Extremes (2018) 21:115–145
DOI 10.1007/s10687-017-0294-4

Multivariate peaks over thresholds models

Holger Rootzén[1] · **Johan Segers**[2] ·
Jennifer L. Wadsworth[3]

Received: 18 October 2016 / Revised: 20 April 2017 / Accepted: 21 April 2017 /
Published online: 23 June 2017

Abstract Multivariate peaks over thresholds modelling based on generalized Pareto distributions has up to now only been used in few and mostly two-dimensional situations. This paper contributes theoretical understanding, models which can respect physical constraints, inference tools, and simulation methods to support routine use, with an aim at higher dimensions. We derive a general point process model for extreme episodes in data, and show how conditioning the distribution of extreme episodes on threshold exceedance gives four 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 is then transformed to the real scale. The third and fourth representations are reformulations of a spectral representation proposed in Ferreira and de Haan (Bernoulli **20**(4), 1717–1737, 2014). Numerically tractable forms of densities and censored densities are found and give tools for flexible parametric likelihood inference. New simulation algorithms, explicit formulas for probabilities and conditional probabilities, and conditions which make the conditional distribution of weighted component sums generalized Pareto are derived.

✉ Holger Rootzén
hrootzen@chalmers.se

Johan Segers
johan.segers@uclouvain.be

Jennifer L. Wadsworth
j.wadsworth@lancaster.ac.uk

[1] Chalmers and Gothenburg University, Göteborg, Sweden

[2] Université catholique de Louvain, Louvain-la-Neuve, Belgium

[3] Lancaster University, Lancaster, UK

Keywords Extreme values · Multivariate generalized Pareto distribution · Peaks over threshold likelihoods · Simulation of extremes

AMS 2000 Subject Classifications 62G32 · 60G70 · 62E10

1 Introduction

Peaks over thresholds (PoT) modelling was introduced in the hydrological literature (NERC 1975). The philosophy is simple: extreme events, perhaps extreme water levels, are often quite different from ordinary everyday behaviour, and ordinary behaviour 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; Smith 1984; Davison and Smith 1990).

Since then, numerous 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), and McNeil et al. (2015). The method has also been presented in a number of books, see e.g. Coles (2001), Beirlant et al. (2004), and Dey and Yan (2015).

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 levees were breached. However, many others held, and damage was determined by which levees held and which were flooded (Andersen et al. 2007). Extreme rain can lead to devastating landslides, and can be caused by one day with very extreme rainfall, or by two or more consecutive days with smaller, but still extreme rain amounts (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 long sequence of high minimum nightly temperatures which led to increasing fatigue and eventually to death (Grynszpan 2003). These and very many other important societal problems underline the importance of statistical methods which can handle multivariate extreme episodes.

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 all the componentwise differences in an extreme episode are modelled, both the overshoots and the undershoots. 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 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.

The aim of this paper is to contribute probabilistic understanding, physically motivated models, likelihood tools, and simulation methods, all of which are needed for multivariate PoT modelling of extreme episodes via multivariate GP distributions. Specifically, the key contributions are: new representations of GP distributions conducive to model construction; density formulas for each of these representations; new properties of multivariate GP distributions; and simulation tools. Many of these results are oriented towards enabling improved statistical modelling, but here we restrict ourselves to a probabilistic study. A companion paper (Kiriliouk et al. 2016) addresses practical modelling aspects.

We begin by deriving the basic properties of the class of multivariate GP distributions. We then pursue the following program:

(i) to exhibit the possible point process limits 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 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 four 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), and the fourth one a simple reformulation of this representation aimed at aiding model construction. A useful, and to us surprising, discovery is that it is possible to derive the density also for the fourth representation, and that this density in fact is simpler than the densities for the other two first representations. The importance of (iii) is that likelihood inference makes it possible to incorporate covariates, e.g. temporal or spatial trends, 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 modelling realism and computational tractability which together will make possible routine use, also in dimensions higher than two. The limiting factor is the number of parameters rather than the number of variables. The models mentioned in Example 3 may be a case in point. The formulas for probabilities, conditional probabilities and conditional densities given in Sections 5 and 6, 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). A closely related approximation was used in Smith et al. (1997). 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; Einmahl et al. 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, compared to the point process approach, 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: for a survey in the spatial context see Davison et al. (2012). However, componentwise maxima may occur at different times for different components, and in many situations the focus is on the PoT structure: extremes which occur simultaneously. Additionally, likelihood inference for GEV distributions is complicated by a lack of tractable analytic expressions for high-dimensional densities, 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. The most important special case of GP models are those for which all variables can be simultaneously extreme, and there is no mass placed on hyperplanes (see Section 2 for details of the support); this is a typical modelling assumption. Further comment on the situation of asymptotic independence, where this does not hold, is made in Section 8, as well as in Kiriliouk et al. (2016).

Section 2 derives and exemplifies the basic properties of the GP cumulative distribution functions (cdf-s). In Section 3 we develop a point process model of extreme episodes, and Section 4 shows how conditioning on exceeding high thresholds leads to three basic representations of the GP distributions. Section 5 exhibits the fourth representation and derives densities and censored likelihoods, while Section 6 gives formulas for probabilities and conditional probabilities in GP distributions. Finally, Section 7 contributes simulation algorithms for multivariate GP distributions and Section 8 discusses parametrization issues and gives a concluding overview.

2 Multivariate generalized Pareto distributions

This section first briefly recalls and adapts existing theory for multivariate GEV distributions, and then derives a number of the basic properties of GP distributions.

Throughout we 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, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ and $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^d$. Operations and relations involving such vectors are meant componentwise, with shorter vectors being recycled if necessary. For instance $\boldsymbol{a}\boldsymbol{x} + \boldsymbol{b} = (a_1x_1 + b_1, \ldots, a_dx_d + b_d)$, $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_j \leq y_j$ for $j = 1, \ldots, d$, and $t^{\boldsymbol{\gamma}} = (t^{\gamma_1}, \ldots, t^{\gamma_d})$. 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. That $\boldsymbol{X} \sim F$ means that $\boldsymbol{X}$ has distribution F, and $\xrightarrow{d}$ denotes convergence in distribution. The symbol $\mathbb{1}$ is the indicator function: $\mathbb{1}_A$ 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, if $\gamma = 0$. This convention also applies componentwise to expressions of the form $(\boldsymbol{x}^{\boldsymbol{\gamma}} - 1)/\boldsymbol{\gamma}$ and $(1 + \boldsymbol{\gamma x})^{1/\boldsymbol{\gamma}}$.

Below we repeatedly use that if $\boldsymbol{X}$ is a d-dimensional vector with $P(\boldsymbol{X} \nleq \boldsymbol{u}) > 0$ and $\boldsymbol{s} > \boldsymbol{0}$ then

$$\mathrm{P}[\boldsymbol{s}(\boldsymbol{X}-\boldsymbol{u}) \le \boldsymbol{x} \mid \boldsymbol{X}-\boldsymbol{u} \nleq \boldsymbol{0}] = \frac{\mathrm{P}[\boldsymbol{X} \le \boldsymbol{x}/\boldsymbol{s} + \boldsymbol{u}] - \mathrm{P}[\boldsymbol{X} \le (\boldsymbol{x} \wedge \boldsymbol{0})/\boldsymbol{s} + \boldsymbol{u}]}{\mathrm{P}[\boldsymbol{X} \nleq \boldsymbol{u}]}. \quad (2.1)$$

2.1 Background: multivariate generalized extreme value distributions

Throughout, G denotes a d-variate GEV distribution, so that in particular G has non-degenerate margins. 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 location-scale normalized maxima*, i.e., the distributions which are limits

$$\mathrm{P}[\boldsymbol{a}_n^{-1}(\textstyle\bigvee_{i=1}^{d} \boldsymbol{X}_i - \boldsymbol{b}_n) \le \boldsymbol{x}] \xrightarrow{d} G(\boldsymbol{x}), \quad \text{as} \quad n \to \infty, \quad (2.2)$$

of normalized maxima of independent and identically distributed (i.i.d.) vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \sim F$, for $\boldsymbol{a}_n > \boldsymbol{0}$ and $\boldsymbol{b}_n$; and (M2) *It is the class of max-stable distributions*, i.e., distributions such that taking maxima of i.i.d. vectors from the distribution only leads to a location-scale change of the distribution. By (M1) the class of GEV distributions is closed under location and scale changes.

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

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

for $x \in \mathbb{R}$ such that $\alpha_j + \gamma_j(x-\mu_j) > 0$. We will use this parametrization throughout. The parameter range is $(\gamma_j, \mu_j, \alpha_j) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$. Define

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

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

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

while G is supported by a (subset of) the rectangle $\tilde{I}_1 \times \cdots \times \tilde{I}_d$. 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 $\boldsymbol{x} \in \tilde{I}_1 \times \cdots \times \tilde{I}_d$ as $\boldsymbol{\gamma x} + \boldsymbol{\sigma} > \boldsymbol{0}$.

Below we assume that $0 < G(\boldsymbol{0}) < 1$. This inequality is equivalent to $G_j(0) > 0$ for *all* $j \in \{1, \ldots, d\}$ and $G_j(0) < 1$ for *some* $j \in \{1, \ldots, d\}$. The equivalence follows from positive quadrant dependence, $G(\boldsymbol{0}) \ge \prod_{j=1}^{d} G_j(0)$ (Marshall and Olkin 1983).

PoT models are determined by the difference between the thresholds and the location parameters of the observations, and not by their individual values. Hence, it does not entail any loss of generality to shift the location parameters $\{\mu_i\}$ to make the assumption $0 < G(\mathbf{0}) < 1$ hold.

We will often use the stronger condition that $\boldsymbol{\sigma} > \mathbf{0}$, i.e., that $\sigma_j > 0$ for *all* $j \in \{1, \dots, d\}$. By Eq. 2.4, this 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 $j \in \{1, \dots, d\}$. This is an additional restriction only for $\gamma_j < 0$: if $\gamma_j = 0$, then $\sigma_j = \alpha_j > 0$, while if $\gamma_j > 0$ then $G(\mathbf{0}) > 0$ 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 for each $t > 0$ there exist scale and location vectors $\boldsymbol{a}_t \in (0, \infty)^d$ and $\boldsymbol{b}_t \in \mathbb{R}^d$ such that $G(\boldsymbol{a}_t \boldsymbol{x} + \boldsymbol{b}_t)^t \equiv G(\boldsymbol{x})$ (Resnick 1987, Equation (5.17)). It follows from Eq. 2.3 that these parameters are given by

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

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 $\nu(\{\boldsymbol{y};\, \boldsymbol{y} \not\leq \boldsymbol{x}\}) = -\log G(\boldsymbol{x})$ for $\boldsymbol{x} \in [-\infty, \infty)^d$, with the convention that $-\log(0) = \infty$ (Resnick 1987, Proposition 5.8). The measure ν is called *intensity measure* because, by (M1), the limit of the expected number of location-scale normalized points, say $\boldsymbol{a}_n^{-1}(\boldsymbol{X}_i - \boldsymbol{b}_n)$, $i \in \{1, \dots, n\}$, in a Borel set A which is bounded away from $\boldsymbol{\eta}$ and such that $\nu(\partial A) = 0$, is equal to $\nu(A)$. The intensity measure ν determines the limit distribution of the sequence of point processes $\sum_{i=1}^{n} \delta_{\boldsymbol{a}_n^{-1}(\boldsymbol{X}_i - \boldsymbol{b}_n)}$, 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 $0 < \nu(\{\boldsymbol{y};\, \boldsymbol{y} \not\leq \mathbf{0}\}) < \infty$, so that we can define a probability measure supported by the set $\{\boldsymbol{y};\, \boldsymbol{y} \not\leq \mathbf{0}\}$ by restricting the intensity measure ν to that set and normalizing it. The result is the *generalized Pareto (GP)* distribution associated to G. Its cdf H may be expressed as

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

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 prove (2.6) in the Appendix. For points $\boldsymbol{x} \in [-\infty, \infty)^d$ with $\boldsymbol{x} \geq \boldsymbol{\eta}$ and $x_j = \eta_j$ for some j, the value of $H(\boldsymbol{x})$ is determined by right-hand continuity. Below it is shown that $\boldsymbol{\eta}$ is determined by the values of $H(\boldsymbol{x})$ for $\boldsymbol{x} \geq \mathbf{0}$.

The probability that the j-th 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 of main interest is when all components have a positive probability of being an exceedance, or equivalently when $H_j(0) < 1$ for all $j \in \{1, \ldots, d\}$, or, again equivalently, when $\boldsymbol{\sigma} > \mathbf{0}$.

