Journal of Multivariate Analysis 165 (2018) 117–131

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Multivariate generalized Pareto distributions: Parametrizations, representations, and properties

Holger Rootzén [a], Johan Segers [b,*], Jennifer L. Wadsworth [c]

[a] Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden
[b] Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
[c] Mathematics and Statistics, Lancaster University, Fylde College, Lancaster, UK

ARTICLE INFO

Article history:
Received 22 May 2017
Available online 15 December 2017

AMS subject classifications:
62G32
62E15

Keywords:
Exceedances
Maxima
Stable tail dependence function
Tail copula
Linear combination

ABSTRACT

Multivariate generalized Pareto distributions arise as the limit distributions of exceedances over multivariate thresholds of random vectors in the domain of attraction of a max-stable distribution. These distributions can be parametrized and represented in a number of different ways. Moreover, generalized Pareto distributions enjoy a number of interesting stability properties. An overview of the main features of such distributions is given, expressed compactly in several parametrizations, giving the potential user of these distributions a convenient catalogue of ways to handle and work with generalized Pareto distributions.

1. Introduction

A core theme in univariate extreme-value analysis is to fit a generalized Pareto (GP) distribution to a sample of excesses over a high threshold. Since univariate GP distributions can be described in terms of a scale parameter and a shape parameter, statistical inference using frequentist or Bayesian likelihood techniques is straightforward, at least for values of the shape parameter at which the Fisher information matrix is finite.

For multivariate extremes, matters are more complicated. First, there is no universal definition of an exceedance of a multivariate threshold. Second, whatever the definition that is selected, the family of distributions proposed by asymptotic theory is no longer parametric.

Following Rootzén and Tajvidi [17], we say that a sample point $\boldsymbol{y} \in \mathbb{R}^d$ exceeds a multivariate threshold $\boldsymbol{u} \in \mathbb{R}^d$ as soon as one of its coordinates exceeds the corresponding threshold coordinate, i.e., $y_j > u_j$ for at least one $j \in \{1, \ldots, d\}$. In dimension $d = 2$, the shape of the excess region $\{\boldsymbol{y} \in \mathbb{R}^d : \boldsymbol{y} \not\leq \boldsymbol{u}\}$ is that of the letter L upside-down; here and in what follows, inequalities between vectors are meant component-wise. The excess region covers a larger part of the sample space than the one for most other threshold exceedance definitions, for instance, that $\boldsymbol{y}$ exceeds $\boldsymbol{u}$ when $\boldsymbol{y} > \boldsymbol{u}$, i.e., $y_j > u_j$ for all $j \in \{1, \ldots, d\}$.

The class of GP distributions that arises from the first definition of a multivariate exceedance is derived directly from the family of multivariate generalized extreme-valued (GEV) or max-stable distributions; see, e.g., Beirlant et al. [2, Section 8.3] or Rootzén and Tajvidi [17]. Still, such multivariate GP distributions have enjoyed much less popularity than their univariate

* Corresponding author.
E-mail addresses: hrootzen@chalmers.se (H. Rootzén), johan.segers@uclouvain.be (J. Segers), j.wadsworth@lancaster.ac.uk (J.L. Wadsworth).

https://doi.org/10.1016/j.jmva.2017.12.003
0047-259X/

counterparts. One reason may be that the multivariate versions are mathematically more involved. Their support is (a subset of) $\{\boldsymbol{x} : \boldsymbol{x} \not\leqslant \boldsymbol{0}\}$, the complement of the negative orthant, where $\boldsymbol{x} = \boldsymbol{y} - \boldsymbol{u}$ represents the excess vector, at least one coordinate of which is positive by definition. The unusual shape of the support introduces a nontrivial dependence structure uncommon to other families of multivariate distributions.

Our aim is to facilitate manipulation of multivariate GP distributions for the analysis of multivariate extremes. Rootzén et al. [16] revisited multivariate GP distributions with an eye towards modelling. To facilitate the incorporation of physical constraints in the construction of GP models, these distributions were connected to a number of point process representations. In Kiriliouk et al. [11], the representations were used for the construction and calibration of parametric models admitting explicit density formulas.

To complete the picture, we focus here on a number of analytic properties of multivariate GP distributions. Our view is that a GP distribution is derived from a max-stable distribution from which it inherits the marginal parameters and the dependence structure after a suitable transformation. This construction directly motivates a number of stochastic representations of GP random vectors. Moreover, it leads to compact expressions and direct proofs of some interesting properties of multivariate GP distributions.

After recalling some basic definitions and properties in Section 2, we introduce a number of parametrizations and stochastic representations in Sections 3 and 4, respectively. These results then provide the background against which we present compact formulas for probability densities (Section 5), marginal distributions (Section 6), and copula-related objects (Section 8). Finally, the family of multivariate GP distributions is stable with respect to conditional marginalization and thresholding (Section 7) and, provided the margins have equal shape parameters, to certain linear transformations (Section 9). All proofs are deferred to the Appendix.

Notation. Throughout, the expressions $(1 + \gamma x)^{1/\gamma}$, $\ln(1 + \gamma x)/\gamma$, and $(x^\gamma - 1)/\gamma$ are to be read as their limits $\exp(x)$, x, and $\ln(x)$, respectively, if $\gamma = 0$. When applied to vectors, mathematical operations such as addition, multiplication and exponentiation are to be interpreted component-wise, where scalars are recycled if necessary; for instance, for $\boldsymbol{\gamma}, \boldsymbol{x} \in \mathbb{R}^d$, we write $(1+\boldsymbol{\gamma x})^{1/\boldsymbol{\gamma}}$ for the vector $((1+\gamma_1 x_1)^{1/\gamma_1}, \ldots, (1+\gamma_d x_d)^{1/\gamma_d})$, with the earlier mentioned convention for $\gamma_j = 0$ applied to each component. We let $a \wedge b$ and $a \vee b$ denote $\min(a, b)$ and $\max(a, b)$, respectively, whereas for vectors, the minimum and the maximum are taken component-wise. Order relations between vectors are to be interpreted component-wise, too. We write $\mathcal{L}(\xi)$ for the law of the random variable or vector ξ and we let $\rightsquigarrow$ denote convergence in distribution. Bold face symbols denote vectors, usually of length d. Likewise, $\boldsymbol{0} = (0, \ldots, 0)$ and $\boldsymbol{1} = (1, \ldots, 1)$, and $\boldsymbol{\infty} = (\infty, \ldots, \infty)$. For a vector $\boldsymbol{x}$, we write $\max(\boldsymbol{x}) = \max(x_1, \ldots, x_d)$. The indicator variable of the set A is denoted by $\mathbf{1}(A)$.

2. Basics

Let $\boldsymbol{X}$ be a d-variate random vector with cumulative distribution function (cdf) F. Suppose that there exist sequences of vectors $\boldsymbol{a}_n \in (0, \infty)^d$ and $\boldsymbol{b}_n \in \mathbb{R}^d$ and a d-variate cdf G with non-degenerate margins such that, as $n \to \infty$,

$$F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \rightsquigarrow G(\boldsymbol{x}). \tag{1}$$

The weak limit G in (1) is a d-variate max-stable or generalized extreme-value (GEV) distribution. The margins, $G_1, \ldots, G_d$, of G are continuous, see (7), so that the convergence in (1) takes place for every $\boldsymbol{x} \in \mathbb{R}^d$. In particular, (1) implies that, for all $\boldsymbol{x} \in \mathbb{R}^d$ such that $G(\boldsymbol{x}) > 0$, we have

$$\lim_{n\to\infty} n\{1 - F(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n)\} = -\ln G(\boldsymbol{x}). \tag{2}$$

We refer to Beirlant et al. [2, Chap. 8] or de Haan and Ferreira [3, Chap. 6] for background on multivariate GEV distributions and their domains of attraction.

