

DISCUSSION PAPER

2016/18

Multivariate peaks over thresholds models

Rootzen, H., Segers, J. and J. Wadsworth

Multivariate peaks over thresholds models

Holger Rootzén, Johan Segers, and Jenny Wadsworth

March 23, 2016

Abstract

Multivariate peaks over thresholds modeling based on generalized Pareto distributions has up to now only been used in few and mostly low-dimensional situations. This paper contributes to the theoretical understanding, physically based models, inference tools, and simulation methods needed to support routine use, also in high dimensions. We derive a general model for extreme episodes in data, and show how conditioning the distribution of extreme episodes on threshold exceedance gives three basic representations of the family of generalized Pareto distributions. The first representation is constructed on the real scale of the observations. The second one starts with a model on a standard exponential scale which then is transformed to the real scale. The third is a spectral representation proposed in A. Ferreira and L. de Haan [Bernoulli 20 (2014) 1717-1737]. Numerically tractable forms of densities and censored densities are derived, and give tools for flexible parametric likelihood inference. Our results are aimed at statistical analysis of complex and multidimensional extreme episodes, such as simultaneous flooding of many dykes; landslides caused by heavy rainfall over one or several days; or deaths caused by a sequence of very hot nights during a heat wave.

1 Introduction

Peaks over thresholds (PoT) modelling was introduced in the hydrological literature. The philosophy is simple: extreme events, perhaps extreme water levels, are often quite different from ordinary everyday behavior, and ordinary behavior then has little to say about extremes, so that only other extreme events give useful information about future extreme events. To make this idea operational, one defines an extreme event as a value, say a water level, which exceeds some high threshold, and only uses the sizes of the excesses over this threshold, the “peaks over the threshold”, for statistical inference. This idea was given a theoretical foundation by combining it with asymptotic arguments motivating that the natural model is that exceedances occur according to a Poisson process and that excess sizes follow a generalized Pareto (GP) distribution (Balkema and de Haan, 1974; Pickands, 1975; Davison and Smith, 1990).

Since then thousands of papers have used one-dimensional PoT models (though often not under this name), in areas ranging from earth and atmosphere science to finance, see e.g. Kyselý et al. (2010), Katz et al. (2002), McNeil et al. (2015), and the method has

been presented in a number of books. As just one example, the book by Coles (2001) gives a useful practical introduction to the method.

However, often it is not just one extreme event which is important, but an entire extreme episode. In the 2005 flooding of New Orleans caused by windstorm Katrina, more than 50 dykes were breached. However, many others held, and the damage was determined by which dykes held and which were flooded (Wikipedia, 2016b). Extreme rain can lead to devastating landslides, and can be caused by one day with very extreme rainfall, or by smaller, but still extreme rain amounts during two or more consecutive days (Guzzetti et al. (2007)). The 2003 heat-wave in central Europe is estimated to have killed between 25 000 and 70 000 people. Many deaths, however, were not caused by one extremely hot day, but rather by a sequence of very high minimum nightly temperatures which led to increasing fatigue, and eventually to death (Wikipedia, 2016a). These, and very many other important societal problems underline the importance of statistical methods which can handle multivariate extreme episodes, such as the water levels at many different locations during a flood, or the minimum night temperatures during a number of consecutive days.

The aim of this paper is to contribute to the probabilistic understanding, physically based models, likelihood tools, and simulation methods which are needed for multivariate PoT modelling of extreme episodes, also in high dimensions. Using the same philosophy as for extreme events in one dimension, PoT modelling of extreme episodes proceeds by choosing a high threshold for each component of the episode, and then to consider an episode as extreme if at least one component exceeds its threshold. One then only models the difference between the values of the components and their respective thresholds. However, in the multivariate case both the heights of the peaks over the thresholds and the depths of the troughs under the thresholds are modeled. For instance, in a rainfall episode affecting a number of catchments, both the amount of rain in the catchments where rainfall exceeds the threshold and in catchments where the threshold is not exceeded are important. Additionally, the inclusion of troughs, or undershoots, increases the amount of information that can be used for inference.

Just as in one dimension, the natural model is that extreme episodes occur according to a Poisson process and that overshoots and undershoots (or undershoots larger than a censoring threshold) jointly follow a multivariate GP distribution. Below we derive the basic properties of this class of distributions. We then pursue the following program:

- (i) to model the possible asymptotic forms of extreme episodes in data;
- (ii) to show how conditioning on threshold exceedances transforms the distribution of the extreme episodes to GP distributions, and to use this to find physically motivated and interpretable representations of the multivariate GP distributions; and
- (iii) to derive likelihoods and censored likelihoods for the representations in ii).

In part ii) of the program, we develop three representations. The first one is in the same units as the observations, i.e., on the real scale, and in the second one the model is built on a standard exponential scale and then transformed to the real observation scale. The third is a spectral representation proposed in Ferreira and de Haan (2014).

A useful, and to us surprising, discovery is that it is possible to derive the density for the Ferreira–de Haan representation, and that this density in fact is simpler than the densities for the other two representations. The importance of iii) is that likelihood inference makes it possible to incorporate covariates, e.g. temporal or spatial trends, in the modelling in a flexible and practical way.

The insights and results obtained in carrying out this program, we believe, will lead to new models, new computational techniques, and new ways to make the necessary compromises between modeling realism and computational tractability which together will make possible routine use, also in high dimensions. The models mentioned in Example 5.2 may be a case in point. The formulas for probabilities, conditional probabilities and conditional densities given in Section 5.3, together with the discovery that weighted sums of components of GP distributions conditioned to be positive also have a GP distribution, add to the usefulness of the methods.

Simulation of GP distributions is needed for several reasons, including computation of the probabilities of complex dangerous events and goodness of fit checking. The final contribution of this paper is a number of simulation algorithms for multivariate GP distributions.

The multivariate GP distributions were introduced in Tajvidi (1996), Beirlant et al. (2004, Chapter 8), and Rootzén and Tajvidi (2006); see also Falk et al. (2010, Chapter 5). The literature on applications of multivariate PoT modelling is rather sparse (Brodin and Rootzén, 2009; Michel, 2009; Aulbach et al., 2012). Some earlier papers use point process models which are closely related to the PoT/GP approach (Coles and Tawn, 1991; Joe et al., 1992). Other papers consider nonparametric or semiparametric rank-based PoT methods focusing on the dependence structure but largely ignoring modelling the margins (de Haan et al., 2008; Einmahl et al., 2012, 2016). However, the GP approach has the advantages that it provides complete models for the threshold excesses, that it can use well-established model checking tools, and that it leads to more natural parametrizations of trends in the Poisson process which governs the occurrence of extreme episodes.

There is an important literature on modelling componentwise, perhaps yearly, maxima with multivariate generalized extreme value (GEV) distributions. A survey in the spatial context is given in Davison et al. (2012). Componentwise maxima are sometimes mandated by the practical problem under study, or by the fact that only data on componentwise maxima are available. However, the componentwise maxima may occur at different times for different components, and in many situations focus is on the structure modeled by the PoT approach: extremes which occur simultaneously. Additionally, likelihood inference for GEV distributions is complicated by the fact that analytic expressions for high-dimensional densities typically are not available, so that inference often is much easier, and perhaps more efficient, in GP models; see Huser et al. (2015) for a survey and an extensive comparison.

Section 2 recalls some basic facts about multivariate GEV and GP distributions, and derives and exemplifies the basic properties of the class of GP cumulative distribution functions (cdf-s). In Section 3 we derive a point process model of the structure of extreme episodes, and Section 4 shows how conditioning on exceeding high thresholds

leads to three basic representations of the GP distributions. In Section 5 we derive the densities and censored densities which correspond to the three representations, discuss some parametrization issues, and give formulas for probabilities and conditional probabilities in GP distributions. Finally, Section 6 contributes simulation algorithms for multivariate GP distributions and Section 7 gives a concluding overview of results and future challenges.

2 Multivariate generalized Pareto distributions

This section first briefly recalls and adapts parts of existing theory for multivariate GEV and GP distributions, then derives a number of basic properties of GP distributions, and ends with a brief comment on work by Falk and coauthors.

We throughout use notation as follows. The maximum and minimum operators are denoted by the symbols $\vee$ and $\wedge$, respectively. Bold symbols denote d -variate vectors. For instance, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ and $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$. Operations and relations involving such vectors are meant componentwise, with shorter vectors being recycled if necessary. For instance $\boldsymbol{a}\boldsymbol{x} + \boldsymbol{b} = (a_1x_1 + b_1, \dots, a_dx_d + b_d)$, $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_j \leq y_j$ for $j = 1, \dots, d$, and $t^\gamma = (t^{\gamma_1}, \dots, t^{\gamma_d})$. If F is a cdf then we write $\bar{F} = 1 - F$ for its tail function, and also write F for the probability distribution determined by the cdf. The expression $\boldsymbol{X} \sim F$ means that $\boldsymbol{X}$ has distribution F , and $\xrightarrow{d}$ denotes convergence in distribution. The symbol $\mathbf{1}$ is the indicator function: $\mathbf{1}_A$ equals 1 on the set A and 0 otherwise.

For fixed $\gamma \in \mathbb{R}$, the functions $x \mapsto (x^\gamma - 1)/\gamma$ (for $x > 0$) and $x \mapsto (1 + \gamma x)^{1/\gamma}$ are to be interpreted as their limits $\log(x)$ and $\exp(x)$, respectively, in case $\gamma = 0$. This convention also applies componentwise to expressions of the form $(\boldsymbol{x}^\gamma - 1)/\gamma$ and $(1 + \gamma \boldsymbol{x})^{1/\gamma}$.

Below we will repeatedly use that if $\boldsymbol{X}$ is a d -dimensional vector with $P(\boldsymbol{X} \not\leq \boldsymbol{u}) > 0$ and $\boldsymbol{s} > \mathbf{0}$ then

$$P(\boldsymbol{s}(\boldsymbol{X} - \boldsymbol{u}) \leq \boldsymbol{x}) \mid \boldsymbol{X} \not\leq \boldsymbol{u} = \frac{P(\boldsymbol{X} \leq \boldsymbol{x}/\boldsymbol{s} + \boldsymbol{u}) - P(\boldsymbol{X} \leq (\boldsymbol{x} \wedge \mathbf{0})/\boldsymbol{s} + \boldsymbol{u})}{P(\boldsymbol{X} \not\leq \boldsymbol{u})}. \quad (2.1)$$

2.1 Background: multivariate generalized extreme value distributions

Throughout, G denotes a d -variate GEV distribution. The class of GEV distributions has the following equivalent characterizations, see e.g. Beirlant et al. (2004).

(M1) *It is the class of limit distributions of maxima:* Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ be an i.i.d. sequence with cdf F . If, for some scaling and location sequences $\boldsymbol{a}_n \in (0, \infty)^d$ and $\boldsymbol{b}_n \in \mathbb{R}^d$,

$$P\left(\boldsymbol{a}_n^{-1}(\vee_{i=1}^d \boldsymbol{X}_i - \boldsymbol{b}_n) \leq \boldsymbol{x}\right) \xrightarrow{d} G(\boldsymbol{x}), \quad \text{as } n \rightarrow \infty, \quad (2.2)$$

where G has non-degenerate margins, then G is a GEV distribution. Conversely all GEV distributions may be obtained in this way.

(M2) *It is the class of max-stable distributions:* Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be i.i.d. with cdf G which has non-degenerate margins. If there are scaling and location sequences $\mathbf{a}_n \in (0, \infty)^d$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$P(\mathbf{a}_n^{-1}(\bigvee_{i=1}^n \mathbf{X}_i - \mathbf{b}_n) \leq \mathbf{x}) = G(\mathbf{x}), \quad \text{for } n = 1, 2, \dots,$$

then G is a GEV distribution. Conversely all GEV distributions have this property.

If (M1) holds then F is said to be in the (maximum) domain of attraction of G , and we write $F \in D(G)$. It follows from (M1) that the class of GEV distributions is closed under location and scale changes.

The marginal distribution functions $G_1, \dots, G_d$ of G may be written as

$$G_j(x) = \exp \left\{ - \left(1 + \gamma_j \frac{x - \mu_j}{\alpha_j} \right)^{-1/\gamma_j} \right\}, \quad (2.3)$$

for $x \in \mathbb{R}$ such that $\alpha_j + \gamma_j(x - \mu_j) > 0$. We will use this parametrization throughout. The case $\gamma_j = 0$ is to be interpreted as the limit $\gamma_j \rightarrow 0$, that is, $G_j(x) = \exp[-\exp\{-(x - \mu_j)/\alpha_j\}]$ for all $x \in \mathbb{R}$. The parameter range is $(\gamma_j, \mu_j, \alpha_j) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$. Define

$$\boldsymbol{\sigma} = \boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{\mu},$$

so that $\sigma_j = \alpha_j - \gamma_j\mu_j$, $j \in \{1, \dots, d\}$. Then G_j is supported on the interval

$$I_j = \begin{cases} (-\sigma_j/\gamma_j, \infty) & \text{if } \gamma_j > 0, \\ (-\infty, \infty) & \text{if } \gamma_j = 0, \\ (-\infty, -\sigma_j/\gamma_j) & \text{if } \gamma_j < 0, \end{cases} \quad (2.4)$$

while G is supported by a (subset of) the rectangle $I_1 \times \dots \times I_d$. The lower and upper endpoints of G_j are denoted by $\eta_j \in \mathbb{R} \cup \{-\infty\}$ and $\omega_j \in \mathbb{R} \cup \{+\infty\}$, respectively. One may alternatively write the condition $\mathbf{x} \in I_1 \times \dots \times I_d$ as $\boldsymbol{\gamma}\mathbf{x} + \boldsymbol{\sigma} > \mathbf{0}$.

Below we will assume that $0 < G(\mathbf{0}) < 1$. This inequality is equivalent to $G_j(0) > 0$ for all $j \in \{1, \dots, d\}$ and $G_j(0) < 1$ for some $j \in \{1, \dots, d\}$. The property underlying this equivalence is the fact that from positive quadrant dependence (Marshall and Olkin, 1983, Equation (13)), we have that $G(\mathbf{0}) \geq \prod_{j=1}^d G_j(0)$.

We will often also use the stronger condition that $\boldsymbol{\sigma} > \mathbf{0}$, i.e., that $\sigma_j > 0$ for all $j \in \{1, \dots, d\}$. By (2.4), this condition is equivalent to assuming that 0 is in the interior of the support of every one of the d margins $G_1, \dots, G_d$, i.e., that $\eta_j < 0 < \omega_j$ and thus $0 < G_j(0) < 1$ for all $j \in \{1, \dots, d\}$. Also, to assume that $\sigma_j > 0$ is an additional restriction only in case $\gamma_j < 0$: if $\gamma_j = 0$, then $\sigma_j = \alpha_j > 0$ anyway, while if $\gamma_j > 0$, then $G_j(0) > 0$ implies $\eta_j < 0$ and thus $\sigma_j = -\gamma_j\eta_j > 0$.

