Peaks over thresholds modelling with multivariate generalized Pareto distributions

Holger Rootzén
Chalmers University of Technology, Department of Mathematical Sciences
SE-412 96 Gothenburg, Sweden. E-mail: hrootzen@chalmers.se

Anna Kiriliouk Johan Segers
Université catholique de Louvain
Institut de Statistique, Biostatistique et Sciences Actuarielles
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
E-mail: anna.kiriliouk@uclouvain.be, johan.segers@uclouvain.be

Jennifer L. Wadsworth
Lancaster University, Deparment of Mathematics and Statistics
Fylde College LA1 4YF, Lancaster, England. E-mail: j.wadsworth@lancaster.ac.uk

Abstract

The multivariate generalized Pareto distribution arises as the limit of a suitably normalized vector conditioned upon at least one component of that vector being extreme. Statistical modelling using multivariate generalized Pareto distributions constitutes the multivariate analogue of peaks over thresholds modelling with the univariate generalized Pareto distribution. We introduce a construction device which allows us to develop a variety of new and existing parametric tail dependence models. A censored likelihood procedure is proposed to make inference on these models, together with a threshold selection procedure and several goodness-of-fit diagnostics. We illustrate our methods on two data applications, one concerning the financial risk stemming from the stock prices of four large banks in the United Kingdom, and one aiming at estimating the yearly probability of a rainfall which could cause a landslide in northern Sweden.

1 Introduction

1.1 Background

Peaks over threshold modelling of univariate time series has been common practice since the seminal paper of Davison and Smith (1990), who advocated the use of the asymptotically motivated generalized Pareto (GP) distribution (Balkema and de Haan, 1974; Pickands, 1975) as a model for exceedances over high thresholds. The multivariate generalized Pareto distribution was introduced in Tajvidi (1996), Beirlant et al. (2004, Chapter 8), and Rootzén and Tajvidi (2006), but still, statistical modelling using this approach has thus far received relatively little attention. Partially this is because theoretically equivalent dependence modelling approaches, based on the so-called "point process approach", have already been in existence for some time (Coles and Tawn, 1991). Nonetheless, the multivariate GP distribution has conceptual advantages over that of the point process representation, in so much as it represents a proper multivariate distribution on an "L-shaped" region, where at least one variable is extreme, $\{\boldsymbol{y} \in \mathbb{R}^d : \boldsymbol{y} \not\leq \boldsymbol{u}\}$; see Figure 1. Furthermore, the GP distribution permits modelling of data on this region without the need to perform any marginal transformation, which is common in other extremal dependence modelling approaches. The justification for the use of multivariate GP distributions as models for the tails of essentially arbitrary distributions is the limit property in (2.2).

There is a growing body of probabilistic literature devoted to multivariate GP distributions (e.g. Rootzén and Tajvidi (2006), Falk and Guillou (2008), Falk and Michel (2009), Ferreira

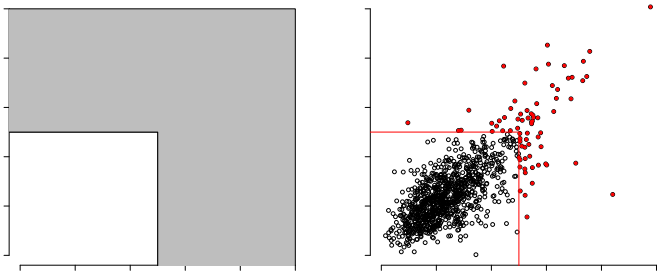

Figure 1: Illustration of the support of the multivariate GP distribution as an "L-shaped" region in two dimensions. Left-hand side: grey area represents the support of the GP distribution; right-hand side: filled points represent values that might be modelled by a GP distribution.

and de Haan (2014), Rootzén et al. (2016)). To our knowledge, however, there are only a few papers that exploit these as a statistical model (Thibaud and Opitz, 2015; Huser et al., 2016).

This paper forms a companion to Rootzén et al. (2016), which collates new and existing probabilistic results on multivariate GP distributions. We briefly recall necessary elements from this paper, but our main focus is on statistical modelling and inference.

1.2 Extremal dependence types

The premise of statistical modelling using limiting distributions such as the GP distribution is that *we assume the limit model holds sufficiently well in a sufficiently extreme region.* Care is needed, however, to ensure that the postulated model is indeed appropriate.

An important consideration when modelling multivariate extremes is the concept of asymptotic dependence. Random variables Y_1 and Y_2 with distribution functions F_1 and F_2, respectively, are said to be asymptotically dependent if

$$\chi_{12}(q) := \mathbb{P}[F_1(Y_1) > q \mid F_2(Y_2) > q] = \frac{\mathbb{P}[F_1(Y_1) > q, F_2(Y_2) > q]}{1-q} \to \chi_{1:2} > 0, \qquad q \to 1.$$

Existence of the limit is an assumption, and the limit being positive characterizes asymptotic dependence, whereas $\chi_{1:2} = 0$ defines asymptotic independence. An extension of $\chi_{1:2}$ to a general dimension $d \geq 2$ is given by

$$\chi_{1:d}(q) := \frac{\mathbb{P}[F_1(Y_1) > q, \ldots, F_d(Y_d) > q]}{1-q} \to \chi_{1:d}. \tag{1.1}$$

When $\chi_{1:d} > 0$, there is a positive probability of the most extreme events occurring simultaneously in all d variables. Multivariate GP distributions are useful mostly when $\chi_{1:d} > 0$. When $\chi_{1:d} = 0$ then the corresponding *limiting* GP distribution does not place any mass in d-dimensional space, with all the mass lying in some lower dimensional subspace instead. This situation is challenging to deal with, as in practice the data do belong to the full d-dimensional space. Whilst modelling of such possibilities is not precluded in the censored likelihood framework that we adopt, we do not consider it further here.

Yet other subtleties arise, since even when $\chi_{1:d} > 0$, this does not rule out the possibility of a distribution placing mass on some lower-dimensional subspace. However, for the remainder of the paper, we consider the simplified situation where there is no mass on any lower-dimensional subspace.

1.3 Outline

The contributions of this paper are the following. In Section 2 we present some of the key results and properties of multivariate GP distributions that are useful for statistical modelling. In Section 3 we introduce a construction device for GP-distributed random vectors and use it to develop a variety of new and existing models in Section 4. Inference using a censored likelihood-based approach is detailed in Section 5, together with threshold selection and goodness-of-fit diagnostics. In Section 6 we fit GP models to returns of four UK-based banks and to rainfall data in the context of landslide risk estimation, showing that multivariate GP modelling can be more useful for financial risk handling than one-dimensional methods, and that our models can respect physical constraints in a way which was not possible before. We conclude with a discussion in Section 7.

2 Background on GP distributions

Let $\boldsymbol{Y}$ be a random vector in $\mathbb{R}^d$ with distribution function F. A common and broadly-applicable assumption on $\boldsymbol{Y}$ is that it is in the so-called *max-domain of attraction* of a multivariate max-stable distribution, G. This means that if $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are independent and identically distributed copies of $\boldsymbol{Y}$, then one can find sequences $\boldsymbol{a}_n \in (0,\infty)^d$ and $\boldsymbol{b}_n \in \mathbb{R}^d$ such that

$$\mathbb{P}[\{\max_{1\leq i\leq n} \boldsymbol{Y}_i - \boldsymbol{b}_n\}/\boldsymbol{a}_n \leq \boldsymbol{x}] \to G(\boldsymbol{x}), \tag{2.1}$$

with G having non-degenerate margins. In (2.1) and throughout, vectors are boldface, and operations involving vectors are to be interpreted componentwise, with shorter vectors being recycled if necessary. The resulting max-stable distribution G has marginal location, scale and shape parameters denoted $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively, and the lower endpoints of its support are determined by these parameters. If $\boldsymbol{\eta}$ denotes this vector of lower endpoints, and $\boldsymbol{\sigma} = \boldsymbol{\mu} - \boldsymbol{\alpha}\boldsymbol{\gamma}$, its components are $\eta_j = -\sigma_j/\gamma_j$ if $\gamma_j > 0$ and $\eta_j = -\infty$ otherwise. We assume $\boldsymbol{\sigma} > \boldsymbol{0}$, which is always possible through appropriate choice of $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$. Max-stable distributions are not a primary concern of this paper, but the above mild assumption leads us to an analogous convergence theorem for multivariate threshold exceedances. Specifically, if convergence (2.1) holds, then

$$\max\left\{\frac{\boldsymbol{Y}-\boldsymbol{b}_n}{\boldsymbol{a}_n}, \boldsymbol{\eta}\right\} \mid \boldsymbol{Y} \not\leq \boldsymbol{b}_n \xrightarrow{d} \boldsymbol{X}, \qquad \text{as } t \to \infty, \tag{2.2}$$

where $\boldsymbol{X}$ follows a multivariate GP distribution (Beirlant et al., 2004; Rootzén et al., 2016). We let H denote the distribution function of $\boldsymbol{X}$, and $H_1, \ldots, H_d$ its marginal distributions. Typically the margins H_j are not univariate GP, due to the difference between the conditioning events $\{Y_j > b_{n,j}\}$ and $\{\boldsymbol{Y} \not\leq \boldsymbol{b}_n\}$ in the one-dimensional and d-dimensional limits. However, the marginal distributions conditioned to be positive are GP distributions. That is, writing $a_+ = \max(a, 0)$, we have

$$\overline{H}_j^+(x) := \mathbb{P}[X_j > x \mid X_j > 0] = (1 + \gamma_j x/\sigma_j)_+^{-1/\gamma_j}, \tag{2.3}$$

where σ_j and γ_j are as previously defined, giving the link between marginal parameterizations for the two convergences. The full link between H and G is $H(\boldsymbol{x}) = \{\log G(\min(\boldsymbol{x}, \boldsymbol{0})) - \log G(\boldsymbol{x})\}/\{\log G(\boldsymbol{0})\}$, and we say that such a H and G are *associated*. The support of the multivariate GP distribution H is included in the set

$$\{\boldsymbol{x} \in \mathbb{R}^d : x_j \geq \eta_j \text{ for all } j = 1, \ldots, d, \text{ and } x_j > 0 \text{ for some } j\}.$$

The dependence structure of H does not have a finite-dimensional parameterization (Beirlant et al., 2004), but it does satisfy certain properties that can be used to construct flexible parametric models. In Section 4 we give several examples, some of which are GP models that are associated to well-known max-stable models.

Following common practice in the statistical modelling of extremes, H may be used as a model for data which arise as multivariate threshold excesses in the sense $\boldsymbol{Y} \not\leq \boldsymbol{u}$. In particular, if $\boldsymbol{u} \in \mathbb{R}^d$ is a threshold that is "sufficiently high" in each margin, then from (2.2), $\max\{\boldsymbol{Y} - \boldsymbol{u}, \boldsymbol{\eta}\} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ can be approximated by a member of the class of multivariate GP distributions, with $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$, the marginal exceedance probabilities $\mathbb{P}(Y_j > u_j)$, and the dependence structure to be estimated. In practice the truncation by the unknown vector $\boldsymbol{\eta}$ is only relevant when dealing with mass on lower-dimensional subspaces. Following the discussion in Section 1.2, we suppose that $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ is to be approximated by a GP distribution. A member of the class of GP distributions $\boldsymbol{X}$ has a representation on $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \boldsymbol{0}\}$ as

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\sigma} \frac{e^{\boldsymbol{X}_0 \boldsymbol{\gamma}} - \boldsymbol{1}}{\boldsymbol{\gamma}}, \tag{2.4}$$

where $\boldsymbol{X}_0$ is a "standard form" GP random vector, that is, a GP vector on a standardized scale $\boldsymbol{\gamma} = \boldsymbol{0}$ and $\boldsymbol{\sigma} = \boldsymbol{1}$. Its construction will be further discussed in Section 3. For $\gamma_j = 0$, the corresponding component of the right-hand side of equation (2.4) is simply $\sigma_j X_{0,j}$.

The following are useful properties of the GP distributions; for further details and proofs we refer to Rootzén et al. (2016).

Threshold stability. GP distributions are *threshold stable*, meaning that if $\boldsymbol{X} \sim H$ follows a GP distribution with marginal parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ then for $\boldsymbol{w} \geq \boldsymbol{0}$ such that $H(\boldsymbol{w}) < 1$ and $\boldsymbol{\sigma} + \boldsymbol{\gamma}\boldsymbol{w} > \boldsymbol{0}$,

$$\boldsymbol{X} - \boldsymbol{w} \mid \boldsymbol{X} \not\leq \boldsymbol{w} \text{ is GP with parameters } \boldsymbol{\sigma} + \boldsymbol{\gamma}\boldsymbol{w} \text{ and } \boldsymbol{\gamma}.$$

This property states that if we increase, or at least do not decrease, the level of the threshold in each margin, then the distribution of conditional excesses is still GP, with a new set of scale parameters, but retaining the same vector of shape parameters.

A special role is played by the levels $\boldsymbol{w} = \boldsymbol{w}_t := \boldsymbol{\sigma}(t^{\boldsymbol{\gamma}} - 1)/\boldsymbol{\gamma}$: these have the stability property that for any set $A \subset \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \boldsymbol{0}\}$ it holds that, for $t \geq 1$,

$$\mathbb{P}[\boldsymbol{X} \in \boldsymbol{w}_t + t^{\boldsymbol{\gamma}} A] = \mathbb{P}[\boldsymbol{X} \in A]/t, \tag{2.5}$$

where $\boldsymbol{w}_t + t^{\boldsymbol{\gamma}} A = \{\boldsymbol{w}_t + t^{\boldsymbol{\gamma}} \boldsymbol{x} : \boldsymbol{x} \in A\}$. This follows from equation (2.4) along with the representation of $\boldsymbol{X}_0$ to be given in equation (3.1). The j-th component of $\boldsymbol{w}_t$, $\sigma_j(t^{\gamma_j} - 1)/\gamma_j$, is the $1 - 1/t$ quantile of H_j^+. Equation (2.5) provides one possible tool for checking if a multivariate GP is appropriate; see Section 5.3.

Lower dimensional conditional margins. Lower dimensional margins of GP distributions are typically not GP, as the conditioning event leading to the distribution involves all d variables, and not only the variables in a lower dimensional margin. Let $J \subset \{1, \ldots, d\}$, with $\boldsymbol{X}_J = (x_j : j \in J)$, and similarly for other vectors. Then $\boldsymbol{X}_J \mid \boldsymbol{X}_J \not\leq \boldsymbol{0}_J$ does follow a GP distribution. Combined with the threshold stability property above, we also have that if $\boldsymbol{w}_J \in \mathbb{R}^{|J|}$ is such that $H_J(\boldsymbol{w}_J) < 1$ and $\boldsymbol{\sigma}_J + \boldsymbol{\gamma}_J \boldsymbol{w}_J > \boldsymbol{0}$ then $\boldsymbol{X}_J - \boldsymbol{w}_J \mid \boldsymbol{X}_J \not\leq \boldsymbol{w}_J$ follows a GP distribution.

Constant conditional exceedances. For a GP distribution H we define the conditional dependence measure $\chi_H(q)$ by

$$\chi_H(q) := \frac{\mathbb{P}[H_1(X_1) > q, \ldots, H_d(X_d) > q]}{1 - q}, \quad q \in (0, 1). \tag{2.6}$$

By property (2.5), for a suitable choice of A, it holds that $\chi_H(q)$ is constant for q sufficiently large such that $H_j(X_j) > q$ implies $X_j > 0$ for $j \in \{1, \ldots, d\}$. As $q \uparrow 1$, the limit of (2.6) coincides with the d-dimensional dependence measure $\chi_{1:d}$ in (1.1) and thus choices of H for which this matches the observed value are candidate models. This property can also help inform threshold selection; see Section 5.3.