By an appropriate choice of the sequences $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$, we can always ensure that

$$\forall_{j\in\{1,\ldots,d\}} \quad 0 < G_j(0) < 1. \tag{3}$$

Multivariate GEV distributions being positive quadrant dependent [12], we then have $0 < G(\boldsymbol{0}) < 1$. By (2) and some elementary calculations, we find that, for all $\boldsymbol{x} \in \mathbb{R}^d$ such that $G_j(x_j) > 0$ for all $j \in \{1, \ldots, d\}$,

$$\lim_{n\to\infty} \Pr\{\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n) \leqslant \boldsymbol{x} \mid \boldsymbol{X} \not\leqslant \boldsymbol{b}_n\} = \frac{\ln G(\boldsymbol{x} \wedge \boldsymbol{0}) - \ln G(\boldsymbol{x})}{\ln G(\boldsymbol{0})}. \tag{4}$$

Let $\boldsymbol{\eta} \in [-\infty, 0)^d$ denote the vector of lower endpoints of the marginal distributions $G_1, \ldots, G_d$. From (4), it follows that, as $n \to \infty$,

$$\mathcal{L}\{\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n) \vee \boldsymbol{\eta} \mid \boldsymbol{X} \not\leqslant \boldsymbol{b}_n\} \rightsquigarrow H, \tag{5}$$

where H is the multivariate generalized Pareto (GP) distribution associated to G; this is denoted $H = \mathrm{GP}(G)$. The support of H is included in, but not necessarily equal to, the set $[\boldsymbol{\eta}, \boldsymbol{\infty}) \setminus [\boldsymbol{\eta}, \boldsymbol{0}]$, the set of all $\boldsymbol{x}$ such that $x_j \geqslant \eta_j$ for all $j \in \{1, \ldots, d\}$ and $x_j > 0$ for at least one j. The function H is determined by

$$\forall_{\boldsymbol{x}\in(\boldsymbol{\eta},\boldsymbol{\infty})} \quad H(\boldsymbol{x}) = \frac{\ln G(\boldsymbol{x} \wedge \boldsymbol{0}) - \ln G(\boldsymbol{x})}{\ln G(\boldsymbol{0})}. \tag{6}$$

If $x_j = \eta_j$ for some $j \in \{1, \ldots, d\}$, then the value of $H(\boldsymbol{x})$ is determined by continuity from the right. Note that H may assign positive mass to the lower boundaries $\{\boldsymbol{x} : x_j = \eta_j\}$, even if $\eta_j = -\infty$; see Proposition 17. Since (4) only covers those $\boldsymbol{x}$ such that $G_j(x_j) > 0$ (i.e., $x_j > \eta_j$) for all $j \in \{1, \ldots, d\}$, it does not provide information on what happens below $\boldsymbol{\eta}$. Without the truncation at $\boldsymbol{\eta}$, statement (5) is no longer true in general, as the following example shows.

Example 1. Consider $\boldsymbol{X} = (X, -X)$, where the law of X is symmetric around the origin and has tail function $\Pr(X < -x) = \Pr(X > x) \sim 1/x$ as $x \to \infty$. Then (2) holds with $\boldsymbol{a}_n = \boldsymbol{b}_n = (n, n)$ and $G(x_1, x_2) = \exp\{-(1 + x_1)^{-1} - (1 + x_2)^{-1}\}$ for $x_1 > -1$ and $x_2 > -1$, so $\boldsymbol{\eta} = (-1, -1)$. The corresponding GP distribution H in (5) and (6) is concentrated on the union of the two half-lines $(0, \infty) \times \{-1\}$ and $\{-1\} \times (0, \infty)$; see Example 1 in Rootzén et al. [16]. Instead, the limit distribution $\tilde{H}$ of $\boldsymbol{a}_n^{-1}(\boldsymbol{X} - \boldsymbol{b}_n) = (n^{-1}X - 1, -n^{-1}X - 1)$ given the event $\{\boldsymbol{X} \not\leq \boldsymbol{b}_n\} = \{|X| > n\}$ is concentrated on $\{(y - 1, -y - 1) : |y| > 1\}$. However, Eq. (2) does not provide any information about limits for points $\boldsymbol{x} = (x_1, x_2)$ such that $x_1 < -1$ or $x_2 < -1$, since $G(\boldsymbol{x}) = 0$ for such $\boldsymbol{x}$. Also, $\tilde{H}$ cannot be a GP distribution, since its margins would have to have heavy upper and lower tails, which is in contradiction with representation (18).

The margins of G are three-parameter generalized extreme-value distributions:

$$G_j(x_j) = \begin{cases} \exp[-\{1 + \gamma_j(x_j - \mu_j)/\alpha_j\}^{-1/\gamma_j}] & \text{if } \gamma_j \neq 0, \\ \exp[-\exp\{-(x_j - \mu_j)/\alpha_j\}] & \text{if } \gamma_j = 0, \end{cases} \tag{7}$$

for $j \in \{1, \ldots, d\}$ and $x_j \in \mathbb{R}$ such that $\alpha_j + \gamma_j(x_j - \mu_j) > 0$; the parameter range is $\gamma_j \in \mathbb{R}$, $\mu_j \in \mathbb{R}$, and $\alpha_j \in (0, \infty)$. The dependence structure (i.e., the copula) of G can be described in many ways. In this paper we opt for the description in terms of the stable tail dependence function (stdf) $\ell : [0, \infty)^d \to [0, \infty)$; see [5]. For $\boldsymbol{x} \in \mathbb{R}^d$ such that $G_j(x_j) > 0$ for all $j \in \{1, \ldots, d\}$, we have

$$G(\boldsymbol{x}) = \exp[-\ell\{-\ln G_1(x_1), \ldots, -\ln G_d(x_d)\}]. \tag{8}$$

The distribution G is thus determined by the parameter vectors $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, and $\boldsymbol{\alpha}$ together with the stdf ℓ; notation $G = \mathrm{GEV}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \ell)$.

