviour of (4.4.16) and g_i for each $i = 1, \ldots, d$ and as $v \downarrow 0$, we conduct the second half of the transformation given by (4.2.7), namely $\mathbf{W} \to \mathbf{Z}$, where $W_j = V^{Z_j}$ for $Z_j \in (0, \infty)$ and $j = 1, \ldots, d-1$; note that w_d becomes $1 - \sum_{j=1}^{d-1} v^{z_j}$.

The joint density of $(V, \mathbf{Z})$ is

$$\begin{aligned} f_{V,Z}(v,\mathbf{z}) &\sim K_3 \left[-\log(v)\right]^{d-1} v^{\left\{\sum_i 1/\xi_i\right\}-1} \left\{\prod_{j=1}^{d-1} v^{z_j/\xi_j}\right\} \left\{1-\sum_{j=1}^{d-1} v^{z_j}\right\}^{1/\xi_d-1} \\ &\times \left(v\left\{1-\sum_{j=1}^{d-1} v^{z_j}\right\}\right)^{-(d+1)/\xi_d} \left\{-V_{\mathbf{x}}\left(g_1(v,\mathbf{z}),\ldots,g_d(v,\mathbf{z})\right)\right\} \\ &\sim K_3[-\log(v)]^{d-1} v^{\left\{\sum_i 1/\xi_i\right\}-1} \left\{\prod_{j=1}^{d-1} v^{z_j/\xi_j}\right\} v^{-(d+1)/\xi_d} \\ &\times \left\{-V_{\mathbf{x}}\left(g_1(v,\mathbf{z}),\ldots,g_d(v,\mathbf{z})\right)\right\}, \end{aligned} \tag{4.4.17}$$

for $\mathbf{z} = (z_1, \ldots, z_{d-1}) \in (0,\infty)^{d-1}$ and as $v \downarrow 0$ and where

$$\begin{aligned} g_i(v,\mathbf{z}) &= \left[\frac{-\xi_i r^F}{\sigma_i}\right]^{1/\xi_i} \left[\frac{-\xi_d r^F}{\sigma_d}\right]^{-1/\xi_d} \frac{v^{(1+z_i)/\xi_i - 1/\xi_d}}{\left(1-\sum_{j=1}^{d-1} v^{z_j}\right)^{1/\xi_d}} \\ &\sim \left[\frac{-\xi_i r^F}{\sigma_i}\right]^{1/\xi_i} \left[\frac{-\xi_d r^F}{\sigma_d}\right]^{-1/\xi_d} v^{(1+z_i)/\xi_i - 1/\xi_d}, \end{aligned} \tag{4.4.18}$$

for $i = 1, \ldots, d$. We note that the last lines of (4.4.17) and (4.4.18) follow as $1 - \sum_{j=1}^{d-1} v^{z_j} \to 1$ as $v \downarrow 0$ and that $g_d(v,\mathbf{z}) = 1$. Note that when we consider $v \downarrow 0$, the asymptotic behaviour of each g_i depends on z_i only, rather than the full vector $\mathbf{z}$; from here we rewrite these functions as $g_i(v, z_i)$. Now, consider

$$\begin{aligned} f_V(v) &\sim K_3[-\log(v)]^{d-1} v^{\left\{\sum_i 1/\xi_i\right\}-1} v^{-(d+1)/\xi_d} \\ &\times \int_0^\infty \cdots \int_0^\infty \left\{\prod_{j=1}^{d-1} v^{z_j/\xi_j}\right\} \left\{-V_{\mathbf{x}}\left(g_1(v,z_1),\ldots,g_{d-1}(v,z_{d-1}),1\right)\right\} \mathrm{d}z_1 \ldots z_{d-1}, \end{aligned} \tag{4.4.19}$$

as $v \downarrow 0$. To evaluate the integral in (4.4.19), we consider the integral as an iterative procedure; we begin by integrating $f_{V,\mathbf{Z}}$ with respect to z_1 and then determine the first-order behaviour of the expression as $v \downarrow 0$. We then integrate this first-order

expression with respect to z_2. Continuing in this fashion, it is possible to derive the first-order behaviour of $\int_{\mathbf{z}} f_{V,\mathbf{Z}}(z,\mathbf{z})\mathrm{d}\mathbf{z}$ as $v \downarrow 0$.

We first consider the integral

$$\begin{aligned}
&\int_0^\infty \left\{\prod_{j=1}^{d-1} v^{z_j/\xi_j}\right\} \left\{-V_{\mathbf{x}}\left(g_1(v,z_1),\ldots,g_{d-1}(v,z_{d-1}),1\right)\right\}\mathrm{d}z_1 \\
&= K_1^* \frac{v^{1/\xi_d - 1/\xi_1}}{[-\log(v)]} \left\{\prod_{j=2}^{d-1} v^{z_j/\xi_j}\right\} \Bigg\{ -V_{\mathbf{x}_{-1}}\left(\infty, g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right) \\
&+ V_{\mathbf{x}_{-1}}\left(g_1(v,0), g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right) \Bigg\}, \qquad (4.4.20)
\end{aligned}$$

as $v \downarrow 0$ and for constant $K_1^* > 0$, and where $V_{\mathbf{x}_{-1}}$ denotes the $(d-1)$-th order partial derivative of V with respect to all components except the first. To evaluate the first order behaviour of (4.4.20) as $v \downarrow 0$, we note that the two terms

$$V_{\mathbf{x}_{-1}}\left(\infty, g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right), \quad \text{and } V_{\mathbf{x}_{-1}}\left(g_1(v,0), g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right),$$

differ in only their first component, and so we consider their behaviour as $v \downarrow 0$ with the other $(d-1)$ components treated as fixed. From the assumptions made in Section 4.2.2, we have that $0 < -V_{\mathbf{x}_{-1}}\left(\infty, g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right) < \infty$ and we can show that

$$V_{\mathbf{x}_{-1}}\left(g_1(v,0), g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right) \to 0, \qquad (4.4.21)$$

as $v \downarrow 0$; this follows as $g_1(v,0) \to 0$ as $v \downarrow 0$ as $\xi_d > \xi_j$ for all $j = 1,\ldots,d-1$ and so $1/\xi_1 - 1/\xi_d > 0$. It then follows from Section 4.2.2 that (4.4.21) holds. Hence,

$$\begin{aligned}
&\int_0^\infty \left\{\prod_{j=1}^{d-1} v^{z_j/\xi_j}\right\} \left\{-V_{\mathbf{x}}\left(g_1(v,z_1),\ldots,g_{d-1}(v,z_{d-1}),1\right)\right\}\mathrm{d}z_1 \\
&\sim K_1^* \frac{v^{1/\xi_d - 1/\xi_1}}{[-\log(v)]} \left\{\prod_{j=2}^{d-1} v^{z_j/\xi_j}\right\} \left\{-V_{\mathbf{x}_{-1}}\left(\infty, g_2(v,z_2),\ldots,g_{d-1}(v,z_{d-1}),1\right)\right\}, \qquad (4.4.22)
\end{aligned}$$

as $v \downarrow 0$. We now let $V_{\mathbf{x}_{-(1,2)}}$ denote the $(d-2)$-th order partial derivative of V with respect to all components except the first and second From (4.4.22), we have that

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \left\{ \prod_{j=1}^{d-1} v^{z_j/\xi_j} \right\} \{-V_{\mathbf{x}} \left(g_1(v,z_1), \ldots, g_{d-1}(v,z_{d-1}), 1\right)\} \mathrm{d}z_1 z_2 \\
&\sim K_1^* \int_0^\infty \frac{v^{1/\xi_d - 1/\xi_1}}{[-\log(v)]} \left\{ \prod_{j=2}^{d-1} v^{z_j/\xi_j} \right\} \{-V_{\mathbf{x}_{-1}} \left(\infty, g_2(v,z_2), \ldots, g_{d-1}(v,z_{d-1}), 1\right)\} \mathrm{d}z_2 \\
&\sim K_2^* \frac{v^{2/\xi_d - 1/\xi_1 - 1/\xi_2}}{[-\log(v)]^2} \left\{ \prod_{j=3}^{d-1} v^{z_j/\xi_j} \right\} \Bigg\{ - V_{\mathbf{x}_{-(1,2)}} \left(\infty, \infty, g_3(v,z_3), \ldots, g_{d-1}(v,z_{d-1}), 1\right) \\
&+ V_{\mathbf{x}_{-(1,2)}} \left(\infty, g_2(v,0), g_3(v,z_3), \ldots, g_{d-1}(v,z_{d-1}), 1\right) \Bigg\} \\
&\sim K_2^* \frac{v^{2/\xi_d - 1/\xi_1 - 1/\xi_2}}{[-\log(v)]^2} \left\{ \prod_{j=3}^{d-1} v^{z_j/\xi_j} \right\} \{-V_{\mathbf{x}_{-(1,2)}} \left(\infty, \infty, g_3(v,z_3), \ldots, g_{d-1}(v,z_{d-1}), 1\right)\}
\end{aligned}$$

as $v \downarrow 0$ and for constant $K_2^* > 0$; the last line follows for the same reasoning as provided above for (4.4.22) and with assumptions made in Section 4.2.2. That is, we have that $-V_{\mathbf{x}_{-(1,2)}} \left(\infty, \infty, g_3(v,z_3), \ldots, g_{d-1}(v,z_{d-1}), 1\right)$ is a finite, positive constant and

$$V_{\mathbf{x}_{-1}} \left(\infty, g_2(v,0), g_3(v,z_3), \ldots, g_{d-1}(v,z_{d-1}), 1\right) \to 0,$$

as $v \downarrow 0$. Continuing in this manner, it follows that

$$\int_{\mathbf{Z}} \left\{ \prod_{j=1}^{d-1} v^{z_j/\xi_j} \right\} \{-V_{\mathbf{x}} \left(g_1(v,\mathbf{z}), \ldots, 1\right)\} \mathrm{d}\mathbf{z} \sim K^* \frac{v^{(d-1)/\xi_d - \sum_{j=1}^{d-1} 1/\xi_j}}{[-\log(v)]^{d-1}} \{-V_{x_d} \left(\infty, \ldots, \infty, 1\right)\}, \tag{4.4.23}$$

as $v \downarrow 0$ and for constant $K^* > 0$ and where $V_{x_d} \left(\infty, \ldots, \infty, 1\right) = -1$ denotes the first-order partial derivative of V with respect to the d-th component. Combining (4.4.19) and (4.4.23), the marginal density of V is $f_V(v) \sim K_2^* v^{-1/\xi_d - 1}$ as $v \downarrow 0$ and for constant $K_2^* > 0$. Transforming V back to R_d and deriving the survival function,

it can be shown that

$$\Pr\{R_d \geq r\} \sim K^*_{R_d} \left(1 - \frac{r}{r^F}\right)^{-1/\xi_d}$$

as $r \to r^F$ and for constant $K^*_{R_d} > 0$. Note that the choice $\xi_d = \max_{i=1,\ldots,d} \xi_i$ was arbitrary, and so we can simply replace ξ_d with $\max_{i=1,\ldots,d} \xi_i$ to achieve the desired result.

4.4.4 Inverted extreme value copula

Let $R_2 = X_1 + X_2$, where $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $\xi_i \neq 0, sgn(\xi_1) = sgn(\xi_2)$ and $i = 1, 2$, and with dependence in (X_1, X_2) induced by the inverted extreme value copula described in Section 4.1. By combining (4.1.1) and (4.1.4) and differentiating with respect to all arguments, the joint density of (X_1, X_2) is

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) = {} & \frac{\xi_1^2 \xi_2^2}{\sigma_1 \sigma_2} \left\{\log(\tilde{x}_1) \log(\tilde{x}_2)\right\}^{-2} e^{-V\left(\frac{\xi_1}{\log(\tilde{x}_1)}, \frac{\xi_2}{\log(\tilde{x}_2)}\right)} \left\{\tilde{x}_1 \tilde{x}_2\right\}^{-1} \\ & \times \left\{ V_{x_1}\left(\frac{\xi_1}{\log(\tilde{x}_1)}, \frac{\xi_2}{\log(\tilde{x}_2)}\right) V_{x_2}\left(\frac{\xi_1}{\log(\tilde{x}_1)}, \frac{\xi_2}{\log(\tilde{x}_2)}\right) - V_{\mathbf{x}}\left(\frac{\xi_1}{\log(\tilde{x}_1)}, \frac{\xi_2}{\log(\tilde{x}_2)}\right) \right\}, \end{aligned} \tag{4.4.24}$$

where $\tilde{x}_i = 1 + \xi_i x_i / \sigma_i$ for $i = 1, 2$ and V is defined in (4.1.3), with V_{x_1} and V_{x_2} denoting the first-order partial derivatives of V with respect to the first and second components, respectively, and $V_{\mathbf{x}}$ denoting the second-order partial derivative of V with respect to both components. Throughout we make the assumption that $1 < V(1, 1) \leq 2$; we omit the case $V(1, 1) = 1$, corresponding to perfect dependence, as this is covered by the proofs in Section 4.4.1.

Strictly positive marginal shapes

We now set $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$. Furthermore, we constrain this proof to the case $d = 2$ only, but assume that the relationship holds for R_d. To justify this, we

illustrate that the form for $\Pr\{R_2 > s\}$ as $s \to \infty$ for this case is exactly that of the case where $\mathbf{X}$ exhibits independence, and so the same relationship will hold for R_d.

We rewrite (4.4.24) with $\xi_1 = \xi_2 = \xi$ as

$$f_{X_1,X_2}(x_1,x_2) = \frac{(\tilde{x}_1\tilde{x}_2)^{-1}}{\sigma_1\sigma_2} e^{-\frac{1}{\xi}V\left(\frac{1}{\log(\tilde{x}_1)},\frac{1}{\log(\tilde{x}_2)}\right)}$$
$$\times \left\{ V_{x_1}\left(1, \frac{\log(\tilde{x}_1)}{\log(\tilde{x}_2)}\right) V_{x_2}\left(\frac{\log(\tilde{x}_2)}{\log(\tilde{x}_1)}, 1\right) - \xi\log(\tilde{x}_1)\log(\tilde{x}_2) V_{\mathbf{x}}\left(\log(\tilde{x}_2), \log(\tilde{x}_2)\right) \right\},$$

which follows as V_{x_1} and V_{x_2} are homogeneous functions of order -2. We then use the transformation $(X_1, X_2) \to (R_2, W)$ given by (4.2.1); the joint density of (R_2, W) is

$$\begin{aligned} f_{R_2,W}(r,w) = \frac{r}{\sigma_1\sigma_2} e^{-\frac{1}{\xi}V\left(\frac{1}{\log\left(1+\xi\frac{rw}{\sigma_1}\right)},\frac{1}{\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)}\right)} \left\{\left(1+\xi\frac{rw}{\sigma_1}\right)\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)\right\}^{-1} \\ \times \left\{ V_{x_1}\left(1, \frac{\log\left(1+\xi\frac{rw}{\sigma_1}\right)}{\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)}\right) V_{x_2}\left(\frac{\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)}{\log\left(1+\xi\frac{rw}{\sigma_1}\right)}, 1\right) \right. \\ - \xi\log\left(1+\xi\frac{rw}{\sigma_1}\right)\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right) \\ \left. \times V_{\mathbf{x}}\left(\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right), \log\left(1+\xi\frac{rw}{\sigma_1}\right)\right) \right\}, \end{aligned} \tag{4.4.25}$$

for $w \in [0,1]$ and $r \in [0,\infty)$. We now note that

$$\begin{aligned} f_{R_2}(r) &= \int_0^1 f_{R_2,W}(r,w)\mathrm{d}w \\ &= \int_{1/2}^1 f_{R_2,W}(r,w)\mathrm{d}w + \int_0^{1/2} f_{R_2,W}(r,w)\mathrm{d}w, \end{aligned} \tag{4.4.26}$$

and so we consider only the second integral; the first can be derived by symmetric arguments. To derive the second integral in (4.4.26), we apply the transformation $(R_2, W) \to (R_2, Z)$ for $Z \in (\log(r)/\log(2), \infty)$ given by (4.2.6). The joint density of

(R_2, Z) is

$$
\begin{aligned}
f_{R_2,Z}(r,z) &= \frac{r^{1-z}\log(r)}{\sigma_1\sigma_2} g_r(z) \left\{ \left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right)\right\}^{-1} \\
&\times \left\{ V_{x_1}\left(1, \frac{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}{\log\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right)}\right) V_{x_2}\left(\frac{\log\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}, 1\right)\right. \\
&- \xi\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\log\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right) \\
&\left.\times V_{\mathbf{x}}\left(\log\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right), \log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\right)\right\} \\
&\sim \frac{r^{-z}\log(r)}{\sigma_1\xi} g_r(z)\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)^{-1} \\
&\times \left\{ V_{x_1}\left(1, \frac{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}{\log(r)}\right) V_{x_2}\left(\frac{\log(r)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}, 1\right)\right. \\
&\left.- \xi\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\log(r)\, V_{\mathbf{x}}\left(\log(r), \log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\right)\right\} \qquad (4.4.27)
\end{aligned}
$$

as $r \to \infty$ and where

$$
\begin{aligned}
g_r(z) &= \exp\left\{ -\frac{1}{\xi}\log\left(1+\frac{\xi r}{\sigma_2}\right) V\left(\frac{\log\left(1+\frac{\xi r}{\sigma_2}\right)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_2}\right)}, \frac{\log\left(1+\frac{\xi r}{\sigma_2}\right)}{\log\left(1+\xi\frac{r(1-r^{-z})}{\sigma_2}\right)}\right)\right\} \\
&\sim \exp\left\{ -\frac{1}{\xi}\log\left(1+\frac{\xi r}{\sigma_2}\right) V\left(\frac{\log(r)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_2}\right)}, 1\right)\right\}.
\end{aligned}
$$

To integrate (4.4.27) with respect to Z, we consider the derivative of

$$
G_r^{(1)}(z) = g_r(z) V_{x_2}\left(\frac{\log(r)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}, 1\right)
$$

with respect to z, which is

$$
\frac{\mathrm{d}}{\mathrm{d}z} G_r^{(1)}(z) = -\frac{r^{1-z}\log(r)}{\sigma_1} g_r(z)\left(1+\xi\frac{r^{1-z}}{\sigma_2}\right)^{-1}
$$

$$\times \left\{ V_{x_1}\left(1, \frac{\log\left(1+\frac{\xi r^{1-z}}{\sigma_1}\right)}{\log(r)}\right) V_{x_2}\left(\frac{\log(r)}{\log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)}, 1\right) \right.$$
$$\left. - \xi \log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\log(r)\, V_{\mathbf{x}}\left(\log(r), \log\left(1+\xi\frac{r^{1-z}}{\sigma_1}\right)\right) \right\}$$
$$\sim -r\xi f_{R_2,Z}(r,z),$$

as $r \to \infty$. Hence $\int_{\log(2)/\log(r)}^{\infty} f_{R_2,Z}(r,z)\mathrm{d}z = \int_0^{1/2} f_{R_2,W}(r,w)\mathrm{d}w$ is

$$\sim -\xi^{-1}r^{-1}\left[G_r^{(1)}(\infty) - G_r^{(1)}(\log(2)/\log(r))\right]$$
$$\sim -\xi^{-1}r^{-1}\left[\exp\left\{-\frac{1}{\xi}\log\left(1+\frac{\xi r}{\sigma_2}\right)V(\infty,1)\right\} V_{x_2}(\infty,1) \right.$$
$$\left. - \exp\left\{-\frac{1}{\xi}\log\left(1+\frac{\xi r}{\sigma_2}\right)V\left(\frac{\log\left(1+\frac{\xi r}{\sigma_2}\right)}{\log\left(1+\xi\frac{r}{2\sigma_1}\right)}, 1\right)\right\} V_{x_2}\left(\frac{\log\left(\xi\frac{r}{\sigma_2}\right)}{\log\left(1+\xi\frac{r}{2\sigma_1}\right)}, 1\right)\right]$$
$$\sim \xi^{-1}r^{-1}\left[\left(1+\xi\frac{r}{\sigma_2}\right)^{-1/\xi} + V_{x_2}(1,1)\left(1+\xi\frac{r}{\sigma_2}\right)^{-V(1,1)/\xi}\right]$$
$$\sim \xi^{-1}r^{-1}\left(\frac{\xi r}{\sigma_2}\right)^{-1/\xi} \tag{4.4.28}$$

as $r \to \infty$ and where the penultimate line follows as $V(\infty,1) = V(1,\infty) = 1$ and $V_{x_2}(\infty,1) = V_{x_1}(1,\infty) = -1$; the last line follows as $V(1,1) > 1$. A symmetric argument can be used to show that

$$\int_{1/2}^{1} f_{R_2,W}(r,w)\mathrm{d}w \sim \xi^{-1}r^{-1}\left(\frac{\xi r}{\sigma_1}\right)^{-1/\xi},$$

as $r \to \infty$, and hence

$$f_{R_2}(r) \sim \xi^{-1}r^{-1}\left(\frac{\xi r}{\sigma_2}\right)^{-1/\xi} + \xi^{-1}r^{-1}\left(\frac{\xi r}{\sigma_1}\right)^{-1/\xi} = \xi^{-1}r^{-1}\left[\sigma_1^{1/\xi} + \sigma_2^{1/\xi}\right](\xi r)^{-1/\xi},$$

as $r \to \infty$. We then have

$$\Pr\{R_2 \geq s\} = \int_s^\infty f_{R_2}(r)\mathrm{d}r \sim \left\{ \frac{\xi}{\left[\sigma_1^{1/\xi} + \sigma_2^{1/\xi}\right]^{-\xi}} \right\}^{-1/\xi} s^{-1/\xi},$$

as $s \to \infty$, and so we have the required result; this is equivalent to the result derived for the independence case (see Section 4.4.2), which justifies our assumption that similar structures are found in R_d as R_2, and so we claim that $\Pr\{R_d \geq s\}$ is of the form

$$\Pr\{R_d \geq s\} \sim \left\{ \frac{\xi}{\left[\sum_{i=1}^d \sigma_i^{1/\xi}\right]^{-\xi}} \right\}^{-1/\xi} s^{-1/\xi},$$

as $s \to \infty$.

Strictly negative marginal shapes

We now let $\xi_i < 0$ for $i = 1, 2$. From (4.4.24), we apply the transformations $(X_1, X_2) \to (U_1, U_2)$ and $(U_1, U_2) \to (R_2, W)$ given by (4.2.2) and (4.2.3), respectively, and consider (R_2, W) for $R_2 > t$, with t defined in (4.2.5). The joint density of (R_2, W) for $w \in [0, 1]$ is

$$\begin{aligned} f_{R_2,W}(r,w) = (\sigma_1\sigma_2)^{-1} \left(r^F - r\right)^{-1} & \left\{ \log\left(-\xi_1 \left(r^F - r\right) \frac{w}{\sigma_1}\right) \log\left(-\xi_2 \left(r^F - r\right) \frac{(1-w)}{\sigma_2}\right) \right\}^{-2} \\ & \times \exp\left\{ -V\left(\frac{\xi_1}{\log\left(-\xi_1 \left(r^F - r\right) \frac{w}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 \left(r^F - r\right) \frac{(1-w)}{\sigma_2}\right)} \right) \right\} \\ & \times \{w(1-w)\}^{-1} \left\{ V_{x_1}\left(\frac{\xi_1}{\log\left(-\xi_1 \left(r^F - r\right) \frac{w}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 \left(r^F - r\right) \frac{(1-w)}{\sigma_2}\right)} \right) \right. \\ & \times V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 \left(r^F - r\right) \frac{w}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 \left(r^F - r\right) \frac{(1-w)}{\sigma_2}\right)} \right) \\ & \left. - V_{\mathbf{x}}\left(\frac{\xi_1}{\log\left(-\xi_1 \left(r^F - r\right) \frac{w}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 \left(r^F - r\right) \frac{(1-w)}{\sigma_2}\right)} \right) \right\}, \end{aligned}$$

where $r^F = -\sigma_1/\xi_1 - \sigma_2/\xi_2$ is the upper-endpoint of R_2. We now apply another transformation $(R_2, W) \to (V, Z)$ for $V \in [1 - t/r^F, 1]$ and $Z \in (0, \infty)$, given by (4.2.7) and consider $v \downarrow 0$. The joint density of (V, Z) is

$$
\begin{aligned}
f_{V,Z}(v,z) &= -(r^F)^{-2}(\sigma_1\sigma_2)^{-1}v^{-1}\log(v)\left\{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)\log\left(-\xi_2 v r^F \frac{(1-v^z)}{\sigma_2}\right)\right\}^{-2} \\
&\quad \times \exp\left\{-V\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F \frac{(1-v^z)}{\sigma_2}\right)}\right)\right\}(1-v^z)^{-1} \\
&\quad \times \left\{V_{x_1}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F \frac{(1-v^z)}{\sigma_2}\right)}\right)\right. \\
&\quad \times V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F \frac{(1-v^z)}{\sigma_2}\right)}\right) \\
&\quad \left. - V_{\mathbf{x}}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F \frac{(1-v^z)}{\sigma_2}\right)}\right)\right\} \\
&\sim -(r^F)^{-2}(\sigma_1\sigma_2)^{-1}v^{-1}\log(v)\left\{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)\log\left(-\xi_2 v r^F/\sigma_2\right)\right\}^{-2} \\
&\quad \times \exp\left\{-V\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\} \\
&\quad \times \left\{V_{x_1}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right. \\
&\quad \times V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right) \\
&\quad \left. - V_{\mathbf{x}}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\},
\end{aligned}
$$

where the last line follows as $1 - v^z \to 1$ for any fixed $z \neq 0$ and as $v \downarrow 0$. To find the marginal density of V, we consider the derivative of

$$G_v(z) = V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right) \times \exp\left\{-V\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\},$$

with respect to z, which is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}G_v(z) = {} & \frac{\xi_1 \log(v)}{\left[\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)\right]^2} \exp\left\{-V\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\} \\ & \times \left\{ - V_{\mathbf{x}}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right. \\ & + V_{x_1}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right) \\ & \left. \times V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\} \\ & \sim K_1 v \left[\log\left(-\xi_2 v r^F/\sigma_2\right)\right]^2 f_{V,Z}(v,z),\end{aligned}$$

as $v \downarrow 0$ and for constant $K_1 = -\sigma_1\sigma_2(r^F)^2/\xi_1 > 0$. Hence,

$$\begin{aligned} f_V(v) = \int_0^\infty f_{V,Z}(v,z)\mathrm{d}z & \sim \left[\frac{1}{K_1} v^{-1}\left[\log\left(-\xi_2 v r^F/\sigma_2\right)\right]^{-2} G_v(z)\right]_0^\infty \\ & \sim \frac{1}{K_1}\left[v^{-1}\left[\log\left(-\xi_2 v r^F/\sigma_2\right)\right]^{-2} V_{x_2}\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right. \\ & \left. \times \exp\left\{-V\left(\frac{\xi_1}{\log\left(-\xi_1 v r^F \frac{v^z}{\sigma_1}\right)}, \frac{\xi_2}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)\right\}\right]_0^\infty \end{aligned}$$

$$\sim \frac{1}{K_1}\left[v^{-1}V_{x_2}\left(\frac{-\xi_1\log\left(-\xi_2 v r^F/\sigma_2\right)}{\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)},-\xi_2\right)e^{\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)V\left(-\xi_1,\frac{-\xi_2\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)}\right]_0^\infty$$

$$\sim -\frac{1}{K_1}v^{-1}V_{x_2}\left(\frac{-\xi_1\log\left(-\xi_2 v r^F/\sigma_2\right)}{\log\left(-\xi_1 v r^F/\sigma_1\right)},-\xi_2\right)e^{\log\left(-\xi_1 v r^F/\sigma_1\right)V\left(-\xi_1,\frac{-\xi_2\log\left(-\xi_1 v r^F/\sigma_1\right)}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)}$$

$$\sim -\frac{1}{K_1}v^{-1}V_{x_2}\left(-\xi_1,-\xi_2\right)e^{\log(v)V(-\xi_1,-\xi_2)} \sim K_2 v^{-1}v^{V(-\xi_1,-\xi_2)} = K_2 v^{V(-\xi_1,-\xi_2)-1},$$

as $v \downarrow 0$ and for constant $K_2 = -(K_1)^{-1}V_{x_2}(-\xi_1,-\xi_2) > 0$, and where the third line follows as V and V_{x_2} are homogeneous functions of order -1 and -2, respectively; the fourth line follows as

$$V_{x_2}\left(\frac{-\xi_1\log\left(-\xi_2 v r^F/\sigma_2\right)}{\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)},-\xi_2\right) \to V_{x_2}(\infty,-\xi_2),$$

as $z \to \infty$ and where $V_{x_2}(\infty,-\xi_2) < 0$ is a constant, and

$$e^{\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)V\left(-\xi_1,\frac{-\xi_2\log\left(-\xi_1 v r^F\frac{v^z}{\sigma_1}\right)}{\log\left(-\xi_2 v r^F/\sigma_2\right)}\right)} \sim v^{zV(-\xi_1,\infty)} \to 0,$$

as $z \to \infty$ and for any $v < 1$. Hence, the product of the two terms equals zero as $z \to \infty$, and the result follows. Transforming back to R_2, it can be shown that

$$\Pr\{R_2 \geq s\} \sim K_3\left(1-\frac{s}{r^F}\right)^{V(-\xi_1,-\xi_2)},$$

as $s \to r^F$ and for constant $K_3 = K_2(r^F)^{2+V(-\xi_1,-\xi_2)}/V(-\xi_1,\xi_2) > 0$.

4.4.5 Standard Gaussian copula

We begin by considering the 2−dimensional random vector $\mathbf{Y} \sim N_2(\mathbf{0}_2,\Sigma)$, where $\mathbf{0}_2$ denotes a 2-vector of zeroes and Σ is a 2×2 positive definite matrix, where $\Sigma_{ii} = 1$ and $\Sigma_{12} = \Sigma_{21} = \rho \in (0,1)$. To perform the transformation $\mathbf{Y} \to \mathbf{X}$ where

$X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $i = 1, 2$, we let

$$\bar{\Phi}(Y_i) = \left(1 + \frac{\xi_i}{\sigma_i} X_i\right)^{-1/\xi_i},$$

for $i = 1, 2$. Here $\bar{\Phi}(\cdot)$ denotes the survival function of the univariate standard Gaussian distribution. By Mill's ratio (Grimmett, 2020), we have that $\bar{\Phi}(x) \sim \frac{\phi(x)}{x}$ as $x \to \infty$ and where ϕ denotes the density of the univariate standard Gaussian distribution. Thus, we can approximate the transformation $Y_i \to X_i$ by finding a solution to

$$\frac{1}{y}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \sim \left\{1 + \frac{\xi}{\sigma} x\right\}^{-1/\xi_i}, \tag{4.4.29}$$

which holds as $x \to x^F$, where x^F denotes the upper endpoint of X. To solve (4.4.29), we begin with an initial solution y_0 that solves

$$\left(1 + \frac{\xi}{\sigma} x\right)^{-1/\xi} = e^{-\frac{1}{2}y^2},$$

which gives $y_0 = \sqrt{\frac{2}{\xi}\log\left\{1 + \frac{\xi}{\sigma}x\right\}}$, where $\xi^{-1}\log(1 + \xi x\sigma) \geq 0$. Now, let $y_1 = x_0 + \epsilon$ where $\epsilon = o\left(\sqrt{\log\left\{1 + \frac{\xi}{\sigma}x\right\}}\right)$. Substituting this into (4.4.29) gives

$$\begin{aligned}
\left(1 + \frac{\xi}{\sigma} x\right)^{-1/\xi} &\sim \frac{1}{\sqrt{\frac{2}{\xi}\log\left\{1 + \frac{\xi}{\sigma}x\right\}} + \epsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}} e^{-\epsilon\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}} e^{-\frac{1}{2}\epsilon^2} \\
&\sim \frac{1}{\sqrt{\frac{2}{\xi}\log\left\{1 + \frac{\xi}{\sigma}x\right\}} + \epsilon} \frac{1}{\sqrt{2\pi}} \left\{1 + \frac{\xi}{\sigma}x\right\}^{-1/\xi - 1} e^{-\epsilon\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}} e^{-\frac{1}{2}\epsilon^2},
\end{aligned}$$

as $x \to x^F$ and thus

$$\begin{aligned}
1 &\sim \frac{1}{\sqrt{\frac{2}{\xi}\log\left\{1 + \frac{\xi}{\sigma}x\right\}} + \epsilon} \frac{1}{\sqrt{2\pi}} e^{-\epsilon\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}} e^{-\frac{1}{2}\epsilon^2} \\
&\sim \frac{1}{\sqrt{\frac{2}{\xi}\log\left\{1 + \frac{\xi}{\sigma}x\right\}}} \frac{1}{1 + \frac{\epsilon}{\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}}} \frac{1}{\sqrt{2\pi}} e^{-\epsilon\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}} e^{-\frac{1}{2}\epsilon^2}
\end{aligned}$$

$$\sim \frac{1}{\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}}\frac{1}{\sqrt{2\pi}}e^{-\epsilon\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}}e^{-\frac{1}{2}\epsilon^2}, \tag{4.4.30}$$

where the last line follows as

$$\frac{1}{1+\frac{\epsilon}{\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}}} \to 1,$$

as $x \to x^F$. Taking logs of both sides of (4.4.30) and rearranging gives

$$\begin{aligned}\epsilon &\sim -\frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}\right)}{2\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x_P\right\}}} - \frac{\epsilon^2}{2\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}} \\ &\sim -\frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}\right)}{2\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}},\end{aligned}$$

as $x \to x^F$. Thus, an approximate solution to (4.4.29) is

$$y = \sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}} - \frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}\right)}{2\sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma}x\right\}}}\left[1+o(1)\right]$$

as $x \to x^F$, and which can be applied to all components of $\mathbf{Y}$. To calculate the determinant of the Jacobian of this marginal transformation, we note that

$$\begin{aligned}\frac{\partial y_i}{\partial x_i} &\sim \sqrt{\frac{\xi_i}{2\sigma_i^2}\left(\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}\right)^{-1}}\left(1+\frac{\xi_i}{\sigma_i}x_i\right)^{-1} \\ &\quad - \frac{\xi_i}{2\sigma_i}\sqrt{\frac{\xi_i}{2}\left(\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}\right)^{-3}}\left(1+\frac{\xi_i}{\sigma_i}x_i\right)^{-1}\left\{1-\frac{1}{2}\log\left(\frac{4\pi}{\xi_i}\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}\right)\right\} \\ &\sim \sqrt{\frac{\xi_i}{2\sigma_i^2}\left(\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}\right)^{-1}}\left(1+\frac{\xi_i}{\sigma_i}x_i\right)^{-1},\end{aligned}$$

as $x_i \to x_i^F$ and for $i = 1, 2$. Note that if $\xi_i < 0$, then $0 \leq x_i \leq -\sigma_i/\xi_i$, and so $\xi_i/\log(1+\xi_i x_i \sigma_i) \geq 0$. We now make the assumption that as $R \to r^F$ that

all components of $\mathbf{X}$ tends to their respective upper-endpoints with associated rates determined by the following; we assume that $\bar{F}_1(x_1)/\bar{F}_2(x_2) \sim c$ as $\bar{F}_1(x_1) \to 0$ for constant $c > 0$, and that $R \to r^F \Rightarrow \bar{F}_1(x_1) \to 0$. The joint density of $\mathbf{X}$ is then

$$f_{\mathbf{X}}(\mathbf{x}) \sim (2\pi)^{-1}|\Sigma|^{-1/2}2^{-1}\exp\left\{-\frac{1}{2}(A-B)^T\Sigma^{-1}(A-B)\right\} \times \prod_{i=1}^{2}\sigma_i^{-1}\sqrt{\frac{\xi_i}{\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}}}\left(1+\frac{\xi_i}{\sigma_i}x_i\right)^{-1}, \tag{4.4.31}$$

as each component of $\mathbf{x}$ tends to its respective upper-endpoint, and where

$$A = \begin{pmatrix} \sqrt{\frac{2}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}} \\ \\ \sqrt{\frac{2}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_2\right\}} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}\right)}{2\sqrt{\frac{2}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}}} \\ \\ \frac{\log\left(\frac{4\pi}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_2\right\}\right)}{2\sqrt{\frac{2}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_2\right\}}} \end{pmatrix}.$$

We now consider the two cases: $\xi_1 = \xi_2 = \xi > 0$ and $\max\{\xi_1, \xi_2\} < 0$.

Strictly positive marginal shapes

We now set $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$. Furthermore, we constrain this proof to the case where $d = 2$ only. To justify this, we illustrate that the form for $\Pr\{R_2 > s\}$ as $s \to \infty$ for this case is exactly that of the case where $\mathbf{X}$ exhibits independence, and so the same relationship will hold for R_d.

We first perform the two-step transformation $(X_1, X_2) \to (R_2, W) \to (R_2, Z)$ given by (4.2.1) and (4.2.6); from these transformations, and from (4.4.31), the joint density of (R_2, Z) for $z \in (0, \infty)$ is

$$f_{R_2,Z}(r,z) \sim r^{1-z}\log(r)\sigma_1^{-1}\sigma_2^{-1}(2\pi)^{-1}|\Sigma|^{-1/2}2^{-1}\exp\{G_r(z)\} \times H_r(z)\sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}}$$

$$\times \sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}}}\left(1+\frac{\xi}{\sigma_1}r^{1-z}\right)^{-1}\left(1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right)^{-1}, \tag{4.4.32}$$

as $r \to \infty$ and where $G_r(z) = -\frac{1}{2}A_{r,z}^T\Sigma^{-1}A_{r,z}$ and

$$H_r(z) = \exp\left\{-\frac{1}{2}\left[-B_{r,z}^T\Sigma^{-1}A_{r,z} - A_{r,z}^T\Sigma^{-1}B_{r,z} + B_{r,z}^T\Sigma^{-1}B_{r,z}\right]\right\} \tag{4.4.33}$$

with

$$A_{r,z} = \sqrt{\frac{2}{\xi}}\begin{pmatrix}\sqrt{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}} \\ \\ \sqrt{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}}\end{pmatrix} \quad \text{and} \quad B_{r,z} = \frac{1}{2}\sqrt{\frac{\xi}{2}}\begin{pmatrix}\frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}\right)}{\sqrt{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}} \\ \\ \frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}\right)}{\sqrt{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}}}\end{pmatrix}.$$

We now note that the marginal density of R_2 can be written as

$$f_{R_2}(r) = \int_{\log(2)/\log(r)}^{\infty} f_{R_2,Z}(r,z)\mathrm{d}z + \int_0^{\log(2)/\log(r)} f_{R_2,Z}(r,z)\mathrm{d}z,$$

and we focus on the first integral only, which we approximate using Laplace's method Laplace (1986); the second integral follows by symmetry. To use Laplace's method, we require the second derivative of $G_r(z)$ and its minimum, which we derive now.

We first note that on $z \in (\log(2)/\log(r), \infty)$, we have

$$A_{r,z} = \sqrt{\frac{2}{\xi}}\begin{pmatrix}\sqrt{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}} \\ \\ \sqrt{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}}\end{pmatrix} = \sqrt{\frac{2}{\xi}\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}\begin{pmatrix}\sqrt{\frac{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}{\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}} \\ \\ \sqrt{\frac{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}}{\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}}\end{pmatrix}$$

$$\sim \sqrt{\frac{2}{\xi}\log(r)} \begin{pmatrix} \sqrt{\frac{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}{\log(r)}} \\ \\ 1 \end{pmatrix}$$

and

$$B_{r,z} = \frac{1}{2}\sqrt{\frac{\xi}{2}\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}^{-1} \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}\right)}{\sqrt{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}/\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}} \\ \\ \frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}\right)}{\sqrt{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z})\right\}/\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}} \end{pmatrix}$$

$$\sim \frac{1}{2}\sqrt{\frac{\xi}{2\log(r)}} \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi}\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}\right)}{\sqrt{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}/\log(r)}} \\ \\ \log\left(\frac{4\pi}{\xi}\log(r)\right) \end{pmatrix},$$

as $r \to \infty$, and which follows as $1 - r^{-z} \sim 1$ for any fixed $z \in (0, \infty)$. Now, let $\Sigma_{1,2} = \rho \in (0, 1)$ and consider

$$\begin{aligned} G_r(z) &= -\frac{1}{\xi|\Sigma|}A_{r,z}^T \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} A_{r,z} \\ &\sim -\frac{1}{\xi|\Sigma|}\log\left(\frac{\xi}{\sigma_2}r\right) \begin{pmatrix} \sqrt{\frac{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}{\log\left(\frac{\xi}{\sigma_2}r\right)}} \\ \\ 1 \end{pmatrix}^{-T} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \sqrt{\frac{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}{\log\left(\frac{\xi}{\sigma_2}r\right)}} \\ \\ 1 \end{pmatrix} \\ &\sim -\frac{1}{\xi|\Sigma|}\log\left(\frac{\xi}{\sigma_2}r\right)(K_r(z)^2 - 2\rho K_r(z) + 1), \end{aligned}$$

as $r \to \infty$ and where

$$K_r(z) = \sqrt{\frac{\log\left\{1 + \frac{\xi}{\sigma_1} r^{1-z}\right\}}{\log\left(\frac{\xi}{\sigma_2} r\right)}} > 0,$$

for all $z \in (0, \infty)$. Recall that, as we are using Laplace's method, we require the minimum of $G_r(z)$ with respect to z; trivially, this is minimised when $K_r(z^*) = \rho$ or, equivalently, where

$$\begin{aligned} z^* &= 1 - \frac{\log\left\{\frac{\sigma_1}{\xi}\left[e^{\left(\rho^2 \log\left(\frac{\xi}{\sigma_2} r\right)\right)} - 1\right]\right\}}{\log\left(\frac{\xi}{\sigma_2} r\right)} \sim 1 - \frac{\log\left\{\frac{\sigma_1}{\xi}\left[r^{\rho^2} - 1\right]\right\}}{\log(r)} \\ &\sim 1 - \frac{\rho^2 \log(r)}{\log(r)} \sim 1 - \rho^2, \end{aligned}$$

as $r \to \infty$ and so

$$G_r(z^*) \sim -\frac{1}{\xi|\Sigma|} \log\left(\frac{\xi}{\sigma_2} r\right)(1 - \rho^2) = -\frac{1}{\xi} \log\left(\frac{\xi}{\sigma_2} r\right),$$

as $|\Sigma| = 1 - \rho^2$; note that we consider $r \to \infty$, and so we have that $\log(r)/\log(r) < 1 - \rho^2$ for $\rho < 1$, hence $1 - \rho^2 \in (\log(2)/\log(r), \infty)$. The derivative of $G_r(z)$ with respect to z is

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}z} G_r(z) &\sim \frac{\mathrm{d}}{\mathrm{d}z}\left[-\frac{1}{\xi|\Sigma|}\left\{\log\left\{1 + \frac{\xi}{\sigma_1} r^{1-z}\right\} - 2\rho\sqrt{\log(r)\log\left\{1 + \frac{\xi}{\sigma_1} r^{1-z}\right\}} + \log\left(\frac{\xi}{\sigma_2} r\right)\right\}\right] \\ &\sim -\frac{1}{\xi|\Sigma|}\left\{\frac{-\xi \log\left(\frac{\xi}{\sigma_2} r\right) r^{1-z}}{\sigma_1\left(1 + \frac{\xi}{\sigma_1} r^{1-z}\right)} + \rho\sqrt{\frac{\log\left(\frac{\xi}{\sigma_2} r\right)}{\log\left\{1 + \frac{\xi}{\sigma_1} r^{1-z}\right\}}} \frac{\xi \log\left(\frac{\xi}{\sigma_2} r\right) r^{1-z}}{\sigma_1\left(1 + \frac{\xi}{\sigma_1} r^{1-z}\right)}\right\} \\ &\sim \frac{1}{\sigma_1|\Sigma|} \frac{\log\left(\frac{\xi}{\sigma_2} r\right)}{\left(r^{z-1} + \frac{\xi}{\sigma_1}\right)} \left\{1 - \rho K_r^{-1}(z)\right\}, \end{aligned}$$

as $r \to \infty$ and its second derivative, which we denote $G_r''(z)$, is

$$\begin{aligned} G_r''(z) &\sim -\frac{\left\{\log\left(\frac{\xi}{\sigma_2}r\right)\right\}^2 r^{z-1}}{\sigma_1|\Sigma|}\left(r^{z-1}+\frac{\xi}{\sigma_1}\right)^{-2}\left\{1-\rho K_r^{-1}(z)\right\} \\ &\quad + \frac{\rho\xi}{2\sigma_1^2|\Sigma|}\frac{\left\{\log\left(\frac{\xi}{\sigma_2}r\right)\right\}^2 r^{2-2z}}{\left(1+\frac{\xi}{\sigma_1}r^{1-z}\right)^2}\sqrt{\frac{\log\left(\frac{\xi}{\sigma_2}r\right)}{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}}}\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z}\right\}^{-1} \\ &\sim -\frac{\left\{\log\left(\frac{\xi}{\sigma_2}r\right)\right\}^2 r^{z-1}}{\sigma_1|\Sigma|}\left(r^{z-1}+\frac{\xi}{\sigma_1}\right)^{-2}\left\{1-\rho K_r^{-1}(z)\right\} \\ &\quad + \frac{\rho\xi}{2\sigma_1^2|\Sigma|}\frac{\log(r)r^{2-2z}}{\left(1+\frac{\xi}{\sigma_1}r^{1-z}\right)^2}K_r(z)^{-3}, \end{aligned}$$

as $r \to \infty$. Substituting $z = z^*$ into $G_r''(z)$ gives $G_r''(z^*) \sim (2\xi\rho^2|\Sigma|)^{-1}\log(\xi r/\sigma_2)$, as $r \to \infty$. Now consider the function $H_r(z)$, defined in (4.4.33). Denoting A_{r,z^*} and B_{r,z^*} as $A_{r,z}$ and $B_{r,z}$, respectively, but with $z = z^*$, it follows that

$$A_{r,z^*} \sim \sqrt{\frac{2}{\xi}\log(r)}\begin{pmatrix}\rho\\1\end{pmatrix}, \quad \text{and } B_{r,z^*} \sim \frac{1}{2}\sqrt{\frac{\xi}{2\log(r)}}\begin{pmatrix}\frac{1}{\rho}\log\left(\frac{4\pi}{\xi}\rho^2\log(r)\right)\\ \\ \log\left(\frac{4\pi}{\xi}\log(r)\right)\end{pmatrix},$$

as $r \to \infty$, and so $A_{r,z^*}^T\Sigma^{-1}B_{r,z^*} + B_{r,z^*}^T\Sigma^{-1}A_{r,z^*}$ is asymptotically (as in $\sim$)

$$\begin{aligned} &\frac{1}{2|\Sigma|}\left\{\begin{pmatrix}\rho\\1\end{pmatrix}^T\begin{pmatrix}1&-\rho\\-\rho&1\end{pmatrix}B_{r,z^*} + {B_{r,z^*}}^T\begin{pmatrix}1&-\rho\\-\rho&1\end{pmatrix}\begin{pmatrix}\rho\\1\end{pmatrix}^T\right\} \\ &\sim \frac{1}{2|\Sigma|}\left\{\begin{pmatrix}0\\1-\rho^2\end{pmatrix}^T B_{r,z^*} + {B_{r,z^*}}^T\begin{pmatrix}0\\1-\rho^2\end{pmatrix}^T\right\} \\ &\sim \frac{2(1-\rho^2)}{2|\Sigma|}\log\left(\frac{4\pi}{\xi}\log(r)\right) \sim \log\left(\frac{4\pi}{\xi}\log(r)\right), \end{aligned} \tag{4.4.34}$$

and we have

$$B_{r,z^*}{}^T\Sigma^{-1}B_{r,z^*} \sim \frac{\xi}{8|\Sigma|\log(r)}\left[\frac{1}{\rho^2}\left\{\log\left(\frac{4\pi}{\xi}\rho^2\log(r)\right)\right\}^2+\left\{\log\left(\frac{4\pi}{\xi}\log(r)\right)\right\}^2 - 2\log\left(\frac{4\pi}{\xi}\rho^2\log(r)\right)\log\left(\frac{4\pi}{\xi}\log(r)\right)\right], \tag{4.4.35}$$

as $r\to\infty$. By comparing the leading terms of (4.4.34) and (4.4.35), we have that

$$(A_{r,z^*}{}^T\Sigma^{-1}B_{r,z^*}+B_{r,z^*}{}^T\Sigma^{-1}A_{r,z^*})+B_{r,z^*}{}^T\Sigma^{-1}B_{r,z^*} \sim A_{r,z^*}{}^T\Sigma^{-1}B_{r,z^*}+B_{r,z^*}{}^T\Sigma^{-1}A_{r,z^*},$$

as $r\to\infty$, and so

$$H_r(z^*)\sim\exp\left\{\frac{1}{2}\log\left(\frac{4\pi}{\xi}\log(r)\right)\right\}=\left\{\frac{4\pi}{\xi}\right\}^{1/2}(\log(r))^{1/2},$$

as $r\to\infty$. By using Laplace's method, the integral $\int_{\log(2)/\log(r)}^{\infty} f_{R_2,Z}(r,z)\mathrm{d}z$ is asymptotically (as in $\sim$)

$$\begin{aligned}
&r^{1-z^*}\log(r)\sigma_1^{-1}\sigma_2^{-1}(2\pi)^{-1}|\Sigma|^{-1/2}2^{-1}\exp\{G_r(z^*)\}H_r(z^*)\\
&\times\sqrt{\frac{2\pi}{|G_r''(z^*)|}}\sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z^*}\right\}}}\sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_2}r(1-r^{-z^*})\right\}}}\\
&\times\left(1+\frac{\xi}{\sigma_1}r^{1-z^*}\right)^{-1}\left(1+\frac{\xi}{\sigma_2}r(1-r^{-z^*})\right)^{-1}\\
&\sim r^{1-z^*}\log(r)\sigma_1^{-1}\sigma_2^{-1}(2\pi)^{-1}|\Sigma|^{-1/2}2^{-1}\exp\{G_r(z^*)\}\left\{\frac{4\pi}{\xi}\right\}^{1/2}\{\log(r)\}^{1/2}\\
&\times\sqrt{2\pi}(2\xi\rho^2|\Sigma|)^{1/2}\left\{\log\left(\frac{\xi}{\sigma_2}r\right)\right\}^{-1/2}\sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_1}r^{1-z^*}\right\}}}\sqrt{\frac{\xi}{\log\left\{1+\frac{\xi}{\sigma_2}r\right\}}}\\
&\times\sigma_1\xi_1^{-1}r^{-1+z^*}\left(1+\frac{\xi}{\sigma_2}r\right)^{-1}\\
&\sim\left\{\frac{\log(r)}{\log\left(\frac{\xi}{\sigma_2}r\right)}\right\}^{3/2}\sigma_2^{-1}\exp\{G_r(z^*)\}\frac{\rho}{K_r(z^*)}\left(1+\frac{\xi}{\sigma_2}r\right)^{-1}
\end{aligned}$$

$$\sim \xi^{-1}\left(\frac{\xi}{\sigma_2}\right)^{-1/\xi} r^{-1/\xi-1},$$

as $r \to \infty$. Symmetric arguments can be used to show that

$$\int_0^{\log(2)/\log(r)} f_{R_2,Z}(r,z)\mathrm{d}z \sim \xi^{-1}\left(\frac{\xi}{\sigma_1}\right)^{-1/\xi} r^{-1/\xi-1}.$$

By integrating $f_{R_2}(r)$, it follows that

$$\Pr\{R_2 \geq s\} \sim \left\{\frac{\xi}{\left[\sigma_1^{1/\xi}+\sigma_2^{1/\xi}\right]^{-\xi}}\right\}^{-1/\xi} s^{-1/\xi},$$

for $s \to \infty$; this is equivalent to the result derived for the independence case (see Section 4.4.2).

Strictly negative marginal shapes

We now consider $\max\{\xi_1,\xi_2\} < 0$ and, without loss of generality, we assume that $\xi_1 \geq \xi_2$. Recall from (4.4.31) that the joint density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) \sim K(\sigma_1\sigma_2)^{-1}\exp\left\{-\frac{1}{2}(A_{\mathbf{x}}-B_{\mathbf{x}})^T\Sigma^{-1}(A_{\mathbf{x}}-B_{\mathbf{x}})\right\}$$
$$\times\prod_{i=1}^{2}\left\{\sqrt{\frac{\xi_i}{\log\left\{1+\frac{\xi_i}{\sigma_i}x_i\right\}}}\left(1+\frac{\xi_i}{\sigma_i}x_i\right)^{-1}\right\}$$

for constant $K = 2^{-2}(\pi)^{-1}|\Sigma|^{-1/2} > 0$ and as $x_i \to x_i^F$ for $i = 1,2$, where

$$A_{\mathbf{x}}-B_{\mathbf{x}} = \begin{pmatrix}\sqrt{\frac{2}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}}\\ \sqrt{\frac{2}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_2\right\}}\end{pmatrix} - \begin{pmatrix}\frac{\log\left(\frac{4\pi}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}\right)}{2\sqrt{\frac{2}{\xi_1}\log\left\{1+\frac{\xi_1}{\sigma_1}x_1\right\}}}\\ \frac{\log\left(\frac{4\pi}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_2\right\}\right)}{2\sqrt{\frac{2}{\xi_2}\log\left\{1+\frac{\xi_2}{\sigma_2}x_{d2}\right\}}}\end{pmatrix}.$$

We now apply the two-stage transformation $\mathbf{X} \to \mathbf{U} \to (R_2, W)$ given by (4.2.2) and (4.2.3), respectively, and consider (R_2, W) for $R_2 > t$, with t defined in (4.2.5). The joint density of (R_2, W) is

$$f_{R_2,w}(r,w) \sim K(\xi_1\xi_2)^{-1}\left(r - r^F\right)^{-1} \exp\left\{-\frac{1}{2}(A_{r,w} - B_{r,w})^T\Sigma^{-1}(A_{r,w} - B_{r,w})\right\} \times \prod_{i=1}^{2}\sqrt{\frac{\xi_i}{\log\left\{\frac{\xi_i(r-r^F)w_i}{\sigma_i}\right\}}}w_i^{-1}$$

as $r \to r^F$ and where $w_1 = w,\ w_2 = 1 - w_1$ and where

$$A_{r,w} - B_{r,w} = \begin{pmatrix} \sqrt{\frac{2}{\xi_1}\log\left\{\frac{\xi_1(r-r^F)w_1}{\sigma_1}\right\}} \\ \\ \sqrt{\frac{2}{\xi_2}\log\left\{\frac{\xi_2(r-r^F)w_2}{\sigma_2}\right\}} \end{pmatrix} - \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi_1}\log\left\{\frac{\xi_1(r-r^F)w_1}{\sigma_1}\right\}\right)}{2\sqrt{\frac{2}{\xi_1}\log\left\{\frac{\xi_1(r-r^F)w_1}{\sigma_1}\right\}}} \\ \frac{\log\left(\frac{4\pi}{\xi_2}\log\left\{\frac{\xi_2(r-r^F)w_2}{\sigma_2}\right\}\right)}{2\sqrt{\frac{2}{\xi_2}\log\left\{\frac{\xi_2(r-r^F)w_2}{\sigma_2}\right\}}} \end{pmatrix}.$$

We now apply another transformation $(R_2, W) \to (V, Z)$ for $V \in [1 - t/r^F, 1]$ and $Z \in (0, \infty)$, given by (4.2.7), and consider $v \downarrow 0$. The joint density of (V, Z) as $v \downarrow 0$ is

$$\begin{aligned} f_{V,Z}(v,z) &\sim K(\xi_1\xi_2)^{-1}v^{-1}(-\log(v))\exp\left\{-\frac{1}{2}(A_{v,z} - B_{v,z})^T\Sigma^{-1}(A_{v,z} - B_{v,z})\right\} \\ &\quad \times \sqrt{\frac{\xi_1}{\log\left\{\frac{-\xi_1 v r^F v^z}{\sigma_1}\right\}}}\sqrt{\frac{\xi_2}{\log\left\{\frac{-\xi_2 v r^F(1-v^z)}{\sigma_2}\right\}}}(1 - v^z) \\ &\sim K(\xi_1\xi_2)^{-1}v^{-1}(-\log(v))\exp\left\{-\frac{1}{2}(A_{v,z} - B_{v,z})^T\Sigma^{-1}(A_{v,z} - B_{v,z})\right\} \\ &\quad \times \sqrt{\frac{\xi_1}{(1+z)\log(v)}}\sqrt{\frac{\xi_2}{\log(v)}} \\ &\sim K(\xi_1\xi_2)^{-1/2}v^{-1}(1+z)^{-1/2}\exp\left\{-E(v,z)\right\}, \end{aligned}$$

and where the second line follows as $1 - v^z \sim 1$ as $v \downarrow 0$ for all $z \in (0, \infty)$, and where

$$\begin{aligned} E_{v,z} =& \frac{1}{2}(A_{v,z} - B_{v,z})^T \Sigma^{-1} (A_{v,z} - B_{v,z}) \\ =& \frac{1}{2}[A_{v,z}^T \Sigma^{-1} A_{v,z} - A_{v,z}^T \Sigma^{-1} B_{v,z} - B_{v,z}^T \Sigma^{-1} A_{v,z} + B_{v,z}^T \Sigma^{-1} B_{v,z}]. \end{aligned}$$

We now note that

$$\begin{aligned} A_{v,z} &= \begin{pmatrix} \sqrt{\frac{2}{\xi_1} \log\left\{ -\frac{\xi_1 r^F v^{1+z}}{\sigma_1} \right\}} \\ \\ \sqrt{\frac{2}{\xi_2} \log\left\{ -\frac{\xi_2 r^F v(1-v^z)}{\sigma_2} \right\}} \end{pmatrix} \sim \begin{pmatrix} \sqrt{\frac{2}{\xi_1}(1+z)\log(v)} \\ \\ \sqrt{\frac{2}{\xi_2}\log(v)} \end{pmatrix} \\ &\sim \sqrt{-2\log(v)} \begin{pmatrix} (-\xi_1)^{-1/2}(1+z)^{1/2} \\ \\ (-\xi_2)^{-1/2} \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} B_{v,z} &= \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi_1}\log\left\{\frac{-\xi_1 r^F v^{1+z}}{\sigma_1}\right\}\right)}{2\sqrt{\frac{2}{\xi_1}\log\left\{\frac{-\xi_1 r^F v^{1+z}}{\sigma_1}\right\}}} \\ \frac{\log\left(\frac{4\pi}{\xi_2}\log\left\{\frac{-\xi_2 r^F v(1-v^z)}{\sigma_2}\right\}\right)}{2\sqrt{\frac{2}{\xi_2}\log\left\{\frac{-\xi_2 r^F v(1-v^z)}{\sigma_2}\right\}}} \end{pmatrix} \sim \begin{pmatrix} \frac{\log\left(\frac{4\pi}{\xi_1}(1+z)\log(v)\right)}{2\sqrt{\frac{2}{\xi_1}(1+z)\log(v)}} \\ \frac{\log\left(\frac{4\pi}{\xi_2}\log(v)\right)}{2\sqrt{\frac{2}{\xi_2}\log(v)}} \end{pmatrix} \sim 2^{-3/2} \begin{pmatrix} \frac{\log(-\log(v))}{\sqrt{\xi_1^{-1}(1+z)\log(v)}} \\ \frac{\log(-\log(v))}{\sqrt{\xi_2^{-1}\log(v)}} \end{pmatrix} \\ &\sim 2^{-3/2}(-\log(v))^{-1/2}\log(-\log(v)) \begin{pmatrix} (-\xi_1)^{1/2}(1+z)^{-1/2} \\ (-\xi_2)^{1/2} \end{pmatrix}, \end{aligned}$$

as $v \downarrow 0$. Hence, we have that

$$\frac{1}{2}A_{v,z}^T \Sigma^{-1} A_{v,z} \sim \frac{\log(v)}{1-\rho^2}\left\{ \frac{1+z}{\xi_1} + 2\rho\left(\frac{1+z}{\xi_1\xi_2}\right)^{1/2} + \frac{1}{\xi_2} \right\} := \frac{\log(v)}{1-\rho^2} g_1(z),$$

for $g_1(z) = \xi_1^{-1}(1+z) + 2\rho(\xi_1\xi_2)^{-1/2}(1+z)^{1/2} + \xi_2^{-1}$ and

$$\begin{aligned}
&\frac{1}{2}[B_{v,z}^T \Sigma^{-1} A_{v,z} + A_{v,z}^T \Sigma^{-1} B_{v,z}] \\
&\sim \frac{\log(-\log(v))}{2(1-\rho^2)} \left\{ 2 - \rho \left(\frac{\xi_1}{\xi_2} \right)^{1/2} (1+z)^{-1/2} - \rho \left(\frac{\xi_1}{\xi_2} \right)^{-1/2} (1+z)^{1/2} \right\} \\
&\sim \frac{\log(-\log(v))}{2(1-\rho^2)} g_2(z)
\end{aligned}$$

for $g_2(z) = 2 - \rho(\xi_1\xi_2^{-1})^{1/2}(1+z)^{-1/2} - \rho(\xi_1\xi_2^{-1})^{-1/2}(1+z)^{1/2}$ and

$$\begin{aligned}
&\frac{1}{2}[B_{v,z}^T \Sigma^{-1} B_{v,z}] \\
&\sim (1-\rho^2)^{-1} 2^{-4} (-\log(v))^{-1} [\log(-\log(v))]^2 [-\xi_1(1+z)^{-1} - 2\rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2} - \xi_2] \\
&\sim (1-\rho^2)^{-1} 2^{-4} (-\log(v))^{-1} [\log(-\log(v))]^2 g_3(z)
\end{aligned}$$

for $g_3(z) = -\xi_1(1+z)^{-1} - 2\rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2} - \xi_2$ as $v \downarrow 0$. Comparing the dominant terms in each component, we have that

$$E(v,z) \sim \frac{\log(v)}{1-\rho^2} g_1(z) - \frac{\log(-\log(v))}{2(1-\rho^2)} g_2(z) + \frac{[\log(-\log(v))]^2}{2^4(1-\rho^2)} (-\log(v))^{-1} g_3(z),$$

as $v \downarrow 0$. We now proceed under the assumption that the asymptotic behaviour of $\partial E(v,z)/\partial z$ as $v \downarrow 0$ can be derived by taking partial derivatives of a function that is asymptotically equivalent to $E(v,z)$ as $v \downarrow 0$. That is, we assume that

$$\begin{aligned}
\frac{\partial E(v,z)}{\partial z} &\sim \frac{\log(v)}{1-\rho^2} \frac{\partial g_1(z)}{\partial z} - \frac{\log(-\log(v))}{2(1-\rho^2)} \frac{\partial g_2(z)}{\partial z} + \frac{[\log(-\log(v))]^2}{2^4(1-\rho^2)} (-\log(v))^{-1} \frac{\partial g_3(z)}{\partial z} \\
&\sim \frac{\log(v)}{1-\rho^2} \frac{\partial g_1(z)}{\partial z} = \frac{\log(v)}{1-\rho^2} [\xi_1^{-1} + \rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2}],
\end{aligned}$$

as $v \downarrow 0$. Hence it follows that

$$f_V(v) \sim \int_0^\infty K(\xi_1\xi_2)^{-1/2} v^{-1} (1+z)^{-1/2} \exp\{-E(v,z)\} \, \mathrm{d}z$$

$$
\begin{aligned}
&\sim \left[-K_2(v\log(v))^{-1} \left\{ \frac{(1+z)^{-1/2}}{\xi_1^{-1} + \rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2}} \right\} \exp\{-E(v,z)\} \right]_0^\infty \\
&+ K_2(v\log(v))^{-1} \int_0^\infty \left\{ \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{(1+z)^{-1/2}}{\xi_1^{-1} + \rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2}} \right) \right\} \exp\{-E(v,z)\} \\
&\sim \left[-K_2(v\log(v))^{-1} \left\{ \frac{(1+z)^{-1/2}}{\xi_1^{-1} + \rho(\xi_1\xi_2)^{1/2}(1+z)^{-1/2}} \right\} \exp\{-E(v,z) \right]_0^\infty
\end{aligned}
$$

as $v \downarrow 0$ for constant $K_2 = (\xi_1\xi_2)^{-1/2}(1-\rho^2) > 0$. The last line follows by comparing the dominant terms of v in the two integrands; the latter integrand includes a $(\log(v))^{-1}$ which satisfies $(\log(v))^{-1} \to 0$ as $v \downarrow 0$, hence the term on lines four dominates as $v \downarrow 0$. It follows that

$$
\begin{aligned}
f_V(v) &\sim K_3(-v\log(v))^{-1} \exp\{-E(v,0)\} \\
&\sim K_3(-v\log(v))^{-1} \exp\left\{ -\frac{\log(v)}{1-\rho^2} g_1(0) + \frac{\log(-\log(v))}{2(1-\rho^2)} g_2(0) \right. \\
&\left. - \frac{[\log(-\log(v))]^2}{2^4(1-\rho^2)} (-\log(v))^{-1} g_3(0) \right\} \\
&\sim K_3(-\log(v))^{\alpha_R - 1} v^{-1/\xi_R - 1}
\end{aligned}
$$

as $v \downarrow 0$ where

$$
\xi_R = (1-\rho^2)(\xi_1^{-1} + 2\rho(\xi_1\xi_2)^{-1/2} + \xi_2^{-1})^{-1}, \text{ and } \alpha_R = \frac{2 - \rho(\xi_1\xi_2^{-1})^{1/2} - \rho(\xi_1\xi_2^{-1})^{-1/2}}{2(1-\rho^2)} \leq 1,
$$

and where $K_3 = -K_2(\xi_1^{-1} + \rho(\xi_1\xi_2)^{1/2})^{-1} > 0$; this follow as $\rho < 1$ and $\xi_1 \geq \xi_2$. We now make the transformation back to R_2. The marginal density of R_2 is

$$
f_{R_2}(r) \sim K_4 \left[-\log\left(1 - \frac{r}{r^F}\right) \right]^{\alpha_R - 1} \left(1 - \frac{r}{r^F}\right)^{-1/\xi_R - 1},
$$

as $r \to r^F$ and for constant $K_4 = K_3 r^F > 0$. It follows that

$$
\Pr\{R_2 \geq s\} \sim K_4 \int_s^{r^F} \left[-\log\left(1 - \frac{r}{r^F}\right) \right]^{\alpha_R - 1} \left(1 - \frac{r}{r^F}\right)^{-1/\xi_R - 1} \mathrm{d}r
$$

$$\begin{aligned}
&\sim K_5 \left[-\log\left(1-\frac{s}{r^F}\right)\right]^{\alpha_R-1}\left(1-\frac{s}{r^F}\right)^{-1/\xi_R} \\
&- \xi_R(\alpha_R-1)K_4\int_s^{r^F}\left[-\log\left(1-\frac{r}{r^F}\right)\right]^{\alpha_R-2}\left(1-\frac{r}{r^F}\right)^{-1/\xi_R-1}\mathrm{d}r \\
&\sim K_5 \left[-\log\left(1-\frac{s}{r^F}\right)\right]^{\alpha_R-1}\left(1-\frac{s}{r^F}\right)^{-1/\xi_R}
\end{aligned}$$

as $s \to r^F$ and for constant $K_5 = -\xi_R r^F K_4 > 0$. Hence, the survival function of R_2 is

$$\Pr\{R_2 \geq s\} \sim K_5 L(s)\left(1-\frac{s}{r^F}\right)^{-1/\xi_R},$$

where $L(s)$ is a slowly-varying function and as required. Note that in the case where $\xi_1 = \xi_2 = \xi$, we have that

$$\xi_R = (1-\rho^2)(\xi^{-1}+2\rho((-\xi)^2)^{-1/2}+\xi^{-1})^{-1} = (1-\rho^2)\xi(2-2\rho)^{-1} = \xi(1+\rho)/2 = \xi\eta_2,$$

or equivalently, $\xi_R = \xi\{\mathbb{1}_2\Sigma^{-1}\mathbb{1}_2\}^{-1}$, see Table 4.3.2.

5

Modelling extremes of spatial aggregates of precipitation using conditional methods

5.1 Introduction

Fluvial flooding is typically not caused by high intensity extreme rainfall at single locations, but by the extremes of precipitation events which are aggregated over spatial catchment areas. Accurate modelling of such events can help to mitigate the financial impacts associated with floods, especially if river defences are built to withstand a $T-$year event of this kind. Approaches to quantifying the tail behaviour of spatial aggregates exist in the literature; however, these techniques are often simplistic or make unrealistic assumptions about the behaviour of the process for which they are trying to model. We present a novel methodology for making inference on the tail behaviour of spatial aggregates, which we apply in the context of extreme precipitation aggregates.