Sum-stability under shape constraints If $\boldsymbol{X}$ follows a multivariate GP distribution, with scale parameter $\boldsymbol{\sigma}$ and shape parameter $\boldsymbol{\gamma} = \gamma \mathbf{1}$, then for weights $a_j > 0$ such that $\sum_{j=1}^d a_j X_j > 0$ with positive probability, we have

$$\textstyle\sum_{j=1}^d a_j X_j \mid \sum_{j=1}^d a_j X_j > 0 \sim \mathrm{GP}(\sum_{j=1}^d a_j \sigma_j, \gamma). \tag{2.7}$$

That is, weighted sums of components of a multivariate GP distribution with equal shape parameters, conditioned to be positive, follow a univariate GP distribution with the same shape parameter and with the scale parameter equal to the weighted sum of the marginal scale parameters. The dependence structure of the GP distribution does not affect this result, but it does affect the probability of the conditioning event, i.e., the probability that the sum of components is positive. Further details can be found in Rootzén et al. (2016).

3 Model construction

3.1 Standard form densities

We focus on how to construct suitable densities for the random vector $\boldsymbol{X}_0$, which through equation (2.4), leads to densities for the multivariate GP distribution with marginal parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$.

Let E be a unit exponential random variable and let $\boldsymbol{T}$ be a d-dimensional random vector, independent of E. Define $\max(\boldsymbol{T}) = \max_{1 \le j \le d} T_j$. Then the random vector

$$\boldsymbol{X}_0 = E + \boldsymbol{T} - \max(\boldsymbol{T}) \tag{3.1}$$

has the required properties to be a GP vector with support included in the set $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \mathbf{0}\}$ and with $\boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$ (interpreted as the limit for $\gamma_j \to 0$ for all j). Moreover, *every* such GP vector can be expressed in this way (Rootzén et al., 2016). The probability of the j-th component being positive is $\mathbb{P}[X_{0,j} > 0] = \mathbb{E}[e^{T_j - \max(\boldsymbol{T})}]$, which, in terms of the original data vector $\boldsymbol{Y}$, corresponds to the marginal exceedance probability $\mathbb{P}[Y_j > u_j \mid \boldsymbol{Y} \not\leq \boldsymbol{u}]$, i.e., the probability that the j-th component exceeds its corresponding threshold given that one of the d components does.

Suppose $\boldsymbol{T}$ has a density $f_{\boldsymbol{T}}$ on $(-\infty, \infty)^d$. By Theorem 5.1 of Rootzén et al. (2016), the density of $\boldsymbol{X}_0$ is given by

$$h_{\boldsymbol{T}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = \frac{\mathbb{1}\{\max(\boldsymbol{x}) > 0\}}{e^{\max(\boldsymbol{x})}} \int_0^\infty f_{\boldsymbol{T}}(\boldsymbol{x} + \log t) t^{-1} \, \mathrm{d}t. \tag{3.2}$$

One way to construct models therefore is to assume different distributions for $\boldsymbol{T}$, which provide flexible forms for $h_{\boldsymbol{T}}$, and for which ideally the integral in (3.2) can be evaluated analytically.

One further construction of GP random vectors is given in Rootzén et al. (2016). If $\boldsymbol{U}$ is a d-dimensional random vector with density $f_{\boldsymbol{U}}$ and such that $\mathbb{E}[e^{U_j}] < \infty$ for all $j = 1, \ldots, d$, then the following function also defines the density of a GP distribution:

$$h_{\boldsymbol{U}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = \frac{\mathbb{1}\{\max(\boldsymbol{x}) > 0\}}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} \int_0^\infty f_{\boldsymbol{U}}(\boldsymbol{x} + \log t) \, \mathrm{d}t. \tag{3.3}$$

The marginal exceedance probabilities are now $\mathbb{P}[X_{0,j} > 0] = \mathbb{E}[e^{U_j}]/\mathbb{E}[e^{\max(\boldsymbol{U})}]$ for $j = 1, \ldots, d$. Formulas (3.2) and (3.3) can be obtained from one another via a change of measure linking $f_{\boldsymbol{T}}$ and $f_{\boldsymbol{U}}$.

Where $f_{\boldsymbol{T}}$ and $f_{\boldsymbol{U}}$ take the same form, then the similarity in integrals between (3.2) and (3.3) means that if one can be evaluated, then typically so can the other; several instances of this are given in the models presented in Section 4. What is sometimes more challenging is calculation of the normalization constant $\mathbb{E}[e^{\max(\boldsymbol{U})}] = \int_0^\infty \{1 - F_{\boldsymbol{U}}(\log t\mathbf{1})\}\, \mathrm{d}t$ in (3.3). Nonetheless, the model in (3.3) has the particular advantage over that of (3.2) that it behaves better across various dimensions: if the density of the GP vector $\boldsymbol{X}$ is $h_{\boldsymbol{U}}$ and if $J \subset \{1, \ldots, d\}$, then the density of the GP subvector $\boldsymbol{X}_J \mid \boldsymbol{X}_J \not\leq \mathbf{0}_J$ is simply $h_{\boldsymbol{U}_J}$. This property is advantageous when moving to the spatial setting, since the model retains the same form when numbers of sites change, which is useful for spatial prediction.

3.2 Densities on the observed scale

In Section 3.1 we presented densities in the standardized form $\boldsymbol{\sigma} = \mathbf{1}, \boldsymbol{\gamma} = \mathbf{0}$. Using transformation (2.4), we obtain a general form for the corresponding densities for threshold excesses, $\boldsymbol{X}$, which in terms of the original data, is an approximation for the conditional density of $\boldsymbol{Y} - \boldsymbol{u}$ given that $\boldsymbol{Y} \not\leq \boldsymbol{u}$:

$$h(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\gamma}) = h\left(\tfrac{1}{\boldsymbol{\gamma}} \log(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma}); \mathbf{1}, \mathbf{0}\right) \prod_{j=1}^{d} \frac{1}{\sigma_j + \gamma_j x_j}. \tag{3.4}$$

In (3.4), h may be either $h_{\boldsymbol{T}}$ or $h_{\boldsymbol{U}}$. Observe that it is always possible to transform this density to be suitable for large values of $\boldsymbol{Y}$ itself, i.e., to a density for the conditional distribution of $\boldsymbol{Y}$ given that $\boldsymbol{Y} \not\leq \boldsymbol{u}$ by replacing $\boldsymbol{x}$ by $\boldsymbol{x} - \boldsymbol{u}$ in the expression above.

The models detailed in Section 3.1 are built on a standardized scale, and then transformed to the observed, or "real" scale. Alternatively a model can be constructed directly on the real scale, which gives the possibility of respecting structures, say additive structures, in a way which is not possible with the other two models; this approach will be used to model ordered data in Section 6.2. One way of presenting this is to define the random vector $\boldsymbol{R}$ in terms of $\boldsymbol{U}$ in (3.3) through the componentwise transformation

$$R_j = \begin{cases} (\sigma_j/\gamma_j)\exp(\gamma_j U_j), & \gamma_j \neq 0, \\ \sigma_j U_j, & \gamma_j = 0, \end{cases} \tag{3.5}$$

and develop suitable models for $\boldsymbol{R}$. This gives the GP density

$$h_{\boldsymbol{R}}(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \frac{\mathbb{1}\{\max(\boldsymbol{x}) > 0\}}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} \int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\boldsymbol{R}}\left(\left(g(t; x_j, \sigma_j, \gamma_j)\right)_{j=1}^d\right) \mathrm{d}t, \tag{3.6}$$

where $f_{\boldsymbol{R}}$ denotes the density of $\boldsymbol{R}$ and where

$$g(t; x_j, \sigma_j, \gamma_j) = \begin{cases} t^{\gamma_j}\left(x_j + \sigma_j/\gamma_j\right), & \gamma_j \neq 0, \\ x_j + \sigma_j \log t, & \gamma_j = 0. \end{cases}$$

The d components of $\boldsymbol{U}$ are found by inverting equation (3.5). For $\boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$, the densities (3.3) and (3.6) are the same.

4 Parametric models

Here we provide details of certain probability distributions for $\boldsymbol{T}$, $\boldsymbol{U}$ and $\boldsymbol{R}$ that generate tractable multivariate GP distributions. Below we refer to these probability distributions as

"generators". Several articles have previously focused on the use of random vectors to generate dependence structures for extremes, for example Segers (2012) and Aulbach et al. (2015) among others. The literature on max-stable modelling for spatial extremes also relies heavily on this device (Schlather, 2002; Davison et al., 2012).

In order to control bias when using a multivariate GP distribution, we often need to use censored likelihood (see Section 5) and thus not just to be able to calculate densities, but also integrals of those densities. These considerations guide our choice of models presented below. For each model we give the uncensored densities in the subsequent subsections, whilst their censored versions are given in Appendix A.

In Sections 4.1 and 4.2 we consider particular instances of densities $f_{\boldsymbol{T}}$ and $f_{\boldsymbol{U}}$ to evaluate the corresponding densities $h_{\boldsymbol{T}}$ and $h_{\boldsymbol{U}}$ in (3.2) and (3.3). As noted in Section 3.1, even if $f_{\boldsymbol{T}} = f_{\boldsymbol{U}}$, the GP densities $h_{\boldsymbol{T}}$ and $h_{\boldsymbol{U}}$ are still different in general. Thus we will focus on the density of a random vector $\boldsymbol{V}$ with density $f_{\boldsymbol{V}}$ and create two GP models per $f_{\boldsymbol{V}}$ by setting $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$ and then $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$, in the latter case with the restriction $\mathbb{E}[e^{U_j}] < \infty$. The support for each GP density given in Sections 4.1 and 4.2 is $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \boldsymbol{0}\}$. In Section 4.3 we exhibit a construction of $h_{\boldsymbol{R}}$ in (3.6) with support depending on $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$.

In all models, identifiability issues occur if $\boldsymbol{T}$ or $\boldsymbol{U}$ have unconstrained location parameters $\boldsymbol{\beta}$, or if $\boldsymbol{R}$ has unconstrained scale parameters $\boldsymbol{\lambda}$. Indeed, replacing $\boldsymbol{\beta}$ or $\boldsymbol{\lambda}$ by $\boldsymbol{\beta} + k$ or $c\boldsymbol{\lambda}$, respectively, with $k \in \mathbb{R}$ and $c > 0$, lead to the same GP distribution (Rootzén et al., 2016, Proposition 1). A single constraint, such as fixing the first parameter in the vector, is sufficient to restore identifiability.

4.1 Generators with independent components

Let $\boldsymbol{V} \in \mathbb{R}^d$ be a random vector with independent components and density $f_{\boldsymbol{V}}$, so that $f_{\boldsymbol{V}}(\boldsymbol{v}) = \prod_{j=1}^d f_j(v_j)$, where f_j are unspecified densities of real-valued random variables. The dependence structure of the associated GP distributions is determined by the relative heaviness of the tails of the marginal distributions: roughly speaking, if one component has a high probability of being "large" compared to the others, then the dependence is weaker than if all components have a high probability of taking similar values. Throughout, $\boldsymbol{x} \in \mathbb{R}^d$ is such that $\max(\boldsymbol{x}) > 0$.

Generators with independent Gumbel components

Let

$$f_j(v_j) = \alpha_j \exp\{-\alpha_j(v_j - \beta_j)\} \exp[-\exp\{-\alpha_j(v_j - \beta_j)\}], \qquad \alpha_j > 0,\, \beta_j \in \mathbb{R}.$$

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. Density (3.2) is

$$h_{\boldsymbol{T}}(\boldsymbol{x}; \boldsymbol{1}, \boldsymbol{0}) = e^{-\max(\boldsymbol{x})} \int_0^\infty t^{-1} \prod_{j=1}^d \alpha_j \left(te^{x_j - \beta_j}\right)^{-\alpha_j} e^{-(te^{x_j-\beta_j})^{-\alpha_j}} \, \mathrm{d}t. \tag{4.1}$$

If $\alpha_1 = \ldots = \alpha_d = \alpha$ then the integral can be explicitly evaluated:

$$h_{\boldsymbol{T}}(\boldsymbol{x}; \boldsymbol{1}, \boldsymbol{0}) = e^{-\max(\boldsymbol{x})} \alpha^{d-1} \frac{\Gamma(d) \prod_{j=1}^d e^{-\alpha(x_j-\beta_j)}}{\left(\sum_{j=1}^d e^{-\alpha(x_j-\beta_j)}\right)^d}.$$

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. The marginal expectation of the exponentiated variable is

$$\mathbb{E}[e^{U_j}] = \begin{cases} e^{\beta_j}\Gamma(1 - 1/\alpha_j), & \alpha_j > 1, \\ \infty, & \alpha_j \leq 1. \end{cases}$$

For $\min_{1\le j\le d}\alpha_j > 1$, density (3.3) is

$$h_{\boldsymbol{U}}(\boldsymbol{x};\mathbf{1},\mathbf{0}) = \frac{\int_0^\infty \prod_{j=1}^d \alpha_j \left(te^{x_j-\beta_j}\right)^{-\alpha_j} e^{-(te^{x_j-\beta_j})^{-\alpha_j}}\,\mathrm{d}t}{\int_0^\infty \left(1-\prod_{j=1}^d e^{-(t/e^{\beta_j})^{-\alpha_j}}\right)\mathrm{d}t}. \tag{4.2}$$

If $\alpha_1 = \ldots = \alpha_d = \alpha$ then this simplifies to:

$$h_{\boldsymbol{U}}(\boldsymbol{x};\mathbf{1},\mathbf{0}) = \frac{\alpha^{d-1}\Gamma(d-1/\alpha)\prod_{j=1}^d e^{-\alpha(x_j-\beta_j)}}{\left(\sum_{j=1}^d e^{-\alpha(x_j-\beta_j)}\right)^{d-1/\alpha}\Gamma(1-1/\alpha)\left(\sum_{j=1}^d e^{\beta_j\alpha}\right)^{1/\alpha}}.$$

Observe that if in addition to $\alpha_1 = \ldots = \alpha_d = \alpha$, we also have $\beta_1 = \ldots = \beta_d = 0$, then this is the multivariate GP distribution associated to the well-known *logistic* max-stable distribution.

Generators with independent reverse Gumbel components

Let

$$f_j(v_j) = \alpha_j \exp\{\alpha_j(v_j-\beta_j)\}\exp[-\exp\{\alpha_j(v_j-\beta_j)\}], \qquad \alpha_j > 0,\ \beta_j \in \mathbb{R}.$$

As the Gumbel case leads to the multivariate GP distribution associated to the logistic max-stable distribution, the reverse Gumbel leads to the multivariate GP distribution associated to the *negative logistic* max-stable distribution[1]. Calculations are very similar to the Gumbel case, and hence omitted.