Similarly to the characterizations (M1) and (M2) of the GEV distributions, the class of GP distributions H such that $H_j(0) < 1$ for all $j \in \{1, \ldots, d\}$ has the following characterizations (Rootzén and Tajvidi 2006).[1] The functions $\boldsymbol{\sigma}_t, \boldsymbol{u}_t$ in the characterizations are assumed to be continuous, and additionally $\boldsymbol{u}_t$ is assumed increasing.

(T1) *The GP distributions are limits of distributions of threshold excesses:* Let $\boldsymbol{X} \sim F$. If there exist scaling and threshold functions $\boldsymbol{s}_t \in (0, \infty)^d$ and $\boldsymbol{u}_t \in \mathbb{R}^d$ with $F(\boldsymbol{u}_t) < 1$ and $F(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

$$\mathrm{P}[\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \vee \mathbf{0} \leq \cdot \mid \boldsymbol{X} \nleq \boldsymbol{u}_t] \xrightarrow{d} H_+, \qquad \text{as } t \to \infty,$$

for some cdf H_+ with nondegenerate margins, then the function $H_+(\boldsymbol{x}); \boldsymbol{x} > \mathbf{0}$ can be uniquely extended to a GP cdf $H(\boldsymbol{x}); \boldsymbol{x} \in \mathbb{R}^d$, and

$$\mathrm{P}[\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \vee \boldsymbol{\eta} \leq \cdot \mid \boldsymbol{X} \nleq \boldsymbol{u}_t] \xrightarrow{d} H, \qquad \text{as } t \to \infty. \tag{2.7}$$

(T2) *The GP distributions are threshold-stable:* Let $\boldsymbol{X} \sim H$ where H has nondegenerate margins on $\mathbb{R}_+$. If there exist scaling and threshold functions $\boldsymbol{s}_t \in (0, \infty)^d$ and $\boldsymbol{u}_t \in \mathbb{R}^d$, with $\boldsymbol{u}_1 = \mathbf{0}$ and $H(\boldsymbol{u}_t) \to 1$ as $t \to \infty$, such that

$$\mathrm{P}[\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t) \leq \boldsymbol{x} \mid \boldsymbol{X} \nleq \boldsymbol{u}_t] = H(\boldsymbol{x}) \tag{2.8}$$

for $\boldsymbol{x} \geq \mathbf{0}$ then there is an uniquely determined GP cdf $\tilde{H}$ such that $\tilde{H}(\boldsymbol{x}) = H(\boldsymbol{x})$ for $\boldsymbol{x} > \boldsymbol{\eta}$. Conversely, all GP distributions H for which $H_j(0) < 1$ for all $j \in \{1, \ldots, d\}$ satisfy (2.8) for all $\boldsymbol{x} \in \mathbb{R}^d$.

We use the term "threshold-stable" for property (T2) in analogy with the terms "sum-stable" and "max-stable". A distribution is sum- or max-stable if the sum or maximum, respectively, of independent variables with this distribution has the same distribution, up to a location-scale change. Analogously, a distribution is threshold-stable if conditioning on the exceedance of suitable higher thresholds leads to distributions which, up to scale changes, are the same as the original distribution. This property is illustrated in Fig. 1, with $\boldsymbol{u}_t$ and $\boldsymbol{s}_t$ as given below in Theorem 1(viii).

If (T1) holds we say that F belongs to the (threshold) domain of attraction of H. In contrast to the limit in (M1), 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.

The GP distribution H is supported by the set

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

[1] In the article, the truncation factor "$\vee \boldsymbol{\eta}$" is missing in Theorem 2.2 and Theorem 2.3 (ii). A correction note is forthcoming.

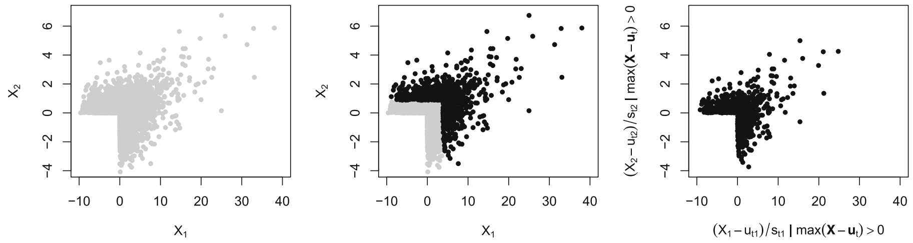

Fig. 1 Illustration of (T2). *Left panel:* points from a two-dimensional multivariate GP distribution with parameters $\boldsymbol{\sigma} = (2, 0.5)$ and $\boldsymbol{\gamma} = (0.2, 0.1)$. *Centre:* black points denote exceedances of the threshold $\boldsymbol{u}_t = \boldsymbol{\sigma}(t^{\boldsymbol{\gamma}} - 1)/\boldsymbol{\gamma}$, for $t = 5$. *Right:* excesses of $\boldsymbol{u}_t$ rescaled by $\boldsymbol{s}_t = t^{\boldsymbol{\gamma}}$ have the same distribution as points in the left panel, but are five times fewer. In particular, extremes in the right plot are therefore smaller

It may assign positive mass to the hyperplanes $\{\boldsymbol{y} : y_j = \eta_j\}$, even if $\eta_j = -\infty$; see Example 1 below.

For a non-empty subset J of $\{1, \ldots, d\}$, let H_J denote the corresponding $|J|$-variate marginal distribution of H. Further, let H_J^+ denote H_J conditioned to have at least one positive component; this presupposes that $\sigma_j > 0$ for some $j \in J$, where $\sigma_j = \alpha_j - \mu_j \gamma_j$ as before. By Theorem 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

$$H_j^+(x) = 1 - \left(1 + \gamma_j \tfrac{x}{\sigma_j}\right)^{-1/\gamma_j}, \qquad \text{for } x \geq 0 \text{ such that } \sigma_j + \gamma_j x > 0. \tag{2.9}$$

This proves the intuitively appealing result that H_j^+ is a one-dimensional GP distribution, and shows that $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$ and then also $\boldsymbol{\eta}$ are determined by the values of $H(\boldsymbol{x})$ for $\boldsymbol{x} \geq \mathbf{0}$.

If J is a non-empty subset of $\{1, \ldots, d\}$ and $\boldsymbol{x} \in [-\infty, \infty]^J$, then $\bar{\boldsymbol{x}} \in [-\infty, \infty]^d$ is defined by $\bar{x}_j = x_j$ if $j \in J$ and $\bar{x}_j = \infty$ if $j \notin J$. Thus, if $\boldsymbol{X} \sim H$, then the marginal distribution, H_J, of $(X_j : j \in J)$ is given by $H_J(\boldsymbol{x}) = H(\bar{\boldsymbol{x}})$ for $\boldsymbol{x} \in [-\infty, \infty]^J$, and if $H_J(\mathbf{0}) < 1$ then

$$H_J^+(\boldsymbol{x}) = \frac{H_J(\boldsymbol{x}) - H_J(\boldsymbol{x} \wedge \mathbf{0})}{\bar{H}_J(\mathbf{0})} \tag{2.10}$$

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

Theorem 1 *Let G be a GEV with margins* (2.3) *and suppose $H \leftrightarrow G$.*

(i) *Let $J \subset \{1, \ldots, d\}$. If $H_J(\mathbf{0}) < 1$ then H_J^+ is a GP cdf too, with $H_J^+ \leftrightarrow G_J$, and if $\sigma_j > 0$ then* Eq. 2.9 *holds. Further, H_J is a GP distribution if and only if $H_J(\mathbf{0}) = 0$.*
(ii) *A scale transformation of H is also a GP distribution.*
(iii) *Let $\boldsymbol{X} \sim H$. If $\boldsymbol{\sigma} > \mathbf{0}$ and $\boldsymbol{u} \geq \mathbf{0}$ with $H(\boldsymbol{u}) < \mathbf{1}$, then the conditional distribution of $\boldsymbol{X} - \boldsymbol{u}$ given that $\boldsymbol{X} \nleq \boldsymbol{u}$ is a GP distribution with the same shape parameter $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$ replaced by $\boldsymbol{\sigma} + \boldsymbol{\gamma}\boldsymbol{u}$.*
(iv) *If $\{H_n\}$ is a sequence of GP distributions with all components of the vectors $\boldsymbol{\sigma}_n$ bounded away from 0 and if $H_n \xrightarrow{d} \tilde{H}$ then $\tilde{H}$ is a GP distribution too.*

(v) *A finite or infinite mixture of GP distributions with the same $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ is a GP distribution.*

(vi) *We have $H \leftrightarrow G^t$ for all $t > 0$. Conversely, if $H \leftrightarrow G_*$ for some G_* and if $\boldsymbol{\sigma} > \mathbf{0}$ then $G_* = G^{t_1}$ for some $t_1 > 0$.*

(vii) *If $G(\mathbf{0}) = e^{-v}$ then $G(\boldsymbol{x}) = \exp\{-v\bar{H}(\boldsymbol{x})\}$, $\boldsymbol{x} \geq \mathbf{0}$, and if $\boldsymbol{\sigma} > \mathbf{0}$ this determines G.*

(viii) *If $\boldsymbol{\sigma} > \mathbf{0}$, the scaling and threshold functions in the (T2) characterization of GP distributions may be taken as $\boldsymbol{s}_t = t^{\boldsymbol{\gamma}}$ and $\boldsymbol{u}_t = \boldsymbol{\sigma}(t^{\boldsymbol{\gamma}} - \mathbf{1})/\boldsymbol{\gamma}$, for $t \geq 1$.*

(ix) *The parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$ are identifiable from H.*

In words, Theorem 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 (2.9)), 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. Theorem 1 (ii)−(v) also establish closure properties of the class of GP distributions. By (vi) and (vii) 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. Regarding (vii), note that a GEV distribution G such that $0 < G_j(0) < 1$ for all j is determined by its values for $\boldsymbol{x} \geq \mathbf{0}$ (proof in the Appendix). Finally, (viii) identifies the affine transformations which leave H unchanged, and (ix) establishes identifiability of the marginal parameters.

Proof (i) Let $\bar{\mathbf{0}}$ denote $\bar{\boldsymbol{x}}$ for the special case when $\boldsymbol{x} = \mathbf{0} \in (-\infty, \infty)^J$ and let $G_J(\boldsymbol{x}) = G(\bar{\boldsymbol{x}})$ be the marginal distribution of G. Clearly $\bar{\boldsymbol{x}} \wedge \mathbf{0} = \overline{(\boldsymbol{x} \wedge \mathbf{0})} \wedge \mathbf{0}$ and hence, for $\boldsymbol{x} > \boldsymbol{\eta}$,

$$
\begin{aligned}
&H_J(\boldsymbol{x}) - H_J(\boldsymbol{x} \wedge \mathbf{0})\\
&= \frac{1}{\log G(\mathbf{0})} \log\left(\frac{G(\bar{\boldsymbol{x}} \wedge \mathbf{0})}{G(\bar{\boldsymbol{x}})}\right) - \frac{1}{\log G(\mathbf{0})} \log\left(\frac{G(\overline{(\boldsymbol{x} \wedge \mathbf{0})} \wedge \mathbf{0})}{G(\overline{\boldsymbol{x} \wedge \mathbf{0}})}\right)\\
&= \frac{1}{\log G(\mathbf{0})} \log\left(\frac{G_J(\boldsymbol{x} \wedge \mathbf{0})}{G_J(\boldsymbol{x})}\right)
\end{aligned}
$$

and

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

so that

$$
H_J^+(\boldsymbol{x}) = \frac{1}{\log G_J(\mathbf{0})} \log\left(\frac{G_J(\boldsymbol{x} \wedge \mathbf{0})}{G_J(\boldsymbol{x})}\right).
$$

Inserting (2.3) into the equation above for $J = \{j\}$ together with straightforward calculation proves (2.9), and hence completes the proof of the first assertion.

If $H_J(\mathbf{0}) = 0$, then $H_J = H_J^+$ and it follows from the first assertion that H_J is a GP distribution function. Further, GP distributions are supported by

$\{\boldsymbol{y};\, \boldsymbol{y} \not\leq \mathbf{0}\}$ and hence if $H_J(\mathbf{0}) > 0$ then H_J is not a GP cdf. This proves the second assertion.