We define a spatial process $\{Y(s) : s \in \mathcal{S}\}$ for some spatial domain $\mathcal{S}$. Our interest

lies in the upper tail behaviour of the spatial aggregate $R_{\mathcal{A}}$ on regions $\mathcal{A} \subset \mathcal{S} \subset \mathbb{R}^2$,

$$R_{\mathcal{A}} = \int_{\mathcal{A}} Y(s)\mathrm{d}s, \tag{5.1.1}$$

for different, possibly overlapping, $\mathcal{A}$, and the joint behaviour of $(R_{\mathcal{A}}, R_{\mathcal{B}})$ for $\mathcal{A}, \mathcal{B} \subset \mathcal{S}$. Typically, the data we would have available for inference are realisations of $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_d))$ for $t = 1, \ldots, n$, which are observations of said process $\{Y(s)\}$ at d sampling locations $\mathbf{s} = (s_1, \ldots, s_d) \subset \mathcal{S}$ at n sampling times. Note that $\mathbf{s}$ need not be point locations; they can instead be non-overlapping grid-boxes. Data produced by climate models are often available in this form and observations of $Y(s_i),\ i = 1, \ldots, d$, correspond to spatial aggregates themselves, as they are typically presented as an average over the grid box s_i. In these circumstances, the integral in (5.1.1) can be replaced with the equivalent summation, but our methodology is still applicable; see Section 5.4. We assume that both the full marginal behaviour, and dependence, of $\{Y_t(s)\}$ is stationary with respect to time. Marginally, the upper tail behaviour of $Y(s)$ is assumed to be characterised by a generalised Pareto distribution (GPD) with scale and shape parameters, $\upsilon(s) > 0$ and $\xi(s)$, respectively, that vary smoothly over $s \in \mathcal{S}$ (Davison and Smith, 1990). Dependence in $\{Y(s)\}$ is characterised through a marginal transformation to the process $\{X(s) : s \in \mathcal{S}\}$, which has standardised margins; further details are given in Section 5.2.2. Thus, we can rewrite (5.1.1) as

$$R_{\mathcal{A}} = \int_{\mathcal{A}} Y(s)\mathrm{d}s = \int_{A} F^{-1}_{Y(s)} \left\{ F_X \left[X(s) \right] \right\} \mathrm{d}s,$$

where $F_{Y(s)}(\cdot)$ and $F_X(\cdot)$ are the marginal CDFs of $\{Y(s)\}$ and $\{X(s)\}$, respectively. We will focus on the situation where $\{X(s)\}$ is a stationary process; an assumption that we find holds well for our application (see Section 5.4.4). If this assumption did not hold, a wide-range of literature exists for both non-parametric, and parametric, methods that account for non-stationarity in extremal dependence (Huser and Genton,

2016; Richards and Wadsworth, 2021), and these methods can easily be incorporated into our methodology.

There are three main existing modelling approaches for inference on the upper tails of $R_{\mathcal{A}}$: univariate methods, spatial approaches that focus on modelling all of the data, and spatial approaches that focus on modelling only the extremes; our approach falls in the latter class, making less restrictive assumptions than previous methods of this type.

We first consider the univariate case. Within an extreme value analysis framework, univariate methods for estimating the size of $T-$year events are well studied and cemented in asymptotic theory (Coles, 2001). If we can create a sample of observations of $R_{\mathcal{A}}$, we can use univariate methods to make inference on its upper tail, i.e., fit a GPD to exceedances of a sample of $R_{\mathcal{A}}$ above some fixed threshold and then extrapolate to high quantiles. However, creating this sample can be challenging. If $\mathbf{s}$ are regularly spaced point locations, or contiguous non-overlapping grid-boxes, then (5.1.1) can be approximated using the sum of the elements of $\mathbf{Y}_t$. However, if $\mathbf{s}$ are irregularly spaced, we may be required to compute a weighted sum to approximate (5.1.1), with the weights to be determined somehow. Further complications arise if we have partially missing observations of $\{Y(s)\}$. Even if these issues are overcome, when using univariate methods we lose the information present in the margins of $\{Y(s)\}$ and dependence of $\{X(s)\}$. If the process we are considering is precipitation, this can lead to inference that is not self-consistent and may be physically unrealistic; a trait that can be undesirable to practitioners. To explain this further, observe that, for precipitation, $\{Y(s)\}$ is non-negative everywhere, i.e., $Y(s) \geq 0$ for all $s \in \mathcal{S}$. Trivially it follows that $R_{\mathcal{A}} \geq R_{\mathcal{B}}$ for all $\mathcal{B} \subseteq \mathcal{A} \subset \mathcal{S}$ and hence return levels should be similarly ordered. This natural ordering may not follow if we take a simple univariate approach to modelling the upper tail behaviour of the $R_{\mathcal{A}}$ and $R_{\mathcal{B}}$ aggregates separately (Nadarajah et al., 1998). To prevent this from occurring, we fit a model

for the process $\{Y(s)\}$, of which observations may be partially missing or complete. We then simulate from our model for $\{Y(s)\}$ for $s \in \mathcal{S}$ and compute realisations of $R_{\mathcal{A}}$.

In the context of precipitation aggregates, one richly studied approach has developed a class of stationary stochastic processes to model the whole precipitation intensity process, continuous in both time and space. These models typically describe the intensity as the accumulation, at each point in time and space, over a random number of simple shaped individual stochastic rain cells, which cluster in time and space, and move on stochastic trajectories. These models were first developed for a single site by Rodriguez-Iturbe et al. (1987), then developed spatially by Northrop (1998) and some of the more recent methods are summarised by Wheater et al. (2005). These models are typically estimated by optimising the fit against a range of characteristics of observed fields. As a result, these models can capture well the features of typical precipitation fields. However, for deriving the distribution of quantities like the upper tail of $R_{\mathcal{A}}$, the models and their inference have limitations as there is no guarantee that models for the body of a process fit well to the extremes. Yet it is precipitation fields that are extreme somewhere in $\mathcal{A}$ that yield extremes of $R_{\mathcal{A}}$ unless $\mathcal{A}$ is very large relative to the range of spatial dependence, but in that case their method's assumption of stationarity is likely to be unreliable.

A typical approach to modelling extreme fields is the use of max-stable models, see Padoan et al. (2010); Westra and Sisson (2011); Reich and Shaby (2012). These models are predominately fit to component-wise block maxima, typically annual maxima, at sampling locations, but cannot be used to make inference about the extremal dependence structure of individual precipitation fields as they cannot account for zeros, and hence neither can be used to describe the distribution of the aggregate. Typically annual maxima do not occur concurrently for different sampling locations and so aggregating over realisations from a max-stable process is not appropriate for inference

on aggregates.

Coles (1993) rectified some of these issues by using a point-process representation of a max-stable field to derive the profile of concurrent events. Coles and Tawn (1996) used this formulation to derive closed form results for the tail behaviour of $R_{\mathcal{A}}$ where the tail parameters are determined by the marginal GPD parameters of $\{Y(s)\}$ and its dependence structure; Ferreira et al. (2012) formalise these results and provide some non-parametric extensions. Further extensions of this framework by Engelke et al. (2019a) relate not only the extremal behaviour of $\{Y(s)\}$ and the aggregates $R_{\mathcal{A}}$, but also the joint behaviour of aggregates over different regions, $\mathcal{A}$. All of these modelling approaches rely on the marginal shape parameters of $\{Y(s)\}$ to be spatially homogeneous i.e., $\xi(s) = \xi$ for all $s \in \mathcal{A}$, for each $\mathcal{A}$ of interest. This assumption is unlikely to hold for applications to larger regions. When $\xi(s)$ varies over a region, models based on the limiting behaviour of the aggregates of $\{Y(s)\}$ are likely to fail. For example, Richards and Tawn (2021) show that when $\xi(s) > 0$ for at least one $s \in \mathcal{S}$, the tail behaviour of $R_{\mathcal{A}}$ will be driven solely by the upper tail behaviour at locations $\mathbf{s} = \arg\max\{\xi(s) : s \in \mathcal{S}\}$. We construct a sub-asymptotic spatial model that avoids the spatial homogeneity constraint. For non-homogeneous shape parameters, de Fondeville and Davison (2020) use functional Pareto processes to model the dependence in $\{Y(s)\}$ and Palacios-Rodríguez et al. (2020) illustrate non-parametric Pareto process modelling to simulate extreme precipitation fields, re-sampling event profiles from observed, gridded data. Both approaches have major limitations for applications due to their dependence structure, as described next.

A particular restriction of using models based on max-stable, or Pareto, processes is that they allow for a restrictive class of dependence structures only. Asymptotic dependence describes the co-occurrence of extremal events and is often quantified through the upper tail index $\chi(s_A, s_B)$ (Joe, 1997) for all $s_A, s_B \in \mathcal{S}$, which can be

defined for $\{Y(s)\}$ as $\chi(s_A, s_B) = \lim_{q \uparrow 1} \chi_q(s_A, s_b)$, where

$$\chi_q(s_A, s_B) = \Pr\{Y(s_B) > F^{-1}_{Y(s_B)}(q) | Y(s_A) > F^{-1}_{Y(s_A)}(q)\}. \tag{5.1.2}$$

In practice, we cannot estimate $\chi(s_A, s_B)$ as $q \uparrow 1$. Instead, estimates are provided by fixing some high threshold $q < 1$ and approximating $\chi(s_A, s_B)$ using $\chi_q(s_A, s_B)$. Max-stable, or Pareto, processes are asymptotically dependent (Coles et al., 1999; Coles, 2001), or perfectly independent, at all spatial distances. That is, for any max-stable, or Pareto, process exhibiting positive spatial association, we have $\chi(s_A, s_B) > 0$ for all $s_A, s_B \in \mathcal{S}$. These models are then unable to account for cases where we have positive association, but $\chi(s_A, s_B) = 0$ for some $s_A, s_B \in \mathcal{S}$ which holds for all Gaussian processes when $s_A \neq s_B$; we refer to this scenario as asymptotic independence, i.e., the tendency for extreme events to occur increasingly independently as the magnitude of the events gets larger. Extensions of max-stable processes, such as max-infinitely divisible processes (Huser et al., 2020), for component-wise maxima can account for asymptotic independence in data. Bopp et al. (2020) illustrate good fits for these models to block-maxima data, but they are not appropriate for precipitation event data.

Wadsworth and Tawn (2019) have developed a conditional approach to spatial extremes. They provide a spatial extension of the multivariate Heffernan and Tawn (2004) model, which enables the modelling of processes given that at least one location in the process is extreme. Dependence parameters within the Heffernan and Tawn (2004) model are represented as smooth functions, parametric or splines, of distance between variables at the site of interest and the conditioning site, and the residual process is driven by a latent Gaussian process, see Section 5.2.2. This modelling approach allows for both asymptotic dependence and asymptotic independence at different spatial distances. We adapt this approach for modelling extreme precipitation fields. Our model outperforms approaches that restrict the dependence in $\{Y(s)\}$

to asymptotic dependence when the true process is asymptotically independent; this is illustrated in Section 5.4.4. However, even if the true process was asymptotically dependent, our model is able to capture this behaviour.

Extensions of Wadsworth and Tawn (2019) are provided by Tawn et al. (2018); Shooter et al. (2021); Simpson and Wadsworth (2021); Simpson et al. (2020); Huser and Wadsworth (2020). These papers cover extremal modelling of air and sea temperature fields and spatial wave heights. Most of these applications use a small numbers of sampling locations ($d < 300$), and full inference is computationally feasible. One exception is Simpson et al. (2020) who detail an approach for fitting the conditional spatial extremes model with much larger d using INLA; however, this imposes restrictions on the dependence structure parameters that are not appropriate in our application. We have $d = 934$ and so find that some non-parametric approaches to modelling dependence parameters are infeasible; we explore novel parametric forms for these. Due to the high value of d, we explore a stratified sampling scheme for model fitting and develop a novel bootstrap method that allows us to estimate uncertainty for estimated parameters. We find that when using Monte-Carlo methods to approximate $\Pr\{R_{\mathcal{A}} > r\}$, that the position and size of $\mathcal{A}$ within $\mathcal{S}$ are important considerations that must be taken into account, as we observe edge effects on this inference for $R_{\mathcal{A}}$.

The model is applied to precipitation data from the 2018 UK climate projections (Lowe et al., 2018). The data are from a convection permitting model, and we find that the extremal behaviour of the underlying process is driven by spatially-localised events consistent with intensive convective rainfall. We observe high variability in the fitted model for $\{Y(s)\}$ as we move further away from the centre of an event, which corresponds to the observed roughness in events that generate extreme precipitation. We further find that the dependence model for $\{Y(s)\}$ fits well and that we are able to comfortably handle zeroes in the data. We find strong indication that we can

replicate the empirical distribution of $R_{\mathcal{A}}$ using Monte-Carlo methods, and so we have evidence to suggest that the further inferences we make about the tail behaviour of the aggregates are well-founded.

The layout of this paper is as follows: Section 5.2 describes our model for the process $\{Y(s)\}$. We describe methods for model inference and simulation of events in Section 5.3, which includes our censoring technique for handling zero values. In Section 5.4 we discuss the marginal and dependence model fits for the precipitation data, and inference on the tail behaviour of spatial aggregates of these data. We compare the results from our approach with those using GPD fitted to the sample aggregates and using a spatial asymptotically dependent model in Section 5.4.4. We end with further discussion and model extensions in Section 5.5.

5.2 Modelling the extremes of the spatial process

5.2.1 Marginal model

The site-wise marginals of $\{Y(s)\}$ can be modelled using a fitted GPD distribution above some high threshold and the empirical distribution below (Coles, 2001). We extend this approach by incorporating a third component, which we denote $p(s)$, that describes the probability that there is no rain at site s. The marginal distribution function of $Y(s)$ for each $s \in \mathcal{S}$ is

$$F_{Y(s)}(y) = \begin{cases} p(s) & \text{if } y = 0 \\ \frac{1-\lambda(s)-p(s)}{F_{Y_+(s)}(q(s))} F_{Y_+(s)}(y) + p(s) & \text{if } 0 < y \leq q(s) \\ 1 - \lambda(s) \left[1 + \frac{\xi(s)(y-q(s))}{\upsilon(s)}\right]_+^{-1/\xi(s)} & \text{if } y > q(s), \end{cases} \tag{5.2.1}$$

where $\upsilon(s) > 0$ and $F_{Y_+(s)}(y)$ denotes the distribution function of strictly positive values of $Y(s)$ and $p(s) \geq 0,\ \lambda(s) > 0$ and $p(s) + \lambda(s) < 1$; this ensures that the

marginal distribution is continuous across components. We expect spatial smoothness over $F_{Y(s)}$ and so define full spatial models for the three components of (5.2.1) which also enable us to make inference about $F_{Y(s)}$ for all $s \in \mathcal{S}$, i.e., including where $Y(s)$ is not observed.

We first consider the distribution of $Y(s)$ above $q(s)$. Following the approach of Youngman (2019), we fit a generalised additive GPD model (GAM) to exceedances $Y(s) - q(s)$. This allows us to represent the GPD scale and shape parameters, $\upsilon(s)$ and $\xi(s)$, respectively, through a basis of smooth splines. We set $\lambda(s) = \lambda$ for all $s \in \mathcal{S}$, allowing us to estimate $q(s)$ for $s \in \mathcal{S}$. We use a non-parametric approach, and simply fit a thin-plate spline to point-wise estimates of $q(s)$ for the associated λ, as the parametric method of Youngman (2019) fails as we have point masses below $q(s)$, caused by rounding of data.

We estimate $p(s) = \Pr\{Y(s) = 0\}$ for $s \in \mathcal{S}$ as a spatially smooth surface by using a logistic GAM (Fasiolo et al., 2020); that is, we fit $\text{logit}\{\mathbb{E}[p(s)]\} = g(s)$ where $g(\cdot)$ is a smoothing spline. The degree of smoothness in the surface $p(s)$ is determined by the choice of spline used for g.

For the distribution for $0 < y \leq q(s)$, we estimate $F_{Y_+}(s)$ using the empirical distribution of strictly positive values of $Y(s)$, which we denote $\tilde{F}_{Y_+(s)}(\cdot)$. We use the site-wise empirical distribution for fitting and recognise that, should we require simulation of $Y(s)$ for $s \in \mathcal{S} \setminus \mathbf{s}$, we can use additive quantile regression (Fasiolo et al., 2020) to compute the empirical distributions at unobserved locations. This can be performed, if necessary, using only a local neighbourhood of sampling locations, as this method is too computationally expensive for large d and so is not viable without dimension reduction.

We use (5.2.1) to perform a site-wise standardisation of the margins of the data. For modelling dependence within a process $\{X(s) : s \in \mathcal{S}\}$ using the Wadsworth and Tawn (2019) conditional extremes framework, we require its margins to have standard

exponential upper tails, i.e., $\Pr\{X(s) > x\} \sim C\exp(-x)$ for some $C > 0$, as $x \to \infty$ and for all $s \in \mathcal{S}$. We follow Keef et al. (2013) and Tawn et al. (2018) and use Laplace margins.

5.2.2 Dependence modelling

Wadsworth and Tawn (2019) model the underlying extremal dependence in our standardised process $\{X(s)\}$, given that it is extreme for some $s \in \mathcal{S}$, by first conditioning on the process being above some high threshold u at a specified site $s_O \in \mathcal{S}$. We introduce the function $h(s_A, s_B) = \|s_A - s_B\|$ for $s_A, s_B \in \mathcal{S}$, where $\|\cdot\|$ is some distance metric (we use the anisotropic measure (5.2.11)). Under the assumption that there exists normalising functions $\{a : (\mathbb{R}, \mathbb{R}_+) \to \mathbb{R}\}$, with $a(x, 0) = x$, and $\{b : (\mathbb{R}, \mathbb{R}_+) \to (0, \infty)\}$, such that as $u \to \infty$, they assume that for each $s_O \in \mathcal{S}$

$$\left(\left\{\frac{X(s) - a\{X(s_O), h(s, s_O)\}}{b\{X(s_O), h(s, s_O)\}} : s \in \mathcal{S}\right\}, X(s_O) - u\right) \Bigg| \Big(X(s_O) > u\Big) \xrightarrow{d} \left(\Big\{Z(s|s_O) : s \in \mathcal{S}\Big\}, E\right), \tag{5.2.2}$$

where E is a standard exponential variable and process $\{Z(s|s_O)\}$ which is non-degenerate for all $s \in \mathcal{S}$ where $s \neq s_O$. That is, there is convergence in distribution of the normalised process to $\{Z(s|s_O) : s \in \mathcal{S}\}$, termed the residual process, which is independent of E, and $Z(s_O|s_O) = 0$ almost surely. Characterisations of the normalising functions, a and b and the residual process $Z(s|s_O)$ are given in Section 5.2.2.

To make inference on the upper tail of $R_{\mathcal{A}}$ for any $\mathcal{A} \subset \mathcal{S}$ we require the process $\{X(s)\}$ given an extreme value somewhere in the domain $\mathcal{S}$, i.e.,

$$\Big\{X(s) : s \in \mathcal{S}\Big\} \Big| \Big(\max_{s \in \mathcal{S}} X(s) > u\Big) \tag{5.2.3}$$

for large u. Limit (5.2.2) conditions only on observing an exceedance at a spe-

cific site $s_O \in \mathcal{S}$, so cannot be immediately used. However, this limit provides a core building block for what is required when combined with a limiting model for $\{X(s_O) > u : s_O \in \mathcal{S}\}|(\max_{s\in\mathcal{S}} X(s) > u)$ as $u \to \infty$. For a stationary process, this limiting model will be invariant to s_O for all $s_O \in \mathcal{S}$ which are sufficiently far from the boundaries of $\mathcal{S}$. Wadsworth and Tawn (2019) show how simulation from process (5.2.3) can be achieved through an importance sampling method; the outline of this is given in Section 5.3.4.

For the process $\{X(s)\}$ to be ergodic over $\mathbb{R}^2$, we need conditions on a and b and $Z(s|s_O)$ so that independence is achieved as $h = h(s, s_O) \to \infty$ for any $s_O \in \mathcal{S}$ and suitably distanced $s \in \mathcal{S}$. This requires for any fixed $x > 0$, that $a(x, h) \to 0$ and $b(x, h) \to 1$ as $h \to \infty$. Further, the residual process $Z(s|s_O)$ must have identical margins to $X(s)$ as $h \to \infty$; in particular, we require standard Laplace margins for $Z(s|s_O)$ as $h \to \infty$.

Normalising functions

For inference, we assume parametric forms for the a and b normalising functions in limit (5.2.2). Wadsworth and Tawn (2019) provide a discussion of these normalising functions and provide some suggestions for their possible parametric forms. We considered the range of parametrics forms discussed by Wadsworth and Tawn (2019); Tawn et al. (2018); Shooter et al. (2021), but for brevity we report only the models that provided the best fit. We let

$$a(x, h) = x\alpha(h), \quad \text{with} \quad \alpha(h) = \begin{cases} 1, & h \leq \Delta, \\ \exp(-\{(h - \Delta)/\kappa_{\alpha_1}\}^{\kappa_{\alpha_2}}), & h > \Delta, \end{cases} \tag{5.2.4}$$

where $\Delta \geq 0$ and $\kappa_{\alpha_1}, \kappa_{\alpha_2} > 0$ which allows $\{X(s)\}$ to be asymptotically dependent up to distance Δ from s_O, and asymptotically independent thereafter. We also take

$$b(x,h) = x^{\beta(h)}, \quad \text{with} \quad \beta(h) = \kappa_{\beta_3} \exp(-\{h/\kappa_{\beta_1}\}^{\kappa_{\beta_2}}) \tag{5.2.5}$$

for $\kappa_{\beta_1}, \kappa_{\beta_2} > 0$ and $\kappa_{\beta_3} \in [0,1]$, and so $b(0,x) = x^{\kappa_{\beta_3}}$. Ergodicity holds for $\{X(s)\}$ whatever the parameters of a and b, as $a(x,h) \to 0$ and $b(x,h) \to 1$ as $h \to \infty$ for fixed $x > 0$.

Residual process $\{Z(s|s_O)\}$

We follow Shooter et al. (2021) by imposing that the residual process $\{Z(s|s_O)\}$ has delta-Laplace margins; a random variable follows a delta-Laplace distribution, i.e., $DL(\mu, \sigma, \delta)$, with location, scale and shape parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\delta > 0$, respectively, if its density is

$$f(z) = \frac{\delta}{2k\sigma\Gamma\left(\frac{1}{\delta}\right)} \exp\left\{ -\left| \frac{z-\mu}{k\sigma} \right|^{\delta} \right\}, \quad (z \in \mathbb{R}) \tag{5.2.6}$$

with $\Gamma(\cdot)$ as the standard gamma function and $k^2 = \Gamma(1/\delta)/\Gamma(3/\delta)$. The scaling by k is used to improve identifiability between σ and δ, as the random variable has expectation μ and variance σ^2 regardless of the value of δ. Use of the delta-Laplace distribution introduces flexibility in the marginal choice for $Z(s|s_O)$, as for $\delta = 1$ or 2, we have the Laplace or Gaussian distributions respectively. As with the normalising functions, we parametrise the delta-Laplace parameters as smooth functions of distance from the conditioning site s_O. That is, $Z(s|s_O) \sim DL(\mu\{h(s,s_O)\}, \sigma\{h(s,s_O)\}, \delta\{h(s,s_O)\})$,

with

$$\begin{aligned}
\mu(h) &= \kappa_{\mu_1} h^{\kappa_{\mu_2}} \exp\{-h/\kappa_{\mu_3}\}, && (\kappa_{\mu_2} > 0, \kappa_{\mu_3} > 0),\\
\sigma(h) &= \sqrt{2}\left(1 - \exp\{-(h/\kappa_{\sigma_1})^{\kappa_{\sigma_2}}\}\right), && (\kappa_{\sigma_1} > 0, \kappa_{\sigma_2} > 0),\\
\delta(h) &= 1 + (\kappa_{\delta_1} h^{\kappa_{\delta_2}} - \kappa_{\delta_4}) \exp\{-h/\kappa_{\delta_3}\}, && (\kappa_{\delta_1} \geq 0, \kappa_{\delta_2} > 0, \kappa_{\delta_3} > 0, \kappa_{\delta_4} \geq 0),
\end{aligned} \tag{5.2.7}$$

for $h \geq 0$. These functions satisfy the constraint that $\mu(0) = \sigma(0) = 0$, which ensures that $Z(s_O|s_O) = 0$ holds and provides a flexible modelling choice for δ. We do not constrain any particular value of $\delta(0)$ and instead let $\delta(0) = 1 - \kappa_{\delta_4} > 1$. Note that we require that $\delta(h) \geq 1$ for all h to ensure that the residual process $Z(s|s_O)$ does not have heavier upper-tails than $X(s)$ for any $s \in \mathcal{S}$. Furthermore, ergodicity of $X(s)$ is achieved as $\mu(h) \to 0$, $\sigma^2(h) \to 2$, and $\delta(h) \to 1$ as $h \to \infty$, where the variance of a standard Laplace random variable is 2.

Following the approach of Shooter et al. (2021), dependence in $\{Z(s|s_O)\}$ is induced by first considering the process

$$W(s|s_O)\} = \{W(s)|(W(s_O) = 0)\} \tag{5.2.8}$$

for all $s \in \mathcal{S}$, where $\{W(s)\}$ is a standard stationary Gaussian process with correlation function $\rho(h)$. We set $\{Z(s|s_O)\} = \{F^{-1}_{Z(s|s_O)}\{\Phi[W(s|s_O)]\}\}$ for all $s \in \mathcal{S}$, where $\Phi(\cdot)$ and $F_{Z(s|s_O)}$ are the CDFs of a standard Gaussian distribution and $Z(s|s_O)$, respectively. The corresponding density function to $F_{Z(s|s_O)}$ is $f_{Z(s|s_O)}$, defined by (5.2.6) and (5.2.7).

To illustrate the dependence in $\{Z(s|s_O) : s \in \mathcal{S}\}$, we consider the joint distribution of $Z(s|s_O)$ in a finite-dimensional setting, which we achieve by using a Gaussian copula model. Consider any $s_O \in (s_1, \ldots, s_d)$, and without loss of generality, rewrite the sampling locations as $s_O, (s_1, \ldots, s_{d-1})$, i.e., here we illustrate with $s_O = s_d$. The

joint distribution of $\{Z(s_1|s_O), \ldots, Z(s_{d-1}|s_O)\}$ for s_O is, for $\mathbf{z} = (z_1, \ldots, z_{d-1})$,

$$F_{s_O}(\mathbf{z}) = \Phi_{d-1}\left\{\Phi^{-1}(F_{Z(s_1|s_O)}(z_1)), \ldots, \Phi^{-1}(F_{Z(s_{d-1}|s_O)}(z_{d-1})); \mathbf{0}, \Sigma\right\}, \tag{5.2.9}$$

where $\Phi_{d-1}(\cdot; \mathbf{0}, \Sigma)$ is the CDF of a $(d-1)-$dimensional Gaussian distribution with mean $\mathbf{0}$. The correlation matrix Σ must account for the conditioning $W(s)|(W(s_O) = 0)$. To create Σ, we initialise a stationary correlation matrix Σ^* using correlation function $\rho(\cdot)$ evaluated for all pairwise distances, and we condition on observing $W(s_O) = 0$. That is, the correlation matrix Σ has (i, j)-th element

$$\Sigma_{ij} = \frac{\Sigma^*_{ij} - \Sigma^*_{i0}\Sigma^*_{j0}}{(1 - \Sigma^{*2}_{i0})^{1/2}(1 - \Sigma^{*2}_{j0})^{1/2}}. \tag{5.2.10}$$

Note the elements of Σ are normalised such that the diagonal elements are equal to one. In our application, $\rho(\cdot)$ is taken to be the Matérn correlation function

$$\rho(h) = \frac{1}{2^{\kappa_{\rho_2}-1}\Gamma(\kappa_{\rho_2})}\left(\frac{2h\sqrt{\kappa_{\rho_2}}}{\kappa_{\rho_1}}\right)^{\kappa_{\rho_2}} K_{\kappa_{\rho_2}}\left(\frac{2h\sqrt{\kappa_{\rho_2}}}{\kappa_{\rho_1}}\right), \quad (\kappa_{\rho_1} > 0, \kappa_{\rho_2} > 0),$$

where $K_{\kappa_{\rho_2}}(\cdot)$ is the modified Bessel function of the second kind of order κ_{ρ_2}.

To account for spatial anisotropy in the extremal dependence structure of $\{X(s)\}$ we use the transformation of coordinates

$$s^* = \begin{pmatrix} 1 & 0 \\ 0 & 1/L \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} s, \tag{5.2.11}$$

where $\theta \in [-\pi/2, 0]$ controls rotation and $L > 0$ controls the coordinate stretching effect; with $L = 1$ recovering the isotropic model. We define our distance metric $\|s_A - s_B\| = \|s^*_A - s^*_B\|_*$, where $\|\cdot\|_*$ denotes great-circle, or spherical, distance.

Extensions for censored precipitation data

The non-zero probability of zeroes for precipitation $Y(s)$ causes the $X(s)$, for all $s \in \mathcal{S}$, to have non-zero mass at a finite lower endpoint. Consequently, the Gaussian copula and delta-Laplace marginal model described in (5.2.9) are not appropriate for the transformed precipitation data in these lower tail regions. To circumvent this issue, we apply censoring at all points where $Y_t(s) = 0$ for all $s \in \mathcal{S}$ and $t = 1, \ldots, n$. A spatially-varying censoring threshold $c(s)$ is attained by transforming $p(s)$ in Section 5.2.1 to the Laplace scale, using $c(s) = F_L^{-1}\{p(s)\}$, where $F_L(\cdot)$ is the standard Laplace CDF. We then assert that $Y_t(s) = 0 \Leftrightarrow X_t(s) \leq c(s)$, which in turn implies that $Z_t(s|s_O) \leq c_t^{(s_O)}(s)$ where $c_t^{(s_O)}(s)$ is dependent on the value observed at the conditioning site and given in (5.3.1).

For inference, the number of censored components varies at each time point; the number, locations and censoring values can all vary, with a maximum value of $d - 1$ locations with censoring. If the number of censored components is large, it is clear from (5.2.9) that evaluation of a censored distribution function will be computationally expensive. We take a pseudo-likelihood approach to inference, which we detail in Section 5.3.1; this requires only a bivariate density. We also detail its multivariate analogue; although this is not used for inference, it has a variety of uses, i.e., infilling of extreme events with missing observations or inference at sites $s \in \mathcal{S} \setminus \mathbf{s}$. These features are detailed in Appendix C.1.

5.3 Inference and simulation

5.3.1 Model fitting

Our censored triplewise likelihood approach for model fitting is based on the pseudo-likelihood approach of Padoan et al. (2010); their pairwise approach provides unbiased estimation of model dependence parameters. Recall that some observations are

right-censored at different sampling locations with varying rate of occurrence over time. To define a single likelihood contribution at time t, we begin by considering a single conditioning site amongst the observed sites $s_i \in (s_1, \ldots, s_d)$ such that $y_t(s_i) > F^{-1}_{Y(s_i)}\{F_L(u)\}$ and hence $x_t(s_i) > u$. We then define the set of all such times by $\mathcal{T}^{(s_i)} = \{t = 1, \ldots, n : x_t(s_i) \geq u\}$. For the observed sites, we define $h_{i,j} = h(s_i, s_j)$ for $i, j = 1, \ldots, d$ with $i \neq j$. Then for each site s_j , $j = 1, \ldots, d, j \neq i$, we define the residual for time $t \in \mathcal{T}^{(s_i)}$ for conditioning site s_i as

$$z_t^{(s_i)}(s_j) = \begin{cases} [x_t(s_j) - a\{x_t(s_i), h_{i,j}\}]/b\{x_t(s_i), h_{i,j},\} & \text{if } x_t(s_j) > c(s_j), \\ c_t^{(s_i)}(s_j), & \text{otherwise,} \end{cases} \tag{5.3.1}$$

where a and b are described in Section 5.2.2, and $c_t^{(s_i)}(s_j) = \frac{c(s_j) - a\{x_t(s_i), h_{i,j}\}}{b\{x_t(s_i), h_{i,j}\}}$ is the censored residual for site s_j with conditioning site s_i and $t \in \mathcal{T}^{(s_i)}$. The full pseudo-likelihood is given, for residuals $\mathbf{z}_t^{(s_i)}$ with conditioning site s_i and parameter vector $\boldsymbol{\psi}$, by

$$\begin{aligned} L_{CL}(\boldsymbol{\psi}) &= \prod_{i=1}^{d} \prod_{t \in \mathcal{T}^{(s_i)}} L_{CL}^{s_i}(\boldsymbol{\psi}; \mathbf{z}_t^{(s_i)}) \\ &= \prod_{i=1}^{d} \prod_{t \in \mathcal{T}^{(s_i)}} \prod_{\forall j<k; j\&k\neq i} \frac{g_{s_i}(z_t^{(s_i)}(s_j), z_t^{(s_i)}(s_k), c_t^{(s_i)}(s_j), c_t^{(s_i)}(s_k))}{J(z_t^{(s_i)}(s_j), z_t^{(s_i)}(s_k), c_t^{(s_i)}(s_j), c_t^{(s_i)}(s_k))}, \end{aligned} \tag{5.3.2}$$

where $L_{CL}^{s_i}$ is the censored likelihood contribution for s_i and the bivariate density g is defined in Appendix C.1. The Jacobian term $J\left(z_t^{(s_i)}(s_j), z_t^{(s_i)}(s_k), c_t^{(s_i)}(s_j), c_t^{(s_i)}(s_k)\right)$ is

$$b\{x_t(s_i), h_{i,j}\}^{1\{z_t(s_j) > c_t^{(s_i)}(s_j)\}} b\{x_t(s_i), h_{i,k}\}^{1\{z_t(s_k) > c_t^{(s_i)}(s_k)\}},$$

where $1\{\cdot\}$ is the indicator function. The parameter vector $\boldsymbol{\psi}$ contains all parameters of the normalising functions a and b, the marginal parameter functions μ, σ and δ,

the correlation ρ and the anisotropy. Estimation of $\boldsymbol{\psi}$ can be achieved by maximising (5.3.2).

5.3.2 Stratified sampling regime

Clearly, maximising (5.3.2) is computationally infeasible if d is large, as evaluation of (5.3.2) requires $(d-1)(d-2)\sum_{i=1}^{d}|\mathcal{T}^{(s_i)}|/2$ evaluations of $g_{(s_i)}$, which can require double integrals and grows as $O(d^3n)$. We detail a stratified sampling regime to create a pseudo-likelihood that circumvents the computational issue. To construct a sub-sample of data that can be used to estimate the parameters $\boldsymbol{\psi}$ via pseudo-likelihood estimation, we first need to consider what these parameters represent. These parameters control characteristics of the dependence functions described in Section 5.2.2, which are functions of either distance to the conditioning site $h_{i,j}$ or pairwise distances $h_{j,k}$ for $s_i, s_j, s_k \in \mathbf{s}$. Thus, we construct our sub-sample by drawing triples of sites $(s_i, s_j, s_k) \in \mathbf{s}$; there are $d(d-1)(d-2)/2$ possible triples of sites, and our sub-sample must adequately represent the distribution of the distances in the full data. However, not all distances can be represented in the sub-sample. If we pick triples randomly, then we are more likely to pick sites with larger pairwise distances. This has two disadvantages: pairs with larger pairwise distances are not informative about the dependence parameters, as at these distances the process may exhibit near-independence, and we are also unlikely to learn about the dependence for small distances, as less pairs with smaller pairwise distances are sampled. To ensure this is not the case, each triple (s_i, s_j, s_k) is chosen so that the distances $h_{i,j}, h_{i,k}$ and $h_{j,k}$ do not exceed a specified threshold. This is a natural extension of the approach of Huser and Davison (2013) who suggest using only pairs of locations that are within some low distance $h_{max} > 0$ of each other - we instead impose this constraint on triples.

To sub-sample $d_s \ll d(d-1)(d-2)/2$ triples of locations for inference, we begin by uniformly sampling a conditioning site $s_i \in (s_1, \ldots, s_d)$. A pair of sites are then

drawn randomly from the set $\{s_j, s_k \in \mathbf{s} \setminus s_i, j < k : \max\{h_{i,j}, h_{i,k}\} < h_{max}\}$ without replacement, and the process is repeated. In sampling in this way, only sites within distance h_{max} of the conditioning site are used for inference. That is, we estimate the spatial functions of the dependence parameters for $h_{i,j} < h_{max}$ and $h_{i,k} < h_{max}$, and so $h_{j,k} < 2h_{max}$ only, and then extrapolate to larger distances. There is a trade-off involved in choosing the value of h_{max}: if too low, then extrapolations to larger distances are likely to be poor; if too high, fit at small distances is compromised. In Section 5.4.3, we describe a heuristic technique for choosing h_{max} and we find this works well in practice. The number of triples d_s is chosen to be as large as possible, whilst pseudo-likelihood estimation remains computationally feasible.

5.3.3 Scaled bootstrap sampling distributions

Our aim is to estimate the sampling distribution of $\hat{\boldsymbol{\psi}}_{d_s}$, the parameter estimates of the model fit using the d_s triples sampled under the regime described in Section 5.3.2. However, to derive this using a bootstrapped sample of $\hat{\boldsymbol{\psi}}_{d_s}$, denoted $\hat{\boldsymbol{\psi}}^*_{d_s}$, may be computationally infeasible; if d_s is large, as it is computationally expensive to get the required number of replicated values. We detail a bootstrapping regime that uses less data, and hence is computationally feasible, but it is still reliable.

We denote $\hat{\boldsymbol{\psi}}_m$ of length Q as parameter estimates using a sample of $m \in \mathbb{N}$ stratified sampled triples with $m \leq d_s$. We further denote $\hat{\boldsymbol{\psi}}^*_m$ as a bootstrap samples of parameter estimates achieved via maximum pseudo-likelihood estimation of (5.3.2) with m triples, and V_m as the $Q \times Q$ variance matrix of the estimator of $\hat{\boldsymbol{\psi}}_m$ which are calculated using $\hat{\boldsymbol{\psi}}^*_m$. We take a scaling approach to approximate the sampling distribution of $\hat{\boldsymbol{\psi}}_{d_s}$. This is achieved by creating a bootstrap sample of $\hat{\boldsymbol{\psi}}^*_{d_s}$; although we cannot compute this directly, we can compute a bootstrap sample $\hat{\boldsymbol{\psi}}^*_{d_s/w}$ for $w > 1$, such that $d_s/w \in \mathbb{N}$; that is, a sample created using $d_s/w < d_s$ triples where the replicates of $\hat{\boldsymbol{\psi}}^*_{d_s/w}$ can be estimated in a feasible time-frame. We then apply a linear

transformation to $\hat{\boldsymbol{\psi}}^*_{d_s/w}$ to create an approximate sample of $\hat{\boldsymbol{\psi}}^*_{d_s}$. To illustrate this, let $\tilde{\boldsymbol{\psi}}_{d_s/w}$ be the component-wise mean of the $\hat{\boldsymbol{\psi}}^*_{d_s/w}$ replicates. Then each replicate is transformed to give

$$\hat{\boldsymbol{\psi}}^*_{d_s} = \hat{\boldsymbol{\psi}}_{d_s} + V_{d_s}^{-1/2} V_{d_s/w}^{1/2} \left(\hat{\boldsymbol{\psi}}^*_{d_s/w} - \tilde{\boldsymbol{\psi}}_{d_s/w} \right) = \hat{\boldsymbol{\psi}}_{d_s} + \lambda^{-1/2} \left(\hat{\boldsymbol{\psi}}^*_{d_s/w} - \tilde{\boldsymbol{\psi}}_{d_s/w} \right), \quad (5.3.3)$$

where $\lambda > 0$ is to be defined in (5.3.5) and V_{d_s} is to be specified below. This ensures that the bootstrap sample $\hat{\boldsymbol{\psi}}^*_{d_s}$ has expectation $\hat{\boldsymbol{\psi}}_{d_s}$ and variance V_{d_s}. The sampling distribution of $\hat{\boldsymbol{\psi}}_{d_s}$ is then approximated empirically from $\hat{\boldsymbol{\psi}}^*_{d_s}$.

To estimate λ, we begin by estimating V_{d_s}; although direct computation is infeasible, we can estimate $V_{d_s/w}$ for $d_s > d_s/w \in \mathbb{N}$, i.e., the variance of the parameter estimates for the model fit using d_s/w triples. As long as the same sampling mechanism is used to create the sub-sampled triples of size d_s and d_s/w, i.e., that described in Section 5.3.2, it follows that $V_{d_s} \approx \lambda V_{d_s/w}$ for some $\lambda > 1$. If observations in both samples are truly independent of one another, we have that $\lambda = w$. However, this is unlikely to be the case as observations will exhibit spatial dependence. To estimate λ, we note that

$$|V_{d_s}| \approx |\lambda V_{d_s/w}| = \lambda^Q |V_{d_s/w}|, \quad (5.3.4)$$

where $|\cdot|$ denotes the matrix determinant and Q is the size of $\boldsymbol{\psi}$; this follows from the property $|\lambda M| = \lambda^Q |M|$ for constant $\lambda > 0$ and M a $Q \times Q$ matrix. Ideally, we would rewrite (5.3.4) to approximate λ; however, we cannot compute V_{d_s} directly. Instead we estimate $V_{d_s/w}$ and $V_{d_s/(2w)}$, and use (5.3.4) to estimate λ_2 such that $V_{d_s/w} \approx \lambda_2 V_{d_s/(2w)}$. It follows that

$$\lambda = \lambda_2^{\log_2(w)} = \left(\frac{|V_{d_s/w}|}{|V_{d_s/(2w)}|} \right)^{\log_2(w)/Q}, \quad (5.3.5)$$

and we can use this to estimate (5.3.3). The exponent in (5.3.5) follows as λ_2 corresponds to the variance matrix scaling factor if the sample size doubles; it would take $\log_2(w)$ repetitions of doubling d_s/w to reach d_s.

5.3.4 Simulation of an event

We now detail a technique that will allow us to draw realisations of $\{Y(s) : s \in \mathcal{S}\}$. First, we note that the model in Section 5.2.2 does not describe the dependence in all of $\{Y(s)\}$; instead, it describes

$$\begin{aligned} &\left\{Y(s) : s \in \mathcal{S}\right\} \Big| \left(\max_{s \in \mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} > v\right) \\ \equiv &\left\{F_{Y(s)}^{-1}(F_L\{X(s)\}) : s \in \mathcal{S}\right\} \Big| \left(\max_{s \in \mathcal{S}} \{X(s)\} > v\right), \end{aligned} \tag{5.3.6}$$

for $v \geq u$ with u used for fitting in Section 5.3.1. Thus, to create a realisation of $\{Y(s) : s \in \mathcal{S}\}$, we draw realisations of (5.3.6) with probability

$$\Pr\left\{\max_{s \in \mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} > v\right\}, \tag{5.3.7}$$

and otherwise draw realisations of

$$\left\{Y(s) : s \in \mathcal{S}\right\} \Big| \left(\max_{s \in \mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} < v\right). \tag{5.3.8}$$

As we do not expect realisations of (5.3.8) to contribute to the tail behaviour of $R_{\mathcal{A}}$, we simply draw realisations of (5.3.8) from the observed data. We estimate (5.3.7) empirically; although this could be inferred using the parametric model of Section 5.2. If $\mathcal{S}$ does not correspond to the set of sampling locations, then we would have to approximate (5.3.8) though some form of infilling, i.e., using the quantile regression technique (Fasiolo et al., 2020) discussed in Section 5.2.1.

We now describe a simulation technique that will allow us to draw realisations of (5.3.6). That is, the field $\{Y(s) : s \in \mathcal{S}\}$ given that an extreme value above a threshold is observed anywhere in the domain. This threshold varies with s and corresponds to the relative quantile v on the Laplace scale. Wadsworth and Tawn (2019) detail the procedure for achieving this. There are three steps: drawing conditioning sites $s_O \in \mathcal{S}$,

simulating the fields $\{Y(s) : s \in \mathcal{S}\}|(F_L^{-1}(F_{Y(s_O)}\{Y(s_O)\}) > v)$ using the fitted model described in Section 5.2, and then using importance sampling to approximate (5.3.6), see Algorithm 1, Step 2. The first step requires random sampling of conditioning sites s_O for some $s_O \in \mathcal{S}$; we do this uniformly, which provides a good first approximation of the occurrence of these sites in $\mathcal{S}$ and then improve on this via the importance sampling regime described below.

To simulate N realisations from process (5.3.6), we follow Wadsworth and Tawn (2019) and draw an initial $N' > N$ realisations of the process $\{X(s) : s \in \mathcal{S}\}$ on the Laplace scale. Then, using importance sampling, we sub-sample N realisations from $\{X(s) : s \in \mathcal{S}\}|\max_{s\in\mathcal{S}} X(s) > v$, and transform the margins of the sample to the original scale, $\{Y(s)\}$. The sub-sampling regime adds extra weight to realisations for which the conditioning site is near the boundary of the domain. This is to alleviate the edge effect caused by not using conditioning sites outside of the boundaries of $\mathcal{S}$. A discussion of a related issue is given in Section 5.3.5. We found that setting $N' \approx 5N$ was sufficient for our application, although this may be dependent on the size of $\mathcal{S}$ and value of N.

5.3.5 Inference on spatial aggregates

Using the sample of realisations of $\{Y(s) : s \in \mathcal{S}\}$ generated in Section 5.3.4, we make inference about the tail behaviour of $R_{\mathcal{A}}$ in (5.1.1) or the corresponding sum; here we focus on the latter, but a discussion of the integral is given in Section 5.5. The possible size of the aggregation region $\mathcal{A}$ in relation to the region $\mathcal{S}$ is of particular interest. Trivially, we require $A \subseteq \mathcal{S}$. However, we cannot have $\mathcal{A} = \mathcal{S}$, as if we did, the simulation algorithm will never generate an event for which the conditioning site lies outside of the boundaries of $\mathcal{S}$, but we still observe an extreme event somewhere inside $\mathcal{A}$. To avoid such edge-effects, we require the boundaries of $\mathcal{A}$ to be far enough inside the interior of $\mathcal{S}$, such that the distribution (5.3.6) does not change if the size

Algorithm 1 Simulating (5.3.6)

1. For $i = 1, \ldots, N'$ with $N' > N$:
   (a) Draw a conditioning location $s_O^{(i)}$ from $\mathcal{S}$ with uniform probability density $1/|\mathcal{S}|$.
   (b) Simulate $E^{(i)} \sim Exp(1)$ and set $x_i(s_O^{(i)}) = v + E^{(i)}$.
   (c) Simulate a field $\{z_i(s|s_O^{(i)}) : s \in \mathcal{S}\}$ from the residual process model defined in Section 5.2.2.
   (d) Set $\{x_i(s) : s \in \mathcal{S}\} = a\{x_i(s_O^{(i)}), h(s, s_O^{(i)})\} + b\{x_i(s_O^{(i)}), h(s, s_O^{(i)})\} \times \{z_i(s|s_O^{(i)}) : s \in \mathcal{S}\}$.
2. Assign each simulated field $\{x_i(s) : s \in \mathcal{S}\}$ an importance weight of
$$\left\{ \int_{\mathcal{S}} 1\{x_i(s) > v\} \mathrm{d}s \right\}^{-1},$$
for $i = 1, \ldots, N^{'}$, and sub-sample N realisations from the collection with probabilities proportional to these weights.
3. Transform each $\{x_i(s) : s \in \mathcal{S}\}$ to $\{y_i(s) : s \in \mathcal{S}\}$ using the marginal transformation (5.2.1). If $x_i(s) \leq c(s)$, set $y_i(s) = 0$, where for some $s^{'} \in \mathcal{S}$, $y_i(s^{'})$ is above its $F_L(v)$-th quantile.

of $\mathcal{S}$ increases. Informally, we require a buffer zone between the boundaries of $\mathcal{A}$ and $\mathcal{S}$ which is large enough, such that any event with conditioning site outside of $\mathcal{S}$ has negligible effect on the distribution within $\mathcal{A}$. We select the width τ of this buffer zone by using the measure $\chi_q(s_A, s_B)$ given in (5.1.2) and stationarity. We choose τ such that for any $s \in \mathcal{A}$ and $s_O \in \mathbb{R}^2 \setminus \mathcal{S}$, such that for $h(s, s_O) > \tau$, we have $\chi_q(s_O, s) < \gamma$ for small $\gamma > 0$ and for all large q; that is, we have small probability less than γ of observing a large event at s given that there is an extreme event at any site outside of $\mathcal{S}$. This measure can be evaluated empirically or by simulating from the fitted model; we take the latter approach in Section 5.4.4.

5.4 Application

5.4.1 Data

We consider data consisting of average hourly precipitation rate (mm/hour) taken from the UK convection-permitting climate model projections 2018 (UKCP18) (Lowe et al., 2018). Data are from a model which produces values over hourly intervals between the years 1980 and 2000, using the observed atmospheric conditions. The sampling locations are $(5km)^2$ grid boxes corresponding to the British National Grid from Ordnance Survey (OSGB). The spatial domain $\mathcal{S}$ of interest is East-Anglia, UK (see Figure 5.4.1) and only data sampled over land have been included, leaving 934 sampling locations. Each observation corresponds to the average over the assigned spatio-temporal grid-box. The data represent the average in each grid-box, and so a natural quantity of interest is $\bar{R}_{\mathcal{A}} := R_{\mathcal{A}}/|\mathcal{A}|$, rather than $R_{\mathcal{A}}$, but we present results on $R_{\mathcal{A}}$ as this variable must satisfy the ordering constraints discussed in Section 5.1. To remove any seasonal effect observed in the data, we use summer, i.e., July-August, observations only, leaving 43200 fields[1] We chose to take summer precipitation events as these typically exhibit higher intensity than winter events (Sharkey and Winter, 2019). We treat the centre of each grid box as a sampling location, and as the grid-boxes are non-overlapping and contiguous, we can approximate the integral $R_{\mathcal{A}}$ in (5.1.1) using a sum. We use the great-circle distance as our distance metric described in Section 5.2.2.

5.4.2 Marginal analysis

Initial analysis shows that the data consists of 8.7% hours with zero precipitation, but much of the data with non-zero values exhibits noise around zero produced by the climate model. Thus, the data less than 1×10^{-5}mm/hour were set to zero[2],

[1]Note that the UKCP18 data uses a 360 day calendar, and so each month is composed of 30 days.
[2]A level which would be recorded as zero by a rain gauge.

increasing the average number of dry hours to 83.7%. Figure 5.4.1 gives a spatial map of the estimated probability of zero precipitation $p(s)$ within a given hour; this is estimated using the logistic regression GAM detailed in Section 5.2.1. We observe some spatial variation in $p(s)$, with slightly lower estimates being found along the north-east coast.

We fit the spatial marginal model detailed in Section 5.2.1. We take $\lambda(s) = 0.995$ for all $s \in \mathcal{S}$ in (5.2.1) and the corresponding GPD threshold $q(s)$, estimated using a thin-plate spline, is illustrated in Figure 5.4.1 with $q(s)$ varying roughly over $\mathcal{S}$; larger values are found along the east coast. The GPD GAM model with spatially smooth estimate parameters is then fit to site-wise exceedances above $\hat{q}(s)$ at each site; a spatial map of the shape parameters are given in Figure 5.4.1. We take the approach of Youngman (2019) and use as many knots as is computationally feasible in the thin-plate splines, which is 300. This creates a potentially overly rough spline which may over-fit the data and not capture true physical smoothness; however, our primary interest is in the dependence structure when studying aggregates as this is the novel element of our model, so we chose this approach to ensure that the empirical marginal distributions are as well modelled as possible. We observe $\hat{\xi}(s) > 0$ for all $s \in \mathcal{S}$, and so the marginal upper tails are unbounded at each site. Q-Q plots of the marginal fits at five randomly sampled locations are presented in Figure C.3.2, all showing good fits. To evaluate the fit over all locations, we use a pooled Q-Q plot, transforming all data onto standard exponential margins using the fitted model, see Figure C.3.2. Again the fit is remarkably good, although confidence intervals are not provided due to the spatial dependence in the pooled data.

5.4.3 Dependence model

All dependence models are fitted by taking the exceedance threshold u in (5.2.2) to be the standard Laplace 98% quantile. This leaves 864 fields for fitting the ex-

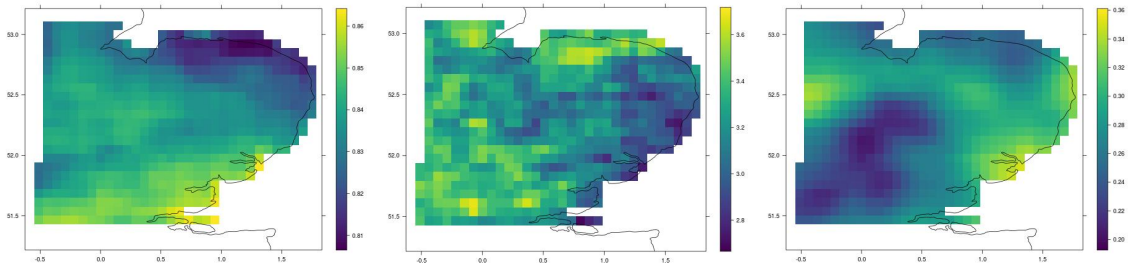

Figure 5.4.1: Spatially smoothed marginal distribution parameter estimates for East Anglia. Left: $\hat{p}(s)$, centre: $\hat{q}(s)$, right: $\hat{\xi}(s)$. $\hat{v}(s)$ is illustrated in Figure C.3.1.

tremal dependence model given an observed extreme at a single conditioning site. The empirical estimate (and 95% confidence interval) for the probability in (5.3.7) is 0.273 $(0.257, 0.290)$; this corresponds to the proportion of all observed fields used for fitting when we pool over all 934 conditioning sites. Confidence intervals for (5.3.7) were created using the approach of Politis and Romano (1994) with 1000 stationary bootstrap samples with expected block size of 48 hours. A lower threshold u was considered; however, we found that this leads to poorer model fits as the data exhibits a partial mixing of dependence structures. We believe this is due to the presence of multiple data generating processes in the climate model. Precipitation is typically generated by either high intensity events with localised spatial profiles, i.e., convective cells, or low intensity events with much large spatial profiles, i.e., frontal storms (Thomassen et al., 2020). In the absence of covariates to distinguish between these events in the data, we use a higher exceedance threshold to remove any frontal events; this is discussed further in Section 5.5.

We proceed with an initial analysis by fitting a simple version of the model of Section 5.2.2 to these data. We fit the model with two caveats: we make the temporary assumption that the residual process $\{Z(s|s_O)\}$ is independent at all distances; and evaluate a sequence of "free" pairwise parameter estimates (Wadsworth and Tawn, 2019) for the normalising functions and those functions that describe the marginal

characteristics of $\{Z(s|s_O)\}$. That is, we fit individual parameters, i.e., $\alpha_{s_i}^{(s_O)}$ etc. for $i = 1, \ldots, d$ with $s_i \neq s_O$, rather than a spatial function $\alpha\{h(s, s_O)\}$, and we do this for seven different conditioning sites s_O sampled randomly over $\mathcal{S}$. This approach can be used to assess non-stationarity in $\{X(s)\}$; if we observe clear disagreement in the parameter estimates for the different conditioning sites, then the assumption of stationarity of $\{X(s)\}$ is unlikely to be appropriate. We find no evidence for non-stationarity in the parameter estimates presented in Figure 5.4.2; while we observe some volatility in the free parameter estimates, the general patterns appear to be the same regardless of the choice of conditioning site. We use the spatial structure in the free estimates to motivate our choice for the forms of the parameter functions detailed in Section 5.2.2.

Using the sampling method described in Section 5.3.2, the full spatial fit uses $d_s = 5000$ triples of sites with each sampling location being used as a conditioning site at least once. As the estimates of α and β in Figure 5.4.2 decay quickly with increasing spatial distance, i.e., for any distance greater than $25km$, $\alpha \approx 0$ and $\beta < 0.5$; this illustrates that the underlying process $Y(s)$ exhibits fairly localised strong extremal dependence. This suggests that we should focus on modelling extremal dependence locally, as this will be the driving factor of the aggregate behaviour. A distance of $25km$ in the anisotropic setting corresponds to an approximate distance of $28km$ in the original setting, and so we set $h_{max} = 28km$. Although 5000 triples of sites represents a very small proportion of all possible triples, we observe a good model fit from Figure 5.4.2, which shows that, even at distances greater than $28km$, the fitted parametric functions for the dependence parameters correspond well to the sequences of free estimates. We further investigate the choice of h_{max} after fitting the model

Figure 5.4.2 can be used to make further inference about the underlying dependence structure of the precipitation process. For example, we find that Δ in (5.2.4) can be taken to be zero without restricting the quality of the fit and similarly we can

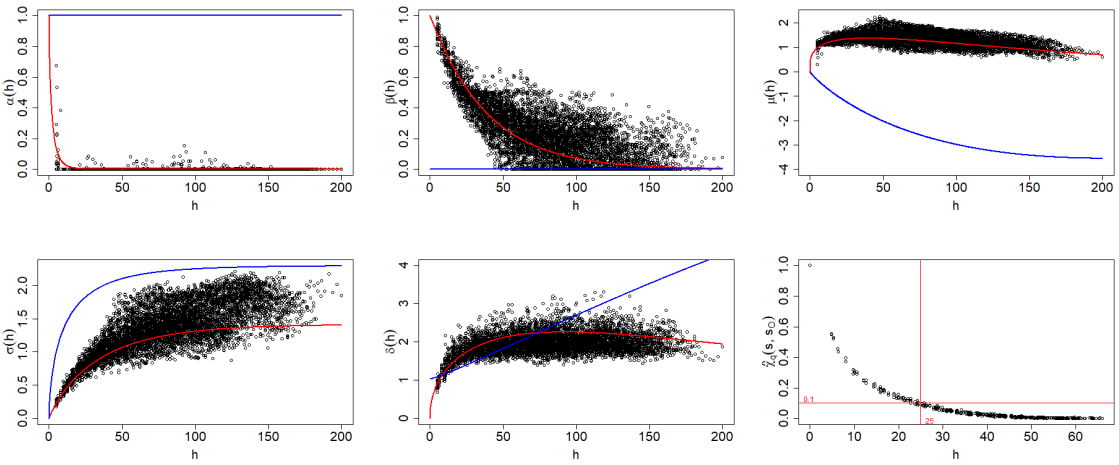

Figure 5.4.2: Estimates of parameters that determine the extremal dependence structure plotted against inter-site distance h, which is calculated under the anisotropy transformation for the full spatial model. Estimates from the free fits described in Section 5.4.3 are given by the black points, parametric spatial functions are given in red (asymptotically independent model) and blue (asymptotically dependent model). Bottom right: estimates from model for $\chi_q(s, s_O)$ in (5.1.2) with $q = 1/(24 \times 90)$. Distances (km) are given in the spatial anisotropy setting.

set $\kappa_{\beta_3} = 1$. We further note that we set $\kappa_{\delta_4} = 1$ for this application, which does not satisfy the constraint that $\kappa_{\delta_4} \leq 0$, given in (5.2.7), and hence implies that the upper-tails of the fitted residual process are too heavy; we found that this did not cause any issues in our analysis and so we chose to take $\kappa_{\delta_4} = 1$ as this provided a better model fit. The estimate $\Delta = 0$ suggests that the process is asymptotically independent at even the closest distances as asymptotic dependence requires both $\alpha(h) = 1$ and $\beta(h) = 0$ for all h, which the estimates in Figure 5.4.2 suggest is not the case; a fit imposing asymptotic dependence is discussed later. Furthermore, we found that incorporating spatial anisotropy into the dependence model improved the overall fit; stronger extremal dependence was found along an approximate $-10°$ bearing, reducing by at most 7% over different directions. Parameter estimates (and standard errors) are provided in Table C.2.1. Although not illustrated in Figure 5.4.2, ρ decays quickly with distance, with $\rho(100) \approx 0.2$. Standard errors are estimated using the bootstrap scheme described in Section 5.3.3 with $w = 20$ and the use of 250 bootstrap

samples. We estimate $\hat{\lambda}_2 = 1.042$ in (5.3.5).

To further support the choice of h_{max}, we estimate $\chi_q(s, s_O)$ in (5.1.2) for $s_O \in \mathcal{S}$ in the centre of $\mathcal{S}$, taking q corresponding to a one-year return level probability, and look to see how far away s must be for $\chi_q(s, s_O)$ to be less than γ for small $\gamma > 0$ (see Section 5.3.5). We estimate $\chi_q(s, s_O)$ by simulating 5×10^4 replications, using Algorithm 1, from the fitted model and is illustrated in Figure 5.4.2, bottom-right panel; for $\gamma = 0.1$, we find that a distance of h_{max} is sufficient and so we set $\tau = h_{max}$ in Section 5.3.5, discussed further in Section 5.4.4.

Figure 5.4.3 illustrates six extreme fields: three realisations from the model defined in (5.3.6) and three observations from the data. Fields are chosen such that the site in the centre of $\mathcal{S}$ exceeds its 99.9%-quantile but the maximum over the entire field does not exceed 30mm/hr; this is to make it easier to compare the spatial structure in the fields. Realisations from the model appear to replicate the roughness in observed events well. Furthermore, in Figure 5.4.3 we observe that replications from the model are able to exhibit some of the different physical properties of extreme precipitation. For example, the top-left panel displays a spatially flat event whilst the other two illustrate localised extreme events; multiple events in the top-middle and a single event in the top-right.

As α quickly goes to zero with distance, all extremal dependence is instead exhibited through β. This is atypical of fits of this model for other applications (Wadsworth and Tawn, 2019; Shooter et al., 2021; Simpson and Wadsworth, 2021), where the α function drives the extremal behaviour of the considered processes, e.g., temperature and sea wave heights. Having β controlling extremal dependence would suggest that the process that generates the extreme precipitation we are modelling is somewhat rough; this concurs with the observed fields containing an extreme value shown in Figure 5.4.3, and consistent with the spatial nature of strong convective rainfall. To illustrate this, we note that, for small $h = h(s, s_O)$, we have $\alpha(h) \approx 0$ and $\beta(h) \approx 1$,

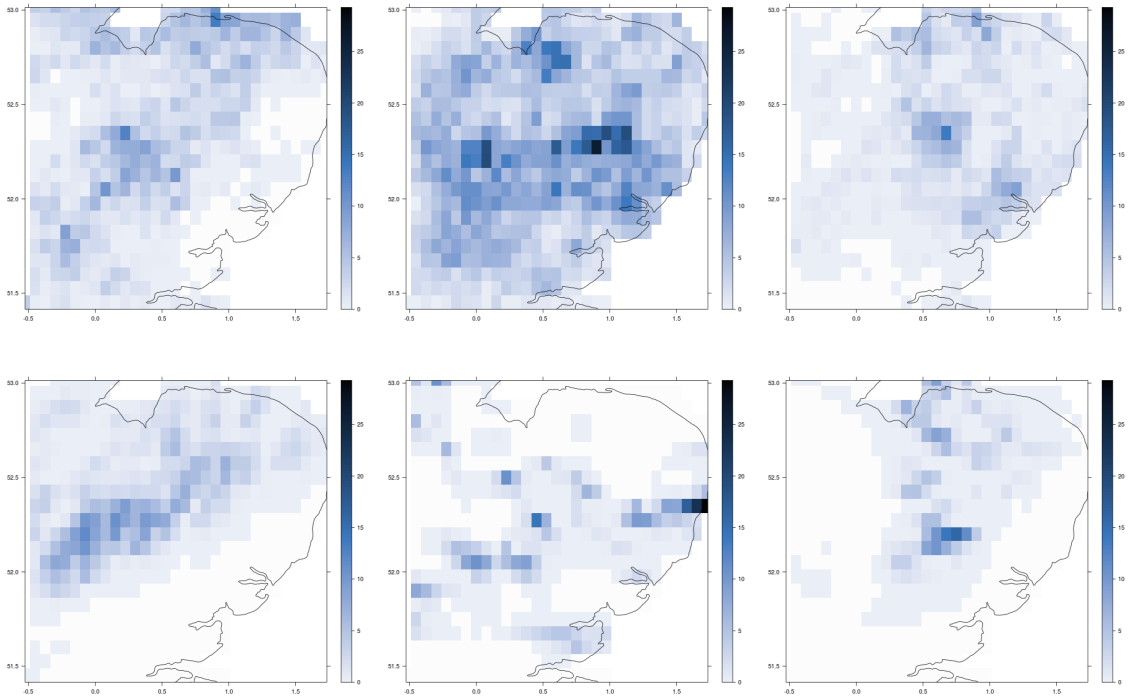

Figure 5.4.3: Extreme precipitation fields (mm/hr). Top-row: realisations from the fitted model described in (5.3.6). Bottom-row: observed fields from the data.

then $\mathbb{E}[X(s)|X(s_O) = x] \approx \mu(h)x$ and $\text{var}\,(X(s)|X(s_O) = x) \approx x^2\sigma^2(h)$, and so the largest events at s_O are the most variable. This has not been observed in other applications as the extremal dependence in these processes is typically quite smooth with $\text{var}\,(X(s)|X(s_O) = x) \approx \sigma^2(h)$ as $\beta(h) \approx 0$ when h is small. Even at the largest h, the process $\{X(s)\}$ does not exhibit independence; although α and β tend to zero the residual process does not attain standard Laplace margins with $\delta(h) = 1$.

Existing literature for approaches that rely on modelling the underlying process to make inference the extremal behaviour of spatial aggregates of precipitation typically use models that only allow for asymptotic dependence (Coles, 1993; Coles and Tawn, 1996; Buishand et al., 2008). We fit such a model to illustrate that imposing asymptotic dependence may lead to poor inference for the tails of spatial aggregates. We term this the "AD model" and the model described above as the "AI model". To specify the AD model, we fix $\alpha(h) = 1$ and $\beta(h) = 0$ for all h and we change $\sigma(h)$ in

(5.2.7) to $\sigma(h) = \kappa_{\sigma_3}\left(1 - \exp\{-(h/\kappa_{\sigma_1})^{\kappa_{\sigma_2}}\}\right)$ with $\kappa_{\sigma_1}, \kappa_{\sigma_2}, \kappa_{\sigma_3} > 0$, as we no longer require that $\sigma(h) \to \sqrt{2}$ as $(h) \to \infty$. The corresponding $\mu(\cdot)$ and $\delta(\cdot)$ functional forms remain the same and the spatial anisotropy setting described in (5.2.11) is still used. To fully capture the behaviour of $\mu(\cdot)$, we found we had to take $h_{max} = 75km$. The estimated spatial functions for the AD model are illustrated in Figure 5.4.2. With α and β fixed, we observe that the other parameters are forced to compensate for this misspecification. For example, we observe a strictly negative μ function; this is to compensate for fixing the α value too large for the data. Given this, we re-estimated the free parameters with $\alpha = 1$ and $\beta = 0$ fixed and observed good agreement between the spatial functions and these new estimates. However, this does not imply that the model as a whole fits well, this is emphasised in Section 5.4.4 where spatial aggregates of simulated fields $\{Y(s)\}$ are studied.

5.4.4 Diagnostics and tails of spatial aggregates

Q-Q plots, presented in Figure 5.4.4, assess how well the tails of the simulated distributions compare against the tails of the empirical distribution of the spatial averages. Confidence intervals given for the simulated quantiles are derived using the bootstrap; for each of the 250 bootstrap parameter estimates discussed in Section 5.3.3, we draw 5×10^5 realisations of $\{Y(s) : s \in \mathcal{S}\}$ using the regime described in Section 5.3.4. Then $\bar{R}_{\mathcal{A}}$ is calculated for each sample and for each region $\mathcal{A}$; per the discussion in Section 5.4.3, each $\mathcal{A}$ is at least $\tau = 28km$ away from the boundaries of $\mathcal{S}$. Figure 5.4.4 illustrates generally good fits for the tails of $\bar{R}_{\mathcal{A}}$ with nested regions $\mathcal{A}$. The AI model appears to slightly underestimate the true magnitude of the largest aggregates for the largest regions; while this may suggest that the model is not capturing dependence at further distances from the conditioning site, Figure 5.4.2 suggests that the model fits well even at the furthest distances. This leads us to suspect that there is a mixture of events present in the data when we consider large spatial regions for aggregation,

and that the model is not flexible enough to capture these mixtures, see Section 5.5. To illustrate the benefits of using our approach, Figure 5.4.4 also illustrates the same diagnostics for the AD model described in Section 5.4.3 for the smallest and largest regions. The AD model provides much poorer fits than the AI models, as it always overestimates the quantiles; this suggests that the AD model overestimates the dependence within the original process even for the smallest aggregation regions. A similar plot for non-overlapping regions is illustrated in Figure C.3.5; here we observe some underestimation in the largest estimated return levels for the regions closest to the east coast, which is a possible indication of non-stationarity along this coast.