Generators with independent reverse exponential components

Let

$$f_j(v_j) = \exp\{(v_j+\beta_j)/\alpha_j\}/\alpha_j, \qquad v_j \in (-\infty, -\beta_j),\ \alpha_j > 0,\ \beta_j \in \mathbb{R}.$$

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. Density (3.2) is

$$\begin{aligned} h_{\boldsymbol{T}}(\boldsymbol{x};\mathbf{1},\mathbf{0}) &= e^{-\max(\boldsymbol{x})}\int_0^{e^{-\max(\boldsymbol{x}+\boldsymbol{\beta})}} t^{-1}\prod_{j=1}^d \frac{1}{\alpha_j}(te^{x_j+\beta_j})^{1/\alpha_j}\,\mathrm{d}t \\ &= \frac{e^{-\max(\boldsymbol{x})-\max(\boldsymbol{x}+\boldsymbol{\beta})\sum_{j=1}^d 1/\alpha_j}}{\sum_{j=1}^d 1/\alpha_j}\prod_{j=1}^d \frac{1}{\alpha_j}(e^{x_j+\beta_j})^{1/\alpha_j}. \end{aligned} \tag{4.3}$$

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. The expectation of the exponentiated variable is $\mathbb{E}[e^{U_j}] = 1/\left\{e^{\beta_j}(\alpha_j+1)\right\}$, which is finite for all permitted parameter values. Density (3.3) is

$$\begin{aligned} h_{\boldsymbol{U}}(\boldsymbol{x};\mathbf{1},\mathbf{0}) &= \frac{1}{\mathbb{E}[e^{\max(\boldsymbol{U})}]}\int_0^{e^{-\max(\boldsymbol{x}+\boldsymbol{\beta})}} \prod_{j=1}^d \frac{1}{\alpha_j}(te^{x_j+\beta_j})^{1/\alpha_j}\,\mathrm{d}t \\ &= \frac{(e^{-\max(\boldsymbol{x}+\boldsymbol{\beta})})^{\sum_{j=1}^d 1/\alpha_j+1}}{\mathbb{E}[e^{\max(\boldsymbol{U})}]}\frac{1}{1+\sum_{j=1}^d 1/\alpha_j}\prod_{j=1}^d \frac{1}{\alpha_j}(e^{x_j+\beta_j})^{1/\alpha_j}. \end{aligned} \tag{4.4}$$

The normalization constant may be evaluated as

$$\mathbb{E}[e^{\max(\boldsymbol{U})}] = \int_0^\infty \left(1-\prod_{j=1}^d \min(e^{\beta_j}t,1)^{1/\alpha_j}\right)\mathrm{d}t$$

[1]The authors are grateful to Clément Dombry for having pointed out this connection.

$$= e^{-\beta_{(d)}} - \frac{\prod_{j=1}^{d} e^{\beta_j/\alpha_j}}{\sum_{j=1}^{d} 1/\alpha_j + 1} e^{-\beta_{(1)}(\sum_{j=1}^{d} 1/\alpha_j + 1)} + \sum_{i=1}^{d-1} \frac{\prod_{j=i+1}^{d} e^{\beta_{(j)}/\alpha_{[j]}}}{\sum_{j=i+1}^{d} 1/\alpha_{[j]} + 1} \left(e^{-\beta_{(i+1)}(\sum_{j=i+1}^{d} 1/\alpha_{[j]} + 1)} - e^{-\beta_{(i)}(\sum_{j=i+1}^{d} 1/\alpha_{[j]} + 1)} \right),$$

where $\beta_{(1)} > \beta_{(2)} > \cdots > \beta_{(d)}$ and where $\alpha_{[j]}$ is the component of $\boldsymbol{\alpha}$ with the same index as $\beta_{(j)}$ (thus the $\alpha_{[j]}$s are not in general ordered). As far as we are aware the associated max-stable model is not well known.

Generators with independent log-gamma components

Let

$$f_j(v_j) = \exp(\alpha_j v_j) \exp\{-\exp(v_j)\}/\Gamma(\alpha_j), \qquad \alpha_j > 0,\ v_j \in (-\infty, \infty),$$

i.e., $e^{V_j} \sim \text{Gamma}(\alpha_j, 1)$.

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. Density (3.2) is

$$h_{\boldsymbol{T}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \prod_{j=1}^{d} \left(\frac{e^{\alpha_j x_j}}{\Gamma(\alpha_j)} \right) \int_0^\infty t^{\sum_{j=1}^{d} \alpha_j - 1} e^{-t \sum_{j=1}^{d} e^{x_j}} \, \mathrm{d}t$$
$$= \frac{\Gamma\left(\sum_{j=1}^{d} \alpha_j\right)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \frac{e^{\sum_{j=1}^{d} \alpha_j x_j - \max(\boldsymbol{x})}}{(\sum_{j=1}^{d} e^{x_j})^{\sum_{j=1}^{d} \alpha_j}}.$$

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. The marginal expectation of the exponentiated variable is $\mathbb{E}[e^{U_j}] = \alpha_j$, hence finite for all permitted parameter values. Density (3.3) is

$$h_{\boldsymbol{U}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = \frac{1}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} \prod_{j=1}^{d} \left(\frac{e^{\alpha_j x_j}}{\Gamma(\alpha_j)} \right) \int_0^\infty t^{\sum_{j=1}^{d} \alpha_j} e^{-t \sum_{j=1}^{d} e^{x_j}} \, \mathrm{d}t$$
$$= \frac{1}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} \frac{\Gamma\left(\sum_{j=1}^{d} \alpha_j + 1\right)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \frac{e^{\sum_{j=1}^{d} \alpha_j x_j - \max(\boldsymbol{x})}}{(\sum_{j=1}^{d} e^{x_j})^{\sum_{j=1}^{d} \alpha_j + 1}}.$$

The normalization constant is

$$\mathbb{E}[e^{\max(\boldsymbol{U})}] = \frac{\Gamma\left(\sum_{j=1}^{d} \alpha_j + 1\right)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \int_{\Delta_{d-1}} \max(u_1, \ldots, u_d) \prod_{j=1}^{d} u_j^{\alpha_j - 1} \, \mathrm{d}u_1 \cdots \mathrm{d}u_{d-1},$$

where $\Delta_{d-1} = \{(u_1, \ldots, u_d) \in [0,1]^d : u_1 + \cdots + u_d = 1\}$ is the unit simplex, and the integral can be easily computed using the R package SimplicialCubature. This GP distribution is associated to the so-called Dirichlet max-stable distribution (Coles and Tawn, 1991; Segers, 2012).

4.2 Generators with multivariate Gaussian components

Let $f_{\boldsymbol{V}}(\boldsymbol{v}) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\{-(\boldsymbol{v} - \boldsymbol{\beta})^T \Sigma^{-1} (\boldsymbol{v} - \boldsymbol{\beta})/2\}$, where $\boldsymbol{\beta} \in \mathbb{R}^d$ is the mean parameter, and $\Sigma \in \mathbb{R}^{d \times d}$ is a positive-definite covariance matrix. As before, $\boldsymbol{x} \in \mathbb{R}^d$ is such that $\max(\boldsymbol{x}) > 0$. For these calculations, it is simplest to make the change of variables $s = \log t$ in equations (3.2) and (3.3).

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. Density (3.2) is

$$h_{\boldsymbol{T}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \int_{-\infty}^{\infty} \frac{(2\pi)^{-d/2}}{|\Sigma|^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\beta} - s\mathbf{1})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\beta} - s\mathbf{1})\right\} \mathrm{d}s$$
$$= \frac{(2\pi)^{(1-d)/2} |\Sigma|^{-1/2}}{(\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\beta})^T A (\boldsymbol{x} - \boldsymbol{\beta}) - \max(\boldsymbol{x})\right\} \tag{4.5}$$

with

$$A = \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{1}\mathbf{1}^T \Sigma^{-1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \tag{4.6}$$

a $d \times d$ matrix of rank $d - 1$.

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. As required, the expectation $\mathbb{E}[e_j^U] = e^{\beta_j + \Sigma_{jj}/2}$ is finite for all permitted parameter values, where Σ_{jj} denotes the jth diagonal element of Σ. Density (3.3) is

$$h_{\boldsymbol{U}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) = \frac{1}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} \int_{-\infty}^{\infty} \frac{(2\pi)^{-d/2}}{|\Sigma|^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\beta} - s\mathbf{1})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\beta} - s\mathbf{1}) + s\right\} \mathrm{d}s$$
$$= \frac{(2\pi)^{(1-d)/2} |\Sigma|^{-1/2}}{\mathbb{E}[e^{\max(\boldsymbol{U})}](\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\beta})^T A (\boldsymbol{x} - \boldsymbol{\beta}) + 2\frac{(\boldsymbol{x} - \boldsymbol{\beta})^T \Sigma^{-1} \mathbf{1} - 1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\right\},$$

with A as in (4.6). The distribution for this case is already known; it is the GP distribution associated to the Brown–Resnick or Hüsler–Reiss max-stable model (Kabluchko et al., 2009; Hüsler and Reiss, 1989). A variant of the density formula with $\mathbb{E}[e^{U_j}] = 1$ (equivalently $\boldsymbol{\beta} = -\mathrm{diag}(\Sigma)/2$) was given in Wadsworth and Tawn (2014). Although they focussed on Poisson point process inference, one can use their expressions to derive the associated GP distribution in a straightforward manner. The normalization constant is $\int_0^\infty [1 - \Phi_d(\log t\mathbf{1} - \boldsymbol{\beta}; \Sigma)] \, \mathrm{d}t$, where $\Phi_d(\cdot; \Sigma)$ is the zero-mean multivariate normal distribution function with covariance matrix Σ. This normalization constant can also be expressed as a sum of multivariate normal distribution functions, see Huser and Davison (2013). Observe that in high dimensions, the normalization constant for the density based on $\boldsymbol{U}$ may be onerous to compute, whilst the density based on $\boldsymbol{T}$ does not require this.

4.3 Generators with structured components

In Sections 4.1 and 4.2 we considered several distributions for $\boldsymbol{U}$ and $\boldsymbol{T}$. Here we present a model for $\boldsymbol{R}$ based on cumulative sums of exponential random variables. Cumulative sums will lead to a vector $\boldsymbol{R}$ whose components are ordered; for the components of the corresponding GP vector to be ordered as well, we will assume that $\boldsymbol{\gamma} = \gamma\mathbf{1}$ and $\boldsymbol{\sigma} = \sigma\mathbf{1}$. This model will be of interest in Section 6.2, where we will focus on modelling cumulative precipitation amounts which may trigger landslides.

Case $\boldsymbol{\gamma} = \mathbf{0}$. By construction, the densities $h_{\boldsymbol{R}}(\,\cdot\,; \mathbf{1}, \mathbf{0})$ and $h_{\boldsymbol{U}}(\,\cdot\,; \mathbf{1}, \mathbf{0})$ coincide since $\boldsymbol{R} = \boldsymbol{U}$. Let $\boldsymbol{R} \in (-\infty, \infty)^d$ be the random vector whose components are defined by

$$R_j = \log\left(\textstyle\sum_{i=1}^{j} E_i\right), \qquad E_j \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda_j), \qquad j = 1, \ldots, d,$$

where $\lambda_1, \ldots, \lambda_d > 0$ are the rate parameters, i.e., $\mathbb{P}[E_j > x_j] = \exp(-\lambda_j x_j)$ for $x_j > 0$. Its density, $f_{\boldsymbol{R}}$, is given by

$$f_{\boldsymbol{R}}(\boldsymbol{r}) = \begin{cases} \left(\prod_{j=1}^d \lambda_j e^{r_j}\right) \exp\left\{-\sum_{j=1}^d (\lambda_j - \lambda_{j+1}) e^{r_j}\right\}, & \text{if } r_1 < \ldots < r_d, \\ 0, & \text{otherwise,} \end{cases}$$

where we set $\lambda_{d+1} = 0$. In view of (3.3), $R_1 < \ldots < R_d$ (or equivalently $U_1 < \ldots < U_d$) implies $X_{0,1} < \ldots < X_{0,d}$. The density of $\boldsymbol{X}_0$ is given as follows: if $x_1 < \ldots < x_d$, then

$$\begin{aligned} h_{\boldsymbol{R}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0}) &= \frac{\mathbb{1}\left(x_d > 0\right)}{\mathbb{E}[e^{R_d}]} \left(\prod_{j=1}^{d} \lambda_j e^{x_j} \right) \int_0^\infty t^d \exp\left\{ -t \left(\sum_{j=1}^{d} (\lambda_j - \lambda_{j+1}) e^{x_j} \right) \right\} \mathrm{d}t \\ &= \frac{\mathbb{1}(x_d > 0)\, d! \prod_{j=1}^{d} \lambda_j e^{x_j}}{\left(\sum_{j=1}^{d} \lambda_j^{-1} \right) \left(\sum_{j=1}^{d} (\lambda_j - \lambda_{j+1}) e^{x_j} \right)^{d+1}}, \end{aligned} \tag{4.7}$$

while $h_{\boldsymbol{R}}(\boldsymbol{x}; \mathbf{1}, \mathbf{0})$ is zero otherwise. The density $h_{\boldsymbol{R}}(\boldsymbol{x}; \boldsymbol{\sigma}, \mathbf{0})$ is then directly obtained from (3.4).

For $d = 2$, the dependence measure $\chi_{1:2}$ from (1.1) takes the simple form $\chi_{1:2} = 1 - \lambda_1/(2(\lambda_1 + \lambda_2))$. Some properties of the structured components model can be inferred from this expression; when $\lambda_1 = \lambda_2$, then $\chi_{12} = 0.75$ regardless of the value of the parameter. If $\lambda_1 \gg \lambda_2$, then $\chi_{12} \to 0.5$; if $\lambda_2 \gg \lambda_1$, then $\chi_{12} \to 1$. It is natural that this model cannot approach asymptotic independence, since it is based on cumulative sums.

Case $\boldsymbol{\gamma} > \mathbf{0}$. Let $\boldsymbol{R} \in (0, \infty)^d$ be the random vector whose components are defined by

$$R_j = \sum_{i=1}^{j} E_i, \qquad E_j \overset{\text{iid}}{\sim} \text{Exp}(\lambda_j), \qquad j = 1, \ldots, d,$$

where $\lambda_1, \ldots, \lambda_d > 0$ are the rate parameters. Its density, $f_{\boldsymbol{R}}$, is given by

$$f_{\boldsymbol{R}}(\boldsymbol{r}) = \begin{cases} \left(\prod_{j=1}^{d} \lambda_j \right) \exp\left\{ -\sum_{j=1}^{d} (\lambda_j - \lambda_{j+1}) r_j \right\}, & \text{if } 0 < r_1 < \ldots < r_d, \\ 0, & \text{otherwise,} \end{cases}$$

where we set $\lambda_{d+1} = 0$. Then

$$\mathbb{E}\left[e^{\max(\boldsymbol{U})} \right] = \mathbb{E}\left[\max_{1 \le j \le d} \left(\frac{\gamma R_j}{\sigma} \right)^{1/\gamma} \right] = \left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \mathbb{E}\left[R_d^{1/\gamma} \right].$$

The distribution of R_d is called generalized Erlang if $\lambda_i \neq \lambda_j$ for all $i \neq j$ (Neuts, 1974), and, letting f_{R_d} denote its density we get

$$\begin{aligned} \left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \mathbb{E}\left[R_d^{1/\gamma} \right] &= \left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \int_0^\infty r^{1/\gamma} f_{R_d}(r)\, \mathrm{d}r \\ &= \left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \sum_{i=1}^{d} \lambda_i \left(\prod_{j=1, j \neq i}^{d} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \int_0^\infty r^{1/\gamma} e^{-r\lambda_i}\, \mathrm{d}r \\ &= \left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \Gamma\left(\frac{1}{\gamma} + 1 \right) \sum_{i=1}^{d} \lambda_i^{-1/\gamma} \left(\prod_{j=1, j \neq i}^{d} \frac{\lambda_j}{\lambda_j - \lambda_i} \right). \end{aligned}$$

If $\lambda_1 = \ldots = \lambda_d$, then R_d follows an Erlang distribution. By (3.6), the density of $\boldsymbol{X}$ becomes, for $x_d > \ldots > x_1 > -\sigma/\gamma$,

$$\begin{aligned} h_{\boldsymbol{R}}(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\gamma}) &= \mathbb{1}(x_d > 0) \frac{\left(\prod_{j=1}^{d} \lambda_j \right) \int_0^\infty t^{d\gamma} \exp\left\{ -t^\gamma \sum_{j=1}^{d} (\lambda_j - \lambda_{j+1})(x_j + \sigma/\gamma) \right\} \mathrm{d}t}{\left(\frac{\gamma}{\sigma} \right)^{1/\gamma} \mathbb{E}\left[R_d^{1/\gamma} \right]} \\ &= \frac{\mathbb{1}(x_d > 0) \left(\prod_{j=1}^{d} \lambda_j \right) \left(\frac{\gamma}{\sigma} \right)^{-1/\gamma} \Gamma\left(d + \frac{1}{\gamma} \right) / \Gamma\left(\frac{1}{\gamma} \right)}{\left(\sum_{j=1}^{d} (\lambda_j - \lambda_{j+1}) x_j + (\sigma/\gamma)\lambda_1 \right)^{d+1/\gamma} \sum_{i=1}^{d} \lambda_i^{-1/\gamma} \left(\prod_{j=1, j \neq i}^{d} \frac{\lambda_j}{\lambda_j - \lambda_i} \right)}. \end{aligned}$$

5 Likelihood-based inference

Working within a likelihood-based framework for inference allows many benefits. Firstly, comparison of nested models can be done using likelihood ratio tests. This is important as the number of parameters can quickly grow large if margins and dependence are fitted simultaneously, thus we can test for simplifications in a principled manner. Secondly, incorporation of covariate effects is straightforward in principle. For univariate peaks over thresholds, such ideas were introduced by Davison and Smith (1990), but nonstationarity in dependence structure estimation has received comparatively little attention; an exception is Huser and Genton (2016) in the context of max-stable processes. Thirdly, such likelihoods could also be exploited for a Bayesian approach to inference if desired.