An easy argument shows that G is max-stable if and only if to each $t > 0$ there exist scale and location vectors $\mathbf{a}_t \in (0, \infty)^d$ and $\mathbf{b}_t \in \mathbb{R}^d$ such that $G(\mathbf{a}_t\mathbf{x} + \mathbf{b}_t)^t \equiv G(\mathbf{x})$ (Resnick, 1987, Equation (5.17)). It follows from (2.3) that these parameters are given by

$$\mathbf{a}_t = t^\gamma, \quad \mathbf{b}_t = \boldsymbol{\sigma}(t^\gamma - 1)/\boldsymbol{\gamma}. \quad (2.5)$$

To a GEV distribution G we can associate a Borel measure ν on $\prod_{j=1}^d [-\eta_j, \infty) \setminus \{\boldsymbol{\eta}\}$ by the formula $\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{x}\}) = -\log G(\mathbf{x})$ for $\mathbf{x} \in [-\infty, \infty)^d$, with the convention that $-\log(0) = \infty$ (Resnick, 1987, Proposition 5.8). The measure ν is called *intensity measure* because, by the domain-of-attraction condition (2.2), the limit of the expected number of points $\mathbf{a}_n^{-1}(\mathbf{X}_i - \mathbf{b}_n)$, $i \in \{1, \dots, n\}$, in a ν -smooth Borel set A is equal to $\nu(A)$. The intensity measure ν determines the limit distribution of the sequence of point processes $\sum_{i=1}^n \delta_{\mathbf{a}_n^{-1}(\mathbf{X}_i - \mathbf{b}_n)}$, see Section 3.

2.2 Generalized Pareto distributions

Let G be a GEV distribution with $0 < G(\mathbf{0}) < 1$ and let ν be the corresponding intensity measure. Then $0 < \nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}\}) < \infty$, so that we can define a probability measure supported on the set $\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}\}$ by restricting the intensity measure ν to that set and normalizing it. The result is the *generalized Pareto (GP)* distribution associated to G . For its cumulative distribution function, H , we obtain the expression

$$H(\mathbf{x}) = \begin{cases} \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G(\mathbf{x} \wedge \mathbf{0})}{G(\mathbf{x})} \right) & \text{if } \mathbf{x} > \boldsymbol{\eta}, \\ 0 & \text{if } x_j < \eta_j \text{ for some } j = 1, \dots, d, \end{cases} \quad (2.6)$$

see Beirlant et al. (2004, Chapter 8) and Rootzén and Tajvidi (2006). If a GEV cdf G and a GP cdf H satisfy (2.6), then we say that they are *associated* and write $H \leftrightarrow G$.

For completeness, we have added a proof of (2.6) in the appendix. For points $\mathbf{x} \in [-\infty, \infty)^d$ such that $\mathbf{x} \geq \boldsymbol{\eta}$ and $x_j = \eta_j$ for some j , the value of $H(\mathbf{x})$ is determined by right-hand continuity¹: $H(\mathbf{x}) = \lim_{\mathbf{h} \searrow \mathbf{0}} H(\mathbf{x} + \mathbf{h})$. The margins of H are denoted by $H_1, \dots, H_d$.

The GP distributions defined by (2.6) constitute the class of limit distributions of threshold excesses for random vectors. The probability that component $j \in \{1, \dots, d\}$ exceeds zero is equal to $1 - H_j(0) = \log G_j(0) / \log G(\mathbf{0})$, which is positive if and only if $G_j(0) < 1$, that is, when $\sigma_j = \alpha_j - \gamma_j \mu_j > 0$. Since $G(\mathbf{0}) < 1$ implies that $G_j(0) < 1$ for some but not necessarily all j , the GP family includes distributions for which one (or several) of the components never exceed their threshold, so that the support of that component lies in $[-\infty, 0]$. This could be useful in some modelling situations, but still, the situation which is of main interest is when all components have a positive probability of being an exceedance, or equivalently when $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$, or, again equivalently, when $\boldsymbol{\sigma} > \mathbf{0}$.

Completely parallel to the class of GEV distributions, the class of GP distributions H such that $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$ has the following equivalent characterizations (Rootzén and Tajvidi, 2006). The threshold functions $\mathbf{u}_t$ in the characterizations are assumed to be continuous and increasing.

¹The cited references did not include this provision, leading to confusion in Ferreira and de Haan (2014).

(T1) *It is the class of limit distributions of threshold excesses:* Let $\mathbf{X} \sim F$. If there exist scaling and threshold functions $\mathbf{s}_t \in (0, \infty)^d$ and $\mathbf{u}_t \in \mathbb{R}^d$, with $F(\mathbf{u}_t) < 1$ and $F(\mathbf{u}_t) \rightarrow 1$ as $t \rightarrow \infty$, such that

$$P(\mathbf{s}_t^{-1}(\mathbf{X} - \mathbf{u}_t) \leq \cdot \mid \mathbf{X} \not\leq \mathbf{u}_t) \xrightarrow{d} H, \quad \text{as } t \rightarrow \infty,$$

where H has nondegenerate margins and $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$, then H is a GP distribution. Conversely, all GP distributions H such that $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$ may be obtained in this way.

(T2) *It is the class of threshold-stable distributions:* Let $\mathbf{X} \sim H$ where H has nondegenerate margins and $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$. If there exist scaling and threshold function $\mathbf{s}_t \in (0, \infty)^d$ and $\mathbf{u}_t \in \mathbb{R}^d$, with $H(\mathbf{u}_t) \rightarrow 1$ as $t \rightarrow \infty$, such that

$$P(\mathbf{s}_t^{-1}(\mathbf{X} - \mathbf{u}_t) \leq \mathbf{x} \mid \mathbf{X} \not\leq \mathbf{u}_t) = H(\mathbf{x}) \quad \text{for all } t \geq 1,$$

then H is a GP distribution. Conversely, all GP distributions H for which $H_j(0) < 1$ for all $j \in \{1, \dots, d\}$ have this property.

In Rootzén and Tajvidi (2006)² it is additionally shown that in (T1) and (T2), it is enough to assume that the convergence holds for $\mathbf{x} > \mathbf{0}$.

If (T1) holds we say that F belongs to the (threshold) domain of attraction of H with threshold function $\mathbf{u}_t$ and write $F \in D(H)$. In contrast to the situation for GEV distributions, different threshold functions can lead to limits which are not location-scale transformations of one another. A cdf F is in a domain of attraction for maxima if and only if it is in a threshold domain of attraction.

Recall the notation $I_j = (\eta_j, \omega_j)$ for the interval supporting G_j . The GP distribution H is supported by the set

$$[\boldsymbol{\eta}, \boldsymbol{\omega}] \setminus [\boldsymbol{\eta}, \mathbf{0}] = \{\mathbf{x} \in \mathbb{R}^d : \omega_j \geq x_j \geq \eta_j \text{ for all } j, \text{ and } x_j > 0 \text{ for some } j\}.$$

The GP distribution H may assign positive mass to the hyperplanes $\{\mathbf{y} : y_j = \eta_j\}$, even if $\eta_j = -\infty$; see Examples 2.3 and 2.4 below.

For a non-empty subset S of $\{1, \dots, d\}$, let H_S denote the corresponding $|S|$ -variate marginal distribution of H . Further, let H_S^+ denote H_S conditioned to have at least one positive component; this presupposes that $\sigma_j > 0$ for some $j \in S$, where $\sigma_j = \alpha_j - \mu_j \gamma_j$ as before. By Proposition 2.1(i) below, if $\boldsymbol{\sigma} > \mathbf{0}$, then $H_j^+ := H_{\{j\}}^+$, the j -th marginal distribution of H , conditioned to be positive, has cdf

$$H_j^+(x) = 1 - \left(1 + \gamma_j \frac{x}{\sigma_j}\right)^{-1/\gamma_j}, \quad \text{for } x \geq 0 \text{ such that } \sigma_j + \gamma_j x > 0. \quad (2.7)$$

This proves the intuitively appealing result that H_j^+ is a one-dimensional GP distribution, and also shows that $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ are parameters of the marginal distributions.

²In Theorem 2.3 (ii) in Rootzén and Tajvidi (2006), “nondegenerate margins” should be changed to “nondegenerate margins on $\mathbb{R}_+$ ”.

In the following proposition we give the general version of this result and collect a number of other properties of GP distributions. If S is a non-empty subset of $\{1, \dots, d\}$ and $\mathbf{x} \in [-\infty, \infty]^S$, then $\mathbf{x}_d \in [-\infty, \infty]^d$ is defined by $(\mathbf{x}_d)_j = x_j$ if $j \in S$ and $(\mathbf{x}_d)_j = \infty$ if $j \notin S$. Thus, if $\mathbf{X} \sim H$, then the marginal distribution, H_S , of $(X_j : j \in S)$ is given by $H_S(\mathbf{x}) = H(\mathbf{x}_d)$ for $\mathbf{x} \in [-\infty, \infty]^S$, and if $H_S(\mathbf{0}) < 1$ then

$$H_S^+(\mathbf{x}) = \frac{H_S(\mathbf{x}) - H_S(\mathbf{x} \wedge \mathbf{0})}{\bar{H}_S(\mathbf{0})} \quad (2.8)$$

is the conditional distribution of $(X_j : j \in S)$ given that $\max_{j \in S} X_j > 0$, see (2.1) above. Recall that G and H are said to be associated, $H \leftrightarrow G$, if they satisfy (2.6).

Proposition 2.1. *Let G be a GEV with margins (2.3) and suppose $H \leftrightarrow G$.*

- (i) *Let $S \subset \{1, \dots, d\}$ be non-empty. If $H_S(\mathbf{0}) < 1$ then H_S^+ is a GP cdf too, with $H_S^+ \leftrightarrow G_S$, and if $\sigma_j > 0$ then (2.7) holds. Furthermore, H_S is a GP distribution if and only if $H_S(\mathbf{0}) = 0$.*
- (ii) *A scale transformation of H is also a GP distribution.*
- (iii) *Let $\mathbf{X} \sim H$. If $\boldsymbol{\sigma} > \mathbf{0}$ and $\mathbf{u} \geq \mathbf{0}$ with $H(\mathbf{u}) < 1$, then the conditional distribution of $\mathbf{X} - \mathbf{u}$ given that $\mathbf{X} \not\leq \mathbf{u}$ is a GP distribution with the same shape parameter γ and $\boldsymbol{\sigma}$ replaced by $\boldsymbol{\sigma} + \gamma\mathbf{u}$.*
- (iv) *We have $H \leftrightarrow G^t$ for all $t > 0$. Conversely, if $H \leftrightarrow G_1$ for some G_1 and if $\boldsymbol{\sigma} > \mathbf{0}$ then $G_1 = G^{t_1}$ for some $t_1 > 0$.*
- (v) *If $G(\mathbf{0}) = e^{-v}$ then $G(\mathbf{x}) = \exp\{-v\bar{H}(\mathbf{x})\}$ for $\mathbf{x} \geq \mathbf{0}$, and if $\boldsymbol{\sigma} > \mathbf{0}$ this determines G .*
- (vi) *If $\boldsymbol{\sigma} > \mathbf{0}$, the scaling and threshold functions in the (T2) characterization of GP distributions may be taken as $\mathbf{s}_t = t^\gamma$ and $\mathbf{u}_t = \boldsymbol{\sigma}(t^\gamma - \mathbf{1})/\gamma$, for $t \geq 1$.*
- (vii) *The parameters γ and $\boldsymbol{\sigma}$ are identifiable from H .*

Proposition 2.2. *If $\{H_n\}$ is a sequence of GP distributions with the vectors $\boldsymbol{\sigma}_n$ bounded away from $\mathbf{0}$ and if $H_n \xrightarrow{d} H$ then H is a GP distribution too.*

In words, Proposition 2.1(i) says that conditional margins of GP distributions are GP, but that marginal distributions of GP distribution are typically not GP. For instance, if H is a two-dimensional GP cdf, then H_1^+ is a one-dimensional GP cdf [given by (2.7)], but typically H_1 is not. Intuitively, the reason is that the conditioning event implicit in $H_1(x)$ also includes the possibility that it is the second component, rather than the first one, that exceeds its threshold.

Proposition 2.1 (i)–(iii) and Proposition 2.2 establish various closure properties of the class of GP distributions. By (iv) and (v) of Proposition 2.1, a GEV distribution specifies the associated GP distribution and conversely a GP distribution specifies a curve of associated GEV distributions in the space of distribution functions. By (iv) it follows that, for a GEV G and a GP H , if $H \leftrightarrow G^{t_0}$ for some $t_0 > 0$, then $H \leftrightarrow G^t$ for all $t > 0$. Moreover, since $G(\mathbf{a}_t\mathbf{x} + \mathbf{b}_t)^t \equiv G(\mathbf{x})$ for $\mathbf{a}_t$ and $\mathbf{b}_t$ given by (2.5), it then also holds that $H \leftrightarrow G(\mathbf{a}_t(\cdot) + \mathbf{b}_t)$. Regarding (v), note that a GEV distribution G such that $0 < G_j(0) < 1$ for all $j \in \{1, \dots, d\}$ is determined by the values of $G(\mathbf{x})$ for $\mathbf{x} \geq \mathbf{0}$ (proof

in the appendix). Finally, the assertion (vi) identifies the affine transformations which leave H unchanged, and (vii) establishes identifiability of the marginal parameters.

Proof of Proposition 2.1. (i) Let $\mathbf{0}_d$ denote $\mathbf{x}_d$ for the special case when $\mathbf{x} = \mathbf{0} \in (-\infty, \infty)^S$ and let $G_S(\mathbf{x}) = G(\mathbf{x}_d)$ be the marginal distribution of G . We have $\mathbf{x}_d \wedge \mathbf{0} = (\mathbf{x} \wedge \mathbf{0})_d \wedge \mathbf{0}$ and hence, for $\mathbf{x} > \boldsymbol{\eta}$,

$$\begin{aligned} H_S(\mathbf{x}) - H_S(\mathbf{x} \wedge \mathbf{0}) &= \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G(\mathbf{x}_d \wedge \mathbf{0})}{G(\mathbf{x}_d)} \right) - \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G((\mathbf{x} \wedge \mathbf{0})_d \wedge \mathbf{0})}{G((\mathbf{x} \wedge \mathbf{0})_d)} \right) \\ &= \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G_S(\mathbf{x} \wedge \mathbf{0})}{G_S(\mathbf{x})} \right) \end{aligned}$$

and

$$\bar{H}_S(\mathbf{0}) = 1 - \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G(\mathbf{0}_d \wedge \mathbf{0})}{G(\mathbf{0}_d)} \right) = \frac{\log G_S(\mathbf{0})}{\log G(\mathbf{0})},$$

so that

$$H_S^+(\mathbf{x}) = \frac{1}{\log G_S(\mathbf{0})} \log \left(\frac{G_S(\mathbf{x} \wedge \mathbf{0})}{G_S(\mathbf{x})} \right).$$

Further, inserting (2.3) into the equation above for $S = \{j\}$ together with straightforward calculation proves (2.7), and hence completes the proof of the first assertion.

If $H_S(\mathbf{0}) = 0$, then $H_S = H_S^+$ and it follows from the first assertion that H_S is a GP distribution function. If instead $H_S(\mathbf{0}) > 0$ then it follows from (2.6) that H_S is not a GP distribution. This proves the second assertion.