For later use we mention the fact that ℓ necessarily satisfies the following properties [5,15]:

$$\left.\begin{array}{l} \bullet\ \ell \text{ is convex;} \\ \bullet\ \max(y_1, \ldots, y_d) \leqslant \ell(\boldsymbol{y}) \leqslant y_1 + \cdots + y_d \text{ for all } \boldsymbol{y} \in [0, \infty)^d; \\ \bullet\ \ell(c\boldsymbol{y}) = c\,\ell(\boldsymbol{y}) \text{ for all } (c, \boldsymbol{y}) \in [0, \infty) \times [0, \infty)^d. \end{array}\right\} \tag{9}$$

A useful fact is also that a function $\ell : [0, \infty)^d \to [0, \infty)$ is a stdf if and only if there exists a random vector $\boldsymbol{V}$ with values in $[0, \infty)^d$ and with $\mathrm{E}(V_j) = 1$ such that

$$\forall_{\boldsymbol{y} \in [0,\infty)^d} \quad \ell(\boldsymbol{y}) = \mathrm{E}\{\max(\boldsymbol{y}\boldsymbol{V})\}. \tag{10}$$

Formula (10) represents ℓ as (the restriction to $[0, \infty)^d$ of) a D-norm [8]. For a given $\boldsymbol{V}$, the function ℓ in (10) is a stdf [21, Lemma 3.1]. Given a stdf ℓ, a possible choice for $\boldsymbol{V}$ in (10) is $\boldsymbol{V} = d\boldsymbol{W}$, where $\boldsymbol{W}$ is a random vector on the unit simplex $\Delta_{d-1} = \{\boldsymbol{w} \in [0, 1]^d : w_1 + \cdots + w_d = 1\}$ whose law is proportional to the angular measure on Δ_{d-1} of the associated GEV distribution: indeed, we have $\ell(\boldsymbol{y}) = d \int_{\Delta_{d-1}} \max(\boldsymbol{y}\boldsymbol{w}) \Pr(\boldsymbol{W} \in \mathrm{d}\boldsymbol{w})$; see [3, Theorem 6.1.14] or [14]. The random vector $\boldsymbol{V}$ generating ℓ is not unique in distribution. Specific constructions will be considered in Section 4.

3. Parametrizations

If H is determined by G and if G is determined by $(\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \ell)$, then so is H. However, this is not a convenient way to parametrize H, because the parameter vectors $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ are not identifiable from H. If $G = \mathrm{GEV}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \ell)$, then $G^t \sim \mathrm{GEV}[\boldsymbol{\mu}(t), \boldsymbol{\gamma}, \boldsymbol{\alpha}(t), \ell]$ for all $t \in (0, \infty)$, where

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu} + \boldsymbol{\alpha}(t^{\boldsymbol{\gamma}} - 1)/\boldsymbol{\gamma}, \quad \boldsymbol{\alpha}(t) = t^{\boldsymbol{\gamma}}\boldsymbol{\alpha}. \tag{11}$$

Still, if $H = \mathrm{GP}(G)$, then also $H = \mathrm{GP}(G^t)$ for all $t \in (0, \infty)$. All GEV distributions G^t thus generate the same GP distribution H. The GP distribution describes the distribution of sample points given that they exceed a high threshold, but not the exceedance probability itself. This explains the loss of one parameter with respect to a full point process model, which has the same number of parameters as the GEV model. The phenomenon already occurs in the univariate case, where GP and GEV distributions have two and three parameters, respectively. Another way to understand the difference in the number of parameters is through the fact that the vector of component-wise maxima of a Poisson number of independent random vectors with common GP distribution H has a distribution function which in its upper tail is equal to the GEV distribution G^t, where t is the expectation of the Poisson random variable.

The lack of identifiability of some of the GEV parameters from the associated GP distribution is one reason why we look for other parametrizations for H. Another reason for doing so is that GP distributions enjoy a number of interesting properties and representations, and some of these are more clearly understood and expressed in other parametrizations.

Let $G = \mathrm{GEV}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \ell)$ and write $\boldsymbol{\sigma} = \boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{\mu}$. The requirement that $0 < G_j(0) < 1$ is equivalent to the requirement that $\sigma_j > 0$. To ensure (3), we will therefore assume that $\boldsymbol{\sigma} \in (\mathbf{0}, \infty) = (0, \infty)^d$. Recall $\boldsymbol{\mu}(t)$ and $\boldsymbol{\alpha}(t)$ in (11) and note that $\boldsymbol{\alpha}(t) - \boldsymbol{\gamma}\boldsymbol{\mu}(t) = \boldsymbol{\sigma}$ for all $t \in (0, \infty)$, i.e., $\boldsymbol{\sigma}$ is a common parameter for all GEV distributions G^t.

Proposition 2. *Let G be $\mathrm{GEV}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \ell)$ with $\boldsymbol{\sigma} = \boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{\mu} \in (\mathbf{0}, \infty)$. Let H be $\mathrm{GP}(G)$. For $\boldsymbol{x} \in \mathbb{R}$ such that $\sigma_j + \gamma_j x_j > 0$ for all $j \in \{1, \ldots, d\}$, we have*

$$H(\boldsymbol{x}) = \ell\{\boldsymbol{\pi}\,(1 + \boldsymbol{\gamma}(\boldsymbol{x} \wedge \mathbf{0})/\boldsymbol{\sigma})^{-1/\boldsymbol{\gamma}}\} - \ell\{\boldsymbol{\pi}\,(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})^{-1/\boldsymbol{\gamma}}\} \tag{12}$$

where $\pi_j = \tau_j/\ell(\boldsymbol{\tau}) \in (0, 1]$ and $\tau_j = -\ln G_j(0) = (1 - \gamma_j\mu_j/\alpha_j)^{-1/\gamma_j} \in (0, \infty)$ for all $j \in \{1, \ldots, d\}$, while $\ell(\boldsymbol{\tau}) = -\ln G(\mathbf{0})$.

Proposition 3. *In Proposition 2, each of $\boldsymbol{\gamma}, \boldsymbol{\sigma}, \boldsymbol{\pi}$, and ℓ are identifiable from H, and $\ell(\boldsymbol{\pi}) = 1$. More precisely, writing $\overline{H} = 1 - H$ and $\overline{H}_j = 1 - H_j$ for $j \in \{1, \ldots, d\}$, we have, for $\boldsymbol{x} \in [\mathbf{0}, \infty)$,*

$$\overline{H}_j(0) = \pi_j, \tag{13}$$

$$\overline{H}_j(x_j)/\overline{H}_j(0) = (1 + \gamma_j x_j/\sigma_j)^{-1/\gamma_j}, \quad \text{provided } \sigma_j + \gamma_j x_j > 0, \tag{14}$$

$$\overline{H}(\boldsymbol{x}) = \ell\{\overline{H}_1(x_1), \ldots, \overline{H}_d(x_d)\}. \tag{15}$$

Furthermore, we have $\boldsymbol{\tau} = \ell(\boldsymbol{\tau})\,\boldsymbol{\pi}$, so that the vector $\boldsymbol{\tau}$ is identifiable up to a constant multiple.

In view of Propositions 2 and 3, we express (12) as

$$H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell). \tag{16}$$

This yields a parametrization of H in terms of $\boldsymbol{\gamma} \in \mathbb{R}^d$, $\boldsymbol{\sigma} \in (0, \infty)^d$, $\boldsymbol{\pi} \in (0, 1]^d$, and a stdf ℓ such that $\ell(\boldsymbol{\pi}) = 1$. All four components of $(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ are identifiable from H. However, the nonlinear constraint $\ell(\boldsymbol{\pi}) = 1$ may perhaps be impractical when doing inference. Therefore, we also propose the alternative parametrization

$$H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \ell), \tag{17}$$

where $\boldsymbol{\pi}$ in (16) has been replaced by a vector $\boldsymbol{\tau} \in (0, \infty)^d$ which is identifiable only up to a positive multiplicative constant. This lack of identifiability can easily be remedied by adding a constraint such as $\tau_1 + \cdots + \tau_d = c$, where c is a positive constant, for instance $c = 1$ or $c = d$. The vector $\boldsymbol{\pi}$ can be reconstructed from $\boldsymbol{\tau}$ and ℓ via $\boldsymbol{\pi} = \boldsymbol{\tau}/\ell(\boldsymbol{\tau})$. Since ℓ is homogeneous, multiplying $\boldsymbol{\tau}$ by a positive constant does not affect $\boldsymbol{\pi}$. A valid choice for $\boldsymbol{\tau}$ would be $\boldsymbol{\pi}$ itself, which justifies the use of the same notation in (16) and (17).