5.1 Censored likelihood

The density (3.4) is the basic ingredient in a likelihood, however, we will use (3.4) as a contribution only when all components of the observed translated vector $\boldsymbol{Y} - \boldsymbol{u}$ are "large", in the sense of exceeding a threshold $\boldsymbol{v}$, with $\boldsymbol{v} \leq \boldsymbol{0}$. The reasoning for this is twofold:

1. For $\gamma_j > 0$, the lower endpoint of the multivariate GP distribution is $-\sigma_j/\gamma_j$. We would like to avoid the situation where small values of a component can affect the fit of the upper tail of that component too strongly.

2. Without censoring, bias in the estimation of parameters controlling the dependence can be larger than for censored estimation, see Huser et al. (2016).

Let $D = \{1, \ldots, d\}$ and $C \subset D$ be the subset of indices denoting which components of $\boldsymbol{Y} - \boldsymbol{u}$ fall below the corresponding component of $\boldsymbol{v}$, i.e., $Y_j - u_j \leq v_j$ for $j \in C$, and $Y_j - u_j > v_j$ for $j \in D \setminus C$, with at least one such $Y_j > u_j$. For each realization of $\boldsymbol{Y}$, we use the likelihood contribution

$$h_C(\boldsymbol{y}_{D\setminus C} - \boldsymbol{u}_{D\setminus C}, \boldsymbol{v}_C; \boldsymbol{\gamma}, \boldsymbol{\sigma}) = \int_{\times_{j\in C}(-\infty, u_j+v_j]} h(\boldsymbol{y} - \boldsymbol{u}; \boldsymbol{\gamma}, \boldsymbol{\sigma})\, \mathrm{d}\boldsymbol{y}_C, \qquad \boldsymbol{y}_C = (y_j)_{j\in C}, \quad (5.1)$$

which is equal to (3.4) with $\boldsymbol{x} = \boldsymbol{y} - \boldsymbol{u}$ if C is empty. Appendix A contains forms of censored likelihood contributions for the models included in Section 3. Thus, for n independent observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ of $\boldsymbol{Y} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$, the form of the censored likelihood function to be optimized is

$$L(\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} h_{C_i}(\boldsymbol{y}_{i,D\setminus C} - \boldsymbol{u}_{D\setminus C}, \boldsymbol{v}_C; \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}), \quad (5.2)$$

where C_i denotes the censoring subset for $\boldsymbol{y}_i$, which may be empty, and $\boldsymbol{\theta}$ represents parameters related to the model that we assumed for the generator.

5.2 Model choice

When fitting multivariate GP distributions to data on the observed scale we have a large variety of potential models and parameterizations. For non-nested models, Akaike's Information Criterion (AIC = $-2 \times$ log-likelihood + $2 \times$ number of parameters) can be used to select a model with a good balance between parsimony and goodness-of-fit. When looking at nested models, e.g., to test for simplifications in parameterization, we can use likelihood ratio tests. Because of the many possibilities for model fitting, to reduce the computational burden we propose the following model-fitting strategy, that we will employ in Section 6.1.

(i) Standardize the data to common exponential margins, $\boldsymbol{Y}_E$, using the rank transformation (i.e., the probability integral transform using the empirical distribution function);

(ii) select a multivariate threshold, denoted $\boldsymbol{u}$ on the scale of the observations, and $\boldsymbol{u}_E$ on the exponential scale, using the method of Section 5.3;

(iii) fit the most complicated standard form model within each class (i.e., maximum number of possible parameters) to the standardized data $\boldsymbol{Y}_E - \boldsymbol{u}_E \mid \boldsymbol{Y}_E \not\leq \boldsymbol{u}_E$;

(iv) select as the standard form model class the one which produces the best fit to the standardized data, in the sense of smallest AIC;

(v) use likelihood ratio tests to test for simplification of models within the selected standard form class, and select a final standard form model;

(vi) fit the GP margins simultaneously with this standard form model, to $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ by maximizing (5.2);

(vii) test for simplifications in the marginal parameterization.

Although this strategy is not guaranteed to result in a final GP model that is globally optimal, in the sense of minimizing an information criterion such as AIC, it should still result in a sensible model whilst avoiding enumeration and fitting of an unfeasibly large number of possibilities. The validity of the fit of the selected model can be assessed via diagnostics.

5.3 Threshold selection and model diagnostics

An important issue that pervades extreme value statistics — in all dimensions — is the selection of a threshold above which the limit model provides an adequate approximation of the distribution of threshold exceedances. Here this amounts to "how can we select a vector $\boldsymbol{u}$ such that $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ is well-approximated by a GP distribution?". There are two considerations to take into account: $Y_j - u_j \mid Y_j > u_j$ should be well-approximated by a univariate GP distribution, for $j = 1, \ldots, d$, and the dependence structure of $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ should be well-approximated by that of a multivariate GP distribution. Marginal threshold selection has a large body of literature devoted to it; see Scarrott and MacDonald (2012) and Caeiro and Gomes (2016) for recent reviews. Threshold selection for dependence models is a much less well studied problem, though recent literature that begins to address this topic includes Lee et al. (2015) who consider threshold selection via Bayesian measures of surprise, and Wadsworth (2016) who examines how to make better use of so-called parameter stability plots, offering a method that can be employed on any parameter, pertaining to the margins or dependence structure.

Here we propose using the "constant conditional exceedance" property of multivariate GP distributions outlined in Section 2 to guide threshold selection for the dependence structure. If $\boldsymbol{Y} \sim F$ and $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u} \sim H$, then on the region $q > \max_j F_j^{-1}(u_j)$, we have $\chi_{1:d}(q) \approx \chi_H(q')$ with $q' = \{q - F(\boldsymbol{u})\}/\{1 - F(\boldsymbol{u})\}$. A consequence of this is that $\chi_{1:d}(q)$ should be constant on the region $\boldsymbol{Y} > \boldsymbol{u}$, if $\boldsymbol{u}$ represents a sufficiently high dependence threshold. Hence if we identify $q^* = \inf\{0 < \tilde{q} < 1 : \chi_{1:d}(q) \equiv \chi \;\forall\; q > \tilde{q}\}$, then $\boldsymbol{u} = (F_1^{-1}(q^*), \ldots, F_d^{-1}(q^*))$ should provide an adequate threshold for the dependence structure. Once suitable thresholds have been identified for margins, $\boldsymbol{u}_{\mathrm{m}}$, and dependence, $\boldsymbol{u}_{\mathrm{d}}$, then a threshold vector which is suitable for the entire multivariate model is $\boldsymbol{u} = \max(\boldsymbol{u}_{\mathrm{m}}, \boldsymbol{u}_{\mathrm{d}})$.

To illustrate this idea, let $\widehat{\chi}_{1:d}(q)$ denote the empirical version of $\chi_{1:d}(q)$, defined by

$$\widehat{\chi}_{1:d}(q) := \frac{\sum_{i=1}^{n} \mathbb{1}\left\{\widehat{F}_1(Y_1) > q, \ldots, \widehat{F}_d(Y_d) > q\right\}}{n(1-q)}, \qquad q \in [0, 1), \tag{5.3}$$

where $\widehat{F}_1, \ldots, \widehat{F}_d$ represent the empirical distribution functions. Figure 2 displays $\widehat{\chi}_{1:d}(q)$ for data in the *domain of attraction* of a GP distribution. From the right-hand panel it appears that the estimate looks reasonably constant after $q = 0.8$. In addition to aiding threshold

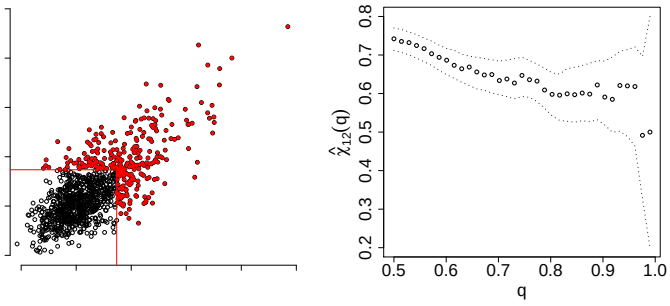

Figure 2: Illustration of dependence threshold selection using $\chi_{1:d}(q)$. Left: two-dimensional data in the domain of attraction of a multivariate GP distribution; right: corresponding $\widehat{\chi}(q)$. This appears constant for $q > 0.8$, and the data satisfying this are given as filled points on the left. Approximate 95% pointwise confidence intervals are obtained by bootstrapping from the sample $\{\boldsymbol{Y}_i : i = 1, \ldots, n\}$

selection, such plots also aid determination of whether the multivariate GP distribution forms a suitable model for the data: if it appears that $\chi_{1:d}(q) \downarrow 0$ as $q \uparrow 1$ then we should explore further whether it is reasonable to assume that the data are asymptotically dependent.

Having identified a multivariate GP model to fit, and a threshold above which to fit it, a key concern is to establish whether the goodness-of-fit is adequate. For the dependence structure, one diagnostic comes from comparing $\widehat{\chi}_{1:d}(q)$ for $q \to 1$ to $\chi_{1:d}$, which for models $h_{\boldsymbol{T}}$ in (3.2) has the form

$$\chi_{1:d} = \mathbb{E}\left[\min_{1 \le j \le d} \frac{e^{T_j - \max(\boldsymbol{T})}}{\mathbb{E}(e^{T_j - \max(\boldsymbol{T})})}\right],$$

whilst for models $h_{\boldsymbol{U}}$ in (3.3) we get the form

$$\chi_{1:d} = \mathbb{E}\left[\min_{1 \le j \le d} \frac{e^{U_j}}{\mathbb{E}(e^{U_j})}\right].$$

The formula of $\chi_{1:d}$ stemming from $h_{\boldsymbol{R}}$ models then follows through equation (3.5). In some cases these expressions may be evaluated analytically, but they can always be evaluated with arbitrary accuracy by simulation; see Rootzén et al. (2016). A further diagnostic could exploit the fact that $\mathbb{P}[X_j > 0] = \mathbb{E}[e^{T_j - \max(\boldsymbol{T})}] = \mathbb{E}[e^{U_j}]/\mathbb{E}[e^{\max(\boldsymbol{U})}]$ (see Section 3.1). Thus one would compare $\mathbb{P}[Y_j > u_j \mid \boldsymbol{Y} \not\leq \boldsymbol{u}]$ with the relevant model-based probability. In the event that the u_j are equal marginal quantiles, $\mathbb{P}[Y_j > u_j \mid \boldsymbol{Y} \not\leq \boldsymbol{u}] = \mathbb{P}[Y_j > u_j]/\mathbb{P}[\boldsymbol{Y} \not\leq \boldsymbol{u}]$ is the same for each margin.

Equation (2.5) suggests a model-free diagnostic of whether a multivariate GP model may be appropriate. To exploit this, one defines a set of interest A, and compares the number of points of $\boldsymbol{Y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ that lie in A to t times the number of points of $(\boldsymbol{Y} - \boldsymbol{u} - \boldsymbol{w}_t)/t^{\gamma} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}$ lying in A for various choices of $t > 1$. According to (2.5), the ratio of these numbers should be approximately equal to 1. We exploit this diagnostic in Section 6.2. Note that setting $A = \{\boldsymbol{x} : \boldsymbol{x} > \boldsymbol{0}\}$ is equivalent to computing χ_H with $H_1, \ldots, H_d$ replaced by $H_1^+, \ldots, H_d^+$.

Finally, in the event that the margins can be modelled with identical shape parameters, one can test property (2.7) by examining the adequacy of the implied univariate GP distribution from a multivariate fit. This will be used in Section 6.1 below.

6 Applications

6.1 UK bank returns

We examine the weekly negative raw returns on the prices of the stocks from four large banks in the UK: HSBC (H), Lloyds (L), RBS (R) and Barclays (B). Data were downloaded from Yahoo Finance. Specifically, letting $Z_{j,t}$, $j \in \{H, L, R, B\}$, denote the closing stock price (adjusted for stock splits and dividends) in week t for bank j, the data we examine are $Y_{j,t} = 1 - Z_{j,t}/Z_{j,t-1}$, so that large positive values of $Y_{j,t}$ correspond to large relative losses for that index. The observation period runs from 29 October 2007 to 17 October 2016, with a total of $n = 470$ datapoints.

Figure 3 displays pairwise plots of the negative returns. There is evidence of strong extremal dependence from these plots, as the largest value of Y_L, Y_R, Y_B occurs simultaneously, with positive association amongst other large values. The largest value of Y_H occurs at a different time, but again there is positive association between other large values. Figure 4 shows an estimate of the function $\chi_{HLRB}(q)$ for these data. As is common in practice the value of $\widehat{\chi}_{HLRB}(q)$ generally decreases as q increases, but is plausibly stable and constant from slightly above $q = 0.8$. Consequently, we proceed with fitting a GP distribution.

Ultimately, we wish to fit a parametric GP model to the raw threshold excesses $\{\boldsymbol{Y}_t - \boldsymbol{u} : \boldsymbol{Y}_t \not\leq \boldsymbol{u}\}$. In view of the large variety of potential models and parameterizations, we use the model selection strategy detailed in Section 5.2. Throughout we use censored likelihood with $\boldsymbol{v} = \boldsymbol{0}$.

We begin by selecting a threshold and finding a standard form model for the marginally transformed data. On the basis of Figure 4, the 0.83 marginal quantile is used as the threshold in each margin; there are a total of 149 observations with at least one exceedance of the corresponding marginal threshold.

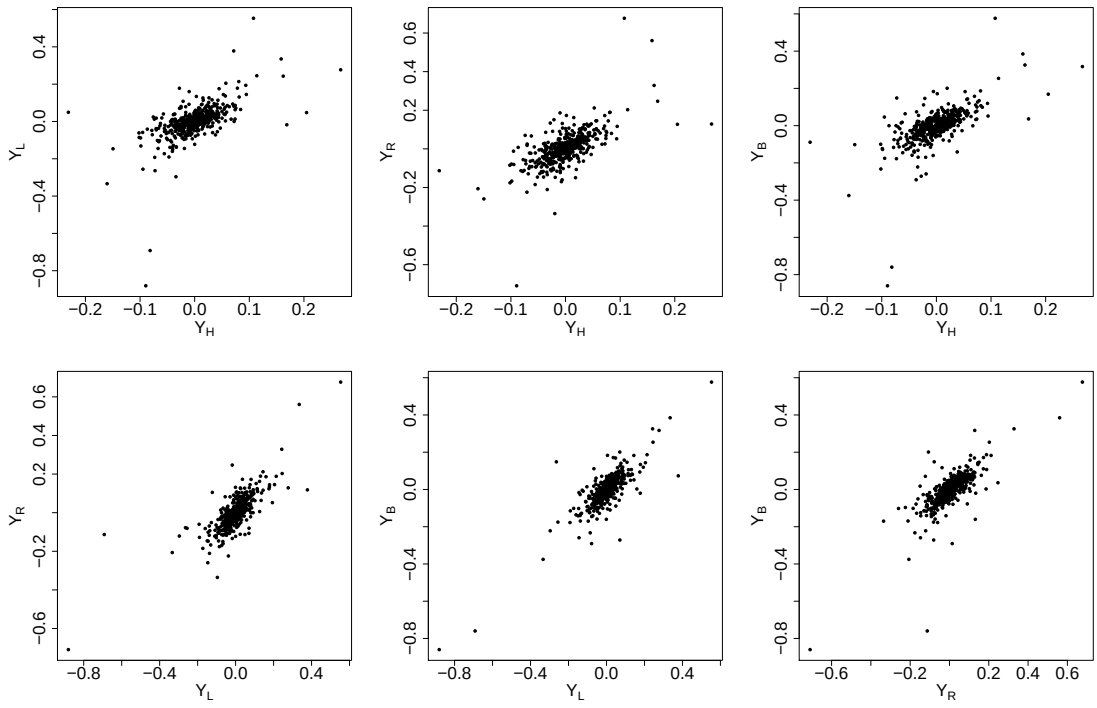

Figure 3: Pairwise scatterplots of the negative weekly returns of the stock prices of four UK banks: HSBC (H), Lloyds (L), RBS (R) and Barclays (B), from 29 October 2007 to 17 October 2016.