(ii) If G is a GEV cdf then, for $\mathbf{s} > \mathbf{0}$, the map $\mathbf{x} \mapsto G(\mathbf{x}/\mathbf{s})$ is a GEV cdf too, and the result then follows from

$$H(\mathbf{x}/\mathbf{s}) = \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G((\mathbf{x}/\mathbf{s}) \wedge \mathbf{0})}{\log G(\mathbf{x}/\mathbf{s})} \right) = \frac{1}{\log G(\mathbf{0}/\mathbf{s})} \log \left(\frac{G((\mathbf{x} \wedge \mathbf{0})/\mathbf{s})}{G(\mathbf{x}/\mathbf{s})} \right).$$

(iii) Proceeding as in the proof of (i), but in the first step instead using that $(\mathbf{x} \wedge \mathbf{0} + \mathbf{u}) = (\mathbf{x} \wedge \mathbf{0} + \mathbf{u}) \wedge \mathbf{0}$, shows that the conditional distribution of $\mathbf{X} - \mathbf{u}$ given that $\mathbf{X} \not\leq \mathbf{u}$ is

$$\frac{H(\mathbf{x} + \mathbf{u}) - H(\mathbf{x} \wedge \mathbf{0} + \mathbf{u})}{\bar{H}(\mathbf{u})} = \frac{1}{\log G(\mathbf{u})} \log \left(\frac{G(\mathbf{x} \wedge \mathbf{0} + \mathbf{u})}{G(\mathbf{x} + \mathbf{u})} \right).$$

The map $\mathbf{x} \mapsto \tilde{G}(\mathbf{x}) = G(\mathbf{x} + \mathbf{u})$ is also a GEV cdf, but with the vector $\boldsymbol{\sigma} = \boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{\mu}$ replaced by $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\alpha} - \boldsymbol{\gamma}(\boldsymbol{\mu} - \mathbf{u}) = \boldsymbol{\sigma} + \boldsymbol{\gamma}\mathbf{u}$.

(iv) The first assertion follows from (2.6). Using the first assertion, we may without loss of generality assume that $-\log G(\mathbf{0}) = -\log G_1(\mathbf{0})$, and then

$$\frac{G(\mathbf{x} \wedge \mathbf{0})}{G(\mathbf{x})} = \frac{G_1(\mathbf{x} \wedge \mathbf{0})}{G_1(\mathbf{x})},$$

and in particular, $G(\mathbf{x}) = G_1(\mathbf{x})$ for $\mathbf{x} \geq \mathbf{0}$. Since a GEV cdf with $\boldsymbol{\sigma} > \mathbf{0}$ is determined by its values for $\mathbf{x} \geq \mathbf{0}$, see the appendix, this completes the proof.

(v) The first part follows from (2.6), and that this determines G again follows from the appendix.

(vi) It follows from the proofs of (iii) and (iv) that the conditional distribution of $\mathbf{s}_t^{-1}(\mathbf{X} - \mathbf{u}_t)$ given that $\mathbf{X} \not\leq \mathbf{u}_t$ is associated with $G(\mathbf{s}_t \mathbf{x} + \mathbf{u}_t) = G(\mathbf{x})^t$, and the result then follows from (v).

(vii) For $t > 0$, let $(\gamma_j(t), \mu_j(t), \mu_j(t)) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$ be the parameter vector of G_j^t , and let $\sigma_j(t) = \alpha_j(t) - \gamma_j(t) \mu_j(t)$. By assertion (iv), the GPD H determines the curve of GEV-s G^t for $t > 0$. It suffices to show that $\gamma_j(t)$ and $\sigma_j(t)$ do not depend on t . But this follows by straightforward calculations from the max-stability property $G(\mathbf{x})^t = G(\mathbf{a}_t^{-1}(\mathbf{x} - \mathbf{b}_t))$ with $\mathbf{a}_t$ and $\mathbf{b}_t$ as in (2.5). $\square$

Proof of Proposition 2.2. Convergence in distribution in $\mathbb{R}^d$ implies convergence of the marginal distributions, and using standard converging subsequence arguments it follows from marginal convergence that there exist $\boldsymbol{\sigma} > \mathbf{0}$ and $\boldsymbol{\gamma}$ such that $\boldsymbol{\sigma}_n \rightarrow \boldsymbol{\sigma}$ and $\boldsymbol{\gamma}_n \rightarrow \boldsymbol{\gamma}$. Define $\mathbf{u}_{n,t}$ and $\mathbf{s}_{n,t}$ from H_n as in Proposition 2.1(vi). Then, since H_n is a GP cdf we have, using first (T2) and (vi), and then the continuous mapping theorem, that

$$\begin{aligned} H_n(\mathbf{x}) &= \frac{H_n(\mathbf{x}/\mathbf{s}_{n,t} + \mathbf{u}_{n,t}) - H_n((\mathbf{x}/\mathbf{s}_{n,t}) \wedge \mathbf{0} + \mathbf{u}_{n,t})}{\bar{H}_n(\mathbf{u}_{n,t})} \\ &\rightarrow_d \frac{H(\mathbf{x}/\mathbf{s}_t + \mathbf{u}_t) - H((\mathbf{x}/\mathbf{s}_t) \wedge \mathbf{0} + \mathbf{u}_t)}{\bar{H}(\mathbf{u}_t)}, \quad \text{as } n \rightarrow \infty, \end{aligned}$$

and since $H_n \rightarrow_d H$ it follows that H satisfies (T2) and hence is a GP cdf. $\square$

In Examples 2.3 and 2.4 below we exhibit two-dimensional GP distributions with positive mass on certain lines, and Example 2.5 provides a distribution where the second assertion in (i) of Proposition 2.1 comes into play. In contrast to scale transformations, it seems likely that if $\boldsymbol{\sigma} > \mathbf{0}$ then a non-trivial location transformation of a MGP distribution never is a GP distribution. Example 2.6 shows one of the exceptional cases where the support of one of the components is contained in $(-\infty, 0)$ and where a location transformation of a GP distribution does give another GP distribution.

Example 2.3. This example rectifies the one on pages 1726–1727 in Ferreira and de Haan (2014). Let $G(x, y) = \exp\{-1/(x+1) - 1/(y+1)\}$ for $(x, y) \in (-1, \infty)^2$, the distribution of two independent unit Fréchet random variables with lower endpoints

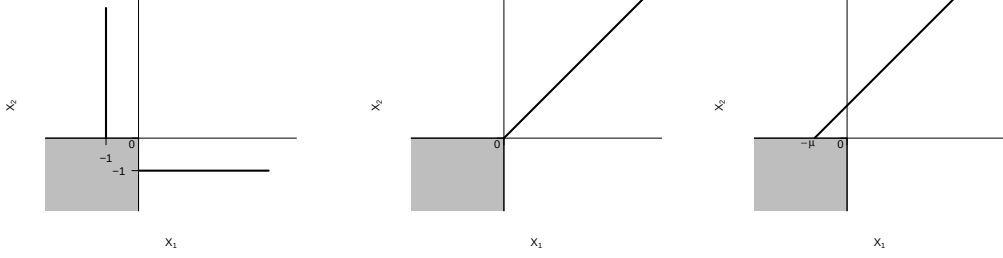


Figure 1: *Supports (solid lines) of the GP distributions H in equations (2.9) (left), (2.10) (middle) and (2.11) (right).*

$\alpha_1 = \alpha_2 = -1$. The corresponding MGP distribution is given by

$$H(x, y) = \begin{cases} \frac{1}{2} \left(1 - \frac{1}{x+1} + 1 - \frac{1}{y+1} \right) & \text{if } (x, y) \in [0, \infty)^2, \\ \frac{1}{2} \left(1 - \frac{1}{x+1} \right) & \text{if } (x, y) \in [0, \infty) \times [-1, 0], \\ \frac{1}{2} \left(1 - \frac{1}{y+1} \right) & \text{if } (x, y) \in [-1, 0] \times [0, \infty), \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

We conclude that H is the distribution function of the random vector (X, Y) given by

$$(X, Y) = \begin{cases} (-1, T) & \text{with probability } 1/2, \\ (T, -1) & \text{with probability } 1/2, \end{cases}$$

where T is a generalized Pareto random variable with $P(T \leq t) = 1 - 1/(t+1)$ for $t \in [0, \infty)$. In particular, H is singular with respect to two-dimensional Lebesgue measure and it is supported by the union of the two lines $\{-1\} \times (0, \infty)$ and $(0, \infty) \times \{-1\}$, see the left-hand panel in Figure 1.

Example 2.4. If we modify Example 2.3 by choosing Gumbel margins rather than Fréchet margins, so that $G(x, y) = \exp(-e^{-x} - e^{-y})$ for $(x, y) \in \mathbb{R}^2$, then the corresponding MGP distribution H is the distribution function of the random pair (X, Y) given by

$$(X, Y) = \begin{cases} (-\infty, T) & \text{with probability } 1/2, \\ (T, -\infty) & \text{with probability } 1/2, \end{cases}$$

where T is a unit exponential random variable, $P(T \leq t) = 1 - e^{-t}$ for $t \in [0, \infty)$. The support of H is given by the union of two half-lines through $-\infty$, that is, the sets $\{-\infty\} \times (0, \infty)$ and $(0, \infty) \times \{-\infty\}$.

Example 2.5. Let $G(x, y) = \exp[-1/\{(x \wedge y) + 1\}]$ for $(x, y) \in (-1, \infty)^2$, the distribution of the random pair (Z, Z) , where Z is a unit Fréchet random variable with lower endpoint given by -1 . The corresponding MGP distribution H is given by

$$H(x, y) = \begin{cases} 1 - \frac{1}{(x \wedge y) + 1} & \text{if } (x, y) \in [0, \infty)^2, \\ 0 & \text{otherwise.} \end{cases} \quad (2.10)$$

We identify H as the distribution of the random pair (T, T) , where $P(T \leq t) = 1 - 1/(t + 1)$ for $t \in [0, \infty)$. The support of H is given by the diagonal $\{(t, t) : 0 < t < \infty\}$, see the middle panel of Figure 1. It follows that, e.g., $H_1(0) = 0$ and hence in this example $H_1 = H_1^+$.

Example 2.6. As a variation of Example 2.5, let $G(x, y) = \exp[-e^{-x \wedge (y + \mu)}]$ for $(x, y) \in \mathbb{R}^2$ be the distribution of the random vector $(Z, Z - \mu)$, with Z a standard Gumbel variable, and with $\mu > 0$. Then the corresponding GP cdf is

$$H(x, y) = e^{-(x \wedge 0) \wedge (y \wedge 0 + \mu)} - e^{-x \wedge (y + \mu)} = e^{-(x \wedge 0) \wedge (y + \mu)} - e^{-x \wedge (y + \mu)}, \quad (2.11)$$

and H is the distribution of $(T, T - \mu)$, where T is a standard exponential variable. Now, for $-\mu < \nu$ the location transformed cdf $H(x, y + \nu)$ equals $e^{-(x \wedge 0) \wedge (y + \nu + \mu)} - e^{-x \wedge (y + \nu + \mu)}$, which is the same as in (2.6), but with μ replaced by $\nu + \mu > 0$. Hence also $H(x, y + \nu)$ is a GP cdf. The support of H is shown in the right-hand panel of Figure 1.

In Falk et al. (2010, Chapter 5), a *multivariate generalized Pareto function* is defined by

$$W(\mathbf{y}) = 1 + \log G^*(\mathbf{y}), \quad \text{for } \mathbf{y} \in (-\infty, 0]^d,$$

where G^* is a d -variate GEV distribution with reverse exponential margins, $G_j^*(y) = \exp(y)$ for $y \in (-\infty, 0]$ and $j \in \{1, \dots, d\}$. Since G^* is continuous and $\exp(dy) \leq G^*(y\mathbf{1}) \leq \exp(y)$ for $y \leq 0$, there exists $-1 \leq y_0 \leq -1/d$ such that $G^*(y_0\mathbf{1}) = \exp(-1)$. Define a GEV distribution G by $G(\mathbf{x}) = G^*(\mathbf{x} + y_0\mathbf{1})$. The margins of G are $G_j(x) = \exp(x + y_0)$ for $x \leq |y_0|$, and we have $0 < G_j(0) = \exp(y_0) < 1$ and $-\log G(\mathbf{0}) = 1$. The GP distribution H associated to G via (2.6) satisfies

$$H(\mathbf{x}) = 1 + \log G(\mathbf{x}), \quad \text{for } \mathbf{x} \in [0, \infty)^d.$$

Hence, if $\mathbf{X}$ is a random vector with cdf H , then $\mathbf{X} + y_0\mathbf{1}$ is a random vector whose cdf satisfies, for all $\mathbf{y} \in (-\infty, 0]^d$ such that $\mathbf{y} \geq y_0\mathbf{1}$, the identity

$$\begin{aligned} P(\mathbf{X} + y_0\mathbf{1} \leq \mathbf{y}) &= H(\mathbf{y} - y_0\mathbf{1}) = 1 + \log G(\mathbf{y} - y_0\mathbf{1}) \\ &= 1 + \log G^*(\mathbf{y}) = W(\mathbf{y}). \end{aligned}$$

Therefore, the restriction of the GP function W to the box $[y_0, 0]^d$ coincides with the cdf of the shifted GP random vector $\mathbf{X} + y_0\mathbf{1}$.

3 Point processes of extremes

The first step in our program for PoT inference is to specify a model for the distribution of extreme episodes in data on the real, physical scale. In this section we do this through a point process model of extreme episodes. We motivate the model by proving that the class of point processes in it coincides with the class of limits of location and scale normalized point processes of extreme values.

Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of independent random vectors with cdf F and marginal cdf-s $F_1, \dots, F_d$. Let $\mathbf{a}_n \in (0, \infty)^d$ and $\mathbf{b}_n \in \mathbb{R}^d$ be scale and location sequences such that $0 < F_j(b_{n,j}) < 1$ and $F_j(b_{n,j}) \rightarrow 1$ as $n \rightarrow \infty$, for all $j = 1, \dots, d$. We consider weak limits of the point processes

$$N_n = \sum_{i=1}^n \delta_{\mathbf{a}_n^{-1}(\mathbf{X}_i - \mathbf{b}_n)},$$

where $\delta_{\mathbf{x}}$ denotes a point mass at $\mathbf{x}$. Define $\bar{I}_j = [0, \infty)$ or $[-\infty, \infty)$ or $[-\infty, 0)$ and $I_j = (0, \infty)$ or $(-\infty, \infty)$ or $(-\infty, 0)$ according to whether $\gamma_j > 0$ or $\gamma_j = 0$ or $\gamma_j < 0$, and set

$$\bar{S}_\gamma = \bar{I}_1 \times \dots \times \bar{I}_d \quad \text{and} \quad S_\gamma = I_1 \times \dots \times I_d.$$

We use the following representation, (R) , of the limiting point process, where the letter ‘ R ’ indicates that the model is constructed on the real scale, i.e., on the space where the observations live.