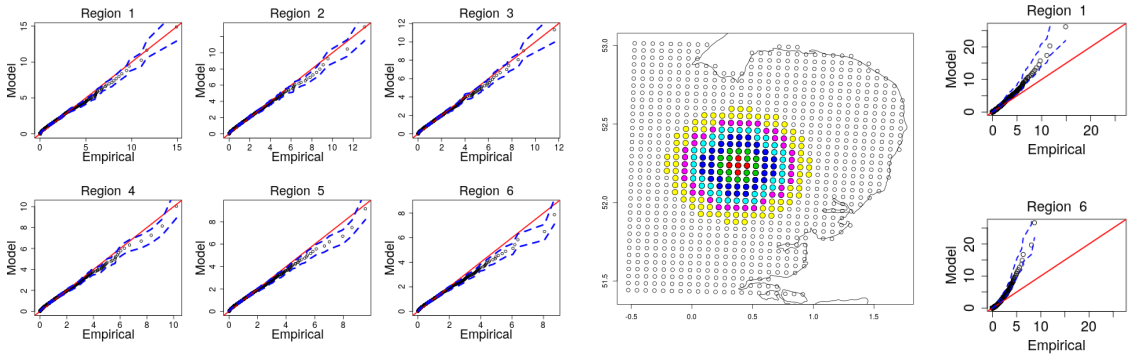

Figure 5.4.4: Q-Q plots for model, and empirical, $\bar{R}_{\mathcal{A}}$ of regions of increasing size. Left: AI model, right: AD model. Probabilities range from 0.7 to a value corresponding to the 20 year return level. 95% confidence intervals are given by the blue dashed lines. Q-Q plots for all six regions for the AD model are given in Figure C.3.4. Centre: aggregate regions $\mathcal{A}$ with corresponding areas $(125, 525, 1425, 2425, 3350, 5425) - km^2$. Regions 1-6 are coloured red, green, blue, cyan, purple, yellow; regions include both the coloured and interior points.

As discussed in Section 5.1, obtaining physically consistent return level estimates of spatial aggregates is essential. We compare two methods for achieving this: (i) performing a long-run simulation from our model, deriving empirical estimates of return-levels from these replicates; (ii) fitting a GPD distribution to the observed aggregate tails and extrapolating to the desired return-level. Ideally, we want to use only the former approach as this mitigates the potential issues with using method (ii) discussed in Section 5.1; however, for computational efficiency, we perform a shorter

run for method (i) with 5×10^5 realisations and use a fitted GPD to extrapolate to the largest return-levels. Figure 5.4.5 presents estimates of return level curves for $R_{\mathcal{A}}$ over the nested regions, illustrated in Figure 5.4.4, using methods (i) and (ii), top-left and top-right panels, respectively. For each region $\mathcal{A}$, a GPD distribution is fitted to exceedances of the respective sample $R_{\mathcal{A}}$ above the 99.9% quantile and return level curves are estimated from these fits. In the top-right panel of Figure 5.4.5, we observe intersection in the return level curves estimated for the two smallest regions using method (ii). This problem does not arise using the computationally efficient version of approach (i), e.g., in the top-left panel of Figure 5.4.5, where we have $5 \times 10^5 \times 20/43200 \approx 231$, and 20, years of data for inference, respectively. Furthermore, the confidence intervals produced by method (i) are tighter, as more data are used for extrapolation; we illustrate this in the bottom-left panel of Figure 5.4.5, where we overlay return level estimates $R_{\mathcal{A}}$ using both methods, for a single region $\mathcal{A}$. Confidence intervals are derived for both methods by fitting a GPD to 250 bootstrap samples of $R_{\mathcal{A}}$: in method (i), these are the samples as described at the top of this section; for (ii), we perform a simple bootstrap of the observed data, assuming temporal independence. A higher exceedance threshold for approach (i) was considered, but we found that the difference in estimates was negligible; to support the use of the 99.9% quantile, we illustrate a pooled Q-Q plot in Figure 5.4.5, transforming exceedances from all 250 bootstrap samples onto standard Exponential margins using their respective GPD fits and observe an excellent overall fit.

A further point of interest for practitioners is inference on the joint behaviour of $(\bar{R}_{\mathcal{A}}, \bar{R}_{\mathcal{B}})$ for different regions $\mathcal{A}, \mathcal{B} \in \mathcal{S}$. We investigate this joint behaviour for the different aggregate regions given in Figure 5.4.4. Figure 5.4.6 illustrates realisations of pairwise $(\bar{R}_{\mathcal{A}}, \bar{R}_{\mathcal{B}})$ for both model and empirical estimates with nested regions $\mathcal{A}, \mathcal{B}$, showing that the model captures the joint distributions well; a similar figure is given for non-overlapping regions in Figure C.3.6, in which we observe that extreme events

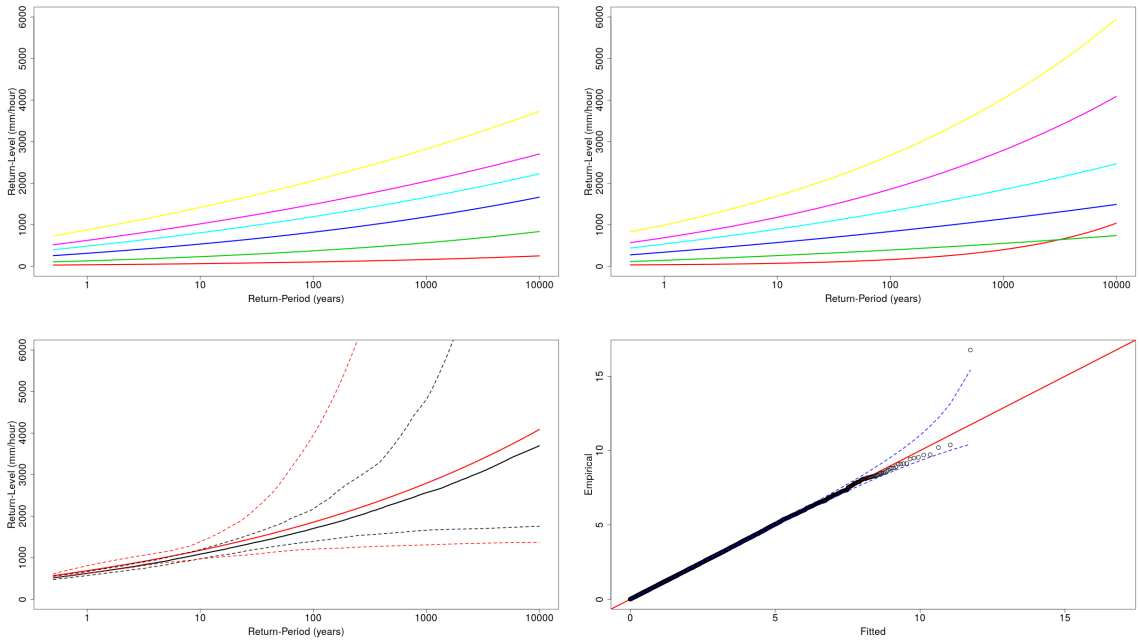

Figure 5.4.5: Top: Estimated return level curves of $R_{\mathcal{A}}$ using the model (left) and observations (right). Colours correspond to the regions illustrated in Figure 5.4.4. Bottom-left: return level estimates for Region 5 in Figure 5.4.4 using methods (i) and (ii) in black and red, respectively. 95% confidence intervals for the methods are given by the coloured dashed lines. Bottom-right: Q-Q plot for pooled GPD fit for approach (i), over all 250 bootstrap samples, on standard Exponential margins. 95% tolerance bounds are given by the dashed lines.

do not typically occur together. This suggests that the extremal behaviour of the aggregates is driven by spatially-localised events. Further evidence for this can be found in Figure 5.4.6 for aggregates over nested regions, as we observe weakening extremal dependence between aggregates over the smallest, and increasingly larger, regions.

5.5 Discussion

We have presented extensions of the Heffernan and Tawn (2004) and Wadsworth and Tawn (2019) models for modelling the extremal dependence for precipitation data. As illustrated in Section 5.2.2, this model provides flexibility over existing models for extreme precipitation as it can capture asymptotic independence. Simulating from

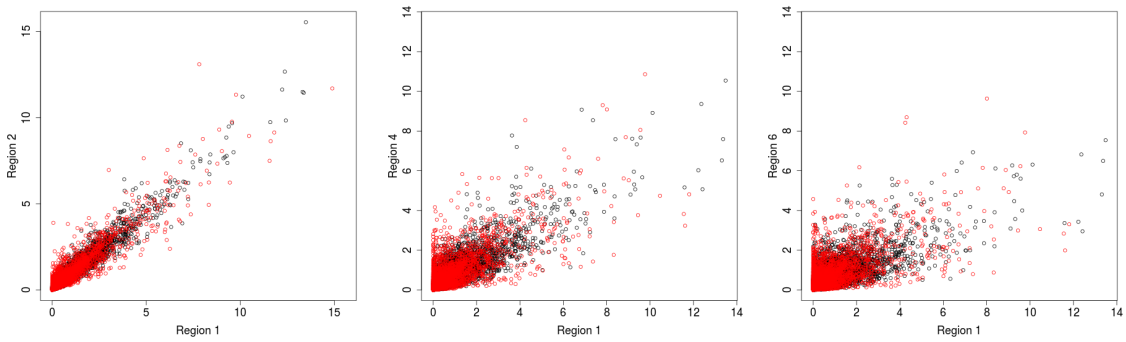

Figure 5.4.6: Plots of 2×10^4 realisations of pairwise $(\bar{R}_{\mathcal{A}}, \bar{R}_{\mathcal{B}})$ for nested regions $\mathcal{A}, \mathcal{B}$, illustrated in Figure 5.4.4. Black points are model estimates, red points are from the data. The regions $\mathcal{A}, \mathcal{B}$ are labelled on the respective panels.

this model is simple, and replications can be used to make reliable inference about the tail behaviour of spatial aggregates of the underlying process once issues linked to edge effects are addressed. This approach circumvents an issue that is common with independent inference on the tails of spatial aggregates over different regions, namely that they run the risk of making inference that is inconsistent with the physical properties of the process.

A particular drawback of our our approach is that inference using the full likelihood is computationally infeasible. To overcome this issue, we proposed methods for model fitting and assessing parameter uncertainty that are based on a pseudo-likelihood approach which requires specification of a hyper-parameter h_{max} and a novel scaling approach respectively. We found that these methods worked well in our application, as we were able to choose a suitable h_{max} quite low for which the model fits well in a reasonable time-frame, see Figure 5.4.2. This is because our data exhibits fairly localised extreme events; in applications where this is not the case, a larger h_{max} will be required which could potentially lead to more samples being required for fitting.

Data used in Section 5.4 are from a climate model, which means that sampling locations are comprised of non-overlapping grid-boxes, rather than point locations. In our application, we take $R_{\mathcal{A}}$ to be the corresponding summations, rather than the

integrals defined in (5.1.1); however, this is not to say that our approach cannot be used if we require inference on the tail of an integral. We have detailed a fully spatial model for both the dependence and marginal behaviour of $\{Y(s)\}$, and so it is possible to create a sample of $\{Y(s) : s \in \mathcal{S}\}$ where $\mathcal{S}$ is not necessarily the sampling locations. We can then approximate $R_{\mathcal{A}}$ by specifying $\mathcal{S}$ as a fine-grid and taking the sum of $\{Y(s) : s \in \mathcal{A}\}$.

When considering spatial aggregates over the largest regions $\mathcal{A}$, we find that our approach slightly underestimates the largest events, see Figure 5.4.4. Whilst this may be caused by boundary effects, it could also be caused by a complexity of the data generating process that is not captured by the model. As the size of $\mathcal{A}$ increases, it becomes less likely that the tail behaviour of $R_{\mathcal{A}}$ is driven by a single type of extreme event. There are two possible areas of complexity that are missed for regions that are sufficiently large: (i) multiple occurrences of localised high-intensity convective events (Schroeer et al., 2018), whereas we modelled single occurrences in Section 5.4; (ii) events consisting of a mixture of localised high-intensity convective and widespread low-intensity non-convective, events. Our data appears to exhibit these; recall in Section 5.4.3 we remarked that we considered a lower threshold u in (5.2.2) for modelling, but we found that this was not feasible as the data exhibits mixtures of dependence. A higher threshold had to be specified to remove observed fields that exhibited long-range spatial dependence to improve model fitting. Further improvements can be made to inference on the tails of $R_{\mathcal{A}}$ by modelling frontal events. To illustrate this, consider that we model $R_{\mathcal{A}}|(\max_{s\in\mathcal{S}} X(s) > v)$, i.e., $R_{\mathcal{A}}$ given an extreme event somewhere in $\mathcal{S}$, and undo said conditioning using the data. We do not model $R_{\mathcal{A}}$ given that there is no extreme event anywhere in $\mathcal{S}$, i.e., such caused by a frontal event. As the size of $\mathcal{A}$ grows, we will increasingly find that these events will drive the extremal behaviour of $R_{\mathcal{A}}$; this could be further explanation behind the underestimation in Figure 5.4.4, and so should be incorporated into the model. We

also considered another measure of the extremal dependence in Appendix C.2, which suggests improvement may be possible using mixture modelling. A possible approach to this problem is to incorporate covariates on precipitation field type into the model.

6

Mixture modelling of extreme precipitation

6.1 Motivation

In Chapter 5, we detailed a methodology for modelling the upper-tail behaviour of spatial aggregates of precipitation. We proposed a model for high-resolution extreme precipitation and fit this model to data, and we found that the extremal behaviour of the underlying process was driven primarily by spatially localised, high-intensity events, which we believed to be caused by convective storm cells (Schroeer et al., 2018). From our fitted model, we simulated events and used these realisations to conduct inference on the extremal behaviour of variables corresponding to averages over spatial regions. We illustrated that this approach was particularly effective for modelling aggregates over small regions (see Figure 5.4.4), but we found that our approach began to underestimate the extremes of the aggregate as the regions grew sufficiently large. For larger spatial regions, we hypothesise that the extremal behaviour of spatial aggregates is not solely driven by convective precipitation events, rather it is driven by a mixture of convective and non-convective events, with the latter being of lower

intensity but having a much larger area of effect (Berg et al., 2013; Gregersen et al., 2013). To investigate this hypothesis, we adapt the modelling approach proposed in Chapter 5; we propose separate extremal dependence models for convective and non-convective precipitation and simulate events from both models to approximate the upper-tails of spatial aggregates.

Using the same notation as given in Chapter 5, our interest lies in the upper-tail behaviour of the aggregate variable

$$\bar{R}_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\int_{\mathcal{A}} Y(s)\mathrm{d}s, \tag{6.1.1}$$

where $\{Y(s) : s \in \mathcal{S}\}$ denotes a spatial process for some spatial domain $\mathcal{S} \subset \mathbb{R}^2$, and for different regions $\mathcal{A} \subset \mathcal{S}$. Data available for inference are realisations $Y_{\mathbf{t}} = (Y_t(s_1), \ldots, Y_t(s_d))$ for times $t = 1, \ldots, n$ and sampling locations $\mathbf{s} = (s_1, \ldots, s_d) \subset \mathcal{S}$. In Chapter 5, we made the assumption that the marginal and dependence behaviour of $\{Y_t(s)\}$ is stationary with respect to time. Here we instead assume that times t can be partitioned into two sets, denoted $\mathcal{C}$ and $\mathcal{NC}$, which correspond to "convective" times and "non-convective" times; that is, if $t \in \mathcal{C}$, then the observed field $\{Y_t(s) : s \in \mathcal{S}\}$ is caused by a convective event, and similarly for non-convective events. We assume that there are two processes $\{Y_t^{\mathcal{C}}(s) : t \in \mathcal{C}\}$ and $\{Y_t^{\mathcal{NC}}(s) : t \in \mathcal{NC}\}$, which denote a convective, and non-convective, process respectively and that these processes have different marginal and dependence structure. For each process, both the marginal behaviour and dependence structures are stationary with respect to the corresponding time sets; that is, the process $\{Y_k^{\mathcal{C}}(s)\}$ is equivalent to $\{Y_l^{\mathcal{C}}(s)\}$ for all $l, k \in \mathcal{C}$, and similarly for the non-convective process and $\mathcal{NC}$. Defining $\bar{R}_{\mathcal{A}}^{\mathcal{C}} = |\mathcal{A}|^{-1}\int_{\mathcal{A}} Y^{\mathcal{C}}(s)\mathrm{d}s$ and $\bar{R}_{\mathcal{A}}^{\mathcal{NC}} = |\mathcal{A}|^{-1}\int_{\mathcal{A}} Y^{\mathcal{NC}}(s)\mathrm{d}s$, we have that

$$\Pr\{\bar{R}_{\mathcal{A}} \leq r\} = \Pr\{\bar{R}_{\mathcal{A}}^{\mathcal{C}} \leq r\}p_{\mathcal{C}} + \Pr\{\bar{R}_{\mathcal{A}}^{\mathcal{NC}} \leq r\}p_{\mathcal{NC}}, \tag{6.1.2}$$

where $p_{\mathcal{C}} = 1 - p_{\mathcal{NC}}$ denotes the probability that a realisation of $\bar{R}_{\mathcal{A}}$ is produced by a convective-event, i.e., $p_{\mathcal{C}} = |\mathcal{C}|/(|\mathcal{C}| + |\mathcal{NC}|)$.

To model $\bar{R}_{\mathcal{A}}^{\mathcal{C}}$ and $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$, we adapt the approach detailed in Chapter 5, and propose separate models for the two variables. Inference for the underlying processes is conducted by first estimating $\mathcal{C}$ and $\mathcal{NC}$ for the observation times, which we do using the algorithm detailed in Section 6.4.2; we then fit separate models to the two classes of data using the pseudo-likelihood framework described in Section 5.3.2. We simulate from our fitted models for $\{Y^{\mathcal{C}}(s)\}$ and $\{Y^{\mathcal{NC}}(s)\}$ using a framework that extends the procedure described in Section 5.3.4, and then combine realisations from both models to estimate the upper-tail behaviour of $\bar{R}_{\mathcal{A}}$; details of both of these techniques are provided in Section 6.3.1. To illustrate the efficacy of this approach to that of Chapter 5, we also fit the single process model described therein and compare inference on aggregates using both approaches; that is, one where we model a single underlying process, which we denote $\{Y^{*}(s)\}$, and one where we model a mixture process with two components, $\{Y^{\mathcal{C}}(s)\}$ and $\{Y^{\mathcal{NC}}(s)\}$.

6.2 Modelling convective and non-convective extreme precipitation

6.2.1 Marginal model

For each of the three processes $\{Y^{\mathcal{C}}(s)\}$, $\{Y^{\mathcal{NC}}(s)\}$ and $\{Y^{*}(s)\}$, we propose a separate marginal model; this is similar to the model described in Section 5.2.1, albeit with two differences. Firstly, we include elevation as a smooth predictor in the components of the marginal distributions, which we denote $\varepsilon \in \mathbb{R}_{+}$. The distribution function for

a general $Y(s), s \in \mathcal{S}$ is

$$F_{Y(s)}(y) = \begin{cases} p(s,\varepsilon), & \text{if } y = 0, \\ \frac{1-\lambda(s,\varepsilon)-p(s,\varepsilon)}{F_{Y_+(s)}(q(s,\varepsilon))} F_{Y_+(s)}(y) + p(s,\varepsilon), & \text{if } 0 < y \leq q(s,\varepsilon), \\ 1 - \lambda(s,\varepsilon)\left[1 + \frac{\xi(y-q(s,\varepsilon))}{\upsilon(s,\varepsilon)}\right]_+^{-1/\xi}, & \text{if } y > q(s,\varepsilon), \end{cases} \tag{6.2.1}$$

for all $s \in \mathcal{S}$, and where $\upsilon(s,\varepsilon) > 0$ and $p(s,\varepsilon) \geq 0$, $\lambda(s,\varepsilon) > 0$ and $p(s,\varepsilon) + \lambda(s,\varepsilon) < 1$; here $F_{Y_+(s)}(y)$ denotes the distribution function of strictly positive values of $Y(s)$, which we estimate this using the empirical estimator. Note that (6.2.1) further differs from (5.2.1) as here we fix $\xi \in \mathbb{R}$ for all $s \in \mathcal{S}$; this is a common approach taken when modelling spatial characteristics of extreme rainfall, see Thibaud et al. (2013); Zheng et al. (2015); Saunders et al. (2017) and Brown (2018). We fix $\lambda(s,\varepsilon) = \lambda$ for all $(s,\varepsilon) \in \mathcal{S} \times \mathbb{R}_+$ and estimate $q(s,\varepsilon)$ for this λ. In Section 5.2.1, we describe a technique for estimating $q(s,\varepsilon)$ whereby we fit a thin-plate spline through point-wise estimates of $q(s,\varepsilon)$ for each $s \in (s_1, \ldots, s_d)$; however, this approach does not account for uncertainty associated with the quantile estimator and so here we used additive quantile regression (Fasiolo et al., 2020) instead. This technique is particularly computationally expensive and so we use a subset of sites for estimating $q(s,\varepsilon)$.

Each of the parameters is represented through a basis of thin-plate splines with separate bases being used for location s and for elevation ε. Recall that in Section 5.2.1 we advocate the use of splines with as many knots as is computationally feasible, i.e., overly rough splines. Here, we instead use as few knots as possible, i.e., four, to ensure that the splines are smooth. This is to avoid over-fitting and makes the marginal fits more interpretable; furthermore, we fully expect the marginal parameters to be functions of elevation, see Coles and Tawn (1996); Cooley et al. (2007); Cooley and Sain (2010).

6.2.2 Dependence model

We use (6.2.1) to perform site-wise standardisation of $\{Y^{\mathcal{C}}(s)\}$, $\{Y^{\mathcal{NC}}(s)\}$ and $\{Y^{*}(s)\}$ to standard Laplace margins; we denote the standardised processes as $\{X^{\mathcal{C}}(s)\}$, $\{X^{\mathcal{NC}}(s)\}$ and $\{X^{*}(s)\}$, respectively. Extremal dependence in $\{X^{\mathcal{C}}(s)\}$ and $\{X^{*}(s)\}$ is characterised using the exact same model described in Section 5.2.2. For $\{X^{\mathcal{NC}}(s)\}$, we use a similar model but with different parametric forms for β and σ to those given by (5.2.5) and (5.2.7), respectively; we denote these new forms by $\beta_{\mathcal{NC}}$ and $\sigma_{\mathcal{NC}}$.

For β, we adopt the approach of Shooter et al. (2021) and let

$$\beta_{\mathcal{NC}}(h) = \frac{\kappa_{\beta_1} h^{\kappa_{\beta_2}} \exp(-h/\kappa_{\beta_3})}{\max_{h>0}\{h^{\kappa_{\beta_2}} \exp(-h/\kappa_{\beta_3})\}}, \quad (\kappa_{\beta_1} \in [0,1], \kappa_{\beta_2} > 0, \kappa_{\beta_3} > 0). \tag{6.2.2}$$

Note that $\beta_{\mathcal{NC}}$ satisfies similar properties to β, e.g., $0 \leq \beta_{\mathcal{NC}}(h) \leq 1$ for all $h \geq 0$ and $\beta_{\mathcal{NC}}(h) \to 0$ as $h \to \infty$. However, they differ in their values at the conditioning site, as $\beta_{\mathcal{NC}}(0) = 0$ whereas $\beta(0) = 1$. Recall from the discussion in Section 5.4.3 that β is an apt choice for the process $Y^{\mathcal{C}}$ as it is exhibits spatial roughness and this can be accommodated into the model by letting $\beta(0) = 1$. This property is not required for $Y^{\mathcal{NC}}$ as we expect this process to be much smoother; to support this claim, we provide example observations of $Y^{\mathcal{C}}$ and $Y^{\mathcal{NC}}$ in Figure 6.4.2.

For $\sigma_{\mathcal{NC}}$, we introduce another parameter into σ and let

$$\sigma_{\mathcal{NC}}(h) = \kappa_{\sigma_3}\left(1 - \exp\{-(h/\kappa_{\sigma_1})^{\kappa_{\sigma_2}}\}\right), \quad (\kappa_{\sigma_1} > 0, \kappa_{\sigma_2} > 0, \kappa_{\sigma_3} > 0), \tag{6.2.3}$$

and so equivalence with σ is achieved by setting $\kappa_{\sigma_3} = \sqrt{2}$. Use of $\sigma_{\mathcal{NC}}$ implies that, if $\kappa_{\sigma_3} \neq \sqrt{2}$, then $\{X^{\mathcal{NC}}(s)\}$ does not satisfy the desirable long-range independence property described in Section 5.2.2, i.e., we do not have independence between $\{X^{\mathcal{NC}}(s_O)\}$ and $\{X^{\mathcal{NC}}(s^*)\}$ as $\|s^* - s_O\| \to \infty$ for $s_O, s^* \in \mathcal{S}$. However, this is not an issue if the domain of interest $\mathcal{S}$ is relatively small, where independence might not be a reasonable assumption at even the largest distances; non-convective events can

have a very large spatial extent, for example, Houze Jr (1997) note that stratiform precipitation can cover a contiguous area of up to $1000km$ in length. We find that for our $\mathcal{S}$, there is little evidence of independence in $X^{\mathcal{NC}}$ at even the largest distances, see Section 6.4.4.

Inference for the three extremal dependence models is conducted using the pseudo-likelihood procedure described in Section 5.3; however, we note that for the stratified sampling regime in Section 5.3.2, we require different values of h_{max} for the three processes that we consider.

6.3 Simulation

6.3.1 Simulating events

To simulate from the three processes $\{Y^{\mathcal{C}}(s)\}$, $\{Y^{\mathcal{NC}}(s)\}$ and $\{Y^{*}(s)\}$, we adapt the procedure detailed in Section 5.3.4 and Algorithm 1; throughout we consider a general process $\{Y(s)\}$, but this approach can be applied to any of the three processes given above. We first note that the techniques described in Section 5.3.4 suffer from the following two limitations: firstly, if $\mathcal{S}$ corresponds to the set of sampling locations $\mathbf{s} = (s_1, \ldots, s_d)$ and d is large, it may be computationally infeasible to simulate a field $\{Y(s)\}$ at all $s \in \mathcal{S}$. Secondly, in Section 5.3.5 we discussed edge effects that occur when considering aggregates over regions $\mathcal{A} \subset \mathcal{S}$, as the simulation procedure never generates events for which the conditioning sites lies outside of the boundaries of $\mathcal{S}$ but an extreme event is still observed within $\mathcal{A}$. To address this issue, we proposed a heuristic for choosing the position of $\mathcal{A}$; we required that the boundaries of $\mathcal{A}$ were at least $\tau > 0$ distance within the boundaries of $\mathcal{S}$. Here τ was chosen so that for any $s \in \mathcal{A}$ and $s_O \in \mathbb{R}^2 \setminus \mathcal{S}$ such that $h(s, s_O) > \tau$, we had $\chi_q(s_O, s) < \gamma$ for small $\gamma > 0$ and for large q, with $\chi_q(s_O, s)$ defined in (5.1.2). In some case, picking a suitable τ may not be feasible. This may occur if the process $\{Y(s)\}$ exhibits particularly strong

extremal dependence even at the largest distances; in our application, we find this to be the case for $\{Y^{\mathcal{NC}}(s)\}$.

To address the two issues above, first consider a region over which we wish to aggregate, denoted $\mathcal{A} \subset \mathcal{S}$, where $\mathcal{S}$ is our entire spatial domain. We now define two sets $\mathcal{D}$ and $\mathcal{O}$ which we use for creating samples of the aggregate $R_{\mathcal{A}}$. Here $\mathcal{D}$ is the set of locations at which we simulate fields, i.e., $\{Y(s) : s \in \mathcal{D}\}$, that satisfies $\mathcal{A} \subset \mathcal{D} \subseteq \mathcal{S}$; recall that if $|\mathcal{S}|$ is large, it may be computationally infeasible to simulate fields $\{Y(s) : s \in \mathcal{S}\}$, and so in this case we would have $\mathcal{D} \subset \mathcal{S}$. Each extreme event we simulate requires a conditioning site s_O; we denote the set of all possible conditioning sites used for simulation by $\mathcal{O}$, which satisfies $\mathcal{D} \subseteq \mathcal{O}$, but not necessarily $\mathcal{O} \subseteq \mathcal{S}$. Illustrations of $\mathcal{A}$, $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{O}$ for our application are presented in Figure 6.3.1 and a heuristic for choosing both $\mathcal{D}$ and $\mathcal{O}$ is given in Section 6.3.2.

To simulate $\{Y(s) : s \in \mathcal{D}\}$, we draw realisations of

$$\begin{aligned} &\left\{Y(s) : s \in \mathcal{D}\right\} \Big| \left(\max_{s\in\mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} > v\right) \\ \equiv &\left\{F_{Y(s)}^{-1}(F_L\{X(s)\}) : s \in \mathcal{D}\right\} \Big| \left(\max_{s\in\mathcal{S}} \{X(s)\} > v\right), \end{aligned} \tag{6.3.1}$$

with probability

$$\Pr\left\{\max_{s\in\mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} > v\right\}, \tag{6.3.2}$$

and otherwise draw realisations of

$$\left\{Y(s) : s \in \mathcal{D}\right\} \Big| \left(\max_{s\in\mathcal{S}} \left\{F_L^{-1}(F_{Y(s)}\{Y(s)\})\right\} < v\right), \tag{6.3.3}$$

from the observed $\{Y(s) : s \in \mathcal{D}\}$. To draw realisations of (6.3.1), we use Algorithm 2. Note that in Step 2, we require that

$$\left\{\int_{\mathcal{D}} 1\{x_i(s) > v\}\mathrm{d}s\right\}^{-1} \approx \left\{\int_{\mathcal{S}} 1\{x_i(s) > v\}\mathrm{d}s\right\}^{-1}, \tag{6.3.4}$$

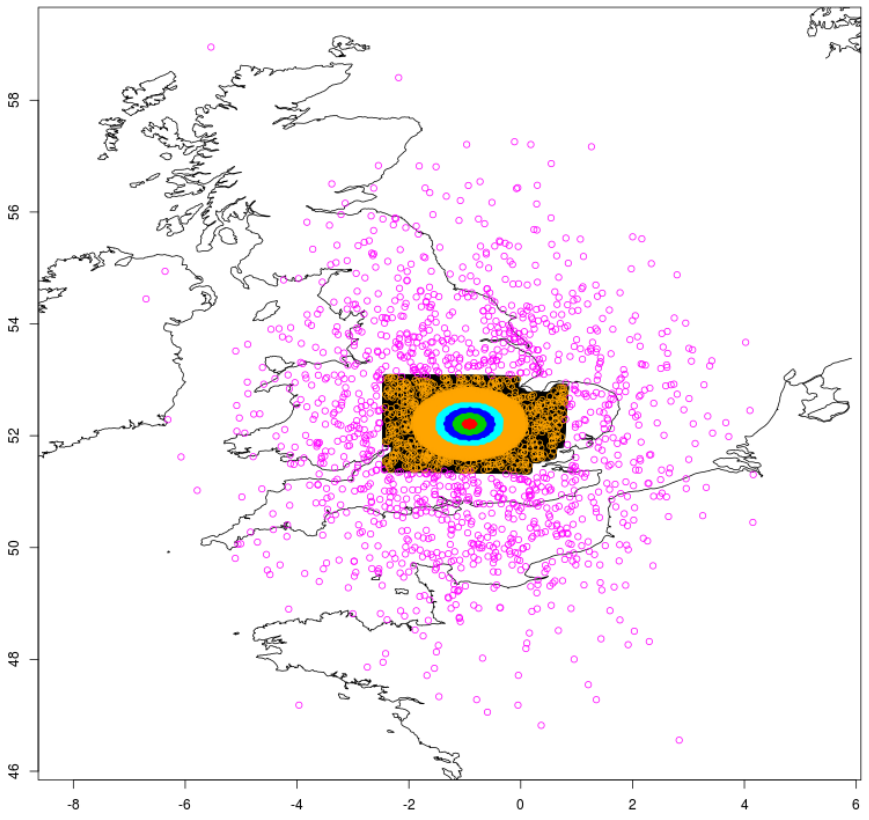

Figure 6.3.1: Regions $\mathcal{A}$, $\mathcal{D}$ and $\mathcal{O}$. Aggregate regions $\mathcal{A}$ with corresponding areas $(179, 1263, 3257, 6200) - km^2$ are coloured red, green, blue, cyan; regions include both the coloured and interior points and are numbered 1 to 4 in Figures 6.4.9 and 6.4.10. The orange and black points denote $\mathcal{D} \setminus \mathcal{A}$ and $\mathcal{S} \setminus \mathcal{D}$, respectively; the purple points outside of the boundaries of $\mathcal{S}$ denote $\mathcal{O} \setminus \mathcal{D}$. Note that $n_{\mathcal{D}} = 500$, $\tau_{\mathcal{D}} = 27.5km$, and $n_{\mathcal{O}} = 500$; these values mean that $|\mathcal{D}| = 3385$ and $|\mathcal{O}| = 4635$.

to account for the conditioning event in (6.3.1). We can improve the accuracy of this approximation by ensuring that sites in $\mathcal{D}$ are sufficiently spread out across $\mathcal{S}$.

Algorithm 2 Simulating (6.3.1)

1. For $i = 1, \ldots, N'$ with $N' > N$:

   (a) Draw a conditioning location $s_O^{(i)}$ from $\mathcal{O}$ with uniform probability density $1/|\mathcal{O}|$.

   (b) Simulate $E^{(i)} \sim \text{Exp}(1)$ and set $x_i(s_O^{(i)}) = v + E^{(i)}$.

   (c) Simulate a field $\{z_i(s|s_O^{(i)}) : s \in \mathcal{D}\}$ from the residual process model.

   (d) Set $\{x_i(s) : s \in \mathcal{D}\} = a\{x_i(s_O^{(i)}), h(s, s_O^{(i)})\} + b\{x_i(s_O^{(i)}), h(s, s_O^{(i)})\} \times \{z_i(s|s_O^{(i)}) : s \in \mathcal{D}\}$.

2. Assign each simulated field $\{x_i(s) : s \in \mathcal{D}\}$ an importance weight of

$$\begin{cases} \left\{\int_{\mathcal{D}} 1\{x_i(s) > v\}\mathrm{d}s\right\}^{-1}, & \text{if } \int_{\mathcal{D}} 1\{x_i(s) > v\}\mathrm{d}s > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{6.3.5}$$

   for $i = 1, \ldots, N'$, and sub-sample N realisations from the collection with probabilities proportional to these weights.

3. Transform each $\{x_i(s) : s \in \mathcal{D}\}$ to $\{y_i(s) : s \in \mathcal{D}\}$ using the marginal transformation (6.2.1). If $x_i(s) \leq c(s)$, set $y_i(s) = 0$, where for some $s' \in \mathcal{D}$, $y_i(s')$ is above its $F_L(v)$-th quantile.

Using the procedure outlined in Algorithm 2, we can draw realisations of $\{Y^{\mathcal{C}}(s) : s \in \mathcal{D}\}$ and $\{Y^{\mathcal{NC}}(s) : s \in \mathcal{D}\}$; we can then use these to derive samples of $\bar{R}_{\mathcal{A}}^{\mathcal{C}}$ and $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$, see (6.1.2). To then acquire a sample of $\bar{R}_{\mathcal{A}}$, we draw from the samples of $\bar{R}_{\mathcal{A}}^{\mathcal{C}}$ and $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$ with probabilities $p_{\mathcal{C}}$ and $1 - p_{\mathcal{C}}$, respectively. Estimates of $p_{\mathcal{C}}$ are derived empirically.

6.3.2 Choosing $\mathcal{D}$ and $\mathcal{O}$

We proceed by assuming that $\mathcal{S} = (s_1, \ldots, s_d)$, i.e., our spatial domain of interest is the set of sampling locations, and that these sampling locations correspond to non-overlapping grid-boxes. However, the heuristic we describe for choosing $\mathcal{D}$ and $\mathcal{O}$ can be extended to a spatially continuous setting. We begin by considering $\mathcal{D}$, which is the set of locations at which we simulate fields. We require that $\mathcal{A} \subset \mathcal{D} \subseteq \mathcal{S}$ and that sites in $\mathcal{D}$ are sufficiently spread out across $\mathcal{S}$ to ensure that the approximation in (6.3.4) is accurate. To this end, we use a two-step procedure to create $\mathcal{D}$. We first set $\mathcal{D} = \{s \in \mathcal{S} : \|s - s_{\mathcal{A}}\| \leq \tau_{\mathcal{D}}, s_{\mathcal{A}} \in \mathcal{A}\}$. That is, we take all points in $\mathcal{A}$ and those in $\mathcal{S}$ that are at most $\tau_{\mathcal{D}} \geq 0$ distance outside the boundaries of $\mathcal{A}$. We do this as we expect events with conditioning sites within this area to have a large effect on the tail behaviour of the aggregate $\bar{R}_{\mathcal{A}}$. We then sample $n_{\mathcal{D}}$ sites uniformly at random across $\mathcal{S} \setminus \mathcal{D}$ and add these to $\mathcal{D}$.

When considering the processes $Y^{\mathcal{C}}$ and Y^*, we found that setting $\mathcal{O} = \mathcal{D}$ was sufficient for simulating events. In some cases, setting $\mathcal{O} = \mathcal{D}$ may be a reasonable choice to make. In our application, we found that this was the case for the processes $Y^{\mathcal{C}}$ and Y^*, i.e., we found no further improvement in our inference on the extremes of spatial aggregates by using a set $\mathcal{O}$ that satisfies $\mathcal{D} \subset \mathcal{O}$. However, in cases where the spatial process we wish to simulate from exhibits strong extremal dependence at even the greatest observed distances between sampling locations, it may not be sufficient to set $\mathcal{O} = \mathcal{D}$; we found this to be the case for $Y^{\mathcal{NC}}$ in our application. For processes of this type, we need a technique to increase $\max_{s \in \mathcal{D}, s_O \in \mathcal{O}}\{\|s - s_O\|\}$. In lieu of obtaining data over a larger spatial domain, we can instead take $\mathcal{O}$ to be a set of "fake" sampling locations, i.e., sites at which we do not observe any data.

We begin by setting $\mathcal{O} = \mathcal{D}$ and then add sites $s^* \in \mathcal{S}^*$, where $\mathcal{S}^* = \{s^* : s^* \notin \mathcal{S}\}$ and $|\mathcal{S}^*| = n_{\mathcal{O}}$. Whilst we note that there exists more elegant solutions, we create $\mathcal{S}^*$ via a brute-force approach; we add independent Gaussian noise to the coordinates

of a site in the centre of $\mathcal{S}$ and remove any new points s^* that are located within the boundaries of $\mathcal{S}$. Note that by simulating events using conditioning sites $s_O \notin \mathcal{S}$, we are likely to simulate events at sites $s \in \mathcal{D}$ such that the distance $\|s - s_O\|$ is not observed in the data, and thus not used for fitting the dependence model. Any inferences made at these distances are from extrapolations of the fitted extremal dependence model at smaller distances, and may not be entirely accurate; however, events of these type are unlikely to affect the tail behaviour of $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$ unless the process $Y^{\mathcal{NC}}$ does not exhibit monotonically decreasing extremal dependence with distance. It is unlikely that data will exhibit non-decreasing extremal dependence in applications and moreover this property is avoided through our model specification in Section 6.2.2.

We note that in both cases $\mathcal{D}$ and $\mathcal{O}$ should be chosen as large as is computationally feasible. Furthermore, suitable values for the hyper-parameters $n_{\mathcal{D}}, \tau_{\mathcal{D}}$ and $n_{\mathcal{O}}$ can be chosen through validation techniques. For example, the simulated aggregate diagnostics (see Figures 6.4.9 and 6.4.10) can be used to validate the choice of the parameter values; if the fits look poor, one can increase the value of the parameters to improve the fits. Note that if increasing the hyper-parameter values does not improve the aggregate fits, there may instead be issues with the dependence parameter estimates themselves.

6.4 Application

6.4.1 Data

Similarly to Chapter 5, we consider data consisting of average hourly precipitation rate (mm/hour) taken from the UK convection-permitting climate model projections 2018 (UKCP18) (Lowe et al., 2018). We conduct the following analysis for the first and fourth ensemble members from these projections; however, we present only the latter analysis. Data are from a model which produces values for hourly intervals,

between the years 1980 and 2000, and use the observed atmospheric conditions in this period. The sampling locations are $(2.2km)^2$ grid boxes and the spatial domain $\mathcal{S}$ of interest is the region of the UK, approximately centred at Northampton, pictured in Figure 6.4.1; only data sampled over land have been included, leaving 7526 sampling locations. Each observation corresponds to the average over the assigned spatio-temporal grid-box and to remove any seasonal effect observed in the data, we use summer (JJA) observations only, which leaves 43200 field. The centre of each grid box is treated as a sampling location, and we use the great-circle distance as our distance metric for the dependence parameter functions. We further follow Section 5.4.2 and set all values of the data less than $1 \times 10^{-5} mm/\text{hour}$ to zero.

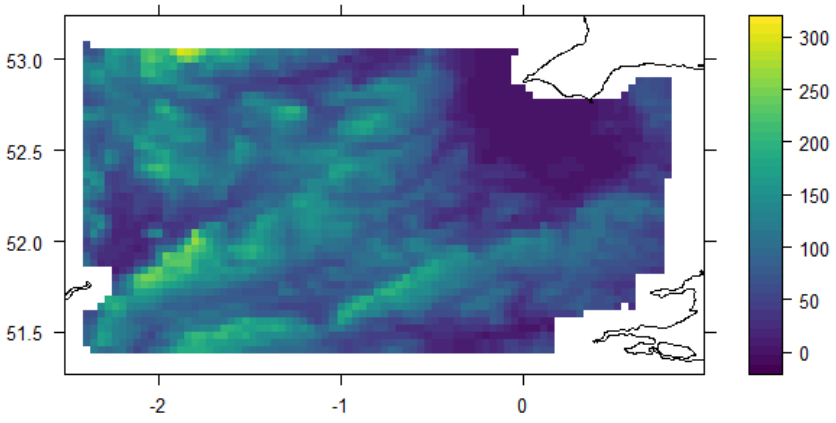

Figure 6.4.1: A map of elevation (m) for the spatial domain $\mathcal{S}$ of interest.

6.4.2 Classification

To classify observation times as convective for the $Y(s)$ field observed over $s \in \mathcal{S}$, we use the following algorithm developed at the Met Office Hadley Centre, UK (Roberts and Kendon, 2020); this procedure is intended for summer precipitation data with gridded sampling locations and has been avocated by Kendon et al. (2012) for its efficacy in identifying convective rainfall. The algorithm identifies a single value $y_t(s_i)$ at time $t \in \{1, \ldots, n\}$ and grid box s_i for $i \in \{1, \ldots, d\}$ as "convective" if the gradient of the surface of $y_t(s)$ in a neighbourhood surrounding s_i is sufficiently steep. This procedure is repeated for all $t = 1, \ldots, n$ and $i = 1, \ldots, d$; we then label the field $\{Y_t(s) : s \in \mathcal{S}\}$ at time t as convective if any $y_t(s_i)$ for $i = 1, \ldots, d$ are identified as convective.

To formally describe the algorithm, we first specify four constant hyper-parameters: $g_l > 0, g_u > g_l, p^* \in [0, 1]$ and $n_s \in \{2d^* - 1, d^* \in \mathbb{N}\}$. We define a neighbourhood $\mathcal{N}_i$ as all s_i for $i = 1, \ldots, d$ that create an $n_s \times n_s$ grid of sampling locations, with s_i at the centre of $\mathcal{N}_i$. Then we proceed with Algorithm 3.

Algorithm 3 Identify convective fields

For all $t = 1, \ldots n$:

1. For all $i = 1, \ldots, d$:
   (a) Identify the $n_s \times n_s$ neighbourhood $\mathcal{N}_i$ for s_i, defined in Section 6.4.2.
   (b) Evaluate all differences $\mathcal{G}_i = \{y_t(s_j) - y_t(s_k) : s_j, s_k \in \mathcal{N}_i\}$.
   (c) Calculate the proportion $p_{g,i} = |\{g \in \mathcal{G}_i : g \geq g_u\}|/|\{g \in \mathcal{G}_i : g \geq g_l\}|$.
   (d) If $p_{g,i} \geq p^*$ and $|\mathcal{N}_i| = n_s^2$, then $y_t(s_i)$ is labelled as convective and hence $t \in \mathcal{C}$. If $|\mathcal{N}_i| < n_s^2$, then $y_t(s_i)$ is labelled as undetermined.
2. If none of $y_t(s_i)$ for all $i = 1, \ldots, d$, are labelled as convective, then $t \in \mathcal{NC}$.

Note that for a value $y_t(s_i)$ to be labelled as either convective or not, we require that $|\mathcal{N}_i| = n_s^2$; we have removed any sampling locations where $|\mathcal{N}_i| < n_s^2$ from the analysis. In our application, we set the hyper-parameters to values provided by the

Met Office, i.e., $g_l = 0.01, g_u = 1$ and $p^* = 0.2$ and $n_s = 9$; these particular values were tuned specifically for the model that generates the data we use in our application, described in Section 6.4.1 (Roberts and Kendon, 2020).

For the data described in Section 6.4.1, we identify 13510 convective hours, i.e., $|\mathcal{C}| = 13510$, which leaves 29690 non-convective hours; from (6.1.2), we then estimate $p_{\mathcal{C}} = 13510/43200 \approx 31.3\%$. Example extreme observations of $\{Y^{\mathcal{C}}(s) : s \in \mathcal{S}\}$ and $\{Y^{\mathcal{NC}}(s) : s \in \mathcal{S}\}$ are presented in Figure 6.4.2. Illustrated fields $\{Y^{\mathcal{C}}(s) : s \in \mathcal{S}\}$ are chosen by randomly sampling from the set of fields for which site-wise maxima are observed, i.e., $\{Y_k^{\mathcal{C}}(s) : s \in \mathcal{S}, k \in \mathcal{K}\}$, such that $\mathcal{K} = \{k \in \mathcal{C} : Y_k^{\mathcal{C}}(s_i) = \max_{t \in \mathcal{C}}\{Y_t^{\mathcal{C}}(s_i)\}, i = 1, \ldots, d\}$ is the set of convective times for which the marginal maxima at any site $s \in \mathcal{S}$ is observed. A similar approach is taken for $\{Y^{\mathcal{NC}}(s) : s \in \mathcal{S}\}$. We observe that observations identified as non-convective appear smoother over space, but with much lower marginal magnitude; note the difference in the scales of Figure 6.4.2.

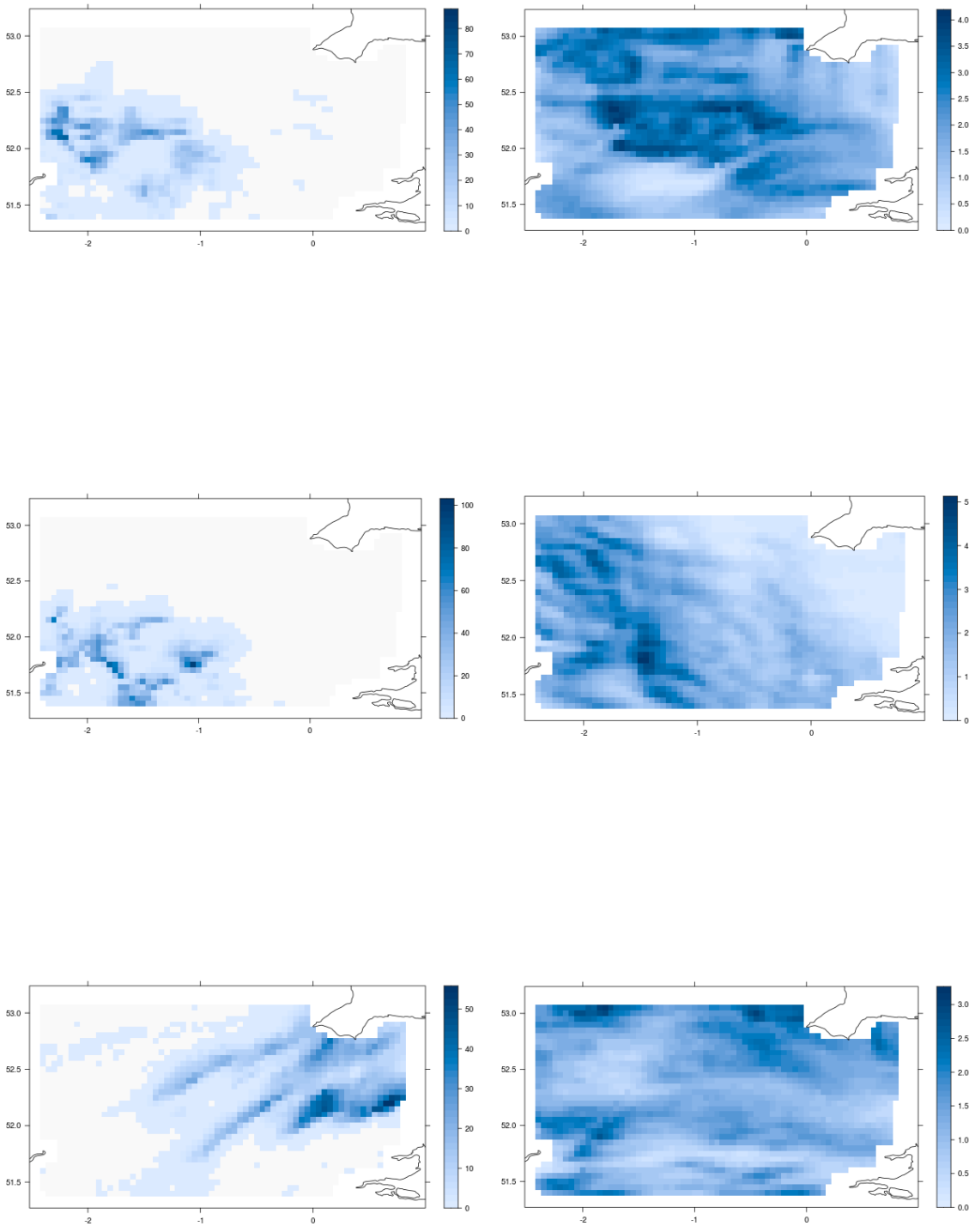

Figure 6.4.2: Observed extreme fields identified as convective (left) and non-convective (right) (mm/hr).

6.4.3 Marginal analysis

Marginal analysis is conducted by fitting the model described in Section 6.2.1 to each of the three datasets, i.e., convective, non-convective, and all, data. Figures 6.4.3, 6.4.4 and 6.4.5 give the estimates for the parameters of the marginal models for $\{Y^{\mathcal{C}}(s)\}$, $\{Y^{\mathcal{NC}}(s)\}$ and $\{Y^{*}(s)\}$, respectively; we further provide 20-year return level estimates for each of the fits. Note that different image colour scales are used across each panel and each of Figures 6.4.3-6.4.5. We set $\lambda = 0.995$ in (6.2.1) for each of the three processes and use a subset of 500 sites sampled randomly over $\mathcal{S}$ to estimate $q(s, \varepsilon)$; to estimate $p(s, \varepsilon)$, $\upsilon(s, \varepsilon)$ and ξ, we use all sampling locations.

We observe similar patterns in estimates of $p(s, \varepsilon)$ and $q(s, \varepsilon)$ for each of the three processes, namely that both are intrinsically linked to elevation, i.e., both $p(s, \varepsilon)$ and $q(s, \varepsilon)$ decrease and increase, respectively, with elevation. For the estimates of $\upsilon(s, \varepsilon)$ and the 20-year return level, we observe differences between the three fits; for $Y^{\mathcal{C}}$, we observe spatially smooth estimates of both, with larger values being found in the east of the domain (see Figure 6.4.3). In Figure 6.4.4, we observe that $\upsilon(s, \varepsilon)$ and the 20-year return level estimates for $Y^{\mathcal{NC}}$ both increase with elevation, suggesting that more intense storms form at higher altitudes. For Y^{*}, we find that elevation has much less of an effect on $\upsilon(s, \varepsilon)$ and the 20-year return level; however, we do note that for the areas with the highest elevations we observe the lowest values of $\upsilon(s, \varepsilon)$, suggesting that, for Y^{*}, we may have less intense storms at higher altitude. Comparing the 20-year return level estimates across the three figures, we observe much higher levels for $Y^{\mathcal{C}}$ and Y^{*} than for $Y^{\mathcal{NC}}$, which suggests that non-convective events are generally of much lower intensity, as expected. Further evidence for this is given by the shape parameter estimates for $Y^{\mathcal{C}}$, $Y^{\mathcal{NC}}$ and Y^{*}, which are 0.226, -0.075 and 0.287, respectively. These estimates suggest that $Y^{\mathcal{C}}(s)$ and $Y^{*}(s)$ have unbounded marginal upper-tails, but the tails of $Y^{\mathcal{NC}}(s)$ are instead bounded above at each site $s \in \mathcal{S}$.

To validate the goodness of fits of the GPD GAM models, we present Q-Q plots of the marginal fits at five randomly sampled locations; we do this for $Y^{\mathcal{C}}$, $Y^{\mathcal{NC}}$ and Y^{*} in Figures D.1.1, D.1.2 and D.1.3, respectively, which can be found in Appendix D. All figures show good individual fits for each of the processes. To evaluate the fit over all locations, we use a pooled Q-Q plot (Heffernan and Tawn, 2001), transforming all data onto standard exponential margins using the fitted model; the respective pooled Q-Q plot is given alongside the individual marginal fits for each process. Again, we observe good fits for each process. Confidence intervals for the Q-Q plots are estimated using the following bootstrap procedure: we create 250 boostrap samples of the data using the stationary bootstrap approach of Politis and Romano (1994) with expected block size of 48 hours. With $q(s, \varepsilon)$ treated as fixed across all samples, we then estimate $\upsilon(s, \varepsilon)$ and ξ for each bootstrap sample. For the pooled diagnostic plot, we apply the marginal transformation to the original data using the 250 estimated GPD parameter sets.

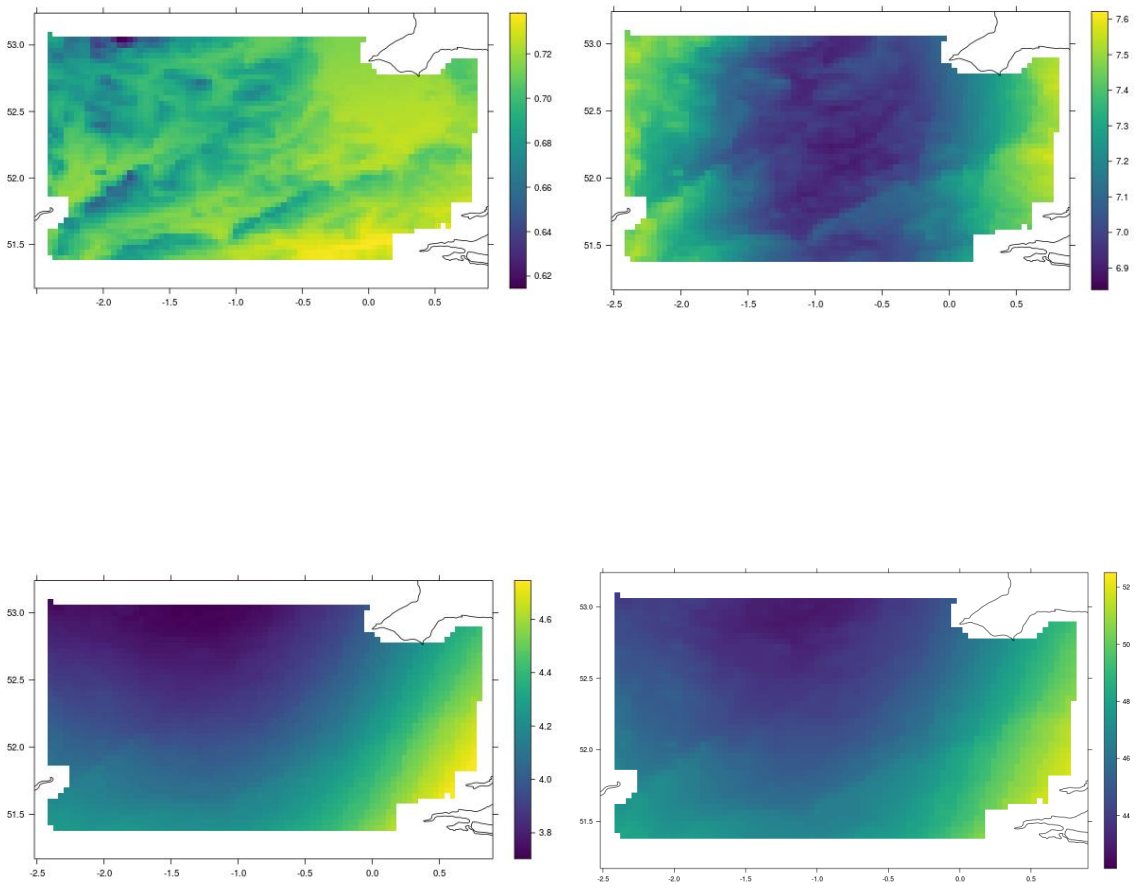

Figure 6.4.3: Spatially smoothed marginal distribution parameter estimates for $\{Y^{\mathcal{C}}(s)\}$, i.e., convective rainfall. Top-left: $\hat{p}(s,\varepsilon)$, top-right: $\hat{q}(s,\varepsilon)$, bottom-left: $\hat{v}(s,\varepsilon)$, bottom-right: 20-year return level estimate.

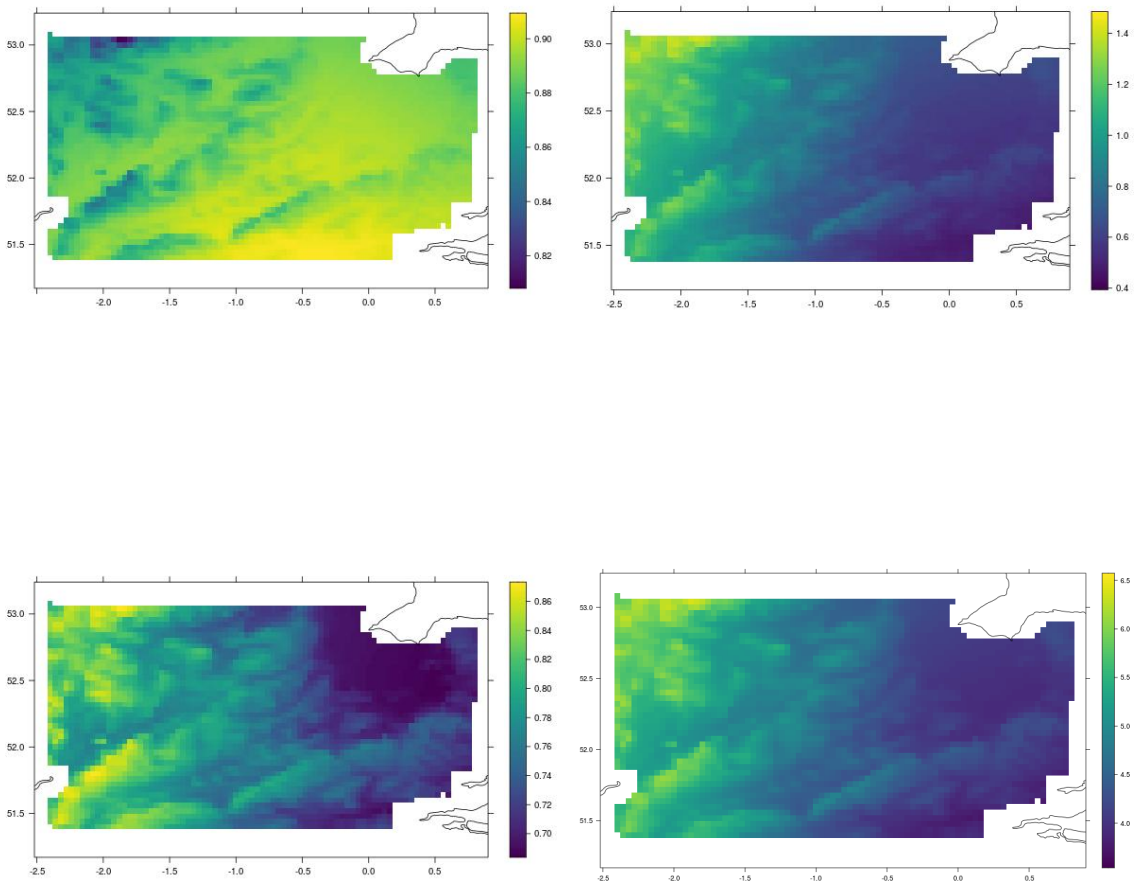

Figure 6.4.4: Spatially smoothed marginal distribution parameter estimates for $\{Y^{\mathcal{NC}}(s)\}$, i.e., non-convective rainfall. Top-left: $\hat{p}(s, \varepsilon)$, top-right: $\hat{q}(s, \varepsilon)$, bottom-left: $\hat{v}(s, \varepsilon)$, bottom-right: 20-year return level estimate.

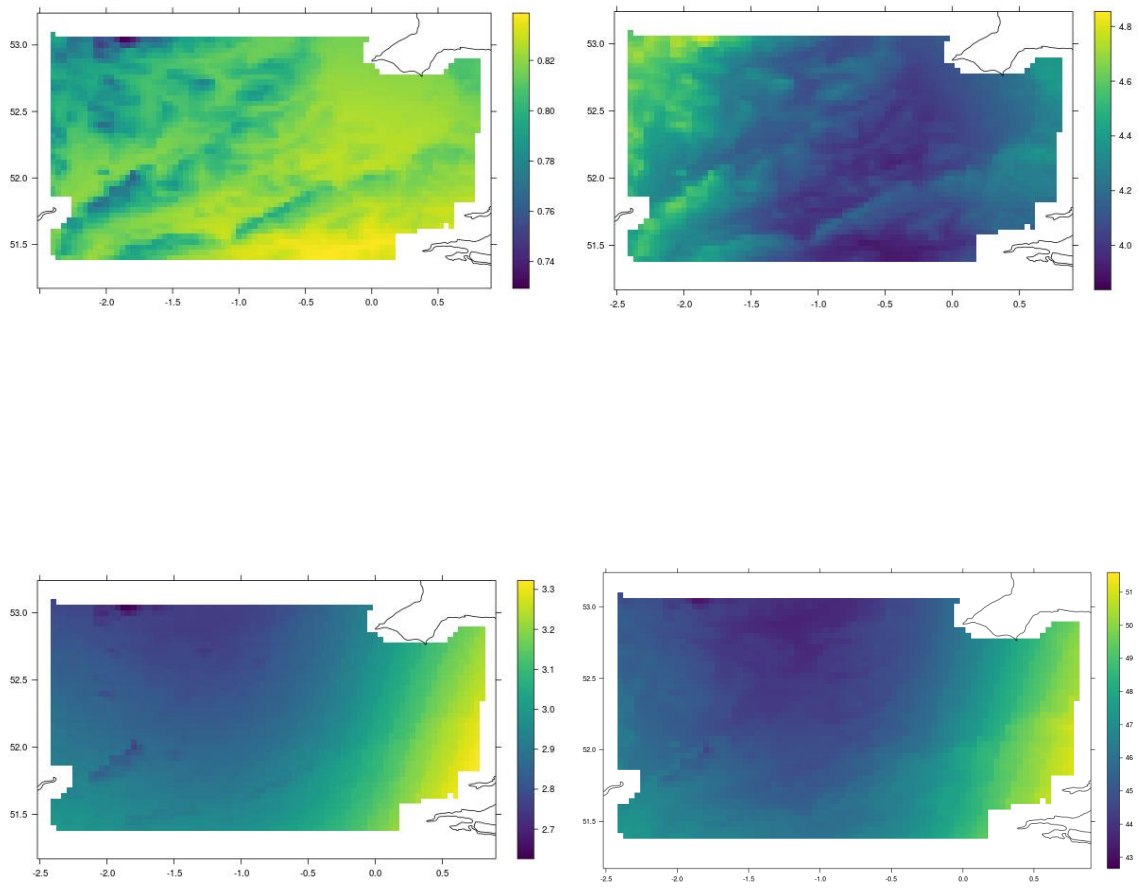

Figure 6.4.5: Spatially smoothed marginal distribution parameter estimates for $\{Y^*(s)\}$, i.e., all rainfall. Top-left: $\hat{p}(s,\varepsilon)$, top-right: $\hat{q}(s,\varepsilon)$, bottom-left: $\hat{v}(s,\varepsilon)$, bottom-right: 20-year return level estimate.

6.4.4 Dependence modelling

We proceed by fitting separately the extremal dependence models described in Section 6.2.2 to $\{X^{\mathcal{C}}(s)\}$, $\{X^{\mathcal{NC}}(s)\}$ and $\{X^*(s)\}$. We use a different exceedance threshold, i.e., u in (5.2.2), for each of the processes; we take u as the 96% and 99% standard Laplace quantiles for $\{X^{\mathcal{C}}(s)\}$ and $\{X^{\mathcal{NC}}(s)\}$, respectively. Ideally for $\{X^*(s)\}$

we would set u such that the number of observations used for inference with both modelling approaches is the same, so as to provide a fair comparison of the two approaches. However, we found that using such a threshold provided poorer inference for $\{X^*(s)\}$. Following the reasoning given in Section 5.4.3, we instead take a higher threshold of u for $\{X^*(s)\}$ as the 99% standard Laplace quantile.

Inference is conducted in the same manner as described in Section 5.3.2 using the stratified sampling regime. Each fit uses $d_s = 6000$ triples of sites, and we set h_{max} to $35km, 50km$ and $250km$ for $\{X^{\mathcal{C}}(s)\}, \{X^*(s)\}$ and $\{X^{\mathcal{NC}}(s)\}$, respectively. We note that for $X^{\mathcal{NC}}$ we conducted inference using both forms of β, i.e., (5.2.5) and (6.2.2); however, we found better results when using the latter and so we present findings with this form only. Following the reasoning given in Section 5.4.3, we allow the fitted δ functions for all three processes to satisfy $\delta(h) \not\geq 1$ for some $h \geq 0$.

We present a comparison of the dependence parameter estimates for the three processes in Figure 6.4.6. As each process has its own anisotropy parameters, we cannot compare estimates of the functions evaluated at h, i.e., the pairwise distance under the anisotropy transformation. Thus, we instead fix a conditioning site s_O in the centre of the domain and evaluate the estimated functions at each $\|s_i^* - s_O^*\|$ for $i = 1, \ldots, d$, where $s_1^*, \ldots, s_d^*$ denote the sampling locations under the estimated anisotropy transformation; we then explore how these estimates change with distance $\|s_i - s_O\|$ for each of the processes.

Figure 6.4.6 suggests that $\{X^{\mathcal{C}}(s)\}$ and $\{X^*(s)\}$ have similar structures in their extremal dependence, which gives evidence to support our claim in Section 5.4.3 that we model predominantly convective events when applying the spatial conditional extremes modelling approach to unclassified precipitation data; thus we can draw similar conclusions about the extremal dependence in $\{X^{\mathcal{NC}}(s)\}$ as we did in Section 5.4.3, namely that all dependence is exhibited through the β function which suggests that the underlying process that generates extreme events is somewhat rough. Figure 6.4.6

gives strong justification for the use of a mixture model approach for extreme precipitation, as we observe widely different structures in the α and β estimates for $\{X^{\mathcal{C}}(s)\}$ and $\{X^{\mathcal{NC}}(s)\}$; for the latter, the α parameter decays much slower with distance and gives no evidence that independence is achieved at even the largest distances within the region $\mathcal{S}$. Moreover, as all dependence is exhibited through the α function, this suggests that extreme realisations of $\{X^{\mathcal{NC}}(s)\}$ are smoother than events from the other two processes.

To compare both the full processes, i.e., both the marginal and dependence structures, we investigate how $\{Y(s)\}|(Y(s_O) = v)$ changes with distance $\|s - s_O\|$ for $s \in \mathcal{S}$; here Y denotes a generic process that we replace with $Y^{\mathcal{C}}$, $Y^{\mathcal{NC}}$ and Y^*. To this end, we take a transect of points $\mathcal{P} \in \mathcal{S}$ and simulate 50000 realisations of $\{Y(s) : s \in \mathcal{P}\}|(Y(s_O) = v^{(l)}_{Y(s_O)})$ where $v^{(l)}_{Y(s_O)}$ denotes the l-year return level for $Y(s_O)$. We use these realisations to estimate the conditional median, and the 2.5% and 97.5% marginal quantiles, of $\{Y(s : s \in \mathcal{P})\}|(Y(s_O) = v^{(l)}_{Y(s_O)})$, which we present in Figure 6.4.7. We consider two values for l, i.e., $l = 1$ and $l = 50$, and take into account the respective length of observation periods when evaluating $v^{(l)}_{Y(s_O)}$; that is, we take $v^{(l)}_{Y^{\mathcal{C}}(s_O)} = F^{-1}_{Y^{\mathcal{C}}(s_O)}(1-1/(l\times|\mathcal{C}|/20))$ and $v^{(l)}_{Y^{\mathcal{NC}}(s_O)} = F^{-1}_{Y^{\mathcal{NC}}(s_O)}(1-1/(l\times|\mathcal{NC}|/20))$ and $v^{(l)}_{Y^*(s_O)} = F^{-1}_{Y^*(s_O)}(1 - 1/(l \times 43200/20))$. We observe that the estimates for $Y^{\mathcal{C}}$ and Y^* are almost identical, with the return-level estimates for $Y(s_O)$ being approximately equal and the conditional medians both decaying quickly with distance. For $Y^{\mathcal{NC}}$, we observe that the conditional median maintains a fairly slow decay rate, with non-zero values at even the largest distances.