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

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

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

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

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

(iv) 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 $\boldsymbol{\sigma}_n \to \boldsymbol{\sigma}$ and $\boldsymbol{\gamma}_n \to \boldsymbol{\gamma}$. Define $\boldsymbol{u}_{n,t}$ and $\boldsymbol{s}_{n,t}$ from H_n as in Eq. 2.8 (vi). Then, since H_n is a GP cdf we have, using first (T2) and (viii), and then the continuous mapping theorem, that

$$\begin{aligned} H_n(\boldsymbol{x}) &= \frac{H_n(\boldsymbol{x}/\boldsymbol{s}_{n,t} + \boldsymbol{u}_{n,t}) - H_n((\boldsymbol{x}/\boldsymbol{s}_{n,t}) \wedge \mathbf{0} + \boldsymbol{u}_{n,t})}{\bar{H}_n(\boldsymbol{u}_{n,t})} \\ &\xrightarrow{d} \frac{\tilde{H}(\boldsymbol{x}/\boldsymbol{s}_t + \boldsymbol{u}_t) - \tilde{H}((\boldsymbol{x}/\boldsymbol{s}_t) \wedge \mathbf{0} + \boldsymbol{u}_t)}{\bar{\tilde{H}}(\boldsymbol{u}_t)}, \qquad \text{as } n \to \infty. \end{aligned}$$

Since $H_n \xrightarrow{d} \tilde{H}$ it follows that $\tilde{H}$ satisfies (T2) and hence is a GP cdf.

(v) We only prove that a mixture of two GP cdf-s with the same $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ is a GP cdf too, using Theorem 4 below (the proof of that theorem does not use the result we are proving now). The proof for arbitrary finite mixtures is the same, and the result for infinite mixtures then follows by taking limits of finite mixtures and using (iv). Let H_1 and H_2 be GP cdf-s with the same marginal parameters $\boldsymbol{\sigma}, \boldsymbol{\gamma}$ and let $p \in (0, 1)$. By Theorem 4 and Eq. 4.2 there exists cdf-s $F_i = F_{Ri}$ such that $H_i(\boldsymbol{x}) = c_i \int_0^\infty \{F_i(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})) - F_i(t^{\boldsymbol{\gamma}}(\boldsymbol{x} \wedge \mathbf{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\}\,\mathrm{d}t$ with $c_i = 1/\int_0^\infty \bar{F}_i(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t$, and with the convention that if $\gamma_i = 0$ then $t^{\gamma_i}(x_i + \frac{\sigma_i}{\gamma_i})$ is interpreted to mean $x_i + \sigma_i \log t$. Writing $F = \frac{pc_1}{pc_1+(1-p)c_2} F_1 + \frac{(1-p)c_2}{pc_1+(1-p)c_2} F_2$ it follows that

$$\begin{aligned} \tilde{H}(\boldsymbol{x}) &:= pH_1(\boldsymbol{x}) + (1-p)H_2(\boldsymbol{x}) \\ &= [pc_1 + (1-p)c_2] \int_0^\infty \left\{F\left(t^{\boldsymbol{\gamma}}(\boldsymbol{x}+\tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\right) - F\left(t^{\boldsymbol{\gamma}}(\boldsymbol{x} \wedge \mathbf{0} + \tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\right)\right\} \mathrm{d}t. \end{aligned}$$

Straightforward calculation shows that $pc_1 + (1-p)c_2 = 1/\int_0^\infty \bar{F}(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t$ so that $\tilde{H}(\boldsymbol{x})$ satisfies Eq. 4.2 and hence is a GP cdf.

(vi) The first assertion follows from Eq. 2.6. Choose t_1 so that $-\log G(\mathbf{0})^{t_1} = -\log G_*(\mathbf{0})$. Then $H \leftrightarrow G$ and $H \leftrightarrow G_*$ imply together that

$$\frac{G(\boldsymbol{x} \wedge \mathbf{0})^{t_1}}{G(\boldsymbol{x})^{t_1}} = \frac{G_*(\boldsymbol{x} \wedge \mathbf{0})}{G_*(\boldsymbol{x})},$$

and in particular, that $G(\boldsymbol{x})^{t_1} = G_*(\boldsymbol{x})$ for $\boldsymbol{x} \geq \mathbf{0}$. Since a GEV cdf with $\boldsymbol{\sigma} > \mathbf{0}$ is determined by its values for $\boldsymbol{x} \geq \mathbf{0}$, see the Appendix, this completes the proof.

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

(viii) By the proofs of (iii) and (vi) the conditional distribution of $\boldsymbol{s}_t^{-1}(\boldsymbol{X} - \boldsymbol{u}_t)$ given that $\boldsymbol{X} \not\leq \boldsymbol{u}_t$ is associated with $G(\boldsymbol{s}_t\boldsymbol{x} + \boldsymbol{u}_t) = G(\boldsymbol{x})^{1/t}$, and the result follows from (vii).

(ix) For $t > 0$, let $(\gamma_j(t), \mu_j(t), \alpha_j(t)) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$ be the parameter vector of G_j^t, and let $\sigma_j(t) = \alpha_j(t) - \gamma_j(t)\,\mu_j(t)$. By assertion (vi), 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 $G(\boldsymbol{x})^t = G(\boldsymbol{a}_t^{-1}(\boldsymbol{x} - \boldsymbol{b}_t))$ with $\boldsymbol{a}_t$ and $\boldsymbol{b}_t$ as in Eq. 2.5.

□

Example 1 below exhibits two-dimensional GP distributions with positive mass on certain lines, and the first part of Example 2 provides a cdf where the second assertion in (i) of Theorem 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 GP cdf never is a GP cdf. The second part of Example 2 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 1 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 $\alpha_1 = \alpha_2 = -1$. The corresponding multivariate generalized Pareto distribution is given by

$$H(x, y) = \begin{cases} \frac{1}{2}\left(1 - \frac{1}{x+1} + 1 - \frac{1}{y+1}\right) & \text{if } (x, y) \in [0, \infty)^2, \\ \frac{1}{2}\left(1 - \frac{1}{x+1}\right) & \text{if } (x, y) \in [0, \infty) \times [-1, 0], \\ \frac{1}{2}\left(1 - \frac{1}{y+1}\right) & \text{if } (x, y) \in [-1, 0] \times [0, \infty), \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

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 generalized Pareto, $\mathrm{P}(T \leq t) = 1 - 1/(t+1)$ for $t \geq 0$. Hence H is supported by the union of the two lines $\{-1\} \times (0, \infty)$ and $(0, \infty) \times \{-1\}$, see Fig. 2, left panel.

If we modify the example by choosing Gumbel rather than Fréchet margins, so that $G(x, y) = \exp(-e^{-x} - e^{-y})$ for $(x, y) \in \mathbb{R}^2$, then the GP cdf H is the cdf of the vector

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

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

Example 2 Let $G(x, y) = \exp[-1/\{(x \wedge y) + 1\}]$ for $(x, y) \geq \mathbf{-1}$, the cdf of (Z, Z) for Z unit Fréchet with lower endpoint -1. The corresponding GP cdf is

$$H(x, y) = \begin{cases} 1 - \frac{1}{(x \wedge y)+1} & \text{if } (x, y) \in [0, \infty)^2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

We identify H as the distribution of the random pair (T, T), where $\mathrm{P}(T \leq t) = 1 - 1/(t+1)$ for $t \in [0, \infty)$. The support of H is the diagonal $\{(t, t) : 0 < t < \infty\}$, see Fig. 2, middle panel. It follows that, e.g., $H_1(0) = 0$ and hence in this example $H_1 = H_1^+$.

As a variation of the example let $G(x, y) = \exp[-e^{-x \wedge (y+\mu)}]$ be the cdf of $(Z, Z - \mu)$, with Z standard Gumbel and $\mu > 0$. The corresponding GP cdf is

$$H(x, y) = e^{-(x \wedge 0) \wedge (y \wedge 0 + \mu)} - e^{-x \wedge (y+\mu)} = e^{-(x \wedge 0) \wedge (y+\mu)} - e^{-x \wedge (y+\mu)}, \tag{2.13}$$

and H is the cdf of $(T, T - \mu)$ with T standard exponential. 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 Eq. 2.13, 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 Fig. 2.

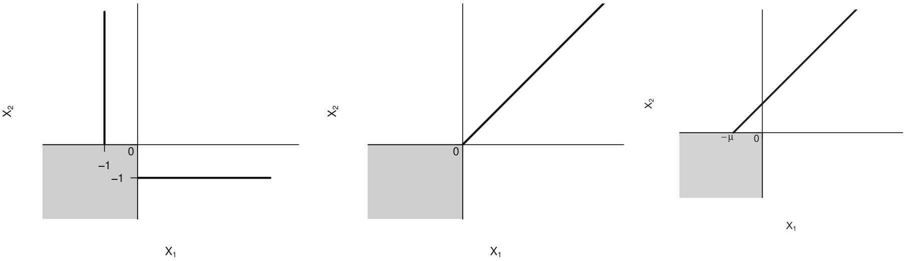

Fig. 2 Supports (*solid lines*) of the GP distributions H in Eqs. 2.11 (*left*), 2.12 (*middle*) and 2.13 (*right*)

3 Point processes of extreme episodes

The first step in our program for PoT inference is to specify a point process model for extreme episodes. This model exhibits extreme episodes as a product process obtained by multiplying a random vector, the "shape" vector, with a random quantity, the "intensity" of the episode. (For $\gamma = 0$, the model instead is a sum.) This is parallel to models commonly used for max-stable processes, see e.g. Schlather (2002). In Section 4 below we obtain basic and physically interpretable representations of the GP distributions by conditioning the product process of extreme episodes on threshold exceedance.

In this and subsequent sections we assume that $\boldsymbol{\sigma} > \boldsymbol{0}$. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be i.i.d. random vectors with cdf F and marginal cdf-s $F_1, \ldots, F_d$, and let $\boldsymbol{a}_n, \boldsymbol{b}_n$ be as in Eq. 2.2. Further, let $\boldsymbol{\eta}$ be the vector of lower endpoints of the limiting GEV distribution, see Eq. 2.4 and the sentences right below it. We consider weak limits of the point processes

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

where $\delta_{\boldsymbol{x}}$ denotes a point mass at $\boldsymbol{x}$. Define $I_j = [-\sigma_j/\gamma_j, \infty)$ or $[-\infty, \infty)$ or $[-\infty, -\sigma_j/\gamma_j)$ according to whether $\gamma_j > 0$ or $\gamma_j = 0$ or $\gamma_j < 0$, and set

$$\bar{S}_{\boldsymbol{\gamma}} = I_1 \times \cdots \times I_d \quad \text{and} \quad S_{\boldsymbol{\gamma}} = \bar{S}_{\boldsymbol{\gamma}} \setminus \{\boldsymbol{\eta}\}.$$

The limit point process is specified as follows: Let $0 < T_1 < T_2 < \ldots$ be the points of a Poisson process on $[0, \infty)$ with unit intensity and let $(\boldsymbol{R}_i)_{i \geq 1}$ be independent copies of a random vector $\boldsymbol{R}$ which satisfies Condition 2 below. Further assume that the vectors $(\boldsymbol{R}_i)_{i \geq 1}$ are independent of $(T_i)_{i \geq 1}$, and define the point process

$$P_r = \sum_{i \geq 1} \delta_{(\boldsymbol{R}_i / T_i^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma})}, \tag{3.1}$$

where, by convention, $R_{i,j}/T_i^0 - \sigma_j/0$ is interpreted to mean $R_{i,j} - \sigma_j \log T$. The condition on $\boldsymbol{R}$ is as follows.

Condition 2 *The components of the random vector $\boldsymbol{R}$ satisfy $R_j \in [0, \infty)$ if $\gamma_j > 0$, $R_j \in [-\infty, \infty)$ if $\gamma_j = 0$, and $R_j \in [-\infty, 0)$ if $\gamma_j < 0$, and furthermore $0 < \mathrm{E}[|R_j|^{1/\gamma_j}] < \infty$ if $\gamma_j \neq 0$ and $\mathrm{E}[\exp(R_j/\sigma_j)] < \infty$ if $\gamma_j = 0$, for $j = 1, \ldots, d$.*

Let $F_{\boldsymbol{R}}$ be the cdf of $\boldsymbol{R}$. For $\gamma_j \neq 0$, the moment restriction in Condition 2 can be seen to be equivalent to requiring that $0 < \int_0^\infty \mathrm{P}(R_j > t^{\gamma_j} x_j)\,\mathrm{d}t < \infty$, if $x_j \in (0, \infty)$ and $\gamma_j > 0$ or if $x_j \in (-\infty, 0)$ and $\gamma_j < 0$. For $\gamma_j = 0$, the moment condition is instead equivalent to $0 < \int_0^\infty \mathrm{P}(R_j > \sigma_j \log t + x_j)\,\mathrm{d}t < \infty$, for $x_j \in (-\infty, \infty)$. For example, if $\gamma_j < 0$, then $\int_0^\infty P(R_j > t^{\gamma_j} x_j)\,\mathrm{d}t = \int_0^\infty \mathrm{P}(|R_j|^{1/\gamma_j} > t)\,\mathrm{d}t\,|x_j|^{-1/\gamma_j} = \mathrm{E}(|R_j|^{1/\gamma_j})\,|x_j|^{-1/\gamma_j}$. Since $\mathrm{P}(R_j > t^{\gamma_j} x_j) \leq \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}} \boldsymbol{x}) \leq \sum_{i=1}^d \mathrm{P}(R_i > t^{\gamma_i} x_i)$, it in turn follows that the moment condition implies that $0 < \int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma}))\,\mathrm{d}t < \infty$, if the components x_j of $\boldsymbol{x}$ are as above, and where we have used the convention that $t^0(x_j + \sigma_j/0)$ should be replaced by $\sigma_j \log t + x_j$.