- (R) Let the d -dimensional random vector $\mathbf{R} \in \bar{S}_\gamma$ satisfy $0 < \mathbb{E}[|R_j|^{1/\gamma_j}] < \infty$ if $\gamma_j \neq 0$ and $\mathbb{E}[\exp(R_j)] < \infty$ if $\gamma_j = 0$, for $j = 1, \dots, d$. Let $0 < T_1 < T_2 < \dots$ be the points of a homogeneous Poisson process on $[0, \infty)$ with unit intensity, let $(\mathbf{R}_i)_{i \geq 1}$ be independent copies of $\mathbf{R}$ which are independent of $(T_i)_{i \geq 1}$, and define the point process

$$P_r = \sum_{i \geq 1} \delta_{\mathbf{R}_i / T_i^\gamma}, \tag{3.1}$$

where, by convention, $R_{i,j}/T_i^0$ is interpreted to mean $R_{i,j} - \log T_i$.

Let $F_{\mathbf{R}}$ be the cdf of $\mathbf{R}$. For $\gamma_j \neq 0$ the moment condition in (R) can be seen to be equivalent to requiring that $0 < \int_0^\infty P(R_j > t^{\gamma_j} x) dt < \infty$, for $x \in I_j$. For $\gamma_j = 0$ the moment condition instead is equivalent to $0 < \int_0^\infty P(R_j > \log t + x) dt < \infty$, for $x \in I_j$. For example, if $\gamma_j < 0$, then $\int_0^\infty P(R_j > t^{\gamma_j} x) dt = \int_0^\infty P(|R_j|^{1/\gamma_j} > t) dt |x|^{-1/\gamma_j} = \mathbb{E}(|R_j|^{1/\gamma_j}) |x|^{-1/\gamma_j}$. Since $P(R_j > t^{\gamma_j} x_j) \leq \bar{F}_{\mathbf{R}}(t^\gamma \mathbf{x}) \leq \sum_{i=1}^d P(R_i > t^{\gamma_i} x_i)$, it in turn follows that the moment condition implies that $0 < \int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \mathbf{x}) dt < \infty$, for $\mathbf{x} \in S_\gamma$, where we have used the convention that $t^0 x$ should be replaced by $\log t + x$.

In the theorem below, the arrow ‘ $\rightarrow_d$ ’ denotes weak convergence of point processes.

Theorem 3.1. *Suppose that $F \in D(G)$ for some G . Then there exists an $\mathbf{R}$ which satisfies (R) and sequences $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n$, with $0 < F_j(b_{n,j}) < 1$ and $F_j(b_{n,j}) \rightarrow 1$ for each $j = 1, \dots, d$, such that*

$$N_n \rightarrow_d P_r \quad \text{on } S_\gamma, \quad \text{as } n \rightarrow \infty. \tag{3.2}$$

Conversely, for any point process P_r given by (3.1) there exist a GEV cdf G and sequences $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n$, with $0 < G_j(b_{n,j}) < 1$ and $G_j(b_{n,j}) \rightarrow 1$ for each $j = 1, \dots, d$, such that (3.2) holds for $F = G$.

Proof. The result is closely related to earlier results by S.I. Resnick, so we only give a brief outline of the proof.

By choosing $\mathbf{b}_n$ suitably we may without loss of generality assume that $\mathbf{a}_n^{-1}(\bigvee_{i=1}^n \mathbf{X}_i - \mathbf{b}_n)$ converges to a GEV cdf G , the margins of which have support I_j , or equivalently that the parameters satisfy $\gamma_j \mu_j = \alpha_j$ for all $j = 1, \dots, d$.

Let $\mathbf{Y} \sim G$ and define $\mathbf{Y}^*$ by $Y_j^* = |Y_j|^{1/\gamma_j}$ if $\gamma_j \neq 0$ and $Y_j^* = e^{Y_j}$ if $\gamma_j = 0$, so that $\mathbf{Y}^*$ is a GEV cdf with margins given by (2.3) with $\boldsymbol{\mu} = \boldsymbol{\alpha}$ and $\boldsymbol{\gamma} = \mathbf{1}$. It follows as in Theorem 5 of Penrose (1992) (see also de Haan and Ferreira (2006) and Schlather (2002)) that there exists a random vector $\mathbf{R}^* \in [0, \infty)^d$ with $E(R_j^*) < \infty$ such that $\mathbf{Y}^*$ has the same distribution as $\sup_{i \geq 1} \mathbf{R}_i^*/T_i$ where the random vectors $\mathbf{R}_i^*$ are i.i.d. copies of $\mathbf{R}^*$ and independent of the homogeneous Poisson process $(T_i)_{i \geq 1}$. Reversing the transformation which led from $\mathbf{Y}$ to $\mathbf{Y}^*$, it follows that $\mathbf{Y}$ has the same distribution as $\sup_{i \geq 1} \mathbf{R}_i/T_i^\gamma$, with $\mathbf{R}_i$ as in (R).

For ν the intensity measure of P_r , we have $G(\mathbf{x}) = \exp\{-\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{x}\})\}$. By standard reasoning, the weak convergence of $\mathbf{a}_n^{-1}(\bigvee_{i=1}^n \mathbf{X}_i - \mathbf{b}_n)$ is equivalent to $n P(\mathbf{a}_n^{-1}(\mathbf{X}_1 - \mathbf{b}_n) \not\leq \mathbf{x}) \rightarrow -\log G(\mathbf{x}) = \nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{x}\})$, and this in turn implies that $n P(\mathbf{a}_n^{-1}(\mathbf{X}_1 - \mathbf{b}_n) \in \cdot)$ converges vaguely to $\nu(\cdot)$. By Theorem 5.3 of Resnick (2007) this proves (3.2).

Conversely, given the point process P_r , define a d -dimensional cdf G by $G(\mathbf{x}) = P(\sup_{i \geq 1} \mathbf{R}_i/T_i^\gamma \leq \mathbf{x})$. The j -th marginal distribution, G_j , of G may be obtained as the zero-probability of the marginal Poisson point process $(R_{i,j}/T_i^{\gamma_j} : i \geq 1)$ for the interval (x, ∞) , so that (see the discussion just after Condition (R)), for x strictly between the two endpoints of I_j ,

$$-\log G_j(x) = \int_0^\infty P(R_j/t^{\gamma_j} > x) dt = \begin{cases} |x|^{-1/\gamma_j} E(|R_j|^{1/\gamma_j}) & \text{for } \gamma_j \neq 0, \\ \exp\{-x + \log E(e^{R_j})\} & \text{for } \gamma_j = 0. \end{cases}$$

Furthermore, easy calculations, such as for instance the ones in Schlather (2002, Theorem 2), show that G is max-stable. It follows that there exist sequences $\mathbf{a}_n$ and $\mathbf{b}_n$ with the stated properties such that for independent random vectors $\mathbf{X}_1, \mathbf{X}_2, \dots$ with common distribution G , the distribution of $\mathbf{a}_n^{-1}(\bigvee_{i=1}^n \mathbf{X}_i - \mathbf{b}_n)$ is equal to G too. By the first part of the theorem, this proves (3.2). $\square$

In the proof of Theorem 3.1 we obtained part of the following result, which we record here for completeness. The proof is the same as the proof of the last displayed equation in the previous proof, so we omit it. Let F_{R_j} denote the j -th marginal distribution of $F_{\mathbf{R}}$.

Corollary 3.2. *Suppose $\mathbf{R}$ satisfies (R). Then $G(\mathbf{x}) = P(\sup_{i \geq 1} \mathbf{R}_i/T_i^\gamma \leq \mathbf{x})$ is a GEV cdf and*

$$G(\mathbf{x}) = \exp \left\{ - \int_0^\infty \bar{F}_{\mathbf{R}}(r^\gamma \mathbf{x}) dr \right\} \quad \text{for } \mathbf{x} \in S_\gamma,$$

and to any GEV cdf for which $\gamma\mu = \alpha$ there exists an $\mathbf{R}$ which satisfies this equation. The marginal distributions are

$$G_j(x) = \begin{cases} e^{-x^{-1/\gamma_j} \int_0^\infty \bar{F}_{R_j}(r^{\gamma_j}) dr} = e^{-x^{-1/\gamma_j} E(R_j^{1/\gamma_j})} & \text{for } \gamma_j > 0, x \geq 0, \\ e^{-e^{-x} \int_0^\infty \bar{F}_{R_j}(\log r) dr} = e^{-e^{-x} + \log E(e^{R_j})} & \text{for } \gamma_j = 0, \\ e^{-(-x)^{-1/\gamma_j} \int_0^\infty \bar{F}_{R_j}(-r^{\gamma_j}) dr} = e^{-(-x)^{-1/\gamma_j} E(|R_j|^{1/\gamma_j})} & \text{for } \gamma_j < 0, x \leq 0, \end{cases}$$

for $j = 1, \dots, d$.

4 Representations of multivariate GP distributions

This section contains the second step in the program for PoT inference. We show how conditioning on a threshold exceedance in the point process (3.1) gives three widely useful representations of the class of multivariate GP distributions. These representations give new intuition for modelling and lead to tractable expressions for densities. In this and subsequent sections we assume that $\sigma > \mathbf{0}$.

The first representation, (R) , directly corresponds to the point process P_r in (3.1). In the second representation, (R_s) , the basic model is constructed on a standardized scale and then transformed to the scale of the observations. The third representation, $(R_{\tilde{s}})$, is equivalent to representation in Ferreira and de Haan (2014), and is obtained by adding the assumption that the random vector $\mathbf{S}$ in the (R_s) representation satisfies $\bigvee_{j=1}^d S_j = 0$.

In the literature the standard modelling scale is typically chosen as one of the following three scales: standard Pareto scale corresponding to $\gamma = \mathbf{1}$, standard Weibull scale corresponding to $\gamma = -\mathbf{1}$, or the standard Gumbel/exponential scale where $\gamma = \mathbf{0}$. Here we choose the last alternative, $\gamma = \mathbf{0}$, because of the simple additive structure which the model then acquires. It is straightforward to translate the (R_s) and $(R_{\tilde{s}})$ representations to representations based on the standard Pareto or standard Weibull scales.

The points in the limit process for (R) have the simplest and perhaps most intuitive structure which, informally, describes extreme events as a random profile, $\mathbf{R}_i$, divided componentwise by random scaling variables $T_i^{\gamma_j}$, or, if $\gamma_j = 0$, shifted by subtracting a random location variable $\log T_i$. In the (R_s) representation, the general model is instead obtained using a componentwise inverse Box–Cox transformation of the (R) model on the standard Gumbel/exponential scale, and the distribution of this limit point process then depends smoothly on the shape parameter vector γ . The $(R_{\tilde{s}})$ model leads to the simplest expressions for likelihoods and for simulation but is a bit further away from physical interpretation.

For the (R_s) representation we make the following assumption.

(R_s) Let $\mathbf{S}$ be a d -dimensional random vector taking values in $[-\infty, \infty)^d$ such that $0 < E(e^{S_j}) < \infty$ for $j = 1, \dots, d$, let $\mathbf{S}_1, \mathbf{S}_2, \dots$ be independent copies of $\mathbf{S}$, and

let $(T_i)_{i \geq 1}$ be as in (R) , and be independent of $(\mathbf{S}_i)_{i \geq 1}$. Define the point process P_s by

$$P_s = \sum_{i \geq 1} \delta_{(\exp\{\gamma(\mathbf{S}_i - \log T_i)\} - 1)/\gamma}.$$

If $\gamma = \mathbf{0}$ then P_r has the simple additive form $P_s = \sum_{i \geq 1} \delta_{\mathbf{S}_i - \log T_i}$, which is obtained as the limit of P_s as $\gamma \rightarrow 0$. Further, if $\gamma_j \neq 0$, then setting $R_{i,j} = \exp(\gamma_j S_{i,j})/\gamma_j$ it is seen that P_s is obtained by moving all the points in the corresponding process P_r componentwise a distance $1/\gamma_j$, and hence that a point process convergence theorem similar to Theorem 3.1 holds also for P_s .

We next obtain the GP representation based on (R) , Equation (4.2) below, through a limiting argument, use this argument to make a basic intuitive interpretation of the representation, and finally, in Theorem 4.1, prove that all GP distributions are obtained through this representation.

For this, we first replace P_r by a truncated point process $\bar{P}_r$ where the unit rate Poisson process $\{T_i\}$ on $\mathbb{R}_+$ is replaced by a unit rate Poisson process $\{\bar{T}_i\}$ on a bounded interval $[0, U]$. Recalling the standard representation of a Poisson process as a Poisson distributed number of $\text{Unif}[0, U]$ variables, a typical point in $\bar{N}_r$ has the form $R/\bar{T}^\gamma$ with $\bar{T} \sim \text{Unif}[0, U]$. Now, by (2.1), and using that $P(R/\bar{T}^\gamma \leq \mathbf{x}) = \frac{1}{U} \int_0^U F_{\mathbf{R}}(t^\gamma \mathbf{x}) dt$, the conditional distribution of $R/\bar{T}^\gamma - \frac{\sigma}{\gamma}$ given that $R/\bar{T}^\gamma \not\leq \frac{\sigma}{\gamma}$ is

$$\begin{aligned} & \frac{\frac{1}{U} \int_0^U \{F_{\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\sigma}{\gamma})) - F_{\mathbf{R}}(t^\gamma(\mathbf{x} \wedge \mathbf{0} + \frac{\sigma}{\gamma}))\} dt}{\frac{1}{U} \int_0^U \bar{F}_{\mathbf{R}}(t^\gamma \frac{\sigma}{\gamma}) dt} \\ &= \frac{\int_0^U \{F_{\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\sigma}{\gamma})) - F_{\mathbf{R}}(t^\gamma(\mathbf{x} \wedge \mathbf{0} + \frac{\sigma}{\gamma}))\} dt}{\int_0^U \bar{F}_{\mathbf{R}}(t^\gamma \frac{\sigma}{\gamma}) dt}. \end{aligned} \quad (4.1)$$

By the assumptions in (R) , the limit as $U \rightarrow \infty$ of this expression,

$$H_r(\mathbf{x}) = \frac{\int_0^\infty \{F_{\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\sigma}{\gamma})) - F_{\mathbf{R}}(t^\gamma(\mathbf{x} \wedge \mathbf{0} + \frac{\sigma}{\gamma}))\} dt}{\int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \frac{\sigma}{\gamma}) dt}, \quad (4.2)$$

exists (cf the discussion just before Theorem 3.1), and it is also immediate that $H_r(\infty) = 1$, so that H_r is a cdf on $[-\infty, \infty)^d$. Here, if $\gamma_i = 0$ then $t^{\gamma_i}(x_i + \frac{\sigma_i}{\gamma_i})$ should be interpreted to mean $x_i/\sigma_i + \log t$.

Theorem 4.1 below shows that the class of cdf-s given by (4.2) is the same as the class of all GP cdf-s. In the following we write $\text{GP}_r(\sigma, \gamma, F_{\mathbf{R}})$ for the distributions defined by (4.2). It should be noted that the location normalization made in Theorem 3.1 to set the support of the point process to S_γ does not restrict the generality of (4.2): an exceedance of σ/γ by $\mathbf{R}/\bar{T}^\gamma$ is the same as an exceedance of $\tilde{\sigma}/\gamma$ by $\mathbf{R}/\bar{T}^\gamma + \mathbf{c}$ if $\tilde{\sigma} = \sigma + \mathbf{c}\gamma$, for any vector $\mathbf{c}$ of constants.