Figure 6.4.8 presents realisations of $\{Y^{\mathcal{C}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{C}}(s_O) > v), \{Y^{\mathcal{NC}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{NC}}(s_O) > v)$ and $\{Y^*(s) : s \in \mathcal{S}\}|(Y^*(s_O) > v)$, where $v = F_L^{-1}(0.99)$. The conditioning sites s_O were sampled uniformly at random over $\mathcal{S}$. We observe similarities between realisations of $\{Y^{\mathcal{C}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{C}}(s_O) > v)$ and $\{Y^*(s) : s \in \mathcal{S}\}|(Y^*(s_O) > v)$; both models produce realisations that have characteristics we would expect to

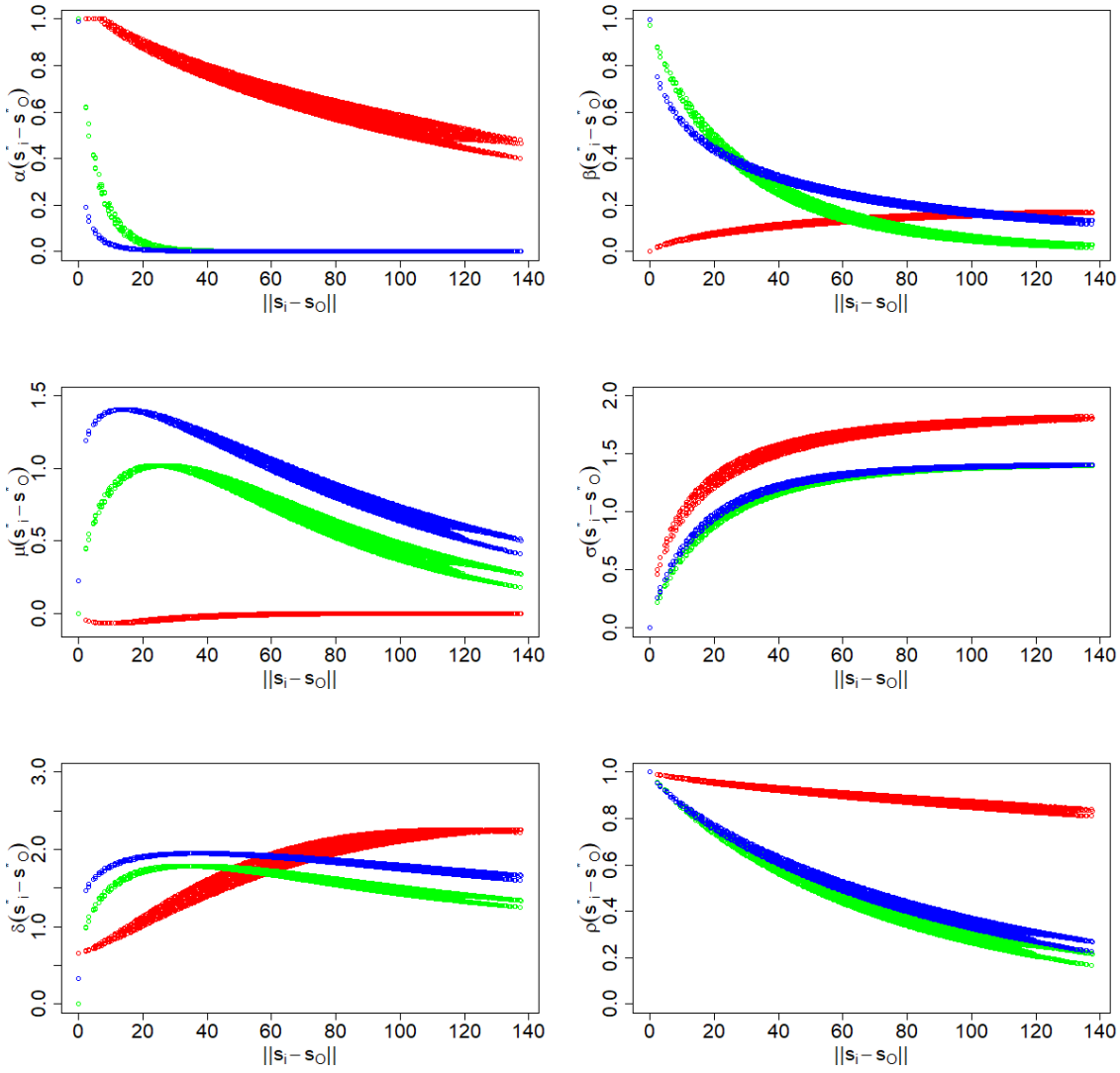

Figure 6.4.6: Estimates of extremal dependence functions evaluated at $\|s_i^* - s_O^*\|$ for $i = 1, \ldots, d$, i.e., anisotropic distances, against original distances $\|s_i - s_O\|$, which are given in km. The conditioning site s_O is in the centre of the spatial domain $\mathcal{S}$. The colours correspond to the estimates for the different spatial processes; these are green, red and blue for $\{X^{\mathcal{C}}(s)\}, \{X^{\mathcal{NC}}(s)\}$ and $\{X^*(s)\}$, respectively.

observe for convective rainfall, e.g., these are high intensity, spatially localised events with a large proportion of the domain $\mathcal{S}$ being dry. We further observe that the model for $\{Y^{\mathcal{NC}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{NC}}(s_O) > v)$ produces events that are much smoother than those produced by the other two models; moreover, these events are lower in their intensity and cover a much larger area.

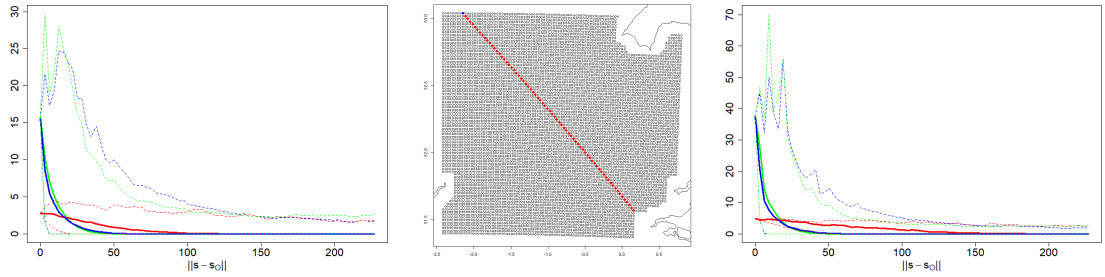

Figure 6.4.7: Summary statistics for $\{Y(s) : s \in \mathcal{P}\}|(Y(s_O) = v^{(l)}_{Y(s_O)})$ against distance $\|s - s_O\|$ with $l = 1$ and $l = 50$ in the left and right plots, respectively. Solid lines correspond to estimates for conditional medians, dashed lines denote 95% confidence intervals. Lines are coloured green, red and blue for $Y^{\mathcal{C}}$, $Y^{\mathcal{NC}}$ and Y^*, respectively. Centre: red and blue points denote $\mathcal{P}$ and s_O.

6.4.5 Inference on spatial aggregates

For each of the three processes, we draw 5×10^5 realisations using the procedure detailed in Section 6.3. For $Y^{\mathcal{C}}$ and Y^*, we used $N' = 8N$ for N' in Algorithm 2; for $Y^{\mathcal{NC}}$, we found that a larger N' was required, and so we set $N' = 20N$. Realisations were created using the regions illustrated in Figure 6.3.1, i.e., $\mathcal{D}$ and $\mathcal{O}$ as described in Section 6.3.2. Note that the regions $\mathcal{A}$ and $\mathcal{D}$ do not change between the three processes, but we take $\mathcal{O}$ as the empty set for $Y^{\mathcal{C}}$ and Y^* and specify $\mathcal{O}$ for $Y^{\mathcal{NC}}$ using the heuristic described in Section 6.3.2; the purple points in Figure 6.3.1 denote $\mathcal{O} \setminus \mathcal{D}$ and are created by adding noise to the coordinate in the centre of $\mathcal{S}$. Using these regions, we create samples of $\bar{R}^{\mathcal{C}}_{\mathcal{A}}$, $\bar{R}^{\mathcal{NC}}_{\mathcal{A}}$ and $\bar{R}^*_{\mathcal{A}} = |\mathcal{A}|^{-1} \int_{\mathcal{A}} Y^*(s)\mathrm{d}s$ for different $\mathcal{A}$ (see Figure 6.3.1); we then create a sample $\bar{R}_{\mathcal{A}}$ by drawing from $\bar{R}^{\mathcal{C}}_{\mathcal{A}}$ and $\bar{R}^{\mathcal{NC}}_{\mathcal{A}}$ with probability $p_{\mathcal{C}}$ and $1 - p_{\mathcal{C}}$, respectively.

Recall that $\bar{R}_{\mathcal{A}}$ is created using our new modelling approach whilst $\bar{R}^*_{\mathcal{A}}$ uses the single process approach detailed in Chapter 5. To illustrate how well the two modelling approaches can capture the extremal behaviour of spatial aggregates, we present Q-Q plots in Figures 6.4.9 and 6.4.10 comparing the quantiles of the simulated aggregates against their empirical equivalents. In Figure 6.4.9, we use just convective and just non-convective data for $\bar{R}^{\mathcal{C}}_{\mathcal{A}}$ and $\bar{R}^{\mathcal{NC}}_{\mathcal{A}}$, respectively; for Figure 6.4.10, we use all data

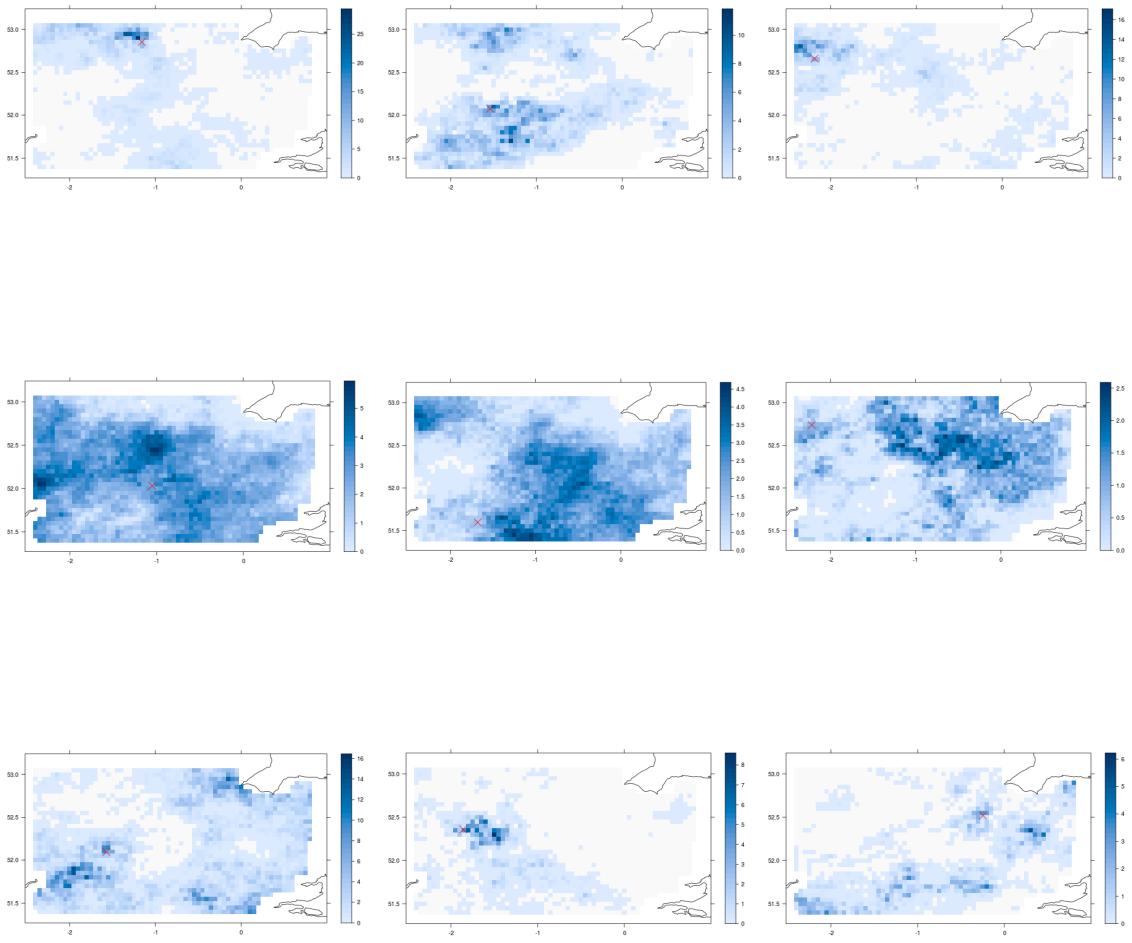

Figure 6.4.8: Extreme precipitation fields (mm/hr). Realisations from the fitted models for $\{Y^{\mathcal{C}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{C}}(s_O) > v)$, $\{Y^{\mathcal{NC}}(s) : s \in \mathcal{S}\}|(Y^{\mathcal{NC}}(s_O) > v)$ and $\{Y^{*}(s) : s \in \mathcal{S}\}|(Y^{*}(s_O) > v)$ in the top, middle and bottom rows, respectively. The conditioning sites s_O are given by the red crosses. Scales differ within each panel and row.

for both $\bar{R}^{*}_{\mathcal{A}}$ and $\bar{R}_{\mathcal{A}}$. Note that due to computational constraints, we were unable to produce tolerance intervals for the estimated quantiles; however, should these be required then the procedure described in Section 5.3.3 can be used separately for each of the three processes.

In Figures 6.4.9 and 6.4.10, we observe good fits for all components of both mod-

elling approaches, with some slight underestimation for the smaller quantiles of $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$. By eye, we seem to have that $\bar{R}_{\mathcal{A}}^{*}$ provides slightly better fits than the corresponding $\bar{R}_{\mathcal{A}}$ for most of the tail. However, observe that for the larger regions, numbered 3 and 4, that $\bar{R}_{\mathcal{A}}^{*}$ overestimates the largest quantiles; this is less of an issue for $\bar{R}_{\mathcal{A}}$, as we observe better better estimates for the very largest quantiles, suggesting that we have made some improvements to the approach in Chapter 5 by including the non-convective component into the model. To further support this claim, we estimate the proportion of non-convective events that contribute to $\bar{R}_{\mathcal{A}} > v^*$, where v^* denotes the 99% and 99.5% quantile of the simulated $\bar{R}_{\mathcal{A}} > v^*$. Both proportions increase with the size of $\mathcal{A}$; for the 99% quantile, this proportion ranges from 0.013 to 0.069, and for the 99.5%, the values range between 0.001 and 0.050. Whilst we observe good fits for the largest regions, we observe that neither the models for $\bar{R}_{\mathcal{A}}$ nor $\bar{R}_{\mathcal{A}}^{*}$ are able to capture the very largest empirical quantile. When we investigated this, we found that this discrepancy was caused by two large events at consecutive hours that are not captured by either dependence model; a discussion of potential extensions to the model that may help to capture these events is given in Section 6.5.

We require some diagnostic metric to provide support for oue claim that $\bar{R}_{\mathcal{A}}$ provides better fits than model $\bar{R}_{\mathcal{A}}^{*}$. We adapt a measure proposed by Varty et al. (2021); we begin by denoting $Q(p; Z) : [0, 1] \to \mathbb{R}_+$ as the sample quantile function of some random variable Z, which is evaluated at a probability $p \in [0, 1]$. Note that we take Z to be one of three variables; either $\bar{R}_{\mathcal{A}}$ or $\bar{R}_{\mathcal{A}}^{*}$, which are taken to be realisations from the model proposed in this chapter and Chapter 5, respectively, or the observed spatial aggregate which we denote $\tilde{R}_{\mathcal{A}}$. Let $p_{min} \in [0, 1]$ and then let $\{p_j = p_{min} + j(1 - p_{min})/(m + 1) : j = 1, \ldots, m\}$ for $m \in \mathbb{N}_+$ be equally spaced probabilities, such that $p_j \in (p_{min}, 1)$ for all j; as our interest lines in the tails of the spatial aggregate, we take $p_{min} > 0$ to be close to one. Then we define the two

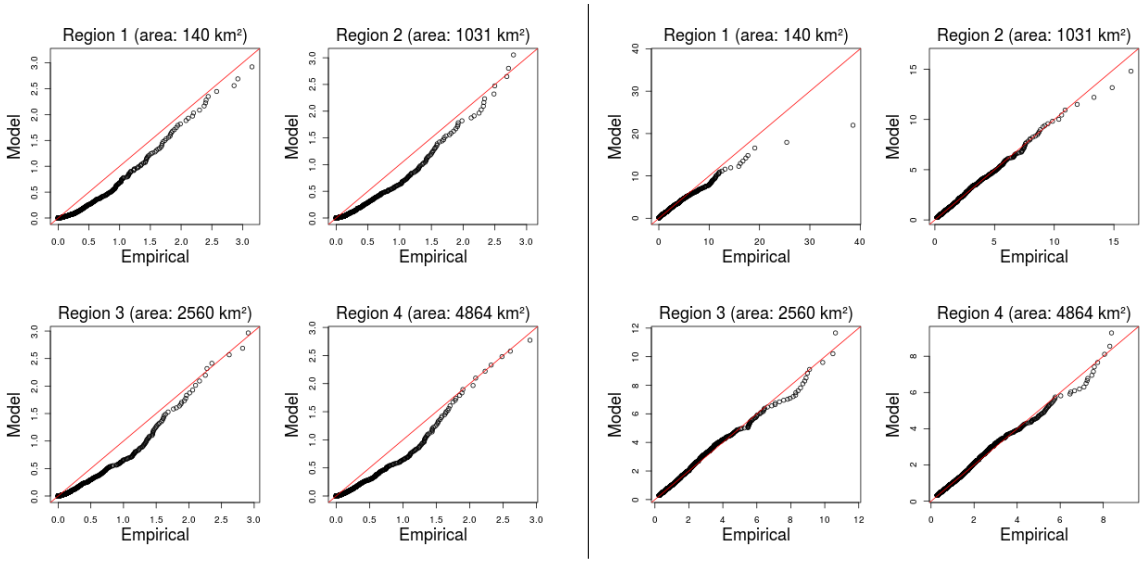

Figure 6.4.9: Q-Q plots for model, and empirical, aggregates of regions of increasing size. Left: $\bar{R}_{\mathcal{A}}^{\mathcal{NC}}$, right: $\bar{R}_{\mathcal{A}}^{\mathcal{C}}$. Probabilities range from 0.8 to a value corresponding to the respective 20 year return level. Regions 1-4 correspond to those illustrated in Figure 6.3.1.

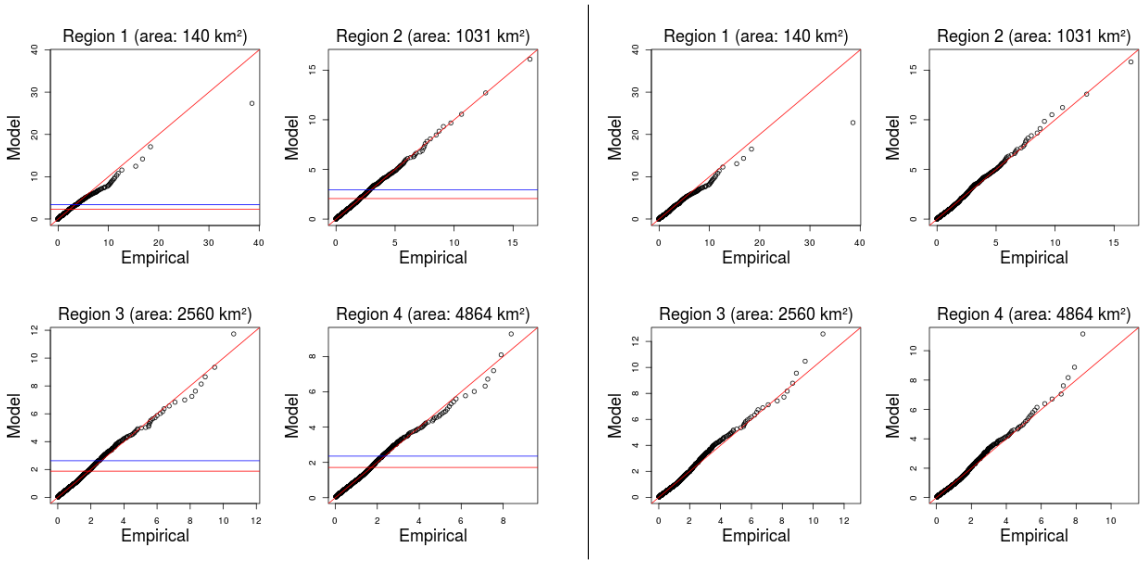

Figure 6.4.10: Q-Q plots for model, and empirical, aggregates of regions of increasing size. Left: $\bar{R}_{\mathcal{A}}$, right: $\bar{R}_{\mathcal{A}}^{*}$, Probabilities range from 0.8 to a value corresponding to the respective 20 year return level. Regions 1-4 correspond to those illustrated in Figure 6.3.1. The blue and red horizontal lines denote the 99% and 99.5% quantiles of the respective simulated aggregates.

measures

$$\Lambda_1(Z) = \frac{1}{m}\sum_{j=1}^{m} |Q(p_j; Z) - Q(p_j; \tilde{R}_{\mathcal{A}})|, \quad \Lambda_2(Z) = \frac{1}{m}\sum_{j=1}^{m} (Q(p_j; Z) - Q(p_j; \tilde{R}_{\mathcal{A}}))^2, \tag{6.4.1}$$

as the expected deviance in the Q-Q plot for Z against $\tilde{R}_{\mathcal{A}}$ from the line $y = x$, with deviation being described through the mean absolute, and mean squared, distances in Λ_1 and Λ_2, respectively. We evaluate Λ_1 and Λ_2 for $Z := \bar{R}_{\mathcal{A}}$ and $Z := \bar{R}^*_{\mathcal{A}}$, and for each region $\mathcal{A}$ illustrated in Figure 6.3.1. These metrics are estimated using $p_{min} = 0.95$ and $m = (1 - 0.95) * 43200 = 2160$, i.e., the number of observations $\tilde{R}_{\mathcal{A}}$ that exceed the 0.95-quantile; estimates are tabulated in Table 6.4.1. Although by eye it appears that $\bar{R}^*_{\mathcal{A}}$ outperforms $\bar{R}_{\mathcal{A}}$ for the largest regions, the estimates in Table 6.4.1 suggest otherwise; observe that the latter model provides lower estimates for both diagnostics, Λ_1 and Λ_2, for the three largest regions.

	$\mathcal{A}$			
	1	2	3	4
$\Lambda_1(\bar{R}_{\mathcal{A}})$	0.132	**0.030**	**0.043**	**0.035**
$\Lambda_1(\bar{R}^*_{\mathcal{A}})$	**0.098**	0.075	0.087	0.086
$\Lambda_2(\bar{R}_{\mathcal{A}})$	**0.142**	**0.003**	**0.007**	**0.005**
$\Lambda_2(\bar{R}^*_{\mathcal{A}})$	0.169	0.010	0.016	0.015

Table 6.4.1: Estimates of aggregate diagnostics Λ_1 and Λ_2 defined in (6.4.1) to 3 d.p. Bold values denote the lower of the two estimates.

6.5 Discussion

We have presented a simple but effective extension of the approach proposed in Chapter 5 for modelling the extremes of spatial aggregates of precipitation. We proposed a two-step mixture modelling approach whereby we first classify observed fields as being either convective or non-convective, and then fit separate spatial models to either dataset; we then detail an approach to simulate from both models and combine samples to explore the upper-tail behaviour of $\bar{R}_{\mathcal{A}}$. Our dual-process modelling approach was compared against the single process approach and we found that the former was able to better capture the extremal behaviour of aggregates over very large spatial regions. We now discuss some further extensions that can be made to improve the

model.

In Section 6.4.5, we identified two consecutive convective events that provided the two largest values of the empirical $\bar{R}_{\mathcal{A}}$ over the smallest aggregate region $\mathcal{A}$, denoted region 1; neither the model for $\bar{R}_{\mathcal{A}}$ nor $\bar{R}^*_{\mathcal{A}}$ was able to capture either of these two anomalous events. As these particular events were at consecutive time points, this leads us to suspect that there is a temporal aspect to the data that we cannot currently capture with the model. The data themselves are aggregates of precipitation over a temporal interval of one hour and a spatial grid-box. Hence, the data values, which are produced by storms moving through a grid-box, are likely to be more extreme if said storm is moving at a slower rate. A potential model extension may be to incorporate the speed of a storm as a covariate in the model. Another natural extension is to incorporate a temporal component into the dependence models; Simpson and Wadsworth (2021) have already proposed a spatio-temporal extension of the conditional extremes framework, and we may be able to adapt this model to allow for modelling of extreme precipitation.

We found that proposing a simple mixture of only two processes for extremal precipitation was sufficient for modelling the extremal behaviour of spatial aggregates; however, this approach is still somewhat unrealistic and does not capture the true underlying physical properties of extreme rainfall. In Section 6.4.2, we detailed the algorithm that we use to classify observed fields; recall that we identify an entire field as convective if any single grid-box within that field is identified as convective. That is, we model under the assumption that the properties of the process that generates extreme convective rainfall are the same regardless of the amount of convective rainfall within $\mathcal{S}$, or the presence/proportion of non-convective rainfall within the same spatial domain. A more realistic approach is to accommodate the effect of mixing of convective and non-convective rainfall in our model, as we become more and more likely to observe both within a single field as the spatial domain $\mathcal{S}$ grows in size. A

simple approach to tackling this issue is to incorporate the proportion of non-zero rainfall within a field, identified as convective, as a covariate in at least one of the marginal and dependence components of the process $Y^{\mathcal{C}}$; a similar idea could be used for $Y^{\mathcal{NC}}$.

We find evidence to suggest that the identification algorithm detailed in Section 6.4.2 works well, as we observe distinctly different structures in the extremal dependence models fitted to the two classified datasets. However, the technique for identification is deterministic in nature, and so improvements could potentially be made by adopting a more probabilistic approach, i.e., through the use of mixture process models and a Bayesian framework for inference. For example, we could construct the residual process $Z(s|s_O)$ using a Dirichlet mixture of Gaussian processes (see Duan et al. (2007)), rather than a single Gaussian process. This would have the added benefit of the model not being limited to only two mixture components: convective and non-convective. Instead, a number of mixture components could be used, each with their own dependence structure.Hazra and Huser (2019) advocate the use of Dirichlet mixtures of Student-t processes for modelling extremal dependence in sea surface temperature data. They propose a computationally inexpensive inference procedure that relies on low-rank approximations of the correlation matrices for each mixture component. However, due to the censored approach we take to inference, a Bayesian approach may be computationally infeasible.

7

Spatial deformation for non-stationary extremal dependence

7.1 Introduction

Statistical methodology for spatial extremes can increasingly handle data sampled at more observation locations. If these observations are taken over large domains with complex features, then there is a strong chance that the data will exhibit spatial non-stationarity in both the marginal distributions and dependence structure. Marginal non-stationarity can often be dealt with by site-wise modelling and transformation. However, there are currently few methods to deal with non-stationarity in extremal dependence structures, and a typical approach is to falsely assume stationarity when fitting spatial extremes models. This may be appropriate when modelling data sampled over small and/or homogeneous regions in space, but as we will illustrate through the examples in Section 7.3, this assumption is not realistic for many datasets with larger spatial domains.

Beyond site-wise transformation of margins, marginal non-stationarity can be handled by jointly modelling marginal parameters as functions of covariates. This can either be achieved parametrically (Mannshardt-Shamseldin et al., 2010; Davison et al., 2012; Ribatet, 2013; Eastoe, 2019) or semiparametrically (Jonathan et al., 2014; Ross et al., 2017; Youngman, 2019, 2020; Zanini et al., 2020) through the use of splines. Another widely applied approach is the use of Bayesian hierarchical models, in which the marginal parameters are assumed to come from some non-stationary latent process (Casson and Coles, 1999; Cooley et al., 2007; Sang and Gelfand, 2010; Opitz et al., 2018).

Non-stationarity in the spatial dependence structure has been studied by Huser and Genton (2016) in the context of max-stable models, through incorporation of a non-stationary variogram. However, this approach requires knowledge of relevant covariates, and asymptotically dependent max-stable models for spatial extremes have been shown to be too inflexible for many spatial datasets (Wadsworth and Tawn, 2012; Davison et al., 2013; Huser et al., 2017; Huser and Wadsworth, 2019). Another approach is to assume local stationarity for model fitting, see Blanchet and Creutin (2017); Castro-Camilo and Huser (2020). This framework is well-suited to modelling processes with short-range dependence but is unlikely to fully capture dependence at large distances. Cooley et al. (2007) and Blanchet and Davison (2011) account for non-stationarity by transforming their spatial domain of interest to some new 'climate space' in which observation locations with similar characteristics are grouped closer together. Again, this approach requires access to relevant covariates and a deeper understanding of the processes which are being modelled.

In this work we develop a computationally quick and simple method, which does not require prior knowledge of covariates and which can be applied before fitting any model suited to spatial extremes. Our method uses spatial deformation and is based on the work of Sampson and Guttorp (1992) and Smith (1996), which has not been

fully adapted for use in a spatial extremes framework. The deformation methodology may reveal physical features and/or covariates that can be incorporated into a spatial extremes model, removing the need for models with complex dependence structures.

Wadsworth and Tawn (2019) applied the deformation method of Smith (1996) before fitting a conditional spatial extremes model to the same Australian summer temperatures data that we explore in Section 7.4.1. However, because this method is not tailored to extremal dependence, it was neccesary to assume that patterns in non-stationarity were similar for both the extremal and non-extremal dependence structures. Youngman (2020) and Chevalier et al. (2020) provide extensions of the Sampson and Guttorp (1992) methodology and fit models for spatial extremes using deformations: a Gaussian process using a censored pairwise likelihood and a max-stable model, respectively. Although these models may be reasonable for some processes, use of either puts restrictions on the types of dependence that the process can exhibit. We look to develop a method that makes no strong assumptions on the extremal dependence structure.

The remainder of this section provides an overview of existing methodology for spatial deformation and modelling of spatial extremes. Our developments of the spatial deformation methodology are detailed in Section 7.2. We present a simulation study in Section 7.3, which is usually absent from the literature on spatial deformations. This study is used to convey that our adaptations to the deformation methodology are necessary when considering extremal dependence and that our method can be used for different processes with a wide range of extremal dependence structures. Finally, we apply our method to temperature and precipitation datasets in Section 7.4, and conclude with a discussion in Section 7.5.

7.1.1 Non-stationary spatial processes

The spatial deformation approach for handling non-stationarity in spatial processes was first proposed by Sampson and Guttorp (1992) and Guttorp and Sampson (1994), with further developments by Meiring et al. (1997), see Sampson (2010, Ch. 9.5). The underlying principle of their approach is that a smooth non-linear transformation can be used to map the sampling locations of a process from a geographical plane, or G-plane, to some latent space, which they name a D-plane, or dispersion plane. Within the D-plane, the dependence structure of the process is assumed to be both stationary and isotropic, and the usual statistical inferences can be made using stationary geostatistical models. To obtain the D-plane, optimisation techniques are used to minimise some objective function which is associated with a stationary geostatistical model. Here Sampson and Guttorp (1992) use multi-dimensional scaling and a stationary spatial dispersion function, whereas further work proposed by Smith (1996) uses the likelihood for a stationary Gaussian process. Our approach is to change this objective function for one which is associated with a stationary spatial extremes model, such as the max-stable, or inverted max-stable, processes.

We begin by assuming we have realisations $\mathbf{Z} = \{Z_1, \ldots, Z_N\}$ from a spatial field observed at sampling locations $s_1, \ldots, s_d$, and so we have $Z_k = \{Z_k(s_1), \ldots, Z_k(s_d)\}$ for all $k = 1, \ldots, N$. We require some smooth mapping function from the G-plane to the D-plane, given by $f(s_i) = s_i^*$ for $i = 1, \ldots, d$, where $s_i = (x_i, y_i)$ and $s_i^* = (x_i^*, y_i^*)$ are the corresponding locations in the D-plane. Both Sampson and Guttorp (1992) and Smith (1996) propose the use of thin-plate splines to achieve this mapping. However, we note that under certain conditions on the correlation structure, analytical forms for $f(\cdot)$ do exist. Perrin and Meiring (1999) prove that this mapping is identifiable assuming differentiability of the stationary and isotropic correlation function used for fitting and Perrin and Senoussi (2000) derive analytical forms for $f(\cdot)$ under the same assumption, with extensions to anisotropic correlation structures. As these

results are available only for correlation functions, and not for extremal dependence functions, we instead use the more flexible thin-plate spline approach.

A thin-plate spline is a mapping function $f(\cdot)$, passing through a finite number of data points $f_i^* = f^*(x_i, y_i), (i = 1, \ldots, n)$, minimising the bending energy

$$J(f) = \iint_{\mathbb{R}^2} \left\{ \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right\} \mathrm{d}x\mathrm{d}y.$$

Here we have denoted f^* the 'true' function that we wish to estimate with the thin-plate spline, f, and f_i^* are observations. Green and Silverman (1994) give a solution to this problem in the form

$$f(x, y) = a + bx + cy + \sum_{i=1}^{n} \delta_i g_i(x, y), \tag{7.1.1}$$

where

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \delta_i x_i = \sum_{i=1}^{n} \delta_i y_i = 0, \tag{7.1.2}$$

and $g_i(x, y) = h_i^2 \log h_i$, with h_i the Euclidean distance between (x, y) and (x_i, y_i). This represents f as the sum of linear terms and n radial basis functions with centres at the observed data locations (x_i, y_i) and the constraints are in place to ensure that the system of equations does not become overdetermined. An interpolating spline satisfies $f_i^* = f(x_i, y_i)$ for all $i = 1, \ldots, n$, whereas we desire a smoothing spline; this can be created by minimising

$$S(f) = \sum_{i=1}^{n} \{f_i^* - f(x_i, y_i)\}^2 + \alpha J(f),$$

for some smoothing parameter $\alpha > 0$. Sampson and Guttorp (1992) give a method for estimating α in the context of multidimensional scaling, but here we take the approach of Smith (1996), who uses a restricted representation of (7.1.1) instead. A subset of m radial basis functions is used and so we let $\delta_i = 0$ for all $i \notin \{i_1, \ldots, i_m\}$.

The choice of this subset is discussed in Section 7.3.

The function in (7.1.1) maps $\mathbb{R}^2$ to $\mathbb{R}$, so the spline is applied twice with different parameter estimates to produce both components. Smith (1996) gives a parametrisation as

$$f^{(1)}(x,y) = b_1^2 x + \rho b_1 b_2 y + \sum_{i=1}^{m} \delta_i^{(1)} g_i(x,y) \tag{7.1.3}$$

$$f^{(2)}(x,y) = b_2^2 y + \rho b_1 b_2 x + \sum_{i=1}^{m} \delta_i^{(2)} g_i(x,y), \tag{7.1.4}$$

where $b_1 > 0,\ b_2 > 0,\ \rho \in \mathbb{R}$ and each of the sequences $\delta^{(1)},\ \delta^{(2)}$ satisfy the constraint in (7.1.2). The introduction of the parameters b_1, b_2 and ρ is to ensure that the model is invariant under orthogonal rotations when $m = 0$. Overall, this yields a spline with $2m - 3$ free parameters whenever $m \geq 3$.

The resulting spline is then used to map the sampling locations s_i to locations s_i^* in a latent space. Parameters are estimated by minimising some objective function provided by a stationary model. As previously mentioned, Sampson and Guttorp (1992) use a stationary spatial dispersion model and multidimensional scaling, the details of which are not given here. Instead, we focus on the approach by Smith (1996), who uses a stationary Gaussian likelihood. It is assumed that $(Z(s_1^*), \ldots, Z(s_d^*)) \sim N_d(\mu, \Omega)$, where μ and Ω are the mean vector and a stationary covariance matrix, respectively. As we are only interested in measuring the dependence structure, it is assumed that the means and variances at each location are known. Analysis is then simplified to only considering the minimisation of the negative log likelihood given by

$$-\log L(\Omega) = \frac{N}{2} \log |\Omega| + \frac{N-1}{2} \operatorname{tr}\left(\Omega^{-1}\hat{\Omega}\right), \tag{7.1.5}$$

where Ω and $\hat{\Omega}$ are the theoretical, and sample, correlation matrices and $\operatorname{tr}(\cdot)$ and $|\cdot|$ are the trace and determinant operators, respectively. The entries of the theoretical

correlation matrix are produced by using a stationary covariance function. Smith (1996) uses the Matérn covariance function, and so

$$\Omega_{ij} = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{2h_{ij}^*\sqrt{\theta_2}}{\theta_1} \right)^{\theta_2} K_{\theta_2} \left(\frac{2h_{ij}^*\sqrt{\theta_2}}{\theta_1} \right), \tag{7.1.6}$$

where $\theta_1 > 0,\ \theta_2 > 0$ and $K_{\theta_2}(\cdot)$ is the modified Bessel function of the second kind of order θ_2 and $h_{ij}^* = \|s_i^* - s_j^*\|$ is the Euclidean distance between locations s_i^* and s_j^* in the D-plane. It is noted that θ_1 can be set to 1 as the spatial scaling of the locations is controlled by the spline.

7.1.2 Spatial extremes

Before describing an extension of the spatial deformation methodology tailored to spatial extremes, we first provide a brief review of methods for modelling spatial extremes.

Max-stable and inverted max-stable processes

Max-stable processes were introduced by de Haan (1984) and developed further by Smith (1990) and Schlather (2002), who suggested models that were first fitted by pairwise composite likelihood in Padoan et al. (2010). They are usually described by a spectral construction. Suppose $\{r_i; i \geq 1\}$ are points of a Poisson process on $(0, \infty)$ with unit intensity. Let $S \subseteq \mathbb{R}^2$ be a spatial index set, and $\{W_i(s); s \in S, i \geq 1\}$ be independent and identically distributed copies of a non-negative stochastic process satisfying $\mathbb{E}[W(s)] = 1$. Then

$$Z(s) = \max_{i \geq 1} \{W_i(s)/r_i\} \tag{7.1.7}$$

is a max-stable process with unit Fréchet margins. The d-dimensional joint distribution function for Z is

$$\Pr\{Z(s_1) \leq z_1, \ldots, Z(s_d) \leq z_d\} = \exp\{-V(z_1, \ldots, z_d)\}, \tag{7.1.8}$$

where the exponent is

$$V(z_1, \ldots, z_d) = \mathbb{E}\left[\max\left\{\frac{W(s_1)}{z_1}, \ldots, \frac{W(s_d)}{z_d}\right\}\right]. \tag{7.1.9}$$

Careful specification of the stochastic process $W(s)$ leads to a limited selection of parametric models for the max-stable process. A particularly flexible model is the Brown-Resnick model (Brown and Resnick, 1977; Kabluchko et al., 2009). This involves specifying $W(s) = \exp\{U(s) - \gamma^*(s, 0)\}$ for $U(s)$ a centred Gaussian process with semivariogram $\gamma^*(\cdot, \cdot)$ and where $U(0) = 0$ almost surely. This leads to a 2-dimensional joint distribution with exponent function

$$V(z_i, z_j) = \frac{1}{z_i}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_i}{z_j}\right)\right\} + \frac{1}{z_j}\Phi\left\{\frac{a}{2} - \frac{1}{a}\log\left(\frac{z_j}{z_i}\right)\right\}, \tag{7.1.10}$$

where $a = [2\gamma^*(s_i, s_j)]^{1/2}$ and $\Phi(\cdot)$ denotes the standard normal distribution function. Note that for a stationary and isotropic Brown-Resnick process, $\gamma^*(s_i, s_j)$ is dependent on $h_{ij} = \|s_i - s_j\|$ only. For clarity, we write $\gamma(h_{ij})$ when Z is stationary and isotropic, and $\gamma^*(s_i, s_j)$, otherwise. Representations for (7.1.10) in higher dimensions exist (see Huser and Davison (2013) or Wadsworth and Tawn (2014)), but due to their computational complexity, inference for max-stable processes is typically done pairwise, providing a reasonable balance between computation time and efficiency.

Max-stable processes are inherently asymptotically dependent, or perfectly independent. That is, $Z(s_i)$ and $Z(s_j)$ are asymptotically dependent, or perfectly independent, for all $s_i, s_j \in S$. Here we characterise asymptotic dependence using the

upper tail index χ (Joe, 1997). Assuming $Z(s_i) \sim F_i, Z(s_j) \sim F_j$, we have

$$\chi(s_i, s_j) = \lim_{q \to 1} \Pr\{F_i\{Z(s_i)\} > q | F_j\{Z(s_j)\} > q\}, \tag{7.1.11}$$

where the process is asymptotically independent at locations s_i and s_j if $\chi(s_i, s_j) = 0$, and asymptotically dependent otherwise. Here we write $\chi(s_i, s_j)$ as Z is not necessarily stationary; henceforth, we write $\chi(h_{ij})$ for $h_{ij} = \|s_i - s_j\|$ when it is assumed that χ is a function of distance only. As this measure is theoretically non-zero at all spatial lags for any max-stable process exhibiting positive spatial association i.e., $\chi(s_i, s_j) > 0$ for all $s_i, s_j \in S$, we require other modelling approaches to deal with processes that may exhibit asymptotic independence.

Wadsworth and Tawn (2012) introduced the inverted max-stable process as that obtained by applying a monotonically decreasing marginal transformation to a max-stable process. For example, with Z as defined in (7.1.7), taking $Y(s) = 1/Z(s)$ gives an inverted max-stable process with exponential margins and joint survival function

$$\Pr\{Y(s_1) \geq y_1, \ldots, Y(s_d) \geq y_d\} = \exp\{-V(1/y_1, \ldots, 1/y_d)\}, \tag{7.1.12}$$

where V is as given in (7.1.9). Such a process is asymptotically independent with $\chi(s_i, s_j) = 0$ for all $s_i \neq s_j$, but can accommodate a variety of flexible extremal dependences structures exhibiting positive association. The dependence in asymptotically independent processes may be characterised by a pre-limiting version of (7.1.11). Specifically, under an assumption of hidden regular variation (Ledford and Tawn, 1996; Resnick, 2002),

$$\chi_q(s_i, s_j) = \Pr\{F_i\{Z(s_i)\} > q | F_j\{Z(s_j)\} > q\} = L(1-q)(1-q)^{1/\eta(s_i, s_j)-1}, \tag{7.1.13}$$

with $L(\cdot)$ slowly varying at 0 and $\eta(s_i, s_j) \in (0, 1]$ the coefficient of tail dependence.

For an inverted max-stable process, $\chi_q(s_i, s_j) = (1-q)^{V(1,1)-1}$.

We fit both max-stable and inverted max-stable models after applying our deformation method for non-stationary spatial extremes. Note that although max-stable processes are typically taken to represent the limiting behaviour of maxima, in practice they, along with inverted max-stable processes, can be used for all extreme values through specification of a censored likelihood; see Section 7.2.5. Inference on these models can then be used to determine the efficacy of our deformation method.

Conditional extremes

An alternative approach to modelling spatial extremes is to condition on the behaviour of the process when it is extreme at a single site. Here we give a brief overview of modelling the extremal behaviour of the process at two sites using this approach. For a full characterisation, see Wadsworth and Tawn (2019) or Shooter et al. (2019). We suppress some of the notation used by Wadsworth and Tawn (2019) and Shooter et al. (2019) as we are only considering a discrete pairwise fit, that we will employ in Section 7.4 as a diagnostic measure. For further details of the discrete approach, see Heffernan and Tawn (2004). Winter et al. (2016) apply this same methodology to a dataset of Australian temperatures, which we revisit in Section 7.4.1.

We begin by assuming that $\{X(s) : s \in S \subset \mathbb{R}^2\}$ is a stationary and isotropic process with exponential-tailed marginals and denote $X(s_i) = X_i$. Conditioning on $X_i = x_i > u$ being large and considering $X_j, i \neq j$, Heffernan and Tawn (2004) assume that there exist normalising functions $a(x_i) : \mathbb{R} \to \mathbb{R}, b(x_i) : \mathbb{R} \to \mathbb{R}_+$, for which

$$\lim_{x_i \to \infty} [\Pr(X_j \leq a(x_i) + b(x_i)z | X_i = x_i)] = G(z),$$

where G is non-degenerate. Re-writing $Z = \{X_j - a(x_i)\}/b(x_i)$ as the standardised residual, and making the assumption that the limit holds above some high threshold

u, we have

$$\Pr(Z \leq z | X_i = x_i) = G(z), \quad x_i > u,$$

where $X_i | X_i > u \sim \text{Exp}(1)$ is independent of Z. Inference on G is often simplified by making the working assumption that $Z \sim N(\mu, \sigma^2)$ and using a specified parametric form for the normalising functions $a(\cdot), b(\cdot)$. For positively dependent data, we simplify the normalising functions to $a(x_i) = \alpha x_i$ for $\alpha \in [0, 1]$ and $b(x_i) = x_i^{\beta}$ for $\beta \in [0, 1)$. The bivariate form of the conditional model can thus be expressed

$$X_j | \left(X_i = x_i\right) = \alpha x_i + x_i^{\beta} Z, \quad x_i > u.$$

The conditional model holds some useful advantages over joint modelling using max-stable, or inverted max-stable, processes. For one, it is able to handle both asymptotically dependent, or asymptotically independent, data. Parameter estimates for α and β can indicate the nature of the dependence between X_j and X_i. For example, asymptotic dependence between X_j and X_i is implied by estimates $\alpha = 1, \beta = 0$. Within the class of asymptotically independent variables, $\alpha < 1, \beta > 0$, with $\alpha = \beta = 0$ giving near extremal independence.

The spatial extensions of this model (Wadsworth and Tawn, 2019; Shooter et al., 2019) specify α and β as functions of distance between sites, when the underlying process is stationary and isotropic. As such, we can use these parameter estimates as diagnostics, to determine whether our deformation method has created a process that has a more stationary extremal dependence structure. We are motivated to use these estimates as our deformation method does not use a conditional extremes approach for fitting.

7.2 Spatial deformation for extremes

In this section, we discuss our adaptations of the deformation methodology for application in a spatial extremes framework. We begin in Section 7.2.1 by proposing a new objective function to that of (7.1.5). Instead, we consider minimising the difference between theoretical and empirical χ measures, where the former are produced through specification of a stationary max-stable dependence structure for the process in the D-plane. This does not in fact mean that this method will not work for asymptotically independent data; on the contrary, in Sections 7.2.2 and 7.2.3 we show that the model choice for $\chi(\cdot)$ is somewhat arbitrary and a single, simple parametric form works well for both classes of extremal dependence. Section 7.2.4 follows with some practical advice for choosing the anchor points used in estimating the thin-plane spline and we conclude with details of model fitting and selection using censored pairwise likelihoods in Section 7.2.5. To assess the efficacy of the deformations we produce, we fit full max-stable, and inverted max-stable, dependence models.

7.2.1 Objective function

To adapt the methodology of Sampson and Guttorp (1992) and Smith (1996) to better suit a spatial extremes framework, we change the objective function given in (7.1.5) to the Frobenius norm of the difference between theoretical and empirical pairwise dependence matrices $\mathrm{X} := [\chi(h^*_{ij})]$ and $\hat{\mathrm{X}} := [\hat{\chi}(h^*_{ij})]$. That is, we estimate the parameters of the thin plate spline through computing

$$\min \|\mathrm{X} - \hat{\mathrm{X}}\|_F = \min \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \left\{\chi(h^*_{ij}) - \hat{\chi}(h^*_{ij})\right\}^2}, \tag{7.2.1}$$

where $\chi(h^*_{ij})$, defined in (7.1.11), is the upper tail index calculated between the process at locations s^*_i and s^*_j in the D-plane and $\hat{\chi}(h^*_{ij})$ is its empirical estimate. Recall that

we assume stationarity in the D-plane, and so write $\chi(h^*_{ij})$, rather than $\chi(s^*_i, s^*_j)$. In practice, this measure cannot be estimated in the limit as $q \to 1$. As such, we estimate $\hat{\chi}(h^*_{ij})$ by fixing some high threshold $q < 1$ and calculating

$$\hat{\chi}_q(h^*_{ij}) = \Pr\{\hat{F}_i\{Z(s^*_i)\} > q | \hat{F}_j\{Z(s^*_j)\} > q\} = \Pr\{\hat{F}_i\{Z(s_i)\} > q | \hat{F}_j\{Z(s_j)\} > q\}, \tag{7.2.2}$$

where $\hat{F}_k(\cdot)$ is the empirical distribution of observations $Z(s^*_k) = Z(s_k)$. Under asymptotic dependence, we assume that $\chi_q(h^*_{ij}) \equiv \chi(h^*_{ij})$ for large enough q. Under asymptotic independence, although $\chi_q(h^*) \to 0$ as $q \to 1$, we typically have $\chi_q(h^*) > 0$ for $q < 1$ and spatial structure in this measure that makes it informative about non-stationarity.

We now focus on a choice of function $\chi(h^*)$, which we only require to be monotonically decreasing from 1 to 0. This leaves several options, including specific parametric forms for $\chi(h^*)$ and $\chi_q(h^*)$ from max-stable, and inverted max-stable, processes. We remark that while we have used χ to measure extremal dependence, other extremal dependence measures exist, and can also be used in this framework. For example, the coefficient of tail dependence, $\eta(h^*_{ij})$, from (7.1.13) can also be used to characterise the strength of asymptotic independence in extremes. This can be estimated separately from $\chi(h^*_{ij})$, however, we found that due to the high variance of the estimator for $\eta(h^*_{ij})$, it was often outperformed by using $\chi(h^*_{ij})$.

7.2.2 Asymptotic dependence versus asymptotic independence

As a parametric model for $\chi(h^*)$ we take the form implied by the stationary Brown-Resnick process,

$$\chi(h^*_{ij}) = 2 - \theta(h^*_{ij}) = 2 - 2\Phi\left\{\frac{[2\gamma(h^*_{ij})]^{1/2}}{2}\right\}, \tag{7.2.3}$$

where $\theta(\cdot)$ is the extremal coefficient function (Schlather and Tawn, 2003) and $\theta(h^*_{ij}) = V(1,1)$, with $V(\cdot,\cdot)$ defined in (7.1.10). The semivariogram $\gamma(h^*_{ij})$ controls the dependence of the max-stable field and a typical choice for the semivariogram would be

$$\gamma(h^*_{ij}) = (h^*_{ij}/\lambda)^{\kappa}, \tag{7.2.4}$$

where $\lambda > 0$ is a scaling parameter and $\kappa \in (0, 2]$ is a smoothing parameter. Note that setting $\kappa = 2$ yields the Smith process (Smith, 1990), a special case of the Brown-Resnick process. As previously mentioned when discussing the Smith (1996) methodology for spatial deformation, we can set the scaling parameter λ to 1, as the spatial scaling of locations is controlled by the deformation itself. Note that the motivation for using the Brown-Resnick process as a parametric model is that $\chi(h^*) \to 0$ as $h^* \to \infty$, unlike other popular parametric models. For a stationary inverted Brown-Resnick process, we have

$$\chi_q(h^*_{ij}) = (1-q)^{\theta(h^*_{ij})-1}. \tag{7.2.5}$$

We denote the dependence measures in (7.2.3) and (7.2.5) as χ^{BR} and χ^{IBR}_q, respectively. Note that although these two measures have different parametric forms, and are applicable to different dependence structures, they often approximate each other very closely when used within a deformation framework; this is illustrated in Figure 7.2.1. Here we create deformations for a simulated dataset as described in Section 7.3.1 using both χ^{BR} and χ^{IBR}_q. The plots show that both methods give very similar deformations when considering the non-stationarity in the $\chi(h^*_{ij})$ estimates. This seems to be the case for both asymptotically dependent and asymptotically independent data. Hence, for the sake of simplicity we only use χ^{BR} to create deformations in the case studies in Section 7.4, as it appears to be flexible enough to capture non-stationarity in both classes of extremal dependence.

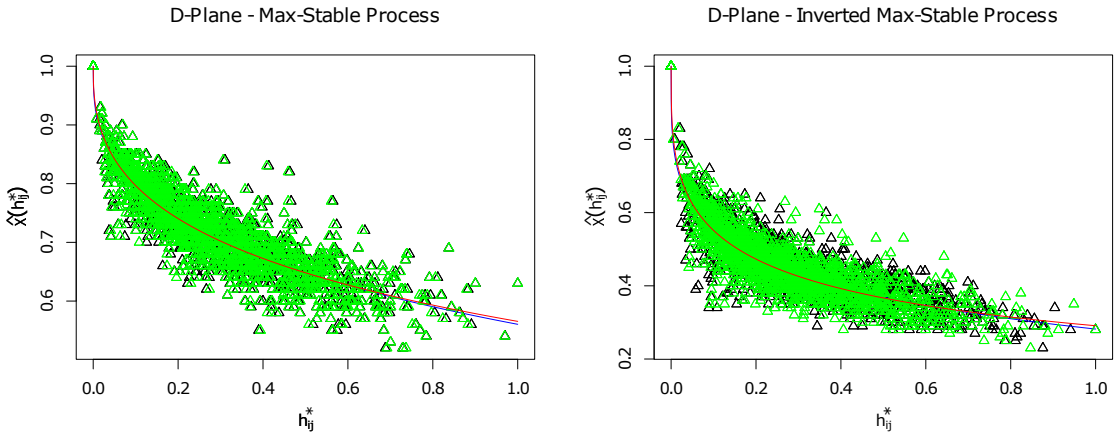

Figure 7.2.1: Comparison of deformations created using both parametric forms χ^{BR} and χ_q^{IBR} for $\chi(\cdot)$ for both max-stable data (left) and inverted max-stable data (right). Plots show empirical $\chi(h^*_{ij})$ estimates against distance, where the black triangles correspond to those created using $\chi(\cdot)$ given by (7.2.3) and green triangles for those created using (7.2.5). The blue and red lines give the fitted function from (7.2.3) and (7.2.5), respectively. Distances are normalised so that the maximum distance is consistent between deformations.

7.2.3 Choice of parametric model for $\chi(h^*)$

We have also found that the function $\chi(h^*)$ from a Brown-Resnick process is sufficiently flexible to create suitable deformations for a variety of different extremal dependence structures. This is for similar reasons to above; different functions $\chi(h^*)$ which decrease to zero as $h^* \to \infty$ can approximate each other well. To illustrate this, we also considered the Gaussian-Gaussian process (Wadsworth and Tawn, 2012), which encompasses different dependence structures to the Brown-Resnick process, but for which $\chi(h^*) \to 0$ as $h^* \to \infty$. Its theoretical form is

$$\chi^{GG}(h^*) = 1 - \frac{1}{2}\int_{\mathbb{R}^2} \{\phi(u)^2 - 2\rho(h^*)\phi(u)\phi(u-\tilde{h}) + \phi(u-\tilde{h})^2\}^{1/2} \mathrm{d}u,$$

where $\rho(h^*)$ is a stationary correlation function and $\tilde{h} = (h^*, h^*)^T$ and $\phi(\cdot)$ is the bivariate Gaussian density function with mean 0 and covariance matrix $\Sigma = \mathrm{diag}(\sigma^2, \sigma^2)$. Note that using a Matérn correlation function given in (7.1.6) with parameters $\theta_1 > 0$

and $\theta_2 > 0$, this function has one extra parameter than $\chi^{BR}(h^*)$, namely $\sigma > 0$.

We chose not to use this parametric form for $\chi(h^*)$, due to the high computational cost required to compute the double integral for each pair of locations. However, we have found that the deformation method described in Section 7.2 appears fairly robust to the choice of $\chi(h^*)$. As Figure E.1.1 in Appendix E shows, the much simpler $\chi^{BR}(h^*)$ can approximate the more complex $\chi^{GG}(h^*)$ very closely for much of $h^* \in \mathbb{R}_+$.

7.2.4 Practical aspects for creating deformations

We now comment on practical aspects of creating the deformations, including choosing a subset of radial basis functions for the thin-plate spline and reducing the chances of producing a non-bijective transformation.

We found that there is no simple robust method for picking the number m, or configuration, of the anchor points used in the deformation splines given in (7.1.3). As detailed in Sampson and Guttorp (1992), there is a trade-off in picking m. Larger values provide "better" deformations, in the sense that the objective function to be calculated is lower and the deformations seem to capture more of the non-stationarity in the process. However, this comes at the price of computational cost, the risk of over-fitting and the phenomenon in which the D-plane folds on to itself. This provides a non-bijective transformation, which is physically unrealistic. Iovleff and Perrin (2004) detail an approach to ensure that the deformation is always bijective through use of a simulated annealing algorithm, with later extensions by Youngman (2020). These approaches add further constraints into the modelling procedure, which we have chosen to avoid. Instead we use a more heuristic approach for avoiding non-bijectivity.

We begin by randomly sampling m_0 initial anchor points with index set given by $I_0 = \{i_1, \ldots, i_{m_0}\}$. There is no single best way to choose I_0; however, we found that ensuring that the anchor points are spread out over the spatial domain helped

to create better deformations. Performing a deformation with I_0 yields parameter estimates $\hat{\boldsymbol{\psi}}_0 = (\hat{b}_1, \hat{b_2}, \hat{\rho}, \hat{\kappa}, \hat{\delta}_4^{(1)}, \hat{\delta}_4^{(2)}, \ldots, \hat{\delta}_{m_0}^{(1)}, \hat{\delta}_{m_0}^{(2)})$. Recall that we have parameters $\delta_i^{(1)}$, $\delta_i^{(2)}$ indexed by $i \geq 4$ as those indexed by $i = 1, 2, 3$ are uniquely determined by the constraints given in (7.1.2). If the deformation for I_0 is bijective, we create a new set of indices $I_1 = \{I_0, i_{m_0+1}\}$, where i_{m_0+1} is sampled from the remaining indices. A deformation is then created using I_1, but with initial parameters in the optimisation program given by $\hat{\boldsymbol{\psi}}_1 = \{\hat{\boldsymbol{\psi}}_0, \delta_{m_0+1}^{(1)} = 0, \delta_{m_0+1}^{(2)} = 0\}$. This ensures that the initial input into the optimisation program creates a deformation that is already bijective. We then continue in this fashion until we have created a deformation using m^* anchor points. Bijectivity is checked by eye.

Using this approach reduces the chances of the D-plane folding as m increases and provides a deformation with m^* anchor points. Here we set m^* as approximately a quarter of the sampling locations as we have not found a clear way to optimize this aspect. Typically this approach can be used for a number of initial index sets. However, in the interest of reducing computational cost, the simulation studies in Section 7.3 are conducted using the same initial index set for each deformation method. We also ensure that the new index sampled at each iteration is consistent across different samples, processes and deformation methods.

7.2.5 Model fitting and selection

To determine whether the deformation has created a process that is more stationary in the extremal dependence structure, and to compare between deformation methods, we look to fitting max-stable and inverted max-stable models to the data using the sampling locations in both the G-plane and the D-plane. In Section 7.1, the computational complexities of the max-stable and inverted max-stable models were discussed. To accommodate for this, we take a pairwise composite likelihood approach and assume independence between pairs (Padoan et al., 2010). The joint distribution for a

Brown-Resnick process is given in (7.1.8) and the joint survival function for an inverted Brown-Resnick process is given in (7.1.12). Note that the former is on standard Fréchet margins, whereas the latter is on standard exponential. To compare between the asymptotically dependent and asymptotically independent structures provided by the two models, we calculate all likelihoods on exponential margins, by first using a site-wise empirical transformation.

Given realisations $\{z_1, \ldots, z_N\}$ from a spatial field, observed at sampling locations $s_1, \ldots, s_d$, the censored composite likelihood is

$$L_{CL}(\lambda, \kappa) = \prod_{i=1}^{N} L_{CL}(\lambda, \kappa; z_i) = \prod_{i=1}^{N} \prod_{k=2}^{d} \prod_{l<k} g_u(z_i(s_k), z_i(s_l); \lambda, \kappa), \tag{7.2.6}$$

where

$$g_u(z_i(s_k), z_i(s_l); \lambda, \kappa) = \begin{cases} f(z_i(s_k), z_i(s_l); \lambda, \kappa) & \text{if } \min(z_i(s_k), z_i(s_l)) > u, \\ \frac{\partial}{\partial z_i(s_k)} F(z_i(s_k), u; \lambda, \kappa) & \text{if } z_i(s_k) > u, z_i(s_l) \leq u, \\ \frac{\partial}{\partial z_i(s_l)} F(u, z_i(s_l); \lambda, \kappa) & \text{if } z_i(s_k) \leq u, z_i(s_l) > u, \\ F(u, u; \lambda, \kappa) & \text{if } \max(z_i(s_k), z_i(s_l)) \leq u, \end{cases} \tag{7.2.7}$$

with u a high threshold and $F(\cdot)$ and $f(\cdot)$ the bivariate joint distribution and density functions for the model. Note that although we set $\lambda = 1$ when producing the deformation, here we treat it as a free parameter. Although the likelihoods give a good indication of the performance of the deformation methods, we use the Composite Likelihood version of the Akaike Information Criterion (CLAIC) for model selection. As given in Varin et al. (2011), the CLAIC is

$$-2\{\log L(\hat{\lambda}, \hat{\kappa}) - \operatorname{tr}(J(\hat{\lambda}, \hat{\kappa}) H^{-1}(\hat{\lambda}, \hat{\kappa}))\}, \tag{7.2.8}$$

where $(\hat{\lambda}, \hat{\kappa})$ are the maximum likelihood estimates from (7.2.6), $H(\cdot)$ is the Hessian

matrix and $J(\cdot)$ is the variance of the score function, i.e.

$$J(\hat{\lambda}, \hat{\kappa}) = \text{var}\, \nabla \log L_{CL}(\hat{\lambda}, \hat{\kappa}) = \text{var} \sum_{i=1}^{N} \nabla \log L_{CL}(\hat{\lambda}, \hat{\kappa}; z_i).$$

In practice, we estimate $J(\cdot)$ by using numerical methods to find $\Delta_i = \nabla \log L_{CL}(\hat{\lambda}, \hat{\kappa}; z_i)$, and then estimating the variance of the score function by setting a block of length $b < N$ and computing

$$\hat{J}(\hat{\lambda}, \hat{\kappa}) = \frac{N}{b} \times \text{var} \sum_{i=1}^{b} \Delta_i, \ldots, \sum_{i=N-b+1}^{N} \Delta_i. \tag{7.2.9}$$

The block sizes are chosen such that each block of data is more reasonably assumed approximately independent. This is usually specific to the data and will be given alongside any results.

7.3 Simulation study

We conduct three simulation studies to illustrate the efficacy of the deformation framework for modelling extremal dependence of non-stationary spatial processes. These studies are designed to highlight the following:

- When fitting a stationary model to the extremal dependence of non-stationary spatial data, using a deformation method will improve the fit when compared to using the original sampling locations in the G-plane;
- The deformation methodology described in Section 7.2.1 is more effective than the original Smith (1996) method when modelling non-stationary extremal dependence, as the latter is tailored towards modelling dependence in the body of the data rather than the extremes;
- It is often necessary to use a deformation method that is tailored explicitly to

extremal dependence, rather than dependence throughout the body; especially for processes that exhibit different degrees of non-stationarity throughout their extremal and central dependence structures.

In order to illustrate these points, we consider five different processes. These processes are chosen as they each exhibit different behaviour in their respective extremal dependence structures. In Section 7.3.1, we consider two processes: a non-stationary Brown-Resnick process and a non-stationary inverted Brown-Resnick process. In Section 7.3.2, we consider two more processes which are both mixtures of stationary and non-stationary processes. We term these max-mixture process and one exhibits asymptotic dependence whilst the other exhibits asymptotic independence. A final process is considered in Section 7.3.3, which is an asymptotically independent Gaussian mixture process.

For each setting, we begin with a sample of 1000 realisations of a spatial process. For this sample, we create four separate deformations using the procedure set out in Section 7.2.4. The first two deformations are created using the approach detailed in Section 7.2.1; with χ^{BR} from (7.2.3) and χ_q^{IBR} from (7.2.5) as the dependence measures used in the objective function in (7.2.1). The latter two are correlation-based deformation methods: one of these is the original Smith (1996) methodology, while the other method replaces $\chi(h_{ij}^*)$ in (7.2.1) with pairwise correlation $\rho(h_{ij}^*)$ as the dependence measure, and replaces the theoretical $\chi(h^*)$ function with the stationary Matérn correlation function detailed in (7.1.6). Note that in both of the latter two methods, correlation is estimated on a Gaussian marginal scale, and for the former two methods, we set $q = 0.9$ in (7.2.2) and (7.2.5).

As detailed in Section 7.2.5, we evaluate the efficacy of each of the four deformations by fitting a model to the extremal dependence of the sample. We fit the same dependence model five times: once using the sampling locations in the original G-plane and then once for each of the respective D-plane sampling locations given from

the four deformations. For each fitted model, we calculate the CLAIC given in (7.2.8). Ordering of the CLAIC allows us to determine which deformation method (if any) was the most effective in accounting for the non-stationarity in that sample. As the underlying process from which the sample is drawn is known, we fit a stationary extremal dependence model of an appropriate class. That is, for processes that are asymptotically dependent, we fit a stationary Brown-Resnick model, and for processes that are asymptotically independent, we fit a stationary inverted Brown-Resnick model.

This procedure is repeated for 50 different samples of a single process. In this simulation study, each deformation for each sample is created using the same anchor points. For each sample, we determine which deformation method was the most effective and the proportion of times this occurred over all samples is reported, with the results in Tables 7.3.1, 7.3.2 and 7.3.3. These results show that stationary dependence models for non-stationary spatial processes routinely provide a better fit if the deformation methodology is used as a preprocessing step. We also show that the original Smith (1996) deformation is outperformed by our extensions.

7.3.1 Non-stationary Brown-Resnick and inverted Brown-Resnick process

The first setting we consider consists of replications of a non-stationary Brown-Resnick, and inverted Brown-Resnick, process sampled at 64 equally spaced locations on $[-1, 1] \times [-1, 1]$. We use a non-stationary variogram in the exponent function in (7.2.3) to ensure that $\chi(h_{ij})$ is not simply a function of distance. In the context of non-stationary Gaussian processes, Fouedjio et al. (2015) propose a semivariogram of the form $\gamma^*(s_i, s_j)$ where

$$\gamma^*(s_i, s_j) = \gamma(\|\psi(s_j) - \psi(s_j)\|), \tag{7.3.1}$$

and

$$\psi(s) = o + (s - o)\|s - o\|$$

is a radial basis function with some centre point o and $\gamma(\cdot)$ is the stationary and isotropic semivariogram given in (7.2.4). The use of the radial basis function $\psi(s)$ within this semivariogram causes pairs that are closer to o to be more strongly dependent than those pairs that are further away. From (7.2.3) and (7.2.4), the Brown-Resnick process with this semivariogram has theoretical $\chi(s_i, s_j)$ given by

$$\chi(s_i, s_j) = 2 - 2\Phi\left\{\frac{\|\psi(s_i) - \psi(s_j)\|^{\kappa/2}}{\lambda^{\kappa/2}\sqrt{2}}\right\}, \tag{7.3.2}$$

for locations s_i, s_j and $\kappa \in (0, 2], \lambda > 0$. For this study, we take the centre o to be the origin and use scale and shape parameters $\lambda = 2$ and $\kappa = 0.8$ in (7.3.2). To illustrate the process a high resolution realisation is given in Figure E.2.1. Simulations are produced using the method of Dieker and Mikosch (2015).

Process (G-plane)	Fitted Model (D-plane)	Deformation Method	Proportion of lowest CLAIC	
Non-stationary Brown-Resnick	Stationary Brown-Resnick	None	0	
		χ^{BR}	0.22	0.34
		χ_q^{IBR}	0.12	
		ρ	0.44	0.66
		Smith (1996)	0.22	
Non-stationary Inverted Brown-Resnick	Stationary Inverted Brown-Resnick	None	0	
		χ^{BR}	0.24	0.56
		χ_q^{IBR}	0.32	
		ρ	0.28	0.44
		Smith (1996)	0.16	

Table 7.3.1: Proportion of lowest CLAIC estimates provided by fitting models to deformations for 50 realisations of non-stationary Brown-Resnick and inverted Brown-Resnick processes. The CLAIC has been estimated with a block size of $b = 1$, corresponding to temporal independence. Composite likelihoods are estimated with the threshold in (7.2.7) as the 90% empirical quantile, which is also used for estimating $\chi(h_{ij}^*)$ in (7.2.2).