4. Stochastic representations

In the parametrization $\boldsymbol{X} \sim \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$, the parameter vectors $\boldsymbol{\sigma} \in (\mathbf{0}, \infty)$ and $\boldsymbol{\gamma} \in \mathbb{R}^d$ represent marginal scale and shape vectors, respectively.

Proposition 4. *We have $\boldsymbol{X} \sim \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ if and only if*

$$\boldsymbol{X} = \boldsymbol{\sigma}\,(e^{\boldsymbol{\gamma}\boldsymbol{Z}} - 1)/\boldsymbol{\gamma}, \quad \text{with } \boldsymbol{Z} \sim \mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell). \tag{18}$$

The support of $\boldsymbol{Z}$ is contained in $[-\infty, \infty) \setminus [-\infty, \mathbf{0}]$ and its cdf is given, for all $\boldsymbol{z} \in \mathbb{R}^d$, by

$$\Pr(\boldsymbol{Z} \leqslant \boldsymbol{z}) = \ell(\boldsymbol{\pi}e^{-(\boldsymbol{z} \wedge 0)}) - \ell(\boldsymbol{\pi}e^{-\boldsymbol{z}}). \tag{19}$$

In view of Proposition 4, we can reduce the study of many aspects of general GP distributions to the special case of GP distributions with $\boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$.

Rootzén [16, Sections 4 and 5] introduced a number of stochastic representations of (standardized) GP random vectors. The representations were derived from that of a multivariate GEV distribution as the law of the vector of component-wise maxima of the points of certain point processes. Here, we derive these representations from scratch via that of the stdf ℓ in (10). We also connect the representations to the parametrization in terms of $\boldsymbol{\pi}$ and ℓ.

Definition 5. A random vector $\boldsymbol{S}$ taking values in $[-\infty, 0]^d$ is called a *spectral random vector* if the following two conditions hold:

(S1) $\Pr\{\max(S_1, \ldots, S_d) = 0\} = 1$;
(S2) $\Pr(S_j > -\infty) > 0$ for all $j \in \{1, \ldots, d\}$.

Theorem 6. *Let $\boldsymbol{S}$ be a spectral random vector and let E be a unit exponential random variable independent of $\boldsymbol{S}$. Then $\boldsymbol{S} + E \sim \mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell)$, where $\boldsymbol{\pi}$ and ℓ are given by*

$$\forall_{j\in\{1,\dots,d\}} \quad \pi_j = \mathrm{E}(e^{S_j}), \tag{20}$$

$$\forall_{\boldsymbol{y}\in[0,\infty)^d} \quad \ell(\boldsymbol{y}) = \mathrm{E}\{\max(\boldsymbol{y}e^{\boldsymbol{S}}/\boldsymbol{\pi})\}. \tag{21}$$

The associated cdf is given by $H(\boldsymbol{z}) = 1 - \mathrm{E}\{1 \wedge e^{\max(\boldsymbol{S}-\boldsymbol{z})}\}$.

Theorem 7. *For every pair $(\boldsymbol{\pi}, \ell)$ where $\boldsymbol{\pi} \in (0, 1]^d$ and where ℓ is a stdf with $\ell(\boldsymbol{\pi}) = 1$, there exists a spectral random vector $\boldsymbol{S}$, unique in distribution, such that $\boldsymbol{Z} \sim \mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell)$ can be represented in distribution as*

$$\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{S} + E, \tag{22}$$

with E a unit exponential random variable, independent of $\boldsymbol{S}$.

Remark 1 (*Pareto Processes*)**.** Setting $\boldsymbol{\gamma} = \boldsymbol{\sigma} = \mathbf{1}$ in (18), we obtain the random vector $\boldsymbol{W} = \boldsymbol{X} + 1 = e^{\boldsymbol{Z}}$, and the representation (22) states that $\max(\boldsymbol{W}) = e^E$ is a unit Pareto variable which is independent of $\boldsymbol{W}/\max(\boldsymbol{W}) = e^{\boldsymbol{S}}$. According to Dombry and Ribatet [4, Theorem 2 and Definition 4], these properties ensure that $\boldsymbol{W}$ is a simple Pareto process with respect to the homogeneous cost functional $\max(\cdot)$. In their Proposition 1, these authors showed that such processes arise as weak limits of rescaled processes given that the cost functional exceeds a large threshold. Ferreira and de Haan [9, Theorem 2.1] already studied simple Pareto processes when the cost functional is the supremum norm of a continuous, nonnegative function. For stationary regularly varying time series, Basrak and Segers [1, Theorem 3.1] had already established the independence property for the case that the cost functional is an arbitrary norm applied to the initial state of the time series.

In view of Theorems 6 and 7, there is a one-to-one relation between pairs $(\boldsymbol{\pi}, \ell)$ satisfying $\ell(\boldsymbol{\pi}) = 1$ on the one hand and distributions, ν, of spectral random vectors $\boldsymbol{S}$ on the other hand. We therefore write $H = \mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ for the GP distribution $H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ with $(\boldsymbol{\pi}, \ell)$ determined by $\nu = \mathcal{L}(\boldsymbol{S})$.

Combining (18) and (22), we find that a general GP random vector $\boldsymbol{X} \sim \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ can be represented as

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\sigma}\,\{e^{\boldsymbol{\gamma}(\boldsymbol{S}+E)} - 1\}/\boldsymbol{\gamma}, \tag{23}$$

where $\boldsymbol{S}$ is the spectral random vector associated to $(\boldsymbol{\pi}, \ell)$ and where E is a unit exponential random variable independent of $\boldsymbol{S}$. Representation (23) is convenient for model construction and Monte Carlo simulation provided we have a handle on the spectral random vector $\boldsymbol{S}$.

The requirement that $\max(\boldsymbol{S}) = 0$ almost surely in Definition 5 may perhaps look difficult to ensure, but actually, it is not. We describe two constructions for doing so.

Proposition 8. *Let $\boldsymbol{T}$ be a random vector taking values in $[-\infty, \infty)^d$ such that the following two conditions hold:*

(T1) $\Pr(T_j > -\infty) > 0$ *for all* $j \in \{1, \dots, d\}$;
(T2) $\Pr\{\max(\boldsymbol{T}) > -\infty\} = 1$.

Then $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$ is a spectral random vector (Definition 5) and the associated GP distribution, $\mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell) = \mathrm{GP}_S[\mathbf{0}, \mathbf{1}, \mathcal{L}(\boldsymbol{S})]$, is determined by

$$\forall_{j\in\{1,\dots,d\}} \quad \pi_j = \mathrm{E}\{e^{T_j - \max(\boldsymbol{T})}\}, \tag{24}$$

$$\forall_{\boldsymbol{y}\in[0,\infty)^d} \quad \ell(\boldsymbol{y}) = \mathrm{E}\left[\max_{j\in\{1,\dots,d\}}\left[y_j \frac{e^{T_j - \max(\boldsymbol{T})}}{\mathrm{E}\{e^{T_j - \max(\boldsymbol{T})}\}}\right]\right]. \tag{25}$$

The associated cdf is given by

$$H(\boldsymbol{z}) = 1 - \mathrm{E}\{1 \wedge e^{\max(\boldsymbol{T}-\boldsymbol{z}) - \max(\boldsymbol{T})}\}. \tag{26}$$

Proposition 9. *Let $\boldsymbol{U}$ be a random vector taking values in $[-\infty, \infty)^d$ such that the following condition holds:*

(U) $0 < \mathrm{E}(e^{U_j}) < \infty$ *for all* $j \in \{1, \dots, d\}$.