Theorem 3 *Suppose F satisfies (2.2). Then, for some $\boldsymbol{R}$ which satisfies Condition 2,*

$$N_n \xrightarrow{d} P_r \quad on \quad S_{\boldsymbol{\gamma}}, \quad as \quad n \to \infty. \tag{3.2}$$

Conversely, for any P_r given by Eq. 3.1 *there exist a GEV cdf G and $\boldsymbol{a}_n > \boldsymbol{0}$ and $\boldsymbol{b}_n$, with $0 < G_j(b_{n,j}) < 1$ and $G_j(b_{n,j}) \to 1$ for $j = 1, \ldots, d$, such that* Eq. 3.2 *holds for $F = G$.*

Proof Let $\boldsymbol{Y} \sim G$ and define $\boldsymbol{Y}^*$ by $Y_j^* = (1 + \frac{\gamma_j}{\alpha_j}(Y_j - \mu_j))^{1/\gamma_j}$ if $\gamma_j \neq 0$ and $Y_j^* = \exp\{(Y_j - \mu_j)/\alpha_j\}$ if $\gamma_j = 0$, for $\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\gamma}$ given by Eq. 2.3 so that the marginal cdf-s of $\boldsymbol{Y}^*$ are standard Fréchet. 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 $\boldsymbol{R}^* \in [0, \infty)^d$ with $\mathrm{E}(R_j^*) < \infty$ such that $\boldsymbol{Y}^*$ has the same distribution as $\sup_{i \geq 1} \boldsymbol{R}_i^*/T_i$ where the random vectors $\boldsymbol{R}_i^*$ are i.i.d. copies of $\boldsymbol{R}^*$ and independent of the unit rate Poisson process $(T_i)_{i \geq 1}$. Reversing the transformation which led from $\boldsymbol{Y}$ to $\boldsymbol{Y}^*$, it follows that $\boldsymbol{Y}$ has the same distribution as $\sup_{i \geq 1}(\frac{\boldsymbol{\alpha}}{\boldsymbol{\gamma}}(\boldsymbol{R}_i^*)^{\boldsymbol{\gamma}}/T_i^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})$. Setting $\boldsymbol{R} = \frac{\boldsymbol{\alpha}}{\boldsymbol{\gamma}}(\boldsymbol{R}^*)^{\boldsymbol{\gamma}}$ it follow that $\boldsymbol{Y}$ has the same distribution as $\sup_{i \geq 1}(\boldsymbol{R}_i/T_i^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})$ with the $\boldsymbol{R}_i$ satisfying Condition 2, and where throughout expressions should be interpreted as specified after Eq. 3.1 for $\gamma_j = 0$.

For ν the intensity measure of P_r, we have $G(\boldsymbol{x}) = \exp\{-\nu(\{\boldsymbol{y} : \boldsymbol{y} \nleq \boldsymbol{x}\})\}$. By standard reasoning, convergence in distribution of $\boldsymbol{a}_n^{-1}(\bigvee_{i=1}^n \boldsymbol{X}_i - \boldsymbol{b}_n)$ is equivalent to $n\,\mathrm{P}[\boldsymbol{a}_n^{-1}(\boldsymbol{X}_1 - \boldsymbol{b}_n) \nleq \boldsymbol{x}] \to -\log G(\boldsymbol{x}) = \nu(\{\boldsymbol{y} : \boldsymbol{y} \nleq \boldsymbol{x}\})$, which implies that $n\,\mathrm{P}[\boldsymbol{a}_n^{-1}(\boldsymbol{X}_1 - \boldsymbol{b}_n) \in \cdot\,]$ converges vaguely to $\nu(\,\cdot\,)$ on $S_{\boldsymbol{\gamma}}$. By Theorem 5.3 of Resnick (2007) this proves (3.2).

Conversely, given P_r, define the cdf G by $G(\boldsymbol{x}) = \mathrm{P}[\sup_{i \geq 1}(\boldsymbol{R}_i/T_i^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) \leq \boldsymbol{x}]$. Straightforward calculation as in Schlather (2002, Theorem 2) show that G is max-stable. Hence there exist sequences $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ with the stated properties such that for independent random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ with common distribution G, the distribution of $\boldsymbol{a}_n^{-1}(\bigvee_{i=1}^n \boldsymbol{X}_i - \boldsymbol{b}_n)$ is equal to G too. By the first part of the proof, this proves (3.2). □

In the proof of Theorem 3 we obtained part of the following result, which we record here for completeness.

Corollary 1 *Suppose $\boldsymbol{R}$ satisfies Condition 2. Then $G(\boldsymbol{x}) = \mathrm{P}[\sup_{i \geq 1}(\boldsymbol{R}_i/T_i^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) \leq \boldsymbol{x}]$ is a GEV cdf and*

$$G(\boldsymbol{x}) = \exp\left\{-\int_0^\infty \bar{F}_{\boldsymbol{R}}\left(t^{\boldsymbol{\gamma}}\left(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\right)\right) \mathrm{d}t\right\} \quad for \quad \boldsymbol{x} \in S_{\boldsymbol{\gamma}}, \tag{3.3}$$

and to any GEV cdf there exists an $\boldsymbol{R}$ which satisfies this equation. Here we use the convention that $t^0(x_j + \sigma_j/0)$ should be replaced by $\sigma_j \log t + x_j$.

Proof Writing ν for the intensity measure of P_r we have $\nu(\{\boldsymbol{y}; \boldsymbol{y} \nleq \boldsymbol{x}\}) = \int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\,\mathrm{d}t$. The right-hand side of Eq. 3.3 is therefore equal to the probability that P_r has no points in the set $\{\boldsymbol{y};\, \boldsymbol{y} \not\leq \boldsymbol{x}\}$. The result now follows from the proof of Theorem 3. □

4 Representations of multivariate GP distributions

This section contains the second step in the program for PoT inference. We show how conditioning on threshold exceedances in the point process (3.1) gives four widely useful representations of the class of multivariate GP distributions. The first representation, (R) is on the real scale and corresponds to the point process P_r in Eq. 3.1 with points obtained as products of shape vectors and intensity variables. In the second representation, (U), the basic model is constructed on a standard scale and then transformed to the real scale. The third representation, (S), is equivalent to the spectral representation in Ferreira and de Haan (2014). A fourth representation, (T), which is a variation of the (S) representation, is introduced in Section 5.

In the literature the standard scale is chosen as one of the following: Pareto scale, $\boldsymbol{\gamma} = \boldsymbol{1}$, uniform scale, $\boldsymbol{\gamma} = -\boldsymbol{1}$, or exponential scale, $\boldsymbol{\gamma} = \boldsymbol{0}$. Here we choose the $\boldsymbol{\gamma} = \boldsymbol{0}$ scale because of the simple additive structure it leads to. For all four representations, it is straightforward to switch from one scale to another one.

To understand the GP representation (R) we first approximate P_r by a truncated point process $\bar{P}_r$ where $\{T_i\}$ is replaced by a unit rate Poisson process $\{\bar{T}_i\}$ on a bounded interval $[0, K]$. Recalling the representation of $\{\bar{T}_i\}$ as a Poisson distributed number of Unif $[0, K]$ variables, $\bar{P}_r$ consists of a Poisson number of vectors $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$ with $\bar{T} \sim$ Unif $[0, K]$. Thus, for large n, a point $\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n)$ in N_n has approximately the same distribution as $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$. Hence, by Eq. 2.1,

$$\begin{aligned}
\mathrm{P}[\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n) \leq \boldsymbol{x} \mid \boldsymbol{X} - \boldsymbol{b}_n \not\leq \boldsymbol{0}] &\approx \mathrm{P}[\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}} \leq \boldsymbol{x} \mid \boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}} \not\leq \boldsymbol{0}] \\
&= \frac{\frac{1}{K}\int_0^K \{F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})) - F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} \wedge \boldsymbol{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\}\,\mathrm{d}t}{\frac{1}{K}\int_0^K \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t} \\
&= \frac{\int_0^K \{F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})) - F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} \wedge \boldsymbol{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\}\,\mathrm{d}t}{\int_0^K \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t}. \qquad (4.1)
\end{aligned}$$

By the assumptions in Condition 2, the limit as $K \to \infty$ of this expression,

$$H_R(\boldsymbol{x}) = \frac{\int_0^\infty \{F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})) - F_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} \wedge \boldsymbol{0} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\}\,\mathrm{d}t}{\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t}, \qquad (4.2)$$

exists (cf the discussion just before Theorem 3), and it is also immediate that $H_R(\infty) = 1$, so that H_R is a cdf on $[-\infty, \infty)^d$. If $\gamma_i = 0$ then $t^{\gamma_i}(x_i + \frac{\sigma_i}{\gamma_i})$ should be interpreted to mean $x_i + \sigma_i \log t$. We write $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ for the cdf (4.2) and call it the (R) representation. Theorem 4 shows that the class of such cdf-s is the same as the class of all GP cdf-s with $\boldsymbol{\sigma} > \boldsymbol{0}$.

Heuristically, for simplicity assuming that $\gamma_j \neq 0$, $j = 1, \ldots d$, the calculations above proceed by equating $\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n)$ with $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$ so that extremes of $\boldsymbol{X}$

asymptotically have the form $\boldsymbol{a}_n\boldsymbol{R}/T^{\boldsymbol{\gamma}}+\boldsymbol{b}_n-\boldsymbol{a}_n\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$. Setting $\boldsymbol{b}=\boldsymbol{b}_n-\boldsymbol{a}_n\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$ and noting that $\boldsymbol{R}$ satisfies Condition 2 if and only if $\boldsymbol{a}_n\boldsymbol{R}$ satisfies Condition 2, the intuition is that, asymptotically, extremes of $\boldsymbol{X}$ have the form

$$\boldsymbol{X}^{\infty}=\boldsymbol{R}/T^{\boldsymbol{\gamma}}+\boldsymbol{b}, \tag{4.3}$$

for some constant $\boldsymbol{b}$ and a random vector $\boldsymbol{R}$ which satisfies Condition 2. The interpretation is that the vector $\boldsymbol{R}$ is the shape of the extreme episode, say a storm, and that $T^{-\boldsymbol{\gamma}}$ is the intensity of the storm. Here T represents a pseudo random variable with an improper uniform distribution on $(0,\infty)$. Although such a $\boldsymbol{X}^{\infty}$ therefore does not have a proper distribution, one can verify that the cdf H_R in Eq. 4.2 is derived as if it were the conditional distribution of $\boldsymbol{X}^{\infty}-\boldsymbol{u}$, given that $\boldsymbol{X}^{\infty}\nleq\boldsymbol{u}$, for $\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}=\boldsymbol{u}-\boldsymbol{b}$, i.e., as the formal conditional distribution of $\boldsymbol{R}/T^{\boldsymbol{\gamma}}-\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$ given that $\boldsymbol{R}/T^{\boldsymbol{\gamma}}-\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\nleq\boldsymbol{0}$. In statistical application one would assume that $\boldsymbol{u}$ is large enough to make it possible to use the cdf H_R as a model for threshold excesses. The parameters of $\boldsymbol{R}$ and the parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ are then estimated from the observed threshold excesses. The heuristic interpretation in case one or more of the γ_j equals 0 is the same, one only has to write $R_j-\sigma_j\log T$ instead of $R_j/T^0-\sigma_j/0$.