The intuition is that a typical point in P_r has the form

$$\mathbf{X} = \mathbf{R}/T^\gamma,$$

where T represents a pseudo random variable with an improper uniform distribution on $(0, \infty)$. Although such a $\mathbf{X}$ therefore does not have a proper distribution, one can verify that (4.2) can be derived as if it were the conditional distribution of $\mathbf{X} - \boldsymbol{\sigma}/\gamma$, given that $\mathbf{X} \not\leq \boldsymbol{\sigma}/\gamma$. In Section 5.3 below is seen that probabilities and conditional probabilities of general events in the distribution (4.2) may be derived in the same way.

Using parallel reasoning for the limit process P_s , the (formal) conditional cdf of

$$\frac{\boldsymbol{\sigma}}{\gamma}(\exp\{\gamma(\mathbf{S} - \log T)\} - \mathbf{1})$$

given that $\frac{\boldsymbol{\sigma}}{\gamma}(\exp\{\gamma(\mathbf{S} - \log T)\} - \mathbf{1}) \not\leq \mathbf{0}$ equals

$$H_s(\mathbf{x}) = \frac{\int_0^\infty \{F_{\mathbf{S}}(\frac{1}{\gamma} \log(\frac{\gamma}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1}) + \log t) - F_{\mathbf{S}}(\frac{1}{\gamma} \log(\frac{\gamma}{\boldsymbol{\sigma}}(\mathbf{x} \wedge \mathbf{0}) + \mathbf{1}) + \log t)\} dt}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt}, \quad (4.3)$$

for $\mathbf{x}$ such that $\frac{\gamma}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1} > \mathbf{0}$, and where $F_{\mathbf{S}}$ is the cdf of $\mathbf{S}$. We write $\text{GP}_s(\boldsymbol{\sigma}, \gamma, F_{\mathbf{S}})$ for the distribution defined by (4.3). Again, in Theorem 4.1 it is proved that H_s indeed is a distribution function, and that the class of distribution functions (4.3) equals the class of GP cdf-s. For later use, note that in this construction, a general GP_s vector is obtained from a $\text{GP}_s(\mathbf{1}, \mathbf{0}, F_{\mathbf{S}})$ vector $\mathbf{X}_0$ through the transformation

$$\mathbf{X} = \boldsymbol{\sigma} \frac{e^{\gamma \mathbf{X}_0} - \mathbf{1}}{\gamma} \quad (4.4)$$

For $\boldsymbol{\sigma} = \mathbf{1}$ and $\gamma = \mathbf{0}$ the cdf-s H_s and H_r are the same, and are given by

$$H_s(\mathbf{x}) = \frac{\int_0^\infty \{F_{\mathbf{S}}(\mathbf{x} + \log t) - F_{\mathbf{S}}(\mathbf{x} \wedge \mathbf{0} + \log t)\} dt}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt}. \quad (4.5)$$

Suppose now that $\mathbf{S} = \tilde{\mathbf{S}}$ where $\tilde{\mathbf{S}}$ satisfies $\bigvee_{j=1}^d \tilde{S}_j = 0$ and that $\boldsymbol{\sigma} = \mathbf{1}$. It is straightforward to see that then $F_{\tilde{\mathbf{S}}}(\mathbf{x} + \log t) - F_{\tilde{\mathbf{S}}}(\mathbf{x} \wedge \mathbf{0} + \log t) = \mathbf{1}_{\{0 \leq t < 1\}} F_{\tilde{\mathbf{S}}}(\mathbf{x} + \log t)$, and, in particular, that $\bar{F}_{\tilde{\mathbf{S}}}(\log t) = \mathbf{1}_{\{0 \leq t < 1\}}$. Inserting this into (4.5) then gives the cdf

$$H_{\tilde{s}}(\mathbf{x}) = \int_0^1 F_{\tilde{\mathbf{S}}}(\mathbf{x} + \log t) dt = \int_0^\infty F_{\tilde{\mathbf{S}}}(\mathbf{x} - v) e^{-v} dv, \quad (4.6)$$

where the second equality follows from the change of variable $\log t = -v$. Further, using the transformation (4.4) which connects (4.5) with (4.3), it follows more generally that if $\mathbf{S} = \tilde{\mathbf{S}}$, where $\bigvee_{j=1}^d \tilde{S}_j = 0$, then for general $\boldsymbol{\sigma}, \gamma$ one obtains the cdf

$$H_{\tilde{s}}(\mathbf{x}) = \int_0^1 F_{\tilde{\mathbf{S}}} \left(\frac{1}{\gamma} \log \left(\frac{\gamma}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1} \right) + \log t \right) dt = \int_0^\infty F_{\tilde{\mathbf{S}}} \left(\frac{1}{\gamma} \log \left(\frac{\gamma}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1} \right) - v \right) e^{-v} ds. \quad (4.7)$$

We write $\text{GP}_s(\boldsymbol{\sigma}, \gamma, F_{\tilde{\mathbf{S}}})$ for the distribution (4.7), and call it the $(R_{\tilde{s}})$ representation.

In Section 5 below, the random vector $\tilde{\mathbf{S}}$ will be constructed from a general random vector $\mathbf{S}$ by setting $\tilde{\mathbf{S}} = \mathbf{S} - \bigvee_{j=1}^d S_j$. Further, the last expression in (4.6) can be given

an interpretation in terms of random variables: it is the cdf of $\tilde{S} + E$, where E is a standard exponential variable which is independent of $\tilde{S}$, and then (4.7) is the cdf of $\frac{\sigma}{\gamma}(e^{\gamma(\tilde{S}+E)} - 1)$. This is the representation derived in Ferreira and de Haan (2014), translated to the exponential scale.

Theorem 4.1. *Suppose $\sigma > \mathbf{0}$. The classes $GP_r(\sigma, \gamma, F_{\mathbf{R}})$, $GP_s(\sigma, \gamma, F_{\mathbf{S}})$, and $GP_s(\sigma, \gamma, F_{\tilde{\mathbf{S}}})$ defined by (4.2), (4.3), and (4.7), are all equal to the class of all GP distributions with $\sigma > \mathbf{0}$. The conditional marginal distributions have the GP form (2.7), for all three classes.*

Proof. By Corollary 3.2, to any GEV cdf G with $\gamma\mu = \alpha$ there exists a random vector $\mathbf{R}$ which satisfies (R) and for which $G_0(\mathbf{x}) = \exp\{-\int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \mathbf{x}) dt\}$, for $\mathbf{x} \in S_\gamma$. The general GEV distribution is thus obtained from a location transformation of G_0 , and by a suitable choice of σ may be written as $G(\mathbf{x} + \frac{\sigma}{\gamma}) = \exp\{-\int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\sigma}{\gamma})) dt\}$. Further, requiring that $0 < G_0(\frac{\sigma}{\gamma}) < 1$ can be seen to be equivalent to $\sigma > \mathbf{0}$. Inserting this into (2.6) shows that the class of cdf-s $GP_r(\sigma, \gamma, F_{\mathbf{R}})$ is equal to the class of all GP distributions with $\sigma > \mathbf{0}$.

By definition, the class of $GP_s(\sigma, \gamma, F_{\mathbf{S}})$ cdf-s is the same as the class of $GP_r(\sigma, \gamma, F_{\mathbf{R}})$ cdf-s, and hence the same conclusion holds for the $GP_s(\sigma, \gamma, F_{\mathbf{S}})$ cdf-s. Since the $GP_s(\sigma, \gamma, F_{\tilde{\mathbf{S}}})$ cdf-s are $GP_s(\sigma, \gamma, F_{\mathbf{S}})$ cdf-s, it follows that they are GP distributions.

To prove the full statement about the class $GP_s(\sigma, \gamma, F_{\tilde{\mathbf{S}}})$ we first note that by the construction of the $GP_s(\sigma, \gamma, F_{\tilde{\mathbf{S}}})$ cdf-s, it is enough to prove that the statement holds for GP distributions with $\gamma = \mathbf{0}$ and $\sigma = \mathbf{1}$. However, then the result follows by combining (T2) with the discrete version of de Haan and Ferreira (2006), Theorem 2.1, translated to the exponential scale (i.e with $\mathbf{W}$ replaced by e^{bS+e}), and with $\omega_0 = 1$.

It is straightforward to verify that the conditional marginal distributions have the form (2.7). As an example we show the calculations for the $GP_r(\sigma, \gamma, F_{\mathbf{R}})$ class for the case $\gamma_j \neq 0$; $j = 1, \dots, d$. Let F_j be the marginal distribution of the j -th component of $\mathbf{R}$. It follows from (2.8) and (4.2) that the distribution of the j -th component of H_r conditional on being positive, denoted H^+ , is given by

$$H_j^+(x) = 1 - \frac{\int_0^\infty \bar{F}_j(t^{\gamma_j}(x + \frac{\sigma_j}{\gamma_j})) dt}{\int_0^\infty \bar{F}_j(t^{\gamma_j} \frac{\sigma_j}{\gamma_j}) dt} = 1 - (1 + \frac{\gamma_j}{\sigma_j} x)^{-1/\gamma_j},$$

where the second equality follows from making the change of variables $t(\frac{\gamma_j}{\sigma_j}x + 1)^{1/\gamma_j} \rightarrow t$ in the numerator. $\square$

It may again be noted that since $F_{\mathbf{R}}$, $F_{\mathbf{S}}$, and $F_{\tilde{\mathbf{S}}}$ are cdf-s then also H_r , H_s , and $H_{\tilde{s}}$ are cdf-s, so that in contrast to equation (2.6) the equations (4.2), (4.3), and (4.5) hold for all $\mathbf{x} \in (-\infty, \infty)^d$, subject to the extra provision that (4.3) and (4.5) only apply for $\gamma\mathbf{x} + \sigma > \mathbf{0}$.

5 Likelihoods, censored likelihoods, and conditional probabilities for multivariate generalized Pareto distributions

The central aim of this paper is to provide the basis for likelihood inference in multivariate peaks over thresholds modelling. In this section we find the densities and censored likelihoods which are needed for this, discuss parametrization issues, and give formulas for probabilities and conditional probabilities of general sets.

5.1 Densities and censored densities

To derive the densities for the (R) and (R_s) models we use the assumption that the random vectors $\mathbf{R}$ and $\mathbf{S}$ have a density with respect to Lebesgue measure on $\mathbb{R}^d$, and for the $(R_{\tilde{s}})$ model we assume that $\tilde{\mathbf{S}}$ is obtained as $\tilde{\mathbf{S}} = \mathbf{S} - \bigvee_{j=1}^d S_j$, where $\mathbf{S}$ has a density with respect to Lebesgue measure on $\mathbb{R}^d$. The form of the densities for (R) and (R_s) are just as would be obtained if the $\mathbf{R}$ and $\mathbf{S}$ cdf-s were continuously differentiable and interchange of differentiation and integration was allowed. However they in fact do not require any assumptions beyond absolute continuity with respect to d -dimensional Lebesgue measure.

The support of the vector $\tilde{\mathbf{S}}$ is contained in the $(d-1)$ dimensional subset $\{\mathbf{s} \in \mathbb{R}^d \mid \bigvee_{j=1}^d s_j = 0\}$ and hence $\tilde{\mathbf{S}}$ does not have a density with respect to Lebesgue measure on $\mathbb{R}^d$. The requirement made in (R_s) of finite exponential moment is not necessary for an $\mathbf{S}$ used to construct $\tilde{\mathbf{S}}$ since $\tilde{\mathbf{S}}$ always has finite exponential moment, so in $(R_{\tilde{s}})$ the only requirement on $\mathbf{S}$ is that it has a density.

Theorem 5.1. *Suppose $\boldsymbol{\sigma} > \mathbf{0}$. If $F_{\mathbf{R}}$ has density $f_{\mathbf{R}}$ with respect to Lebesgue measure on $\mathbb{R}^d$, then H_r has the density h_r given below, and if $F_{\mathbf{S}}$ has density $f_{\mathbf{S}}$ with respect to Lebesgue measure on $\mathbb{R}^d$, then H_s and $H_{\tilde{s}}$ have the densities h_s and $h_{\tilde{s}}$ below:*

$$h_r(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{1}{\int_0^\infty \bar{F}_{\mathbf{R}}(t\boldsymbol{\gamma} \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt} \int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\mathbf{R}}(t\boldsymbol{\gamma}(\mathbf{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})) dt, \quad (5.1)$$

$$h_s(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{\prod_{j=1}^d (\gamma_j x_j + \sigma_j)^{-1}}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt} \int_0^\infty f_{\mathbf{S}}(\frac{1}{\boldsymbol{\gamma}} \log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1}) + \log t) dt, \quad (5.2)$$

$$h_{\tilde{s}}(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{\prod_{j=1}^d (\gamma_j x_j + \sigma_j)^{-1}}{\bigvee_{j=1}^d (\frac{\gamma_j}{\sigma_j} x_j + 1)^{1/\gamma_j}} \int_0^\infty t^{-1} f_{\mathbf{S}}(\frac{1}{\boldsymbol{\gamma}} \log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}} \mathbf{x} + \mathbf{1}) + \log t) dt, \quad (5.3)$$

for $\boldsymbol{\gamma} \mathbf{x} + \boldsymbol{\sigma} > \mathbf{0}$, and where the densities are 0 otherwise. If $\gamma_j = 0$ then for h_r the argument $t^{\gamma_j} \frac{\sigma_j}{\gamma_j}$ in $F_{\mathbf{R}}$ should be replaced by $\log t$ and the argument $t^{\gamma_j} (x_j + \frac{\sigma_j}{\gamma_j})$ in $f_{\mathbf{R}}$ should be replaced by $x_j/\sigma_j + \log t$. Similarly, for h_s and h_{S_0} , if $\gamma_j = 0$, the expressions $\frac{1}{\gamma_j} \log(\frac{\gamma_j}{\sigma_j} x_j + 1)$ should be replaced by their limits x_j/σ_j .

Proof. We first prove (5.1) for the special case when $\boldsymbol{\sigma} = \boldsymbol{\gamma} = \mathbf{1}$, and for $\mathbf{x} \not\leq \mathbf{0}$, $\mathbf{x} + \mathbf{1} > \mathbf{0}$.