Let $\boldsymbol{S}$ be defined in distribution by

$$\Pr(\boldsymbol{S} \in \cdot\,) = \frac{\mathrm{E}[\mathbf{1}\{\boldsymbol{U} - \max(\boldsymbol{U}) \in \cdot\,\}\, e^{\max(\boldsymbol{U})}]}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}}. \tag{27}$$

Then $\boldsymbol{S}$ *is a spectral random vector as in Definition 5 and the associated GP distribution,* $\mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell) = \mathrm{GP}_S[\mathbf{0}, \mathbf{1}, \mathcal{L}(\boldsymbol{S})]$*, is determined by*

$$\forall_{j\in\{1,\dots,d\}} \quad \pi_j = \frac{\mathrm{E}(e^{U_j})}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}}, \tag{28}$$

$$\forall_{\boldsymbol{y}\in[0,\infty)^d} \quad \ell(\boldsymbol{y}) = \mathrm{E}\left[\max_{j\in\{1,\dots,d\}}\left\{y_j \frac{e^{U_j}}{\mathrm{E}(e^{U_j})}\right\}\right]. \tag{29}$$

The associated cdf is given by

$$H(\boldsymbol{z}) = 1 - \frac{\mathrm{E}\{e^{\max(\boldsymbol{U})} \vee e^{\max(\boldsymbol{U}-\boldsymbol{z})}\}}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}}. \tag{30}$$

If ν is the law of the random vector $\boldsymbol{T}$ in Proposition 8, we write

$$\mathrm{GP}_T(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu) = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell) \tag{31}$$

with $(\boldsymbol{\pi}, \ell)$ determined by $\boldsymbol{T}$ as in (24)–(25). Similarly, if ν is the law of the random vector $\boldsymbol{U}$ in Proposition 9, we write

$$\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu) = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell) \tag{32}$$

with $(\boldsymbol{\pi}, \ell)$ determined by $\boldsymbol{U}$ as in (28)–(29).

Practical modelling considerations led us in Rootzén et al. [16] to consider multivariate GP distribution functions of the form

$$H_R(\boldsymbol{x}) = \frac{\int_0^\infty [F_{\boldsymbol{R}}\{t^{\boldsymbol{\gamma}}(\boldsymbol{x}+\boldsymbol{\sigma}/\boldsymbol{\gamma})\} - F_{\boldsymbol{R}}[t^{\boldsymbol{\gamma}}\{(\boldsymbol{x}\wedge\mathbf{0})+\boldsymbol{\sigma}/\boldsymbol{\gamma}\}]]\,\mathrm{d}t}{\int_0^\infty \overline{F}_{\boldsymbol{R}}(t^{\boldsymbol{\gamma}}\boldsymbol{\sigma}/\boldsymbol{\gamma})\,\mathrm{d}t}, \tag{33}$$

a distribution denoted as $\mathrm{GP}_R[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{R})]$. The marginal parameter vectors are $\boldsymbol{\sigma}, \boldsymbol{\gamma} \in (\mathbf{0}, \boldsymbol{\infty})$, while $\boldsymbol{R}$ is a random vector on $[\mathbf{0}, \boldsymbol{\infty})$ such that $0 < \mathrm{E}(R_j^{1/\gamma_j}) < \infty$ for all $j \in \{1, \dots, d\}$. (The cases where $\gamma_j = 0$ or $\gamma_j < 0$ for some or all j were considered in Rootzén et al. [16] as well.) Further, $F_{\boldsymbol{R}}$ is the cdf of $\boldsymbol{R}$ and $\overline{F}_{\boldsymbol{R}} = 1 - F_{\boldsymbol{R}}$, while the argument, $\boldsymbol{x}$, in (33) is such that $\sigma_j + \gamma_j x_j > 0$ for all $j \in \{1, \dots, d\}$.

Proposition 10. *For* $\boldsymbol{\sigma}, \boldsymbol{\gamma} \in (\mathbf{0}, \boldsymbol{\infty})$*, the function* H_R *in (33) is the cdf of the* $\mathrm{GP}_U[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{U})]$ *distribution, where* $\boldsymbol{U} = \boldsymbol{\gamma}^{-1}\ln(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})$*.*

Remark 2 (*Identifying* ν)**.** By the uniqueness statement in Theorem 7, the law, ν, of $\boldsymbol{S}$ in the GP_S representation can be identified from the GP distribution. This is not the case for the GP_T or GP_U, representations, however: adding a common, independent random variable ξ with $\mathrm{E}(e^\xi) < \infty$ to all components T_j or U_j in Propositions 8 or 9 does not change the resulting spectral random vector $\boldsymbol{S}$ nor the GP parameter $(\boldsymbol{\pi}, \ell)$. Still, if the laws of $\boldsymbol{T}$ or $\boldsymbol{U}$ are known up to some finite-dimensional parameter, then this parameter may well be identifiable from the associated GP distribution. This is to be investigated on a case-by-case basis, for instance by inspection of the density functions (Section 5).

Remark 3 (*The Role of* ν)**.** If ν is already the law of a spectral random vector $\boldsymbol{S}$, then the transformations in Propositions 8 and 9 leave ν invariant, so that $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$, $\mathrm{GP}_T(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ and $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ are all the same. In that sense, the GP_S notation is redundant. Still, we keep using it because of the uniqueness of $\mathcal{L}(\boldsymbol{S})$ and because of its role as common starting point for the GP_T and GP_U constructions. If ν is not the distribution of a spectral random vector, however, then $\mathrm{GP}_S(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ is not defined, whereas $\mathrm{GP}_T(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ and $\mathrm{GP}_U(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \nu)$ are different in general.

Remark 4 (*Model Construction*)**.** The probability measure ν in (31) and (32) need only satisfy conditions (T1)–(T2) or (U) in Propositions 8 or 9, respectively. This makes the GP_T and GP_U representations attractive for model construction. Some examples include the multivariate normal distribution or distributions with independent components. More involved models arise when ν is the joint distribution of a vector of partial sums. Specific constructions and case studies are worked out in Kiriliouk et al. [11]. The possibilities are endless and constitute a potentially fruitful research avenue.

Remark 5 (*D-Norms*)**.** Formulas (21), (25) and (29) represent the stdf ℓ as a D-norm as in (10). In Falk et al. [8], D-norms are linked to multivariate GP distributions as well.

5. Densities

For statistical inference, it is highly useful to know the probability density functions of the GP distributions constructed via the methods in Section 4. Most of the results of this section can also be found in Rootzén et al. [16], but for completeness, we give self-contained proofs in the Appendix. Theorem 15 is new.