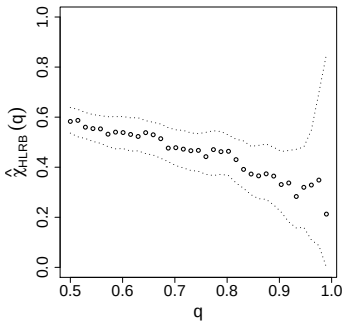

Figure 4: Negative UK bank returns: estimate of $\chi_{HLRB}(q)$ with HSBC (H), Lloyds (L), RBS (R) and Barclays (B). Approximate 95% pointwise confidence intervals are obtained by bootstrapping from $\{\boldsymbol{Y}_t : t = 1, \ldots, n\}$.

We fit five dependence models from Section 4, specifically the models with densities (4.1), (4.2), (4.3), (4.4) and (4.5), to the standardized data. The smallest AIC is given by model (4.1), i.e., where $f_{\boldsymbol{T}}$ is the density of independent Gumbel random variables. We therefore select this class and proceed with item (v) of the procedure in Section 5.2 to test for simplifications within this class. The results are summarized in Table 1, indicating that in fact a model with a single parameter is adequate, since both possible sequences of likelihood ratio tests between nested models lead to M4 when adopting a 5% significance level. This is a useful simplification, since with relatively few data points it places us in a better position for fitting a full GP distribution.

Table 1: Negative UK bank returns: parameterizations of model (4.1) for the standardized data and their maximized log likelihoods.

Model	Parameters	Number	Maximized log-likelihood
M1	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \lambda_1, \lambda_2, \lambda_3$	7	-917.0
M2	$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	4	-918.2
M3	$\alpha, \lambda_1, \lambda_2, \lambda_3$	4	-920.8
M4	α	1	-921.0

Finally we use the selected dependence model to fit a full GP distribution, and test for simplifications in the marginal parameterization. Marginal parameter stability plots suggest that the 0.83 quantile is adequate for these purposes, which is also supported by diagnostics from the fitted model (Appendix B, Figure 12). At a 5% significance level, a likelihood ratio test for the hypothesis of $\gamma_H = \gamma_L = \gamma_R = \gamma_B$ provides no evidence to reject the null hypothesis, whereas a further test for $\sigma_H = \sigma_L = \sigma_R = \sigma_B$ is rejected. The maximum likelihood estimates of parameters from the final model are displayed in Table 2.

To scrutinize the fit of the model, we examine marginal, dependence, and joint diagnostics. Quantile-quantile (QQ) plots for each of the univariate GP distributions implied for $Y_{t,j} - u_j \mid Y_{t,j} > u_j$ are displayed in Figure 12 of Appendix B, indicating reasonable fits in each

Table 2: Negative UK bank returns: maximum likelihood estimates (MLE) and standard errors (SE) of parameters from the final model for the original data.

	α	σ_H	σ_L	σ_R	σ_B	γ
MLE	1.29	0.020	0.041	0.038	0.035	0.43
SE	0.14	0.0026	0.0053	0.0052	0.0049	0.082

case. Estimates of the pairwise $\chi_{ij}(q)$, $i \neq j \in \{H, L, R, B\}$, are plotted in Figure 5, with the corresponding fitted value and threshold indicated; tripletwise plots were also examined and showed similarly good agreement. Since the model has a single parameter, all pairs are exchangeable and hence have the same fitted value of χ for any fixed dimension. The fitted value for the quadruple χ_{HLRB} is 0.40, which can be seen on Figure 4 to be appropriate.

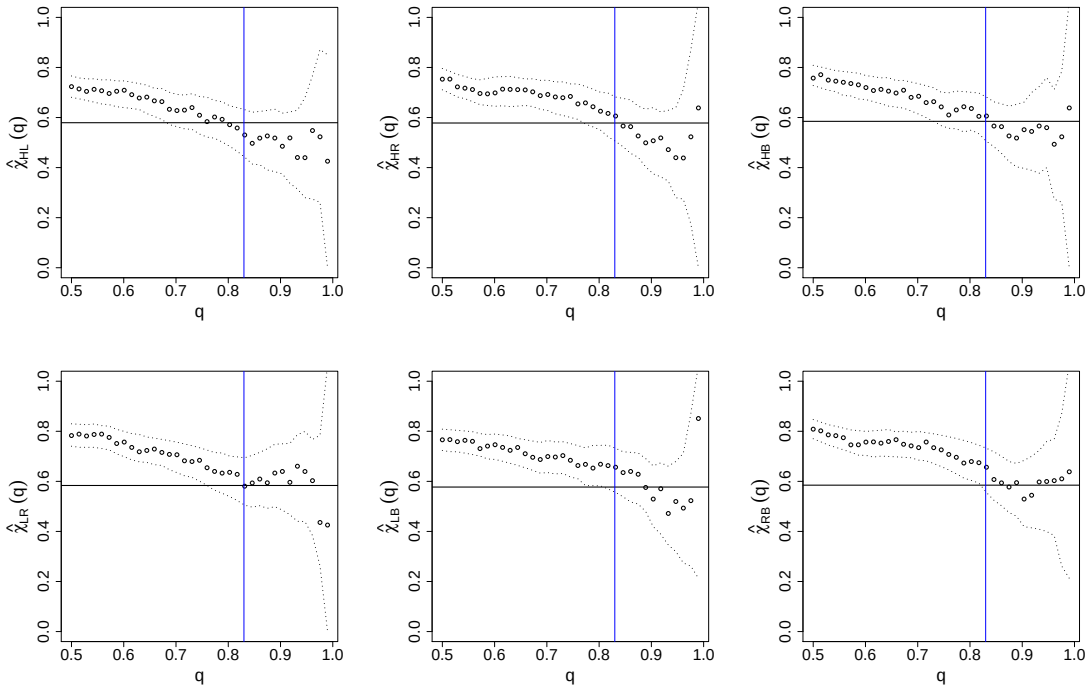

Figure 5: Negative UK bank returns: estimates of pairwise $\chi_{ij}(q)$ with fitted pairwise χ_{ij} (horizontal line), for HSBC (H), Lloyds (L), RBS (R) and Barclays (B). Clockwise from top left: χ_{HL}, χ_{HR}, χ_{HB}, χ_{RB}, χ_{LB}, χ_{LR}. The vertical line represents the threshold used. Approximate 95% pointwise confidence intervals are obtained by bootstrapping from $\{\boldsymbol{Y}_t : t = 1, \ldots, n\}$.

Other model diagnostics are available either by theoretical properties of the GP distribution, or by simulation from it (Rootzén et al., 2016); recall Section 5.3. One useful property outlined in Section 2 is that of sum-stability under shape constraints, given in equation (2.7). As a diagnostic, we fit a univariate GP distribution to

$$\sum_{j \in \{H,L,R,B\}} (Y_{t,j} - u_j) \,\Big|\, \sum_{j \in \{H,L,R,B\}} (Y_{t,j} - u_j) > 0, \tag{6.1}$$

with scale parameter estimate (standard error) obtained as 0.10 (0.021), and shape parameter estimate 0.45 (0.17). QQ plots suggest that the fit is good; see Figure 13 of Appendix B. For comparison, we have $\sum_{j \in \{H,L,R,B\}} \hat{\sigma}_j = 0.13$ with standard error 0.014 obtained using the delta method, whilst the maximized univariate GP log-likelihood is 63.5, and that for the parameters obtained via the multivariate fit is 62.2, showing that the theory holds well.

We now use the fitted multivariate GP model, combined with property (2.7), to investigate assessing portfolio risk, noting that weighted sums of raw stock returns correspond to portfolio performance. We consider two commonly-used risk measures, *Value at Risk* (VaR) and *Expected Shortfall* (ES) to describe the losses one could expect to see for different portfolio configurations containing these stocks. VaR is an estimate of the loss that is exceeded with probability p over a specified time horizon; since we have weekly returns, we consider a time horizon of one week. Thus, if the conditional distribution of $\sum_j a_j(Y_{t,j} - u_j)$ given the event $\sum_j a_j(Y_{t,j} - u_j) > 0$ is

$\text{GP}(\sum_j a_j\sigma_j, \gamma)$, then

$$\text{VaR}(p) = \sum_j a_j u_j + \frac{\sum_j a_j\sigma_j}{\gamma}\left\{\left(\frac{\phi}{p}\right)^{\gamma} - 1\right\}, \tag{6.2}$$

where $\phi = \mathbb{P}[\sum_j a_j(Y_{t,j} - u_j) > 0]$ and $0 < p < \phi$, so that (6.2) is the unconditional $1-p$ quantile of $\sum_j a_j Y_{t,j}$. The probability ϕ is estimated by maximum likelihood using the binomial assumption $\sum_t \mathbb{1}\{\sum_j a_j(Y_{t,j} - u_j) > 0\} \sim \text{Bin}(n, \phi)$, and in the univariate model it is orthogonal to the GP parameters of the conditional excess distribution. In the multivariate model, one can in theory express

$$\begin{aligned}\mathbb{P}\left[\textstyle\sum_j a_j(Y_{t,j} - u_j) > 0\right] &= \mathbb{P}\left[\textstyle\sum_j a_j(Y_{t,j} - u_j) > 0 \mid \boldsymbol{Y}_t \not\leq \boldsymbol{u}\right]\mathbb{P}[\boldsymbol{Y}_t \not\leq \boldsymbol{u}] \\ &= p(\boldsymbol{\theta})\,\tilde{\phi},\end{aligned} \tag{6.3}$$

where $p(\boldsymbol{\theta})$ is an expression involving the parameters of the multivariate GP model, and $\tilde{\phi}$ is the proportion of points for which $\boldsymbol{Y}_t \not\leq \boldsymbol{u}$. The expression $p(\boldsymbol{\theta})$ is not tractable here, thus we continue to estimate ϕ as the binomial maximum likelihood estimate, and as a working assumption treat it as orthogonal to the other parameters. However, an estimate of $p(\boldsymbol{\theta})$ could be obtained by simulation given known or estimated parameters $\boldsymbol{\theta}$, and the utility of this will be demonstrated in Figure 7.

The expected shortfall is defined as the expected loss given that a particular VaR threshold has been exceeded. Under the GP model, and provided $\gamma < 1$, it is given by

$$\begin{aligned}\text{ES}(p) &= \mathbb{E}\left[\textstyle\sum_j a_j Y_{t,j} \mid \sum_j a_j Y_{t,j} > \text{VaR}(p)\right] \\ &= \text{VaR}(p) + \frac{\sum_j a_j\sigma_j + \gamma\left[\text{VaR}(p) - \sum_j a_j u_j\right]}{1-\gamma}.\end{aligned} \tag{6.4}$$

Asymptotic theory suggests that a univariate GP model fit directly to $\sum_j a_j(Y_{t,j} - u_j)$ or the implied $\text{GP}(\sum_j a_j\sigma_j, \gamma)$ model obtained from the multivariate fit could be used. An advantage of using the $\text{GP}(\sum_j a_j\sigma_j, \gamma)$ model derived from the multivariate fit is reduced uncertainty, combined with consistent estimates across different portfolio combinations.

Figures 6 displays VaR curves and delta-method 95% pointwise confidence intervals for two different weight combinations and for both the univariate fits and implied multivariate fits, whilst Figure 14 of Appendix B displays the corresponding ES curves. For both sets of plots, empirical counterparts are shown for model validation. For VaR (Figure 6), the univariate fit is closer in the body whilst the multivariate fit apprears to be closer to the data in the tails. The reduction in uncertainty is clear and potentially quite useful for smaller p, though we note that at such levels the uncertainty is likely to be asymmetric, and profile likelihood-based confidence intervals would perhaps be more useful; however, we do not pursue this here. For ES (Appendix B, Figure 14) the univariate fit estimates smaller values than the multivariate fit in each case, and seems to reflect the observed data better. However, the empirical ES values fall within the 95% (approximate) confidence intervals obtained via the multivariate model, suggesting that the model is still consistent with the data for this quantity. For the univariate fit the clear shortcomings of delta-method confidence intervals can be seen with the coverage of negative values.

Finally, Figure 7 illustrates how use of a multivariate model provides more consistent estimates of VaR across different portfolio combinations compared to use of multiple univariate models. To produce the figures, we suppose that $\sum_j a_j = 100$ represents the total amount available to invest. The value $a_H = 10$ is fixed, with other weights varying, but with each $a_j \geq 1$. Two estimates making use of the multivariate model are provided: one for which a model-based estimate of $p(\boldsymbol{\theta})$ from (6.3) is used (with estimation based on 100 000 draws from

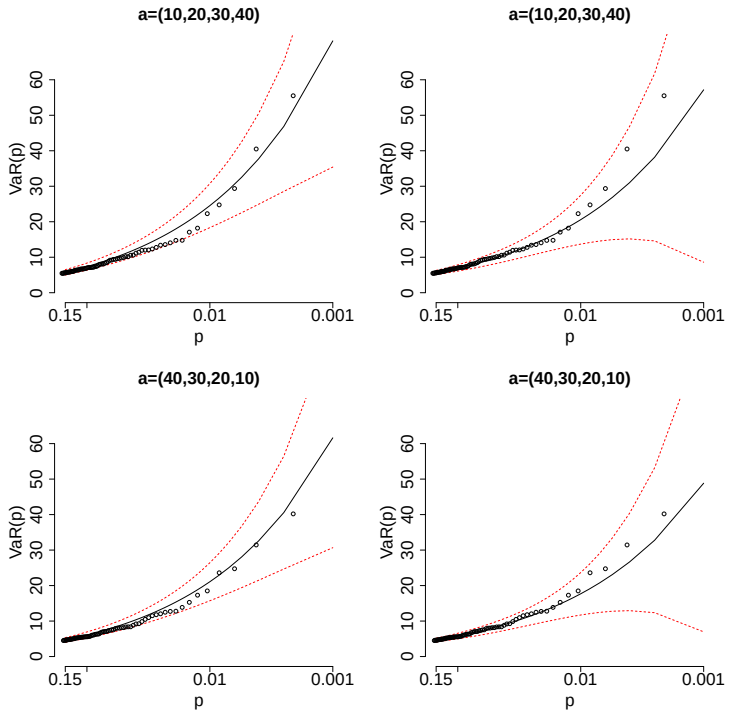

Figure 6: VaR estimates and pointwise 95% delta-method confidence intervals for portfolio losses based on the weights given as percentages invested in HSBC, Lloyds, RBS and Barclays as in the figure title. Estimates based on the multivariate GP fit are on the left; estimates based on the univariate fit are on the right.

the fitted model), and one where the empirical estimate of ϕ is used, as in Figure 6 and in Figure 14 of Appendix B. We observe that both sets of multivariate estimates suggest much more consistent behaviour across portfolio combinations than the use of univariate fits. In particular, the behaviour is very smooth once a model-based estimate for $p(\boldsymbol{\theta})$ is included.

6.2 Landslides

Rainfall can cause ground water pressure build-up which, if very high, can trigger a landslide. The cause can be either short periods with extreme rain intensities, or longer periods of more moderate, but still high rain intensities. Guzzetti et al. (2007) consolidate many previous studies and find that the majority of rainfall events that cause landslides have a duration between one hour and three days (Guzzetti et al., 2007, Figure 6A). Moreover, they propose threshold functions such that rainfall below these thresholds are unlikely to cause landslides. These functions link duration in hours, D, with total rainfall in millimeters, P. For highland climates in Europe this threshold function is

$$P = 7.56 \times D^{0.52}. \tag{6.5}$$

Thus, e.g., a one-hour rainfall amount below 7.5 mm, a one-day amount below 39.5 mm, a two-day amount below 56.6 mm, or a three-day amount below 69.9 mm are all unlikely to cause a landslide.