Figure 3 illustrates how the multivariate GP distribution is derived from the Poisson process representation. Each realization of the Poisson process (3.1) yields a small, Poisson distributed, number of points in the region $\{\boldsymbol{x};\boldsymbol{x}\nleq\boldsymbol{0}\}$. The expected number of such points is $\mathrm{E}[\bigvee_{j=1}^{d}(\gamma_jR_j/\sigma_j)^{1/\gamma_j}]$, where if $\gamma_j=0$ the component is to be interpreted as e^{R_j/σ_j}, and thus depends on the distribution of $\boldsymbol{R}$ and the parameters $\boldsymbol{\sigma},\boldsymbol{\gamma}$.

Defining $\boldsymbol{U}$ by $\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}e^{\boldsymbol{\gamma}\boldsymbol{U}}=\boldsymbol{R}$, where we use the convention that if $\gamma_j=0$ then the j-th component is given by $\sigma_jU_j=R_j$, we can write (4.2) as

$$H_U(\boldsymbol{x})=\frac{\int_0^{\infty}\left\{F_U\left(\frac{1}{\boldsymbol{\gamma}}\log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x}+1)+\log t\right)-F_U\left(\frac{1}{\boldsymbol{\gamma}}\log(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}(\boldsymbol{x}\wedge\boldsymbol{0})+1)+\log t\right)\right\}\mathrm{d}t}{\int_0^{\infty}\ \bar{F}_U(\log t)\,\mathrm{d}t}, \tag{4.4}$$

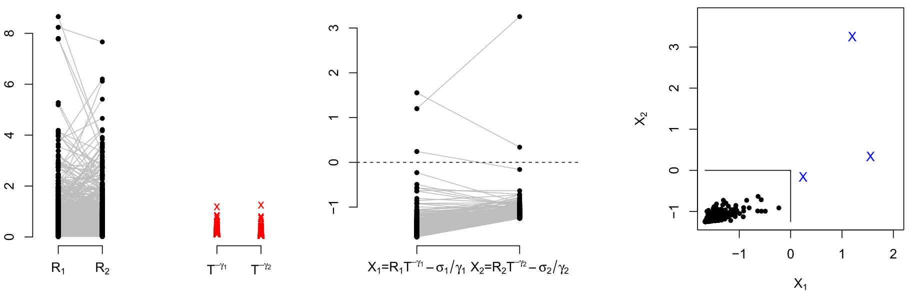

Fig. 3 Deriving the GP from the Poisson process representation. *Left:* two-dimensional illustrations of "shape vectors" $\boldsymbol{R}$ and the 1000 largest "intensities" $T^{-\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma}=(0.3,0.4)$. *Centre:* points $\boldsymbol{X}=\boldsymbol{R}/T^{\boldsymbol{\gamma}}-\boldsymbol{\sigma}/\boldsymbol{\gamma}$, where $\boldsymbol{\sigma}=(0.5,0.5)$ against index, with a horizontal line at zero. *Right:* points X_2 versus X_1 with exceedances of zero in at least one coordinate highlighted

for $\boldsymbol{x}$ such that $\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x} + \mathbf{1} > \mathbf{0}$, and where $F_{\boldsymbol{U}}$ is the cdf of $\boldsymbol{U}$. Here we assume that $0 < \mathrm{E}(e^{U_j}) < \infty$ for $j = 1, \ldots, d$, which is equivalent to assuming that $\boldsymbol{R}$ satisfies Condition 2. We write $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$ for the cdf defined by Eq. 4.4 and call it the (U) representation. The intuition parallel to Eq. 4.3 is that H_U is the formal conditional distribution of

$$\boldsymbol{\sigma}\frac{e^{\boldsymbol{\gamma}(\boldsymbol{U}-\log T)} - 1}{\boldsymbol{\gamma}}$$

given that $\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}(e^{\boldsymbol{\gamma}(\boldsymbol{U}-\log T)} - 1) \nleq \mathbf{0}$ or, equivalently, given that $\boldsymbol{U} - \log T \nleq \mathbf{0}$.

For later use we note that a $\mathrm{GP}_U(\mathbf{1}, \mathbf{0}, F_{\boldsymbol{U}})$ vector $\boldsymbol{X}_0$ has the cdf

$$H_U(\boldsymbol{x}) = \frac{\int_0^\infty \{F_{\boldsymbol{U}}(\boldsymbol{x} + \log t) - F_{\boldsymbol{U}}(\boldsymbol{x} \wedge \mathbf{0} + \log t)\}\,\mathrm{d}t}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)\,\mathrm{d}t}, \tag{4.5}$$

and that a general GP_U vector $\boldsymbol{X}$ is obtained from $\boldsymbol{X}_0$ through the transformation

$$\boldsymbol{X} = \boldsymbol{\sigma}\frac{e^{\boldsymbol{\gamma}\boldsymbol{X}_0} - 1}{\boldsymbol{\gamma}}. \tag{4.6}$$

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

$$H_S(\boldsymbol{x}) = \int_0^1 F_{\boldsymbol{S}}(\boldsymbol{x} + \log t)\,\mathrm{d}t = \int_0^\infty F_{\boldsymbol{S}}(\boldsymbol{x} - v)\,e^{-v}\,\mathrm{d}v, \tag{4.7}$$

where the second equality follows from the change of variable $\log t = -v$. Further, using the transformation (4.6) which connects (4.5) with Eq. 4.4, it follows more generally that if $\boldsymbol{U} = \boldsymbol{S}$, where $\bigvee_{j=1}^{d} S_j = 0$, then for general $\boldsymbol{\sigma}, \boldsymbol{\gamma}$ one obtains the cdf

$$H_S(\boldsymbol{x}) = \int_0^1 F_{\boldsymbol{S}}\left(\frac{1}{\boldsymbol{\gamma}}\log\left(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x} + 1\right) + \log t\right)\mathrm{d}t = \int_0^\infty F_{\boldsymbol{S}}\left(\frac{1}{\boldsymbol{\gamma}}\log\left(\frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x} + 1\right) - v\right)e^{-v}\,\mathrm{d}s. \tag{4.8}$$

We write $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{S}})$ for the cdf (4.8), and call it the (S) representation.

The last expression in Eq. 4.7 can be given an interpretation in terms of random variables: it is the cdf of $\boldsymbol{S} + E$, where E is a standard exponential variable which is independent of $\boldsymbol{S}$, and then Eq. 4.8 is the cdf of $\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}\left(e^{\boldsymbol{\gamma}(\boldsymbol{S}+E)} - \mathbf{1}\right)$. This is the Ferreira and de Haan (2014) spectral representation transformed to the exponential scale.

Theorem 4 *Suppose* $\boldsymbol{\sigma} > \mathbf{0}$. *The* $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$, $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$, *and* $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{S}})$ *classes defined by* Eqs. 4.2, 4.4, *and* 4.8, *are all equal to the class of all GP distributions with* $\boldsymbol{\sigma} > \mathbf{0}$. *For each class the conditional marginal distributions are given by* Eq. 2.9.

Proof The assertion for the $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ distributions follows from combining Eq. 3.3 with Eq. 2.6.

By definition, the class of $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$ cdf-s is the same as the class of $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ cdf-s, and thus the same conclusion holds for the $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$

cdf-s. Since $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{S}})$ cdf-s are $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$ cdf-s, it follows that they are GP distributions.

To prove the full statement about the class $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{S}})$ we first note that by the construction of the $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{S}})$ cdf-s, it is enough to prove that the statement holds for GP distributions with $\boldsymbol{\gamma} = \mathbf{0}$ and $\boldsymbol{\sigma} = \mathbf{1}$. 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 their $\boldsymbol{W}$ replaced by $e^{\boldsymbol{S}+E}$), and with $\omega_0 = 1$.

The last assertion follows by straightforward calculation. As an example we prove it for the $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ class for the case $\gamma_j \neq 0$; $j = 1, \ldots, d$. Let F_j be the marginal distribution of the j-th component of $\boldsymbol{R}$. It follows from Eqs. 2.10 and 4.2 that H_j^+, the distribution of the j-th component of $H_{\boldsymbol{R}}$ conditioned to be positive, is given by

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

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

It may be noted that since $F_{\boldsymbol{R}}$, $F_{\boldsymbol{U}}$, and $F_{\boldsymbol{S}}$ are cdf-s then also $H_{\boldsymbol{R}}$, $H_{\boldsymbol{U}}$, and $H_{\boldsymbol{S}}$ are cdf-s, so that in contrast to Eq. 2.6 the Eqs. 4.2, 4.4, and 4.8 hold for all $\boldsymbol{x} \in [-\infty, \infty)^d$, subject to the provision that Eqs. 4.4 and 4.5 only apply for $\boldsymbol{\gamma x} + \boldsymbol{\sigma} > \mathbf{0}$.

The distributions of the random vectors $\boldsymbol{R}$ and $\boldsymbol{U}$ are not uniquely determined by the corresponding GP distributions $H_{\boldsymbol{R}}$ and $H_{\boldsymbol{U}}$ in Eqs. 4.2 and 4.4, respectively. The next proposition is a generalization of Theorem 1 (vi).

Proposition 1 *Suppose that the random variable Z is strictly positive, has finite mean and is independent of $\boldsymbol{R}$ or $\boldsymbol{U}$. Then $GP_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{Z^{\boldsymbol{\gamma}}\boldsymbol{R}}) = GP_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ and $GP_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}+\log Z}) = GP_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$, where $Z^{\gamma_j}R_j$ should be interpreted to mean $R_j + \sigma_j \log Z$ if $\gamma_j = 0$.*

Proof We only prove the assertion for $\boldsymbol{R}$, since the one for $\boldsymbol{U}$ follows from it. Replacing $F_{\boldsymbol{R}}$ by $F_{Z^{\boldsymbol{\gamma}}\boldsymbol{R}}$ in the numerator and denominator of Eq. 4.2 yields, after an application of Fubini's theorem and a change of variables, a factor $\mathrm{E}(Z)$ coming out in front the integrals both in the numerator and denominator. Upon simplification, the random variable Z is seen to have had no effect on $H_{\boldsymbol{R}}$. □

Usually one would let the model for $\boldsymbol{U}$ include free location parameters for each component, and the model for $\boldsymbol{R}$ a free scale parameter for each component, in order to let data determine the relative sizes of the components. However, as one consequence of the proposition, one should then, e.g., fix the location parameter for one of the components of $\boldsymbol{U}$, or fix the sum of the components, to ensure parameter identifiability. Similarly, if $\gamma_j \neq 0$ for $j = 1, \ldots, d$ and if the model for $\boldsymbol{R}$ includes a free scale parameter for each component, then one should, e.g., fix one of these scale parameters.

5 Densities, likelihoods, and censored likelihoods

5.1 Densities

To find the densities for the (R) and (U) representations, we assume that $\boldsymbol{R}$ and $\boldsymbol{U}$ have densities with respect to Lebesgue measure on $\mathbb{R}^d$. For the (S) representation, we make the assumption that $\boldsymbol{S}$ is obtained from a vector $\boldsymbol{T}$ by setting $\boldsymbol{S} = \boldsymbol{T} - \bigvee_{j=1}^d T_j$. We write $\mathrm{GP}_T(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{T}})$ for these distributions, write H_T for the cdf-s, and call it the (T) representation. Clearly, the class of GP_T distributions is the same as the class of GP_S distributions, and hence is equal to the class of all GP distributions with $\boldsymbol{\sigma} > \boldsymbol{0}$. The densities for the (R) and (U) representations are just as would be obtained if the $\boldsymbol{R}$ and $\boldsymbol{U}$ 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 $\boldsymbol{T} - \bigvee_{j=1}^d T_j$ is contained in the $(d-1)$-dimensional set $\{\boldsymbol{x};\ \bigvee_{j=1}^d x_j = 0\}$ and hence $\boldsymbol{T} - \bigvee_{j=1}^d T_j$ does not have a density with respect to Lebesgue measure on $\mathbb{R}^d$. Nevertheless, the density of H_T exists and can be computed if $\boldsymbol{T}$ has a density with respect to Lebesgue measure on $\mathbb{R}^d$.