By Proposition 4, it is sufficient to consider the standardized case $\boldsymbol{\sigma} = \mathbf{1}$ and $\boldsymbol{\gamma} = \mathbf{0}$. The density in case of general $(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ can then be found by the component-wise increasing transformation $\boldsymbol{z} \mapsto \boldsymbol{x} = \boldsymbol{\sigma}(e^{\boldsymbol{\gamma}\boldsymbol{z}} - 1)/\boldsymbol{\gamma}$.

Lemma 11. *If $\boldsymbol{Z} \sim \mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell)$ is absolutely continuous with Lebesgue density $h_{\boldsymbol{Z}}$, then $\boldsymbol{X} \sim \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ in (18) is absolutely continuous with Lebesgue density $h_{\boldsymbol{X}}$ given by*

$$h_{\boldsymbol{X}}(\boldsymbol{x}) = h_{\boldsymbol{Z}}\{\boldsymbol{\gamma}^{-1} \ln(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})\} \prod_{j=1}^{d} \frac{1}{\sigma_j + \gamma_j x_j}, \tag{34}$$

for $\boldsymbol{x} \in \mathbb{R}^d$ such that $\boldsymbol{x} \not\leqslant \mathbf{0}$ and $\sigma_j + \boldsymbol{\gamma}_j x_j > 0$ for all $j \in \{1, \ldots, d\}$.

Next we give expressions for the density of $\boldsymbol{Z} = \boldsymbol{S} + E$ in Theorem 6 for the stochastic representations of $\boldsymbol{S}$ via $\boldsymbol{T}$ and $\boldsymbol{U}$ in Propositions 8 and 9, respectively.

Theorem 12. *If $\nu = \mathcal{L}(\boldsymbol{T})$, with $\boldsymbol{T}$ as in Proposition 8, has support included in $\mathbb{R}^d$ and is absolutely continuous with Lebesgue density $f_{\boldsymbol{T}}$, then the $\mathrm{GP}_T(\mathbf{1}, \mathbf{0}, \nu)$ distribution has Lebesgue density h given by*

$$h(\boldsymbol{z}) = \mathbf{1}(\boldsymbol{z} \not\leqslant \mathbf{0}) \frac{1}{e^{\max(\boldsymbol{z})}} \int_{-\infty}^{\infty} f_{\boldsymbol{T}}(\boldsymbol{z} + t)\, \mathrm{d}t. \tag{35}$$

Theorem 13. *If $\nu = \mathcal{L}(\boldsymbol{U})$, with $\boldsymbol{U}$ as in Proposition 9, has support included in $\mathbb{R}^d$ and is absolutely continuous with Lebesgue density $f_{\boldsymbol{U}}$, then the $\mathrm{GP}_U(\mathbf{1}, \mathbf{0}, \nu)$ distribution has Lebesgue density h given by*

$$h(\boldsymbol{z}) = \mathbf{1}(\boldsymbol{z} \not\leqslant \mathbf{0}) \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_{-\infty}^{\infty} f_{\boldsymbol{U}}(\boldsymbol{z} + t)\, e^t\, \mathrm{d}t. \tag{36}$$

Combining Propositions 9 and 10, we also find the density of H_R in (33) in terms of the one of $\boldsymbol{R}$.

Corollary 14. *Let $\boldsymbol{\sigma}, \boldsymbol{\gamma} \in (\mathbf{0}, \boldsymbol{\infty})$ and let $\boldsymbol{R}$ be a random vector with values in $(\mathbf{0}, \boldsymbol{\infty})$ such that $\mathrm{E}(R_j^{1/\gamma_j}) < \infty$ for all $j \in \{1, \ldots, d\}$. If $\boldsymbol{R}$ is absolutely continuous with Lebesgue density $f_{\boldsymbol{R}}$, then the GP cdf H_R in (33) is absolutely continuous with Lebesgue density*

$$h(\boldsymbol{x}) = \mathbf{1}(\boldsymbol{x} \not\leqslant \mathbf{0}) \frac{1}{\mathrm{E}[\max\{(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})^{1/\gamma}\}]} \int_0^{\infty} f_{\boldsymbol{R}}\{t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma})\}\, t^{\sum_{j=1}^d \gamma_j}\, \mathrm{d}t$$

for $\boldsymbol{x}$ such that $\sigma_j + \gamma_j x_j > 0$ for all $j \in \{1, \ldots, d\}$.

By definition, a spectral random vector $\boldsymbol{S}$ is supported on the Lebesgue null set $\{\boldsymbol{s} : \max(\boldsymbol{s}) = 0\}$. Still, it may be absolutely continuous with respect to the $(d-1)$-dimensional Lebesgue measure on that set with some density function $f_{\boldsymbol{S}}$, and then the associated GP_S distribution is absolutely continuous with respect to the d-dimensional Lebesgue measure.

Theorem 15. *Let $\boldsymbol{S}$ be a spectral random vector with Lebesgue density $f_{\boldsymbol{S}}$ defined on $\{\boldsymbol{s} \in \mathbb{R}^d : \max(\boldsymbol{s}) = 0\}$. Let $\boldsymbol{Z} = \boldsymbol{S} + E$ be the associated GP random vector. Then the density of $\boldsymbol{Z}$ is given by*

$$h(\boldsymbol{z}) = \mathbf{1}(\boldsymbol{z} \not\leqslant \mathbf{0}) f_{\boldsymbol{S}}\{\boldsymbol{z} - \max(\boldsymbol{z})\}\, e^{-\max(\boldsymbol{z})}.$$

6. Margins and lower boundaries

The margins of a multivariate GP distribution are in general not univariate GP distributions; indeed, their supports are not necessarily included in $[0, \infty)$. Still, by (14), their conditional versions are univariate GP.

Proposition 16. *For $j \in \{1, \ldots, d\}$, the jth marginal distribution function, H_j, of a GP distribution $H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell) = \mathrm{GP}_S[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{S})]$ with $(\sigma_j, \gamma_j) = (1, 0)$ is given as follows: for $z_j \in \mathbb{R}$,*

$$H_j(z_j) = \ell\big(\pi_1, \ldots, \pi_{j-1}, \pi_j e^{-(z_j \wedge 0)}, \pi_{j+1}, \ldots, \pi_d\big) - \pi_j e^{-z_j} \tag{37}$$

$$= \mathrm{E}\{\max(e^{S_1}, \ldots, e^{S_{j-1}}, e^{S_j - (z_j \wedge 0)}, e^{S_{j+1}}, \ldots, e^{S_d})\} - e^{-z_j}\ \mathrm{E}(e^{S_j}). \tag{38}$$

- *If $z_j \geqslant 0$, the right-hand sides simplify to $1 - \pi_j e^{-z_j} = 1 - e^{-z_j}\ \mathrm{E}(e^{S_j})$.*
- *For $H = \mathrm{GP}_T[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{T})]$, replace S_k by $T_k - \max(\boldsymbol{T})$ for all $k \in \{1, \ldots, d\}$.*
- *For $H = \mathrm{GP}_U[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{U})]$, replace S_k by U_k for all $k \in \{1, \ldots, d\}$ and divide everything by $\mathrm{E}\{e^{\max(\boldsymbol{U})}\}$.*
- *For general $(\sigma_j, \gamma_j) \in (0, \infty) \times \mathbb{R}$, replace z_j by $(1 + \gamma_j x_j/\sigma_j)^{-1/\gamma_j}$ for real x_j such that $\sigma_j + \gamma_j x_j > 0$.*

Recall that η_j is the lower endpoint of G_j, the jth margin of the GEV G generating H:

$$\eta_j = \begin{cases} -\sigma_j/\gamma_j & \text{if } \gamma_j > 0, \\ -\infty & \text{if } \gamma_j \leqslant 0. \end{cases} \tag{39}$$

Note that this lower endpoint is common for all GEV distributions G^t with $t \in (0, \infty)$. Letting z_j decrease to $-\infty$ in (37) yields the probability mass $H_j(\{\eta_j\})$ assigned by $H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ to the hyperplane $\{\boldsymbol{x} : x_j = \eta_j\}$.