Table 7.3.1 contains some interesting results. Most notably, in all cases a deformation has aided in model fitting when compared to using the original simulation grid. For both the max-stable, and inverted max-stable, cases, improvements on the efficacy of the original Smith (1996) method are made by utilising the Frobenius norm in the objective function. However, it is not entirely clear whether use of an extremal dependence measure for creating deformations is necessary in this case. We often found that deforming the space using measures for dependence throughout the distribution created better deformations than those using extremal dependence measures. We believe that this is because the variance of the estimator for $\rho(h_{ij}^*)$ is much lower than that of $\chi(h_{ij}^*)$, as we use all of the data to estimate correlation, and that there are strong similarities in patterns of spatial non-stationarity for the central- and extremal-

dependence structures of this process. We next consider other processes with more complicated dependence structures.

7.3.2 Max-mixture process

We now consider the hybrid dependence model, detailed in full by Wadsworth and Tawn (2012). Let $X(s)$ be a max-stable process and $Y(s)$ an asymptotically independent spatial process, both with standard Fréchet margins. For $\omega \in [0,1]$, $H(s) = \max\{\omega X(s), (1-\omega)Y(s)\}$ is an asymptotically dependent spatial process with standard Fréchet margins. In particular, we take $X(s)$ to be the non-stationary Brown-Resnick process detailed in Section 7.3.1 and $Y(s)$ to be a marginally transformed stationary Gaussian process with the Matérn correlation structure detailed in (7.1.6).

It can be shown that the theoretical $\chi(h_{ij})$ values for $H(s)$ are the same as for $X(s)$, but multiplied by ω. There is no closed form for the correlation for $H(s)$ on the Gaussian scale. Computationally, it can be shown that it is a mixture of the correlation from both $X(s)$ and $Y(s)$. As such, we would expect the extremal dependence and central dependence of $H(s)$ to be mixtures of those coming from $X(s)$ and $Y(s)$, with different amounts of mixing occurring for both. We set ω to be 0.3 and take $(\theta_1, \theta_2) = (1, 1.2)$ in (7.1.6).

By construction of $H(s)$, taking its reciprocal creates an asymptotically independent process on standard exponential margins, as with the inverted max-stable process. As in Section 7.3.1, the simulation study is repeated separately for the asymptotically dependent and asymptotically independent mixtures. The results are given in Table 7.3.2.

Process (G-plane)	Fitted Model (D-plane)	Deformation Method	Proportion of lowest CLAIC	
Asymptotically-dependent Max-mixture	Stationary Brown-Resnick	None	0.06	
		χ^{BR}	0.14	0.78
		χ_q^{IBR}	0.64	
		ρ	0.16	0.16
		Smith (1996)	0	
Asymptotically-independent Max-mixture	Stationary Inverted Brown-Resnick	None	0	
		χ^{BR}	0.42	0.90
		χ_q^{IBR}	0.48	
		ρ	0.06	0.10
		Smith (1996)	0.04	

Table 7.3.2: Proportion of lowest CLAIC estimates provided by fitting models to deformations of 50 realisations of asymptotically dependent and asymptotically independent max-mixture processes. The CLAIC has been estimated with a block size of $b = 1$, corresponding to temporal independence. Composite likelihoods are estimated with the threshold in (7.2.7) as the 90% empirical quantile, which is also used for estimating $\chi(h_{ij}^*)$ in (7.2.2).

In contrast to the results given in Table 7.3.1, Table 7.3.2 shows a clearer need for an extremal dependence-based approach when creating deformations for a process that exhibits more complicated dependence structures. Here this max-mixture process is designed to represent a process with a mixture of stationarity in both the extremal dependence and dependence throughout the distribution. We now consider a process that has non-stationary extremal dependence, but is nearly stationary in the body.

7.3.3 Gaussian mixture process

With previous simulations, we found it is sometimes sufficient to simply use measures of central dependence when deforming the spatial domain to create a process with a more stationary extremal dependence structure. This is because the central- and

extremal-dependence structures of these processes are closely related and using either approach typically creates similar deformations. In applications, we may find that these structures are not so closely related. As such, we are motivated to consider a process that is designed to have completely different dependence in the body to the tails.

Let $Y_{\mathrm{S}}(s), Y_{\mathrm{NS}}(s)$ be stationary and non-stationary Gaussian processes, respectively, each with standard Gaussian margins. We then consider the process

$$Y^*(s) = \begin{cases} Y_{\mathrm{S}}(s), & \text{if } \Phi(Y(s_0)) \leq p \\ Y_{\mathrm{NS}}(s), & \text{if } \Phi(Y(s_0)) > p \end{cases}, \tag{7.3.3}$$

where $s_0 \in S$ is a fixed location, $\Phi(\cdot)$ is the standard Gaussian cdf, and $p \in [0,1]$ is a probability. By specifying $Y^*(s)$ in this manner, we create a process with an extremal dependence structure determined mostly by the correlation structure of Y_{NS} and with dependence through the body determined mostly by Y_{S}. Simulation of this process is simple; we draw $Y(s_0) \sim N(0,1)$ and then simulate the rest of the field conditioning on that value and whether $\Phi(Y(s_0)) \leq p$ or $\Phi(Y(s_0)) > p$.

For this particular study, we use replications of this Gaussian mixture sampled at 81 equally spaced locations in $[-1,1]\times[-1,1]$. We take s_0 to be the origin and $p = 0.9$. Both Y_{S} and Y_{NS} are specified to have the Matérn correlation structure given in (7.1.6), with respective parameter sets $\boldsymbol{\theta}^{(\mathrm{S})} = (\theta_1^{(\mathrm{S})}, \theta_2^{(\mathrm{S})})$ and $\boldsymbol{\theta}^{(\mathrm{NS})} = (\theta_1^{(\mathrm{NS})}, \theta_2^{(\mathrm{NS})}, o)$. Note that $\boldsymbol{\theta}^{(\mathrm{NS})}$ contains an extra parameter as we use the difference of the radial basis functions given in (7.3.1) and detailed by (Fouedjio et al., 2015) as a measure of pairwise distance, rather than Euclidean distance. The parameters for this study are set to $\boldsymbol{\theta}^{(\mathrm{S})} = (2,1)$ and $\boldsymbol{\theta}^{(\mathrm{NS})} = (2, 0.8, (0,0))$. Results are given in Table 7.3.3.

Process (G-plane)	Fitted Model (D-plane)	Deformation Method	Proportion of lowest CLAIC	
Gaussian Mixture	Stationary Inverted Brown-Resnick	None	0	
		χ^{BR}	0.08	1
		χ_q^{IBR}	0.92	
		ρ	0	0
		Smith (1996)	0	

Table 7.3.3: Proportion of lowest CLAIC estimates provided by fitting models to deformations of 50 realisations of the Gaussian mixture process, see (7.3.3). The CLAIC has been estimated with a block size of $b = 1$, corresponding to temporal independence. Composite likelihoods are estimated with the threshold in (7.2.7) as the 90% empirical quantile, which is also used for estimating $\chi(h_{ij}^*)$ in (7.2.2).

Table 7.3.3 highlights a clear need for extremal dependence-based methods when creating deformations for processes that have different patterns of non-stationarity in their central- and extremal dependence structures. In contrast to the results given in the previous studies, here using $\chi(h_{ij}^*)$ or $\chi_q(h_{ij}^*)$ is always favoured.

7.4 Case studies

We present two case studies using our deformation methodology. In both cases, we follow the procedure set out in Section 7.2.4. However, as we consider relatively large spatial domains we use Great Earth distance in place of Euclidean distance for h and h^*. We consider 30 different initial index sets, taking the best deformation over all sets. Here we define the best deformation to be that which provides the lowest objective value in (7.2.1) whilst remaining a bijective mapping. When using extremal dependence measures, we focus on deformations based on χ^{BR} only, following the justification in Section 7.2.2. We then fit max-stable and inverted max-stable models to the data in the G-plane and D-plane, comparing the model fits using CLAIC estimates. For both studies, all pairs of sampling locations are used in model fitting

and the block size in (7.2.9) corresponds to a season. We propose two diagnostics for scrutinising the model fits and deformations.

7.4.1 Australian summer temperatures

Data consist of daily summer (DJF) maximum near-surface air temperatures taken from the HadGHCND global gridded dataset (Caesar et al., 2006) and interpolated to 72 grid point locations covering Australia, for the period 1957-2014. Previous analysis of this data has been conducted using the multivariate conditional extremes model, detailed in Section 7.1.2 (Winter et al., 2016) and its spatial extension (Wadsworth and Tawn, 2019). Figure 7.4.1 shows the original sampling locations and estimated pairwise $\chi(h_{ij})$ against distances. We estimate $\chi(h_{ij})$ by setting $q = 0.98$ in (7.2.2). The deformation was produced using $m^* = 18$, i.e. a quarter of the original sampling locations. These are presented as the blue points on Figures 7.4.1 and 7.4.2, where the latter figure depicts the sampling locations in the D-plane. Figure 7.4.2 also presents $\hat{\chi}(h^*_{ij})$ against distance in the deformed space. We observe that the deformation has created a process that appears to be much more stationary with regards to the $\chi(h^*_{ij})$ estimates in the new coordinates.

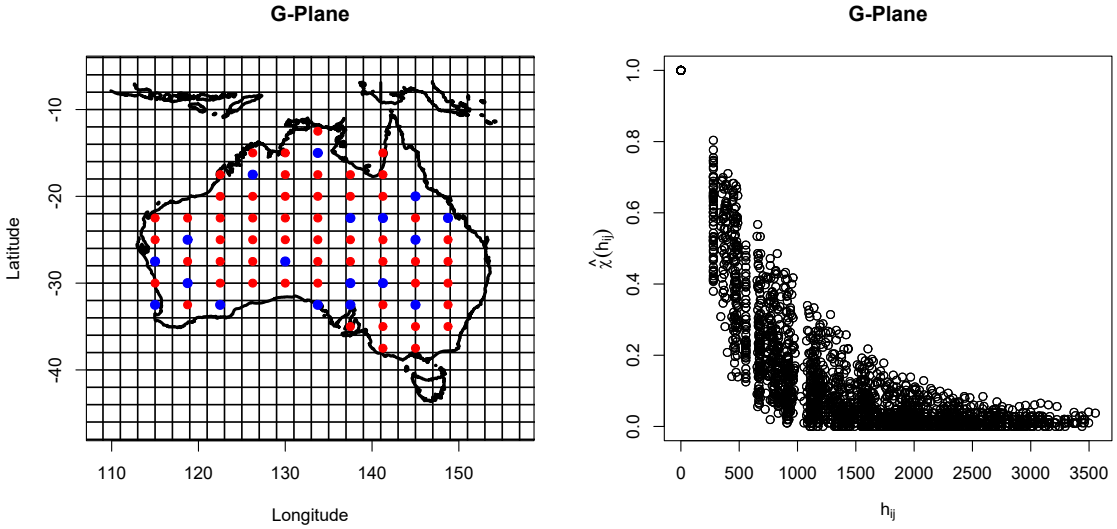

Figure 7.4.1: Australia summer temperatures. Left: the original 72 sampling locations. The blue points are the anchor points used for the thin-plate splines. Right: empirical $\chi(h_{ij})$ measures against distance (km). Estimates $\hat{\chi}(h_{ij})$ are calculated above a threshold given by the 98% empirical quantile.

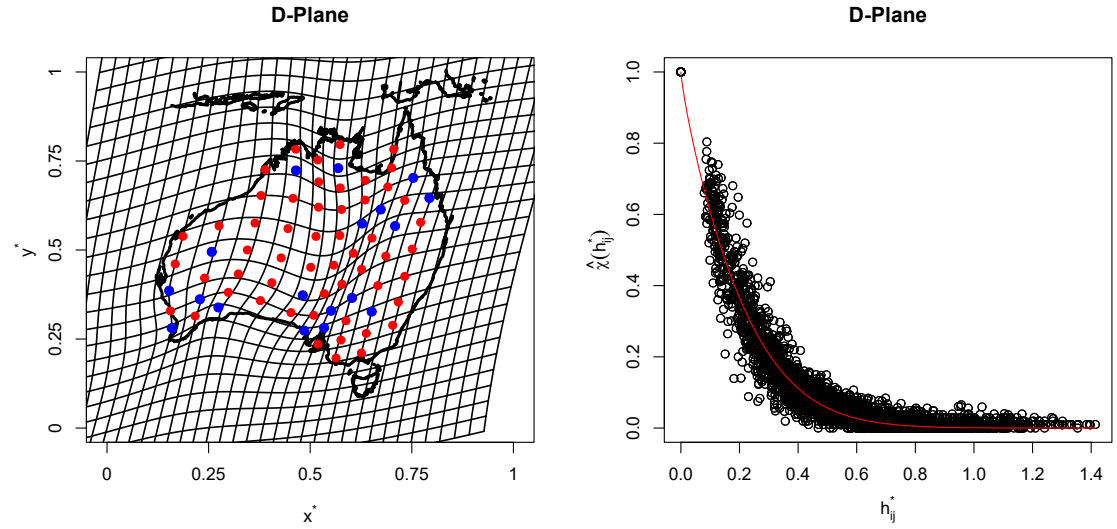

Figure 7.4.2: Australia summer temperatures. Left: the 72 sampling locations in the D-plane. The blue points are the anchor points used for the thin-plate splines. The coordinates have been scaled to $[0, 1] \times [0, 1]$, which equals the aspect ratio of the left plot in Figure 7.4.1. Right: empirical $\chi(h^*_{ij})$ measures against distance in the D-plane. The red line gives the fitted function $\chi(h^*)$ used in the deformation.

	Model	Negative Composite Log-Likelihood ($\times 10^6$)	$(\hat{\kappa}, \hat{\lambda})$ (2 d.p.)	CLAIC ($\times 10^7$)
G-Plane	IMSP*	3.078	(2.00, 1048.20)	6.157
	MSP	3.078	(1.59, 358.30)	6.157
D-Plane	IMSP*	3.074	(2.00, 2.61)	6.148
	MSP	3.073	(1.71, 0.95)	**6.146**

Table 7.4.1: Model parameters and diagnostics for the Australian summer temperatures data. Composite likelihoods are estimated with the threshold in (7.2.7) as the 98% empirical quantile. (* estimated using Smith process likelihood). CLAIC and negative composite log-likelihood estimates are given to four significant figures.

The fits of the max-stable and inverted max-stable models are summarised in Table 7.4.1. The CLAIC estimates suggest that a max-stable model is more appropriate for the data. This becomes even more apparent when we consider that fitting an inverted Brown-Resnick model yields an inverted Smith model as the best fit. These processes are typically quite smooth and often provide unrealistic representations of actual data. However, we note that when naively fitting models on the G-plane, the inverted Smith model provided the lowest CLAIC estimate. This is further evidence that non-stationarity in this data should be incorporated into the modelling procedure.

We use two diagnostics to scrutinise the deformation and the model fit. As our deformation method is tailored to $\chi(h^*_{ij})$, we seek to use other extremal dependence measures to verify that the resulting deformation is not subject to overfitting. To do this, the conditional extremes model described in Section 7.1.2 is fitted pairwise and the parameter estimates are used to calculate the conditional expectation of one variable when the other variable is at the modelling threshold u, taken as the 98% quantile of the marginal distribution. For each pair, $(X(s_i), X(s_j)), i \neq j$, we have

$$\mathbb{E}\left[X(s_j)|X(s_i) = u\right] = \hat{\alpha}u + u^{\hat{\beta}}\hat{\mu},$$

where $(\hat{\alpha}, \hat{\beta}, \hat{\mu})$ are the maximum likelihood estimates for the model. For a stationary and isotropic process, we would expect this measure to be a smooth function of Euclidean distance. The conditional expectation is plotted against distance for both the process on the G-plane and the D-plane.

A second diagnostic is used to evaluate the best model fit in the D-plane. As we have used $\chi(h_{ij}^*)$ to create the deformations, we compare the theoretical triple-wise χ, which we denote $\chi(s_i^*, s_j^*, s_k^*) = \chi(s_i, s_j, s_k)$, from the model fits against empirical estimates. The triple-wise χ is defined as

$$\begin{aligned}\chi(s_i^*, s_j^*, s_k^*) &= \lim_{q\to 1} \Pr[F_i\{Z(s_i^*)\} > q, F_j\{Z(s_j^*)\} > q | F_k\{Z(s_k^*)\} > q] \\ &= \lim_{q\to 1} \Pr[F_i\{Z(s_i)\} > q, F_j\{Z(s_j)\} > q | F_k\{Z(s_k)\} > q] = \chi(s_i, s_j, s_k)\end{aligned}$$

for $i \neq j \neq k$. For a Brown-Resnick process, the theoretical value for this measure is

$$\chi(s_i^*, s_j^*, s_k^*) = 3 - V_2(1,1;i,j) - V_2(1,1;i,k) - V_2(1,1;j,k) + V_3(1,1,1),$$

where $V_2(\cdot,\cdot;l,m)$ is the pairwise exponent given in (7.1.10) and $V_3(\cdot,\cdot,\cdot)$ is the triple-wise exponent measure, for which the parametric form is given in Huser and Davison (2013); recall that if the process is stationary, both of these are functions of Euclidean distance. A similar parametrisation can be given for $\chi_q(s_i^*, s_j^*, s_k^*)$ for an inverted Brown-Resnick process, which is $\chi_q(s_i^*, s_j^*, s_k^*) = (1-q)^{V_3(1,1,1)-1}$.

Standard errors for empirical estimates of $\chi_q(s_i^*, s_j^*, s_k^*)$ are estimated using a stationary bootstrap (Politis and Romano, 1994). We begin by drawing a random block size B from a geometric distribution with mean K. The bootstrap sample for locations $s_i, s_j, s_k,\ i \neq j \neq k$ is built by drawing a random starting time τ and creating a block of observations

$$\{\mathbf{z}_\tau^*, \ldots, \mathbf{z}_{\tau+B-1}^*\}, \quad \text{where} \quad \mathbf{z}_t^* = \{z_t(s_i), z_t(s_j), z_t(s_k)\},$$

which we add to the bootstrap sample. This procedure is repeated and the bootstrap is built up iteratively until it has length n. We then estimate $\chi(s_i^*, s_j^*, s_k^*)$ for that sample and repeat for a number of samples. When choosing locations to compare empirical and theoretical values of $\chi(s_i^*, s_j^*, s_k^*)$, we take advantage of the gridded structure of the coordinates in the G-plane, and ensure that each set of points share roughly the same configuration and pairwise distances. This is used to evaluate the stationarity of the dependence structure on the original G-plane, as we would expect the empirical values of $\chi(s_i^*, s_j^*, s_k^*)$ to be consistently similar across sets of locations with the same configuration.

Diagnostics for the deformations and best model fit are given in Figure 7.4.3. For the estimation of $\chi(s_i^*, s_j^*, s_k^*)$, 30 sets of three adjacent locations along the north/south transect in the G-plane are randomly selected and a stationary bootstrap with mean block size $K = 14$ and 1000 samples is used to create 95% confidence intervals for the empirical estimates of $\chi(s_i^*, s_j^*, s_k^*)$. Empirical estimates of $\chi(s_i^*, s_j^*, s_k^*)$ are calculated above the 98% quantile. The right panel of Figure 7.4.3 displays estimates for the conditional expectation from the conditional extremes model, where distances are normalised so that the average distance is equal for both the values in the G-plane and the D-plane.

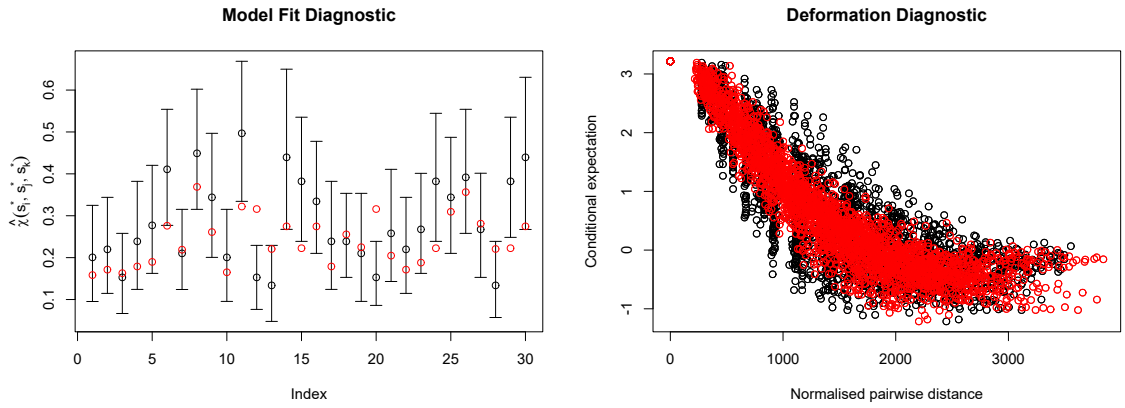

Figure 7.4.3: Australian summer temperatures diagnostics. Left: estimates of $\chi(s_i^*, s_j^*, s_k^*)$ (black dots) and 95% confidence intervals using the stationary bootstrap. Red dots are the respective theoretical values suggested by the model fit. Right: conditional expectation from conditional extremes model. Red points denote estimates for the process on the D-plane; black points are those on the G-plane.

The diagnostic based on $\chi(s_i^*, s_j^*, s_k^*)$ from Figure 7.4.3 suggests that a max-stable model is a reasonable fit for the data in the deformed space, as the patterns of the theoretical $\chi(s_i^*, s_j^*, s_k^*)$ values follow the empirical estimates. The large variability in the bootstrap estimates across sets of locations with similar configurations suggests that the process on the original plane is highly non-stationary. Estimates from the conditional extremes model provide further evidence that the deformation has produced something more stationary with regards to the dependence structure, especially at smaller distances. The use of a measure for extremal dependence that is not used for fitting lends credibility to the $\chi(h_{ij}^*)$ plot in Figure 7.4.2 and suggests that the deformation has worked well.

7.4.2 UK precipitation rate

Data consist of hourly precipitation rate (mm/day) observed at locations on two 10×10 grids; the first is centred in Snowdonia, Wales and the second is centred in the Scottish Highlands. Observations are taken from the UK climate projections 2018 (UKCP18) (Lowe et al., 2018) which contain values produced at hourly intervals on

$2.2 \times 2.2\text{km}^2$ grid boxes between the years 1980 and 2000. We have treated the centre of each grid box as a sampling location and we take every fifth grid box to create the 10×10 grid of sampling locations. Observations are aggregated to 12-hr intervals, beginning at 12pm, and to remove the seasonal effect often observed in precipitation data, we have taken only winter observations (DJF). This leaves 3600 observations at each sampling location.

Figure 7.4.4 shows both sets of original sampling locations and their respective estimates of $\chi(h_{ij})$ against distances. In both cases, we estimate $\chi(h_{ij})$ by setting $q = 0.95$ in (7.2.2). Both deformations are produced using $m^* = 25$ and these are presented as the blue points in Figure 7.4.4. Figure 7.4.5 presents both deformations and estimates of $\hat{\chi}(h^*_{ij})$ against distance in the respective deformed spaces. We observe that both deformations have created a process that appears to be much more stationary with regards to their respective $\chi(h^*_{ij})$ estimates in the new coordinates. In both cases, deformations are more prominent around areas of higher elevation.

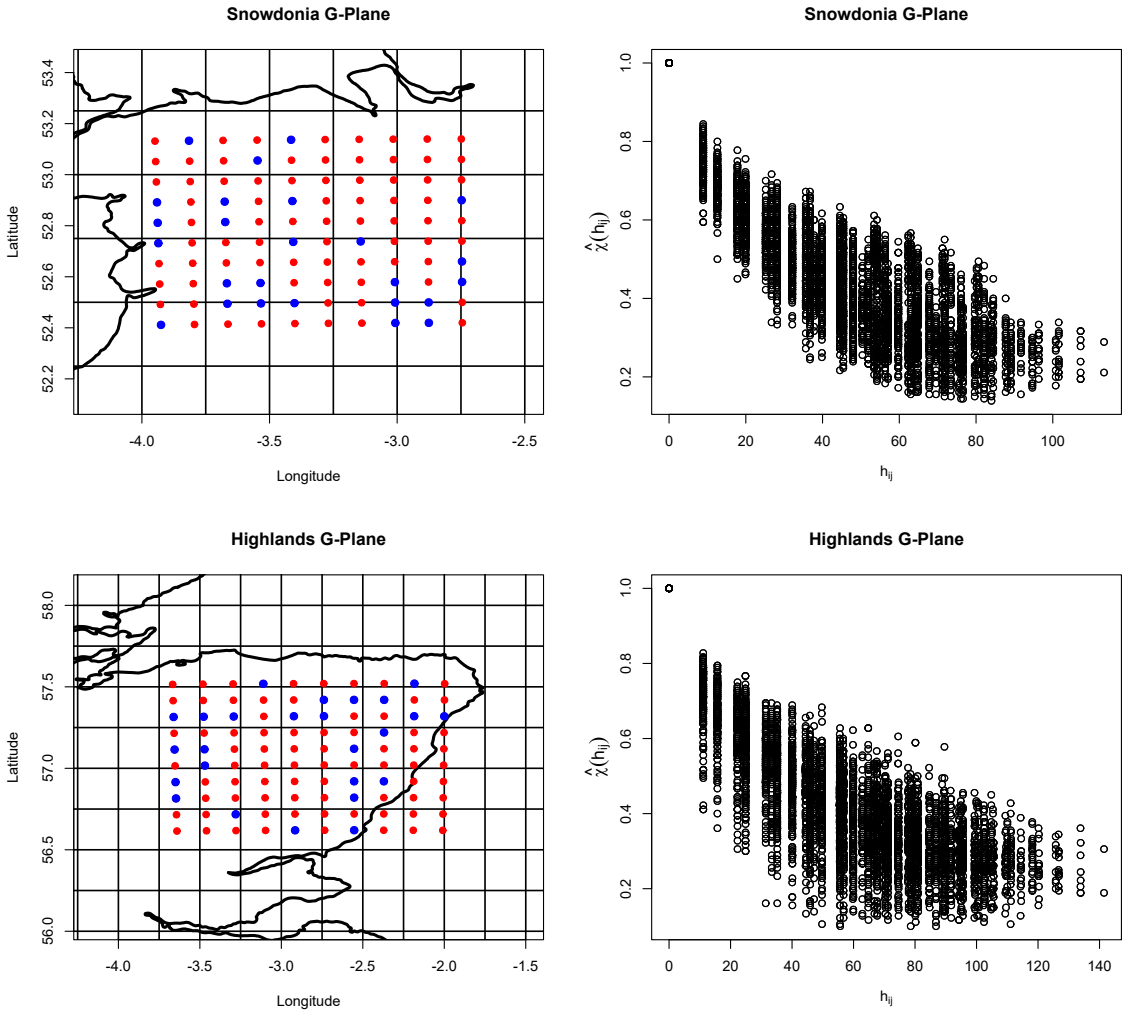

Figure 7.4.4: Top row: Snowdonia. Bottom row: Scottish Highlands. Left: the original 100 sampling locations. The blue points are the anchor points used for the thin-plate splines. Right: empirical $\chi(h_{ij})$ measures against distance (km) in the respective G-planes. Estimates $\hat{\chi}(h_{ij})$ are calculated above a threshold given by the 95% empirical quantile.

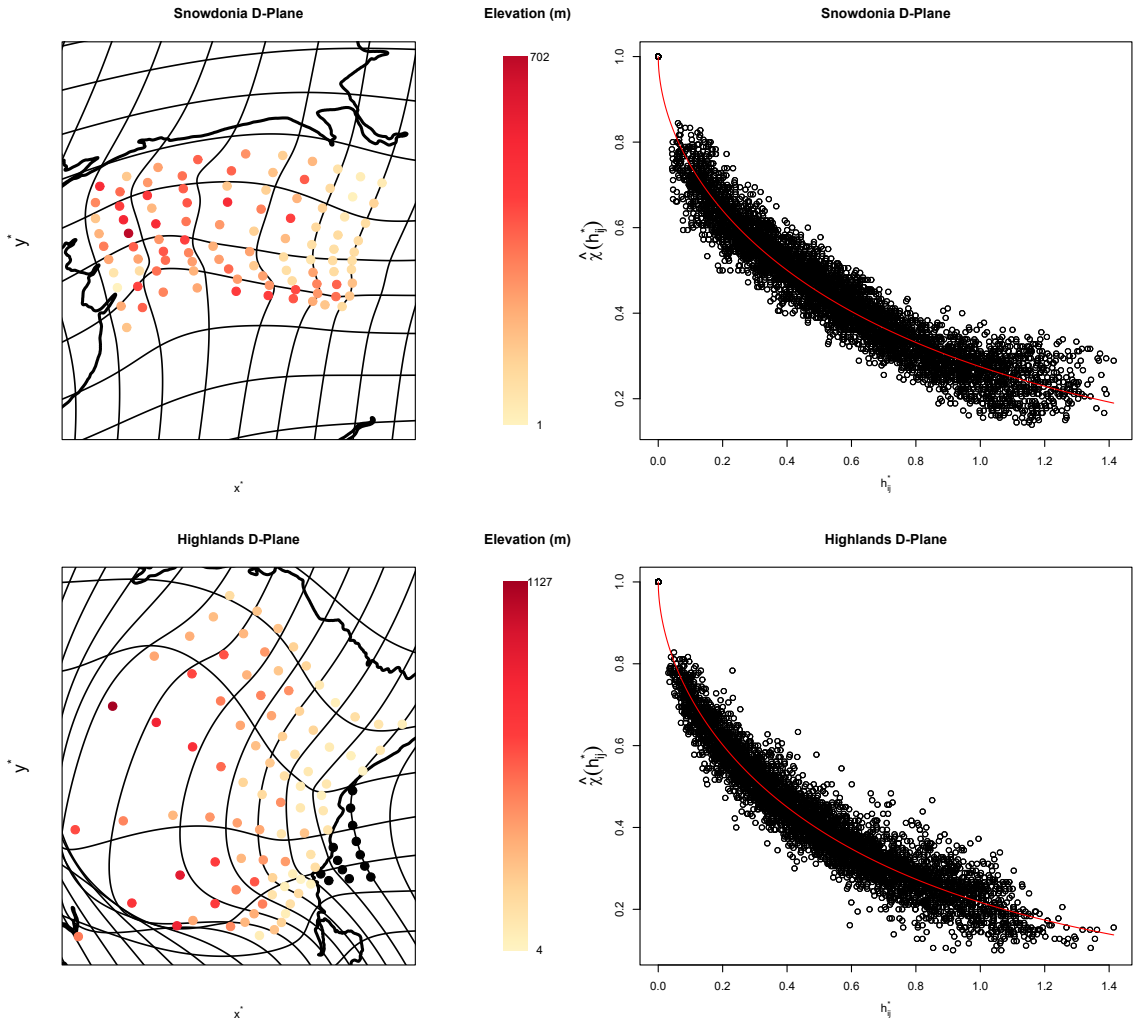

Figure 7.4.5: Top row: Snowdonia. Bottom row: Scottish Highlands. Left: the 100 sampling locations in their respective D-planes. The points are coloured such that darker points correspond to sampling locations with higher elevation and black points correspond to locations over sea. The coordinates have been scaled to $[0, 1] \times [0, 1]$, which equals the aspect ratio of the left plots in Figure 7.4.4. Right: empirical $\chi(h^*_{ij})$ measures against distance in the D-plane. The red line gives the fitted function $\chi(h^*)$ used in the deformation.

		Model	Negative Composite Log-Likelihood ($\times 10^6$)	$(\hat{\kappa}, \hat{\lambda})$ (2 d.p.)	CLAIC ($\times 10^7$)
Snowdonia	G-Plane	IMSP	8.023	(1.40, 111.84)	1.605
		MSP	8.050	(1.00, 25.96)	1.610
	D-Plane	**IMSP**	8.011	(1.29, 3.33)	**1.602**
		MSP	8.037	(0.93, 0.69)	1.607
Highlands	G-Plane	IMSP	8.099	(1.25, 143.77)	1.620
		MSP	8.124	(0.87, 27.37)	1.625
	D-Plane	**IMSP**	8.076	(1.30, 3.34)	**1.615**
		MSP	8.099	(0.93, 0.69)	1.620

Table 7.4.2: Model parameters and diagnostics for the UK precipitation data. Composite likelihoods are estimated with the threshold in (7.2.7) as the 95% empirical quantile. CLAIC and negative composite log-likelihood estimates are given to four significant figures.

Table 7.4.2 summarises the fits for the Brown-Resnick and inverted Brown-Resnick models for both sets of sampling locations. The CLAIC estimates in Table 7.4.2 suggest that an inverted max-stable model is the most appropriate for both the Snowdonia and Highlands data. Both see improved fits using the sampling locations mapped to the respective D-planes. In Figures 7.4.6 and 7.4.7, we present diagnostics for the deformations and best model fits using the same measures described in Section 7.4.1. As the best fitting model for both datasets is the inverted Brown-Resnick process, Figure 7.4.6 compares empirical estimates and model-based values of $\chi_q(s_i^*, s_j^*, s_k^*)$ with $q = 0.95$. Confidence intervals for the empirical estimates of $\chi_q(s_i^*, s_j^*, s_k^*)$ are calculated by randomly selecting 30 sets of three adjacent locations along the east/west transect and a using stationary bootstrap with mean block size $K = 14$ and 1000 samples. For the diagnostic given in Figure 7.4.7, the 95% quantile is used for fitting the conditional extremes model and we plot the pairwise conditional expectation

estimates against distance. Distances are normalised so that the average distance is equal for both the values in the G-plane and the D-plane.

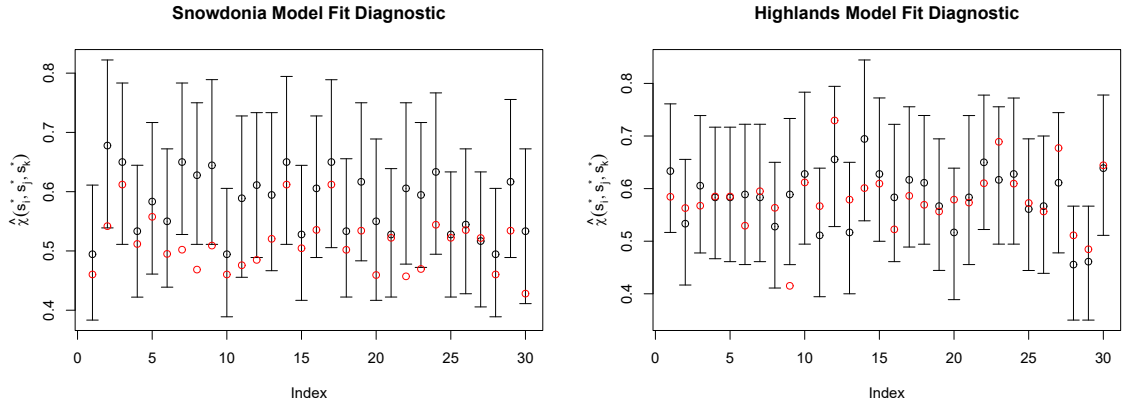

Figure 7.4.6: UK precipitation model fit diagnostics. Estimates of $\chi_q(s_i^*, s_j^*, s_k^*)$ (black dots) with $q = 0.95$ and 95% confidence intervals using the stationary bootstrap. Red dots are the respective theoretical values suggested by the model fits. Left: Snowdonia. Right: Highlands.

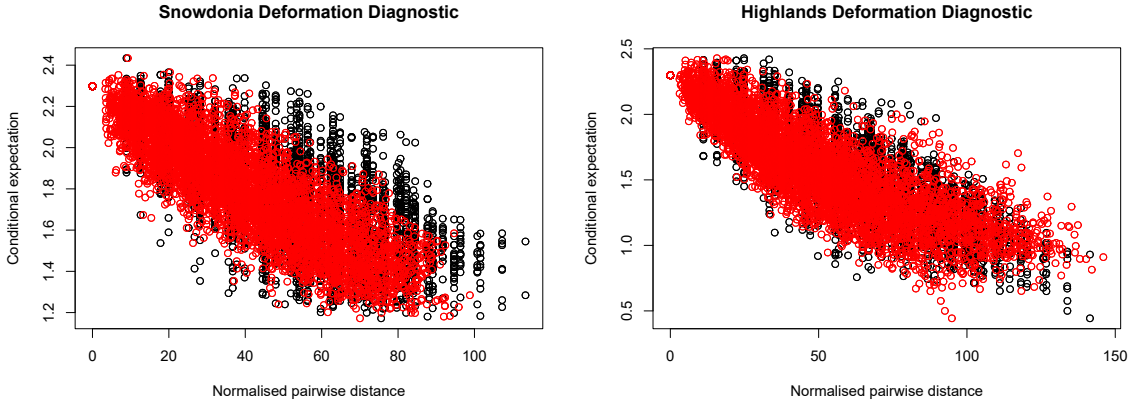

Figure 7.4.7: UK precipitation deformation diagnostics. Conditional expectation from conditional extremes model. Red points denote estimates for the process on the D-plane; black points are those on the G-plane. Left: Snowdonia. Right: Highlands.

Figure 7.4.6 shows that the inverted max-stable model gives a relatively good fit to the extremal dependence of both datasets with sampling locations mapped to their respective D-planes, but the fit appears better for the Scottish Highlands. The low variability in the $\chi_q(s_i^*, s_j^*, s_k^*)$ estimates suggests that the original process may not be

highly non-stationary. The pairwise conditional expectation estimates in Figure 7.4.7 suggest that both deformations have produced a more stationary process, albeit more so in the case of the Snowdonia D-plane. The small change in the Highlands estimates may suggest that overfitting to the $\chi(h_{ij})$ values has occurred, especially when compared to the Snowdonia estimates. This may also explain the stronger agreement of the $\chi_q(s_i^*, s_j^*, s_k^*)$ measures in Figure 7.4.7. To investigate the possibility of overfitting, we recreated the diagnostic using deformations created with fewer anchor points, but this did not show any improvements.

7.5 Discussion

In this paper, we presented a simple yet effective approach to modelling non-stationary extremal dependence. This approach extends that of Sampson and Guttorp (1992) and Smith (1996) to be applicable for modelling extremal dependence, rather than dependence throughout the body. We do this by replacing the objective function in these methods with the Frobenius norm of the difference between empirical, and theoretical, pairwise dependency matrices, with the theoretical measures coming from a stationary dependence model. Although most of our focus is on $\chi(h_{ij}^*)$ as the dependence measure, we have also shown that this is easily replaced by other measures, such as $\chi_q(h_{ij}^*)$ and correlation. Model selection is carried out using pairwise composite likelihoods and CLAIC estimation and we propose diagnostics for evaluating these model fits.

We presented two case studies; in each scenario, we showed that when modelling the extremal dependence of the data using stationary models, better fits are provided using our methodology. Here we have fit very simple models to the data. However, in practice these deformations may be used as a pre-processing step to reveal covariates or orography that can be incorporated into the modelling procedure. Two diagnostics

were introduced and used to provide evidence that our approach has produced a process which is more stationary with regards to the extremal dependence.

As with many areas of extreme value analysis, there is a bias-variance trade-off present when estimating χ_q. Using values of q closer to 1 puts greater focus on extremal dependence at the expense of increased variance of the estimator. In Sections 7.3 and 7.4, we choose q close to 1 whilst preserving some initial spatial structure observed in the χ estimates. However, if q is too high then it is possible that any structure is masked by the high variability of the estimators and the deformation methodology is likely to fail in such circumstances. We have not considered the effect of estimator variability on the deformation, but note this could form a future research direction.

A further issue that could be considered is the possible non-bijectivity of the mapping used in the deformation. We detail an approach to reduce this in Section 7.3, however, this method is not particularly robust. Bijectivity of deformations must be checked by eye which can become cumbersome when a large number are produced. To avoid this necessary supervision, the G-plane can be represented as a Delaunay triangulation, see Iovleff and Perrin (2004) and Youngman (2020). Incorporating this extra computational aspect into the model adds to the complexity, and so as to preserve the simplicity of our approach, we leave this as a future consideration.

8

Conclusions and further work

The research presented in this thesis develops on methodology for modelling the extremal behaviour of aggregates of random variables and spatial processes. We consider the thesis to be composed of two main aspects of work: the first half takes a theoretical approach to describing the extremal behaviour of aggregates by providing results for the first order upper-tail behaviour of weighted sums of random variables; the latter half takes a data-driven approach and provides statistical modelling techniques for spatial aggregates, with particular interest being taken in precipitation aggregates. The thesis concludes with a brief study of non-stationarity in the extremal dependence structure of spatial processes and proposes a methodology for handling this issue if it presents itself in an application; this is increasingly likely to be an issue when modelling aggregates over increasingly large spatial regions. We now provide summaries of the chapters of this thesis and propose some avenues for further work.

Chapters 3 and 4 detail theoretical results for the first order behaviour of the survival function $\Pr\{R_d \geq r\}$ as $r \to r^F$, where r^F is the upper-endpoint of the distribution for the random variable $R_d = \sum_{i=1}^{d} X_i$; the marginals of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ are described by $X_i \sim \text{GPD}(\sigma_i, \xi_i)$ for $\sigma_i > 0, \xi_i \in \mathbb{R}$ and all $i = 1, \ldots, d$. The two main differences between Chapters 3 and 4 are as followed:

firstly, we consider $d = 2$ and $d \in \mathbb{N}$ in Chapters 3 and 4, respectively; the second difference is in how dependence within $\mathbf{X}$ is described. In Chapter 3, we model extremal dependence in $\mathbf{X}$ using the limiting dependence characterisations of Ledford and Tawn (1996) and Heffernan and Tawn (2004); in Chapter 4, we instead fully model dependence in $\mathbf{X}$ using one of five copulas: perfect dependence, independence, the standard Gaussian copula and both the upper, and lower, joint tails of the extreme value copula. In both chapters, the form of $\Pr\{R_d \geq r\}$ as $r \to r^F$ is shown to be driven both by the extremal dependence class of the associated $\mathbf{X}$ and the value of the marginal shape parameters ξ_i; furthermore we illustrate that, in certain cases, the driving behaviour of the upper-tails of R_d is heavily linked to the d-dimensional coefficient of tail dependence, η_d, first proposed by Ledford and Tawn (1996) for $d = 2$ and extended to $d > 2$ by Eastoe and Tawn (2012).

The theorems described in Section 3.3.2 apply only when specific assumptions are met by the extremal dependence structure of $\mathbf{X}$. Whilst these conditions may appear quite restrictive, they are often met by widely applied copula models; moreover, situations where the conditions are not met may instead be covered by the results derived in Chapter 4. However, that is not to say that the results across both Chapters 3 and 4 comprehensively describe all possible dependence models for $\mathbf{X}$; on the contrary, there exists structures that are not covered by Chapter 4 and do not satisfy the conditions given in 3.3.2, e.g., the Gumbel (1960) type 1 distribution or Morgenstern (1956) distribution. Thus, a natural area for extensions of Chapters 3 and 4 is to weaken the constraints in Section 3.3.2 or cover further examples of different copulas in Chapter 4, thereby expanding the catalogue of distributions for $\mathbf{X}$ for which results on R_d can be derived.

Although we consider $d > 2$ in Chapter 4, we constrain our focus to $d = 2$ when deriving the results in Chapter 3. However, whilst we present the two-dimensional representations of the Ledford and Tawn (1997) and Heffernan and Tawn (2004)

models in Section 3.3.1, d-dimensional variants do exist (see Eastoe and Tawn (2012)), and so it is reasonable to assume that the results given in Section 3.3.2 can be extended to any finite d.

Chapters 3 and 4 provide an interesting study into the driving factors of the extremal behaviour of aggregates, but extensions can be given for the initial model for $\mathbf{X}$. For example, we assert that X_i follows a GPD for all $X_i > 0$, rather than using the typical exceedance modelling framework, i.e., $(X_i - u_i)|X_i > u_i$ follows a GPD where $u_i \geq 0$ is some fixed threshold; whilst we argue in Section 3.2 that this does not affect the driving behaviour of $\Pr\{R_d \geq r\}$ as $r \to r^F$, certain scaling constants will be affected by the value of u_i. A potential avenue for future work is to generalise the marginals for X_i and allow for $u_i > 0$. However, this will require that some assumptions are proposed for the distribution of $X_i < u_i$, which will contribute to the complexity of the derivations. It may also be desirable to model the aggregate in this framework as well; that is, we model $(R_d - u_R)|R_d > u_R$ for fixed u_R, which we may assume is GPD. Deriving the relationship between each $u_i, i = 1, \ldots, d$ and u_R may also be an interesting study to conduct. Finally, we assert throughout Chapters 3 and 4 that the marginal components of $\mathbf{X}$ are positive and we allow for positive association only; weakening these two constraints may also provide interesting behaviour in the upper-tail of R_d.

The results in Chapters 3 and 4 describe the first-order behaviour of $\Pr\{R_d \geq r\}$ in the limit as $r \to r^F$. Although many extreme value models are underpinned by these such limit properties of random variables, they only characterise the behaviour of the most extreme values of the aggregate; from a practical stance, this may not be too useful for statistical modelling, as we will never observe data that achieve these limits, e.g., we will never observe precipitation aggregates that attain their physical maximum prescribed by the underlying hydrological processes. In practice, we are often much more interested in the aggregate at a sub-asymptotic level, and so instead

require a model for $\Pr\{R_d \geq r\}$ for some large, but finite, r; as analytical results for these probabilities do not exist in closed form, we instead rely on inference using data-driven approaches.

In Chapter 5, we propose extensions of the spatial conditional extremes model (Wadsworth and Tawn, 2019), which include novel parametric forms for the dependence parameters and a framework for deriving aggregates of spatial processes over spatial regions. For our model, we introduce a novel censored pseudo-likelihood approach to inference which has two benefits: firstly, we are able to address an issue commonly attributed to modelling rainfall, namely that data often contain multiple zero values corresponding to spatio-temporal periods without rain, and secondly, we are able to fit our model to high-dimensional data. We detail a procedure for simulating from our model and discuss methods for avoiding edge-effects; a common issue with inference using spatial models. From our fitted model, we simulate fields that can be used for inference on the extremal behaviour of spatial aggregates, and we illustrate that this approach reduces the uncertainty of return level estimates. Moreover, it preserves self-consistency of return level estimates for aggregates over different nested regions. That is, for a spatial process $\{Y(s) : s \in \mathcal{S}\}$ which is non-negative everywhere, i.e., $Y(s) \geq 0$ for all $s \in \mathcal{S}$, and any nested regions $\mathcal{B} \subset \mathcal{A} \subset \mathcal{S}$, then our approach guarantees that estimates of return-levels for $\int_{\mathcal{A}} Y(s)\mathrm{d}s$ and $\int_{\mathcal{B}} Y(s)\mathrm{d}s$ satisfy their natural ordering, regardless of the return period.

Chapter 6 provides an extension of the methodology detailed in Chapter 5. We use a deterministic algorithm to classify data as being either generated by a convective, or a non-convective, process, and then fit extensions of the marginal and extremal dependence models described in Chapter 5 to both clusters of data, separately. We detail extensions of the simulation procedure proposed in Section 5.3.4 which allow us to draw realisations from both the non-convective and convective fitted models; we then combine these into a single sample, which we use for inference on the upper-tail

behaviour of spatial aggregates. Whilst we find that this mixture model approach provides some improvements over the single process approach proposed in Chapter 5, there are some limitations to both approaches which may require further consideration.

The modelling approach we propose in Chapters 5 and 6 is underpinned by an assumption that the underlying data generating process and, hence, the extremal dependence structure is stationary; that is, the strength of extremal dependence between sites is a function of their displacement. Although this assumption seems appropriate in our application, we may find that the assumption is not realistic when considering rainfall aggregates over much larger spatial domains; this issue has been identified by Blanchet and Creutin (2017) and Castro-Camilo and Huser (2020) for regions in southern France and the contiguous USA, respectively. Literature on incorporating non-stationary extremal dependence into the spatial conditional extremes model is limited; Wadsworth and Tawn (2019) use the deformation methodlogy that we discuss in Chapter 7. However, this approach can be computationally burdensome and so may not be appropriate for high-dimensional data. A further extension of Chapters 5 and 6, and more generally the spatial conditional extremes framework, may be to adapt the model described in Section 5.2.2 to allow for non-stationary data. This may involve weakening the constraint that the process $\{W(s)\}$, defined in (5.2.8), is stationary; a possible alternative could be a Gaussian process with non-stationary correlation function, see Stein (2005) and Paciorek and Schervish (2006). Another possible route is to allow the dependence parameters, e.g., α and β, to be dependent on the conditioning site s_O; this could be achieved by incorporating covariates into α and β that relate to s_O, e.g., distance of s_O to a feature of orography, such as the open sea (Vandeskog et al., 2021), or by having an individual set of dependence parameters for each conditioning site s_O.

Full inference for our model is computationally infeasible for any reasonable num-

ber of sampling locations d, as it would require evaluations of k-dimensional conditional Gaussian distribution functions where $k \leq d-1$, i.e., k-variate integrals, see Sections 5.2.2 and 5.3.2, and (C.1.4) in Appendix C.1.3. Whilst full computation of these integrals may not be feasible, we could instead consider using approximations. For example, in Section 4.2.3 we detail Laplace's method (Laplace, 1986) for approximating multivariate integrals of a particular form, which can be applied to evaluate (C.1.4); thus, this technique could be used here to reduce the computational time of inference. Application of Laplace's method for approximating Gaussian CDFs has already been applied successfully in the literature; a class of latent Gaussian models rely on the integrated nested Laplace approximation (INLA) for high-dimensional Bayesian inference (Rue et al., 2009), and INLA is a method that relies on approximating posterior distributions via Laplace's method.

We note that even if evaluation of (C.1.4) was computationally feasible, we may still be constrained to smaller dimension d due the computational cost of inverting the $(d-1) \times (d-1)$ correlation matrix, required when evaluating the Gaussian density given by (C.1.1). This is a common problem with models for high-dimensional data that rely on Gaussian processes, but these issues can often be circumvented with "sparse" modelling techniques, which rely on low-rank approximations of the correlation matrix, which are computationally easier to invert. Such techniques include the stochastic partial differential equation (SPDE) approach, proposed by Lindgren et al. (2011), which approximates a Gaussian process by a finite-dimensional Gaussian Markov random field; when combined with INLA, this proves to be a particularly powerful tool for conducting inference for spatial models of high-dimensional data, and has been applied in the context of extreme quantile regression by Castro-Camilo et al. (2019) and Castro-Camilo et al. (2021), with Opitz et al. (2018) and Vandeskog et al. (2021) using this framework to model precipitation extremes. Moreover, Simpson et al. (2020) use the INLA-SPDE framework to conduct inference using the spatial

conditional extremes framework, which presents a potential means of accommodating large d in the modelling approach described in Chapter 5. However, to extend the work of Simpson et al. (2020) to allow sparse modelling of extreme preciptation, certain shortcomings must first be addressed: firstly, they assert that the dependence parameters $\beta(h)$ and $\sigma(h)$, given in Section 5.2.2, are constant for all distances, i.e., $\beta(h) = \beta$ and $\sigma(h) = \sigma$ for all $h = \|s - s_O\|$, which would be inappropriate for modelling precipitation following the discussion given in Section 5.4.3; secondly, they model the residual process $Z(s|s_O)$ with Gaussian margins, i.e., $\delta(h) = 2$ for all $h \geq 0$, which again is not an appropriate modelling choice for precipitation, see Figure 5.4.2; and finally, Simpson et al. (2020) fit their model to uncensored data, and so adaptations would need to be made to accommodate the censoring techniques we describe in Section 5.2.2. Although the INLA-SPDE approach is an appealing framework for inducing sparsity in spatial models, it is not the only method; Hazra and Huser (2019) propose a low-rank t-process for modelling extreme sea surface temperatures that is constructed using a low-rank Gaussian process (Wikle, 2010). Similar techniques may be applicable to the residual process $Z(s|s_O)$ in our model, as it is also constructed from a Gaussian process.

Fluvial floods are generally caused by extreme rainfall aggregated over both spatial regions and temporal periods, and whilst the model in Chapter 5 addresses the first aspect, as it stands it cannot be used for inference on spatio-temporal aggregates; this would require a temporal component in the model, that describes dependence between $X_t(s_O)|X_t(s_O) > u$ and $\{X_{t+\tau}(s) : s \in \mathcal{S}\}$ for $\tau > 0$. Simpson and Wadsworth (2021) propose a spatio-temporal extension of the spatial conditional extremes model which could be applied to precipitation using the inference techniques we propose in Section 5.3.1. However, it is not immediately clear how we would use replications from a spatio-temporal model to obtain a sample of spatio-temporal aggregates, and so extending the technique we describe in Section 5.3.5 could be potential further

work.

We find that our modelling approach in Chapter 5 is able to capture the extremal behaviour of spatial aggregates well, see Section 5.4.4. That is, when we average over realisations from our fitted model for extreme rainfall, we find similar characteristics to those present in the data. For the purposes of this work that is sufficient, as our interest lies in the average behaviour of the underlying process, rather than the process itself. However, when we consider single realisations from this model, we find that they may not appear realistic when compared to observed extreme rainfall events, see Figure 5.4.3; this suggests that there are some improvements that can be made to the model so that it can capture more realistic properties of rainfall. Firstly, our model assumes that rainfall is generated by a single underlying process. However, we know that in real life this is not the case; extreme precipitation is generated by a mixture of processes, e.g., high-intensity and spatially localised, convective events, and low-intensity, frontal events that cover a much larger spatial area (Schroeer et al., 2018), and that these processes will have their own respective marginal and dependence behaviour. Incorporating separate models for these two types of processes in our approach will be particularly useful when we consider modelling spatial aggregates, as we are likely to find that the respective rates at which these two types of process contribute to the extremal behaviour of spatial aggregates will vary with the region size, i.e., the extremal behaviour of aggregates for smaller and larger regions will be driven mostly by convective and frontal events, respectively. We begin to explore mixture modelling in Chapter 6 by taking a very simple approach and assuming that rainfall in a given hour is generated by one of two processes, and that these processes are independent of one another, with separate margins and dependence structure. We assume that all values in an observed field can be classified as coming from only one of these two processes. This is likely to be a good approximation over regions of the size that we study, but for larger regions we might actually expect that there is a

mixture of rainfall and convective events within a single field. A possible extension of our model could be to introduce extra spatial processes to help to account for events that are not generated by the first two processes; this would require either adapting the identification algorithm we use to allow for more than two labels or adopting a means of probabilistic clustering using a Bayesian framework.

Another way of improving the realism in our model is to remove the constraint that dependence within the residual process $Z(s|s_O)$ is stationary, but anisotropic, and Gaussian, as this constraint causes realised events to exhibit a generally consistent orientation and elliptical profile. In fact, extreme precipitation often occurs along weather fronts, i.e., boundaries that separate air masses with different atmospheric properties (Egger and Hoinka, 1992); this results in events forming as bands of rainfall, where the orientation is affected by wind direction, rather than elliptical events At present, our dependence model cannot capture events of this type; however, a first step may be to relate the anisotropy parameters (θ, L), given in (5.2.11), to covariates that describe atmospheric conditions. An alternative, non-parametric approach could be to use the empirical distribution of the residuals, rather than a parametric model, which could be achieved using the following two-step procedure: we first make the working assumption that $Z(s|s_O)$ follows the parametric model we propose in Section 5.2.2, and then fit the model under this assumption to attain estimates for the parameter functions α and β; these estimates can then be used to get an empirical sample of residuals, which are then used for simulation. We could then further assume that the residual process is a mixture of processes, and apply dimensionality reduction techniques, e.g., empirical orthogonal function (EOF) analysis, to identify classes of dependence structures that the residuals exhibit; simulation of an event may then require choosing one of these dependence classes with some probability. To simulate using the empirical residuals, we would be required to employ resampling techniques, which have been used in the context of spatial extremes by Palacios-Rodríguez et al.

(2020) and Opitz et al. (2021), with the former applied to extreme precipitation.

In Chapter 7, we propose an approach to modelling non-stationary extremal dependence, which extends the spatial deformation methodology of Sampson and Guttorp (1992) and Smith (1996). We do this by replacing the objective function in these methods with the Frobenius norm of the difference between empirical, and theoretical, pairwise dependency matrices, with the theoretical measures coming from a stationary dependence model. Most of our focus is on $\chi(h^*_{ij})$ as the dependence measure, where $\chi(h^*_{ij}) = \lim_{q\to 1} \chi_q(h^*_{ij})$ for $\chi_q(h^*_{ij})$ defined in (7.2.2) and $h^*_{ij} = \|s^*_i - s^*_j\|$ denotes pairwise distance between sites s^*_i and s^*_j in the deformed space; however, we show that $\chi(h^*_{ij})$ can easily be replaced by other measures, such as $\chi_q(h^*_{ij})$ with $q < 1$, correlation or the coefficient of tail dependence, η. As with many areas of extreme value analysis, there is a bias-variance trade-off present when estimating χ_q. Using values of q closer to 1 puts greater focus on extremal dependence at the expense of increased variance of the estimator. In Sections 7.3 and 7.4, we choose q close to 1 whilst preserving some initial spatial structure observed in the χ estimates. However, if q is too high then it is possible that any structure is masked by the high variability of the estimators and the deformation methodology is likely to fail in such circumstances. We have not considered the effect of estimator variability on the deformation, but note this could form a future research direction.

A further issue that could be considered is the possible non-bijectivity of the mapping used in the deformation. We detail an approach to reduce the chances of a non-bijective deformation occuring in Section 7.3, however, this method is not particularly robust. Without prior specification of a triangulation on the coordinate system, bijectivity of deformations must be checked by eye which can become cumbersome when a large number are produced. To avoid this necessary supervision, the G-plane can be represented as a Delaunay triangulation, see Iovleff and Perrin (2004) and Youngman (2020). Incorporating this extra computational aspect into the model

adds to the complexity, and so as to preserve the simplicity of our approach, we leave this as a future consideration.

Appendix A

Supplementary material for Chapter 3

A.1 Outline

Appendix A.2 provides justification for the tail formulation for R in (3.2.3) in the main text by linking this characterisation to the GPD tail model, (3.1.2) in the main text. The rest of the Supplementary Material then follows with proofs for Theorem 3.3.1-3.3.4 which are detailed in Section 3.3.2 in the main text. Appendix A.4 and A.5 provide the proofs of Theorems 3.3.2 and 3.3.3, respectively; both proofs follow a similar outline to that of the proof for the $\xi < 0$ case for Theorem 3.3.1, which is given in Appendix A.3.1 of the main text. Appendix A.6 concludes with the proof of Theorem 3.3.4.

A.2 Linking (3.2.3) to the usual GPD modelling framework

Assume that (3.2.3) holds in equality, rather than asymptotically (as in = not $\sim$), for $r \geq u_R$ for fixed $u_R \geq 0$. If $\xi_R > 0$, we have $\Pr\{R \geq r\} = K_1 r^{-1/\xi_R}$ for $r \geq u_R$, and then for $r > 0$

$$\Pr\{R \geq r + u_R | R > u_R\} = \frac{K_1(r+u_R)^{-1/\xi_R}}{K_1 u_R^{-1/\xi_R}} = \left(1 + \frac{r}{u_R}\right)^{-1/\xi_R} = \left(1 + \frac{\xi_R r}{u_R \xi_R}\right)^{-1/\xi_R}.$$

It follows that $(R - u_R) \,|\, (R > u_R) \sim \text{GPD}(\sigma_R, \xi_R)$, with $\sigma_R = u_R \xi_R$. A similar approach can be used to show that if $\xi_R = 0$, then $(R - u_R) \,|\, (R > u_R)$ is $\text{GPD}(\sigma_R, 0)$. For $\xi_R < 0$ and $r > 0$ with $r + u_R < r^F$, we have

$$\begin{aligned}\Pr\{R \geq r + u_R | R > u_R\} = \frac{K_3\left(1 - \frac{r+u_R}{r^F}\right)^{-1/\xi_R}}{K_3\left(1 - \frac{u_R}{r^F}\right)^{-1/\xi_R}} &= \left(1 - \frac{r}{(r^F - u_R)}\right)^{-1/\xi_R} \\ &= \left(1 + \xi_R \frac{r}{(-\xi_R)(r^F - u_R)}\right)^{-1/\xi_R},\end{aligned}$$

and so $(R - u_R) \,|\, (R > u_R) \sim \text{GPD}(\sigma_R, \xi_R)$, with $\sigma_R = (-\xi_R)(r^F - u_R)$. Note that we have made no assumptions about r^F as this is fully determined by the marginal upper-endpoints.

A.3 Proof of Theorem 3.3.1

A.3.1 Negative Shape Case: $\xi < 0$

The general framework of the proof is as followed: we begin by deriving the joint density of (X_1, X_2) implied by the dependence model given in (3.3.1), which we give on GPD margins. We use the probability integral transform to perform a transformation $(X_1, X_2) \to (R, W)$, where $R = X_1 + X_2$ and W is an auxiliary variable, chosen as its

support does not depend on R for R greater than some threshold. We integrate out W to give the density of R and derive its survival function.

Combining (3.3.1) and (3.1.2) with $\xi_1 = \xi_2 = \xi < 0$, we have that

$$\Pr\left\{1-\left(1+\xi\frac{X_1}{\sigma_1}\right)^{-1/\xi} > 1-\frac{1}{x_1}, 1-\left(1+\xi\frac{X_2}{\sigma_2}\right)^{-1/\xi} > 1-\frac{1}{x_2}\right\} = \frac{\mathcal{L}(x_1+x_2)}{(x_1x_2)^{\frac{1}{2\eta}}} g\left(\frac{x_1}{x_1+x_2}\right),$$

as $x_1, x_2 \to \infty$ such that the limit of $x_1/(x_1+x_2)$ is bounded by $(0,1)$. Under the assumption that $\mathcal{L}(y)$ acts as a constant which can be absorbed by g for $y > v$ for some $v > 0$, we have $\Pr\{X_1 > x_1, X_2 > x_2\} \sim \tilde{x}_1^{-\frac{1}{2\eta\xi}} \tilde{x}_2^{-\frac{1}{2\eta\xi}} g(\omega_x)$ for $x_1 \to x_1^F$ and $x_2 \to x_2^F$, such that $\omega_x = \tilde{x}_1^{1/\xi}/(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}) \to \omega_x^* \in (0,1)$ and where $\tilde{x}_i = (1+\xi x_i/\sigma_i)$ for $i = 1, 2$. Assuming that the first and second derivatives of g exist, then the density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1,x_2) \sim \frac{(\tilde{x}_1\tilde{x}_2)^{-\frac{1}{2\eta\xi}-1}}{\sigma_1\sigma_2} \tag{A.3.1}$$

$$\times \left[\frac{g(\omega_x)}{4\eta^2} + (\tilde{x}_1\tilde{x}_2)^{\frac{1}{\xi}} \frac{\tilde{x}_1^{1/\xi} - \tilde{x}_2^{1/\xi}}{\left(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}\right)^3} g'(\omega_x) - \frac{(\tilde{x}_1\tilde{x}_2)^{\frac{2}{\xi}}}{\left(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}\right)^4} g''(\omega_x)\right], \tag{A.3.2}$$

as $x_1 \to x_1^F$ and $x_2 \to x_2^F$ such that $\omega_x \to \omega_x^* \in (0,1)$. We now apply the transformation $(X_1, X_2) \to (R, W)$, where

$$R = X_1 + X_2, \quad W = \frac{(\sigma_1 + \xi X_1)}{(\sigma_1 + \xi X_1) + (\sigma_2 + \xi X_2)},$$

where $(1+\xi X_1/\sigma_1) = -\frac{\xi}{\sigma_1}(r^F - R)W$ and $(1+\xi X_2/\sigma_2) = -\frac{\xi}{\sigma_2}(r^F - R)(1-W)$ for $r^F = -(\sigma_1/\xi + \sigma_2/\xi)$ the upper-endpoint of R and where $(r^F - r)$ is the determinant

of the Jacobian of the transformation. The density of (R, W) as $r \to r^F$, is

$$f_{R,W}(r,w) \sim \frac{(-\xi)^{-\frac{1}{\eta\xi}-1}}{\eta(\sigma_1+\sigma_2)^{-\frac{1}{\eta\xi}}}(r^F-r)^{-\frac{1}{\eta\xi}-1}g^*(w)$$

where

$$g^*(w) = \frac{\eta(\sigma_1+\sigma_2)^{\frac{1}{\eta\xi}}}{(-\xi)(\sigma_1\sigma_2)^{-\frac{1}{\eta\xi}}}\left\{w(1-w)\right\}^{-\frac{1}{2\eta\xi}-1}\left[\frac{g\left(t_w\right)}{4\eta^2}\right.$$
$$\left.+\left(\frac{w(1-w)}{\sigma_1\sigma_2}\right)^{\frac{1}{\xi}}\frac{\left(\frac{w}{\sigma_1}\right)^{1/\xi}-\left(\frac{1-w}{\sigma_2}\right)^{1/\xi}}{\left(\left(\frac{w}{\sigma_1}\right)^{1/\xi}+\left(\frac{1-w}{\sigma_2}\right)^{1/\xi}\right)^3}g^{'}\left(t_w\right)-\frac{\left(\frac{w(1-w)}{\sigma_1\sigma_2}\right)^{\frac{2}{\xi}}}{\left(\left(\frac{w}{\sigma_1}\right)^{1/\xi}+\left(\frac{1-w}{\sigma_2}\right)^{1/\xi}\right)^4}g^{''}\left(t_w\right)\right],$$

and

$$t_w = (w/\sigma_1)^{1/\xi}\left\{(w/\sigma_1)^{1/\xi}+((1-w)/\sigma_2)^{1/\xi}\right\}^{-1} \in (0,1).$$

We now show that the support of W is independent of R given that R is above $u = \max\{x_1^F, x_2^F\}$. Let $x^{max} = \max\{-\sigma_1/\xi, -\sigma_2/\xi\}$ and $x^{min} = \min\{-\sigma_1/\xi, -\sigma_2/\xi\}$. As $X_1+X_2 \leq -(\sigma_1/\xi+\sigma_2/\xi) = x^{max}+x^{min}$, there exists a random variable $P \in [0,1]$, such that $X_1+X_2|(X_1+X_2 > x^{max}) = x^{max}+x^{min}P$. Now for $i = \arg\max_{j=1,2}\{-\sigma_j/\xi\}$, let $X_i|(X_1+X_2 > x^{max}) = x^{max}Q_i$ for random $0 < Q_i \leq 1$. Then as

$$0 \leq (X_1+X_2-X_i)|(X_1+X_2 > x^{max}) \leq x^{min} \Rightarrow 0 \leq x^{max}+x^{min}P-x^{max}Q_i \leq x^{min},$$

it follows that Q_i must satisfy

$$-\frac{P}{1-P} < 0 \leq \frac{x^{max}(1-Q_i)}{x^{min}(1-P)} < 1.$$

Now consider W. If we have $i = 1$, then

$$W = \frac{\frac{\sigma_1}{\xi}+X_1}{\frac{\sigma_1}{\xi}+\frac{\sigma_2}{\xi}+X_1+X_2} = \frac{-x^{max}+x^{max}Q_1}{-x^{max}-x^{min}+x^{max}+x^{min}P} = \frac{x^{max}(1-Q_1)}{x^{min}(1-P)},$$

and so $W \in [0,1]$ as needed. However, if $i = 2$ we instead have $W = 1 - x^{max}(1 - Q_2)/\{x^{min}(1-P)\}$; this also gives $W \in [0,1]$ and so the support of W does not depend on R when $R > u$. Now consider the survival function of R as $s \to r^F$, so $s > u$, then

$$\begin{aligned}\Pr\{R \geq s\} &\sim \int_s^\infty \int_0^1 \frac{(-\xi)^{-\frac{1}{\eta\xi}-1}}{\eta(\sigma_1+\sigma_2)^{-\frac{1}{\eta\xi}}}(r^F - r)^{-\frac{1}{\eta\xi}-1} g^*(w)\mathrm{d}w\mathrm{d}r \\ &\sim K\int_s^\infty \frac{(-\xi)^{-\frac{1}{\eta\xi}-1}}{\eta(\sigma_1+\sigma_2)^{-\frac{1}{\eta\xi}}}(r^F - r)^{-\frac{1}{\eta\xi}-1}\mathrm{d}r \\ &\sim K\frac{(-\xi)^{-\frac{1}{\eta\xi}}}{(\sigma_1+\sigma_2)^{-\frac{1}{\eta\xi}}}(r^F - s)^{-\frac{1}{\eta\xi}} = K\left(1+\xi\frac{s}{\sigma_1+\sigma_2}\right)^{-\frac{1}{\eta\xi}},\end{aligned}$$

where the limits hold as $f_{R,W}$ is a valid probability density, and K is a constant, given by

$$K = \int_0^1 g^*(w)\mathrm{d}w < \infty. \tag{A.3.3}$$

A.3.2 Positive Shape Case: $\xi > 0$

We begin by noting that

$$\Pr\{R \geq r\} = \Pr\{R \geq r \cap X_1 > u_1\} + \Pr\{R \geq r \cap X_2 > u_2\} - \Pr\{R \geq r \cap X_1 > u_1 \cap X_2 > u_2\}, \tag{A.3.4}$$

for any fixed constants $u_1, u_2 > 0$. To derive $\Pr\{R \geq r\}$, we consider each of the terms in (A.3.4) in turn. We first derive $\Pr\{R \geq r \cap X_1 > u_1\}$ for large u_1, which we do by starting with the dependence model in (3.3.2) and deriving the joint density of (Y_1, Y_2) on standard Exponential margins. We then transform these to heavy tailed GPD margins, X_1 and X_2, and we perform the transformation $(X_1, X_2) \to (R, W)$, where W is an auxiliary variable. To integrate out W, we perform another transformation $W \to T$ where T is an auxiliary variable chosen so that it is possible to integrate over and derive the marginal density of $R|X_1 > u_1$; we then use this to determine the asymptotic behaviour of $f_{R \geq r \cap X_1 > u_1}$ and derive its survival function. The $\Pr\{R \geq$

$r \cap X_2 > u_2\}$ follows by symmetry, and we then evaluate $\Pr\{R \geq r \cap X_1 > u_1 \cap X_2 > u_2\}$ using (3.3.1).