We will use a long time series of daily precipitation amounts $P_1, \ldots, P_N$ collected by the Abisko Scientific Research Station[2] in northern Sweden between January 1, 1913, and December 31, 2014, to estimate a lower bound for the probability of the occurrence of rainfall events which exceed the threshold in (6.5), and which may hence lead to landslides.

[2]http://polar.se/en/abisko-naturvetenskapliga-station/vaderdata/

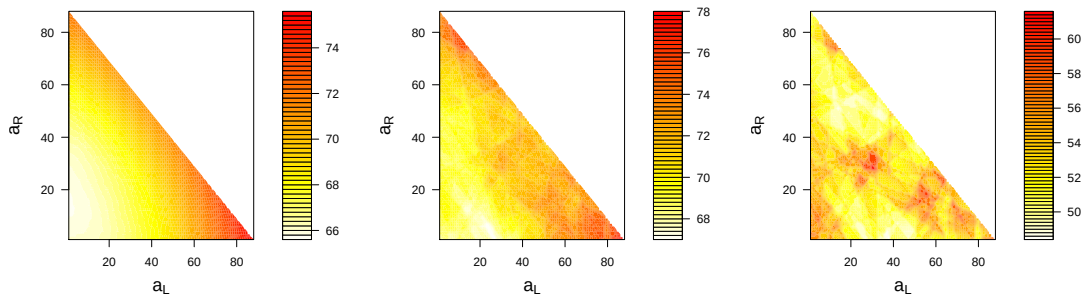

Figure 7: Maximum likelihood estimates of VaR(0.001) for $\sum_j a_j Y_{t,j}$ with $a_H = 10$ and $a_B = 90 - a_L - a_R$ for four UK banks, HSBC, Lloyds, RBS and Barclays. Left: from multivariate model including simulation to estimate $p(\boldsymbol{\theta})$ from (6.3); centre: from multivariate model using purely empirical estimate of ϕ; right: from univariate model fit to each combination separately. Note the different colour scales on each panel.

The total cost of landslides in Sweden is typically around SEK 200 million/year. There have been several landslides in the Abisko area in the past century, for instance in October 1959, August 1998, and July 2004 (Rapp and Strömquist, 1976; Jonasson and Nyberg, 1999; Beylich and Sandberg, 2005). The rain causing the landslides are clearly visible in the data, with 24.5 mm of rain on October 5, 1959, 21.0 mm of rain on August 24, 1998, and 61.9 mm of rain on July 21, 2004. The 2004 rain amount is well above the 1-day risk threshold, whereas the 1959 and 1998 rain amounts are below the 1-day threshold. The explanation may be that the durations of the latter two rain events were actually substantially shorter than 24 hours, and that the threshold in (6.5) was still exceeded. However, the 1-day data resolution is not sufficient to be able to verify this.

We wish to construct a dataset $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \in \mathbb{R}^3$, for $n < N$, whose components represent daily, two-day, and three-day extreme rainfall amounts respectively. We limit ourselves to $d = 3$ because of the findings in Guzzetti et al. (2007). Based on a mean residual life plot and parameter stability plots (not shown here) for the daily rainfall amounts $P_1, \ldots, P_N$, we choose the threshold $u = 12$, which corresponds roughly to the 99% quantile. Figure 8 shows the cumulative three-day precipitation amounts $P_i + P_{i+1} + P_{i+2}$ for $i \in \{1, \ldots, N-2\}$. The threshold u chosen above is used to extract clusters of data containing extreme episodes; because extreme rainfall occurs when either the daily, or the two-day, or the three-day precipitation amounts are extreme and these quantities are ordered, we consider three-day precipitation amounts exceeding u. The data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are then constructed as follows:

1. Let i correspond to the first sum $P_i + P_{i+1} + P_{i+2}$ which exceeds the threshold u and set $P_{(1)} = \max(P_i, P_{i+1}, P_{i+2})$.

2. Let the first cluster $C_{(1)}$ consist of $P_{(1)}$ plus the five values preceding it and the five values following it.

3. Let Y_{11} be the largest value in $C_{(1)}$, Y_{12} the largest sum of two consecutive non-zero values in $C_{(1)}$, and Y_{13} the largest sum of three consecutive non-zero values in $C_{(1)}$.

4. Find the second cluster $C_{(2)}$ and compute $\boldsymbol{Y}_2 = (Y_{21}, Y_{22}, Y_{23})$ in the same way, starting with the first observation after $C_{(1)}$.

Continuing this way, we obtain a dataset $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$, with $n = 580$.

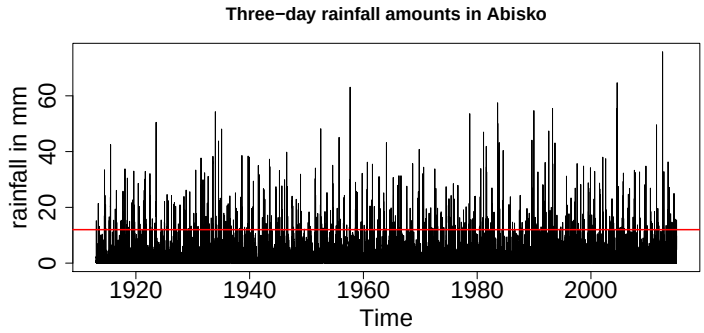

Figure 8: Precipitation data in Abisko: cumulative three-day precipitation amounts $P_i + P_{i+1} + P_{i+2}$ for $i \in \{1, \ldots, N-2\}$ with threshold $u = 12$ in red.

Time trend

A similar dataset has previously been analysed in Rudvik (2012), where a univariate generalized extreme value model with a linear trend in the location parameter was fitted to annual maxima, with the conclusion that there is no significant trend. We investigate the question whether there is a trend in the daily, two-day or three-day rainfall amounts by fitting a univariate GP distribution with a fixed shape parameter γ but a loglinear trend for the scale parameter, $\log \sigma(t) = a + bt$ for $t \in (0, 1]$, to the marginal components of the series $(\boldsymbol{Y}_i)_{i=1}^n$. To this end, we need to select marginal thresholds above which we fit the univariate GP distributions. For the first component, we take $u_1 = 12$ as found previously; for the second and third components, we take $u_2 = 13.5$ and $u_3 = 14$ respectively, based on inspection of parameter stability plots. For the first component, the time t corresponds to the indices $i \in \{1, \ldots, N\}$ for which $P_i > u_1$; for the second and third component, we use the time corresponding to $\max(P_i, P_{i+1})$ and $\max(P_i, P_{i+1}, P_{i+2})$.

In Table 3, we report the parameter estimates for the univariate GP fit above these thresholds. The final line shows the deviance, i.e., -2 times the difference in log-likelihood with respect to a model with $\sigma(t) \equiv \sigma$. We compare to the 95% quantile of a χ^2_1 distribution, given by 3.84. Likelihood ratio tests show that the absence of a linear trend in the logarithm of the scale parameter cannot be rejected.

We do not adopt any trend and Table 4 shows the result of fitting univariate GP distributions to the margins conditional on exceeding the previously mentioned thresholds. Observing that the estimated shape parameters are all around zero, we would like to test if the simpler model $\boldsymbol{\gamma} = \mathbf{0}$ would suffice and we find that such a restricted model cannot be rejected either. Moreover, the overlap in the confidence intervals of $\widehat{\sigma}_j$ suggest the plausibility of a model where $\boldsymbol{\sigma} = \sigma\mathbf{1}$. Note that a common σ and γ only implies that the marginal distributions are equal *conditional on* exceeding the threshold; it does not imply that the unconditional probabilities $\mathbb{P}[Y_j > u_j]$ are equal. In the following analysis, we will set $\boldsymbol{\sigma} = \sigma\mathbf{1}$ and $\boldsymbol{\gamma} = \gamma\mathbf{1}$, and we will fit both a model with $\gamma = 0$ and one with $\gamma > 0$.

Dependence structure

We wish to approximate the distribution of $\boldsymbol{Y}_i - \boldsymbol{u} \mid \boldsymbol{Y}_i \not\leq \boldsymbol{u}$ by a multivariate GP distribution on $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \mathbf{0}\}$; because the components of our data vectors are strictly increasing, the structured components model from Section 4.3 may be appropriate. We consider a model with $\gamma = 0$ and a more general model with $\gamma > 0$. Recall that we need to impose a restriction on the parameters for identifiability; we set $\lambda_1 = 1$ for both models. We estimate the parameters $(\lambda_2, \lambda_3, \sigma)$ in the first case and the parameters $(\lambda_2, \lambda_3, \sigma, \gamma)$ in the second case. We choose $\boldsymbol{u} = u\mathbf{1}$ with $u = 24$ since parameter estimates more or less stabilize for thresholds around this

Table 3: Precipitation data in Abisko: estimates of the parameters of marginal GP models with $\log \sigma(t) = a + bt$ and shape γ for thresholds $u = 12$, $u = 13.5$ and $u = 14$ respectively; standard errors in parentheses.

	$\boldsymbol{Y}_{i1}$	$\boldsymbol{Y}_{i2}$	$\boldsymbol{Y}_{i3}$
$\widehat{\gamma}$	-0.09 (0.06)	-0.05 (0.06)	-0.03 (0.06)
$\widehat{a}$	2.01 (0.12)	2.13 (0.11)	2.21 (0.11)
$\widehat{b}$	0.24 (0.21)	0.24 (0.19)	0.21 (0.17)
deviance	1.17	1.49	1.46

Table 4: Precipitation data in Abisko: estimates of the parameters of marginal GP models for thresholds $u = 12$, $u = 13.5$ and $u = 14$ respectively; standard errors in parentheses.

	$\boldsymbol{Y}_{i1}$	$\boldsymbol{Y}_{i2}$	$\boldsymbol{Y}_{i3}$
$\widehat{\gamma}$	-0.06 (0.05)	-0.02 (0.06)	-0.01 (0.05)
$\widehat{\sigma}$	8.26 (0.69)	9.34 (0.74)	9.96 (0.74)

value and we continue with the 142 datapoints whose third components exceed $u = 24$.

Table 5 shows the parameter estimates obtained from maximizing the censored likelihood in (5.2) with censoring threshold $\boldsymbol{v} = \boldsymbol{0}$. Again, the hypothesis of $\gamma = 0$ cannot be rejected. We see that the estimates of σ are somewhat higher than we saw in the marginal analysis, which is intuitively reasonable since the maximum likelihood estimators for γ and σ are negatively correlated and since $\widehat{\gamma}$ is positive for the second model.

We wish to estimate the probability of a future landslide using formula (6.5), i.e., we wish to calculate $\mathbb{P}[\boldsymbol{Y} \not\leq \boldsymbol{y}]$ where $\boldsymbol{y} = (39.5, 56.6, 69.9)$, which is such that $\boldsymbol{y} > \boldsymbol{u}$. In the first model, $\gamma = 0$, we can write

$$\begin{aligned}\mathbb{P}[\boldsymbol{Y} \not\leq \boldsymbol{y}] &= \mathbb{P}\left[\boldsymbol{Y} - \boldsymbol{u} \not\leq \boldsymbol{y} - \boldsymbol{u} \mid \boldsymbol{Y} \not\leq \boldsymbol{u}\right] \mathbb{P}[\boldsymbol{Y} \not\leq \boldsymbol{u}] = \{1 - H(\boldsymbol{y} - \boldsymbol{u}; \boldsymbol{\sigma}, \boldsymbol{\gamma})\} \, \mathbb{P}[Y_3 > u] \\ &= \left\{1 - H\left(\frac{\boldsymbol{y} - \boldsymbol{u}}{\boldsymbol{\sigma}}; \boldsymbol{1}, \boldsymbol{0}\right)\right\} \mathbb{P}[Y_3 > u]. \end{aligned} \tag{6.6}$$

The first term of (6.6) is calculated directly by integrating the density of the structured components model given in (4.7), where we plugged in the parameter estimates $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3, \widehat{\sigma})$ from the top row of Table 5. Since we are interested in the yearly exceedance probability, we estimate $\mathbb{P}[Y_3 > u]$ by counting the number of years with and without exceedances and computing the empirical probability. We obtain

$$\mathbb{P}[Y_1 > 39.5 \text{ or } Y_2 > 56.6 \text{ or } Y_3 > 69.9] \approx 0.062. \tag{6.7}$$

We find that the probability of rain amounts which could lead to a landslide in any given year is 0.062, which is higher than the result in Rudvik (2012). The data used in Rudvik (2012) is for the period 1913–2008 and based on daily, three-day and five-day precipitation amounts, which might explain the difference.

Goodness-of-fit

For a visual test we consider QQ-plots for each of the marginal conditional GP distributions $(Y_{ij} - u) \mid Y_{ij} > u$, with $u = 24$ as before. Figure 9 shows that for the model with $\gamma = 0$, the fit is less good for the first component due to the restriction $\boldsymbol{\sigma} = \sigma\boldsymbol{1}$. This restriction is used to ensure that the components are ordered. Plots for the model with $\gamma > 0$ (not shown) are very similar.

Table 5: Precipitation data in Abisko: parameter estimates for the structured components model with $u = 24$; standard errors in parentheses.

$u = 24$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\sigma}$	$\widehat{\gamma}$	Maximized log-likelihood
Model 1: $\gamma = 0$	1.00	0.84	1.08	10.17	0	-870.0
	–	(0.13)	(0.18)	(0.80)	–	
Model 2: $\gamma > 0$	1.00	0.83	1.06	9.14	0.11	-868.9
	–	(0.12)	(0.18)	(0.99)	(0.08)	

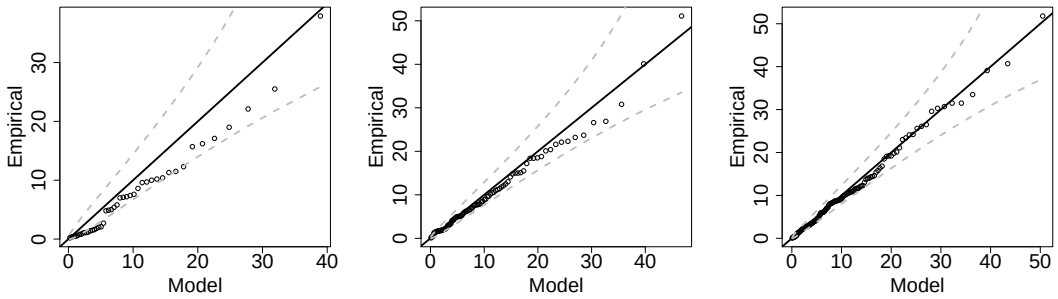

Figure 9: Precipitation data in Abisko: QQ-plots for the univariate GP distributions of the three variables Y_{i1}, Y_{i2} and Y_{i3} (left to right) for the model with $\gamma = 0$ with parameters implied by Table 5. The 95% pointwise confidence intervals are obtained by a transformation of the beta distributed order statistics of a uniform distribution.

For the dependence structure, we start by looking at the goodness-of-fit diagnostic suggested by equation (2.5) (see also Section 5.3). Since we set $\gamma = 0$, we look at the ratio

$$\frac{\mathbb{P}[\boldsymbol{Y} - \boldsymbol{u} \in A \mid \boldsymbol{y} \not\leq \boldsymbol{u}]}{t\,\mathbb{P}[\boldsymbol{Y} - \boldsymbol{u} - \boldsymbol{\sigma} \log t \in A \mid \boldsymbol{Y} \not\leq \boldsymbol{u}]}, \tag{6.8}$$

where $\boldsymbol{\sigma}$ denotes the vector of scale parameter estimates of the marginal GP models above $u = 24$. We consider the sets $A_j = \{\boldsymbol{x} \in \mathbb{R}^3 : x_j > 0\}$ for $j \in \{1, 2, 3\}$. Figure 10 shows that the result is satisfactory, i.e., expression (6.8) is near one, so that a MGP model seems appropriate for this dataset. The result for A_1 shows more variability than the other two since the number of exceedances in the first component is relatively low.