Proposition 17. *Let $H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$, let $j \in \{1, \ldots, d\}$, and let η_j be as in (39). We have*

$$H_j(\{\eta_j\}) = \lim_{y_j \to \infty} \{\ell\big(\pi_1, \ldots, \pi_{j-1}, \pi_j y_j, \pi_{j+1}, \ldots, \pi_d\big) - \pi_j y_j\} \tag{40}$$

$$= \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \{\ell\big(\varepsilon\pi_1, \ldots, \varepsilon\pi_{j-1}, \pi_j, \varepsilon\pi_{j+1}, \ldots, \varepsilon\pi_d\big) - \pi_j\}. \tag{41}$$

If the first-order partial derivatives of ℓ exist and are continuous on a neighbourhood of $\pi_j \boldsymbol{e}_j = (0, \ldots, 0, \pi_j, 0, \ldots, 0)$, then also

$$H_j(\{\eta_j\}) = \sum_{k \in \{1,\ldots,d\} \setminus \{j\}} \pi_k \, \dot{\ell}_k(\pi_j \boldsymbol{e}_j). \tag{42}$$

Finally, in the GP_S, GP_T and GP_U representations, we have, respectively,

$$H_j(\{\eta_j\}) = \begin{cases} \Pr(S_j = -\infty), \\ \Pr(T_j = -\infty), \\ \mathrm{E}\{e^{\max(\boldsymbol{U})} \, \mathbf{1}(U_j = -\infty)\} \,/\, \mathrm{E}\{e^{\max(\boldsymbol{U})}\}. \end{cases} \tag{43}$$

Eq. (37) implies that the margins of H are continuous, except for a possible atom at the lower endpoints $\eta_1, \ldots, \eta_d$. Weak convergence of multivariate threshold exceedances in (5) implies that, for all $j \in \{1, \ldots, d\}$ and all $x_j > \eta_j$, we have

$$\lim_{n \to \infty} \Pr\{a_{n,j}^{-1}(X_{n,j} - b_{n,j}) \leqslant x_j \mid \boldsymbol{X} \not\leq \boldsymbol{b}_n\} = H_j(x_j).$$

Taking the limit as $x_j \downarrow \eta_j$, we obtain a statistical interpretation of the probability mass (if any) assigned by H to $\{\boldsymbol{x} : x_j = \eta_j\}$:

$$H_j(\{\eta_j\}) = \lim_{x_j \downarrow \eta_j} \lim_{n \to \infty} \Pr\{a_{n,j}^{-1}(X_{n,j} - b_{n,j}) \leqslant x_j \mid \boldsymbol{X} \not\leq \boldsymbol{b}_n\}.$$

In words, $H_j(\{\eta_j\})$ represents the probability that the rescaled (negative) excess in the jth component, conditionally on a (positive) excess in some of the other components, drops below a certain level to a region which is not covered by the max-stable model, G, for the upper tail of F in (1).

7. Stability

Lower-dimensional margins of GP distributions, conditionally on having at least one positive component, are GP distributed as well. In addition, conditional distributions of multivariate threshold excesses by GP random vectors are GP distributed, too [17]. These two properties are expressed together in the following result. For a vector $\boldsymbol{x} \in (0, \infty)^d$ and a non-empty subset J of $\{1, \ldots, d\}$, let $\boldsymbol{x}_J$ denote $(x_j)_{j \in J}$. Further, for a d-variate stdf ℓ, let ℓ_J denote the function $[0, \infty)^J \to [0, \infty)$ given by

$$\ell_J(\boldsymbol{y}) = \ell\big(\sum_{j \in J} y_j \boldsymbol{e}_j\big),$$

where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$ are the d canonical unit vectors in $\mathbb{R}^d$. The function ℓ_J is a $|J|$-variate stdf, too; in fact, if ℓ is the stdf of the GEV G, then ℓ_J is the stdf of the J-margin, G_J, of G.

Proposition 18. *Let $\boldsymbol{X} \sim \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$, let $\varnothing \neq J \subset \{1, \ldots, d\}$, and let $\boldsymbol{u} \in [0, \infty)^J$ be such that $\Pr(X_j > u_j) > 0$ for all $j \in J$. Then*

$$\mathcal{L}(\boldsymbol{X}_J - \boldsymbol{u} \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u}) = \mathrm{GP}[\boldsymbol{\sigma}_J + \boldsymbol{\gamma}_J \boldsymbol{u}, \ \boldsymbol{\gamma}_J, \ (\Pr[X_j > u_j \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u}])_{j \in J}, \ \ell_J]. \tag{44}$$

Here are some interesting special cases:

- The special case $J = \{j\}$ reproduces the result that the conditional distribution of X_j given that $X_j > 0$ is univariate GP with parameters (γ_j, σ_j), a fact we already knew from Proposition 3.
- If $u_j = 0$ for all $j \in J$, then we find that

$$\mathcal{L}\{\boldsymbol{X}_J \mid \max(\boldsymbol{X}_J) > 0\} = \mathrm{GP}[\boldsymbol{\sigma}_J, \ \boldsymbol{\gamma}_J, \ \boldsymbol{\pi}_J / \ell_J(\boldsymbol{\pi}_J), \ \ell_J]. \tag{45}$$

- The most compact formula arises in the GP_U representation: if $\boldsymbol{X} \sim \mathrm{GP}_U[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{U})]$, then, by the previous equation and Eqs. (28)–(29), we have

$$\mathcal{L}\{\boldsymbol{X}_J \mid \max(\boldsymbol{X}_J) > 0\} = \mathrm{GP}_U[\boldsymbol{\sigma}_J, \ \boldsymbol{\gamma}_J, \ \mathcal{L}(\boldsymbol{U}_J)].$$

To see this, calculate (28)–(29) with $\boldsymbol{U}$ replaced by $\boldsymbol{U}_J$ and check that the results are equal to $\boldsymbol{\pi}_J / \ell_J(\boldsymbol{\pi}_J)$ and ℓ_J, respectively, which is sufficient by (45).