The change of variables $\mathbf{y} = t(\mathbf{z} + \mathbf{1})$ shows that for this case

$$\begin{aligned} F_{\mathbf{R}}(t(\mathbf{x} + \mathbf{1})) - F_{\mathbf{R}}(t(\mathbf{x} \wedge \mathbf{0} + \mathbf{1})) &= \int_{\mathbf{y}=t(\mathbf{x} \wedge \mathbf{0} + \mathbf{1})}^{t(\mathbf{x} + \mathbf{1})} f_{\mathbf{R}}(\mathbf{y}) d\mathbf{y} \\ &= \int_{\mathbf{z}=\mathbf{x} \wedge \mathbf{0}}^{\mathbf{x}} t^d f_{\mathbf{R}}(t(\mathbf{z} + \mathbf{1})) d\mathbf{z}. \end{aligned}$$

Hence, by (4.2), and using Fubini's theorem,

$$\begin{aligned} H_r(\mathbf{x}) &= \frac{\int_{t=0}^{\infty} \int_{\mathbf{z}=\mathbf{x} \wedge \mathbf{0}}^{\mathbf{x}} t^d f_{\mathbf{R}}(t(\mathbf{z} + \mathbf{1})) d\mathbf{z} dt}{\int_0^{\infty} F_{\mathbf{R}}(t\mathbf{1}) dt} \\ &= \int_{\mathbf{z}=\mathbf{x} \wedge \mathbf{0}}^{\mathbf{x}} \frac{\int_{t=0}^{\infty} t^d f_{\mathbf{R}}(t(\mathbf{z} + \mathbf{1})) dt}{\int_0^{\infty} F_{\mathbf{R}}(t\mathbf{1}) dt} d\mathbf{z}, \end{aligned}$$

which proves that (5.1) holds for $\sigma = \gamma = 1$. The proof of the general form of (5.1) only differs from this case in bookkeeping details, and is omitted.

To prove (5.2), recall that $H_s = H_r$ if $\gamma = \mathbf{0}$, $\sigma = \mathbf{1}$, so that it follows from (5.1) that then

$$h_s(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{\int_{t=0}^{\infty} f_{\mathbf{S}}(\mathbf{x} + \log t) dt}{\int_0^{\infty} F_{\mathbf{S}}(\log t) dt}.$$

Writing $\tilde{H}$ for the corresponding cdf, the general cdf H_s is obtained as $H_s(\mathbf{x}) = \tilde{H}(\frac{1}{\gamma} \log(\frac{\gamma}{\sigma} \mathbf{x} + \mathbf{1}))$ and (5.2) then follows by a chain rule type argument.

We again only prove (5.3) for the case $\sigma = \mathbf{1}, \gamma = \mathbf{0}$, and with $\mathbf{x} \not\leq \mathbf{0}$. Also here extension to the general case is a straightforward chain rule argument. By (4.6) we have that

$$H_{\tilde{s}}(\mathbf{x}) = \int_{t=0}^1 F_{\tilde{\mathbf{S}}}(\mathbf{x} + \log t) = \int_{t=0}^{\infty} \int_{\mathbf{s}} \mathbf{1}_{\{0 \leq t \leq 1, \mathbf{s} - \vee_{j=0}^d s_j \leq \mathbf{x} + \log t\}} f_{\mathbf{S}}(\mathbf{s}) d\mathbf{s} dt.$$

Hence, using first Fubini's theorem, then the change of variables $te^{\vee_{j=1}^d s_j} \rightarrow t$, then Fubini's theorem and the change of variables $\mathbf{s} \rightarrow \mathbf{s} + \log t$, and then Fubini's theorem once more, we have that

$$\begin{aligned} H_{\tilde{s}}(\mathbf{x}) &= \int_{\mathbf{s}} \int_{t=0}^{\infty} e^{-\vee_{j=1}^d s_j} \mathbf{1}_{\{0 \leq te^{-\vee_{j=1}^d s_j} \leq 1, \mathbf{s} \leq \mathbf{x} + \log(t)\}} f_{\mathbf{S}}(\mathbf{s}) dt d\mathbf{s} \\ &= \int_{t=0}^{\infty} \int_{\mathbf{s}} e^{-\vee_{j=1}^d s_j} t^{-1} \mathbf{1}_{\{e^{-\vee_{j=1}^d s_j} \leq 1, \mathbf{s} \leq \mathbf{x}\}} f_{\mathbf{S}}(\mathbf{s} + \log t) d\mathbf{s} dt \\ &= \int_{-\infty}^{\mathbf{x}} \mathbf{1}_{\{\mathbf{s} \not\leq \mathbf{0}\}} e^{-\vee_{j=1}^d s_j} \int_{t=0}^{\infty} t^{-1} f_{\mathbf{S}}(\mathbf{s} + \log t) dt d\mathbf{s}. \end{aligned}$$

This proves that (5.3) holds for $\gamma = \mathbf{0}$ and $\sigma = \mathbf{1}$. □

In $(R_{\tilde{s}})$ the distribution on $\{\mathbf{s} \in \mathbb{R}^d \mid \vee_{j=1}^d s_j = 0\}$ given by $\tilde{\mathbf{S}}$ is obtained by starting with an absolutely continuous distribution on $\mathbb{R}^d$ and then subtracting the maximal

component. It is also possible avoid the detour via $\mathbb{R}^d$ and obtain a GP variable directly from a vector $\tilde{\mathbf{S}}$ which has a density on $\{\mathbf{s} \in \mathbb{R}^d \mid \bigvee_{j=1}^d s_j = 0\}$. Also for this construction it is possible to find the density of $\mathbf{X}$, by arguments as above. We omit the details.

In some cases the integrals in the numerators of (5.1) - (5.3) can be computed explicitly, see the examples below. When this is not possible, the one-dimensional integrals in the densities allow for fast numerical computation as soon as one can compute densities and distribution functions of $\mathbf{R}$ (or $\mathbf{S}$) efficiently. Either way, this can make full likelihood inference possible, also in high dimensions.

Sometimes one doesn't trust the GP distribution to hold on the entire set $\mathbf{x} \not\leq \mathbf{0}$. Then, instead of using a full likelihood obtained as a product of the densities in Theorem 5.1, one can use a censored likelihood which is based on the values of the excesses which exceed some censoring threshold $\mathbf{v} = (v_1, \dots, v_d)$. Write $D = \{1, \dots, d\}$, and let $C \subset D$ be the set of indices which correspond to the components which are censored, i.e. which do not exceed their censoring threshold v_i . Then, using the notation $\mathbf{x}_A = \{x_j; j \in A\}$ and writing h for h_r, h_s or $h_{\tilde{s}}$, the likelihood contribution of the censored observation is

$$h_C(\mathbf{x}_{D \setminus C}) = \int_{\{x_j \in (-\infty, v_j]; j \in C\}} h(\mathbf{x}) d\mathbf{x}_C. \quad (5.4)$$

Example 5.2. The simplest situation is when the components of $\mathbf{R}$ (or $\mathbf{S}$) are independent. This could e.g. be a model for windspeeds over a small area, perhaps a windmill park, with T^γ representing the average (geostrophic) wind and the components being "nuggets": random wind variations caused by local turbulence.

Let f_j be the density function of R_j , the j th component of $\mathbf{R}$, let F_j be the corresponding cdf, write $y_j = x_j + \sigma_j/\gamma_j$, and assume that $v_j \leq 0, j \in C$. The integral which appears in numerator in (5.4) for $h = h_r$ can then be written as

$$\mathbb{1}_{\{\mathbf{x}_{D \setminus C} \not\leq \mathbf{0}\}} \int_0^\infty t^{\sum_{j \in D \setminus C} \gamma_j} \prod_{j \in C} F_j(t^{\gamma_j} v_j) \prod_{j \in D \setminus C} f_j(t^{\gamma_j} y_j) dt$$

and the integral in the denominator equals $\int_0^\infty \{1 - \prod_{j=1}^d F_j(t^{\gamma_j} \sigma_j/\gamma_j)\} dt$. Here both integrands typically can be computed quickly, and then also numerical computation of h_C will be fast.

In a number of cases it is also possible to compute these integrals explicitly, and similarly for the corresponding integrals for $h = h_s$ and $h_{\tilde{s}}$. As a simple example consider (5.3) with $\boldsymbol{\gamma} = \mathbf{0}, \boldsymbol{\sigma} = \mathbf{1}$ and with the components of $\mathbf{S}$ having independent standard Gumbel distributions with cdf $\exp\{-e^{-x}\}$, and consider $h_{\tilde{s}}$. Then, writing c for the number of elements in C , i.e. for the number of uncensored components, and abbreviating

$\mathbb{1}_{\{\mathbf{x}_{D \setminus C} \not\leq \mathbf{0}\}}$ to $\mathbb{1}_{D \setminus C}$, we obtain that

$$\begin{aligned}
h_C(\mathbf{x}) &= \mathbb{1}_{D \setminus C} e^{-\vee_{j=1}^d x_j} \int_0^\infty \prod_{j \in D \setminus C} e^{-x_j - \log t} \exp\{-e^{-x_j - \log t}\} \prod_{j \in C} \exp\{-e^{-v_j - \log t}\} dt \\
&= \mathbb{1}_{D \setminus C} e^{-\vee_{j=1}^d x_j - \sum_{j \in D \setminus C} x_j} \int_0^\infty t^{-(d-c)} \exp\{-t^{-1} (\sum_{j \in D \setminus C} e^{-x_j} + \sum_{j \in C} e^{-v_j})\} dt \\
&= \mathbb{1}_{D \setminus C} (d-c-2)! e^{-\vee_{j=1}^d x_j - \sum_{j \in D \setminus C} x_j} \left(\sum_{j \in D \setminus C} e^{-x_j} + \sum_{j \in C} e^{-v_j} \right)^{-(d-c)+1}.
\end{aligned}$$

Example 5.3. An extreme flow episode in a river network consisting of two tributaries which join to form the main river could be modeled as $\mathbf{R}/T^\gamma = (R_1/T^\gamma, R_2/T^\gamma, (R_1 + R_2 + E)/T^\gamma)$, with $\gamma > 0$, so that $R_3 = R_1 + R_2 + E$. Here the first component corresponds to flow in tributary one, the second component corresponds to flow in tributary two, and the third component is flow in the main river. The very simplest model is that R_1, R_2, E are independent and have a standard exponential distribution. Then,

$$\begin{aligned}
\int_0^\infty t^{3\gamma} f_{\mathbf{R}}(t^\gamma \mathbf{y}) dt &= \mathbb{1}_{\{\mathbf{0} \leq \mathbf{y}, y_1 + y_2 \leq y_3\}} \int_0^\infty t^{3\gamma} e^{-t^\gamma y_3} dt \\
&= \mathbb{1}_{\{\mathbf{0} \leq \mathbf{y}, y_1 + y_2 \leq y_3\}} \gamma^{-1} \Gamma(3 + 1/\gamma) y_3^{-3-1/\gamma}, \tag{5.5}
\end{aligned}$$

and, additionally assuming that $\boldsymbol{\sigma} = (\sigma, \sigma, \sigma)$,

$$\begin{aligned}
\int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \boldsymbol{\sigma}/\gamma) dt &= E(\vee_{j=1}^3 (R_j \gamma/\sigma)^{1/\gamma}) \\
&= (\gamma/\sigma)^{1/\gamma} E(R_3^{1/\gamma}) \\
&= (\gamma/\sigma)^{1/\gamma} \Gamma(3 + 1/\gamma)/2,
\end{aligned}$$

since R_3 is a sum of three exponential variables, and thus has a gamma distribution. It follows from (5.1) and (5.5) with $\mathbf{y} = \mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma}$ that

$$h_r(\mathbf{x}) = \mathbb{1}_{\{\mathbf{x} \not\leq \mathbf{0}, -\boldsymbol{\sigma}/\gamma \leq \mathbf{x}, x_1 + x_2 \leq x_3\}} \gamma^{-1} 2(\sigma/\gamma)^{1/\gamma} (x_3 + \sigma/\gamma)^{-3-1/\gamma}.$$

Example 5.4. Variants of log Gaussian distributions have been used in max-stable modelling e.g. in Huser and Davison (2013), and as a point process model in Wadsworth and Tawn (2014), and they provide an important class of models. As an example, in the (R) model suppose that $\mathbf{0} \leq \boldsymbol{\gamma}$ and that $F_{\mathbf{R}}(\mathbf{x}) = \Phi(\log \mathbf{x})$ where Φ is the d.f. of a d -dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and nonsingular covariance matrix Σ . Write ϕ for the corresponding density and let $A = \Sigma^{-1}$ be the precision matrix. Then, writing $\mathbf{y} = \log(\mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma}) - \boldsymbol{\mu}$,

$$\begin{aligned}
\int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\mathbf{R}}(t^\gamma (\mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma})) dt &= \frac{1}{\prod_{j=1}^d (x_j + \frac{\sigma_j}{\gamma_j})} \int_0^\infty \phi((\gamma \log t + \log(\mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma}))) dt \\
&= \frac{1}{\prod_{j=1}^d (x_j + \frac{\sigma_j}{\gamma_j})} \frac{|A|^{1/2}}{(2\pi)^{d/2}} \int_0^\infty \exp[-\frac{1}{2}(\gamma \log t + \mathbf{y})A(\gamma \log t + \mathbf{y})'] dt.
\end{aligned}$$

Making the change of variables $\log t \rightarrow t$ and completing the square, the integral can be evaluated, to show that

$$h_r(\mathbf{x}) = \mathbb{1}_{\{\mathbf{x} \neq \mathbf{0}\}} \frac{|A|^{1/2}}{[(2\pi)^{(d-1)} \gamma A \gamma']^{1/2}} \frac{1}{\prod_{j=1}^d (x_j + \frac{\sigma_j}{\gamma_j})} \frac{\exp[-\frac{1}{2}(\mathbf{y} A \mathbf{y}' - \frac{(\gamma A \mathbf{y}' - 1)^2}{\gamma A \gamma'})]}{\int_0^\infty \bar{\Phi}[\gamma \log(t) + \log(\boldsymbol{\sigma}/\gamma)] dt}.$$

The integral in the denominator can be expressed as a sum of d components, each of which is a $d - 1$ dimensional Gaussian cdf, see Huser and Davison (2013). However, if d is large then this expression is cumbersome. It is an interesting challenge to develop methods to compute the denominator efficiently for large d .

5.2 Parametrization

Each of the models $GP_r(\boldsymbol{\sigma}, \gamma, F_{\mathbf{R}})$, $GP_s(\boldsymbol{\sigma}, \gamma, F_{\mathbf{S}})$ and $GP_s(\boldsymbol{\sigma}, \gamma, F_{\hat{\mathbf{S}}})$ contains all the GP cdf-s with $\boldsymbol{\sigma} > \mathbf{0}$, so which one to use for likelihood inference is a choice of parametrization which is guided by questions of modelling realism and computational efficiency. In this section we discuss a number of issues connected with the parametrization of GP distributions.

The vectors $\mathbf{R}$ and $\mathbf{S}$ are not uniquely determined by the corresponding GP distributions, as is seen in the following result which is a generalization of Proposition 2.1(iv). It follows that if the model for $F_{\mathbf{S}}$ includes free location parameters for each component, then the model is overparametrized, and one should, e.g., fix the location parameter for one of the components to zero. Similarly, if $\gamma_i \neq 0$ for $i = 1, \dots, d$ and if the model for $F_{\mathbf{R}}$ includes a free scale parameter for each component, then one should e.g. fix one of these scale parameters to 1.