Assume that limit (3.3.2) holds for $a(y) = \alpha(y)$ and $b(y) = y^\beta$ for some $\alpha \in [0,1]$ and $\beta \in [0,1]$ for some large u. We denote the residual distribution by $G_Z(\cdot)$ and assume it is differentiable with density g_Z. Then the joint density of $(Y_1, Y_2)|Y_1 > u$ is

$$f_{(Y_1,Y_2)|Y_1>u}(y_1, y_2) = \exp(-y_1) y_1^{-\beta} g_Z\left(\frac{y_2 - \alpha y_1}{y_1^\beta}\right),$$

for $y_1 > u_1$ and $y_2 \geq 0$. We now transform to heavy tailed marginals $X_1 \sim \text{GPD}(\sigma_1, \xi)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi)$ for $\xi > 0$ through the transformation $(Y_1, Y_2) \to (X_1, X_2)$ where $Y_i = \frac{1}{\xi}\log(1 + \xi X_i/\sigma_i)$ for $i = 1, 2$. We also note that $Y_1 > u$ is equivalent to $X_1 > \frac{\sigma_1}{\xi}\{\exp(\xi u) - 1\} := u_1$ and so we rewrite the condition as $X_1 > u_1$. The joint density of $(X_1, X_2)|X_1 > u_1$ is

$$\begin{aligned} f_{(X_1,X_2)|X_1>u_1}(x_1, x_2) &= |J| \xi^\beta \tilde{x}_1^{-1/\xi} \left\{\log(\tilde{x}_1)\right\}^{-\beta} g_Z(z_x^*) \\ &= \frac{\xi^\beta}{\sigma_1\sigma_2} \tilde{x}_2^{-1} \tilde{x}_1^{-1/\xi - 1} \left\{\log(\tilde{x}_1)\right\}^{-\beta} g_Z(z_x^*), \end{aligned}$$

for large u and where $\tilde{x}_i = 1 + \xi x_i/\sigma_i$ for $i = 1, 2$ and $|J| = (\sigma_1\sigma_2\tilde{x}_1\tilde{x}_2)^{-1}$ is the determinant of the Jacobian of the transformation and $z_x^* = \xi^{\beta-1}\{\log(\tilde{x}_1)\}^{-\beta}[\log(\tilde{x}_2) - \alpha\log(\tilde{x}_1)]$ where $z_x^* \in \mathbb{R}$ if $\beta < 1$ and $z_x^* \geq -\alpha$, otherwise. A final transformation to pseudo-radial and -angular components $(X_1, X_2) \to (R, W = X_1/R)$ is performed, with Jacobian determinant $|J| = R$. The density of $(R, W)|X_1 > u_1$ is

$$\begin{aligned} f_{(R,W)|X_1>u_1}(r, w) = r\frac{\xi^\beta}{\sigma_1\sigma_2} &\left(1 + \xi\frac{r(1-w)}{\sigma_2}\right)^{-1} \left(1 + \xi\frac{rw}{\sigma_1}\right)^{-1/\xi - 1} \\ &\times \left\{\log\left(1 + \xi\frac{rw}{\sigma_1}\right)\right\}^{-\beta} g_Z(z_w^*), \end{aligned}$$

with

$$z_w^* = \xi^{\beta-1} \frac{\log\left(1+\xi\frac{r(1-w)}{\sigma_2}\right) - \alpha\log\left(1+\xi\frac{rw}{\sigma_1}\right)}{\left\{\log\left(1+\xi\frac{rw}{\sigma_1}\right)\right\}^{\beta}}.$$

Note that as we have $X_1 > u_1$, this implies that $w \in [u_1/r, 1]$. However, we can prove that

$$\int_{u_1/r}^{c_1} f_{(R,W)|X_1>u_1}(r,w)\mathrm{d}w \sim \int_0^{c_1} f_{(R,W)|X_1>u_1}(r,w)\mathrm{d}w \tag{A.3.5}$$

as $r \to \infty$ by showing that $f_{(R,W)|X_1>u_1}(r,w)$ goes to infinity at a sufficiently slow rate as $w \to 0$. This follows as

$$f_{(R,W)|X_1>u_1}(r,w) \sim \frac{\xi^{\beta-1}}{\sigma_1}\left\{\log\left(1+\xi\frac{rw}{\sigma_1}\right)\right\}^{-\beta} g_Z\left(z_w^*\right) \sim \frac{\xi^{\beta-1}}{\sigma_1}\left(\xi\frac{r}{\sigma_1}\right)^{-\beta} w^{-\beta} g_Z\left(z_w^*\right),$$

as $w \to 0$. Note that as $w \to 0$, we have that $z_w^* \to \infty$; as g_Z is a valid density, it follows that $g_Z(z_w^*) \to 0$ as $w \to 0$ and so $f_{(R,W)|X_1>u_1}(r,w)$ must go to infinity at a slower rate than $w^{-\beta}$ as $w \to 0$ for $\beta \leq 1$. Hence, $f_{(R,W)|X_1>u_1}(r,w)$ must integrate to a function that goes to zero as $w \to 0$, and so it follows that the asymptotic relationship in (A.3.5) holds.

To integrate W out of $f_{R,W}$, we use the transformation $W \to T$, where $W = 1 - R^{-T}$ for $T \in (0,\infty)$. The determinant of the Jacobian of this transformation is $R^{-T}\log(R)$, and the joint density of (R,T) for $t \in (0,\infty)$ and as $r \to \infty$ is

$$\begin{aligned} f_{(R,T)|X_1>u_1}(r,t) &\sim \frac{\xi^{\beta} r^{1-t}\log(r)}{\sigma_1\sigma_2}\left(1+\xi\frac{r^{1-t}}{\sigma_2}\right)^{-1}\left(1+\xi\frac{r(1-r^{-t})}{\sigma_1}\right)^{-1/\xi-1} \\ &\quad\times\left\{\log\left(1+\xi\frac{r(1-r^{-t})}{\sigma_1}\right)\right\}^{-\beta} g_Z\left(z_t^*\right) \\ &\sim r^{-1/\xi-1}r^{1-t}\{\log(r)\}^{1-\beta}\frac{\xi^{\beta-1}\xi^{-1/\xi}}{\sigma_1^{-1/\xi}\sigma_2}\left(1+\xi\frac{r^{1-t}}{\sigma_2}\right)^{-1} g_Z\left(z_t^*\right) \end{aligned}$$

where

$$z_t^* = \xi^{\beta-1}\frac{\log\left(1+\xi\frac{r^{1-t}}{\sigma_2}\right)-\alpha\log\left(1+\xi\frac{r(1-r^{-t})}{\sigma_1}\right)}{\left\{\log\left(1+\xi\frac{r(1-r^{-t})}{\sigma_1}\right)\right\}^{\beta}} \sim \xi^{\beta-1}\frac{\log\left(1+\xi\frac{r^{1-t}}{\sigma_2}\right)-\alpha\log(r)}{\{\log(r)\}^{\beta}},$$

as $r\to\infty$; this follows as $1-r^{-t}\sim 1$ as $r\to\infty$ and for any $t\in(0,\infty)$. Hence the density of R is

$$\begin{aligned}
f_{R|X_1>u_1}(r) &\sim \int_0^\infty f_{(R,T)|X_1>u_1}(r,w)\mathrm{d}w \\
&\sim \left(\frac{\xi}{\sigma_1}\right)^{-1/\xi} r^{-1/\xi-1}\int_0^\infty r^{1-t}\{\log(r)\}^{1-\beta}\frac{\xi^{\beta-1}}{\sigma_2}\left(1+\xi\frac{r^{1-t}}{\sigma_2}\right)^{-1} g_Z\left(z_t^*\right)\mathrm{d}w \\
&\sim \frac{(\xi r)^{-1/\xi-1}}{\sigma_1^{-1/\xi}}\left[\bar{G}_Z\left(z_t^*\right)\right]_0^\infty \\
&\sim \frac{(\xi r)^{-1/\xi-1}}{\sigma_1^{-1/\xi}}\left[\bar{G}_Z\left(-\alpha\xi^{\beta-1}\{\log(r)\}^{1-\beta}\right)-\bar{G}_Z\left(\xi^{\beta-1}(1-\alpha)\{\log(r)\}^{1-\beta}\right)\right] \\
&\sim K_G\frac{(\xi r)^{-1/\xi-1}}{\sigma_1^{-1/\xi}},
\end{aligned}$$

as $r\to\infty$ and where

$$K_G = \begin{cases} \bar{G}_Z(0), & \text{if } \alpha=0, \\ \bar{G}_Z(-\alpha)-\bar{G}_Z(1-\alpha), & \text{if } \beta=1, \\ 1, & \text{otherwise,} \end{cases} \tag{A.3.6}$$

and so we have $\Pr\{R\geq s|X_1>u_1\}\sim K_G\xi^{-1/\xi}\sigma_1^{1/\xi}s^{-1/\xi}$ as $s\to\infty$. It follows that

$$\Pr\{R\geq s\cap X_1>u_1\}\sim \exp(-u)K_G\xi^{-1/\xi}\sigma_1^{1/\xi}s^{-1/\xi}, \tag{A.3.7}$$

as $s\to\infty$ and for large u as $\Pr(X_1>u_1)=\Pr(Y_1>u)=e^{-u}$. We now assume that Y_1 and Y_2 in (3.3.2) are interchangeable, i.e., a similar limit holds for $(Y_2>u)\equiv(X_2>u_2)$ with $u_2=\frac{\sigma_2}{\xi}\{\exp(\xi u)-1\}$. By a symmetric argument, we can also show

that

$$\Pr\{R \geq s \cap X_2 > u_2\} \sim \exp(-u) K_G \xi^{-1/\xi} \sigma_2^{1/\xi} s^{-1/\xi}, \tag{A.3.8}$$

as $s \to \infty$ and for large u. From (A.3.4), we now require only

$$\Pr\{(R \geq s) \cap (X_1 > u_1) \cap (X_2 > u_2)\}.$$

To derive this term, we first consider $\Pr\{R \geq s | (X_1 > u_1 \cap X_2 > u_2)\}$, which we derive using characterisation (3.3.1). Assuming limit (3.3.1) to hold for fixed u_1 and u_2, we follow the beginning of the proof in Section A.3.1. We derive the joint density of $(X_1, X_2)|(X_1 > u_1 \cap X_2 > u_2)$ and then perform the marginal transformation $(X_1, X_2) \to (R, W)$, where W is an auxiliary variable that is to be integrated out. This leaves us with the marginal density of $R|(X_1 > u_1 \cap X_2 > u_2)$ which allows us to find the survival function of $R|(X_1 > u_1 \cap X_2 > u_2)$.

From (A.3.1), we have the density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) \sim \frac{(\tilde{x}_1 \tilde{x}_2)^{-\frac{1}{2\eta\xi}-1}}{\sigma_1 \sigma_2} \times \left[\frac{g(\omega_x)}{4\eta^2} + (\tilde{x}_1 \tilde{x}_2)^{\frac{1}{\xi}} \frac{\tilde{x}_1^{1/\xi} - \tilde{x}_2^{1/\xi}}{\left(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}\right)^3} g'(\omega_x) - \frac{(\tilde{x}_1 \tilde{x}_2)^{\frac{2}{\xi}}}{\left(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}\right)^4} g''(\omega_x) \right],$$

as $x_1 \to \infty$ and $x_2 \to \infty$ such that $\omega_x = \tilde{x}_1^{1/\xi}/(\tilde{x}_1^{1/\xi} + \tilde{x}_2^{1/\xi}) \to \omega_x^* \in (0, 1)$ and where $\tilde{x}_i = (1 + \xi x_i/\sigma_i)$ for $i = 1, 2$. We now perform the transformation $(X_1, X_2) \to (R, W = X_1/R)$ where $W \in [u_1/R, 1 - u_2/R]$, which has Jacobian determinant $|J| = r$, and it follows that $f_{(R,W)|(X_1>u_1 \cap X_2>u_2)}(r, w) \sim \frac{r}{\sigma_1 \sigma_2} g^*(r, w)$ for $w \in [u_1/r, 1 - u_2/r]$ and as

$r \to \infty$ and where

$$g^*(r,w) = \left(1+\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{2\eta\xi}-1}\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)^{-\frac{1}{2\eta\xi}-1} \times\left[\frac{g\left(t_{r,w}\right)}{4\eta^2}+t_{r,w}(1-t_{r,w})(2t_{r,w}-1)g^{'}(t_{r,w})-t_{r,w}^2(1-t_{r,w})^2g^{''}(t_{r,w})\right], \tag{A.3.9}$$

and

$$t_{r,w} = \frac{\left(1+\xi\frac{rw}{\sigma_1}\right)^{1/\xi}}{\left(1+\xi\frac{rw}{\sigma_1}\right)^{1/\xi}+\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)^{1/\xi}}.$$

Now, recall that we have $W \in [u_1/R, 1-u_2/R]$. It follows that, as $r \to \infty$, that

$$f_{R|(X_1>u_1\cap X_2>u_2)}(r) \sim (\sigma_1\sigma_2)^{-1}rI(r),$$

where $I(r) = \int_{u_1/r}^{1-u_2/r} g^*(r,w)\mathrm{d}w$ for g^* defined in (A.3.9). To evaluate the integral $I(r)$, we make different assumptions on how the tails of g, and hence $g*$, behave; we consider two cases, each with $I(r) < \infty$.

Case 1 We assume Condition 2. Note that $w \to 0 \Rightarrow t_{r,w} \to 0$. We now rewrite the integral $I(r) = I_d(r) + I_1(r) + I_2(r)$, where

$$I_d(r) = \int_{d_1}^{1-d_2} g^*(r,w)\mathrm{d}w,\ I_1(r) = \int_{u_1/r}^{d_1} g^*(r,w)\mathrm{d}w,\ \text{ and }\ I_2 = \int_{1-d_2}^{1-u_2/r} g^*(r,w)\mathrm{d}w,$$

where d_1 and d_2 are constants chosen such that $d_1 > u_1/r, d_2 > u_2/r$ and $d_1 < 1-d_2$. We show that, as $r \to \infty$, we have that $I(r) \sim I_1(r) + I_2(r)$. First, consider $I_d(r)$. As

$r \to \infty$, we have

$$t_{r,w} \sim \frac{\left(\xi \frac{rw}{\sigma_1}\right)^{1/\xi}}{\left(\xi \frac{rw}{\sigma_1}\right)^{1/\xi} + \left(\xi \frac{r(1-w)}{\sigma_2}\right)^{1/\xi}} = \frac{\left(\frac{w}{\sigma_1}\right)^{1/\xi}}{\left(\frac{w}{\sigma_1}\right)^{1/\xi} + \left(\frac{(1-w)}{\sigma_2}\right)^{1/\xi}} = t_w,$$

and it follows that $g^*(r,w) \sim \xi^{-\frac{1}{\eta\xi}-2}(\sigma_1\sigma_2)^{\frac{1}{2\eta\xi}+1} r^{-\frac{1}{\eta\xi}-2} h_w(w)$, where

$$h_w(w) = w^{-\frac{1}{2\eta\xi}-1}(1-w)^{-\frac{1}{2\eta\xi}-1}\left[\frac{g\left(t_w\right)}{4\eta^2} + t_w(1-t_w)(2t_w-1)g^{'}(t_w) - t_w^2(1-t_w)^2 g^{''}(t_{r,w})\right].$$

Thus, we have $I_d(r) \sim K_d r^{-\frac{1}{\eta\xi}-2}$ for constant $K_d = \xi^{-\frac{1}{\eta\xi}-2}(\sigma_1\sigma_2)^{\frac{1}{2\eta\xi}+1}\int_{d_1}^{1-d_2} h_w(w)\mathrm{d}w > 0$. Now, consider $I_1(r)$. We begin by noting that as $r \to \infty$ and for $w \in [u_1/r, c_1]$, we have

$$t_{r,w} \sim \left(1+\xi\frac{rw}{\sigma_1}\right)^{1/\xi}\left(\frac{\xi r}{\sigma_2}\right)^{-1/\xi} \to 0.$$

From (A.3.9), it follows that

$$g^*(r,w) \sim K_g\left(\frac{1}{4\eta^2}-\kappa^2\right)\xi^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-1}(\sigma_2)^{\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}+1} r^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-1}\left(1+\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}-1},$$

and

$$\begin{aligned}
I_1(r) &\sim K_g\left(\frac{1}{4\eta^2}-\kappa^2\right)\xi^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-1}(\sigma_2)^{\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}+1} r^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-1}\int_{u_1/r}^{d_1}\left(1+\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}-1}\mathrm{d}w \\
&\sim \left(\kappa-\frac{1}{2\eta}\right)^{-1} K_g\left(\frac{1}{4\eta^2}-\kappa^2\right)\xi^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-1}(\sigma_2)^{\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}+2} r^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-2} \\
&\times \left[\left(1+\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{2\eta\xi}+\frac{\kappa}{\xi}}\right]_{u_1/r}^{d_1} \\
&\sim K_4 r^{-\frac{1}{2\eta\xi}-\frac{\kappa}{\xi}-2},
\end{aligned}$$

for constant

$$K_4 = \left(\frac{1}{2\eta} - \kappa\right)^{-1} K_g \left(\frac{1}{4\eta^2} - \kappa^2\right) \xi^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 1} (\sigma_2)^{\frac{1}{2\eta\xi} + \frac{\kappa}{\xi} + 2} \left(1 + \xi \frac{u_1}{\sigma_1}\right)^{-\frac{1}{2\eta\xi} + \frac{\kappa}{\xi}} > 0$$

and where the last line follows as $-\frac{1}{2\eta\xi} + \frac{\kappa}{\xi} < 0$. A symmetric argument can be used to show that $I_2(r) \sim K_5 r^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 2}$ where

$$K_5 = \left(\kappa - \frac{1}{2\eta}\right)^{-1} K_g \left(\frac{1}{4\eta^2} - \kappa^2\right) \xi^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 1} (\sigma_1)^{\frac{1}{2\eta\xi} + \frac{\kappa}{\xi} + 2} \left(1 + \xi \frac{u_2}{\sigma_2}\right)^{-\frac{1}{2\eta\xi} + \frac{\kappa}{\xi}} > 0.$$

Comparing the powers on the r term in $I_d(r), I_1(r)$ and $I_2(r)$, and with $-\frac{1}{\eta\xi} < -\frac{1}{2\eta\xi} + \frac{\kappa}{\xi}$ for all η, κ, it follows that $I(r) \sim I_1(r) + I_2(r) \sim (K_4 + K_5) r^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 2}$ as $r \to \infty$. Hence,

$$f_{R|(X_1 > u_1 \cap X_2 > u_2)}(r) \sim (\sigma_1 \sigma_2)^{-1} (K_4 + K_5) r^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 1}$$

and

$$\Pr\{R \geq s | (X_1 > u_1) \cap (X_2 > u_2)\} \sim (\sigma_1 \sigma_2)^{-1} (K_4 + K_5) \int_s^\infty r^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi} - 1} \mathrm{d}r \sim K_6 s^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi}},$$

as $s \to \infty$ and where $K_6 = \xi (\sigma_1 \sigma_2)^{-1} (K_4 + K_5)(1/2\eta + \kappa)^{-1} > 0$.

Recall that $u_i = \frac{\sigma_i}{\xi} \{\exp(\xi u) - 1\}$ for $i = 1, 2$. From (3.3.1), we have $\Pr\{(X_1 > u_1) \cap (X_2 > u_2)\} = \exp(-u/\eta) g(1/2)$ for large u_1, u_2, and hence

$$\Pr\{(R \geq r) \cap (X_1 > u_1) \cap (X_2 > u_2)\} \sim K_6 \exp\left(-\frac{u}{\eta}\right) g\left(\frac{1}{2}\right) r^{-\frac{1}{2\eta\xi} - \frac{\kappa}{\xi}}, \quad \text{(A.3.10)}$$

as $r \to \infty$ and for large u. Combining (A.3.7), (A.3.8) and (A.3.10), we have $\Pr\{R \geq$

$r\} \sim K^+ r^{-1/\xi}$ as $r \to \infty$, where

$$K^+ = \begin{cases} \exp(-u) K_G \left\{ \frac{\xi}{\left(\sigma_1^{1/\xi} + \sigma_2^{1/\xi}\right)^\xi} \right\}^{-1/\xi} - K_6 \exp\left(-\frac{u}{\eta}\right) g\left(\frac{1}{2}\right), & \text{if } \frac{1}{2\eta} + \kappa = 1, \\ \exp(-u) K_G \left\{ \frac{\xi}{\left(\sigma_1^{1/\xi} + \sigma_2^{1/\xi}\right)^\xi} \right\}^{-1/\xi}, & \text{if } \frac{1}{2\eta} + \kappa > 1, \end{cases} \tag{A.3.11}$$

for K_G defined in (A.3.6). Note that as $\eta \in [1/2, 1]$ and $\kappa < 1/(2\eta)$, we have $\frac{1}{2\eta} + \kappa \geq 1$ only.

Case 2 We now assume that $\eta = 1$ and g satisfies Conditions 3 and 3a. Then

$$I(r) = \int_{u_1/r}^{1-u_2/r} g^*(r, w) \mathrm{d}w \sim \int_0^1 g^*(r, w) \mathrm{d}w, \tag{A.3.12}$$

for g^* defined in (A.3.9). To illustrate this, we first show that for fixed r, we have that $g^*(r, w) \sim C_1(r)$ as $w \to 0$ and $g_2^*(w) \sim C_2(r)$ as $w \to 1$, where $C_1(r), C_2(r) > 0$ are constants with respect to w. As g satisfies Condition 3, we have that

$$g^*(r, w) = \left(1 + \xi \frac{rw}{\sigma_1}\right)^{-\frac{1}{\xi}-1} \left(1 + \xi \frac{r(1-w)}{\sigma_2}\right)^{-1} \frac{1 - t_{r,w}}{t_{r,w}} \left\{ -H_{1,2}\left(1, \frac{1 - t_{r,w}}{t_{r,w}}\right) \right\},$$

where $H_{12}(1, \cdot) < 0$. We then have

$$g^*(r, w) \sim \left(1 + \xi \frac{r}{\sigma_2}\right)^{1/\xi - 1} \left\{ -H_{1,2}\left(1, \left(1 + \xi \frac{r}{\sigma_2}\right)^{1/\xi}\right) \right\} := C_1(r)$$

as $w \to 0$, which follows as $1 - w \sim 1$ and $(1 - t_{r,w})/t_{r,w} \sim (1 + \xi r(1-w)/\sigma_1)^{1/\xi}$. Conversely,

$$g^*(r, w) \sim \left(1 + \xi \frac{r}{\sigma_1}\right)^{-\frac{2}{\xi}-1} \left\{ -H_{1,2}\left(1, \left(1 + \xi \frac{r}{\sigma_1}\right)^{-\frac{1}{\xi}}\right) \right\} := C_2(r)$$

as $w \to 1$. Hence, (A.3.12) holds, and we have that, as $r \to \infty$,

$$\begin{aligned} I(r) &\sim \int_0^1 \left(1+\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{\xi}-1}\left(1+\xi\frac{r(1-w)}{\sigma_2}\right)^{-1}\frac{1-t_{r,w}}{t_{r,w}}\left\{-H_{1,2}\left(1,\frac{1-t_{r,w}}{t_{r,w}}\right)\right\}\mathrm{d}w \\ &\sim \int_0^1 \left(\xi\frac{rw}{\sigma_1}\right)^{-\frac{1}{\xi}-1}\left(\xi\frac{r(1-w)}{\sigma_2}\right)^{-1}\frac{1-t_{w}}{t_{w}}\left\{-H_{1,2}\left(1,\frac{1-t_{w}}{t_{w}}\right)\right\}\mathrm{d}w \\ &\sim \xi^{-\frac{1}{\xi}-1}(\sigma_1\sigma_2)^{\frac{1}{2\xi}}r^{-\frac{1}{\xi}-1}\int_0^1 g_2^*(w)\mathrm{d}w, \end{aligned}$$

where

$$t_{r,w} \sim \frac{\left(\frac{w}{\sigma_1}\right)^{1/\xi}}{\left(\frac{w}{\sigma_1}\right)^{1/\xi}+\left(\frac{1-w}{\sigma_2}\right)^{1/\xi}} := t_w$$

as $r \to \infty$ and where

$$g_2^*(w) = \left(\frac{\sigma_1}{\sigma_2}\right)^{\frac{1}{2\xi}}\xi^{-1}w^{-\frac{1}{2\xi}-1}(1-w)^{-1}\frac{1-t_w}{t_w}\left\{-H_{12}\left(1,\frac{1-t_w}{t_w}\right)\right\}.$$

It follows that

$$\Pr\{R \geq s|(X_1 > u_1)\cap(X_2 > u_2)\} \sim K\int_s^\infty \xi^{-\frac{1}{\xi}-1}(\sigma_1\sigma_2)^{\frac{1}{2\xi}}r^{-\frac{1}{\xi}-1}\mathrm{d}r \sim K\xi^{-\frac{1}{\xi}}(\sigma_1\sigma_2)^{\frac{1}{2\xi}}s^{-\frac{1}{\xi}}, \tag{A.3.13}$$

as $s \to \infty$ and where $K = \int_0^1 g_2^*(w)\mathrm{d}w < \infty$. From (3.3.1), we have $\Pr\{(X_1 > u_1)\cap(X_2 > u_2)\} = \exp(-u)g(1/2)$ for large u_1, u_2, and hence

$$\Pr\{(R \geq r)\cap(X_1 > u_1)\cap(X_2 > u_2)\} \sim K\exp(-u)g\left(1/2\right)\xi^{-\frac{1}{\xi}}(\sigma_1\sigma_2)^{\frac{1}{2\xi}}r^{-\frac{1}{\xi}}, \tag{A.3.14}$$

as $r \to \infty$ and for large u. Combining (A.3.7), (A.3.8) and (A.3.14), we have

$$\Pr\{R \geq r\} \sim \exp(-u) K_G \left\{ \frac{\xi}{\left(\sigma_1^{1/\xi} + \sigma_2^{1/\xi}\right)^{\xi}} \right\}^{-1/\xi} r^{-1/\xi} - K \exp(-u)\, g\left(\frac{1}{2}\right) \frac{\xi^{-\frac{1}{\xi}}}{(\sigma_1\sigma_2)^{-\frac{1}{2\xi}}} r^{-\frac{1}{\xi}},$$

as $r \to \infty$ and for K defined in (A.3.13).

Combining Cases 1 and 2 we have that $\Pr\{R \geq s\} \sim K^* s^{-\frac{1}{\xi}}$ as $s \to \infty$, where

$$K^* = \begin{cases} K^+, & \text{for Case 1,} \\ \exp(-u)\xi^{-1/\xi} \left\{ K_G(\sigma_1^{1/\xi} + \sigma_2^{1/\xi}) - \xi K g\left(\frac{1}{2}\right)(\sigma_1\sigma_2)^{\frac{1}{2\xi}} \right\}, & \text{for Case 2,} \end{cases} \tag{A.3.15}$$

and for K^+ and K_G defined in (A.3.11) and (A.3.6), respectively.

A.4 Proof of Theorem 3.3.2

We now provide the proof of Theorem 3.3.2. The general framework of the proof is similar to that of the $\xi < 0$ case for Theorem 3.3.1. We begin by deriving the joint density of (X_1, X_2) implied by the dependence model given in (3.3.1). We use the probability integral transform to perform the transformation $(X_1, X_2) \to (R, W)$, where $R = X_1 + X_2$ and W is an auxiliary variable. This particular transformation does not leave R and W independent and so we make some assumptions about the relationship between R and W, and $g(w)$, which allow us to integrate W out from the joint density analytically and derive the survival function of R. Two cases are presented for assumptions on $g(w)$ that provide different forms for $\Pr\{R \geq r\}$.

From (3.3.1) and (3.1.2), we have that

$$\Pr\left\{1-\exp\left\{-\frac{X_1}{\sigma_1}\right\}>1-\frac{1}{x_1},1-\exp\left\{-\frac{X_2}{\sigma_2}\right\}>1-\frac{1}{x_2}\right\}=\frac{\mathcal{L}(x_1+x_2)}{(x_1x_2)^{\frac{1}{2\eta}}}g\left(\frac{x_1}{x_1+x_2}\right),$$

as $x_1, x_2 \to \infty$ such that the limit of $x_1/(x_1+x_2)$ is bounded by $(0,1)$. Under the assumption that $\mathcal{L}(y)$ acts as a constant which can be absorbed by g for $y > v$ for some $v > 0$, we have $\Pr\{X_1 > x_1, X_2 > x_2\} \sim (\tilde{x}_1\tilde{x}_2)^{-\frac{1}{2\eta}} g(\omega_x)$ as $x_1, x_2 \to \infty$ such that $\omega_x = \tilde{x}_1/(\tilde{x}_1+\tilde{x}_2) \to \omega_x^* \in (0,1)$ and where $\tilde{x}_i = \exp(x_i/\sigma_i)$ for $i = 1, 2$; this implies that $x_2 \sim \sigma_2\left(\frac{x_1}{\sigma_1} + \log\left(\frac{1-\omega_x^*}{\omega_x^*}\right)\right)$ as $x_1 \to \infty$. Under the assumption that the first and second derivatives of g exist, the joint density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1,x_2) \sim \frac{(\tilde{x}_1\tilde{x}_2)^{-\frac{1}{2\eta}}}{\sigma_1\sigma_2}\left[\frac{g(\omega_x)}{4\eta^2} + (\tilde{x}_1\tilde{x}_2)\frac{\tilde{x}_1-\tilde{x}_2}{(\tilde{x}_1+\tilde{x}_2)^3}g'(\omega_x) - \frac{(\tilde{x}_1\tilde{x}_2)^2}{(\tilde{x}_1+\tilde{x}_2)^4}g''(\omega_x)\right],$$

as $x_1, x_2 \to \infty$ such that $\omega_x = \tilde{x}_1/(\tilde{x}_1+\tilde{x}_2) \to \omega_x^* \in (0,1)$. We now apply the transformation $(X_1, X_2) \to (R, W)$, where

$$R = X_1 + X_2, \quad W = \frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2},$$

where $X_1 = (\sigma_1 R + \sigma_1\sigma_2 W)/(\sigma_1+\sigma_2)$ and $X_2 = (\sigma_2 R - \sigma_1\sigma_2 W)/(\sigma_1+\sigma_2)$ with $\sigma_1\sigma_2/(\sigma_1+\sigma_2)$ the determinant of the Jacobian. Note that the limits of W and R are not independent and we have $W \in [-R/\sigma_2, R/\sigma_1]$. The density of (R, W) is

$$f_{R,W}(r,w) \sim \frac{1}{\eta(\sigma_1+\sigma_2)}\exp\left(-\frac{r}{\eta(\sigma_1+\sigma_2)}\right)g^*(w),$$

for $w \in [-r/\sigma_2, r/\sigma_1]$ and as $r \to \infty$, and where

$$g^*(w) = \eta\exp\left(-\frac{(\sigma_2-\sigma_1)w}{2\eta(\sigma_1+\sigma_2)}\right)\left[\frac{g(t_w)}{4\eta^2} + t_w(1-t_w)(2t_w-1)g'(t_w) - t_w^2(1-t_w)^2 g''(t_w)\right] \tag{A.4.1}$$

and we have

$$t_w = \frac{\exp\left(\frac{\sigma_2 w}{\sigma_1+\sigma_2}\right)}{\exp\left(\frac{\sigma_2 w}{\sigma_1+\sigma_2}\right) + \exp\left(-\frac{\sigma_1 w}{\sigma_1+\sigma_2}\right)} = \frac{\exp(w)}{\exp(w)+1} \in (0,1),$$

which follows by multiplying the denominator and numerator of t_w by $\exp(\sigma_1 w/(\sigma_1+\sigma_2))$. It follows that with $I(r) = \int_{-r/\sigma_2}^{r/\sigma_1} g^*(w)\mathrm{d}w$, as $r \to \infty$,

$$f_R(r) = \int_{-r/\sigma_2}^{r/\sigma_1} f_{R,W}(r,w)\mathrm{d}w \sim \frac{I(r)}{\eta(\sigma_1+\sigma_2)} \exp\left\{-\frac{r}{\eta(\sigma_1+\sigma_2)}\right\}. \tag{A.4.2}$$

To evaluate $I(r)$, we make different assumptions on how $g(w)$ behaves; we consider two cases, each with $I(r) < \infty$.

A.4.1 Case 1

We first make the assumption that there exists a fixed $v > 0$ such that $g(w) = 1$ for all $w \in [0,1]$ and for $r > v$. Hence, $I(r)$ in (A.4.2) becomes

$$\begin{aligned} I(r) &= \int_{-r/\sigma_2}^{r/\sigma_1} \frac{1}{4\eta} \exp\left\{-\frac{(\sigma_2-\sigma_1)w}{2\eta(\sigma_1+\sigma_2)}\right\} \mathrm{d}w \\ &= \begin{cases} \frac{(\sigma_1+\sigma_2)r}{4\eta\sigma_1\sigma_2}, & \text{if } \sigma_1 = \sigma_2, \\ \frac{\sigma_1+\sigma_2}{2(\sigma_2-\sigma_1)}\left[\exp\left\{-\frac{(\sigma_1-\sigma_2)r}{2\eta\sigma_2(\sigma_1+\sigma_2)}\right\} - \exp\left\{-\frac{(\sigma_2-\sigma_1)r}{2\eta\sigma_1(\sigma_1+\sigma_2)}\right\}\right], & \text{if } \sigma_1 \neq \sigma_2. \end{cases} \end{aligned}$$

When $\sigma_1 = \sigma_2 = \sigma$ (say), the marginal density of R is $f_R(r) \sim \frac{r}{4\eta^2\sigma^2}\exp\left\{-\frac{r}{2\eta\sigma}\right\}$ as $r \to \infty$, hence, $\Pr\{R \geq r\} \sim \frac{r}{2\eta\sigma}\exp(-r/(2\eta\sigma))$ as $r \to \infty$. Whereas when $\sigma_1 \neq \sigma_2$, we assume, without loss of generality, that $\sigma_2 > \sigma_1$. Then the marginal density of R

is

$$\begin{aligned} f_R(r) &\sim \frac{1}{2\eta(\sigma_2-\sigma_1)}\left[\exp\left\{\frac{(\sigma_2-\sigma_1)r}{2\eta\sigma_2(\sigma_1+\sigma_2)}\right\}-\exp\left\{\frac{(\sigma_1-\sigma_2)r}{2\eta\sigma_1(\sigma_1+\sigma_2)}\right\}\right]\exp\left\{-\frac{r}{\eta(\sigma_1+\sigma_2)}\right\} \\ &\sim \frac{1}{2\eta(\sigma_2-\sigma_1)}\exp\left\{\frac{(\sigma_2-\sigma_1)r}{2\eta\sigma_2(\sigma_1+\sigma_2)}\right\}\exp\left\{-\frac{r}{\eta(\sigma_1+\sigma_2)}\right\} \\ &\sim \frac{1}{2\eta(\sigma_2-\sigma_1)}\exp\left\{-\frac{r}{2\eta\sigma_2}\right\}, \end{aligned}$$

as $r\to\infty$ and so $\Pr\{R\geq s\}\sim\frac{\sigma_2}{\sigma_2-\sigma_1}\exp\left\{-\frac{s}{2\eta\sigma_2}\right\}$ as $s\to\infty$. By symmetry, this can be written as

$$\Pr\{R\geq r\}\sim\frac{\sigma_{max}}{\sigma_{max}-\sigma_{min}}\exp\left\{-\frac{r}{2\eta\sigma_{max}}\right\},$$

as $r\to\infty$ and where $\sigma_{max}=\max\{\sigma_1,\sigma_2\}$ and $\sigma_{min}=\min\{\sigma_1,\sigma_2\}$.

A.4.2 Case 2

We now assume that $\eta=1$ and that Conditions 3 and 3a hold. In this case, we can show that

$$I(r)=\int_{-r/\sigma_2}^{r/\sigma_1}g^*(w)\mathrm{d}w\sim\int_{-\infty}^{\infty}g^*(w)\mathrm{d}w:=K, \tag{A.4.3}$$

for finite constant $K>0$. To show this, we first derive that $g^*(w)\to 0$ at an exponential rate as $w\to\infty$ or as $w\to-\infty$. From Condition 3 and (A.4.1) and with $\eta=1$, it follows that

$$\begin{aligned} g^*(w) &= -\exp\left(-\frac{(\sigma_2-\sigma_1)w}{2(\sigma_1+\sigma_2)}\right)\left(\frac{1-t_w}{t_w}\right)^{3/2}H_{12}\left(1,\frac{1-t_w}{t_w}\right) \\ &= -\exp\left(\left\{\frac{\sigma_1}{(\sigma_1+\sigma_2)}-2\right\}w\right)H_{12}\left(1,\exp(-w)\right)=-\exp(aw)H_{12}\left(1,\exp(-w)\right) \end{aligned}$$

where $a = \sigma_1/(\sigma_1 + \sigma_2) - 2 \in (-2, -1)$. It then follows that

$$g^*(w) = -\exp(aw)H_{12}\left(1, \exp(-w)\right) \sim \begin{cases} K_{H_1}\exp((a-c_1)w) \to 0 \text{ as } w \to \infty, \\ K_{H_2}\exp((a-c_2)w) \to 0 \text{ as } w \to -\infty, \end{cases}$$

where the first limit follows as $a - c_1 < -2 < 0$ and the second follows as $a - c_2 > 0$. Hence, (A.4.3) holds, and it follows that the survival function of R as $s \to \infty$ is

$$\Pr\{R \geq s\} \sim \int_s^\infty K\frac{1}{(\sigma_1+\sigma_2)}\exp\left\{-\frac{r}{(\sigma_1+\sigma_2)}\right\}\mathrm{d}r = K\exp\left\{-\frac{s}{(\sigma_1+\sigma_2)}\right\}.$$

A.5 Proof of Theorem 3.3.3

The general framework of the proof follows that of the $\xi < 0$ case for Theorem 3.3.1. We begin by deriving the joint density of (X_1, X_2) implied by the dependence model given in (3.3.1), on GPD margins. We use the probability integral transform to perform an initial transformation $(X_1, X_2) \to (R, W)$, where $R = X_1 + X_2$ and W is an auxiliary variable, chosen so that we are able to show that the support of W is independent of R for R greater than some threshold. At this point, we make two different assumptions on how $g(w)$ acts as $w \to 1$ and $w \to 0$; for the first case, we can simply marginalise W out of f_{RW} and derive the survival function of R. For the second case, we find that we must perform another transformation $(R, W) \to (V, Z)$ where V is a normalisation of R and Z is an auxiliary variable chosen so that it is possible to integrate over and derive the marginal density of V. We then transform V back to R and derive the survival function.

Combining (3.3.1) and (3.1.2), we have that

$$\Pr\left\{1-\left(1+\xi_1\frac{X_1}{\sigma_1}\right)^{-1/\xi_1}>1-\frac{1}{x_1},1-\left(1+\xi_2\frac{X_2}{\sigma_2}\right)^{-1/\xi_2}>1-\frac{1}{x_2}\right\} = \frac{\mathcal{L}(x_1+x_2)}{(x_1x_2)^{\frac{1}{2\eta}}}g\left(\frac{x_1}{x_1+x_2}\right),$$

as $x_1, x_2 \to \infty$ such that the limit of $x_1/(x_1+x_2)$ is bounded by $(0,1)$. Under the assumption that $\mathcal{L}(y)$ acts as a constant which can be absorbed by g for $y > v$ for some $v > 0$, we have $\Pr\{X_1 > x_1, X_2 > x_2\} \sim \tilde{x}_1^{-\frac{1}{2\eta\xi_1}}\tilde{x}_2^{-\frac{1}{2\eta\xi_2}}g(\omega_x)$ for $x_1 \to x_1^F$ and $x_2 \to x_2^F$, such that $\omega_x = \tilde{x}_1^{1/\xi_1}/(\tilde{x}_1^{1/\xi_1}+\tilde{x}_2^{1/\xi_2}) \to \omega_x^* \in (0,1)$ and where $\tilde{x}_i = (1+\xi_i x_i/\sigma_i)$ for $i = 1, 2$. Assuming that the first and second derivatives of g exist, then the density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1,x_2) \sim \frac{\tilde{x}_1^{-\frac{1}{2\eta\xi_1}-1}\tilde{x}_2^{-\frac{1}{2\eta\xi_2}-1}}{\sigma_1\sigma_2} \times\left[\frac{g(\omega_x)}{4\eta^2}+\tilde{x}_1^{\frac{1}{\xi_1}}\tilde{x}_2^{\frac{1}{\xi_2}}\frac{\tilde{x}_1^{1/\xi_1}-\tilde{x}_2^{1/\xi_2}}{\left(\tilde{x}_1^{1/\xi_1}+\tilde{x}_2^{1/\xi_2}\right)^3}g'(\omega_x)-\frac{\tilde{x}_1^{\frac{2}{\xi_1}}\tilde{x}_2^{\frac{2}{\xi_2}}}{\left(\tilde{x}_1^{1/\xi_1}+\tilde{x}_2^{1/\xi_2}\right)^4}g''(\omega_x)\right],$$

as $x_1 \to x_1^F$ and $x_2 \to x_2^F$ such that $\omega_x \to \omega_x^* \in (0,1)$. We now apply the transformation $(X_1, X_2) \to (R, W)$, where

$$R = X_1 + X_2, \quad W = \frac{\left(\frac{\sigma_1}{\xi_1}+X_1\right)}{\left(\frac{\sigma_1}{\xi_1}+X_1\right)+\left(\frac{\sigma_2}{\xi_2}+X_2\right)},$$

where $1+\xi_1X_1/\sigma_1 = -\frac{\xi_1}{\sigma_1}(r^F-R)W$ and $1+\xi_2X_2/\sigma_2 = -\frac{\xi_2}{\sigma_2}(r^F-R)(1-W)$ for $r^F = -(\sigma_1/\xi_1+\sigma_2/\xi_2)$ the upper-endpoint of R and where the numerator and denominator of W are both negative. The density of (R, W) as $r \to r^F$ is

$$f_{R,W}(r,w) \sim \frac{(-\xi_1)^{-\frac{1}{2\eta\xi_1}-1}(-\xi_2)^{-\frac{1}{2\eta\xi_2}-1}}{\sigma_1^{-\frac{1}{2\eta\xi_1}}\sigma_2^{-\frac{1}{2\eta\xi_2}}}\left(r^F-r\right)^{-\frac{1}{2\eta\xi_1}-\frac{1}{2\eta\xi_2}-1}w^{-\frac{1}{2\eta\xi_1}-1}(1-w)^{-\frac{1}{2\eta\xi_2}-1}$$

$$\times \left[\frac{g\left(t_{r,w}\right)}{4\eta^2} + t_{r,w}(1-t_{r,w})(2t_{r,w}-1)g^{'}\left(t_{r,w}\right) - t_{r,w}^2(1-t_{r,w})^2 g^{''}\left(t_{r,w}\right) \right], \tag{A.5.1}$$

as $r \to r^F$, which includes the determinant, $(r^F - r)$, of the Jacobian of the transformation and where

$$t_{r,w} = \frac{\left\{ \frac{-\xi_1}{\sigma_1} \left(r^F - r\right) w \right\}^{1/\xi_1}}{\left\{ \frac{-\xi_1}{\sigma_1} \left(r^F - r\right) w \right\}^{1/\xi_1} + \left\{ \frac{-\xi_2}{\sigma_2} \left(r^F - r\right) (1-w) \right\}^{1/\xi_2}}.$$

Recall from the proof of Theorem 3.3.1 that we use this transformation to ensure that the support of $W \in [0,1]$ is independent of R, given that R is above a fixed threshold $u > 0$; here we show that this holds when $u = \max\{-\sigma_1/\xi_1, -\sigma_2/\xi_2\}$. This proof is identical in its layout to the similar proof given in Section A.3.1, however, the details are slightly different as here we have $\xi_1 \neq \xi_2$, rather than equal shape parameters. Let $x^{max} = \max\{-\sigma_1/\xi_1, -\sigma_2/\xi_2\}$ and $x^{min} = \min\{-\sigma_1/\xi_1, -\sigma_2/\xi_2\}$. As $X_1 + X_2 \leq -(\sigma_1/\xi_1 + \sigma_2/\xi_2) = x^{max} + x^{min}$, there exists a random variable $P \in [0,1]$, such that

$$X_1 + X_2 | (X_1 + X_2 > x^{max}) = x^{max} + x^{min} P.$$

Now for $i = \underset{j=1,2}{\arg\max}\{-\sigma_j/\xi_j\}$ let $X_i|(X_1 + X_2 > x^{max}) = x^{max} Q_i$ for random $Q_i \leq 1$. Then as

$$0 \leq (X_1 + X_2 - X_i)|(X_1 + X_2 > x^{max}) \leq x^{min} \Rightarrow 0 \leq x^{max} + x^{min} P - x^{max} Q_i \leq x^{min},$$

it follows that Q_i must satisfy

$$-\frac{P}{1-P} < 0 \leq \frac{x^{max}(1-Q_i)}{x^{min}(1-P)} \leq 1.$$

Now consider W. If we have $i = 1$, then

$$W = \frac{\frac{\sigma_1}{\xi_1} + X_1}{\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + X_1 + X_2} = \frac{-x^{max} + x^{max}Q_1}{-x^{max} - x^{min} + x^{max} + x^{min}P} = \frac{x^{max}(1 - Q_1)}{x^{min}(1 - P)},$$

and so $W \in [0, 1]$ as needed. However, if $i = 2$ we instead have $W = 1 - x^{max}(1 - Q_2)/\{x^{min}(1 - P)\}$; this also gives $W \in [0, 1]$. So the support of W does not depend on R when $R > u$.

We now explore how $f_{R,W}(r, w)$ and $f_R(r)$ behave for $r \to r^F$, i.e., hence for r such that $r > u$. Without loss of generality, we assume that $0 > \xi_1 > \xi_2$, with the other case following by symmetry. If $\xi_1 > \xi_2$, we have $t_{r,w} \to 1$ as $r \to r^F$, and so require assumptions on how $g(t)$ behaves as $t \to 1$. We consider two cases:

A.5.1 Case 1

We now assume that Conditions 2 holds. Then the joint density of (R, W) is

$$\begin{aligned} f_{R,W}(r, w) &\sim K_1 \left(r^F - r\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2} - 1} \left[\frac{1}{4\eta^2} - \kappa(2t_{r,w} - 1) - \kappa(\kappa - 2t_{r,w} + 1)\right] g^*(w) \\ &\sim K_1 \left(r^F - r\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2} - 1} \left[\frac{1}{4\eta^2} - \kappa^2\right] g^*(w), \end{aligned}$$

as $r \to r^F$, where $g^*(w) = K_2 w^{-\frac{1}{2\eta\xi_1} - \frac{\kappa}{\xi_1} - 1}(1 - w)^{-\frac{1}{2\eta\xi_2} + \frac{\kappa}{\xi_2} - 1}$, and for constants

$$K_1 = -\frac{\left(\frac{1+2\eta\kappa}{2\eta\xi_1} + \frac{1-2\eta\kappa}{2\eta\xi_2}\right)}{\left(r^F\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2}}} > 0 \text{ and } K_2 = -\frac{K_g(-\xi_1)^{-\frac{1}{2\eta\xi_1} - \frac{\kappa}{\xi_1} - 1}(-\xi_2)^{-\frac{1}{2\eta\xi_2} + \frac{\kappa}{\xi_2} - 1}}{K_1 \sigma_1^{-\frac{1}{2\eta\xi_1} - \frac{\kappa}{\xi_1}} \sigma_2^{-\frac{1}{2\eta\xi_2} + \frac{\kappa}{\xi_2}}} > 0.$$

The survival function of R is

$$\begin{aligned}\Pr\{R \geq s\} &\sim \left[\frac{1}{4\eta^2} - \kappa^2\right] \int_s^\infty \int_0^1 K_1 \left(r^F - r\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2} - 1} g^*(w)\mathrm{d}w\mathrm{d}r \\ &\sim KK_1 \int_s^\infty \left(r^F - r\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2} - 1} \mathrm{d}r \\ &\sim K \left(1 + \xi_1\xi_2 \frac{s}{\sigma_1\xi_2 + \sigma_2\xi_2}\right)^{-\frac{1+2\eta\kappa}{2\eta\xi_1} - \frac{1-2\eta\kappa}{2\eta\xi_2}}, \end{aligned} \tag{A.5.2}$$

as $s \to r^F$, and where

$$K = \left[\frac{1}{4\eta^2} - \kappa^2\right] \int_0^1 g^*(w)\mathrm{d}w = K_2 \left[\frac{1}{4\eta^2} - \kappa^2\right] B\left(-\frac{1}{2\eta\xi_1} - \frac{\kappa}{\xi_1}, -\frac{1}{2\eta\xi_2} + \frac{\kappa}{\xi_2}\right) > 0 \tag{A.5.3}$$

is constant; here $B(\cdot,\cdot)$ denotes the beta function and both of its arguments are positive, and we note that the limits in (A.5.2) hold as $f_{R,W}$ is a valid probability density. The general result follows by replacing ξ_1 and ξ_2 with $\max\{\xi_1, \xi_2\}$ and $\min\{\xi_1, \xi_2\}$ respectively and using the behaviour of g as $t \to 0$ as well as $t \to 1$.

A.5.2 Case 2

We now assume that Conditions 3 and 3b hold. From (A.5.1), the joint density of (R, W) for $w \in [0, 1]$ is

$$\begin{aligned} f_{R,W}(r,w) &\sim \left(r^F - r\right)^{-\frac{1}{\eta\xi_1} - 1} \\ &\quad \times \left[(2\kappa + 2t_{r,w} - 1)t_{r,w}(1 - t_{r,w})L^{'}(t_{r,w}) - t_{r,w}^2(1 - t_{r,w})^2 L^{''}(t_{r,w})\right] g^*(w) \\ &\sim \left(r^F - r\right)^{-\frac{1}{\eta\xi_1} - 1} \left[-\frac{1 - t_{r,w}}{t_{r,w}} H_{12}\left(1, \frac{1 - t_{r,w}}{t_{r,w}}\right)\right. \\ &\quad \left. + (2\kappa - 1)\left\{\frac{t_{r,w}}{1 - t_{r,w}} + \frac{1 - t_{r,w}}{t_{r,w}} H_2\left(1, \frac{1 - t_{r,w}}{t_{r,w}}\right)\right\}\right] g^*(w), \end{aligned} \tag{A.5.4}$$

as $r \to r^F$ and which follows by exploiting the homogeneity properties of H_2 and H_{12}, and

$$g^*(w) = (-\xi_1)^{-\frac{1}{\eta\xi_1}-1}(-\xi_2)^{-1}\sigma_1^{\frac{1}{\eta\xi_1}} w^{-\frac{1}{\eta\xi_1}-1}(1-w)^{-1}.$$

To marginalise W out of (A.5.4), we make the transformation $(R, W) \to (V, Z)$, where $V = (r^F - R)/r^F$ and $W = 1 - V^Z$ for $Z \in (0, \infty)$, and so large R now corresponds to small positive V. The determinant of the Jacobian of this transformation is $(r^F)^{-1} \times \big(-v^Z \log(v)\big)$, and the joint density of (V, Z) for $z \in (0, \infty)$ and as $v \downarrow 0$ is

$$f_{V,Z}(v,z) \sim K_2\left(-\log(v)\right)(vr^F)^{-\frac{1}{\eta\xi_1}-\frac{1}{\xi_1}+\frac{1}{\xi_2}-1} v^{z/\xi_2}\Bigg[- x_r H_{12}\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}\right) + (2\kappa-1)\Bigg\{ x_r^{-1}(vr^F)^{2/\xi_1-2/\xi_2}v^{-2z/\xi_2} + x_r H_2\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}\right)\Bigg\}\Bigg],$$

which follows by exploiting $1 - v^z \sim 1$ as $v \downarrow 0$ for $z > 0$, and for constants

$$K_2 = (-\xi_1)^{-\frac{1}{\eta\xi_1}-1}(-\xi_2)^{-1}\sigma_1^{\frac{1}{\eta\xi_1}}(r^F)^{-1} > 0$$

and $x_i^F = -\xi_i/\sigma_i$ for $i = 1, 2$ and the ratio $x_r = (x_2^F)^{1/\xi_2}/(x_1^F)^{1/\xi_1}$. Consider now the integrals

$$\begin{aligned} I_1(v) &= \frac{x_r}{\xi_2}\int_0^\infty \log(v)(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}H_{12}\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}\right)\mathrm{d}z \\ &= H_1\left(1,\infty\right) - H_1\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}\right), \end{aligned}$$

and

$$\begin{aligned} I_2(v) &= \frac{x_r}{\xi_2}\int_0^\infty \log(v)(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}H_2\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}v^{z/\xi_2}\right)\mathrm{d}z \\ &= H\left(1,\infty\right) - H\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}\right). \end{aligned}$$

It follows that, as $f_V(v) = \int_0^\infty f_{V,Z}(v,z)\mathrm{d}z$ is, as $v \downarrow 0$,

$$\begin{aligned} f_V(v) &\sim K_2\xi_2(vr^F)^{-\frac{1}{\eta\xi_1}-1} \\ &\quad\times\left[I_1(v) - (2\kappa-1)\left\{(\xi_2 x_r)^{-1}\int_0^\infty \log(v)(vr^F)^{1/\xi_1-1/\xi_2}v^{-z/\xi_2}\mathrm{d}z + I_2(v)\right\}\right] \\ &\sim K_2\xi_2(vr^F)^{-\frac{1}{\eta\xi_1}-1}\Bigg[H_1\left(1,\infty\right) - H_1\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}\right) \\ &\quad - (2\kappa-1)\left\{x_r^{-1}(vr^F)^{1/\xi_1-1/\xi_2} + H\left(1,\infty\right) - H\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}\right)\right\}\Bigg] \\ &\sim K_2\xi_2\left\{H_1\left(1,\infty\right) - (2\kappa-1)\right\}(vr^F)^{-\frac{1}{\eta\xi_1}-1}, \end{aligned} \tag{A.5.5}$$

where $H(1,\infty) = 1$. The last line of (A.5.5) follows as $1/\xi_2 - 1/\xi_1 > 0$ and $H_1(1,z) \to 0$ as $z \to 0$; furthermore, as $H(1,z) \sim 1/z$ as $z \to 0$, we have that $H\left(1, x_r(vr^F)^{-1/\xi_1+1/\xi_2}\right) \sim x_r^{-1}(vr^F)^{1/\xi_1-1/\xi_2}$, as $v \downarrow 0$. Transforming back to R, then the density of R is $f_R(r) \sim \xi_2\left\{H_1\left(1,\infty\right) - (2\kappa-1)\right\}K_2 r^F(r^F-r)^{-\frac{1}{\eta\xi_1}-1}$ as $r \to r^F$ and so

$$\Pr\{R \geq r\} \sim K_3\left(1 + \xi_1\xi_2\frac{r}{\sigma_1\xi_2 + \sigma_2\xi_2}\right)^{-\frac{1}{\eta\xi_1}}$$

for constant

$$K_3 = \left|H_1\left(1,\infty\right) - (2\kappa-1)\right|\eta(-\xi_1)^{-\frac{1}{\eta\xi_1}}(r^F)^{-\frac{1}{\eta\xi_1}}\sigma_1^{\frac{1}{\eta\xi_1}} > 0. \tag{A.5.6}$$

The limits in the respective integrals that lead to $\Pr\{R \geq s\}$ are valid as $f_{R,W}$ and $f_{V,Z}$ are valid joint densities. The general result follows by replacing ξ_1 and ξ_2 with $\max\{\xi_1,\xi_2\}$ and $\min\{\xi_1,\xi_2\}$ respectively.

A.6 Proof of Theorem 3.3.4

We show that the result holds for the two limiting cases of positive association between X_1 and X_2, namely perfect dependence and independence; this implies that the results

hold for any cases where X_1 and X_2 have positive association. To illustrate why this is possible, let R_I be R such that X_1 and X_2 are independent, and R_D be R such that X_1 and X_2 are perfectly-dependent, i.e., X_2 is some, possibly non-linear, function of X_1. It is clear that, for any $y > 0$, we have

$$\min\left\{\Pr\{R_I \leq y\}, \Pr\{R_D \leq y\}\right\} \leq \Pr\{R \leq y\} \leq \max\left\{\Pr\{R_I \leq y\}, \Pr\{R_D \leq y\}\right\},$$

and hence, if we have that $\Pr\{R_I \geq y\} \sim C_1 S(y)$ and $\Pr\{R_D \geq y\} \sim C_2 S(y)$ for some function $S(y)$ and constants $C_1, C_2 > 0$ and as $y \to \infty$, we have that $\Pr\{R \geq y\} \sim CS(y)$ for $C \in [C_1, C_2]$ also holds.

The proof follows by considering the limiting cases of perfect dependence and independence, separately. For $X_1 \sim \text{GPD}(\sigma_1, \xi_1)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi_2)$ consider four cases: $(\xi_1 > 0, \xi_2 < 0)$, $(\xi_1 > 0, \xi_2 = 0)$, $(\xi_1 = 0, \xi_2 < 0)$ and $(\xi_1 > \xi_2, \xi_2 > 0)$; the other cases follow by symmetry.

A.6.1 Perfect dependence

We begin by considering those cases where X_1 and X_2 are perfectly dependent; this is induced by letting $X_2 = F_2^{-1}\{F_1(X_1)\}$. We then illustrate that

$$\begin{aligned}\Pr\{R \geq r\} &= \Pr\{X_1 + X_2 \geq r\} = \Pr\{X_1 + F_2^{-1}\{F_1(X_1)\} \geq r\} \\ &= \Pr\{X_1 \geq x^*\} \sim \Pr\{X_1 \geq r\},\end{aligned}$$

as $r \to \infty$ and where x^* solves $r = x^* + F_2^{-1}\{F_1(x^*)\}$. In each case, we find an approximate solution for x^* as $r \to \infty$, using an iterative procedure.

Case 1 Let $X_1 \sim \text{GPD}(\sigma_1, \xi_1 > 0)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi_2 < 0)$. From (3.1.2), we have that

$$X_2 = \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}X_1\right)^{\xi_2/\xi_1}\right].$$

To solve for x^*, we begin with the initial solution $x_0^* = r$ and consider $x_1^* = r + \epsilon$. We then have

$$r = r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}(r+\epsilon)\right)^{\xi_2/\xi_1}\right] \sim r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1} r\right)^{\xi_2/\xi_1}\right],$$

as $r \to \infty$. Hence, $\epsilon = -\frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1} r\right)^{\xi_2/\xi_1}\right]$ and an approximate solution for x^* is

$$x^* \sim r + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1} r\right)^{\xi_2/\xi_1}\right] = r\left\{1 + O\left(r^{\xi_2/\xi_1 - 1}\right)\right\},$$

as $r \to \infty$. It follows that

$$\Pr\{R \geq r\} = \left(1 + \frac{\xi_1}{\sigma_1} r\left\{1 + O\left(r^{\xi_2/\xi_1 - 1}\right)\right\}\right)^{-1/\xi_1} \sim \left(1 + \frac{\xi_1}{\sigma_1} r\right)^{-1/\xi_1} = \Pr\{X_1 \geq r\},$$

as $r \to \infty$, and as $\xi_2/\xi_1 - 1 < 0$.

Case 2 Let $X_1 \sim \text{GPD}(\sigma_1, \xi_1 > 0)$ and $X_2 \sim \text{GPD}(\sigma_2, 0)$. From (3.1.2), we have that

$$X_2 = \frac{\sigma_2}{\xi_1} \log\left(1 + \frac{\xi_1}{\sigma_1} X_1\right).$$

To solve for x^*, we begin with the initial solution $x_0^* = r$ and consider $x_1^* = r + \epsilon$. We then have

$$r = r + \epsilon + \frac{\sigma_1}{\xi_1} \log\left(1 + \frac{\xi_1}{\sigma_1}(r+\epsilon)\right) \sim r + \epsilon + \frac{\sigma_1}{\xi_1} \log\left(1 + \frac{\xi_1}{\sigma_1} r\right),$$

as $r \to \infty$. Hence, $\epsilon = -\frac{\sigma_1}{\xi_1} \log\left(1 + \frac{\xi_1}{\sigma_1}\right)$ and an approximate solution for x^* is

$$x^* \sim r + \frac{\sigma_1}{\xi_1} \log\left(1 + \frac{\xi_1}{\sigma_1}\right) = r\left\{1 + O\left(\frac{\log(r)}{r}\right)\right\},$$

as $r \to \infty$. It follows that

$$\Pr\{R \geq r\} = \left(1 + \frac{\xi_1}{\sigma_1} r \left\{1 + O\left(\frac{\log(r)}{r}\right)\right\}\right)^{-1/\xi_1} \sim \left(1 + \frac{\xi_1}{\sigma_1} r\right)^{-1/\xi_1} = \Pr\{X_1 \geq r\},$$

as $r \to \infty$.

Case 3 Let $X_1 \sim \text{GPD}(\sigma_1, 0)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi_2 < 0)$. From (3.1.2), we have that

$$X_2 = \frac{\sigma_2}{\xi_2}\left[-1 + \exp\left(\frac{X_1}{\xi_2 \sigma_1}\right)\right].$$

To solve for x^*, we begin with the initial solution $x_0^* = r$ and consider $x_1^* = r + \epsilon$. We then have

$$r = r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \exp\left(\frac{r+\epsilon}{\xi_2 \sigma_1}\right)\right] \sim r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \exp\left(\frac{r}{\xi_2 \sigma_1}\right)\right],$$

as $r \to \infty$. Hence, $\epsilon = -\frac{\sigma_2}{\xi_2}\left[-1 + \exp\left(\frac{r}{\xi_2 \sigma_1}\right)\right]$ and an approximate solution for x^* is

$$x^* \sim r - \frac{\sigma_2}{\xi_2}\left[-1 + \exp\left(\frac{r}{\xi_2 \sigma_1}\right)\right] = r\left\{1 - O\left(\frac{\exp(\{\sigma_1 \xi_2\}^{-1} r)}{r}\right)\right\} + \frac{\sigma_2}{\xi_2},$$

as $r \to \infty$. It follows that

$$\begin{aligned}\Pr\{R \geq r\} &= \exp\left(-\frac{r}{\sigma_1}\left\{1 - O\left(\frac{\exp(\{\sigma_1 \xi_2\}^{-1} r)}{r}\right)\right\} - \frac{\sigma_2}{\sigma_1 \xi_2}\right) \\ &\sim C \exp\left(-\frac{r}{\sigma_1}\right) = C \Pr\{X_1 \geq r\}, \end{aligned} \tag{A.6.1}$$

as $r \to \infty$ and for $C = \exp(-\sigma_2(\sigma_1 \xi_2)^{-1}) > 0$, and as $\xi_2 < 0$.

Case 4 Let $X_1 \sim \text{GPD}(\sigma_1, \xi_1 > 0)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi_2 > 0)$ with $\xi_1 > \xi_2$. From (3.1.2), we have that

$$X_2 = \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}X_1\right)^{\xi_2/\xi_1}\right].$$

To solve for x^*, we begin with the initial solution $x_0^* = r$ and consider $x_1^* = r + \epsilon$. We then have

$$r = r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}(r+\epsilon)\right)^{\xi_2/\xi_1}\right] \sim r + \epsilon + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}r\right)^{\xi_2/\xi_1}\right],$$

as $r \to \infty$. Hence, $\epsilon = -\frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}r\right)^{\xi_2/\xi_1}\right]$ and an approximate solution for x^* is

$$x^* \sim r + \frac{\sigma_2}{\xi_2}\left[-1 + \left(1 + \frac{\xi_1}{\sigma_1}r\right)^{\xi_2/\xi_1}\right] = r\left\{1 + O\left(r^{\xi_2/\xi_1 - 1}\right)\right\},$$

as $r \to \infty$. It follows that

$$\Pr\{R \geq r\} = \left(1 + \frac{\xi_1}{\sigma_1}r\left\{1 + O\left(r^{\xi_2/\xi_1 - 1}\right)\right\}\right)^{-1/\xi_1} \sim \left(1 + \frac{\xi_1}{\sigma_1}r\right)^{-1/\xi_1} = \Pr\{X_1 \geq r\},$$

as $r \to \infty$, and as $\xi_2/\xi_1 - 1 < 0$.

A.6.2 Independence

We now consider the cases where X_1 and X_2 are independent. In each case, we consider the joint density of (X_1, X_2) and use a marginal transformation to R and some auxiliary variable W. We then derive the marginal density of R as $r \to \infty$.

Case 1 Let $X_1 \sim \text{GPD}(\sigma_1, \xi_1 > 0)$ and $X_2 \sim \text{GPD}(\sigma_2, \xi_2 < 0)$. The density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sigma_1\sigma_2}\left(1 + \frac{\xi_1}{\sigma_1}x_1\right)^{-1/\xi_1-1}\left(1 + \frac{\xi_2}{\sigma_2}x_2\right)^{-1/\xi_2-1},$$

with $x_1 \in (0, \infty)$ and $x_2 \in (0, -\sigma_2/\xi_2)$. Using the transformation $(X_1, X_2) \to (R, W = X_2)$, the joint density of (R, W) is

$$f_{R,W}(r, w) = \frac{1}{\sigma_1\sigma_2}\left(1 + \frac{\xi_1}{\sigma_1}(r - w)\right)^{-1/\xi_1-1}\left(1 + \frac{\xi_2}{\sigma_2}w\right)^{-1/\xi_2-1},$$

for $r \in (0, \infty)$ and $w \in [0, -\sigma_2/\xi_2]$. Note that as $w = x_2$ has a finite upper-endpoint $-\sigma_2/\xi_2$, we have that

$$f_{R,W}(r, w) \sim \frac{1}{\sigma_1\sigma_2}\left(\frac{\xi_1}{\sigma_1}r\right)^{-1/\xi_1-1}\left(1 + \frac{\xi_2}{\sigma_2}w\right)^{-1/\xi_2-1},$$

as $r \to \infty$. Hence,

$$\begin{aligned} f_R(r) &= \int_0^{-\sigma_2/\xi_2} f_{R,W}(r, w)\mathrm{d}w \sim \int_0^{-\sigma_2/\xi_2} \frac{1}{\sigma_1\sigma_2}\left(\frac{\xi_1}{\sigma_1}r\right)^{-1/\xi_1-1}\left(1 + \frac{\xi_2}{\sigma_2}w\right)^{-1/\xi_2-1}\mathrm{d}w \\ &\sim \frac{1}{\sigma_1}\left(\frac{\xi_1}{\sigma_1}r\right)^{-1/\xi_1-1} \sim f_{X_1}(r), \end{aligned}$$

and hence $\Pr\{R \geq r\} \sim \Pr\{X_1 \geq r\}$ as $r \to \infty$.