Theorem 5 *Suppose $\boldsymbol{\sigma} > \boldsymbol{0}$. If $F_{\boldsymbol{R}}$ has a density $f_{\boldsymbol{R}}$ on $\mathbb{R}^d$, then H_R has the density h_R given below, if $F_{\boldsymbol{U}}$ has density $f_{\boldsymbol{U}}$ on $\mathbb{R}^d$, then H_U has the density h_U below, and if $F_{\boldsymbol{T}}$ has density $f_{\boldsymbol{T}}$ on $\mathbb{R}^d$, then H_T has density h_T below:*

$$h_R(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \not\leq \boldsymbol{0}\}} \frac{1}{\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}} \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t} \int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}))\,\mathrm{d}t, \qquad (5.1)$$

$$h_U(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \not\leq \boldsymbol{0}\}} \frac{\prod_{j=1}^d (\gamma_j x_j + \sigma_j)^{-1}}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)\,\mathrm{d}t} \int_0^\infty f_{\boldsymbol{U}}(\tfrac{1}{\boldsymbol{\gamma}} \log(\tfrac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x} + \boldsymbol{1}) + \log t)\,\mathrm{d}t, \qquad (5.2)$$

$$h_T(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \not\leq \boldsymbol{0}\}} \frac{\prod_{j=1}^d (\gamma_j x_j + \sigma_j)^{-1}}{\bigvee_{j=1}^d (\frac{\gamma_j}{\sigma_j} x_j + 1)^{1/\gamma_j}} \int_0^\infty t^{-1} f_{\boldsymbol{T}}(\tfrac{1}{\boldsymbol{\gamma}} \log(\tfrac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}}\boldsymbol{x} + \boldsymbol{1}) + \log t)\mathrm{d}t, \qquad (5.3)$$

for $\boldsymbol{\gamma x} + \boldsymbol{\sigma} > \boldsymbol{0}$, and where the densities are 0 otherwise. If $\gamma_j = 0$ then for h_R the expressions $t^{\gamma_j}(x_j + \frac{\sigma_j}{\gamma_j})$ should be replaced by $x_j + \sigma_j \log t$. For h_U and h_T, 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 $\boldsymbol{\sigma} = \boldsymbol{\gamma} = \boldsymbol{1}$, and for $\boldsymbol{x} \not\leq \boldsymbol{0}$, $\boldsymbol{x} + \boldsymbol{1} > \boldsymbol{0}$. The change of variables $\boldsymbol{y} = t(\boldsymbol{z} + \boldsymbol{1})$ shows that for this case

$$\begin{aligned} F_{\boldsymbol{R}}(t(\boldsymbol{x} + \boldsymbol{1})) - F_{\boldsymbol{R}}(t(\boldsymbol{x} \wedge \boldsymbol{0} + \boldsymbol{1})) &= \int \mathbb{1}_{\{\boldsymbol{y} \leq t(\boldsymbol{x}+\boldsymbol{1}),\ \boldsymbol{y} \not\leq t(\boldsymbol{x} \wedge \boldsymbol{0} + \boldsymbol{1})\}} f_{\boldsymbol{R}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \\ &= \int \mathbb{1}_{\{\boldsymbol{z} \leq \boldsymbol{x},\ \boldsymbol{z} \not\leq \boldsymbol{x} \wedge \boldsymbol{0}\}} t^d f_{\boldsymbol{R}}(t(\boldsymbol{z} + \boldsymbol{1}))\,\mathrm{d}\boldsymbol{z}. \end{aligned}$$

Hence, by Eq. 4.2, and using Fubini's theorem,

$$\begin{aligned} H_R(\boldsymbol{x}) &= \frac{\int_{t=0}^{\infty} \int \mathbb{1}_{\{\boldsymbol{z}\le\boldsymbol{x},\, \boldsymbol{z}\nleq\boldsymbol{x}\wedge\boldsymbol{0}\}} t^d f_{\boldsymbol{R}}(t(\boldsymbol{z}+\boldsymbol{1}))\,\mathrm{d}\boldsymbol{z}\mathrm{d}t}{\int_0^{\infty} F_{\boldsymbol{R}}(t\boldsymbol{1})\,\mathrm{d}t} \\ &= \int \mathbb{1}_{\{\boldsymbol{z}\le\boldsymbol{x},\, \boldsymbol{z}\nleq\boldsymbol{x}\wedge\boldsymbol{0}\}} \frac{\int_{t=0}^{\infty} t^d f_{\boldsymbol{R}}(t(\boldsymbol{z}+\boldsymbol{1}))\,\mathrm{d}t}{\int_0^{\infty} F_{\boldsymbol{R}}(t\boldsymbol{1})\,\mathrm{d}t} \mathrm{d}\boldsymbol{z} \\ &= \int_{(-\infty,\boldsymbol{x}]} \mathbb{1}_{\{\boldsymbol{z}\nleq\boldsymbol{0}\}} \frac{\int_{t=0}^{\infty} t^d f_{\boldsymbol{R}}(t(\boldsymbol{z}+\boldsymbol{1}))\,\mathrm{d}t}{\int_0^{\infty} F_{\boldsymbol{R}}(t\boldsymbol{1})\,\mathrm{d}t} \mathrm{d}\boldsymbol{z}. \end{aligned}$$

We conclude that Eq. 5.1 holds for $\boldsymbol{\sigma} = \boldsymbol{\gamma} = \boldsymbol{1}$. The proof of the general form of Eq. 5.1 only differs from this case in bookkeeping details, and is omitted.

To prove (5.2), recall that $H_U = H_R$ if $\boldsymbol{\gamma} = \boldsymbol{0}$, $\boldsymbol{\sigma} = \boldsymbol{1}$, so that, by (5.1),

$$h_U(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x}\nleq\boldsymbol{0}\}} \frac{\int_{t=0}^{\infty} f_{\boldsymbol{U}}(\boldsymbol{x}+\log t)\,\mathrm{d}t}{\int_0^{\infty} F_{\boldsymbol{U}}(\log t)\,\mathrm{d}t}.$$

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

We again only prove (5.3) for the case $\boldsymbol{\sigma} = \boldsymbol{1}$, $\boldsymbol{\gamma} = \boldsymbol{0}$, and with $\boldsymbol{x} \nleq \boldsymbol{0}$. Also here extension to the general case is a chain rule argument. It follows from Eq. 4.7 that

$$H_T(\boldsymbol{x}) = \int_{t=0}^{1} F_{\boldsymbol{T}-\bigvee_{j=1}^{d} T_j}(\boldsymbol{x}+\log t) = \int_{t=0}^{\infty}\int_{\boldsymbol{s}} \mathbb{1}_{\{0\le t\le 1,\ \boldsymbol{s}-\vee_{j=0}^{d} s_j \le \boldsymbol{x}+\log t\}} f_{\boldsymbol{T}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\,\mathrm{d}t.$$

Hence, using first Fubini's theorem, then a change of variables from $te^{\bigvee_{j=1}^{d} s_j}$ to t and Fubini's theorem, and finally a change of variables from $\boldsymbol{s}$ to $\boldsymbol{s}+\log t$ and Fubini's theorem,

$$\begin{aligned} H_T(\boldsymbol{x}) &= \int_{\boldsymbol{s}}\int_{t=0}^{\infty} e^{-\bigvee_{j=1}^{d} s_j} \mathbb{1}_{\{0\le te^{-\bigvee_{j=1}^{d} s_j}\le 1,\ \boldsymbol{s}\le\boldsymbol{x}+\log(t)\}} f_{\boldsymbol{T}}(\boldsymbol{s})\,\mathrm{d}t\,\mathrm{d}\boldsymbol{s} \\ &= \int_{t=0}^{\infty}\int_{\boldsymbol{s}} e^{-\bigvee_{j=1}^{d} s_j} t^{-1} \mathbb{1}_{\{e^{-\bigvee_{j=1}^{d} s_j}\le 1,\ \boldsymbol{s}\le\boldsymbol{x}\}} f_{\boldsymbol{T}}(\boldsymbol{s}+\log t)\,\mathrm{d}\boldsymbol{s}\,\mathrm{d}t \\ &= \int_{-\infty}^{\boldsymbol{x}} \mathbb{1}_{\{\boldsymbol{s}\nleq\boldsymbol{0}\}} e^{-\bigvee_{j=1}^{d} s_j} \int_{t=0}^{\infty} t^{-1} f_{\boldsymbol{T}}(\boldsymbol{s}+\log t)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{s}. \end{aligned}$$

This proves that Eq. 5.3 holds for $\boldsymbol{\gamma} = \boldsymbol{0}$ and $\boldsymbol{\sigma} = \boldsymbol{1}$. □

In some cases, the integrals in Eqs. 5.1, 5.2 and 5.3 can be computed explicitly; see the examples below. Otherwise the one-dimensional integrals allow for fast numerical computation as soon as one can compute densities and distribution functions of $\boldsymbol{R}$ or $\boldsymbol{U}$ efficiently. Either way, this can make full likelihood inference possible, also in high dimensions.

5.2 Censored likelihood

Sometimes one does not trust the GP distribution to fit the excesses well on the entire set $\boldsymbol{x} \nleq \boldsymbol{0}$. Then, instead of using a full likelihood obtained as a product of the densities in Theorem 5, one can use a censored likelihood which is based on the values of the excesses which are larger than some censoring threshold $\boldsymbol{v} = (v_1, \ldots, v_d)$. This idea was introduced for multivariate extremes in Smith et al. (1997), and has since become a standard approach to inference. Huser et al. (2015) explore the merits of this and other approaches via simulation.

Write $D = \{1, \ldots, d\}$, 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 $\boldsymbol{x}_A = \{x_j;\, j \in A\}$ and writing h for h_R, h_U or h_T, the likelihood contribution of a censored observation is

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

For certain models, the $|C|-$dimensional integral in Eq. 5.4 can be avoided, which is advantageous from a practical perspective.

Example 3 The simplest situation is when the components of the shape vector $\boldsymbol{R}$ are mutually independent. This could e.g. be a model for windspeeds over a small area, perhaps a wind farm, with $T^{-\boldsymbol{\gamma}}$ representing the intensity of the average geostrophic wind and with the components of $\boldsymbol{R}$ representing random wind variations caused by local turbulence.

Let f_j be the density function of R_j, the jth component of $\boldsymbol{R}$, 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 the numerator in Eq. 5.4 for $h = h_R$ in Eq. 5.1 can then be written as

$$\mathbb{1}_{\{\boldsymbol{x}_{D\setminus C} \nleq \boldsymbol{0}\}} \int_0^\infty t^{\sum_{j\in D\setminus C} \gamma_j} \prod_{j\in C} F_j(t^{\gamma_j} v_j) \prod_{j \in D\setminus C} f_j(t^{\gamma_j} y_j)\,\mathrm{d}t$$

and the integral in the denominator equals $\int_0^\infty \{1 - \prod_{j=1}^d F_j(t^{\gamma_j}\sigma_j/\gamma_j)\}\,\mathrm{d}t$. Here quick numerical computation of both integrals is typically possible.

Sometimes these integrals can also be computed analytically, and similarly for the corresponding integrals for h_U and h_T. As a simple example, consider (5.3) with $\boldsymbol{\gamma} = \boldsymbol{0}$ and $\boldsymbol{\sigma} = \boldsymbol{1}$ and with the components of $\boldsymbol{T}$ having independent standard Gumbel distributions with cdf $\exp\{-e^{-x}\}$. Then, with c the number of elements in

C, i.e., the number of censored components, and abbreviating $\mathbb{1}_{\{\boldsymbol{x}_{D\setminus C} \not\leq \mathbf{0}\}}$ to $\mathbb{1}_{D\setminus C}$, we obtain that

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

Whilst the previous example is a theoretical illustration, the class of GP distributions obtained by letting $\boldsymbol{R}$ (or $\boldsymbol{U}$) have independent components with parametrized marginal distributions does make for a large and flexible class of models. It includes, for example, the GP distributions associated to the commonly used logistic and negative logistic max-stable distributions. For this and further examples, see Kiriliouk et al. (2016).

5.3 Further examples

We illustrate two further constructions with tractable densities. The first is a toy example to exhibit the idea of building process knowledge into a model. The second is a variation on existing extreme value models based on lognormal distributions.