Next, we would like to compare the pairwise and trivariate χ (formulas can be found in Appendix B) to their empirical counterparts; see Section 5.3. Figure 11 shows the results, where the horizontal lines represent the model-based χ's based on $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\lambda}_3$ from the top row in Table 5. Again, the fit is satisfactory.

Finally, we consider the goodness-of-fit diagnostic which consists of comparing the model-based probabilities

$$\mathbb{P}[X_j > 0] = \frac{\mathbb{E}[e^{U_j}]}{\mathbb{E}[e^{\max(\boldsymbol{U})}]} = \frac{\sum_{i=1}^{j} \lambda_i^{-1}}{\sum_{i=1}^{d} \lambda_i^{-1}},$$

with the empirical probabilities $\mathbb{P}[Y_j > u_j \mid \boldsymbol{Y} \not\leq \boldsymbol{u}]$ for $j \in \{1, 2\}$. We find $\mathbb{P}[X_1 > 0] = 0.34$ (0.03) and $\mathbb{P}[X_2 > 0] = 0.63$ (0.03) using the values from the top row in Table 5; the standard errors are obtained using the delta method. For $\boldsymbol{u} = (24, 24, 24)$, the empirical probabilities are 0.32 and 0.69 respectively. Plots of the empirical probabilities for a range of thresholds $\boldsymbol{u} = u\mathbf{1}$ (not shown) confirm the chosen threshold value $u = 24$.

A more formal way to assess the goodness-of-fit of the dependence structure of this model is by calculating the test statistic presented in Einmahl et al. (2016, Corollary 2.5). The test

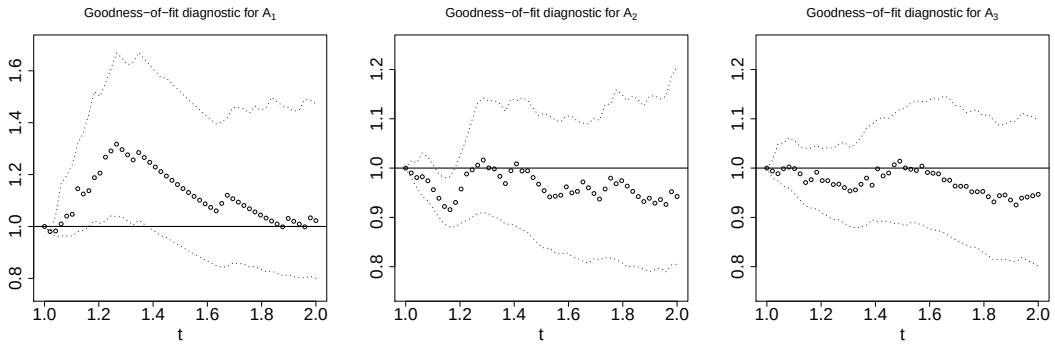

Figure 10: Abisko precipitation data: ratio (6.8) for $A \in \{A_1, A_2, A_3\}$ with $u = 24$. Approximate 95% pointwise confidence intervals are obtained by bootstrapping from $\{\boldsymbol{Y}_i : i = 1, \ldots, \boldsymbol{Y}_n\}$.

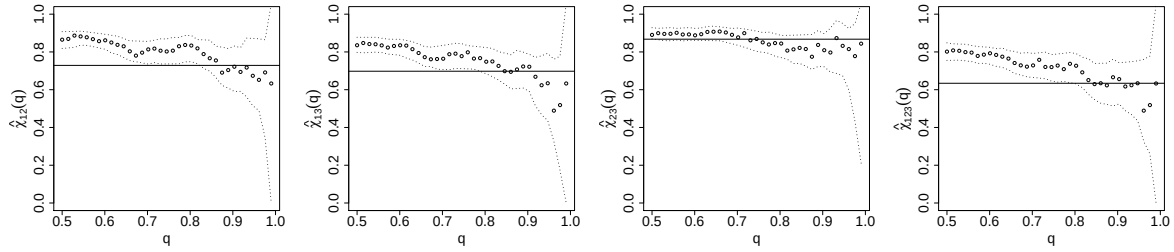

Figure 11: Precipitation data in Abisko: pairwise and trivariate $\widehat{\chi}(q)$ (dots) and model-based limiting χ (horizontal lines) for $u = 24$, with parameters implied by Table 5 for $\boldsymbol{\gamma} = \mathbf{0}$. Approximate 95% pointwise confidence intervals are obtained by bootstrapping from $\{\boldsymbol{Y}_i : i = 1, \ldots, \boldsymbol{Y}_n\}$.

statistic proposed there is based on the difference between the value of $(\chi_{12}, \chi_{13}, \chi_{23}, \chi_{123})$ and an empirical estimator thereof. It depends on a sequence $k \in \{1, \ldots, n\}$ where $k \to \infty$ and $k/n \to 0$ as $n \to \infty$, which represents the threshold value used: a low value of k corresponds to a high threshold. The test statistic converges to a chi-square distribution with 2 degrees of freedom; its 95% quantile is equal to 5.99. Computing the test statistic for $k \in \{50, 75, 100, 125, 150\}$, where we set again $\lambda_1 = 1$, we find the values 1.08, 4.48, 1.17, 5.42, and 0.99 respectively, so that we cannot reject the structured components model for any value of k.

7 Discussion

We have outlined several new models for multivariate GP distributions, along with their uncensored and censored likelihoods. The models were applied to two types of data: stock price returns, where advantages in terms of consideration of a portfolio were demonstrated; and rainfall data, where the context dictated that extremes of ordered cumulative data were the object of interest. Methods to select a multivariate threshold and diagnostics for the fitted models have been considered and demonstrated through the applications.

The threshold selection method suggested in Section 5.3 is relatively conservative in the sense that a multivariate GP model could hold with lower thresholds in some margins. Better ways to select a multivariate threshold, ideally incorporating marginal and dependence considerations simultaneously, remains an interesting problem.

A key issue that was highlighted in Section 1.2 was the idea of asymptotic (in)dependence, and how we can detect when multivariate GP distributions form an appropriate model. In particular, we have not dealt with the situation where the GP distribution places mass on a lower dimensional subspace, with the models outlined in Section 4 placing no mass on lower-dimensional subspaces of the d-dimensional space $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \mathbf{0}\}$. In principle, parameters assigning mass to such lower-dimensional spaces could be introduced and estimated through censored likelihood, although it seems likely that information on these parameters would be weak, and alternative ways of handling this situation are much needed.

If certain subsets of variables display asymptotic dependence, then, depending on questions of interest, it may be worth considering these separately. However, if all pairwise $\chi_{ij} = 0, j > i$, then no asymptotic dependence exists amongst any variables, and no multivariate GP model would be approporiate, since the limiting model places mass only on one-dimensional lines. In this context, methods from Ledford and Tawn (1997), Heffernan and Tawn (2004) or Wadsworth and Tawn (2013) may prove useful.

A Censored likelihoods

Here we detail forms of censored likelihoods for the models detailed in Section 3. For simplicity they are presented in standardized ($\boldsymbol{\sigma} = \mathbf{1}$, $\boldsymbol{\gamma} = \mathbf{0}$) form, i.e.,

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = \int_{\times_{j\in C}(-\infty, v_j]} h(\boldsymbol{x}; \mathbf{1}, \mathbf{0})\, \mathrm{d}\boldsymbol{x}_C, \tag{A.1}$$

for $v_j \leq 0$ and h corresponding to either $h_{\boldsymbol{T}}$ or $h_{\boldsymbol{U}}$. The generalized form of a censored likelihood is easily obtained from (A.1) as

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \boldsymbol{\sigma}, \boldsymbol{\gamma}) = h_C\left(\tfrac{1}{\boldsymbol{\gamma}}\log(1 + \boldsymbol{\gamma}\boldsymbol{x}_{D\setminus C}/\boldsymbol{\sigma}), \tfrac{1}{\boldsymbol{\gamma}}\log(1 + \boldsymbol{\gamma}\boldsymbol{v}_C/\boldsymbol{\sigma}); \mathbf{1}, \mathbf{0}\right) \prod_{j\in D\setminus C} \frac{1}{\sigma_j + \gamma_j x_j}.$$

The support for each density is $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \not\leq \mathbf{0}\}$, and we let $|C|$ denote the cardinality of the set C.

Generators with independent Gumbel components

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$.

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \int_0^\infty t^{-1} \prod_{j\in C} e^{-(te^{v_j-\beta_j})^{-\alpha_j}} \prod_{j\in D\setminus C} \alpha_j \left(te^{x_j-\beta_j}\right)^{-\alpha_j} e^{-(te^{x_j-\beta_j})^{-\alpha_j}}\, \mathrm{d}t.$$

If all α_j are equal to α:

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \frac{\alpha^{d-|C|-1}\Gamma(d-|C|)\prod_{j\in D\setminus C} e^{-\alpha(x_j-\beta_j)}}{\left(\sum_{j\in C} e^{-\alpha(v_j-\beta_j)} + \sum_{j\in D\setminus C} e^{-\alpha(x_j-\beta_j)}\right)^{d-|C|}}.$$

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$.

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = \frac{\int_0^\infty \prod_{j\in C} e^{-(te^{v_j-\beta_j})^{-\alpha_j}} \prod_{j\in D\setminus C} \alpha_j \left(te^{x_j-\beta_j}\right)^{-\alpha_j} e^{-(te^{x_j-\beta_j})^{-\alpha_j}}\, \mathrm{d}t}{\Gamma(1-1/\alpha)\left(\sum_{j=1}^d e^{\beta_j\alpha}\right)^{1/\alpha}}.$$

If all α_j are equal to α:

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = \frac{\alpha^{d-|C|-1}\Gamma(d-|C|-1/\alpha)\prod_{j\in D\setminus C} e^{-\alpha(x_j-\beta_j)}}{\Gamma(1-1/\alpha)\left(\sum_{j=1}^d e^{\beta_j\alpha}\right)^{1/\alpha}\left(\sum_{j\in C} e^{-\alpha(v_j-\beta_j)} + \sum_{j\in D\setminus C} e^{-\alpha(x_j-\beta_j)}\right)^{d-|C|-1/\alpha}}.$$

Generators with independent reverse exponential components

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$.

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \times \int_0^{e^{-\max_{j\in D\setminus C}(x_j+\beta_j)}} t^{-1} \prod_{j\in C} \min(te^{v_j+\beta_j}, 1)^{1/\alpha_j} \prod_{j\in D\setminus C} \frac{1}{\alpha_j} (te^{x_j+\beta_j})^{1/\alpha_j} \, \mathrm{d}t \quad \text{(A.2)}$$

To evaluate this, consider two cases: (i) $\max_{j\in C}(v_j+\beta_j) < \max_{j\in D\setminus C}(x_j+\beta_j)$; and (ii) let $v_{(1)}+\beta_{(1)} > \ldots > v_{(k)}+\beta_{(k)} > \max_{j\in D\setminus C}(x_j+\beta_j) > v_{(k+1)}+\beta_{(k+1)} > \ldots$ for $j \in C$ and $k \leq |C|$. In case (i), we have

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \frac{\prod_{j\in C} e^{(v_j+\beta_j)/\alpha_j} \prod_{j\in D\setminus C} (1/\alpha_j) e^{(\beta_j+x_j)/\alpha_j}}{\left(\sum_{j=1}^d 1/\alpha_j\right) (e^{\max_{j\in D\setminus C}(x_j+\beta_j)})^{\sum_{j=1}^d 1/\alpha_j}},$$

since on the range $t \in (0, e^{-\max_{j\in D\setminus C}(x_j+\beta_j)})$ the term $\prod_{j\in C} \min(te^{v_j+\beta_j}, 1)^{1/\alpha_j}$ in (A.2) is equal to $\prod_{j\in C}(te^{v_j+\beta_j})^{1/\alpha_j}$. In case (ii) this term will vary over that range, and one needs to split the integral as follows:

$$\int_0^{e^{-(v_{(1)}+\beta_{(1)})}} + \int_{e^{-(v_{(1)}+\beta_{(1)})}}^{e^{-(v_{(2)}+\beta_{(2)})}} + \cdots + \int_{e^{-(v_{(k)}+\beta_{(k)})}}^{e^{-\max_{j\in D\setminus C}(x_j+\beta_j)}}.$$

An evaluation of each integral yields that $e^{\max(\boldsymbol{x})} h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0})$ is equal to

$$\begin{aligned}
&\frac{\prod_{j\in C} e^{(v_j+\beta_j)/\alpha_j} \prod_{j\in D\setminus C} (1/\alpha_j) e^{(x_j+\beta_j)/\alpha_j}}{\left(\sum_{j=1}^d 1/\alpha_j\right) (e^{v_{(1)}+\beta_{(1)}})^{\sum_{j=1}^d 1/\alpha_j}} \\
&+ \sum_{i=1}^{k-1} \left\{ \frac{\prod_{j\in C_{(i)}} e^{(v_j+\beta_j)/\alpha_j} \prod_{j\in D\setminus C} (1/\alpha_j) e^{(x_j+\beta_j)/\alpha_j}}{\sum_{j\in C_{(i)}} 1/\alpha_j + \sum_{j\in D\setminus C} 1/\alpha_j} \right. \\
&\qquad \left. \times \left[\left(e^{v_{(i+1)}+\beta_{(i+1)}}\right)^{-\sum_{j\in C_{(i)}} 1/\alpha_j - \sum_{j\in D\setminus C} 1/\alpha_j} - \left(e^{v_{(i)}+\beta_{(i)}}\right)^{-\sum_{j\in C_{(i)}} 1/\alpha_j - \sum_{j\in D\setminus C} 1/\alpha_j} \right] \right\} \\
&+ \frac{\prod_{j\in C_{(k)}} e^{(v_j+\beta_j)/\alpha_j} \prod_{j\in D\setminus C} (1/\alpha_j) e^{(x_j+\beta_j)/\alpha_j}}{\sum_{j\in C_{(k)}} 1/\alpha_j + \sum_{j\in D\setminus C} 1/\alpha_j} \\
&\qquad \times \left[\left(e^{\max_{j\in D\setminus C}(x_j+\beta_j)}\right)^{-\sum_{j\in C_{(k)}} 1/\alpha_j - \sum_{j\in D\setminus C} 1/\alpha_j} - \left(e^{v_{(k)}+\beta_{(k)}}\right)^{-\sum_{j\in C_{(k)}} 1/\alpha_j - \sum_{j\in D\setminus C} 1/\alpha_j} \right]
\end{aligned}$$

with $C_{(i)} = C\setminus\{(1), \ldots, (i)\}$, i.e., with the indices corresponding to the i largest $v_j+\beta_j$ removed.

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. Is found similarly by noting the relation between these two approaches.