Case 2 Let $X_1 \sim \text{GPD}\{\sigma_1, \xi > 0\}$ and let $X_2 \sim \text{GPD}\{\sigma_2, 0\}$. The density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sigma_1\sigma_2}\left(1 + \frac{\xi}{\sigma_1}x_1\right)^{-1/\xi-1}\exp\left\{-\frac{1}{\sigma_2}x_2\right\}, \quad (x_1, x_2 \geq 0).$$

We now use the transformation $(X_1, X_2) \to (R, W = X_1/R)$, which has Jacobian determinant $|J| = R$. The density of (R, W) is

$$f_{R,W}(r,w) = \frac{r}{\sigma_1\sigma_2}\left(1+\frac{\xi}{\sigma_1}rw\right)^{-1/\xi-1}\exp\left\{-\frac{1}{\sigma_2}r(1-w)\right\},$$

for $r \in (0,\infty)$ and $w \in [0,1]$ and it follows that

$$\begin{aligned}
f_R(r) &= \int_0^1 f_{R,W}(r,w)\mathrm{d}w = \frac{r}{\sigma_1\sigma_2}\int_0^1\left(1+\frac{\xi}{\sigma_1}rw\right)^{-1/\xi-1}\exp\left\{-\frac{1}{\sigma_2}r(1-w)\right\}\mathrm{d}w \\
&= \frac{r}{\sigma_1\sigma_2}\left[\frac{\sigma_2}{r}\left(1+\frac{\xi}{\sigma_1}r\right)^{-1/\xi-1} - \frac{\sigma_2}{r}\exp\left\{-\frac{r}{\sigma_2}\right\}\right. \\
&\left. + \frac{\sigma_2\xi}{(1/\xi-1)\sigma_1}\int_0^1\left(1+\frac{\xi}{\sigma_1}rw\right)^{-1/\xi-2}\exp\left\{-\frac{1}{\sigma_2}r(1-w)\right\}\mathrm{d}w\right] \\
&= \frac{r}{\sigma_1\sigma_2}\left[\frac{\sigma_2}{r}\left(1+\frac{\xi}{\sigma_1}r\right)^{-1/\xi_1-1} - \frac{\sigma_2}{r}\exp\left\{-\frac{r}{\sigma_2}\right\}\right. \\
&+ \frac{\sigma_2\xi}{(1/\xi-1)\sigma_1}\left\{\frac{\sigma_2}{r}\left(1+\frac{\xi}{\sigma_1}r\right)^{-1/\xi-2} - \frac{\sigma_2}{r}\exp\left\{-\frac{r}{\sigma_2}\right\}\right. \\
&\left.\left. + \frac{\sigma_2\xi}{(1/\xi-2)\sigma_1}\int_0^1\left(1+\frac{\xi}{\sigma_1}rw\right)^{-1/\xi-3}\exp\left\{-\frac{1}{\sigma_2}r(1-w)\right\}\mathrm{d}w\right\}\right] \\
&= \frac{1}{\sigma_1}\left(1+\frac{\xi}{\sigma_1}r\right)^{-1/\xi_1-1}\left[1 - O\left(r^{1/\xi+1}\exp(-r)\right) - O(r^{-1})\right] \\
&\sim \frac{1}{\sigma_1}\left(1+\frac{\xi}{\sigma_1}r\right)^{-1/\xi_1-1} = f_{X_1}(r),
\end{aligned}$$

as $r \to \infty$, and so $\Pr\{R \geq r\} \sim \Pr\{X_1 \geq r\}$ as $r \to \infty$.

Case 3 We now let $X_1 \sim \mathrm{GPD}(\sigma_1, 0)$ and $X_2 \sim \mathrm{GPD}(\sigma_2, \xi_2 < 0)$. The density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\sigma_1\sigma_2}\exp\left\{-\frac{x_1}{\sigma_1}\right\}\left(1+\frac{\xi_2}{\sigma_2}x_2\right)^{-1/\xi_2-1},$$

with $x_1 \in (0, \infty)$ and $x_2 \in (0, -\sigma_2/\xi_2)$. Using the transformation $(X_1, X_2) \to (R, W = X_2)$, the density of (R, W) is

$$f_{R,W}(r, w) = \frac{1}{\sigma_1 \sigma_2} \exp\left\{-\frac{r-w}{\sigma_1}\right\} \left(1 + \frac{\xi_2}{\sigma_2} w\right)^{-1/\xi_2 - 1},$$

and hence

$$\begin{aligned} f_R(r) &= \int_0^{-\sigma_2/\xi_2} \frac{1}{\sigma_1 \sigma_2} \exp\left\{-\frac{r}{\sigma_1}\right\} \exp\left\{\frac{w}{\sigma_1}\right\} \left(1 + \frac{\xi_2}{\sigma_2} w\right)^{-1/\xi_2 - 1} \mathrm{d}w \\ &= \frac{C}{\sigma_1} \exp\left\{-\frac{r}{\sigma_1}\right\} = C f_{X_1}(r), \end{aligned} \tag{A.6.2}$$

for $C = \int_0^{-\sigma_2/\xi_2} \sigma_2^{-1} \exp\{w/\sigma_1\} (1 + \xi_2 w/\sigma_2)^{-1/\xi_2 - 1} \mathrm{d}w > 0$. Hence, it follows that

$$\Pr\{R \geq r\} = C \Pr\{X_1 \geq r\}.$$

Case 4 We now let $X_1 \sim \mathrm{GPD}(\sigma_1, \xi_1 > 0)$ and $X_2 \sim \mathrm{GPD}(\sigma_2, \xi_2 > 0)$ for $\xi_1 > \xi_2$. The density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sigma_1 \sigma_2} \left(1 + \frac{\xi_1 x_1}{\sigma_1}\right)^{-1/\xi_1 - 1} \left(1 + \frac{\xi_2 x_2}{\sigma_2}\right)^{-1/\xi_2 - 1}, \quad (x_1, x_2 \geq 0).$$

We now apply the transformation $(X_1, X_2) \to (R, W)$, where

$$R = X_1 + X_2, \quad W = \frac{\left(\frac{\sigma_1}{\xi_1} + X_1\right)}{\left(\frac{\sigma_1}{\xi_1} + X_1\right) + \left(\frac{\sigma_2}{\xi_2} + X_2\right)},$$

and for $1 + \xi_1 X_1/\sigma_1 = \frac{\xi_1}{\sigma_1}(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + R)W$ and $1 + \xi_2 X_2/\sigma_2 = \frac{\xi_2}{\sigma_2}(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + R)(1 - W)$. The joint density of (R, W) for $r > 0$ is

$$\begin{aligned} f_{R,W}(r, w) &= \sigma_1^{1/\xi_1} \sigma_2^{1/\xi_2} \xi_1^{-1/\xi_1 - 1} \xi_2^{-1/\xi_2 - 1} \left(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + r\right)^{-1/\xi_1 - 1/\xi_2 - 1} w^{-1/\xi_1 - 1} (1 - w)^{-1/\xi_2 - 1} \\ &\sim \sigma_1^{1/\xi_1} \sigma_2^{1/\xi_2} \xi_1^{-1/\xi_1 - 1} \xi_2^{-1/\xi_2 - 1} r^{-1/\xi_1 - 1/\xi_2 - 1} w^{-1/\xi_1 - 1} (1 - w)^{-1/\xi_2 - 1} \end{aligned}$$

as $r \to \infty$, and for $w \in [t_1(r), 1 - t_2(r)]$ where

$$t_1(r) = \frac{\sigma_1}{\xi_1}\left(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + r\right)^{-1}, \;\; t_2(r) = \frac{\sigma_2}{\xi_2}\left(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + r\right)^{-1}.$$

We can write the marginal density of R as

$$f_R(r) = \int_{t_1(r)}^{c_1} f_{R,W}(r,w)\mathrm{d}w + \int_{c_1}^{1-c_2} f_{R,W}(r,w)\mathrm{d}w + \int_{1-c_2}^{1-t_2(r)} f_{R,W}(r,w)\mathrm{d}w$$

where c_1 and c_2 are constants chosen such that $c_1 > t_1, c_2 > t_2(r)$ and $c_1 < 1 - c_2(r)$. We then have

$$\begin{aligned}
f_R(r) &\sim \sigma_1^{1/\xi_1}\sigma_2^{1/\xi_2}\xi_1^{-1/\xi_1-1}\xi_2^{-1/\xi_2-1}r^{-1/\xi_1-1/\xi_2-1}\left[\int_{t_1(r)}^{c_1} w^{-1/\xi_1-1}(1-w)^{-1/\xi_2-1}\mathrm{d}w\right. \\
&\quad \left. + \int_{c_1}^{1-c_2} w^{-1/\xi_1-1}(1-w)^{-1/\xi_2-1}\mathrm{d}w + \int_{1-c_2}^{1-t_2(r)} w^{-1/\xi_1-1}(1-w)^{-1/\xi_2-1}\mathrm{d}w\right] \\
&\sim \sigma_1^{1/\xi_1}\sigma_2^{1/\xi_2}\xi_1^{-1/\xi_1-1}\xi_2^{-1/\xi_2-1}r^{-1/\xi_1-1/\xi_2-1} \\
&\quad \times \left[\int_{t_1(r)}^{c_1} w^{-1/\xi_1-1}\mathrm{d}w + K_1 + \int_{1-c_2}^{1-t_2(r)} (1-w)^{-1/\xi_2-1}\mathrm{d}w\right],
\end{aligned}$$

as $r \to \infty$ and for constant $K_1 = \int_{c_1}^{1-c_2} w^{-1/\xi_1-1}(1-w)^{-1/\xi_2-1}\mathrm{d}w > 0$. The last line follows as $(1-w) \approx 1$ and $w \approx 1$ on their respective domains. This gives

$$\begin{aligned}
f_R(r) &\sim \sigma_1^{1/\xi_1}\sigma_2^{1/\xi_2}\xi_1^{-1/\xi_1-1}\xi_2^{-1/\xi_2-1}r^{-1/\xi_1-1/\xi_2-1}\left[\xi_1\{t_1(r)\}^{-1/\xi_1} + \xi_2\{t_2(r)\}^{-1/\xi_2} + K_2\right] \\
&\sim \sigma_1^{1/\xi_1}\sigma_2^{1/\xi_2}\xi_1^{-1/\xi_1-1}\xi_2^{-1/\xi_2-1}r^{-1/\xi_1-1/\xi_2-1}\left[\xi_1\left\{\frac{\xi_1}{\sigma_1}\left(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + r\right)\right\}^{1/\xi_1}\right. \\
&\quad \left. + \xi_2\left\{\frac{\xi_2}{\sigma_2}\left(\frac{\sigma_1}{\xi_1} + \frac{\sigma_2}{\xi_2} + r\right)\right\}^{1/\xi_2} + K_2\right] \\
&\sim \sigma_1^{1/\xi_1}\sigma_2^{1/\xi_2}\xi_1^{-1/\xi_1-1}\xi_2^{-1/\xi_2-1}r^{-1/\xi_1-1/\xi_2-1}\xi_2\left\{\frac{\xi_2}{\sigma_2}r\right\}^{1/\xi_2} \\
&\sim \sigma_1^{1/\xi_1}\xi_1^{-1/\xi_1-1}r^{-1/\xi_1-1} \sim f_{X_1}(r)
\end{aligned}$$

as $r \to \infty$, for constant $K_2 = K_1 - \xi_1 c_1^{-1/\xi_1} - \xi_2 c_2^{-1/\xi_2}$, and where the last line follows as $1/\xi_2 < 1/\xi_1$. Hence $\Pr\{R \geq r\} \sim \Pr\{X_1 \geq r\}$ as $r \to \infty$.

Appendix B

Supplementary material for Chapter 4

B.1 Proofs relating to Section 4.2.1

B.1.1 Proof of determinant for Jacobian of pseudo-radial and -angular transformation

Here we derive the determinant of the Jacobian for the transformation given in (4.2.1). The Jacobian is given by the matrix

$$A = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_{d-1} & \left(1 - \sum_{j=1}^{d-1} w_j\right) \\ r & 0 & 0 & \dots & 0 & -r \\ 0 & r & 0 & \dots & 0 & -r \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & r & 0 & -r \\ 0 & 0 & \dots & 0 & r & -r \end{bmatrix}.$$

Note that by conditioning along the top row of the matrix, we have

$$|A| = \sum_{j=1}^{d-1} (-1)^{j+1} w_j |A_j| + (-1)^{d-1} \left(1 - \sum_{j=1}^{d-1} w_j\right) |A_d|$$

where

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -r \\ r & 0 & 0 & \dots & 0 & -r \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & -r \\ 0 & \dots & \dots & 0 & r & -r \end{bmatrix}, \quad A_2 = \begin{bmatrix} r & 0 & 0 & \dots & 0 & -r \\ 0 & 0 & 0 & \dots & 0 & -r \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & -r \\ 0 & \dots & \dots & 0 & r & -r \end{bmatrix}, \dots$$

$$\dots, A_{d-1} = \begin{bmatrix} r & 0 & 0 & \dots & 0 & -r \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & -r \\ 0 & \dots & \dots & 0 & r & -r \\ 0 & 0 & 0 & \dots & 0 & -r \end{bmatrix}, \quad A_d = \begin{bmatrix} r & 0 & 0 & \dots & 0 & 0 \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & 0 \\ 0 & \dots & \dots & 0 & r & 0 \\ 0 & 0 & 0 & \dots & 0 & r \end{bmatrix}.$$

Note that for $i = 1, \dots, d-2$, we have A_i and A_{i+1} are identical matrices with two rows swapped. This implies that $|A_j| = -|A_{j+1}|$ for $j = 1, \dots, d-2$ and thus

$$|A_j| = (-1)^{j-1} |A_1|$$

for $j = 1, \ldots, d-1$, where

$$|A_1| = \begin{vmatrix} 0 & 0 & 0 & \ldots & 0 & -r \\ r & 0 & 0 & \ldots & 0 & -r \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & -r \\ 0 & \ldots & \ldots & 0 & r & -r \end{vmatrix} = (-1)^d(-r) \begin{vmatrix} r & 0 & 0 & \ldots & 0 & 0 \\ 0 & r & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & r & 0 & 0 \\ 0 & \ldots & \ldots & 0 & r & 0 \\ 0 & 0 & 0 & \ldots & 0 & r \end{vmatrix}$$
$$= (-1)^{d-1} r r^{d-2} = (-1)^{d-1} r^{d-1}.$$

Trivially $|A_d| = r^{d-1}$, and so it follows that

$$\begin{aligned} |A| =& \sum_{j=1}^{d-1} (-1)^{j+1} w_j |A_j| + (-1)^{d-1} \left(1 - \sum_{j=1}^{d-1} w_j\right) |A_d| \\ =& \sum_{j=1}^{d-1} (-1)^{j+1} w_j (-1)^{j-1} (-1)^{d-1} r^{d-1} + (-1)^{d-1} \left(1 - \sum_{j=1}^{d-1} w_j\right) r^{d-1} \\ =& (-1)^{d-1} r^{d-1} \sum_{j=1}^{d-1} w_j + (-1)^{d-1} r^{d-1} \left(1 - \sum_{j=1}^{d-1} w_j\right) = (-1)^{d-1} r^{d-1}. \end{aligned}$$

As the determinant of the Jacobian requires the absolute value of $|A|$, this provides the required result.

B.1.2 Proof of (4.2.5)

Here we prove that if $R_d = \sum_{i=1}^d X_i$ with upper end-point $r^F = -\sum_{i=1}^d \sigma_i/\xi_i$ and

$$W_j = \frac{\frac{\sigma_j}{\xi_j} + X_j}{R - r^F},$$

for $j = 1, \ldots, d-1$, then the support of $\mathbf{W}$ is $[0,1]^{d-1}$, independent of the value of R_d for $R_d > t$, where $t = r^F + \max_{1 \le j \le d}\{\sigma_j/\xi_j\} = r^F - \min_{1 \le j \le d}\{-\sigma_j/\xi_j\}$. We begin

by letting $x^{min} = \min_{1\leq j\leq d}\{-\sigma_j/\xi_j\} > 0$ and $x^{max} = r^F - x^{min}$. As $R_d \leq r^F = x^{max} + x^{min}$, there exists a random variable $P \in [0,1]$, such that

$$R_d|(R_d > x^{max}) = x^{max} + x^{min}P.$$

Now for $i = \underset{1\leq k\leq d}{\arg\min}\{-\sigma_k/\xi_k\}$, we let $X_k|(R_d > x^{max}) = -\frac{\sigma_k}{\xi_k}Q_k$, for $k = 1,\ldots,d, k \neq i$, for random $Q_k \leq 1$. Then as

$$0 \leq \left(R_d - \sum_{k\neq i} X_j\right)\Big|(R_d > x^{max}) \leq x^{min} \Rightarrow 0 \leq x^{max} + x^{min}P + \sum_{k\neq i}\frac{\sigma_k}{\xi_k}Q_k \leq x^{min},$$

and

$$\begin{aligned} 0 \leq (R_d - X_k)\,|(R_d > x^{max}) &\leq x^{min} + x^{max} + \frac{\sigma_k}{\xi_k} \\ \Rightarrow 0 \leq x^{max} + x^{min}P + \frac{\sigma_k}{\xi_k}Q_k &\leq x^{min} + x^{max} + \frac{\sigma_k}{\xi_k}, \end{aligned}$$

it follows that Q_k for all $k = 1,\ldots,d, k \neq i$ must satisfy

$$-\frac{P}{1-P} < 0 \leq \frac{x^{max} - \sum_{k\neq i}\frac{\sigma_k}{\xi_k}Q_k}{x^{min}(1-P)} \leq 1,$$

and

$$-\frac{P}{1-P} - \frac{x^{max}}{x^{min}(1-P)} < 0 \leq \frac{-\frac{\sigma_k}{\xi_k}(1-Q_k)}{x^{min}(1-P)} \leq 1.$$

Hence, we have that

$$\begin{aligned} W_i = \frac{\frac{\sigma_i}{\xi_i} + X_i}{R_d - r^F} &= \frac{-x^{min} + R_d - \sum_{k\neq i} X_k}{-x^{max} - x^{min} + x^{max} + x^{min}P} \\ &= \frac{-x^{min} + x^{max} + x^{min}P - \sum_{k\neq i}\frac{\sigma_k}{\xi_k}Q_k}{-x^{min}(1-P)} = 1 - \frac{x^{max} - \sum_{k\neq i}\frac{\sigma_k}{\xi_k}Q_k}{x^{min}(1-P)} \end{aligned}$$

and so $W_i \in [0,1]$ as needed. Conversely, for $j \neq i$, we have that

$$W_j = \frac{\frac{\sigma_j}{\xi_j} + X_j}{R_d - r^F} = \frac{\frac{\sigma_j}{\xi_j} - \frac{\sigma_j}{\xi_j}Q_j}{-x^{min}(1-P)} = \frac{-\frac{\sigma_j}{\xi_j}(1-Q_j)}{x^{min}(1-P)}$$

which gives $W_j \in [0,1]$ for all $j = 1, \ldots, d-1, j \neq i$. Hence, the support of $\mathbf{W}$ does not depend on R when $R > t$.

B.2 Proof of transformation of (4.4.11) to (4.4.13)

Here we provide a proof of the transformation of the integral in (4.4.11), namely

$$\int_{\mathcal{S}_{d-1}} \frac{\xi^{-(d-1)}}{\left[\sum_{i=1}^{d}(w_i/\sigma_i)^{1/\xi}\right]^{d+1}} \left\{\prod_{k=1}^{d} \frac{w_k^{1/\xi-1}}{\sigma_k^{1/\xi}}\right\} \times h\left(\frac{w_1^{1/\xi}}{\sigma_1^{1/\xi}\sum_{i=1}^{d}(w_i/\sigma_i)^{1/\xi}}, \ldots, \frac{w_{d-1}^{1/\xi}}{\sigma_{d-1}^{1/\xi}\sum_{i=1}^{d}(w_i/\sigma_i)^{1/\xi}}\right) \mathrm{d}\mathbf{w}, \tag{B.2.1}$$

for $\mathbf{w} \in \mathcal{S}_{d-1}$, the $(d-1)$-dimensional unit simplex, to the integral in (4.4.13) given by

$$\int_{\Omega} \frac{1}{\sum_{i=1}^{d} f_i(\boldsymbol{\omega})} \times h\left(\omega_1, \ldots, \omega_{d-1}, \omega_d\right) \mathrm{d}\boldsymbol{\omega}, \tag{B.2.2}$$

where

$$\omega_d = 1 - \sum_{i=1}^{d-1}\left\{\sum_{j=1}^{d-1}\frac{\sigma_i}{\sigma_j}\left(\frac{\omega_i}{\omega_j}\right)^{\xi} + \frac{\sigma_d}{\sigma_i}\left(\frac{1-\sum_{k=1}^{d-1}\omega_k}{\omega_i}\right)^{\xi}\right\}^{-1},$$

and

$$f_i(\boldsymbol{\omega}) = \begin{cases} \left\{\sum_{j=1}^{d-1}\frac{\sigma_i}{\sigma_j}\left(\frac{\omega_i}{\omega_j}\right)^{\xi} + \frac{\sigma_d}{\sigma_i}\left(\frac{1-\sum_{k=1}^{d-1}\omega_k}{\omega_i}\right)^{\xi}\right\}^{-1/\xi} \sigma_i^{-1/\xi}, & \text{if } i \neq d, \\ (w_d/\sigma_d)^{1/\xi}, & \text{if } i = d, \end{cases}$$

and

$$\Omega = \Bigg\{ \boldsymbol{\omega} : 0 \leq \omega_i \leq 1, i = 1, \ldots, d,$$
$$\omega_d + \sum_{i=1}^{d-1} \left\{ \sum_{j=1}^{d-1} \frac{\sigma_i}{\sigma_j} \left(\frac{\omega_i}{\omega_j} \right)^{\xi} + \frac{\sigma_d}{\sigma_i} \left(\frac{1 - \sum_{k=1}^{d-1} \omega_k}{\omega_i} \right)^{\xi} \right\}^{-1} = 1 \Bigg\}.$$

Recall that for all $j = 1, \ldots, d-1$ we apply the transformation

$$\omega_j = \frac{w_1^{1/\xi}}{\sigma_1^{1/\xi} \sum_{i=1}^{d} (w_i/\sigma_i)^{1/\xi}} = \frac{w_1^{1/\xi}}{\sigma_1^{1/\xi}} S^{-1},$$

where $S = \sum_{i=1}^{d} (w_i/\sigma_i)^{1/\xi}$. This transformation has partial derivatives

$$\begin{aligned} \frac{\partial \omega_j}{\partial w_j} &= \xi^{-1} \sigma_j^{-1} S^{-1} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi - 1} - S^{-2} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi} \left[\xi^{-1} \sigma_j^{-1} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi - 1} - \xi^{-1} \sigma_d^{-1} \left(\frac{w_d}{\sigma_d} \right)^{1/\xi - 1} \right] \\ &= \xi^{-1} \sigma_j^{-1} S^{-2} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi - 1} \left[S - \left(\frac{w_j}{\sigma_j} \right)^{1/\xi} + \sigma_d^{-1} w_j \left(\frac{w_d}{\sigma_d} \right)^{1/\xi - 1} \right], \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \omega_i}{\partial w_j} &= S^{-2} \left(\frac{w_i}{\sigma_i} \right)^{1/\xi} \left[\xi^{-1} \sigma_j^{-1} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi - 1} - \xi^{-1} \sigma_d^{-1} \left(\frac{w_d}{\sigma_d} \right)^{1/\xi - 1} \right] \\ &= \xi^{-1} S^{-2} \left(\frac{w_i}{\sigma_i} \right)^{1/\xi} \left[\sigma_j^{-1} \left(\frac{w_j}{\sigma_j} \right)^{1/\xi - 1} - \sigma_d^{-1} \left(\frac{w_d}{\sigma_d} \right)^{1/\xi - 1} \right], \end{aligned}$$

for $i \neq j$. The reciprocal of the determinant of the Jacobian of the transformation is

$$|J| = \xi^{(d-1)} \left[\sum_{i=1}^{d} (w_i/\sigma_i)^{1/\xi} \right]^d \left\{ \prod_{k=1}^{d} \frac{\sigma_k^{1/\xi}}{w_k^{1/\xi - 1}} \right\} = \xi^{(d-1)} S^d \left\{ \prod_{k=1}^{d} \frac{\sigma_k^{1/\xi - 1}}{w_k^{1/\xi - 1}} \sigma_k \right\}. \quad \text{(B.2.3)}$$

We prove this analytically for $d = 2$ and $d = 3$, and note that certain software, e.g., Maple, can be used to show that (B.2.3) holds for $d > 3$.

For $d = 2$, we require only that

$$\begin{aligned}
\frac{\partial \omega_1}{\partial w_1} &= \xi^{-1}\sigma_1^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left[S - \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} + \sigma_2^{-1}w_1\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\right] \\
&= \xi^{-1}\sigma_1^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left[\left(\frac{w_2}{\sigma_2}\right)^{1/\xi} + \sigma_2^{-1}w_1\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\right] \\
&= \xi^{-1}\sigma_1^{-1}\sigma_2^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}[w_2 + w_1] \\
&= \xi^{-1}\sigma_1^{-1}\sigma_2^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1},
\end{aligned}$$

as its reciprocal gives (B.2.3). For the case where $d = 3$, we have

$$\begin{aligned}
|J|^{-1} &= \frac{\partial \omega_1}{\partial w_1} \times \frac{\partial \omega_2}{\partial w_2} - \frac{\partial \omega_1}{\partial w_2} \times \frac{\partial \omega_2}{\partial w_1} \\
&= \xi^{-1}\sigma_1^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left[S - \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} + \sigma_3^{-1}w_1\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&\times \xi^{-1}\sigma_2^{-1}S^{-2}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\left[S - \left(\frac{w_2}{\sigma_2}\right)^{1/\xi} + \sigma_3^{-1}w_2\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&- \xi^{-1}S^{-2}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi}\left[\sigma_1^{-1}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1} - \sigma_3^{-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&\times \xi^{-1}S^{-2}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi}\left[\sigma_2^{-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1} - \sigma_3^{-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}S^{-4}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\left\{\left[S - \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} + \frac{w_1}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right]\right. \\
&\times \left[S - \left(\frac{w_2}{\sigma_2}\right)^{1/\xi} + \frac{w_2}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] - w_1w_2\left[\sigma_1^{-1}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1} - \sigma_3^{-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&\times \left.\left[\sigma_2^{-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1} - \sigma_3^{-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right]\right\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}S^{-4}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\left\{S^2 - \frac{w_2}{\sigma_3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right. \\
&- \left.\frac{w_1}{\sigma_3}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} + \left(\frac{w_2w_2}{\sigma_1\sigma_2}\right)^{1/\xi}\right.
\end{aligned}$$

$$
\begin{aligned}
&- \left(\frac{w_2 w_2}{\sigma_1\sigma_2}\right)^{1/\xi} S\left[\left(\frac{w_2}{\sigma_2}\right)^{1/\xi} + \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} - \frac{w_2}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} - \frac{w_1}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right] \\
&+ \frac{w_1 w_2}{\sigma_3^2}\left(\frac{w_3}{\sigma_3}\right)^{2/\xi-2} - \left(\frac{w_1 w_2}{\sigma_1\sigma_2}\right)^{1/\xi} + \frac{w_2}{\sigma_3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} \\
&+ \frac{w_1}{\sigma_3}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} - \frac{w_1 w_2}{\sigma_3^2}\left(\frac{w_3}{\sigma_3}\right)^{2/\xi-2}\Bigg\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}S^{-4}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1} \\
&\times \left\{S^2 - S\left[\left(\frac{w_2}{\sigma_2}\right)^{1/\xi} + \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} - \frac{w_2}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} - \frac{w_1}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right]\right\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}S^{-3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1} \\
&\times \left\{S - \left(\frac{w_2}{\sigma_2}\right)^{1/\xi} - \left(\frac{w_1}{\sigma_1}\right)^{1/\xi} + \frac{w_2}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} + \frac{w_1}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}S^{-3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1} \\
&\times \left\{\left(\frac{w_3}{\sigma_3}\right)^{1/\xi} + \frac{w_2}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} + \frac{w_1}{\sigma_3}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\right\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}\sigma_3^{-1}S^{-3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1}\Big\{w_3 + w_2 + w_1\Big\} \\
&= \xi^{-2}\sigma_1^{-1}\sigma_2^{-1}\sigma_3^{-1}S^{-3}\left(\frac{w_1}{\sigma_1}\right)^{1/\xi-1}\left(\frac{w_2}{\sigma_2}\right)^{1/\xi-1}\left(\frac{w_3}{\sigma_3}\right)^{1/\xi-1} \\
&= \xi^{-2}S^{-3}\prod_{k=1}^{3}\left(\frac{w_k}{\sigma_k}\right)^{1/\xi-1}\sigma_k^{-1},
\end{aligned}
$$

as needed. It can then be shown that for each marginal transformation, we have

$$
w_i = \left\{\sum_{j=1}^{d-1}\frac{\sigma_i}{\sigma_j}\left(\frac{\omega_i}{\omega_j}\right)^{\xi} + \frac{\sigma_d}{\sigma_i}\left(\frac{1-\sum_{k=1}^{d-1}\omega_k}{\omega_i}\right)^{\xi}\right\}^{-1}
$$

for all $i = 1, \ldots, d-1$. Recalling that $w_d = 1 - \sum_{j=1}^{d-1} w_j$ and combining this with the above result for the Jacobian yields the desired integral.

B.3 Proofs for sums of exponential random variables

This section details the proofs of the results detailed in Tables 4.3.4 and 4.3.5 of Section 4.3 which pertain to the upper-tail of $R_d = \sum_{i=1}^{d} X_i$ where $X_i \sim \text{GPD}(\sigma_i, 0) = \text{Exp}(\sigma_i^{-1})$ for $\sigma_i > 0$. The proofs are presented such that each subsection considers a different copula. Note that for the sake of notation, all proofs in this section are written under the assumption that $X_i \sim \text{Exp}(\sigma_i)$ for all $i = 1, \ldots, d$; the results herein can easily be linked to the results given in Section 4.3 by replacing each σ_i for $i = 1, \ldots, d$ with its reciprocal, i.e., $1/\sigma_i$.

B.3.1 Perfect dependence

We induce perfect dependence between d exponentially distributed random variables by letting $X_1 \sim \text{Exp}(\sigma_1)$ and $X_i = \sigma_1 X_1/\sigma_i$ for all $i = 1, \ldots, d$. Then for $R_d = \sum_{i=1}^{d} X_i$, we have

$$\begin{aligned}\Pr\{R_d \geq r\} = \Pr\left\{X_1 \sum_{i=1}^{d} \frac{\sigma_1}{\sigma_i} \geq r\right\} &= \Pr\left\{X_1 \geq r\left(\sum_{i=1}^{d} \frac{\sigma_1}{\sigma_i}\right)^{-1}\right\} \\ &= \exp\left\{-r\left(\sum_{i=1}^{d} \frac{1}{\sigma_i}\right)^{-1}\right\},\end{aligned}$$

as needed. Note that if $\sigma_i = \sigma$ for all $i = 1, \ldots, d$, then $\left(\sum_{i=1}^{d} \frac{1}{\sigma_i}\right)^{-1} = \sigma/d$.

B.3.2 Independence

The proof for the independence copula is split into two cases: in the first case, we assert that $X_i \sim \text{Exp}(\sigma_i)$ where $\sigma_j \neq \sigma_i$ for all $i = 1, \ldots, d$, $j = 1, \ldots, d$ and $i \neq j$. It becomes much more difficult to derive the first-order behaviour of the survival function

of R_d when we remove the constraint that $\sigma_j \neq \sigma_i$ for all $i, j = 1, \ldots, d, i \neq j$; this we do for the second case. In particular, we derive results for $d \leq 4$ and make the assumption that this behaviour can be extended to R_d for $d > 4$.

Case 1

Let $X_i \sim \text{Exp}(\sigma_i)$ for $\sigma_j \neq \sigma_i$ for all $i = 1, \ldots, d$, $j = 1, \ldots, d$ and $i \neq j$. We make the assumption that the density of $R_d = \sum_{i=1}^{d} X_i$ is of the form

$$f_{R_d}(r) = \left(\prod_{i=1}^{d} \sigma_i\right) \sum_{j=1}^{d} \frac{e^{-r\sigma_j}}{\prod_{k \neq j}(\sigma_k - \sigma_j)}, \tag{B.3.1}$$

for $r > 0$. We prove this by induction and begin with the base case, which is $d = 2$. We note that the joint density of (X_1, X_2) is

$$f_{X_1,X_2}(x_1, x_2) = \sigma_1\sigma_2 e^{-\sigma_1 x_1} e^{-\sigma_2 x_2},$$

for $x_1 > 0, x_2 > 0$ and consider the transformation $(X_1, X_2) \to (R_1, R_2)$, where $R_2 = X_1 + X_2$ and $R_1 = X_1$. Note that $R_2 \in [R_1, \infty)$ and $R_1 \in [0, \infty)$. The joint density of (R_1, R_2) is

$$f_{R_1,R_2}(r_1, r_2) = \sigma_1\sigma_2 e^{-\sigma_1 r_1} e^{-\sigma_2(r_2 - r_1)} = \sigma_1\sigma_2 e^{-(\sigma_1 - \sigma_2)r_1} e^{-\sigma_2 r_2},$$

for $0 < r_1 < r_2$ as the determinant of the Jacobian for the transformation is one. The marginal density of R_2 is

$$\begin{aligned} f_{R_2}(r_2) &= \int_0^{r_2} f_{R_1,R_2}(r_1, r_2)\mathrm{d}r_1 = \sigma_1\sigma_2 e^{-\sigma_2 r_2} \int_0^{r_2} e^{-(\sigma_1 - \sigma_2)r_1}\mathrm{d}r_1 \\ &= \frac{\sigma_1\sigma_2}{\sigma_2 - \sigma_1} e^{-\sigma_2 r_2}\left[e^{-(\sigma_1 - \sigma_2)r_2} - 1\right] = \frac{\sigma_1\sigma_2}{\sigma_2 - \sigma_1}\left[e^{-\sigma_1 r_2} - e^{-\sigma_2 r_2}\right], \end{aligned}$$

as needed. We now assume that (B.3.1) holds for R_{d-1} and consider R_d. Note that the joint density of (R_{d-1}, X_d) is

$$f_{R_{d-1},X_d}(r, x_d) = \left(\prod_{i=1}^{d-1}\sigma_i\right)\sum_{j=1}^{d-1}\frac{e^{-r\sigma_j}}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)} \times \sigma_d e^{-\sigma_d x_d},$$

for $r > 0, x_d > 0$ and consider $R_d = R_{d-1} + X_d$. The joint density of (R_d, R_{d-1}) is

$$\begin{aligned} f_{R_d,R_{d-1}}(r_d, r_{d-1}) &= \left(\prod_{i=1}^{d-1}\sigma_i\right)\sum_{j=1}^{d-1}\frac{e^{-r_{d-1}\sigma_j}}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)} \times \sigma_d e^{-\sigma_d(r_d-r_{d-1})} \\ &= \left(\prod_{i=1}^{d}\sigma_i\right)\sum_{j=1}^{d-1}\frac{e^{-r_{d-1}(\sigma_j-\sigma_d)}}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)} \times e^{-\sigma_d r_d}, \end{aligned}$$

for $0 < r_{d-1} < r_d$ and the marginal density of R_d for $r_d > 0$ is then

$$\begin{aligned} f_{R_d}(r_d) &= \left(\prod_{i=1}^{d}\sigma_i\right)e^{-\sigma_d r_d}\sum_{j=1}^{d-1}\frac{1}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)}\int_0^{r_d} e^{-r_{d-1}(\sigma_j-\sigma_d)}\mathrm{d}r_{d-1} \\ &= \left(\prod_{i=1}^{d}\sigma_i\right)e^{-\sigma_d r_d}\sum_{j=1}^{d-1}\frac{1}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)}\frac{1}{\sigma_d-\sigma_j}\left[e^{-r_d(\sigma_j-\sigma_d)}-1\right] \\ &= \left(\prod_{i=1}^{d}\sigma_i\right)\sum_{j=1}^{d-1}\frac{1}{\prod_{k\neq j}^{d-1}(\sigma_k-\sigma_j)}\left[\frac{e^{-\sigma_j r_d}}{\sigma_d-\sigma_j}+\frac{e^{-\sigma_d r_d}}{\sigma_j-\sigma_d}\right] \\ &= \left(\prod_{i=1}^{d}\sigma_i\right)\sum_{j=1}^{d}\frac{e^{-r_d\sigma_j}}{\prod_{k\neq j}^{d}(\sigma_k-\sigma_j)}, \end{aligned}$$

and hence we have proven that (B.3.1) holds for all finite d. The survival function of R_d is then

$$\begin{aligned} \Pr\{R_d \geq s\} &= \int_s^\infty \left(\prod_{i}^{d}\sigma_i\right)\sum_{j=1}^{d}\frac{e^{-r\sigma_j}}{\prod_{k\neq j}(\sigma_k-\sigma_j)}\mathrm{d}r \\ &= \left(\prod_{i}^{d}\sigma_i\right)\sum_{j=1}^{d}\frac{e^{-s\sigma_j}}{\prod_{k\neq j}\sigma_j(\sigma_k-\sigma_j)}, \end{aligned}$$

for $s > 0$. We note that, asymptotically, the term that dominates is the exponential term with the smallest power. Hence, we have

$$\Pr\{R_d \geq s\} \sim \frac{\prod_i^d \sigma_i}{\sigma_J \prod_{i \neq J}^d (\sigma_i - \sigma_J)} e^{-s\sigma_J} = \prod_{i \neq J}^{d} \frac{\sigma_i}{(\sigma_i - \sigma_J)} e^{-s\sigma_J},$$

as $s \to \infty$ and where $\sigma_J = \min_{i=1,\ldots,d} \{\sigma_i\}$. Note that this result only holds if $\sigma_i \neq \sigma_j$ for all $i = 1, \ldots, d$, $j = 1, \ldots, d$ and $i \neq j$.

Case 2

Deriving the first-order behaviour of $\Pr\{R_d \geq s\}$ becomes much more difficult when we consider cases where subsets of the marginal scale parameters are equal. For $X_i \sim \text{Exp}(\sigma_i)$ for $i = 1, \ldots, d$, the number of cases that would need to be considered is the number of partitions of the set $\{1, \ldots, d\}$, i.e., the d-th Bell number. Whilst we are unable to provide a parsimonious proof that can encompass all partitions for any finite d, we explore the case $d = 4$ in Section B.4; we then make the assumption that the results derived therein can be extended to $d > 4$. Under this assumption, we have

$$\Pr\{R_d \geq r\} \sim K r^{|L|-1} \exp\{-r\sigma_{min}\} \tag{B.3.2}$$

as $r \to \infty$, where $\sigma_{min} = \min_{i=1,\ldots,d} \{\sigma_i\}$ and $L = \{j : \sigma_j = \sigma_{min}, j = 1, \ldots, d\}$. Note that clearly this holds in the case where all marginal scales are different, i.e., case 1, and where they are all equal, i.e., $\sigma_i = \sigma$ for all $i = 1, \ldots, d$. The latter follows as the sum of d independent and identically distributed exponential random variables is a gamma random variable with shape and rate parameter, d and scale σ, respectively.

B.3.3 Extreme value copula

Let $X_i \sim \text{Exp}(\sigma_i)$ for all $i = 1, \ldots, d$ and with dependence in $\mathbf{X}$ described using the extreme value copula defined in Section 4.1. Following the assumptions made in

Section 4.2.2 that lead to (4.2.9), the density of $\mathbf{X}$ is of the form

$$f_{\mathbf{X}}(\mathbf{x}) \sim \left\{ \prod_{i=1}^{d} \sigma_i e^{\sigma_i x_i} \right\} \left\{ -V_{\mathbf{x}}(e^{\sigma_1 x_1}, \ldots, e^{\sigma_d x_d}) \right\}, \tag{B.3.3}$$

as each component of $\mathbf{x} = (x_1, \ldots, x_d)$ tends to its respective upper-endpoint and where $V_{\mathbf{x}}$ denotes the d-th partial derivative of V in (4.1.3) with respect to all components. Following Coles and Tawn (1991) and the details provided in Section 4.4.3, we re-write (B.3.3) in terms of the density $h(\cdot)$ for the distribution function H, defined in (4.1.3). Then

$$f_{\mathbf{X}}(\mathbf{x}) \sim \left\{ \prod_{i=1}^{d} \sigma_i \right\} \frac{e^{\sum_{i=1}^{d} \sigma_i x_i}}{\left(\sum_{i=1}^{d} e^{\sigma_i x_i} \right)^{(d+1)}} \times h\left(\frac{e^{\sigma_1 x_1}}{\sum_{i=1}^{d} e^{\sigma_i x_i}}, \ldots, \frac{e^{\sigma_{d-1} x_{d-1}}}{\sum_{i=1}^{d} e^{\sigma_i x_i}} \right),$$

as each component of $\mathbf{x} = (x_1, \ldots, x_d)$ tends to its respective upper-endpoint. We now apply the transformation $\mathbf{X} \to (R_d, \mathbf{W})$, where $R_d = \sum_{i=1}^{d} X_i$ and

$$\mathbf{W} = \{W_k = \sigma_1 X_1 - \sigma_{k+1} X_{k+1}, k = 1, \ldots, d-1\}, \tag{B.3.4}$$

with $W_k \in [-\sigma_{k+1} R_d, \sigma_1 R_d]$ for all $k = 1, \ldots, d-1$. Inverting this transformation and deriving $\mathbf{X}$ as a function of R_d and $\mathbf{W}$ is non-trivial. To illustrate this, we note that $A_d \mathbf{X} = (R_d, \mathbf{W})$ and so $\mathbf{X} = A_d^{-1}(R_d, \mathbf{W})$ where A_d is the $d \times d$ matrix

$$A_d = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ \sigma_1 & -\sigma_2 & 0 & \ldots & 0 \\ \sigma_1 & 0 & -\sigma_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & 0 & 0 & \ldots & -\sigma_d \end{bmatrix}.$$

Whilst the full inverse A_d^{-1} can be evaluated computationally, we require only the first column of A_d^{-1} for this proof, as this provides the contribution of R_d to $\mathbf{X}$ which is all we require for integrating $f_{R_d,\mathbf{W}}$ with respect to R_d and $\mathbf{W}$. To find the first column of A_d^{-1}, we first note that the determinant of A_d is

$$|A_d| = (-1)^{d+1} \sum_{i=1}^{d} \prod_{j \neq i} \sigma_j, \tag{B.3.5}$$

which we prove by induction. The case $d = 2$ is trivial and so is omitted. Now, assume that (B.3.5) holds for $d = k$ and consider A_{k+1}, where

$$A_{k+1} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ \sigma_1 & -\sigma_2 & 0 & \dots & 0 & 0 \\ \sigma_1 & 0 & -\sigma_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_1 & 0 & 0 & \dots & -\sigma_k & 0 \\ \sigma_1 & 0 & 0 & \dots & 0 & -\sigma_{k+1} \end{bmatrix}.$$

We then have

$$\begin{aligned} |A_{k+1}| &= (-1)^k \sigma_1 \begin{vmatrix} 1 & 1 & \dots & 1 & 1 \\ -\sigma_2 & 0 & \dots & 0 & 0 \\ 0 & -\sigma_3 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\sigma_k & 0 \end{vmatrix} - \sigma_{k+1} \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ \sigma_1 & -\sigma_2 & 0 & \dots & 0 \\ \sigma_1 & 0 & -\sigma_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & 0 & 0 & \dots & -\sigma_k \end{vmatrix} \\ &= (-1)^k \sigma_1 (-1)^{k-1} \begin{vmatrix} -\sigma_2 & 0 & \dots & 0 \\ 0 & -\sigma_3 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\sigma_k \end{vmatrix} - \sigma_{k+1} |A_k| \end{aligned}$$

$$
\begin{aligned}
&= (-1)^k \sigma_1 (-1)^{k-1} (-1)^{k-1} \prod_{j=2}^{k} \sigma_j - \sigma_{k+1} |A_k| \\
&= (-1)^k \prod_{j=1}^{k} \sigma_j - \sigma_{k+1} (-1)^{k+1} \sum_{i=1}^{k} \prod_{j \neq i} \sigma_j = (-1)^{k+2} \left(\prod_{j=1}^{k} \sigma_j + \sum_{i=1}^{k} \prod_{j \neq i} \sigma_j \right) \\
&= (-1)^{k+2} \sum_{i=1}^{k+1} \prod_{j \neq i} \sigma_j \, .
\end{aligned}
$$

as needed and so (B.3.5) holds for all finite $d \in \mathbb{N}$. We can now write the inverse of A_d as

$$
A_d^{-1} = \frac{1}{|A_d|} \tilde{A}_d = \frac{1}{|A_d|} ((-1)^{i+j} M_{ji})_{1 \leq i,j \leq d},
$$

where $\tilde{A}_d$ denotes the adjugate matrix and M_{ij} is the (i,j)-th minor of A_d. Hence, the $(1,1)$-th component of A_d^{-1} is

$$
\begin{aligned}
\frac{1}{|A_d|} (-1)^2 M_{11} =& \frac{1}{|A_d|} \begin{vmatrix} -\sigma_2 & 0 & \dots & 0 \\ 0 & -\sigma_3 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\sigma_d \end{vmatrix} = \frac{1}{|A_d|} (-1)^{d-1} \prod_{i=2} \sigma_i \\
=& \frac{(-1)^{d-1} \prod_{i=2} \sigma_i}{(-1)^{d+1} \sum_{i=1}^{d} \prod_{j \neq i} \sigma_j} = \frac{\prod_{i=2} \sigma_i}{\sum_{i=1}^{d} \prod_{j \neq i} \sigma_j},
\end{aligned}
$$

and so we can express X_1 in terms of R_d and a function g_1 of $\mathbf{W}$, i.e.,

$$
X_1 = \frac{\prod_{i=2} \sigma_i}{\sum_{i=1}^{d} \prod_{k \neq i} \sigma_k} R_d + g_1(\mathbf{w}).
$$

A similar strategy can be used to show that all components of $\mathbf{X}$ are of the form

$$
X_j = \frac{\sigma_1 X_1 - W_{j-1}}{\sigma_j} = \phi_j R_d + g_j(\mathbf{W}),
$$

for functions $g_j(\mathbf{W})$, which are functions of $\mathbf{W}$ only, and where

$$\phi_j = \frac{\prod_{i\neq j}\sigma_i}{\sum_{i=1}^{d}\prod_{k\neq i}\sigma_k},$$

for all $j = 1,\ldots,d$. Note that $\sigma_j\phi_j = \sigma_k\phi_k$ for all $j = 1,\ldots,d,\ k = 1,\ldots,d$. As the transformation $\mathbf{X} \to (R_d, \mathbf{W})$ is linear, the corresponding Jacobian has determinant $J = 1/|A_d|$. Then for $\mathbf{w} = (w_1,\ldots,w_{d-1})$, the joint density of R_d and $\mathbf{W}$ is

$$\begin{aligned}
f_{R_d,\mathbf{W}}(r,\mathbf{w}) &\sim |J|\left\{\prod_{i=1}^{d}\sigma_i\right\}\frac{e^{\sum_{j=1}^{d}\sigma_j(\phi_j r+g_j(\mathbf{w}))}}{\left(\sum_{j=1}^{d}e^{\sigma_j(\phi_j r+g_j(\mathbf{w}))}\right)^{(d+1)}}\times \\
&\quad h\left(\frac{e^{\sigma_1(\phi_1 r+g_1(\mathbf{w}))}}{\left(\sum_{j=1}^{d}e^{\sigma_j(\phi_j r+g_j(\mathbf{w}))}\right)},\ldots,\frac{e^{\sigma_{d-1}(\phi_{d-1} r+g_{d-1}(\mathbf{w}))}}{\left(\sum_{j=1}^{d}e^{\sigma_j(\phi_j r+g_j(\mathbf{w}))}\right)}\right) \\
&\sim |J|\left\{\prod_{i=1}^{d}\sigma_i\right\}\frac{e^{d\sigma_1\phi_1 r+\sum_{j=1}^{d}\sigma_j g_j(\mathbf{w})}}{e^{(d+1)\sigma_1\phi_1 r}\left(\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}\right)^{(d+1)}}\times \\
&\quad h\left(\frac{e^{\sigma_1\phi_1 r+\sigma_1 g_1(\mathbf{w})}}{e^{\sigma_1\phi_1 r}\left(\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}\right)},\ldots,\frac{e^{\sigma_1\phi_1 r+\sigma_{d-1} g_{d-1}(\mathbf{w})}}{e^{\sigma_1\phi_1 r}\left(\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}\right)}\right) \\
&\sim |J|\left\{\prod_{i=1}^{d}\sigma_i\right\}e^{-r\sigma_1\phi_1}G(\mathbf{w}),
\end{aligned}$$

as $r \to \infty$ and where $w_k \in [-\sigma_{k+1}r, \sigma_1 r]$ for all $k = 1,\ldots,d-1$, and

$$G(\mathbf{w}) = \frac{e^{\sum_{j=1}^{d}\sigma_j g_j(\mathbf{w}))}}{\left(\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}\right)^{(d+1)}} \times h\left(\frac{e^{\sigma_1 g_1(\mathbf{w})}}{\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}},\ldots,\frac{e^{\sigma_{d-1} g_{d-1}(\mathbf{w}))}}{\sum_{j=1}^{d}e^{\sigma_j g_j(\mathbf{w})}}\right).$$

Now to derive $\Pr\{R_d \geq s\}$, we make the assumption that

$$\begin{aligned}
f_{R_d}(r) &= \int_{-\sigma_d r}^{\sigma_1 r}\cdots\int_{-\sigma_2 r}^{\sigma_1 r} f_{R_d,\mathbf{W}}(r,\mathbf{w})\mathrm{d}w_1\ldots w_{d-1} \\
&\sim |J|\left\{\prod_{i=1}^{d}\sigma_i\right\}e^{-r\sigma_1\phi_1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}G(\mathbf{w})\mathrm{d}\mathbf{w}, \qquad \text{(B.3.6)}
\end{aligned}$$

as $s \to \infty$. The change of limits for the $(d-1)$-dimensional inner integral are required as g_i is unknown for $i = 1, \ldots, d$ and so without this assumption we are unable to integrate $f_{R_d,\mathbf{W}}(r, \mathbf{w})$ with respect to $\mathbf{w}$. Whilst we do not prove that this assumptions holds for any $d > 2$, the proof for $d = 2$ is given in Appendix A.5. To further justify this assumption, we note that the probability that any component of $\mathbf{W}$ is in its respective upper, or lower, tails is approximately zero; hence the integral of $f_{R_d,\mathbf{W}}(r, \mathbf{w})$ with $\mathbf{w}$ in these regions is likely to have negligible effect on the asymptotic behaviour of $f_{R_d}(r)$. To see this, recall that in Section 4.2.2 we made the assumption that as $R \to \infty$, that each component of $\mathbf{X}$ tends to infinity at a similar rate. Combining this with (B.3.4), we observe that there is a very low probability that any component of $\mathbf{W}$ attains its upper or lower endpoint. Under the assumption that (B.3.6) holds, it follows that

$$\Pr\{R_d \geq r\} \sim K \exp\left\{ -r \frac{\prod_{i=1}^{d} \sigma_i}{\sum_{i=1}^{d} \prod_{k \neq i}^{d} \sigma_k} \right\},$$

as $r \to \infty$ and where

$$K = |J| \sum_{i=1}^{d} \prod_{k \neq i}^{d} \sigma_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G(\mathbf{w}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G(\mathbf{w}) > 0.$$

It can be shown that the power in the exponent of $\Pr\{R_d \geq r\}$ as $r \to \infty$ can be re-written as

$$\frac{\prod_{i=1}^{d} \sigma_i}{\sum_{i=1}^{d} \prod_{j \neq i}^{d} \sigma_j} = \left(\sum_{i=1}^{d} \frac{1}{\sigma_i} \right)^{-1},$$

which yields the required result.

B.3.4 Inverted extreme value copula

Let $X_i \sim \text{Exp}(\sigma_i)$ for $\sigma_i > 0$ and all $i = 1, \ldots, d$. By combining (4.1.1) and (4.1.4) and differentiating with respect to all arguments, the joint density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \left\{ \prod_{i=1}^{d} \frac{x_i^{-2}}{\sigma_i} \right\} \exp\{-V((\sigma_1 x_1)^{-1}, \ldots, (\sigma_d x_d)^{-1})\} \times \sum_{\pi \in \mathcal{P}} \prod_{\pi_I \in \pi} -V_{\pi_I}((\sigma_1 x_1)^{-1}, \ldots, (\sigma_d x_d)^{-1}), \tag{B.3.7}$$

where V is defined in (4.1.3). We now apply the transformation given by (4.2.1), namely $\mathbf{X} \to (R_d, \mathbf{W})$. The joint density of $(R_d, \mathbf{W})$ for $\mathbf{w} = (w_1, \ldots, w_{d-1}) \in [0,1]^{d-1}$ is

$$\begin{aligned}
f_{R_d,\mathbf{W}}(r, \mathbf{w}) &= r^{-(d+1)} \left\{ \prod_{i=1}^{d} \frac{w_i^{-2}}{\sigma_i} \right\} \exp\{-V((\sigma_1 r w_1)^{-1}, \ldots, (\sigma_d r w_d)^{-1})\} \\
&\quad \times \sum_{\pi \in \mathcal{P}} \prod_{\pi_I \in \pi} \{-V_{\pi_I}((\sigma_1 r w_1)^{-1}, \ldots, (\sigma_d r w_d)^{-1})\} \\
&= r^{-(d+1)} \left\{ \prod_{i=1}^{d} \frac{w_i^{-2}}{\sigma_i} \right\} \exp\{-rV((\sigma_1 w_1)^{-1}, \ldots, (\sigma_d w_d)^{-1})\} \\
&\quad \times \sum_{\pi \in \mathcal{P}} r^{k_\pi} \prod_{\pi_I \in \pi} \{-V_{\pi_I}((\sigma_1 w_1)^{-1}, \ldots, (\sigma_d w_d)^{-1})\} \\
&\sim r^{-(d+1)} \left\{ \prod_{i=1}^{d} \frac{w_i^{-2}}{\sigma_i} \right\} \exp\{-rV((\sigma_1 w_1)^{-1}, \ldots, (\sigma_d w_d)^{-1})\} \\
&\quad \times r^{2d} \prod_{i=1}^{d} \{-V_{x_i}((\sigma_1 w_1)^{-1}, \ldots, (\sigma_d w_d)^{-1})\},
\end{aligned}$$

as $r \to \infty$ and where $w_d = 1 - \sum_{j=1}^{d-1} w_j$; the last line follows as $\max_{\pi \in \mathcal{P}}\{k_\pi\} = 2d$. To illustrate this we use a similar argument to that provided in Section 4.2.2, except here we require the partition $\pi \in \mathcal{P}$ that minimises the order of homogeneity for the expression $\prod_{\pi_I \in \pi}\{-V_{\pi_I}(z_1, \ldots, z_d)\}$. It can be shown that $\arg\max_{\pi \in \mathcal{P}}\{k_\pi\} = \{\{1\}, \ldots, \{d\}\}$ and hence $\max_{\pi \in \mathcal{P}}\{k_\pi\} = 2d$ as $\prod_{i=1}^{d}\{-V_{x_i}(z_1, \ldots, z_d)\}$ is the product of d functions with

order of homogeneity -2. We then have

$$f_{R_d,\mathbf{W}}(r,\mathbf{w}) \sim r^{d-1}\exp\{-rh(\mathbf{w})\}q(\mathbf{w}),$$

where $h(\mathbf{w}) = V((\sigma_1 w_1)^{-1},\ldots,(\sigma_d w_d)^{-1})$, and

$$q(\mathbf{w}) = -\prod_{i=1}^{d}\frac{w_i^{-2}}{\sigma_i}V_{x_i}((\sigma_1 w_1)^{-1},\ldots,(\sigma_d w_d)^{-1});$$

recall that $w_d = 1-\sum_{j=1}^{d-1} w_j$. To derive the marginal density of R_d, we utilise Laplace's method, see Section 4.2.3. We denote H as the Hessian of h with respect to $\mathbf{w}$ and $\mathbf{w}^* = \underset{\mathbf{w}\in[0,1]^{d-1}}{\arg\min}\, h(\mathbf{w})$, which we assume to be unique; then

$$f_{R_d}(r) = \int_{\mathbf{w}\in[0,1]^{d-1}} f_{R,\mathbf{W}}(r,\mathbf{w})\mathrm{d}\mathbf{w} \sim Kr^{(d-1)/2}\exp\{-rh(\mathbf{w}^*)\},$$

as $r\to\infty$ and where $K = (2\pi)^{(d-1)/2}q(\mathbf{w}^*)|-H(\mathbf{w}^*)|^{-1/2}$. It follows that the survival function of R_d is

$$\Pr\{R_d \geq s\} \sim \int_s^\infty Kr^{(d-1)/2}e^{-rh(\mathbf{w}^*)}\mathrm{d}r \sim K_2 s^{(d-1)/2}e^{-rh(\mathbf{w}^*)},$$

as $s\to\infty$, where $K_2 = K/h(\mathbf{w}^*) > 0$. Note that if $\sigma_i = \sigma$ for all $i = 1,\ldots,d$ and $V(\mathbf{w})$ is symmetric, then

$$h(\mathbf{w}) = \sigma V\left(\frac{1}{w_1},\ldots,\frac{1}{\left(1-\sum_{j=1}^{d} w_j\right)}\right),$$

and $\mathbf{w}^* = (1/d,\ldots,1/d)$. Hence,

$$h(\mathbf{w}^*) = \sigma V\left(d,\ldots,d\right) = \sigma V(1,\ldots,1)/d.$$

B.3.5 Standard Gaussian copula

We begin by considering the $d-$dimensional random vector $\mathbf{Y} \sim N_d(\mathbf{0}_d, \Sigma)$, where $\mathbf{0}_d$ denotes a d-vector of zeroes and Σ is a $d \times d$ positive definite matrix, where $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \Sigma_{ji} = \rho_{ij} \in (0, 1)$ for all $i, j = 1, \ldots, d, i \neq j$. To perform the transformation $\mathbf{Y} \to \mathbf{X}$ where $X_i \sim \text{Exp}(\sigma_i)$ for all $i = 1, \ldots, d$, we let $\bar{\Phi}(Y_i) = e^{-\sigma_i X_i}$ for all $i = 1, \ldots, d$. Here $\bar{\Phi}(\cdot)$ denotes the survival function of the univariate standard Gaussian distribution. By Mill's ratio (Grimmett, 2020), we have that $\bar{\Phi}(x) \sim \frac{\phi(x)}{x}$ as $x \to \infty$ and where ϕ denotes the density of the univariate standard Gaussian distribution. Thus, we can approximate the transformation $Y_i \to X_i$ by finding a solution to

$$\frac{1}{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} \sim e^{-\sigma x}, \tag{B.3.8}$$

which holds as $x \to \infty$.

To solve (B.3.8), we begin with an initial solution y_0 that solves $e^{-\sigma x} = e^{-\frac{1}{2}y_0^2}$, i.e., $y_0 = \sqrt{2\sigma x}$. Now, let $y_1 = \sqrt{2\sigma x} + \epsilon$ where $\epsilon = o(\sqrt{x})$. Substituting this into (B.3.8) gives

$$e^{-\sigma x} \sim \frac{1}{\sqrt{2\sigma x} + \epsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(2\epsilon\sqrt{2\sigma x}+\epsilon^2)}e^{-\sigma x},$$

as $x \to \infty$ and thus

$$1 \sim \frac{1}{\sqrt{2\sigma x}}\frac{1}{1+\frac{\epsilon}{\sqrt{2\sigma x}}}\frac{1}{\sqrt{2\pi}}e^{-\epsilon\sqrt{2\sigma x}}e^{-\frac{1}{2}\epsilon^2} \sim \frac{1}{\sqrt{2\sigma x}}\frac{1}{\sqrt{2\pi}}e^{-\epsilon\sqrt{2\sigma x}}e^{-\frac{1}{2}\epsilon^2},$$

which follows as

$$\frac{1}{1+\frac{\epsilon}{\sqrt{2\sigma x}}} \to 1 \quad \text{as} \quad x \to \infty.$$

Taking logs of both sides and rearranging gives

$$\epsilon \sim -\frac{\log(4\pi\sigma x)}{2\sqrt{2\sigma x}} - \frac{\epsilon^2}{2\sqrt{2\sigma x}} \sim -\frac{\log(4\pi\sigma x)}{2\sqrt{2\sigma x}},$$

as $x \to \infty$. Thus, an approximate solution to (B.3.8) is

$$y = \sqrt{2\sigma x} - \frac{\log(4\pi\sigma x)}{2\sqrt{2\sigma x}}[1 + o(1)],$$

which can be applied to all components of $\mathbf{Y}$. To calculate the determinant of the Jacobian of the marginal transformation, we note that

$$\frac{\partial y_i}{\partial x_i} \sim \frac{\sqrt{2\sigma_i}}{2} x_i^{-1/2} - \frac{2 - \log(4\pi\sigma_i x_i)}{4\sqrt{2\sigma_i} x_i^{3/2}} \sim \sqrt{\frac{\sigma_i}{2x_i}},$$

as $x_i \to \infty$ and for all $i = 1, \ldots, d$. We now make the assumption that as $R \to r^F$ that all components of $\mathbf{X}$ tends to their respective upper-endpoints with associated rates determined by the following; we assume that $\bar{F}_1(x_1)/\bar{F}_j(x_j) \sim c_j$ as $\bar{F}_1(x_1) \to 0$ for constants $c_j > 0$ and for $j = 1, \ldots, d$, and that $R \to r^F \Rightarrow \bar{F}_1(x_1) \to 0$. The joint density of $\mathbf{X}$ is then

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} A_{\mathbf{x}}^T \Sigma^{-1} A_{\mathbf{x}} \right\} \times 2^{-d/2} \prod_{i=1}^{d} \sqrt{\frac{\sigma_i}{x_i}}, \tag{B.3.9}$$

where

$$A_{\mathbf{x}} = \begin{pmatrix} \sqrt{2\sigma_1 x_1} - \frac{\log(4\pi\sigma_1 x_1)}{2\sqrt{2\sigma_1 x_1}} \\ \vdots \\ \sqrt{2\sigma_d x_d} - \frac{\log(4\pi\sigma_d x_d)}{2\sqrt{2\sigma_d x_d}} \end{pmatrix}.$$

We now apply the transformation given by (4.2.1), i.e., $\mathbf{X} \to (R_d, \mathbf{W})$. The joint density of R_d and $\mathbf{W}$ is, for $\mathbf{w} = (w_1, \ldots, w_{d-1}) \in [0,1]^{d-1}$

$$f_{R_d, \mathbf{W}}(r, \mathbf{w}) \sim (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} A_{r,\mathbf{w}}^T \Sigma^{-1} A_{r,\mathbf{w}} \right\} \times 2^{-d/2} r^{d/2-1} \prod_{i=1}^{d} \sqrt{\frac{\sigma_i}{w_i}},$$

as $r \to \infty$, and where $w_d = 1 - \sum_{j=1}^{d-1} w_j$ and

$$A_{r,\mathbf{w}} = \begin{pmatrix} \sqrt{2\sigma_1 r w_1} - \frac{\log(4\pi\sigma_1 r w_1)}{2\sqrt{2\sigma_1 r w_1}} \\ \vdots \\ \sqrt{2\sigma_d r w_d} - \frac{\log(4\pi\sigma_d r w_d)}{2\sqrt{2\sigma_d r w_d}} \end{pmatrix} = \sqrt{2r} B_{\mathbf{w}} - \frac{1}{2\sqrt{2r}} \log(4\pi r) C_{\mathbf{w},1} - \frac{1}{2\sqrt{2r}} C_{\mathbf{w},2},$$

for

$$B_{\mathbf{w}} = \begin{pmatrix} \sqrt{\sigma_1 w_1} \\ \vdots \\ \sqrt{\sigma_d w_d} \end{pmatrix}, \; C_{\mathbf{w},1} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_1 w_1}} \\ \vdots \\ \frac{1}{\sqrt{\sigma_d w_d}} \end{pmatrix} \text{ and } C_{\mathbf{w},2} = \begin{pmatrix} \frac{\log(\sigma_1 w_1)}{\sqrt{\sigma_1 w_1}} \\ \vdots \\ \frac{\log(\sigma_d w_d)}{\sqrt{\sigma_d w_d}} \end{pmatrix}.$$

Now, consider the exponent in (B.3.9) and let $C_{\mathbf{w}} = \log(4\pi r) C_{\mathbf{w},1} + C_{\mathbf{w},2}$. The exponent can be written as

$$\begin{aligned} \frac{1}{2} A_{r,\mathbf{w}}^T \Sigma^{-1} A_{r,\mathbf{w}} &= \frac{1}{2} \left(\sqrt{2r} B_{\mathbf{w}} - \frac{1}{2\sqrt{2r}} C_{\mathbf{w}} \right)^T \Sigma^{-1} \left(\sqrt{2r} B_{\mathbf{w}} - \frac{1}{2\sqrt{2r}} C_{\mathbf{w}} \right) \\ &= r B_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} - \frac{1}{4} \left[C_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w}} \right] + \frac{1}{16r} C_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w}} \\ &= r B_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} - \frac{1}{4} \left[C_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w}} \right] + o(1), \end{aligned}$$

as $r \to \infty$. We then have

$$\begin{aligned} \frac{1}{2} A_{r,\mathbf{w}}^T \Sigma^{-1} A_{r,\mathbf{w}} = r B_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} - \frac{1}{4} \big[&(\log(4\pi r) C_{\mathbf{w},1} + C_{\mathbf{w},2})^T \Sigma^{-1} B_{\mathbf{w}} + \\ &B_{\mathbf{w}}^T \Sigma^{-1} (\log(4\pi r) C_{\mathbf{w},1} + C_{\mathbf{w},2}) \big] \\ = r B_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}} &- \log(4\pi r) \frac{1}{4} \left[C_{\mathbf{w},1}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w},1} \right] \\ &- \frac{1}{4} \left[C_{\mathbf{w},2}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w},2} \right] + o(1), \end{aligned}$$

as $r \to \infty$ and so we can rewrite (B.3.9) as

$$f_{R_d,\mathbf{W}}(r, \mathbf{w}) \sim (2\pi)^{-d/2} |\Sigma|^{-1/2} 2^{-d/2} r^{d/2-1} \left(\prod_{i=1}^{d} \sqrt{\frac{\sigma_i}{w_i}} \right) (4\pi r)^{-g_2(\mathbf{w})}$$

$$\times \exp\{-rg_1(\mathbf{w})\} \exp\{-g_3(\mathbf{w})\},$$

as $r \to \infty$, and where $g_1(\mathbf{w}) = B_{\mathbf{w}}^T \Sigma^{-1} B_{\mathbf{w}}$ and

$$g_2(\mathbf{w}) = \frac{1}{4}\left[C_{\mathbf{w},1}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w},1}\right], \text{ and } g_3(\mathbf{w}) = \frac{1}{4}\left[C_{\mathbf{w},2}^T \Sigma^{-1} B_{\mathbf{w}} + B_{\mathbf{w}}^T \Sigma^{-1} C_{\mathbf{w},2}\right].$$

To derive $\Pr\{R_d \geq s\}$, we require the $(d-1)$-dimensional integral of $f_{R_d,\mathbf{W}}$ with respect to $\mathbf{w} \in [0,1]^{d-1}$; this can be approximated for large r by Laplace's method, which gives

$$\begin{aligned}
f_{R_d}(r) &\sim \int_0^1 \cdots \int_0^1 (2\pi)^{-d/2} |\Sigma|^{-1/2} 2^{-d/2} r^{d/2-1} \left(\prod_{i=1}^d \sqrt{\frac{\sigma_i}{w_i}}\right) (4\pi r)^{-g_2(\mathbf{w})} \\
&\quad \times \exp\{-rg_1(\mathbf{w})\} \exp\{-g_3(\mathbf{w})\} \, \mathrm{d}\mathbf{w} \\
&\sim (2\pi)^{-d/2} |\Sigma|^{-1/2} 2^{-d/2} r^{d/2-1} \left(\prod_{i=1}^d \sqrt{\frac{\sigma_i}{w_i^*}}\right) (4\pi r)^{-g_2(\mathbf{w}^*)} \\
&\quad \times \exp\{-rg_1(\mathbf{w}^*)\} \exp\{-g_3(\mathbf{w}^*)\} \times |G_1(\mathbf{w}^*)|^{-1/2} (2\pi)^{(d-1)/2} r^{-(d-1)/2} \\
&\sim h_2(\mathbf{w}^*) r^{g_2(\mathbf{w}^*)-1/2} \exp\{-rg_1(\mathbf{w}^*)\}, \qquad \text{(B.3.10)}
\end{aligned}$$

as $r \to \infty$ and where

$$h_2(\mathbf{w}) = 2^{2g_2(\mathbf{w})-(d+1)/2} \pi^{g_2(\mathbf{w})-1/2} e^{-g_3(\mathbf{w})} |\Sigma|^{-1/2} |G_1(\mathbf{w}^*)|^{-1/2} \left(\prod_{i=1}^d \sigma_i\right) \prod_{i=1}^d w_i^{-1/2}$$

for $\mathbf{w}^* = \underset{\mathbf{w} \in [0,1]^{d-1}}{\arg\min}\, g_1(\mathbf{w})$ and $G_1(\mathbf{w}^*)$ is the Hessian of $g_1(\cdot)$ evaluated at $\mathbf{w}^*$. It can then be shown that the survival function of R_d is of the form

$$\Pr\{R_d \geq s\} \sim K s^{g_2(\mathbf{w}^*)-1/2} \exp\{-sg_1(\mathbf{w}^*)\} \qquad \text{(B.3.11)}$$

as $s \to \infty$ and for constant $K > 0$, which is the required result.