Example 4 An extreme flow episode in a river network consisting of two tributaries which join to form the main river could be modeled as $\boldsymbol{R}/T^{\boldsymbol{\gamma}} = (R_1/T^\gamma, R_2/T^\gamma, (R_1 + R_2 + E)/T^\gamma)$, with $\gamma > 0$, so that $R_3 = R_1 + R_2 + E$. Here the first component corresponds to flow in tributary number one, the second component to flow in tributary number two, and the third component to flow in the main river. The simplest model is that R_1, R_2, E are independent and have a standard exponential distribution. Then,

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

Assuming in addition that $\boldsymbol{\sigma} = (\sigma, \sigma, \sigma)$, we have

$$\begin{aligned} \int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}} \boldsymbol{\sigma}/\boldsymbol{\gamma})\,\mathrm{d}t &= \mathrm{E}[\textstyle\bigvee_{j=1}^3 (R_j\, \gamma/\sigma)^{1/\gamma}] \\ &= (\gamma/\sigma)^{1/\gamma}\, \mathrm{E}[R_3^{1/\gamma}] \\ &= (\gamma/\sigma)^{1/\gamma}\Gamma(3+1/\gamma)/2, \end{aligned}$$

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

$$h_R(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \nleq \boldsymbol{0},\, -\boldsymbol{\sigma}/\boldsymbol{\gamma} \leq \boldsymbol{x},\, x_1 + x_2 \leq x_3\}}\, \gamma^{-1} 2(\sigma/\gamma)^{1/\gamma}\, (x_3 + \sigma/\gamma)^{-3-1/\gamma}.$$

Example 5 Lognormal distributions have been used in max-stable modelling, e.g., in Huser and Davison (2013), and as point process models in Wadsworth and Tawn (2014), and are an important class of models. As an example, in the (R) representation, suppose that $\boldsymbol{0} \leq \boldsymbol{\gamma}$ and that $F_{\boldsymbol{R}}(\boldsymbol{x}) = \Phi(\log \boldsymbol{x})$, where Φ is the cdf of a d-dimensional normal distribution with mean $\boldsymbol{\mu}$ and nonsingular covariance matrix Σ. Write ϕ for the corresponding density and let $A = \Sigma^{-1}$ be the precision matrix. Then, writing $\boldsymbol{y} = \log(\boldsymbol{x} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}) - \boldsymbol{\mu}$, we have

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

Making the change of variables from $\log t$ to t and completing the square, we can evaluate the integral, finding

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

The integral in the denominator can be expressed as a sum of d components, each of which involves a $(d-1)$-dimensional normal cdf, see Huser and Davison (2013). However, if d is large then this expression is cumbersome. Inference methods for similar high-dimensional models are explored in de Fondeville and Davison (2016).

6 Probabilities and conditional probabilities

Equations 4.2, 4.4, and 4.8 give probabilities of rectangles for GP distributions, on the real scale. In this section they are generalized to expressions for probabilities of general sets and for conditional probabilities. Below, we only consider GP_R models. It is straightforward to derive the corresponding formulas for the other representations.

Let $F = \{\boldsymbol{y};\ \boldsymbol{y} \nleq \boldsymbol{0}\}$, set $A = \{\boldsymbol{y};\ \boldsymbol{y} \leq \boldsymbol{x}\}$, and for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ and a set $B \subset \mathbb{R}^d$ write $\boldsymbol{a}(B + \boldsymbol{b})$ for the set $\{\boldsymbol{a}(\boldsymbol{y} + \boldsymbol{b});\ \boldsymbol{y} \in B\}$. As is easily checked,

$$H_R(\boldsymbol{x}) = H_R(A) = \frac{\int_0^\infty \mathrm{P}[\boldsymbol{R} \in t^{\boldsymbol{\gamma}}(A \cap F + \boldsymbol{\sigma}/\boldsymbol{\gamma})]\, \mathrm{d}t}{\int_0^\infty \mathrm{P}[\boldsymbol{R} \in t^{\boldsymbol{\gamma}}(F + \boldsymbol{\sigma}/\boldsymbol{\gamma})]\, \mathrm{d}t}. \tag{6.1}$$

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

$$H_R(A) = \frac{\int_0^\infty \mathrm{P}[\boldsymbol{R} \in t^{\boldsymbol{\gamma}}(A \cap F + \boldsymbol{\sigma}/\boldsymbol{\gamma})]\,\mathrm{d}t}{\int_0^\infty \mathrm{P}[\boldsymbol{R} \in t^{\boldsymbol{\gamma}}(F + \boldsymbol{\sigma}/\boldsymbol{\gamma})]\,\mathrm{d}t} = \frac{\int_0^\infty \mathrm{P}[\boldsymbol{R}/t^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \in A \cap F]\,\mathrm{d}t}{\int_0^\infty \mathrm{P}[\boldsymbol{R}/t^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \in F]\,\mathrm{d}t}. \tag{6.2}$$

A proof that Eq. 6.2 holds for any set A is immediate: using Fubini's theorem it is seen that the right-hand 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 $\{\boldsymbol{y};\ \boldsymbol{y} \le \boldsymbol{x}\}$, the two distributions are equal. The intuition is that $H_R(A)$ is the (formal) conditional probability of the event $\{\boldsymbol{R}/T^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \in A\}$ given the event $\{\boldsymbol{R}/T^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \not\le \boldsymbol{0}\}$.

Let the random vector $\boldsymbol{X}$ have the distribution H_R in Eq. 4.2. Then $\mathrm{P}[\boldsymbol{X} \in A \mid \boldsymbol{X} \in B] = H_R(A \cap B)/H_R(B)$, and hence (6.2) can also be used to find conditional probabilities. Further, assuming continuity, Eq. 5.1 determines the conditional densities. For instance, writing $f_{\mid X_1 = x}$ for the conditional density of $(X_2, \dots X_d)$ given that $X_1 = x$, we find, for $x > 0$,

$$f_{\mid X_1 = x}(x_2, \dots x_d) = \frac{\int_0^\infty t^{\sum_{j=1}^d \gamma_j} f_{\boldsymbol{R}}\left(t^{\boldsymbol{\gamma}}((x, x_2, \dots x_d) + \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\right)\,\mathrm{d}t}{\int_0^\infty t^{\gamma_1} f_{R_1}\left(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})\right)\,\mathrm{d}t}. \tag{6.3}$$

By further integration, it follows that

$$\begin{aligned} &\mathrm{P}[\boldsymbol{X} \in A \mid X_1 = x] \\ &= \frac{\int_0^\infty t^{\gamma_1} f_{R_1}\left(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})\right) \mathrm{P}[(x, R_2, \dots R_d)/t^{\boldsymbol{\gamma}} - \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}} \in A \mid R_1 = t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})]\,\mathrm{d}t}{\int_0^\infty t^{\gamma_1} f_{R_1}\left(t^{\gamma_1}(x + \frac{\sigma_1}{\gamma_1})\right)\,\mathrm{d}t}. \end{aligned} \tag{6.4}$$

Example 6 In Example 4, extreme flow episodes in the two river tributaries are modelled using $(R_1/T^\gamma, R_2/T^\gamma)$ with R_1 and R_2 independent standard exponential variables and with $\gamma > 0$. Suppose $\boldsymbol{X} \sim H$ where H is the GP distribution obtained from $(R_1/T^\gamma, R_2/T^\gamma)$ and let $s > 0$. Since $R_1 + R_2$ has a gamma distribution, it is straightforward to evaluate (6.2) to find the distribution of the sum of the flows in the two tributaries:

$$\mathrm{P}[X_1 + X_2 > s] = c_1\left(1 + \tfrac{\gamma}{\sigma_1 + \sigma_2}s\right)^{-1/\gamma}, \tag{6.5}$$

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 Eq. 6.3 show that for $x_1, x_2 > 0$

$$f_{\mid X_1 = x_1}(x_2) = c_2\left(1 + \frac{\gamma/(1+\gamma)}{(\gamma x_1 + \sigma_1 + \sigma_2)/(1+\gamma)}\ x_2\right)^{-1-1/[\gamma/(1+\gamma)]}$$

and

$$\mathrm{P}[X_2 > x_2 \mid X_1 = x_1] = c_3 \left(1 + \frac{\gamma/(1+\gamma)}{(\gamma x_1 + \sigma_1 + \sigma_2)/(1+\gamma)}\, x_2\right)^{-1/[\gamma/(1+\gamma)]},$$

for $c_2 = (1+\gamma)(x_1 + \sigma_1/\gamma)^{1+1/\gamma}(\gamma x_1 + \sigma_1 + \sigma_2)^{-2-1/\gamma}$ and $c_3 = (x_1 + \sigma_1/\gamma)^{1+1/\gamma}(\gamma x_1 + \sigma_1 + \sigma_2)^{-1-1/\gamma}$. Hence, dividing (6.5) with the same expression with s set to zero, we find that 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. 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, $\gamma/(1+\gamma)$.

Many of the results in Example 6 hold more generally. For instance, 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 $(R_1/T^\gamma - \sigma_1/\gamma, \ldots, R_d/T^\gamma - \sigma_d/\gamma)$ 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 $\sigma = \sum_{j=1}^d a_j \sigma_j$, with coefficients $a_1, \ldots, a_d$. According to the GP$_R$ representation, provided $a_1, \ldots, a_d \geq 0$, 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. Here, we only prove the one-dimensional result.

Proposition 2 *Let $\boldsymbol{X}$ be a GP random vector with common shape parameter γ for all d margins and with scale parameter $\boldsymbol{\sigma} > \boldsymbol{0}$, and if $\gamma \leq 0$ additionally assume that* $\mathrm{P}[\sum_{j=1}^d a_j X_j > 0] > 0$. *Then, for $\boldsymbol{a} \in [\boldsymbol{0}, \boldsymbol{\infty}) \setminus \{\boldsymbol{0}\}$, 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 $\mathrm{P}[\sum_{j=1}^d a_j X_j > 0] > 0$ holds automatically if $\gamma > 0$ and $\boldsymbol{\sigma} > \boldsymbol{0}$, this condition is satisfied for all values of γ. Let $A_x = \{\boldsymbol{y} \in \mathbb{R}^d \mid \sum_{j=1}^d a_j y_j > x\}$ and as above define $R = \sum_{j=1}^d a_j R_j$. Then, for $x > 0$ and with $F = \{\boldsymbol{y}; \boldsymbol{y} \nleq \boldsymbol{0}\}$, as above, $A_x \cap F = A_x$, and [for $\gamma = 0$ using the convention that $t^0(x + \sigma/0)$ means $x + \sigma \log t$] the numerator in Eq. 6.1 for $A = A_x$ is

$$\int_0^\infty \mathrm{P}[\boldsymbol{R}/t^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma} \in A_x]\,\mathrm{d}t = \int_0^\infty \mathrm{P}[R/t^\gamma - \sigma/\gamma > x]\,\mathrm{d}t,$$

and hence by Eq. 6.1

$$\mathrm{P}\Big[\textstyle\sum_{j=1}^{d} a_j X_j > x \,\Big|\, \sum_{j=1}^{d} a_j X_j > 0\Big] = \frac{H_R(A_x)}{H_R(A_0)} = \frac{\int_0^\infty \mathrm{P}[R/t^\gamma - \sigma/\gamma > x]\,\mathrm{d}t}{\int_0^\infty \mathrm{P}[R/t^\gamma - \sigma/\gamma > 0]\,\mathrm{d}t}$$
$$= (1 + \tfrac{\gamma}{\sigma} x)^{-1/\gamma},$$

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

Example 1 exhibits a situation where the component sum in a GP distribution is identically equal to $-\infty$ and hence the assumption $\mathrm{P}[X_1 + X_2 > 0] > 0$ is not satisfied.

7 Simulation

In this section we outline four methods for sampling from multivariate GP distributions. For Methods 1 to 3 we focus on simulation of a GP vector $\boldsymbol{X}_0$ with $\boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$, since a vector $\boldsymbol{X}$ with general $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$ is obtained at once from the vector $\boldsymbol{X}_0$ through (4.6). Furthermore, using the connection between GP$_U$ and GP$_R$ distributions, GP$_R$ vectors may be obtained by simulating GP$_U$ vectors, and vice versa. Throughout we assume that simulation of $\boldsymbol{U}$ from $F_{\boldsymbol{U}}$ and $\boldsymbol{T}$ from $F_{\boldsymbol{T}}$ is possible. Recall the relation $\boldsymbol{S} = \boldsymbol{T} - \bigvee_{j=1}^{d} T_j$ which was used to define the (T) representation. The first method follows immediately from Eq. 4.7.

Method 1: simulation from the (T) *representation.* Simulate a vector $\boldsymbol{T} \sim F_{\boldsymbol{T}}$ and an independent variable $E \sim \mathrm{Exp}(1)$ and set $\boldsymbol{X}_0 = E + \boldsymbol{T} - \max_{1 \le j \le d} T_j$.