Proposition 5.5. *Suppose that the random variable Z is strictly positive and has finite mean. Then $GP_s(\boldsymbol{\sigma}, \gamma, F_{\mathbf{S} + \log Z}) = GP_s(\boldsymbol{\sigma}, \gamma, F_{\mathbf{S}})$, and $GP_r(\boldsymbol{\sigma}, \gamma, F_{Z\gamma\mathbf{R}}) = GP_r(\boldsymbol{\sigma}, \gamma, F_{\mathbf{R}})$, where $Z^{\gamma_i} R_i$ should be interpreted to mean $R_i + \log Z$ if $\gamma_i = 0$.*

Proof. The first assertion is an immediate consequence of the second one, so we only prove this. Now, conditioning on Z and using Fubini's theorem and the change of variables $t/c \rightarrow t$ shows that

$$\begin{aligned} \int_{t=0}^{\infty} \bar{F}_{Z\gamma\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\gamma}) dt &= \int_{t=0}^{\infty} \int_{c=0}^{\infty} \bar{F}_{\mathbf{R}}((t/c)^\gamma \frac{\boldsymbol{\sigma}}{\gamma}) dF_Z(c) dt \\ &= \int_{c=0}^{\infty} \int_{t=0}^{\infty} \bar{F}_{\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\gamma}) c dt dF_Z(c) \\ &= E(Z) \int_{t=0}^{\infty} \bar{F}_{\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\gamma}) dt. \end{aligned}$$

The same argument shows that

$$\begin{aligned} \int_0^\infty \{F_{Z\gamma\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma})) - F_{Z\gamma\mathbf{R}}(t^\gamma(\mathbf{x} \wedge \mathbf{0} + \frac{\boldsymbol{\sigma}}{\gamma}))\} dt \\ = E(Z) \int_0^\infty \{F_{\mathbf{R}}(t^\gamma(\mathbf{x} + \frac{\boldsymbol{\sigma}}{\gamma})) - F_{\mathbf{R}}(t^\gamma(\mathbf{x} \wedge \mathbf{0} + \frac{\boldsymbol{\sigma}}{\gamma}))\} dt, \end{aligned}$$

and inserting these two equations into (4.2) proves the proposition. $\square$

Computationally the (R_s) densities typically are simplest, and the (R_s) densities may also have some computational advantage over the (R) densities since the integral in the denominator in the (R_s) densities does not depend on the parameters σ, γ . Instead, understanding of how well a model describes a physical situation is best done on the real physical scale, and often requirements of realistic physical modelling lead to use of the (R) model.

One can gain insight into the realism of an (R_s) model by rewriting it as an (R) model. Using the intuition described after the derivation of (4.2), exceedances of $\mathbf{0}$ by $\frac{\sigma}{\gamma}(e^{\gamma(\mathbf{S}-\log T)} - 1)$ are the same as exceedances of $\frac{\sigma}{\gamma}$ by $\frac{\sigma}{\gamma}e^{\gamma\mathbf{S}}/T^\gamma$ which in turn are the same as exceedances of $\frac{\sigma}{\gamma}$ by $\text{sign}(\gamma)e^{\gamma\mathbf{S}+\log(\sigma/|\gamma|)}/T^\gamma$. Thus, if one sets $\mathbf{R}$ equal to $\text{sign}(\gamma)e^{\gamma\mathbf{S}+\log(\sigma/|\gamma|)}$ then, heuristically, the (R) and (R_s) models are the same. It is easy to prove that this in fact is correct, i.e. that this choice of $\mathbf{R}$ makes the righthand side of (4.2) equal to the righthand side (4.3), by checking the inequalities componentwise, separately for the cases $\gamma_j > 0$, $\gamma_j = 0$ and $\gamma_j < 0$, and using the standard conventions for the case $\gamma_j = 0$.

As one consequence of this, if the parametric model for $\mathbf{S}$ doesn't include location parameters, then $e^{\log(\sigma/|\gamma|)}$ determines the relative sizes of the components of the extreme episode. However, σ is strongly influenced by the choice of thresholds and γ is determined by the tails of the marginal distributions, and hence it doesn't seem likely that $e^{\log(\sigma/|\gamma|)}$ will catch the relative component sizes well, and it thus often is desirable to have the model for $\mathbf{S}$ include location parameters.

A main issue is how estimates and confidence intervals for quantities like high quantiles perform when one or more of the γ_j -s could be either positive or negative. For the (R_s) and (R_s) model the transition from positive to negative γ -s is smooth, and such estimates and confidence intervals will work automatically.

The distributions of marginal excesses in (R) models are the same as in the (R_s) models, and hence inference about γ and σ again will work for γ_j -s which could be both positive and negative. However, in (R) models one may have to restrict the allowed range of γ , e.g. to $\gamma > \mathbf{0}$ to obtain useful estimates and confidence intervals for functions of components, such as high quantiles of maxima or sums of components.

One example where this transition from positive to negative γ -s is not smooth in the (R) model is the river network in Example 5.3, where it isn't even clear how to define the model in a useful way for negative γ -s. However, this model cannot be constructed at all as an (R_s) model.

As another example, if one uses a location-scale model for $\mathbf{S}$ in the (R_s) model, so that $\mathbf{S} = \mathbf{s}_s\mathbf{S}^* + \mathbf{m}_s$, where the model for $\mathbf{S}^*$ usually would include additional parameters, then the (R) model $\mathbf{R} = \text{sign}(\gamma)e^{\mathbf{s}_r\mathbf{S}^*+\mathbf{m}_r}$ is just a reparametrization $\gamma\mathbf{s}_s \leftrightarrow \mathbf{s}_r$ and $\gamma\mathbf{m}_s + \log(\sigma/|\gamma|) \leftrightarrow \mathbf{m}_r$ of the (R_s) model. Hence maximum likelihood estimators, say of high quantiles of functions of the components, are the same in both models, and, e.g., delta method confidence intervals are asymptotically the same. However, non-asymptotically, if confidence intervals in the (R) model includes both negative and

positive values for some γ_i then confidence interval for functions of components may become unreasonable. Hence, again, in the (R) model, it may be safest to make the sign of the components γ a modelling assumption.

5.3 Probabilities and conditional probabilities

Equations (4.2), (4.3), and (4.7) give the probabilities of rectangles for GP distributions. In this section they are generalized to give expression for probabilities of general sets, and to expressions for conditional probabilities. We only consider the GP_r models, since such expressions are usually needed on the real scale.

Let $F = \{\mathbf{y}; \mathbf{y} \not\leq \mathbf{0}\}$, set $A = \{\mathbf{y}; \mathbf{y} \leq \mathbf{x}\}$, and for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and a set $B \subset \mathbb{R}^d$ write $\mathbf{a}B + \mathbf{b}$ for the set $\{\mathbf{a}\mathbf{y} + \mathbf{b}; \mathbf{y} \in B\}$. As is easily checked,

$$H_r(\mathbf{x}) = H_r(A) = \frac{\int_0^\infty P(\mathbf{R} \in t^\gamma(A \cap F + \boldsymbol{\sigma}/\gamma)) dt}{\int_0^\infty P(\mathbf{R} \in t^\gamma(F + \boldsymbol{\sigma}/\gamma)) dt}.$$

Now, if in the derivation of (4.2) the special A defined above is replaced by a general set $A \subset \mathbb{R}^d$, the result still is the same,

$$H_r(A) = \frac{\int_0^\infty P(\mathbf{R} \in t^\gamma(A \cap F + \boldsymbol{\sigma}/\gamma)) dt}{\int_0^\infty P(\mathbf{R} \in t^\gamma(F + \boldsymbol{\sigma}/\gamma)) dt} = \frac{\int_0^\infty P(\mathbf{R}/t^\gamma - \boldsymbol{\sigma}/\gamma \in A \cap F) dt}{\int_0^\infty P(\mathbf{R}/t^\gamma - \boldsymbol{\sigma}/\gamma \in F) dt}. \quad (5.6)$$

A proof that (5.6) holds for any set A is immediate: using Fubini's theorem it is seen that the righthand side of the equation is a probability distribution, as function of A , and since it agrees with the distribution H_r on sets of the form $\{\mathbf{y}; \mathbf{y} \leq \mathbf{x}\}$, the two distributions are equal. The intuition is that $H_r(A)$ is the (formal) conditional probability of the event $\{\mathbf{R}/T^\gamma - \boldsymbol{\sigma}/\gamma \in A\}$ given the event $\{\mathbf{R}/T^\gamma - \boldsymbol{\sigma}/\gamma \not\leq \mathbf{0}\}$.

Let the random vector $\mathbf{X} \sim H_r$ have the distribution (4.2). Then $P(\mathbf{X} \in A | \mathbf{X} \in B) = H_r(A \cap B)/H_r(B)$, and hence (5.6) can also be used to find conditional probabilities. Further, assuming continuity, (5.1) determines the conditional densities. E.g., writing $f_{|X_1=x}$ for the conditional density of $(X_2, \dots, X_d)$ given that $X_1 = x$ we get, for $x > 0$, that

$$f_{|X_1=x}(x_2, \dots, x_d) = \frac{\int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\mathbf{R}}(t^\gamma((x, x_2, \dots, x_d) + \frac{\boldsymbol{\sigma}}{\gamma})) dt}{\int_0^\infty t^{\gamma_1} f_{R_1}(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})) dt}, \quad (5.7)$$

and by integrating it further follows that

$$\begin{aligned} & P(\mathbf{X} \in A | X_1 = x) \\ &= \frac{\int_0^\infty t^{\gamma_1} f_{R_1}(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})) P((x, R_2, \dots, R_d)/t^\gamma - \frac{\boldsymbol{\sigma}}{\gamma} \in A | R_1 = t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})) dt}{\int_0^\infty t^{\gamma_1} f_{R_1}(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})) dt}. \end{aligned} \quad (5.8)$$

The one-dimensional integrals in these equations can often be computed numerically. Below, a simple example where they are obtained analytically.

Example 5.6. In Example 5.3 extreme flow episodes in the two river tributaries are modeled as $(R_1/T^\gamma, R_2/T^\gamma)$ with R_1, R_2 independent standard exponential variables and with $\gamma > 0$. Suppose $\mathbf{X} \sim H$ where H is the corresponding GP distribution and let $s > 0$. Using that $R_1 + R_2$ has a gamma distribution, it is straightforward to evaluate (5.6) to find the distribution of the sum of the flows in the two tributaries,

$$P(X_1 + X_2 > s) = c_1 \left(1 + \frac{\gamma}{\sigma_1 + \sigma_2} s\right)^{-1/\gamma},$$

with $c_1 = \gamma^{-1}(1 + \gamma)(\sigma_1 + \sigma_2)^{-1/\gamma} / (\sigma_1^{-1/\gamma} + \sigma_2^{-1/\gamma} - (\sigma_1 + \sigma_2)^{-1/\gamma})$.

Similar computations using (5.7) show that for $x > 0$

$$f_{|X_1=x}(x_2) = c_2 \left(1 + \frac{\gamma/(1 + \gamma)}{(\gamma x + \sigma_1 + \sigma_2)/(1 + \gamma)} x_2\right)^{-1-1/[\gamma/(1+\gamma)]}$$

and

$$P(X_2 > v | X_1 = x) = c_3 \left(1 + \frac{\gamma/(1 + \gamma)}{(\gamma x + \sigma_1 + \sigma_2)/(1 + \gamma)} x_2\right)^{-1/[\gamma/(1+\gamma)]},$$

for $c_2 = (1 + \gamma)(x + \sigma_1/\gamma)^{1+1/\gamma}(\gamma x + \sigma_1 + \sigma_2)^{-2-1/\gamma}$ and $c_3 = (x + \sigma_1/\gamma)^{1+1/\gamma}(\gamma x + \sigma_1 + \sigma_2)^{-1-1/\gamma}$. Hence the sum conditioned to be positive has a GP distribution with the same shape parameter as the marginal distributions but with a larger scale parameter, and the conditional distribution of X_2 given that $X_1 = x > 0$, conditioned to be positive, has a GP distribution with a smaller shape parameter, $\gamma/(1 + \gamma)$, as compared to the marginal parameters.

Much of the results in the example hold more generally. E.g., the conditional GP distribution of sums holds as soon as the marginals have the same shape parameter. The intuition is simple: Suppose the GP_r distribution has been obtained from the vector $(R_1/T^\gamma - \sigma_1/\gamma, \dots, R_d/T^\gamma - \sigma_d/\gamma)$ by (formal) conditioning on at least one of the components being positive. Then a weighted sum of the components equals $R/T^\gamma - \sigma/\gamma$, for $R = \sum_{j=1}^d a_j R_j$ and $\sigma = \sum_{j=1}^d a_j \sigma_j$. According to the GP_r representation, provided $a_1, \dots, a_d \geq 0$, the distribution of $R/T^\gamma - \sigma/\gamma$ conditioned to be positive is a one-dimensional GP distribution with parameters γ and σ . Further, that a sum is positive implies that at least one component is positive, and hence first conditioning on at least one component being positive, and then conditioning on the sum being positive gives the same result as conditioning directly on the sum being positive. Thus the one-dimensional GP distribution holds for component sums in GP distributions. Similar reasoning can be applied to, e.g., joint distributions of several weighted sums and several components. However, we only prove the one-dimensional result.

Proposition 5.7. *Let $\mathbf{X}$ be a GP random vector with common shape parameter γ for all d margins and with scale parameter $\boldsymbol{\sigma} > \mathbf{0}$, and for $\gamma \leq 0$ additionally assume that $P(\sum_{j=1}^d a_j X_j > 0) > 0$. Then, for $\mathbf{a} \in [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$, the conditional distribution of the weighted sum $\sum_{j=1}^d a_j X_j$ given that it is positive is generalized Pareto with shape parameter γ and scale parameter $\sigma = \sum_{j=1}^d a_j \sigma_j$.*

Proof. Since $P(\sum_{j=1}^d a_j X_j > 0) > 0$ holds automatically if $\gamma > 0$ and $\boldsymbol{\sigma} > \mathbf{0}$, this condition is satisfied for all values of γ . Let $A_x = \{\mathbf{y} \in \mathbb{R}^d \mid \sum_{j=1}^d a_j y_j > x\}$ and as above define $R = \sum_{j=1}^d a_j R_j$. Then, for $x > 0$ and with $F = \{\mathbf{y}; \mathbf{y} \not\leq \mathbf{0}\}$, as above, $A_x \cap F = A_x$, and (for $\gamma = 0$ using the standard convention that $t^0(x + \sigma/0)$ means $x/\sigma + \log t$) the numerator in (5.6) for $A = A_x$ is

$$\int_0^\infty P(\mathbf{R}/t^\gamma - \boldsymbol{\sigma}/\gamma \in A_x) dt = \int_0^\infty P(R/t^\gamma - \sigma/\gamma > x) dt,$$

and hence by (5.6)

$$\begin{aligned} P\left(\sum_{j=1}^d a_j X_j > x \mid \sum_{j=1}^d a_j X_j > 0\right) &= \frac{H_r(A_x)}{H_r(A_0)} = \frac{\int_0^\infty P(R/t^\gamma - \sigma/\gamma > x) dt}{\int_0^\infty P(R/t^\gamma - \sigma/\gamma > 0) dt} \\ &= \left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma}, \end{aligned}$$

where the last equality follows from making the change of variables $t(1 + x/\sigma)^{1/\gamma} \rightarrow t$ in the numerator. $\square$

Example 2.4 exhibits a situation where the sum of the components of a GP distribution is identically equal to $-\infty$ and where the assumption $P(X_1 + X_2 > 0) > 0$ hence is not satisfied.