We note that $\mathbf{w}^*$ cannot typically be solved analytically. However, if $\sigma_i = \sigma$ for

all $i = 1, \ldots, d$ and $\rho_{ij} = \rho$ for $i \neq j$, then

$$g_1(\mathbf{w}) = \sigma \begin{pmatrix} \sqrt{w_1} \\ \vdots \\ \sqrt{\left(1 - \sum_j w_j\right)} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \sqrt{w_1} \\ \vdots \\ \sqrt{\left(1 - \sum_j w_j\right)} \end{pmatrix},$$

and it can be shown that

$$\mathbf{w}^* = \underset{\mathbf{w} \in [0,1]^{d-1}}{\arg\min}\ g_1(\mathbf{w}) = (1/d, \ldots, 1/d).$$

Hence, $g_1(\mathbf{w}^*) = \sigma \mathbb{1}^T \Sigma^{-1} \mathbb{1}/d = \sigma(d\eta_d)^{-1}$ and $g_2(\mathbf{w}^*) = \frac{1}{2}\mathbb{1}^T \Sigma^{-1} \mathbb{1} = 1/(2\eta_d)$. To prove this, we first note that the problem of minimising g_1 can be approached from the perspective of a non-linear program; by setting $v_i = \sqrt{\sigma_i w_i}$ for all $i = 1, \ldots, d$, the problem is of the form

$$\begin{aligned} \underset{\mathbf{v}}{\text{minimize}} \quad & g_1(\mathbf{v}) = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}, \\ \text{subject to} \quad & f_i(\mathbf{v}) = -v_i < 0,\ i = 1, \ldots, d, \ \text{ and} \\ & h(\mathbf{v}) = \sum_{i=1}^{d} \frac{v_i^2}{\sigma_i} = 1. \end{aligned}$$

Using this framework, any optimal solution $\mathbf{v}^*$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions (Kuhn and Tucker, 1951), namely

$$-\nabla g_1(\mathbf{v}^*) = \lambda \nabla h(\mathbf{v}^*) + \sum_{i=1}^{d} \mu_i \nabla f_i(\mathbf{v}^*), \tag{B.3.12}$$

$$\mu_i f_i(\mathbf{v}^*) = 0, \quad \text{for } \ i = 1, \ldots, d, \tag{B.3.13}$$

$$\mu_i \geq 0, \quad \text{for } \ i = 1, \ldots, d, \tag{B.3.14}$$

for constant $\lambda \in \mathbb{R}$. Note that we do not require v_i strictly greater than 0 as we can show that the density is non-degenerate in this the case. To satisfy all of the above conditions implies that any optimal solution $\mathbf{v}^* = (v_1^*, \ldots, v_d^*)$ must solve

$$
\begin{aligned}
&-\frac{1}{2}\begin{pmatrix} v_1^* \\ \vdots \\ v_d^* \end{pmatrix}^T \Sigma^{-1} = 2\lambda \begin{pmatrix} v_1^*/\sigma_1 \\ \vdots \\ v_d^*/\sigma_d \end{pmatrix}^T - \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} \\
&\Rightarrow -\frac{1}{2}\sum_{i=1}^{d} v_i^* \Sigma_{ij}^{-1} = 2\lambda v_j^*/\sigma_j - \mu_j \quad \text{for all} \quad j = 1, \ldots, d \\
&\Rightarrow \frac{\sigma_j}{v_j^*}\left(\frac{1}{2}\sum_{i=1}^{d} v_i^* \Sigma_{ij}^{-1} - \mu_j\right) = \frac{\sigma_k}{v_k}\left(\frac{1}{2}\sum_{i=1}^{d} v_i^* \Sigma_{ik}^{-1} - \mu_k\right), \qquad \text{(B.3.15)}
\end{aligned}
$$

for all $j = 1, \ldots, d$ and $k = 1, \ldots, d$. Solving this system of equations is non-trivial and numerical optimisation is difficult as the objective function is multi-modal; furthermore, it is not necessarily clear whether or not the optimal solution lies on the boundaries of the domain of $\mathbf{v}$, i.e., $v_i = 0$ for any $i \in \{1, \ldots, d\}$. However, if $\sigma_i = \sigma$ for all $i = 1, , \ldots, d$ and $\rho_{i,j} = \rho \in (0, 1)$ for all $i \neq j$, then we can illustrate that $\mathbf{v}^* = (1/\sqrt{d}, \ldots, 1/\sqrt{d})$ by solving (B.3.15) explicitly.

We begin by deriving the entries of Σ^{-1}. Note that we can write $\Sigma = (1-\rho)I_d + \rho \mathbb{1}_{d\times d}$, where $\mathbb{1}_{d\times d}$ is a $d \times d$ matrix of ones and I_d denotes the $d \times d$ identity matrix. With Σ of this form, Miller (1981) illustrates that

$$
\begin{aligned}
\Sigma^{-1} &= \frac{1}{(1-\rho)} I_d - \frac{\rho}{1 + tr\left(\frac{\rho}{1-\rho}\mathbb{1}_{d\times d}\right)} \frac{1}{(1-\rho)^2} \mathbb{1}_{d\times d} \\
&= \frac{1}{(1-\rho)} I_d - \frac{\rho}{(1-\rho+d\rho)(1-\rho)} \mathbb{1}_{d\times d},
\end{aligned}
$$

where $tr(\cdot)$ denotes the trace of a matrix. Thus,

$$\Sigma_{ij}^{-1} = \begin{cases} a = \frac{1}{(1-\rho)} - \frac{\rho}{(1-\rho+d\rho)(1-\rho)} = \frac{1-2\rho+d\rho}{(1-\rho)(1-\rho+d\rho)}, & \text{if } i = j, \\ b = -\frac{\rho}{(1-\rho+d\rho)(1-\rho)}, & \text{if } i \neq j. \end{cases} \tag{B.3.16}$$

Substituting this into (B.3.15) with $\sigma_i = \sigma$ for all $i = 1, \ldots, d$ gives the system of equations

$$\begin{aligned} \frac{\sigma}{2v_j^*}\left(\sum_{i=1}^{d} v_i^* \Sigma_{ij}^{-1} - 2\mu_j\right) = \frac{\sigma}{2v_j^*}\left(av_j + b\sum_{i\neq j}^{d} v_i^* - \mu_j\right) &= \frac{\sigma}{2}\left(a + \frac{b}{v_j^*}\sum_{i\neq j}^{d} v_i^* - \mu_j\right) \\ &= \frac{\sigma}{2}\left(a + \frac{b}{v_k^*}\sum_{i\neq k}^{d} v_i^* - \mu_k\right), \end{aligned} \tag{B.3.17}$$

for all $j = 1, \ldots, d$ and $k = 1, \ldots, d$, which simplifies to

$$bv_k^* \sum_{i\neq j}^{d} v_i^* - v_k^* v_j \mu_j = bv_j^* \sum_{i\neq k}^{d} v_i^* - v_j^* v_k^* \mu_k.$$

Now consider constraint (B.3.13); we have that either $\mu_i = 0$ or $v_i^* = 0$ for all $i = 1, \ldots, d$. We consider the latter case first; here (B.3.17) becomes

$$v_k^* \sum_{i\neq j}^{d} v_i^* = 0,$$

which would imply that v_k^* is also zero so as not to break constraint (B.3.14). In a similar fashion, we can show that if v_j^* is zero for any j, then all other v_k^* for $k \neq j$ are also zero; this breaks the constraint that $\sum_{i=1}^{d}(v_i^*)^2 = 1$ and hence does not lead to a feasible solution. Thus, we must instead have $\mu_i = 0$ for all $i = 1, \ldots, d$, and so solve

$$v_k^* \sum_{i\neq j}^{d} v_i^* = v_j^* \sum_{i\neq k}^{d} v_i^*,$$

by letting $v_i^* = v^*$ for all $i = 1, \ldots, d$. With the constraint that $\sum_{i=1}^d (v_i^*)^2 = 1$, it follows that $\mathbf{v}^* = (1/\sqrt{d}, \ldots, 1/\sqrt{d})$, and hence $\mathbf{w}^* = (1/d, \ldots, 1/d)$, as required.

B.4 Pairing independent exponential scale parameters

Here we explore the behaviour of R_4 for $X_i \sim \text{Exp}(\sigma_i)$ for $i = 1, \ldots, 4$. For the scale parameters $\sigma_1, \ldots, \sigma_4$, there are five cases that we need to consider:

- $\sigma_i = \sigma$ for all $i = 1, \ldots, 4$,
- $\sigma_i \neq \sigma_j$ for all $i, j = 1, \ldots, 4$ and $i \neq j$,
- There is one pair of equal scale parameters only,
- There are two pairs of equal scale parameters only,
- There is one triple of equal scale parameters only.

Note that the first two cases are covered in Section B.3.2. Without loss of generality, we cover the third case by letting $\sigma_1 = \sigma_2$, for the fourth case we let $\sigma_1 = \sigma_2, \sigma_3 = \sigma_4$ and $\sigma_1 \neq \sigma_3$, and for the fifth case we let $\sigma_1 = \sigma_2 = \sigma_3$. We begin by re-writing (B.3.1) as

$$\begin{aligned} f_{R_d}(r) &= \left(\prod_i^d \sigma_i\right) \sum_{j=1}^d \frac{e^{-r\sigma_j}}{\prod_{k\neq j}(\sigma_k - \sigma_j)} \\ &= \frac{\left(\prod_i^d \sigma_i\right)}{\prod_{i=2}^d \prod_{k<i}(\sigma_i - \sigma_k)} \left[\sum_{j=1}^d (-1)^{j+1} \prod_{i\neq j, i>1}^d \prod_{i>k} (\sigma_i - \sigma_k) e^{-r\sigma_j}\right] = \prod_{i=1}^d \sigma_i \frac{g(\boldsymbol{\sigma})}{h(\boldsymbol{\sigma})}, \end{aligned}$$

where

$$g(\boldsymbol{\sigma}) = \sum_{j=1}^d (-1)^{j+1} \prod_{i\neq j, i>1}^d \prod_{i>k} (\sigma_i - \sigma_k) e^{-r\sigma_j}, \text{ and } h(\boldsymbol{\sigma}) = \prod_{i=2}^d \prod_{k<i} (\sigma_i - \sigma_k).$$

For $d = 4$, it follows that

$$\begin{aligned} g(\boldsymbol{\sigma}) &= (\sigma_4 - \sigma_3)(\sigma_4 - \sigma_2)(\sigma_3 - \sigma_2)e^{-r\sigma_1} - (\sigma_4 - \sigma_3)(\sigma_4 - \sigma_1)(\sigma_3 - \sigma_1)e^{-r\sigma_2} \\ &\quad + (\sigma_4 - \sigma_2)(\sigma_4 - \sigma_1)(\sigma_2 - \sigma_1)e^{-r\sigma_3} - (\sigma_3 - \sigma_2)(\sigma_3 - \sigma_1)(\sigma_2 - \sigma_1)e^{-r\sigma_4}, \\ h(\boldsymbol{\sigma}) &= (\sigma_4 - \sigma_3)(\sigma_4 - \sigma_2)(\sigma_4 - \sigma_1)(\sigma_3 - \sigma_2)(\sigma_3 - \sigma_1)(\sigma_2 - \sigma_1). \end{aligned}$$

Through repeated use of L'Hôpital's rule, we can explore the effect of different pairings of σ_i.

B.4.1 One pair

Let $\sigma_1 \to \sigma_2$. Note that $g(\sigma_2, \sigma_2, \dots) = h(\sigma_2, \sigma_2, \dots) = 0$, and so by L'Hôpital's rule,

$$f_{R_4}(r) = \prod_{i=1}^{4} \sigma_i \lim_{\sigma_1 \to \sigma_2} \frac{g'_{\sigma_1}(\sigma_1, \sigma_2, \dots)}{h'_{\sigma_1}(\sigma_1, \sigma_2, \dots)},$$

where $g'_{\sigma_1}(\cdot)$ represents the first partial derivative of $g(\cdot)$ with respect to the subscript. Note that from here, we only consider the function $g(\cdot)$, as $h(\cdot)$ contributes only to the proportionality constant of the density. It follows that

$$\begin{aligned} g'_{\sigma_1}(\sigma_1, \sigma_2, \dots) &= -(\sigma_4 - \sigma_3)(\sigma_4 - \sigma_2)(\sigma_3 - \sigma_2)re^{-r\sigma_1} - (\sigma_4 - \sigma_3)(-\sigma_4 - \sigma_3 + 2\sigma_1)e^{-r\sigma_2} \\ &\quad + (\sigma_4 - \sigma_2)(-\sigma_4 - \sigma_2 + 2\sigma_1)e^{-r\sigma_3} - (\sigma_3 - \sigma_2)(-\sigma_3 - \sigma_2 + 2\sigma_1)e^{-r\sigma_4}, \end{aligned}$$

where

$$\begin{aligned} \lim_{\sigma_1 \to \sigma_2} g'_{\sigma_1}(\sigma_1, \sigma_2, \dots) &= -(\sigma_4 - \sigma_3)(\sigma_4 - \sigma_2)(\sigma_3 - \sigma_2)re^{-r\sigma_2} - (\sigma_4 - \sigma_3)(-\sigma_4 - \sigma_3 + 2\sigma_2)e^{-r\sigma_2} \\ &\quad + (\sigma_4 - \sigma_2)(-\sigma_4 + \sigma_2)e^{-r\sigma_3} - (\sigma_3 - \sigma_2)(-\sigma_3 + \sigma_2)e^{-r\sigma_4}. \end{aligned}$$

Hence,

$$f_{R_4}(r) = K_1\Big[-(\sigma_4-\sigma_3)(\sigma_4-\sigma_2)(\sigma_3-\sigma_2)re^{-r\sigma_2} - (\sigma_4-\sigma_3)(-\sigma_4-\sigma_3+2\sigma_2)e^{-r\sigma_2} + (\sigma_4-\sigma_2)(-\sigma_4+\sigma_2)e^{-r\sigma_3} - (\sigma_3-\sigma_2)(-\sigma_3+\sigma_2)e^{-r\sigma_4}\Big],$$

for constant $K_1 = \prod_{i=1}^4 \sigma_i / \lim_{\sigma_1\to\sigma_2} h'_{\sigma_1}(\sigma_1,\sigma_2,\dots)$. The density of R_4 is then

$$f_{R_4}(r) \sim \begin{cases} K_1 re^{-r\sigma_2}, & \text{if } \sigma_2 = \min\limits_{i=1,\dots,4}\sigma_i, \\ K_2 e^{-r\sigma_j}, & \text{for } \sigma_j = \min\limits_{i=1,\dots,4}\{\sigma_i\} \neq \sigma_2, \end{cases}$$

as $r\to\infty$ and for constants $K_1, K_2 > 0$.

B.4.2 Two pairs

We now consider $\sigma_1\to\sigma_2$ and $\sigma_3\to\sigma_4$. Note that

$$\lim_{\sigma_3\to\sigma_4} g'_{\sigma_1}(\sigma_2,\sigma_2,\sigma_3,\sigma_4) = \lim_{\sigma_3\to\sigma_4} h'_{\sigma_1}(\sigma_2,\sigma_2,\sigma_3,\sigma_4) = 0,$$

so we apply L'Hôpital's rule again. We now have

$$g'_{\sigma_1,\sigma_3}(\sigma_2,\sigma_2,\sigma_3,\sigma_4) = -(\sigma_4-\sigma_2)(\sigma_4-\sigma_2-2\sigma_3)e^{-r\sigma_2} - (2\sigma_3-2\sigma_2)e^{-r\sigma_2} - (\sigma_4-\sigma_2)(\sigma_2-\sigma_4)re^{-r\sigma_3} - (2\sigma_2-2\sigma_3)e^{-r\sigma_4},$$

which has limit

$$\lim_{\sigma_3\to\sigma_4} g'_{\sigma_1,\sigma_3}(\sigma_2,\sigma_2,\sigma_3,\sigma_4) = (\sigma_4-\sigma_2)(\sigma_4+\sigma_2)e^{-r\sigma_2} - 2(\sigma_4-\sigma_2)e^{-r\sigma_2} + (\sigma_4-\sigma_2)^2 re^{-r\sigma_4} - 2(\sigma_2-\sigma_4)e^{-r\sigma_4}.$$

Using a similar procedure to as the previous case, it follows that

$$f_{R_4}(r) \sim \begin{cases} K_1 r e^{-r\sigma_2}, & \text{if } \sigma_2 = \min_{i=1,\dots,4} \sigma_i, \\ K_2 r e^{-r\sigma_4}, & \text{if } \sigma_4 = \min_{i=1,\dots,4} \sigma_i, \end{cases}$$

as $r \to \infty$ and for constants $K_1, K_2 > 0$.

B.4.3 One triple

Finally, we consider $\sigma_1 \to \sigma_2$ and $\sigma_2 \to \sigma_3$. We begin by noting that

$$\begin{aligned} g'_{\sigma_1}(\sigma_2, \sigma_2, \dots) = &-(\sigma_4 - \sigma_3)(\sigma_4 - \sigma_2)(\sigma_3 - \sigma_2) r e^{-r\sigma_2} - (\sigma_4 - \sigma_3)(-\sigma_4 - \sigma_3 + 2\sigma_2) e^{-r\sigma_2} \\ &+ (\sigma_4 - \sigma_2)(-\sigma_4 + \sigma_2) e^{-r\sigma_3} - (\sigma_3 - \sigma_2)(-\sigma_3 + \sigma_2) e^{-r\sigma_4}, \end{aligned}$$

and so we now have

$$\begin{aligned} g'_{\sigma_1,\sigma_3}(\sigma_2, \sigma_2, \sigma_3, \sigma_4) = &-(\sigma_4 - \sigma_2)(\sigma_4 + \sigma_2 - 2\sigma_3) r e^{-r\sigma_2} - 2(\sigma_3 - \sigma_2) e^{-r\sigma_2} \\ &- (\sigma_4 - \sigma_2)(-\sigma_4 + \sigma_2) r e^{-r\sigma_3} - 2(\sigma_2 - \sigma_3) e^{-r\sigma_4}. \end{aligned}$$

Note that

$$\lim_{\sigma_2 \to \sigma_3} g'_{\sigma_1,\sigma_3}(\sigma_2, \sigma_2, \sigma_3, \sigma_4) = 0,$$

and it can also be shown that $\lim_{\sigma_2 \to \sigma_3} h'_{\sigma_1,\sigma_3}(\sigma_2, \sigma_2, \sigma_3, \sigma_4) = 0$. Hence, we can apply L'Hôpital's rule a second time. Considering the partial derivative of g with respect to σ_1 once and σ_3 twice, we have

$$g^{1,2}_{\sigma_1,\sigma_3}(\sigma_2, \sigma_2, \sigma_3, \sigma_4) = 2(\sigma_4 - \sigma_2) r e^{-r\sigma_2} - 2e^{-r\sigma_2} + (\sigma_4 - \sigma_2)(-\sigma_4 + \sigma_2) r^2 e^{-r\sigma_3} + 2e^{-r\sigma_4}$$

and

$$\lim_{\sigma_2 \to \sigma_3} g^{1,2}_{\sigma_1,\sigma_3}(\sigma_2,\sigma_2,\sigma_3,\sigma_4) = 2(\sigma_4-\sigma_3)re^{-r\sigma_3}-2e^{-r\sigma_3}+(\sigma_4-\sigma_3)(-\sigma_4+\sigma_3)r^2e^{-r\sigma_3}+2e^{-r\sigma_4}.$$

Hence,

$$f_{R_4}(r) \sim \begin{cases} K_1 r^2 e^{-r\sigma_3} & \text{if } \min_i \sigma_i = \sigma_3, \\ K_2 e^{-r\sigma_4} & \text{if } \min_i \sigma_i = \sigma_4, \end{cases}$$

as $r \to \infty$ and for constants $K_1, K_2 > 0$. We note that integration of the density of R_4 provided for each case leads to the survival function of R_4, which satisfies the general form given by (B.3.2).

Appendix C

Supplementary material for Chapter 5

C.1 Delta-Laplace conditional distribution functions

C.1.1 Connection to main text

This appendix supports the material presented in Sections 5.2.2 and 5.3.1 of the main paper. This appendix describes contributions to the pairwise likelihood function in (5.3.2) given by the conditional distribution associated with the residual distribution described in Section 5.2.2 of the main text.

C.1.2 Bivariate conditional distribution

We note that if we had $Y(s_i) > 0$ for all $i = 1, \ldots, d$, then the corresponding joint density of (5.2.9) for observed residuals $\mathbf{z} = (z_1, \ldots, z_{d-1}) \in ((c_1, \infty) \times \cdots \times (c_{d-1}, \infty))$

is

$$f_{s_O}(\mathbf{z}) = \phi_{d-1}\left\{\Phi^{-1}(F_{Z(s_1|s_O)}(z_1)), \ldots, \Phi^{-1}(F_{Z(s_{d-1}|s_O)}(z_{d-1})); \mathbf{0}, \Sigma\right\} \prod_{i=1}^{d-1} \frac{f_{Z(s_i|s_O)}(z_i)}{\phi(\Phi^{-1}(F_{Z(s_i|s_O)}))}, \tag{C.1.1}$$

where $c_1, \ldots, c_{d-1}$ are the censoring thresholds defined in Section 5.3.1 of the main text, $\phi(\cdot)$ is the PDF of a standard Gaussian distribution and $\phi_{d-1}(\cdot; \mathbf{0}, \Sigma)$ is the CDF of a $(d-1)-$dimensional Gaussian distribution with mean $\mathbf{0}$ and correlation matrix Σ defined in (5.2.10); components of $\mathbf{z}$ are not censored when in this domain.

For censoring thresholds c_1, c_2, the bivariate density with the conditioning constraints is

$$g_{s_O}(z_1, z_2; c_1, c_2) = \begin{cases} f_{s_O}(z_1, z_2) & \text{if } z_1 > c_1, z_2 > c_2, \\ f_{s_O}(z_1) F_{s_O,\, 2|1}\{c_2, z_1\} & \text{if } z_1 > c_1, z_2 \leq c_2, \\ f_{s_O}(z_2) F_{s_O,\, 1|2}\{c_1, z_2\} & \text{if } z_1 \leq c_1, z_2 > c_2, \\ F_{s_O}(c_1, c_2) & \text{if } z_1 \leq c_1, z_2 \leq c_2, \end{cases} \tag{C.1.2}$$

where the conditional distribution $F_{s_O,\, i|j}(c, z)$ is given by

$$F_{s_O,\, i|j}(c_i, z_j) = \Phi\left\{\Phi^{-1}(F_{Z(s_i|s_O)}(c_i)); \mu_{i|j}, \Sigma_{i|j}\right\}, \tag{C.1.3}$$

where $\mu_{i|j} = \Sigma_{ij}\Sigma_{jj}^{-1}\Phi^{-1}\{F_{Z(s_j|s_O)}(z_j)\}$ and $\Sigma_{i|j} = \Sigma_{ij} - \Sigma_{ij}^2\Sigma_{jj}^{-1}$ with Σ defined in (5.2.10). Both $f_{s_O}(\cdot, \cdot)$ and $F_{s_O}(\cdot, \cdot)$ are the bivariate analogues of (5.2.9) and (C.1.1) and $F_{Z(s_i|s_O)}$ is the CDF of a delta-Laplace distribution with parameters given in (5.2.7); these and Σ are all identifiable as they are all fully determined by pairwise distances.

C.1.3 Multivariate extension of (C.1.3)

Suppose we have some $s_O \in \mathcal{S}$ and $\boldsymbol{s}_A \in \mathcal{S} \setminus s_O$, which need not be any of the original sampling locations, and is indexed such that $\mathbf{s}_A = (s_1, \ldots, s_{n_A})$. We partition the indices of set $\boldsymbol{s}_A$ into two sets; indices for censored sites and non-censored sites, A_c and A_{nc}, respectively and without loss of generality we write $A_c = (s_1, \ldots, s_{n_c})$ and $A_{nc} = (s_{n_c+1}, \ldots, s_{n_A})$. If sites in A_c are censored with thresholds $c_1, \ldots, c_{n_c}$, then the conditional distribution function of

$$\Big(Z(s_1)|s_O), \ldots, Z(s_{n_c})|s_O)\Big) \Big| \Big(Z(s_{n_c+1}|s_O) = z_{n_c+1}, \ldots, Z(s_{n_A}|s_O) = z_{n_A}\Big)$$

is $F_{s_O,\, A_c|A_{nc}}(c_1, \ldots, c_{n_c}, z_{n_c+1}, \ldots, z_{n_A})$ equals to

$$\Phi_{n_c}\left\{\Phi^{-1}(F_{Z(s_1|s_O)}(c_1)), \ldots, \Phi^{-1}(F_{Z(s_{n_c}|s_O)}(c_{n_c})); \mu_{A_c|A_{nc}}, \Sigma_{A_c|A_{nc}}\right\}, \tag{C.1.4}$$

where

$$\begin{aligned}
\mu_{A_c|A_{nc}} &= \Sigma_{A_cA_{nc}}\Sigma^{-1}_{A_{nc}A_{nc}}(\Phi^{-1}\{F_{Z(s_{n_c+1}|s_O)}(z_{n_c+1})\}, \ldots, \Phi^{-1}\{F_{Z(s_{n_A}|s_O)}(z_{n_A})\})^T \\
\Sigma_{A_c|A_{nc}} &= \Sigma_{A_cA_c} - \Sigma_{A_cA_{nc}}\Sigma^{-1}_{A_{nc}A_{nc}}\Sigma_{A_{nc}A_c}.
\end{aligned}$$

The notation Σ_{AB} denotes the partition of Σ that takes rows indexed by the set A and columns indexed by the set B.

C.2 Application dependence model evaluation

C.2.1 Connection to main text

This appendix supports the material in Section 5.4.3 of the main text. Appendix C.2.2 details the parameter estimates for the spatial AI model and Appendix C.2.3 details

a diagnostic, based on χ, for evaluating the fitted dependence model.

C.2.2 Table of AI model parameter estimates

Table C.2.1 gives the estimates and standard errors for the model parameters discussed in Section 5.4.3 of the main text. Note that those parameters without standard errors were treated as fixed in the model fit.

Table C.2.1: Parameter estimates (standard errors) to 2 d.p.

$\alpha(h)$			$\beta(h)$		
κ_{α_1}	κ_{α_2}	Δ	κ_{β_1}	κ_{β_2}	κ_{β_3}
1.95 (0.25)	0.73 (0.03)	0.00	38.58 (0.65)	1.02 (0.03)	1.00

$\mu(h)$			$\sigma(h)$	
κ_{μ_1}	κ_{μ_2}	κ_{μ_3}	κ_{σ_1}	κ_{σ_2}
0.65 (0.04)	0.28 (0.02)	140.00 (0.50)	34.22 (0.80)	0.89 (0.01)

$\delta(h)$			
κ_{δ_1}	κ_{δ_2}	κ_{δ_3}	κ_{δ_4}
0.43 (0.59)	0.46 (0.03)	142.14 (0.02)	1.00

$\rho(h)$		(5.2.11)	
κ_{ρ_1}	κ_{ρ_2}	θ	L
58.71 (0.59)	0.53 (0.01)	-0.18 (0.02)	0.93 (0.01)

C.2.3 Probability of no rain diagnostic

We further evaluate the extremal behaviour of our fitted model by proposing a novel statistic for extremal dependence of precipitation processes that is based on χ. We define

$$\chi_q^{(0)}(s, s_O) = \Pr\left\{(Y(s) = 0) \mid \left(Y(s_O) > F^{-1}_{Y(s_O)}(q)\right)\right\}, \tag{C.2.1}$$

for $s, s_O \in S$ and $q \in [0, 1]$. This is the probability of observing no rain at s given that an extreme is observed at s_O. If $\{Y(s)\}$ is truly stationary in both its marginal and dependence structures, we would expect this measure to be a function of distance; results in Section 5.4.2 of the main text suggest the former is not true, but that there is only limited marginal non-stationarity. Figure C.3.3 compares estimates of (C.2.1) from the model against estimates from the data, for different q, with s_O in the centre of $\mathcal{S}$. Model estimates are calculated empirically from 1×10^6 realisations of $\{Y(s) : s \in \mathcal{S}\}$, which are drawn using the scheme described in Section 5.3.4 of the main text. For lower values of q, the model captures the empirical values for the probability reasonably well. However, as q increases, model estimates of (C.2.1) decrease whilst empirical estimates remain broadly the same. This could suggest that our model for $\{Y(s)\}$ is unable to capture some of the dependence behaviour in the data, i.e., that caused by a mixture of processes.

C.3 Supplementary Figures

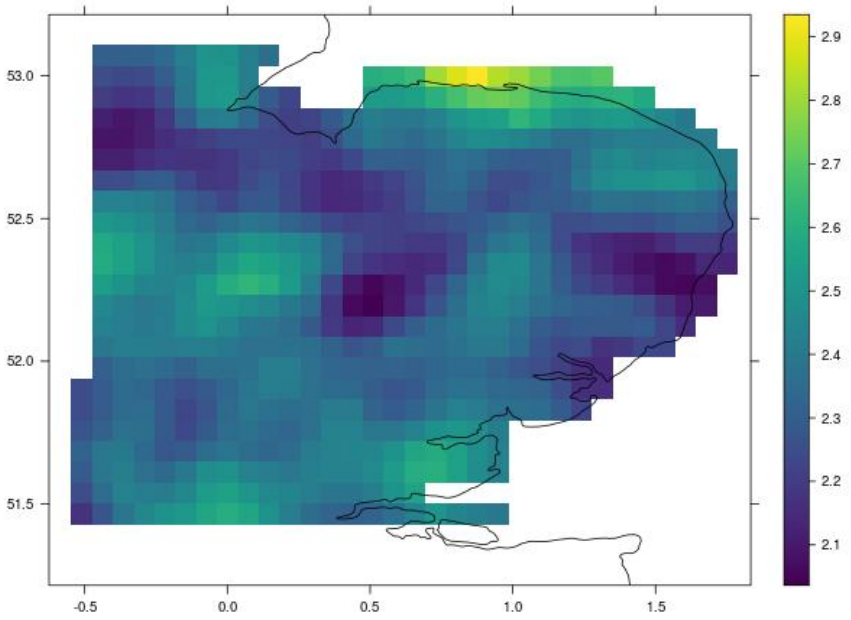

Figure C.3.1: Spatially smoothed estimate of $\upsilon(s)$ for East Anglia.

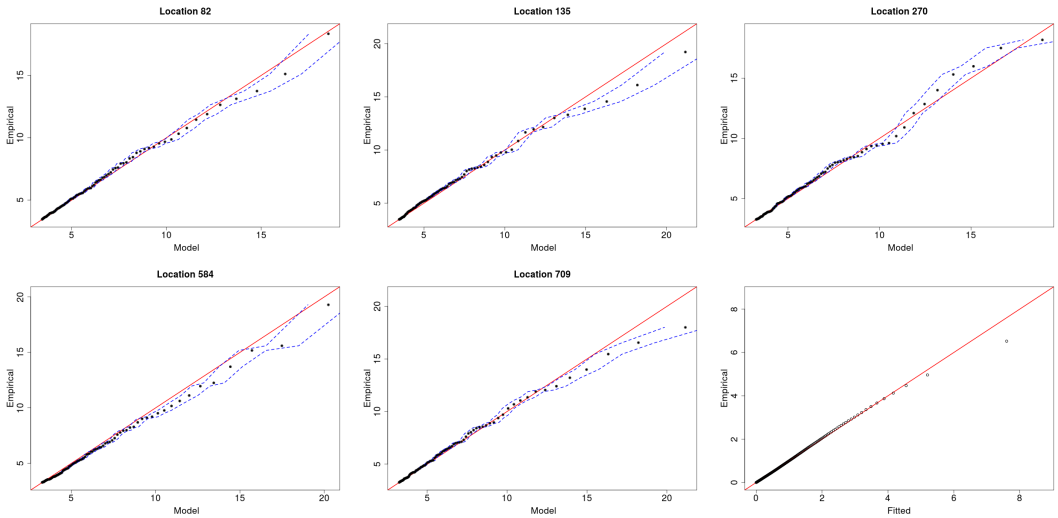

Figure C.3.2: Q-Q plots for the fitted GAM GPD distributions at five randomly sampled sites in $\mathcal{S}$. The 95% confidence intervals are given by the blue dashed lines. Bottom-right: Q-Q plot for pooled marginal transformation over all sites to standard exponential margins.

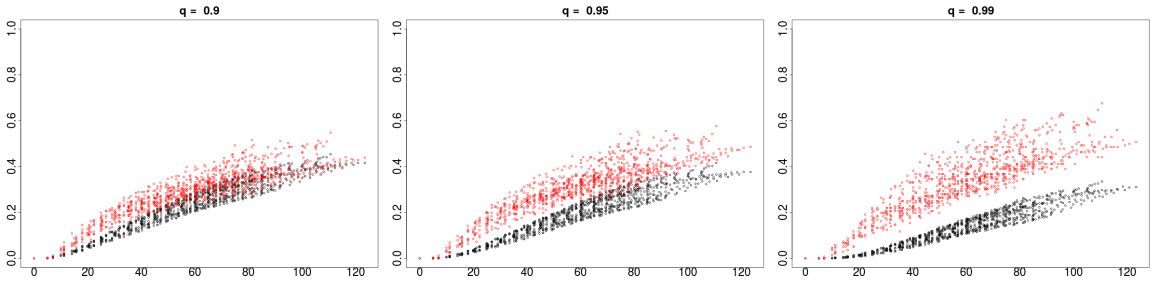

Figure C.3.3: Estimates of $\chi_q^{(0)}(s, s_O)$ against distance $h(s, s_O)$ for $q = (0.9, 0.95, 0.99)$. Red points denote empirical estimates, black points denote estimates derived from simulations from the fitted model.

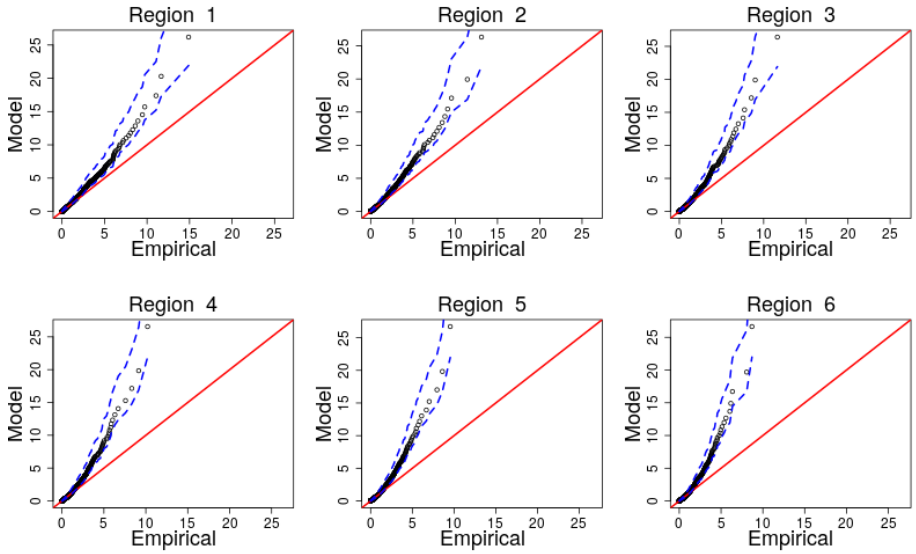

Figure C.3.4: Q-Q plots for AD model, and empirical, $\bar{R}_{\mathcal{A}}$ of regions of increasing size; these are illustrated in Figure 5.4.4 of the main text. Probabilities range from 0.7 to a value corresponding to the 20 year return level, with 95% confidence intervals given by the blue dashed lines.

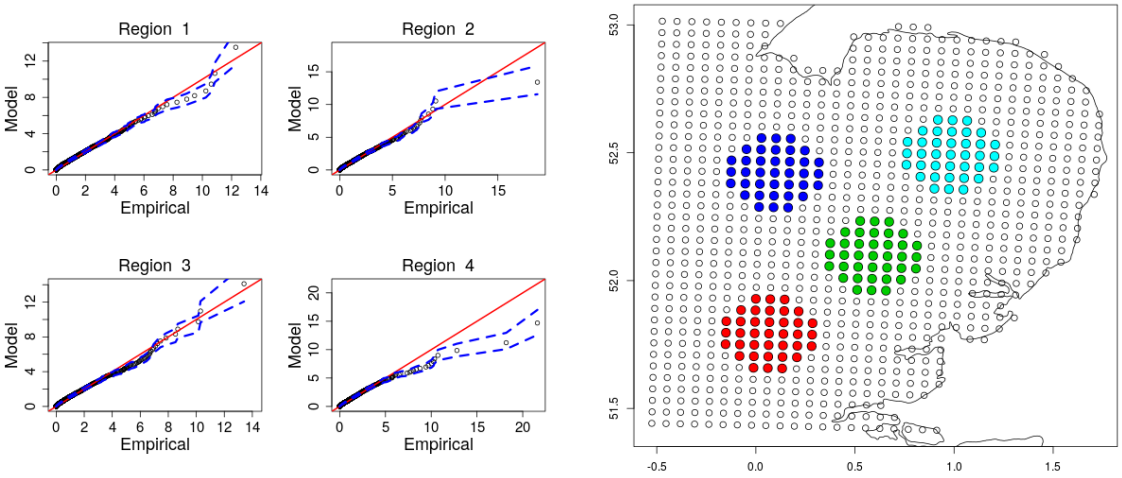

Figure C.3.5: Left: Q-Q plots for AI model, and empirical, $\bar{R}_{\mathcal{A}}$ of four regions in $\mathcal{S}$, with 95% confidence intervals given by the dashed lines. Probabilities range from 0.7 to a value corresponding to the 20 year return level. Right: regions, each with approximate area $925km^2$, are coloured 1-4 red, green, blue, cyan.

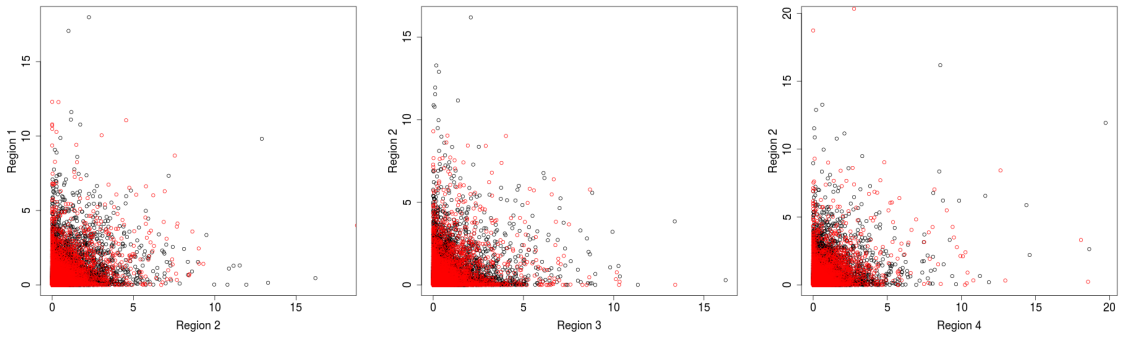

Figure C.3.6: Plots of 2×10^4 realisations of pairwise $(\bar{R}_{\mathcal{A}}, \bar{R}_{\mathcal{B}})$ for non-overlapping regions $\mathcal{A}, \mathcal{B}$, illustrated in Figure C.3.5. Black points are model estimates, red points are from the data. The regions $\mathcal{A}$ and $\mathcal{B}$ are labelled on the respective panels.

Appendix D

Supplementary material for Chapter 6

D.1 Supplementary Figures

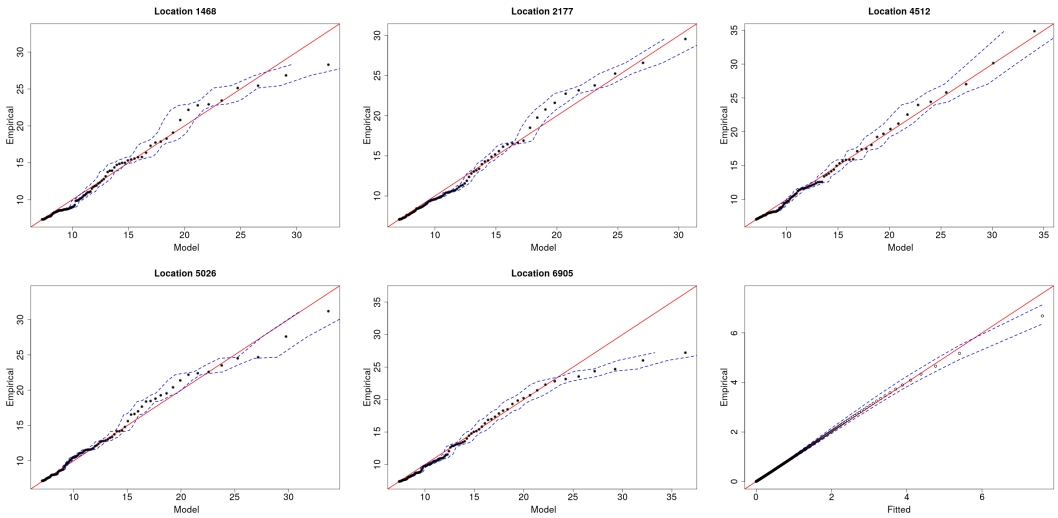

Figure D.1.1: Q-Q plots for the GPD GAM fits for $\{Y^{\mathcal{C}}(s)\}$ at five randomly sampled sites in $\mathcal{S}$. Bottom-right: Q-Q plot for pooled marginal transformation over all sites to standard exponential margins. The 95% confidence intervals are given by the blue dashed lines.

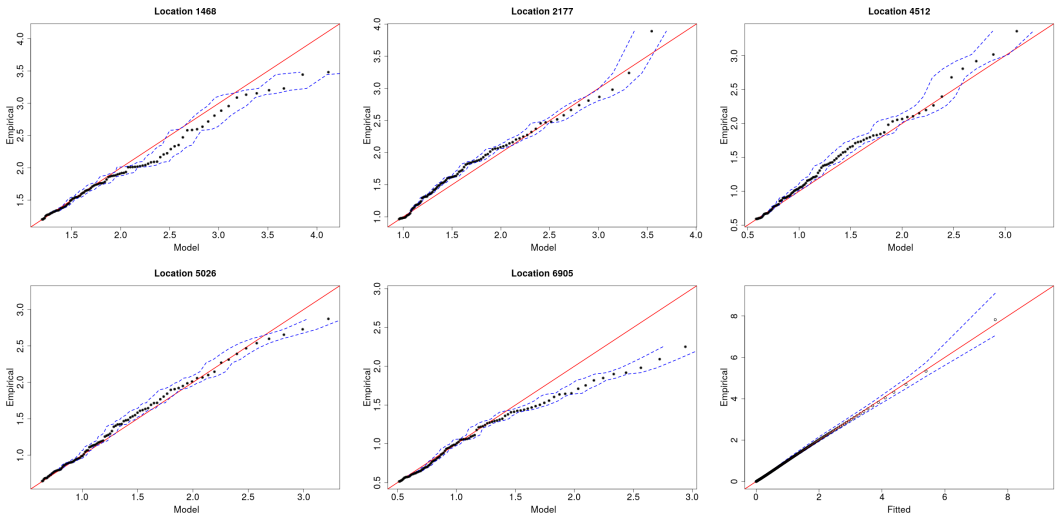

Figure D.1.2: Q-Q plots for the GPD GAM fits for $\{Y^{\mathcal{NC}}(s)\}$ at five randomly sampled sites in $\mathcal{S}$. Bottom-right: Q-Q plot for pooled marginal transformation over all sites to standard exponential margins. The 95% confidence intervals are given by the blue dashed lines.

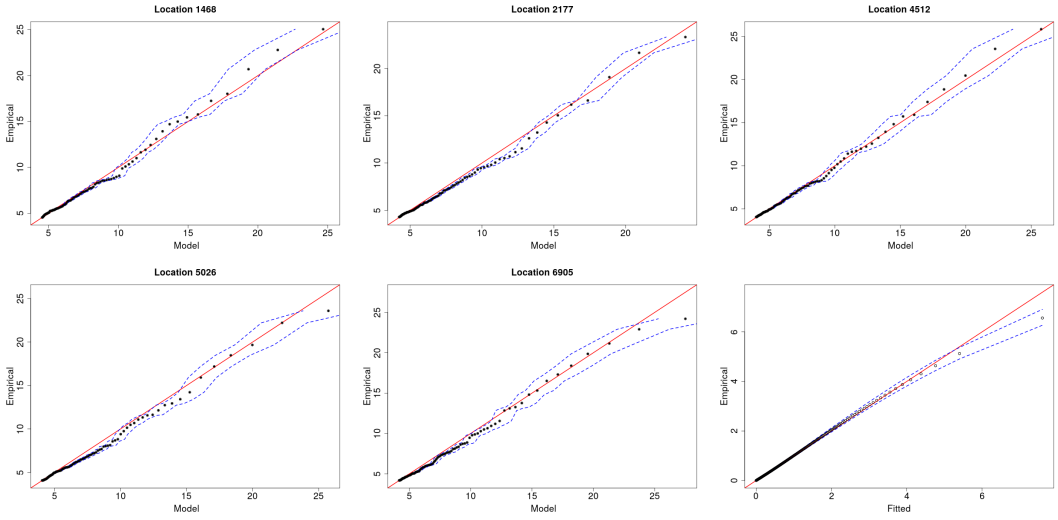

Figure D.1.3: Q-Q plots for the GPD GAM fits for $\{Y^{*}(s)\}$ at five randomly sampled sites in $\mathcal{S}$. Bottom-right: Q-Q plot for pooled marginal transformation over all sites to standard exponential margins. The 95% confidence intervals are given by the blue dashed lines.

Appendix E

Supplementary material for Chapter 7

E.1 Comparison of $\chi^{GG}(h)$ and $\chi^{BR}(h)$

In Section 7.2.2, we discuss using the theoretical $\chi(h)$ function from a Brown-Resnick model rather than a Gaussian-Gaussian model. This is because the former is less computationally expensive to compute and often approximates the latter very closely for $h \in \mathbb{R}_+$. To illustrate this, Figure E.1.1 shows the best approximation of $\chi^{BR}(h)$ to some fixed $\chi^{GG}(h)$ with Matérn correlation function and parameter set $(\theta_1 = 1, \theta_2, \sigma)$. Here θ_1 is set to 1 as this controls spatial scaling only. The functions $\chi^{BR}(h)$ are produced by minimising $\|\chi^{BR}(h) - \chi^{GG}(h)\|_F$ for a sequence of fixed $h \in [0, 10]$. Each figure uses different values of (θ_2, σ).

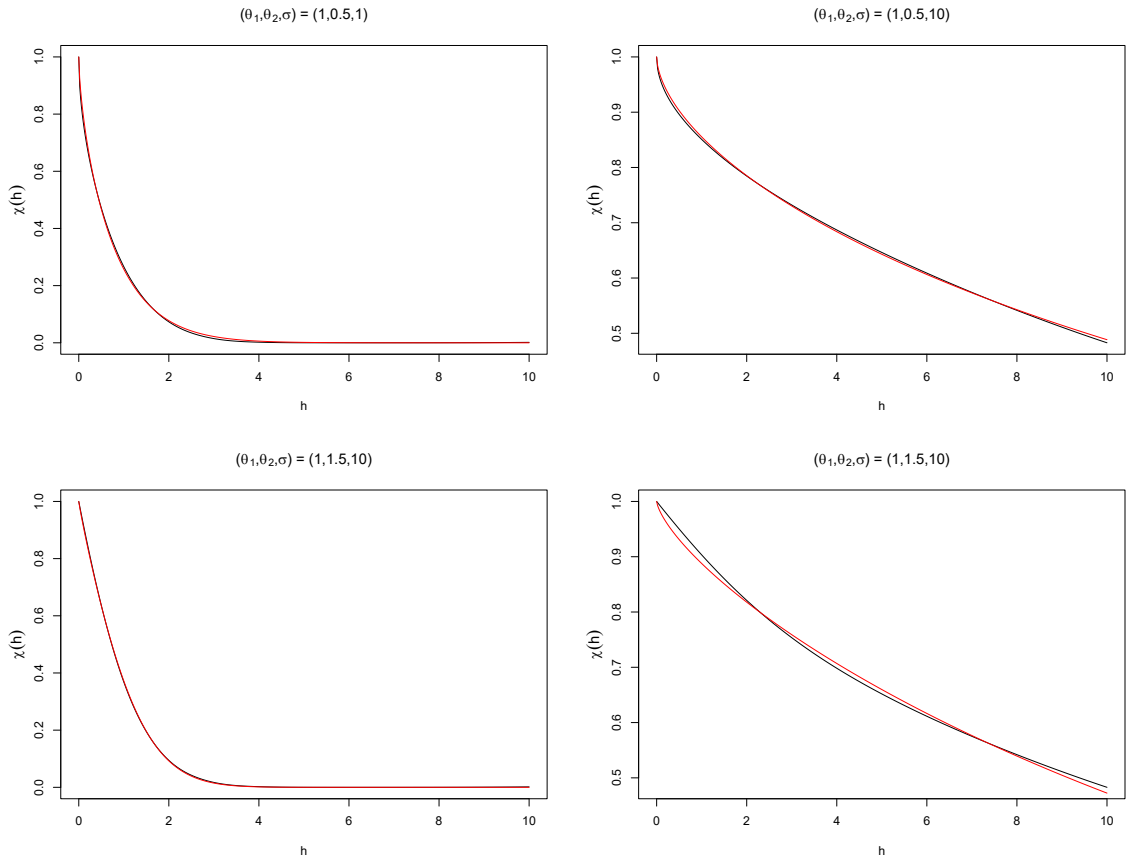

Figure E.1.1: Comparison of $\chi^{GG}(h)$ (black) and $\chi^{BR}(h)$ (red) for different parameter values.

E.2 High resolution heatmap of non-stationary max stable process in Sections 7.3.1 and 7.3.2

Figure E.2.1 gives a high-resolution heatmap of one realisation of a non-stationary Brown-Resnick process, with pairwise $\chi(s_i, s_j)$ given in (7.3.2). This process is used in the simulation studies in Sections 7.3.1 and 7.3.2.

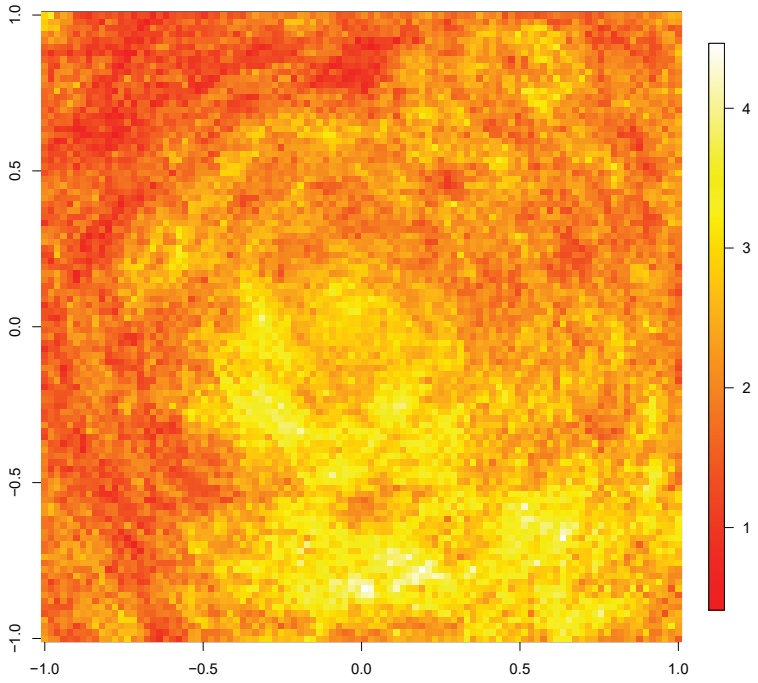

Figure E.2.1: Heatmap of one high-resolution realisation of the max-stable process proposed in Section 7.3.1 on a Gumbel marginal scale. This is sampled at 100×100 equally spaced points in $[-1, 1] \times [-1, 1]$. The parameter values in (7.3.2) are taken to be $o = (0, 0)$ and we have $(\lambda, \kappa) = (2, 0.8)$.

Bibliography

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Berg, P., Moseley, C., and Haerter, J. O. (2013). Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, 6(3):181–185.

Bernard, C., Denuit, M., and Vanduffel, S. (2018). Measuring portfolio risk under partial dependence information. *Journal of Risk and Insurance*, 85(3):843–863.

Blanchet, J. and Creutin, J.-D. (2017). Co-occurrence of extreme daily rainfall in the French Mediterranean region. *Water Resources Research*, 53(11):9330–9349.

Blanchet, J. and Davison, A. C. (2011). Spatial modeling of extreme snow depth. *The Annals of Applied Statistics*, 5(3):1699–1725.

Bopp, G. P., Shaby, B. A., and Huser, R. (2020). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association*, pages 1–14.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Brown, S. J. (2018). The drivers of variability in UK extreme rainfall. *International Journal of Climatology*, 38(S1):e119–e130.

Buishand, T. A., de Haan, L., Zhou, C., et al. (2008). On spatial extremes: With application to a rainfall problem. *The Annals of Applied Statistics*, 2(2):624–642.

Caesar, J., Alexander, L., and Vose, R. (2006). Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres*, 111(D5).

Casson, E. and Coles, S. G. (1999). Spatial regression models for extremes. *Extremes*, 1(4):449–468.

Castro-Camilo, D. and Huser, R. (2020). Local likelihood estimation of complex tail dependence structures, applied to U.S. precipitation extremes. *Journal of the American Statistical Association*, 115(531):1037–1054.

Castro-Camilo, D., Huser, R., and Rue, H. (2019). A spliced Gamma-Generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):517–534.

Castro-Camilo, D., Mhalla, L., and Opitz, T. (2021). Bayesian space-time gap filling for inference on extreme hot-spots: an application to Red Sea surface temperatures. *Extremes*, 24(1):105–128.

Chen, D., Mao, T., Pan, X., and Hu, T. (2012). Extreme value behavior of aggregate dependent risks. *Insurance: Mathematics and Economics*, 50(1):99 – 108.

Chevalier, C., Martius, O., and Ginsbourger, D. (2020). Modeling nonstationary extreme dependence with stationary max-stable processes and multidimensional scaling. *Journal of Computational and Graphical Statistics*, pages 1–11.

Coelho, C. A. S., Ferro, C. A. T., Stephenson, D. B., and Steinskog, D. J. (2008). Methods for exploring spatial and temporal variability of extreme events in climate data. *Journal of Climate*, 21(10):2072–2092.

Coles, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):797–816.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer series in statistics. Springer-Verlag London.

Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1990). Statistics of coastal flood prevention. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 332(1627):457–476.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):1–48.

Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(2):329–347.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cooley, D. and Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of agricultural, biological, and environmental statistics*, 15(3):381–402.

Davis, R. and Resnick, S. I. (1984). Tail estimates motivated by extreme value theory. *Annals of Statistics*, 12(4):1467–1487.

Davison, A. C., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):19. 511–529.

Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial extremes. In *Handbook of Environmental and Ecological Statistics*. CRC Press.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.

de Fondeville, R. and Davison, A. C. (2020). Functional peaks-over-threshold analysis. *arXiv preprint arXiv:2002.02711*.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204.

de Haan, L., Omey, E., and Resnick, S. (1984). Domains of attraction and regular variation in $\mathbb{R}^d$. *Journal of Multivariate Analysis*, 14(1):17–33.

de Melo Mendes, B. V. and Lopes, H. F. (2004). Data driven estimates for mixtures. *Computational statistics & data analysis*, 47(3):583–598.

Dieker, A. B. and Mikosch, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes*, 18(2):301–314.

Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics.* Springer Series in Statistics. Springer.

Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and its Interface*, 8(1):9–17.

Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika*, 94(4):809–825.

Eastoe, E. F. (2019). Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics*, 30(5):e2560.

Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.

Egger, J. and Hoinka, K. P. (1992). Fronts and orography. *Meteorology and Atmospheric Physics*, 48(1):3–36.

Eggert, B., Berg, P., Haerter, J. O., Jacob, D., and Moseley, C. (2015). Temporal and spatial scaling impacts on extreme precipitation. *Atmospheric Chemistry and Physics*, 15(10):5957–5971.

Embrechts, P., Wang, B., and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790.

Engelke, S., de Fondeville, R., and Oesting, M. (2019a). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, 106(1):127–144.

Engelke, S., Opitz, T., and Wadsworth, J. L. (2019b). Extremal dependence of random scale constructions. *Extremes*, 22(4):623–666.

Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., and Goude, Y. (2020). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, to appear.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Ferreira, A., de Haan, L., and Zhou, C. (2012). Exceedance probability of the integral of a stochastic process. *Journal of Multivariate Analysis*, 105(1):241–257.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Fouedjio, F., Desassis, N., and Romary, T. (2015). Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13:45–61.

Fougeres, A.-L. and Mercadier, C. (2012). Risk measures and multivariate extensions of Breiman's theorem. *Journal of Applied Probability*, 49(2):364–384.

Fowler, H. J. and Kilsby, C. G. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, 23(11):1313–1334.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70(351a):674–680.

Ghosh, I. and Banks, O. (2020). A study of bivariate generalized Pareto distribution and its dependence structure among model parameters. *Sankhya B*, pages 1–30.

Goovaerts, M. J., Kaas, R., Laeven, R. J., Tang, Q., and Vernic, R. (2005). The tail probability of discounted sums of Pareto-like losses in insurance. *Scandinavian Actuarial Journal*, 2005(6):446–461.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, United Kingdom.

Gregersen, I. B., Sørup, H. J. D., Madsen, H., Rosbjerg, D., Mikkelsen, P. S., and Arnbjerg-Nielsen, K. (2013). Assessing future climatic changes of rainfall extremes at small spatio-temporal scales. *Climatic Change*, 118(3):783–797.

Grimmett, G. S. (2020). *Probability and random processes*. Oxford university press.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, pages 127–145, Berlin, Heidelberg. Springer Berlin, Heidelberg.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Guttorp, P. and Sampson, P. D. (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. *Handbook of statistics*, 12:661–689.

Hauksson, H. A., Dacorogna, M., Domenig, T., Mller, U., and Samorodnitsky, G. (2001). Multivariate extremes, aggregation and risk estimation. *Quantitative Finance*, 1(1):79–95.

Hazra, A. and Huser, R. (2019). Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model. *arXiv preprint arXiv:1912.05657*.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2001). Extreme value analysis of a large designed experiment: A case study in bulk carrier safety. *Extremes*, 4(4):359–378.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Houze Jr, R. A. (1997). Stratiform precipitation in regions of convection: A meteorological paradox? *Bulletin of the American Meteorological Society*, 78(10):2179–2196.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.

Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):470–491.

Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics*, 21:166–186.

Huser, R., Opitz, T., and Thibaud, E. (2020). Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.

Huser, R. and Wadsworth, J. L. (2020). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, page e1537.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Iovleff, S. and Perrin, O. (2004). Estimating a nonstationary spatial structure using simulated annealing. *Journal of Computational and Graphical Statistics*, 13(1):90–105.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Chapman & Hall.

Jonathan, P., Randell, D., Wu, Y., and Ewans, K. (2014). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88:520–532.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, 25(17):5791–5806.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, 61(1):123–135.

Klüppelberg, C. and Resnick, S. I. (2008). The Pareto copula, aggregation of risks, and the emperor's socks. *Journal of Applied Probability*, 45(1):67–84.

Kole, E., Markwat, T., Opschoor, A., and van Dijk, D. (2017). Forecasting Value-at-Risk under temporal and portfolio aggregation. *Journal of Financial Econometrics*, 15(4):649–677.

Koutsoyiannis, D. (2020). *Stochastics of Hydroclimatic Extremes*, chapter 8, pages 271–273. National Technical University of Athens. http://itia.ntua.gr/2000/. Access date 13 May 2021.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear Programming. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492.

Laplace, P.-S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, 1(3):364 – 378. Original work published 1774.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.

Ledford, A. W. and Tawn, J. A. (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability*, 30(1):197–215.

Li, C., Zwiers, F., Zhang, X., and Li, G. (2019). How much information is required

to well constrain local estimates of future precipitation extremes? *Earth's Future*, 7(1):11–24.

Li, J. (2018). On the joint tail behavior of randomly weighted sums of heavy-tailed random variables. *Journal of Multivariate Analysis*, 164:40–53.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, T., Fosser, G., et al. (2018). UKCP18 science overview report. *Met Office Hadley Centre: Exeter, UK*.

Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, pages 484–502.

Meiring, W., Monestiez, P., Sampson, P. D., and Guttorp, P. (1997). Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. *Geostatistics Wollongong &.96. 1 (1997)*, 1:162.

Miller, K. S. (1981). On the inverse of the sum of matrices. *Mathematics magazine*, 54(2):67–72.

Morbidelli, R., Saltalippi, C., Flammini, A., Corradini, C., Wilkinson, S. M., and Fowler, H. J. (2018). Influence of temporal data aggregation on trend estimation for intense rainfall. *Advances in Water Resources*, 122:304 – 316.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt fur Mathematische Statistik*, 8:234–235.

Nadarajah, S. (2008). A review of results on sums of random variables. *Acta Applicandae Mathematicae*, 103:131–140.

Nadarajah, S., Anderson, C. W., and Tawn, J. A. (1998). Ordered multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):473–496.

Nadarajah, S. and Espejo, M. R. (2006). Sums, products, and ratios for the generalized bivariate Pareto distribution. *Kodai Mathematical Journal*, 29(1):72–83.

Nadarajah, S. and Kotz, S. (2008). The generalized Pareto sum. *Hydrological Processes*, 22(2):288–294.

Nadarajah, S., Zhang, Y., and Pogány, T. K. (2018). On sums of independent generalized Pareto random variables with applications to insurance and CAT bonds. *Probability in the Engineering and Informational Sciences*, 32(2):296–305.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer series in statistics. Springer, New York, 2nd edition.

Nguyen, Q. H. and Robert, C. Y. (2015). Series expansions for convolutions of Pareto distributions. *Statistics & Risk Modeling*, 32(1):49 – 72.

Northrop, P. (1998). A clustered spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1975):1875–1888.

Opitz, T. (2013). Extremal-t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on laplace random fields. *Spatial Statistics*, 16:1–18.

Opitz, T., Allard, D., and Mariethoz, G. (2021). Semi-parametric resampling with extremes. *Spatial Statistics*, 42:100445.

Opitz, T., Huser, R., Bakka, H. C., and Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21:441–462.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506.

Padoan, S. A. (2013). Extreme dependence models based on event magnitude. *Journal of Multivariate Analysis*, 122:1–19.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Palacios-Rodríguez, F., Toulemonde, G., Carreau, J., and Opitz, T. (2020). Generalized Pareto processes for simulating space-time extreme events: an application to precipitation reanalyses. *Stochastic Environmental Research and Risk Assessment*, 34(12):2033–2052.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

Perrin, O. and Meiring, W. (1999). Identifiability for non-stationary spatial structure. *Journal of Applied Probability*, 36(4):1244–1250.

Perrin, O. and Senoussi, R. (2000). Reducing non-stationary random fields to stationarity and isotropy using a space deformation. *Statistics & Probability Letters*, 48(1):23 – 32.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, 8(4):745–756.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981.*

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(1):219–241.

Ramsay, C. M. (2006). The distribution of sums of certain i.i.d. Pareto variates. *Communications in Statistics - Theory and Methods*, 35(3):395–405.

Ramsay, C. M. (2008). The distribution of sums of i.i.d. Pareto random variables with arbitrary shape parameter. *Communications in Statistics - Theory and Methods*, 37(14):2177–2184.

Reich, B. J. and Shaby, R. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Applied probability; vol. 4. Springer-Verlag, New York.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177.

Richards, J. and Tawn, J. A. (2021). On the tail behaviour of aggregated random variables. *arXiv preprint arXiv:2105.11917*.

Richards, J. and Wadsworth, J. L. (2021). Spatial deformation for nonstationary extremal dependence. *Environmetrics*, page e2671.

Roberts, N. and Kendon, E. (2020). personal communication.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London Series A*, 410(1839):269–288.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018a). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018b). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917 – 930.

Ross, E., Randell, D., Ewans, K., Feld, G., and Jonathan, P. (2017). Efficient estimation of return value distributions from non-stationary marginal extreme value models using bayesian inference. *Ocean Engineering*, 142:315 – 328.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Sampson, P. D. (2010). Spatial deformation models. In Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M., editors, *Handbook of Spatial Statistics*, pages 124–126. CRC press.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.

Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):49–65.

Sangati, M. and Borga, M. (2009). Influence of rainfall spatial resolution on flash flood modelling. *Natural Hazards & Earth System Sciences*, 9(2).

Saunders, K., Stephenson, A. G., Taylor, P. G., and Karoly, D. (2017). The spatial distribution of rainfall extremes and the influence of El Niño Southern Oscillation. *Weather and Climate Extremes*, 18:17 – 28.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Schroeer, K., Kirchengast, G., and O, S. (2018). Strong dependence of extreme convective precipitation intensities on gauge network density. *Geophysical Research Letters*, 45(16):8253–8263.

Sharkey, P. and Winter, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics*, 30(1):e2529.

Shooter, R., Ross, E., Tawn, J. A., and Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30(6).

Shooter, R., Tawn, J. A., Ross, E., and Jonathan, P. (2021). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24(2):241–265.

Simpson, E. S., Opitz, T., and Wadsworth, J. L. (2020). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and the SPDE approach. *arXiv e-prints*, arXiv:2011.04486.

Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:100482.

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

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*.

Smith, R. L. (1996). Estimating nonstationary spatial correlations. *Preprint, University of North Carolina*, 76.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and Ten Veldhuis, J. A. E. (2013). A statistical analysis of insurance damage claims related to rainfall extremes. *Hydrology & Earth System Sciences*, 17(3).

Stein, M. L. (2005). Nonstationary spatial covariance functions. *Unpublished technical report*.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Tawn, J. A., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*.

Thibaud, E., Mutzner, R., and Davison, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water Resources Research*, 49(8):4633–4644.

Thomassen, E. D., Sørup, H. J. D., Scheibel, M., Einfalt, T., and Arnbjerg-Nielsen, K. (2020). Data-driven distinction between convective, frontal and mixed extreme rainfall events in radar data. *Hydrology and Earth System Sciences Discussions*, 2020:1–26.

Vandeskog, S. M., Martino, S., Castro-Camilo, D., and Rue, H. (2021). Modelling short-term precipitation extremes with the blended generalised extreme value distribution. *arXiv preprint arXiv:2105.09062*.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Varty, Z., Tawn, J. A., Atkinson, P. M., and Bierman, S. (2021). Inference for extreme

earthquake magnitudes accounting for a time-varying measurement process. *arXiv preprint arXiv:2102.00884*.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2).

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15.

Wadsworth, J. L. and Tawn, J. A. (2019). Higher-dimensional spatial extremes via single-site conditioning. *arXiv preprint arXiv:1912.06560*.

Westra, S. and Sisson, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, 406(1-2):119–128.

Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment*, 19(6):403–416.

Wikle, C. K. (2010). Low-rank representations for spatial processes. In Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M., editors, *Handbook of Spatial Statistics*, pages 114–125. CRC press.

Winter, H., Tawn, J. A., and Brown, S. (2016). Modelling the effect of the El Niño-Southern Oscillation on extreme spatial temperature events over Australia. *Annals of Applied Statistics*, 10(4):2075–2101.

Youngman, B. D. (2019). Generalized additive models for exceedances of high thresholds with an application to return level estimation for U.S. wind gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.

Youngman, B. D. (2020). Flexible models for nonstationary dependence: Methodology and examples. *arXiv preprint arXiv:2001.06642*.

Zaliapin, I. V., Kagan, Y. Y., and Schoenberg, F. P. (2005). Approximating the distribution of Pareto sums. *Pure and Applied geophysics*, 162(6-7):1187–1228.

Zanini, E., Eastoe, E., Jones, M., Randell, D., and Jonathan, P. (2020). Flexible covariate representations for extremes. *Environmetrics*, 31(5):e2624.

Zheng, F., Thibaud, E., Leonard, M., and Westra, S. (2015). Assessing the performance of the independence method in modeling spatial extreme rainfall. *Water Resources Research*, 51(9):7744–7758.