Simulation from the (R) and (U) representations is less direct. We propose three methods: rejection sampling, MCMC sampling, and approximate simulation using (4.1). The idea in Methods 2 and 3 is to use an appropriate change of measure so that Method 1 can be used to simulate from the (T) representation. The GP$_T(\mathbf{1}, \mathbf{0}, F_{\boldsymbol{T}})$ density is

$$h_T(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \not\le \mathbf{0}\}}\, e^{-\bigvee_{j=1}^{d} x_j} \int_0^\infty t^{-1} f_{\boldsymbol{T}}(\boldsymbol{x} + \log t)\,\mathrm{d}t.$$

If in this equation one replaces $\boldsymbol{T}$ by $\boldsymbol{T}_0$ where $\boldsymbol{T}_0$ has density

$$f_{\boldsymbol{T}_0}(\boldsymbol{x}) = \frac{e^{\bigvee_{j=1}^{d} x_j} f_{\boldsymbol{U}}(\boldsymbol{x})}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)\,\mathrm{d}t} \tag{7.1}$$

then

$$h_T(\boldsymbol{x}) = \mathbb{1}_{\{\boldsymbol{x} \not\le \mathbf{0}\}} \frac{e^{-\bigvee_{j=1}^{d} x_j} \int_0^\infty t^{-1} e^{\bigvee_{j=1}^{d}(x_j + \log t)} f_{\boldsymbol{U}}(\boldsymbol{x} + \log t)\,\mathrm{d}t}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)\,\mathrm{d}t}$$
$$= \mathbb{1}_{\{\boldsymbol{x} \not\le \mathbf{0}\}} \frac{\int_0^\infty f_{\boldsymbol{U}}(\boldsymbol{x} + \log t)\,\mathrm{d}t}{\int_0^\infty \bar{F}_{\boldsymbol{U}}(\log t)\,\mathrm{d}t} = h_U(\boldsymbol{x}).$$

Thus, if one can simulate $\boldsymbol{T}_0$ vectors, then these give $\mathrm{GP}_U(\mathbf{1}, \mathbf{0}, F_U)$ vectors via Method 1.

Method 2: simulation of $\boldsymbol{T}_0$ *via rejection sampling.* Let φ be a probability density function which satisfies $f_{\boldsymbol{T}_0}(\boldsymbol{x}) \leq K\varphi(\boldsymbol{x})$, for some constant $K > 0$. Draw a candidate vector $\boldsymbol{T}_0^c$ from φ and accept the candidate with probability $f_{\boldsymbol{T}_0}(\boldsymbol{T}_0^c)/[K\varphi(\boldsymbol{t}_0^c)]$, and repeat otherwise. Use the accepted vector as input $\boldsymbol{T}$ in Method 1.

The acceptance probability is $1/K$, and thus it is advantageous to find a φ such that K is not too large. In high dimensions however, such a φ might be difficult to find.

Method 3: simulation of $\boldsymbol{T}_0$ *via MCMC.* Use a standard Metropolis–Hastings algorithm to simulate from a Markov chain with stationary distribution (7.1). At iteration i, draw a candidate vector $\boldsymbol{T}_0^c$ from the density $f_{\boldsymbol{T}}$ and accept the candidate with probability $\min\{1, \exp(\bigvee_{j=1}^d T_{0,j}^c - \bigvee_{j=1}^d T_{0,j}^{i-1})\}$, where $\boldsymbol{T}_0^{i-1}$ is the current state of the chain. If the candidate is not accepted, then the previous state of the chain is repeated. After a suitable burn-in time, values of the chain should represent dependent samples from Eq. 7.1; the draws can be thinned to produce approximately independent replicates. Use the simulated values of the chain as inputs $\boldsymbol{T}$ to Method 1.

Alternative proposal distributions could be used with appropriate modification of the acceptance probability; for details see e.g. Chib and Greenberg (1995).

By Eq. 4.1, an approximate way to simulate $\boldsymbol{X} \sim \mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ is as follows.

Method 4: approximate simulation from the (R) *representation.* Choose a large $K > 0$. Simulate $\bar{T} \sim \mathrm{Unif}\,[0, K]$ and an independent $\boldsymbol{R} \sim F_{\boldsymbol{R}}$. If $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} \nleq \boldsymbol{\sigma}/\boldsymbol{\gamma}$ set $\boldsymbol{X} = \boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}} - \boldsymbol{\sigma}/\boldsymbol{\gamma}$, and repeat otherwise.

In this algorithm, the probability to keep a simulated $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}}$ value is

$$\frac{1}{K}\int_0^K \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t \approx \frac{1}{K}\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\tfrac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t,$$

so one has to simulate approximately $K/\int_0^\infty \bar{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}})\,\mathrm{d}t$ values of $\boldsymbol{R}/\bar{T}^{\boldsymbol{\gamma}}$ to get one $\boldsymbol{X}$-value. Hence a large K, which ensures that the approximating distribution $H^{(K)}$ is close to H, leads to longer computation times, and a compromise has to be made. As a guide to the compromise, it is often, e.g. for Gaussian or log-Gaussian processes, possible to compute, analytically or numerically, sharp bounds for the approximation errors.

To summarize, Method 1 is simplest, but only produces $\mathrm{GP}_T(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_T)$ vectors. Method 2 and Method 3 provide ways to simulate vectors $\boldsymbol{T}_0$ from distribution (7.1), which can then be inserted into Method 1 to simulate from the $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{U}})$ and $\mathrm{GP}_R(\boldsymbol{\sigma}, \boldsymbol{\gamma}, F_{\boldsymbol{R}})$ distributions. Method 4 is as simple to program as Method 1 and produces i.i.d. vectors, but, similarly to Method 3, only approximates the target distribution.

8 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 behaviour under conditioning; scale change; convergence in distribution; mixing; and connections with generalized extreme value distributions. The main results are a point process limit result which gives a general and concrete description of the behaviour of extreme episodes; new 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 results are illustrated by examples.

We provided four different representations of GP distributions, labelled (R), (S), (T), and (U). Computationally, the (T) densities are simplest, and simulation from the (T) representation also is simpler than simulation from the other representations. On the other hand, it seems impractical to use the (T) representation for prediction or spatial modelling, since taking lower-dimensional margins of it do not simply lead to the proper lower-dimensional (T) representations, and since a d-dimensional (T) representation does not include any prescription for how to extend it to a $(d+1)$-dimensional one. The (S), (T) and (U) representations allow for smooth transitions from positive to negative γ_j, in contrast to the (R) representation. In some situations, however, requirements of realistic physical modelling can nevertheless lead to the use of the (R) representation.

Peaks over thresholds modelling of extremes of a random vector $\boldsymbol{Y}$ first selects a suitable level $\boldsymbol{u}$ and then models the distribution of the over- and undershoots, $\boldsymbol{X} = \boldsymbol{Y} - \boldsymbol{u}$, conditional on the occurrence of at least one overshoot, by a GP distribution. Of course, this GP model also models the conditional distribution of the original vector $\boldsymbol{Y}$, since $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{u}$. Modelling issues which are not treated include choice of the level $\boldsymbol{u}$, perhaps as a function of covariates like time, and modelling of the Poisson process which governs the occurrence of extreme episodes.

A further practical issue, which is outside the scope of the current paper, is that of asymptotic independence of extremes. In the event that the limiting probability of joint occurrence of extremes, conditional upon at least one extreme component, is zero, multivariate GP distributions will typically not represent the best models. Asymptotic independence is usually manifested in practice by the threshold stability properties of multivariate GP distributions not holding. Diagnostics based on these stability properties are presented in Kiriliouk et al. (2016).

The paper gives a basis for understanding and modelling of extreme episodes. We believe it will contribute to the solution of many different and important risk handling problems. However, it still is an early excursion into new territory, and much research remains to be done. Important challenges include incorporating temporal dependence and developing methods for prediction of the unfolding of extreme episodes.

Acknowledgments Anna Kiriliouk has throughout participated in the discussions leading to this paper. We thank her and two referees for many helpful comments. Research supported by the Knut and Alice

Wallenberg foundation. 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.

Appendix

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

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

As a consequence,

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

as required. □

The following property was used in the course of the proof of Theorem 1(vi).

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

References

Andersen, C.F., et al.: The New Orleans Hurricane Protection System: What Went Wrong and Why. A Report by the ASCE Hurricane Katrina External Review Panel. American Society of Civil Engineers (2007)

Aulbach, S., Bayer, V., Falk, M., et al.: A multivariate piecing-together approach with an application to operational loss data. Bernoulli **18**(2), 455–475 (2012)

Balkema, A.A., de Haan, L.: Residual life time at great age. Ann. Probab. **2**(5), 792–804 (1974)

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

Brodin, E., Rootzén, H.: Univariate and bivariate GPD methods for predicting extreme wind storm losses. Insurance: Math. Econ. **44**(3), 345–356 (2009)

Chib, S., Greenberg, E.: Understanding the Metropolis-Hastings algorithm. Am. Stat. **49**(4), 327–335 (1995)

Coles, S.G.: An Introduction to Statistical Modeling of Extreme Values. Springer (2001)

Coles, S.G., Tawn, J.A.: Modelling extreme multivariate events. J. R. Stat. Soc. Ser. B (Stat Methodol.) **53**(2), 377–392 (1991)

Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modeling of spatial extremes. Stat. Sci. **27**(2), 161–186 (2012)

Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds. J. R. Stat. Soc. Ser. B (Stat Methodol.) **52**(3), 393–442 (1990)

de Fondeville, R., Davison, A.C.: High-dimensional peaks-over-threshold inference for the Brown–Resnick process. arXiv:1605.08558 (2016)

de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer (2006)

de Haan, L., Neves, C., Peng, L.: Parametric tail copula estimation and model testing. J. Multivar. Anal. **99**(6), 1260–1275 (2008)

Dey, D.K., Yan, J., Extreme Value Modeling and Risk Analysis: Methods and Applications. Chapman and Hall/CRC (2015)

Einmahl, J., Kiriliouk, A., Krajina, A., Segers, J.: An M-estimator of spatial tail dependence. J. R. Stat. Soc. Ser. B (Stat Methodol.) **78**(1), 275–298 (2016)

Einmahl, J.H.J., Krajina, A., Segers, J.: An M-estimator for tail dependence in arbitrary dimensions. Ann. Stat. **40**(3), 1764–1793 (2012)

Falk, M., Hüsler, J., Reiss, R.-D.: Laws of Small Numbers: Extremes and Rare Events. Springer Science & Business Media (2010)

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

Grynszpan, D.: Lessons from the french heatwave. Lancet **362**, 1169–1170 (2003)

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.: Rainfall thresholds for the initiation of landslides in Central and Southern Europe. Meteorol. Atmos. Phys., 239–267 (2007)

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

Huser, R., Davison, A.C., Genton, M.G.: Likelihood estimators for multivariate extremes. Extremes **19**(1), 79–103 (2015)

Joe, H., Smith, R.L., Weissman, I.: Bivariate threshold methods for extremes. J. R. Stat. Soc. Ser. B Methodol. **54**(1), 171–183 (1992)

Katz, R.W., Parlange, M.B., Naveau, P.: Statistics of extremes in hydrology. Adv. Water Resour. **25**(1), 1287–1304 (2002)

Kiriliouk, A., Rootzén, H., Segers, J., Wadsworth, J.: Peaks over thresholds modelling with multivariate generalized Pareto distributions. arXiv:1612.01773 (2016)

Kyselý, J., Picek, J., Beranová, R.: Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. Global Planet. Change **72**(1–2), 55–68 (2010)

Marshall, A.W., Olkin, I.: Domains of attraction of multivariate extreme value distributions. Ann. Probab. **11**(1), 168–177 (1983)

McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts Techniques and Tools. Princeton University Press, Princeton (2015)

Michel, R.: Parametric estimation procedures in multivariate generalized Pareto models. Scand. J. Stat. **36**(1), 60–75 (2009)

NERC: Flood Studies Report. Natural Environment Research Council, London (1975)

Penrose, M.D.: Semi-minstable processes. Ann. Probab. **20**(3), 1450–1463 (1992)

Pickands, J.I.: Statistical inference using extreme order statistics. Ann. Stat. **3**(1), 119–131 (1975)

Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer (1987)

Resnick, S.I.: Heavy-Tail Phenomena, Probabilistic and Statistical Modelling. Springer (2007)

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

Schlather, M.: Models for stationary max-stable fields. Extremes **5**(1), 61–82 (2002)

Smith, R.L.: Threshold methods for sample extremes. In: de Oliveira, J.T. (ed.) Statistical extremes and applications, pp. 621–638. D. Reidl, Dordrecht (1984)

Smith, R.L., Tawn, J.A., Coles, S.G.: Markov chain models for threshold exceedances. Biometrika **84**(2), 249–268 (1997)

Tajvidi, N.: Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions. Ph. D. thesis Department of Mathematics. Chalmers University of Technology, Göteborg (1996)

Wadsworth, J., Tawn, J.: Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. Biometrika **101**(1), 1–15 (2014)