Generators with independent log-gamma components

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. Let F_j denote the cumulative distribution function of a Gamma$(\alpha_j, 1)$ random variable. Then

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = e^{-\max(\boldsymbol{x})} \prod_{j\in D\setminus C} \frac{e^{\alpha_j x_j}}{\Gamma(\alpha_j)} \int_0^\infty t^{-1} \left(\prod_{j\in D\setminus C} t^{\alpha_j} e^{-te^{x_j}} \right) \left(\prod_{j\in C} F_j(te^{v_j}) \right) \mathrm{d}t.$$

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. Defining $C_d = \int_{\Delta_{d-1}} \max(u_1, \ldots, u_d) \prod_{j=1}^{d} u_j^{\alpha_j - 1} \, \mathrm{d}u_1 \cdots \mathrm{d}u_{d-1}$, we have

$$h_C(\boldsymbol{x}_{D\setminus C}, \boldsymbol{v}_C; \mathbf{1}, \mathbf{0}) = \frac{C_d^{-1}}{\Gamma\left(\sum_{j=1}^{d} \alpha_j + 1\right)} \prod_{j \in D\setminus C} e^{\alpha_j x_j} \prod_{j \in C} \Gamma(\alpha_j) \int_0^\infty \left(\prod_{j \in D\setminus C} t^{\alpha_j} e^{-t e^{x_j}} \right) \left(\prod_{j \in C} F_j(t e^{v_j}) \right) \mathrm{d}t.$$

Generators with multivariate Gaussian components

Case $f_{\boldsymbol{T}} = f_{\boldsymbol{V}}$. For the Gaussian model, using abbreviated notation, the key observation is

$$\int_{\times_{j \in C}(-\infty, v_j]} h(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}_C = h_{D\setminus C}(\boldsymbol{x}_{D\setminus C}) \int_{\times_{j \in C}(-\infty, v_j]} \frac{h(\boldsymbol{x})}{h_{D\setminus C}(\boldsymbol{x}_{D\setminus C})} \, \mathrm{d}\boldsymbol{x}_C, \tag{A.3}$$

and the ratio in the second integral can be written as a proper Gaussian density function (with parameters that depend on $\boldsymbol{x}_{D\setminus C}$). The integrand is

$$\begin{aligned} \frac{h(\boldsymbol{x})}{h_{D\setminus C}(\boldsymbol{x}_{D\setminus C})} = \frac{e^{\max_{j \in D\setminus C} x_j}}{e^{\max(\boldsymbol{x})}} \frac{(\mathbf{1}^T \Sigma_{D\setminus C}^{-1} \mathbf{1})^{1/2}}{(\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{1/2}} \frac{|\Sigma_{D\setminus C}|^{1/2}}{|\Sigma|^{1/2}} \frac{(2\pi)^{(d-|C|-1)/2}}{(2\pi)^{(d-1)/2}} \\ \times \exp\left\{ -\frac{1}{2} \left[(\boldsymbol{x} - \boldsymbol{\beta})^T A (\boldsymbol{x} - \boldsymbol{\beta}) - (\boldsymbol{x}_{D\setminus C} - \boldsymbol{\beta}_{D\setminus C})^T A_{D\setminus C} (\boldsymbol{x}_{D\setminus C} - \boldsymbol{\beta}_{D\setminus C}) \right] \right\} \end{aligned} \tag{A.4}$$

with

$$A_{D\setminus C} = \Sigma_{D\setminus C}^{-1} - \frac{\Sigma_{D\setminus C}^{-1} \mathbf{1}\mathbf{1}^T \Sigma_{D\setminus C}^{-1}}{\mathbf{1}^T \Sigma_{D\setminus C}^{-1} \mathbf{1}}.$$

Firstly note that

$$e^{\max_{j \in D\setminus C} x_j} = e^{\max(\boldsymbol{x})}$$

as the maximum will not occur among the censored components. By a completion of the square it can be shown that expression (A.4) is in fact equal to

$$\frac{(2\pi)^{(|C|-d)/2}}{|\Gamma|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_C - \boldsymbol{\mu})^T \Gamma^{-1} (\boldsymbol{x}_C - \boldsymbol{\mu}) \right\}$$

with

$$\boldsymbol{\mu} = -(K_C^T A K_C)^{-1} K_C A K_{D\setminus C} (\boldsymbol{x}_{D\setminus C} - \boldsymbol{\beta}_{D\setminus C})$$

and

$$\Gamma = (K_C^T A K_C)^{-1},$$

where K_C (respectively $K_{D\setminus C}$) is a $d \times |C|$ [respectively $d \times (d - |C|)$] matrix of 0s with 1s in the $(C_k, l)^{\text{th}}$ position, for C_k the kth index in C and $k = 1, \ldots, |C|$, $l = 1, \ldots, |C|$ (similarly for $K_{D\setminus C}$). Therefore equation (A.3) resolves as

$$h_{D\setminus C}(\boldsymbol{x}_{D\setminus C}) \, \Phi_{|C|}(\boldsymbol{v}_C - \boldsymbol{\beta}_C; \boldsymbol{\mu}, \Gamma)$$

with $\Phi_{|C|}(\,\cdot\,; \boldsymbol{\mu}, \Gamma)$ the cdf of a $|C|$-variate multivariate Gaussian distribution with location vector $\boldsymbol{\mu}$ and covariance matrix Γ.

Case $f_{\boldsymbol{U}} = f_{\boldsymbol{V}}$. Again this can be found similarly to the above noting the relation between these two forms; see also Wadsworth and Tawn (2014).

Generators with structured components

Recall that since this is a model on the random vector $\boldsymbol{R}$, we need to differentiate between $\boldsymbol{\gamma} = \mathbf{0}$ and $\boldsymbol{\gamma} > \mathbf{0}$. We present the case $\boldsymbol{\gamma} = \mathbf{0}$ only, since the case $\boldsymbol{\gamma} > \mathbf{0}$ is very similar. Moreover, we set $\boldsymbol{v} = v\mathbf{1}$ as in Section 6.2.

Case $\boldsymbol{\gamma} = \mathbf{0}$. The censored likelihood has an analytical expression but is tedious to write down. Note that, since the density $h(\boldsymbol{x}; \mathbf{1}, \mathbf{0})$ is non-zero only for $x_1 < \ldots < x_d$, we censor in $|C| = k$ components if $x_1 < \ldots < x_k < v < x_{k+1} < \ldots < x_d$. If $k = 1$, then for $\mathbb{1}(v < x_2 < \ldots < x_d)$ and $\mathbb{1}(x_d > 0)$,

$$\begin{aligned} h_C(x_{2:d}, v; \mathbf{1}, \mathbf{0}) &= \frac{d! \prod_{j=1}^d \lambda_j}{\sum_{j=1}^d \lambda_j^{-1}} \int_{-\infty}^v \frac{\prod_{j=1}^d e^{x_j}}{\left(\sum_{j=1}^d (\lambda_j - \lambda_{j+1}) e^{x_j}\right)^{d+1}} \mathrm{d}x_1 \\ &= \frac{(d-1)! e^{\sum_{j=2}^d x_j} \prod_{j=1}^d \lambda_j}{(\lambda_1 - \lambda_2) \sum_{j=1}^d \lambda_j^{-1}} \left\{ \left(\sum_{j=2}^d (\lambda_j - \lambda_{j+1}) e^{x_j} \right)^{-d} \right. \\ &\quad \left. - \left((\lambda_1 - \lambda_2) e^v + \sum_{j=2}^d (\lambda_j - \lambda_{j+1}) e^{x_j} \right)^{-d} \right\}, \end{aligned}$$

where $x_{2:d} = (x_2, \ldots, x_d)$. Expressions for $k > 1$ follow naturally by repeated integration of the above result.

B Supporting information for Section 6

Supporting information for Section 6.1

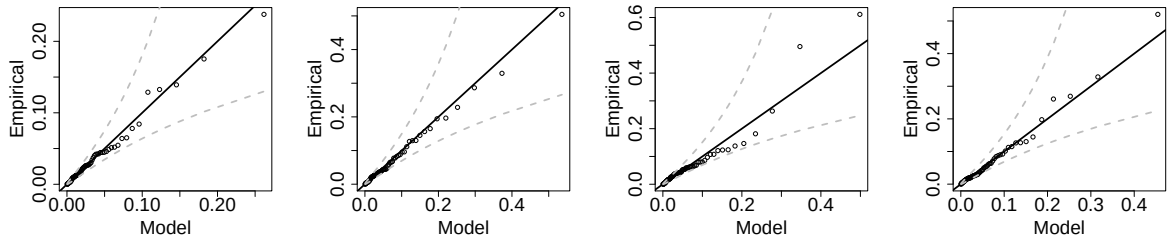

Figure 12: Negative UK bank returns: marginal QQ-plots using the fitted GP distribution. From left to right: HSBC, Lloyds, RBS and Barclays. The 95% pointwise confidence intervals are obtained by a transformation of the beta distributed order statistics of a uniform distribution.

Supporting information for Section 6.2

For the three-dimensional structured components model fitted in Section 6.2, the dependence measures χ_{13}, χ_{23} and χ_{123} are

$$\begin{aligned} \chi_{13} &= 1 - \frac{\lambda_1(\lambda_2 + \lambda_3)^3}{(\lambda_3 + 2\lambda_2)(\lambda_2 + 2\lambda_3)(\lambda_2\lambda_3 + \lambda_1\lambda_3 + \lambda_1\lambda_2)}, \\ \chi_{23} &= 1 - \frac{\lambda_1\lambda_2(\lambda_1 + \lambda_2)^2}{(\lambda_1 + 2\lambda_2)(\lambda_2 + 2\lambda_1)(\lambda_2\lambda_3 + \lambda_1\lambda_3 + \lambda_1\lambda_2)}, \\ \chi_{123} &= 1 - \frac{\lambda_1}{2(\lambda_1 + \lambda_2)} - \frac{\lambda_1\lambda_2(4\lambda_1\lambda_2 + \lambda_1\lambda_3 + 3\lambda_2^2 + \lambda_2\lambda_3)}{3(2\lambda_1 + \lambda_2)(2\lambda_2 + \lambda_3)(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)}. \end{aligned}$$

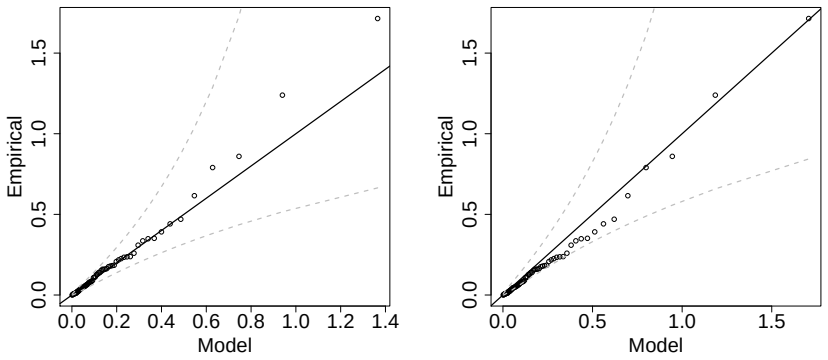

Figure 13: Negative UK bank returns: QQ-plots for GP distribution fitted by maximum likelihood to (6.1) (left) and for GP distribution with scale and shape parameter determined by the multivariate fit and Proposition 5.7 of Rootzén et al. (2016) (right). The 95% pointwise confidence intervals are obtained by a transformation of the beta distributed order statistics of a uniform distribution.

Acknowledgements

The authors gratefully acknowledge the financial support from the following agencies and projects: the Knut and Alice Wallenberg foundation (A. Kiriliouk, H. Rootzén, and J. Wadsworth), the "Fonds de la Recherche Scientifique - FNRS" (A. Kiriliouk) the contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" (A. Kiriliouk and J. Segers), IAP research network Grant P7/06 of the Belgian government (J. Segers), and EPSRC fellowship grant EP/P002838/1 (J. Wadsworth).

We thank the Abisko Scientific Research Station for providing access to their rainfall data, which we used in Section 6.2. These data may be obtained by following the instructions at `http://polar.se/en/abisko-naturvetenskapliga-station/vaderdat`. Data for Section 6.1 can be downloaded from `http://finance.yahoo.com`. Programs for the analyses in Sections 6.1 and 6.2 are available from the authors on request.

References

Aulbach, S., Falk, M., and Zott, M. (2015). The space of D-norms revisited. *Extremes*, 18(1):85–97.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

Beylich, A. A. and Sandberg, O. (2005). Geomorphic effects of the extreme rainfall event of 20–21 july, 2004 in the Latnjavagge catchment, northern Swedish Lapland. *Geografiska Annaler: Series A, Physical Geography*, 87(3):409–419.

Caeiro, F. and Gomes, M. I. (2016). Threshold selection in extreme value analysis. In *Extreme Value Modeling and Risk Analysis: Methods and Applications*. CRC Press.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(2):377–392.

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 comments). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52(3):393–442.

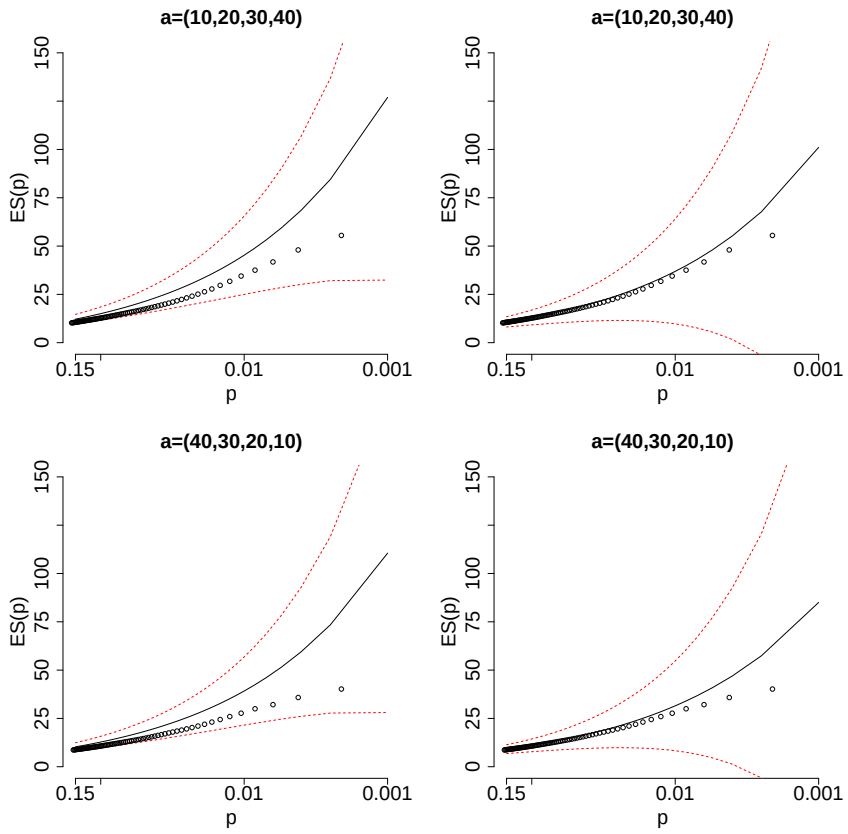

Figure 14: ES estimates and pointwise 95% delta-method confidence intervals for portfolio losses based on the weights given as the figure title. Estimates based on the multivariate GP fit are on the left; estimates based on the univariate fit are on the right.

Einmahl, J. H., Kiriliouk, A., and Segers, J. (2016). A continuous updating weighted least squares estimator of tail dependence in high dimensions. Available at `http://arxiv.org/abs/1601.04826`.

Falk, M. and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis*, 99(4):715–734.

Falk, M. and Michel, R. (2009). Testing for a multivariate generalized Pareto distribution. *Extremes*, 12(1):33–51.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern europe. *Meteorology and atmospheric physics*, 98(3-4):239–267.

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.

Huser, R. and Davison, A. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.

Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103.

Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 1–22.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist. Probab. Lett.*, 7(4):283–286.

Jonasson, C. and Nyberg, R. (1999). The rainstorm of August 1998 in the Abisko area, northern Sweden: preliminary report on observations of erosion and sediment transport. *Geografiska Annaler: Series A, Physical Geography*, 81(3):387–390.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative

definite functions. *Annals of Probability*, 37(5):2042–2065.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59:475–499.

Lee, J., Fan, Y., and Sisson, S. A. (2015). Bayesian threshold selection for extremal models using measures of surprise. *Computational Statistics & Data Analysis*, 85:84–99.

Neuts, M. F. (1974). Probability distributions of phase type. In *Liber Amicorum Professor Emeritus H. Florin*, pages 173–206. University of Louvain, Belgium.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Rapp, A. and Strömquist, L. (1976). Slope erosion due to extreme rainfall in the Scandinavian mountains. *Geografiska Annaler. Series A. Physical Geography*, 58(3):193–200.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2016). Multivariate peaks over thresholds models. Available at `http://arxiv.org/abs/1603.06619`.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.

Rudvik, A. (2012). *Dependence structures in stable mixture models with an application to extreme precipitation*. PhD thesis, Chalmers University of Technology.

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.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT — Statistical Journal*, 10(1):61–92.

Tajvidi, N. (1996). *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions*. PhD thesis, Department of Mathemtics, Chalmers, Göteborg.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.

Wadsworth, J. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wadsworth, J. and Tawn, J. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5):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.