6 Simulation

In this section we briefly outline four different methods for sampling from multivariate GP distributions. For Methods 1–3 we focus on simulation of a GP vector $\mathbf{X}_0$ with $\boldsymbol{\sigma} = \mathbf{1}$, $\gamma = \mathbf{0}$; $\mathbf{X}$ with general $\boldsymbol{\sigma}, \gamma$ is obtained from the vector $\mathbf{X}_0$ through (4.4). Furthermore, using the connection between GP_s and GP_r distributions, simulated GP_r vectors may be obtained by simulating GP_s vectors, and vice versa. Throughout this section we assume that simulation of $\mathbf{S}$ from $f_{\mathbf{S}}$ is possible. Recall the relation $\tilde{\mathbf{S}} = \mathbf{S} - \vee_{j=1}^d S_j$.

Method 1 (simulation from the $(R_{\tilde{\mathbf{s}}})$ model). This is immediate using (4.6): simulate a vector $\mathbf{S}$ from $f_{\mathbf{S}}$ and an independent variable $E \sim \text{Exp}(1)$ and set $\mathbf{X}_0 = E + \mathbf{S} - \max_{1 \leq j \leq d} S_j$.

Simulation from the (R) and (R_s) models is less direct. We develop three methods for this: rejection sampling, MCMC sampling, and approximate simulation using (4.1).

The main idea behind Methods 2 and 3 is to use an appropriate change of measure so that Method 1 can be used to simulate from the (R_s) model. The $\text{GP}_s(\mathbf{1}, \mathbf{0}, F_{\tilde{\mathbf{S}}})$ density is

$$h_{\tilde{\mathbf{s}}}(\mathbf{x}) = \mathbb{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} e^{-\vee_{j=1}^d x_j} \int_0^\infty t^{-1} f_{\mathbf{S}}(\mathbf{x} + \log t) dt.$$

If in this equation one replaces $\mathbf{S}$ by $\mathbf{S}_0$ where $\mathbf{S}_0$ has density

$$f_{\mathbf{S}_0}(\mathbf{x}) = \frac{e^{\sum_{j=1}^d x_j} f_{\mathbf{S}}(\mathbf{x})}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt} \quad (6.1)$$

then

$$\begin{aligned} h_{\tilde{s}}(\mathbf{x}) &= \mathbb{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{e^{-\sum_{j=1}^d x_j} \int_0^\infty t^{-1} e^{\sum_{j=1}^d (x_j + \log t)} f_{\mathbf{S}}(\mathbf{x} + \log t) dt}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt} \\ &= \mathbb{1}_{\{\mathbf{x} \not\leq \mathbf{0}\}} \frac{\int_0^\infty f_{\mathbf{S}}(\mathbf{x} + \log t) dt}{\int_0^\infty \bar{F}_{\mathbf{S}}(\log t) dt} = h_s(\mathbf{x}). \end{aligned}$$

Thus, if one can simulate $\mathbf{S}_0$ vectors, these give $\text{GP}_s(\mathbf{1}, \mathbf{0}, F_s)$ vectors via Method 1.

Method 2 (simulation of $\mathbf{S}_0$ via rejection sampling). Assume that φ is a bounding density which satisfies $f_{\mathbf{S}_0}(\mathbf{x}) \leq K\varphi(\mathbf{x})$, for some constant K . Then one draws a candidate vector $\mathbf{S}_0^c \sim \varphi$ and accepts the candidate with probability $\frac{f_{\mathbf{S}_0}(\mathbf{S}_0^c)}{\varphi(\mathbf{S}_0^c)K}$, and repeats until the required number of draws have been accepted. The probability of acceptance is $1/K$, and thus it is advantageous to find a φ such that K is not large. In high dimensions however, such a K might be difficult to find, and the following method may be preferable.

Method 3 (simulation of $\mathbf{S}_0$ via MCMC). This consists of using a standard Metropolis–Hastings algorithm to simulate from a Markov chain with stationary distribution (6.1). At iteration i , one draws a candidate vector $\mathbf{S}_0^c$ from the density $f_{\mathbf{S}}$ and accepts with probability $\min \left\{ 1, e^{\sum_{j=1}^d S_{0,j}^c - \sum_{j=1}^d S_{0,j}^{i-1}} \right\}$, where $\mathbf{S}_0^{i-1}$ is the current state of the chain. Alternative proposal distributions could be used with appropriate modification of the acceptance probability; for details see e.g Chib and Greenberg (1995).

Method 4 (approximate simulation from the (R) model). Equation (4.1) suggests an approximate way to simulate a vector $\mathbf{X} \sim \text{GP}_r(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\mathbf{R}})$ distribution: one chooses a suitably large U , simulates $\bar{T} \sim \text{Unif}[0, U]$ and an independent $\mathbf{R} \sim F_{\mathbf{R}}$, and if $\mathbf{R}/\bar{T}^\gamma \not\leq \boldsymbol{\sigma}/\boldsymbol{\gamma}$ one sets $\mathbf{X} = \mathbf{R}/\bar{T}^\gamma - \boldsymbol{\sigma}/\boldsymbol{\gamma}$, and if not one repeats the procedure. In this algorithm, the probability to keep a simulated $\mathbf{R}/\bar{T}^\gamma$ value is

$$\frac{1}{U} \int_0^U \bar{F}_{\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt \approx \frac{1}{U} \int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt,$$

so one has to simulate approximately $U / \int_0^\infty \bar{F}_{\mathbf{R}}(t^\gamma \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) dt$ values of $\mathbf{R}/\bar{T}^\gamma$ to get one $\mathbf{X}$ -value. Hence, a large U which ensures that $H^{(U)}$ approximates H well leads to longer computation times, and a compromise has to be made. As a guide to the compromise, it is often possible to compute, analytically or numerically, sharp bounds for the approximation errors. Sometimes, as for Gaussian or log-Gaussian processes, one in particular can use existing sharp bounds for $P(\sum_{i=1}^d |R_i|^{1/\gamma_i} > t)$.

To summarize, Method 1 is the simplest, but can only be used to simulate $\text{GP}_s(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\mathcal{S}})$ vectors in a straightforward manner. Method 2 and Method 3 provide ways to simulate vectors $\mathbf{S}_0$ from distribution (6.1), which can then be inserted into Method 1 to simulate from the $\text{GP}_s(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\mathcal{S}})$ distribution. Method 4 is as simple as Method 1 to implement and produces i.i.d. vectors, but similarly as Method 3, only approximates the target distribution. Which method to use will depend on the specific problem at hand, and on available analytic, programming, and computer resources.

7 Conclusion

This paper studies the probability theory underlying peaks over thresholds modelling of multivariate data using generalized Pareto distributions. We first derive basic properties of the multivariate GP distribution, including behavior under conditioning, scale change, and convergence in distribution, and provide the connections with multivariate generalized extreme value distributions. The main results are a point process limit result which gives a general and concrete description of the behavior of extreme episodes in data; three representations of the cdf-s of multivariate GP distributions, motivated by and derived from the point process result; expressions for likelihoods and censored likelihoods; formulas for probabilities and conditional probabilities of general sets; and algorithms for random sampling from multivariate GP distributions. Throughout, the general results are illustrated by examples.

Peaks over thresholds modeling of extremes of random vectors $\mathbf{Y}$ first selects a suitable level $\mathbf{u}$ and then models the distribution of the over- and undershoots, $\mathbf{X} = \mathbf{Y} - \mathbf{u}$, conditional on the occurrence of at least one overshoot, by a GP distribution. Of course, this GP model at once also gives a model for the original $\mathbf{Y}$ -s, since $\mathbf{Y} = \mathbf{X} + \mathbf{u}$. Remaining modelling components which are not the topic of this paper include choice the level $\mathbf{u}$, perhaps as a function of covariates like time, and modelling of the Poisson process which governs the occurrence of extreme episodes. Specifically, quantile regression provides an appealing approach to choosing $\mathbf{u}$.

The paper gives a basis for understanding and likelihood modelling of extreme episodes, also in high dimensions. We believe it will contribute to the solution of different and important risk handling problems. However, it still is an early excursion into new territory, and much research remains to be done. As examples of this, inclusion of time series dependence into the models, and development of methods for prediction of extremes are important challenges.

Acknowledgement

Anna Kiriliouk has throughout participated in the discussions leading to this paper. We thank her for support and many helpful comments. Research supported by the Knut and Alice Wallenberg foundation. The work of Johan Segers was funded by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy).

References

- Aulbach, S., V. Bayer, M. Falk, et al. (2012). A multivariate piecing-together approach with an application to operational loss data. *Bernoulli* 18(2), 455–475.
- Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *Ann. Probability* 2(5), 792–804.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.
- Brodin, E. and H. Rootzén (2009). Univariate and bivariate gpd methods for predicting extreme wind storm losses. *Insurance: Mathematics and Economics* 44(3), 345–356.
- Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49(4), 327–335.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53(2), 377–392.
- Davison, A. C., S. A. Padoan, and M. Ribatet (2012). Statistical modeling of spatial extremes. *Statist. Sci.* 27(2), 161–186.
- Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* 52(3), 393–442.
- de Haan, L. and A. Ferreira (2006). *Extreme value theory, an introduction*. Springer.
- de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis* 99(6), 1260–1275.
- Einmahl, J., A. Kiriliouk, A. Krajina, and J. Segers (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, in press. arXiv:1403.1975.
- Einmahl, J. H. J., A. Krajina, and J. Segers (2012). An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* 40(3), 1764–1793.
- Falk, M., J. Hüsler, and R.-D. Reiss (2010). *Laws of small numbers: extremes and rare events*. Springer Science & Business Media.
- Ferreira, A. and L. de Haan (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* 20(4), 1717–1737.
- Guzzetti, F., S. Peruccacci, M. Rossi, and C. Stark (2007). Rainfall thresholds for the initiation of landslides in central and southern europe. *Meteorology and Atmospheric Physics*, 239–267.

- Huser, R. and A. Davison (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* 100(2), 511–518.
- Huser, R., A. Davison, and M. Genton (2015). Likelihood estimators for multivariate extremes. *Extremes* 19(1), 79–103.
- Joe, H., R. L. Smith, and I. Weissman (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B (Methodological)* 54(1), 171–183.
- Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25 (2002) 25(1), 1287–1304.
- Kysely, J., J. Picek, and R. Beranová (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change* 72(1–2), 55–68.
- Marshall, A. W. and I. Olkin (1983). Domains of attraction of multivariate extreme value distributions. *Annals of Probability* 11(1), 168–177.
- McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.
- Michel, R. (2009). Parametric estimation procedures in multivariate generalized Pareto models. *Scandinavian Journal of Statistics* 36(1), 60–75.
- Penrose, M. D. (1992). Semi-minstable processes. *Annals of Probability* 20(3), 1450–1463.
- Pickands, III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* 3(1), 119–131.
- Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*. Springer.
- Resnick, S. I. (2007). *Heavy-tail phenomena, probabilistic and statistical modelling*. Springer.
- Rootzén, H. and N. Tajvidi (2006). Multivariate generalized Pareto distributions. *Bernoulli* 12(5), 917–930.
- Schlather, M. (2002). Models for stationary maxstable fields. *Extremes* 5(1), 61–82.
- Tajvidi, N. (1996). *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions*. Ph. D. thesis, Department of Mathematics, Chalmers University of Technology, Göteborg.
- Wadsworth, J. and J. Tawn (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* 101(1), 1–15.
- Wikipedia (2016a). 2003 European heat wave — Wikipedia, the free encyclopedia. [Online; accessed 8-February-2016].

Wikipedia (2016b). Effects of hurricane Katrina in New Orleans — Wikipedia, the free encyclopedia. [Online; accessed 15-January-2016].

Appendix

Proof of (2.6). If $x_j < \eta_j$ for some $j \in \{1, \dots, d\}$, then $\nu(\{\mathbf{y} : \mathbf{y} \leq \mathbf{x}\}) = 0$, so that $H(\mathbf{x}) = 0$ too. Let $\mathbf{x} > \boldsymbol{\eta}$. We have

$$\begin{aligned} \{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}, \mathbf{y} \leq \mathbf{x}\} &= \{\mathbf{y} : \exists j, y_j > 0; \forall j, y_j \leq x_j\} \\ &= \{\mathbf{y} : \exists j, y_j > x_j \wedge 0; \forall j, y_j \leq x_j\} \\ &= \{\mathbf{y} : \mathbf{y} \not\leq \mathbf{x} \wedge \mathbf{0}, \mathbf{y} \leq \mathbf{x}\}. \end{aligned}$$

As a consequence,

$$\begin{aligned} H(\mathbf{x}) &= \frac{\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}, \mathbf{y} \leq \mathbf{x}\})}{\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}\})} \\ &= \frac{\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{x} \wedge \mathbf{0}, \mathbf{y} \leq \mathbf{x}\})}{\nu(\{\mathbf{y} : \mathbf{y} \not\leq \mathbf{0}\})} \\ &= \frac{(-\log G(\mathbf{x} \wedge \mathbf{0})) - (-\log G(\mathbf{x}))}{-\log G(\mathbf{0})} = \frac{1}{\log G(\mathbf{0})} \log \left(\frac{G(\mathbf{x} \wedge \mathbf{0})}{G(\mathbf{x})} \right), \end{aligned}$$

as required. $\square$

Proof that a GEV cdf G with $\boldsymbol{\sigma} > \mathbf{0}$ is determined by its values for $\mathbf{x} \geq \mathbf{0}$. It follows from $\boldsymbol{\sigma} > \mathbf{0}$ that the conditional margins of G has the form (2.3), and hence $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ are determined by the values of $G(\mathbf{x})$ for $\mathbf{x} > \mathbf{0}$. Further by max-stability we have that $G(\mathbf{a}_t \mathbf{x} + \mathbf{b}_t)^t = G(\mathbf{x})$ and hence $G(\mathbf{x})$ is determined for all values of $\mathbf{x}$ such that $\mathbf{a}_t \mathbf{x} + \mathbf{b}_t \geq \mathbf{0}$ i.e. for $\mathbf{x} \geq -\mathbf{a}_t^{-1} \mathbf{b}_t$. Using (2.5), it is seen that if $\gamma_i > 0$ then $-\mathbf{a}_{t,i}^{-1} \mathbf{b}_{t,i} \rightarrow -\sigma_i / \gamma_i = \eta_i$ and if $\gamma_i = 0$ then $-\mathbf{a}_{t,i}^{-1} \mathbf{b}_{t,i} \rightarrow -\infty = \eta_i$ as $t \rightarrow \infty$. Further, if $\gamma_i > 0$ then $-\mathbf{a}_{t,i}^{-1} \mathbf{b}_{t,i} \rightarrow -\infty = \eta_i$ as $t \rightarrow 0$. Thus $G(\mathbf{x})$ is determined for all values in the support of G , and this in turn determines $G(\mathbf{x})$ for all values of $\mathbf{x}$. $\square$