8. Copula and tail dependence coefficients

Like any multivariate distribution function, a GP distribution H with margins $H_1, \ldots, H_d$ has a copula $C : [0, 1]^d \to [0, 1]$, i.e., a distribution function with standard uniform margins such that, for all $\boldsymbol{x} \in \mathbb{R}^d$,

$$H(\boldsymbol{x}) = C\{H_1(x_1), \ldots, H_d(x_d)\}.$$

Since H_j is continuous on (η_j, ∞), with $j \in \{1, \ldots, d\}$ and η_j as in (39), the copula C is unique on $\prod_{j=1}^d [H_j(\eta_j), 1]$ [13,22]. In addition, copulas remain invariant under increasing, component-wise transformations, so that by Proposition 4, the set of copulas associated to GP$(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ does not depend on $(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ but only on $(\boldsymbol{\pi}, \ell)$. Explicit computation of the marginal quantile functions from Proposition 16 seems unfeasible, however. Still, we have the following, partial result. The tail copula [6,20], $R : [0, \infty)^d \to [0, \infty)$, associated to a stdf ℓ is defined by

$$R(\boldsymbol{y}) = \Lambda([0, y_1] \times \cdots \times [0, y_d])$$

where Λ is the unique measure on $[0, \infty]^d \setminus \{\infty\}$ determined by ℓ via $\Lambda([\boldsymbol{0}, \infty] \setminus [\boldsymbol{y}, \infty]) = \ell(\boldsymbol{y})$.

The function R can be expressed directly in terms of the function ℓ by the inclusion–exclusion formula. For instance, for $d = 2$ we have $R(y_1, y_2) = y_1 + y_2 - \ell(y_1, y_2)$, whereas for general dimension d, we have

$$R(\boldsymbol{y}) = \sum_{J : \varnothing \neq J \subset \{1, \ldots, d\}} (-1)^{|J|-1} \, \ell\{(y_j \mathbf{1}_{\{j \in J\}})_{j=1}^d\}. \tag{46}$$

A more insightful formula is that if there exists a random vector $\boldsymbol{V}$ on $[0, \infty)^d$ with $\mathrm{E}(V_j) = 1$ for all $j \in \{1, \ldots, d\}$ such that (10) holds, then the tail copula, R, associated to ℓ is given, for all $\boldsymbol{y} \in [0, \infty)^d$, by

$$R(\boldsymbol{y}) = \mathrm{E}\{\min(\boldsymbol{y}\boldsymbol{V})\}. \tag{47}$$

Eq. (47) follows from (10) and (46) and the minimum–maximum identity. Identity (47) can be applied to either $V_j = e^{S_j}/\mathrm{E}(e^{S_j})$ in (21) or to $V_j = e^{U_j}/\mathrm{E}(e^{U_j})$ in (29).

Proposition 19. *If $\boldsymbol{X} \sim$ GP$(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$, then, for any $\boldsymbol{x} \in [0, \infty)^d$, we have*

$$\Pr(\exists_{j \in \{1, \ldots, d\}} \, X_j > x_j) = \ell\{\Pr(X_1 > x_1), \ldots, \Pr(X_d > x_d)\}, \tag{48}$$

$$\Pr(\forall_{j \in \{1, \ldots, d\}} \, X_j > x_j) = R\{\Pr(X_1 > x_1), \ldots, \Pr(X_d > x_d)\}, \tag{49}$$

where R is the tail copula associated to ℓ.

Eq. (49) says that, on the set $\prod_{j=1}^d [0, \overline{H}_j(0)]$, the survival copula of $H = \mathrm{GP}(\boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \ell)$ is given by the tail copula associated to ℓ.

The functions ℓ and R are homogeneous. As a consequence, for $p \geqslant 0$ such that $p \leqslant \overline{H}_j(0)$ for all $j \in \{1, \ldots, d\}$, we have

$$\Pr\{\exists_{j \in \{1, \ldots, d\}} \, \overline{H}_j(X_j) < p\} = p \, \ell(1, \ldots, 1), \quad \Pr\{\forall_{j \in \{1, \ldots, d\}} \, ; \overline{H}_j(X_j) < p\} = p \, R(1, \ldots, 1).$$

The quantity $\ell(1, \ldots, 1)$ is an example of an extremal coefficient [18], whereas the quantity $R(1, \ldots, 1)$ is an example of a multivariate tail dependence coefficient [19]. One way to exploit the above relations in model checking is to check whether the ratio of marginal and joint exceedance probabilities is indeed constant starting from a certain point on, either via diagnostic plots or via formal hypothesis tests; see for instance Kiriliouk et al. [11].

Remark 6 (*Other Tail Dependence Functions*)**.** The above formulas involving ℓ and R could be generalized to more general exceedance events involving exceedances in some components and non-exceedances in some other components, perhaps using additional conditioning [10].

Remark 7 (*Nonparametric Inference*)**.** By homogeneity, the stdf ℓ is determined by its values on $[0, \varepsilon]^d$ for arbitrarily small, positive $\varepsilon > 0$. Relation (48) could then serve as a basis for nonparametric inference on ℓ, for instance via the empirical copula or the empirical stable tail dependence function [7].

9. Linear combinations

We investigate the joint conditional distribution of linear combinations of the components of a GP random vector with nonnegative coefficients in case all the shape parameters γ_j are identical. To express the formulas compactly, we use matrix notation. The random vector $\boldsymbol{X}$ and the scale parameter vector $\boldsymbol{\sigma}$ are seen as $d \times 1$ column vectors. The ith row of the $m \times d$ matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}_i$. Expressions such as $\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{A}_i\boldsymbol{X}$ are to be interpreted via matrix products.

If $\gamma_j \leqslant 0$, we need to take into account the possibility of masses as $-\infty$. Therefore, we apply the convention that in expressions like $\sum_{j=1}^d a_j X_j$, if $a_j = 0$ and $X_j = -\infty$, then $0 \times (-\infty)$ is to be interpreted as 0, i.e., the jth component of $\boldsymbol{X}$ was not 'selected' in the first place: $\sum_{j=1}^d a_j X_j = \sum\{a_j X_j : a_j \neq 0\}$.

Proposition 20. *Let $\boldsymbol{X} \sim \mathrm{GP}_S[\boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathcal{L}(\boldsymbol{S})]$ be such that $\gamma_1 = \cdots = \gamma_d \equiv \gamma$. Further, let $\boldsymbol{A} = (a_{i,j})_{i,j} \in [0,\infty)^{m\times d}$ be such that $\Pr(\boldsymbol{A}_i\boldsymbol{X} > 0) > 0$ for all $i \in \{1,\dots,m\}$. For $\boldsymbol{x} \in \mathbb{R}^m$ such that $\boldsymbol{A}_i\boldsymbol{\sigma} + \gamma x_i > 0$ for all $i \in \{1,\dots,m\}$, we have*

$$\Pr[\boldsymbol{A}\boldsymbol{X} \not\leqslant \boldsymbol{x}] = \mathrm{E}\left[1 \wedge \max_{i\in\{1,\dots,m\}} \{(1+\gamma x_i/\boldsymbol{A}_i\boldsymbol{\sigma})^{-1/\gamma}\, e^{U_i}\}\right] \tag{50}$$

where $\boldsymbol{U} = (U_1,\dots,U_m)$ is given by

$$U_i = \begin{cases} \gamma^{-1}\ln\Big(\sum_{j=1}^{d} p_{i,j}\, e^{\gamma S_j}\Big) & \text{if } \gamma \neq 0, \\ \sum_{j=1}^{d} p_{i,j}\, S_j & \text{if } \gamma = 0, \end{cases} \tag{51}$$

where $p_{i,j} = a_{i,j}\sigma_j/\boldsymbol{A}_i\boldsymbol{\sigma}$. As a consequence, $\mathcal{L}(\boldsymbol{A}\boldsymbol{X} \mid \boldsymbol{A}\boldsymbol{X} \not\leqslant \boldsymbol{0}) = \mathrm{GP}_U[\boldsymbol{A}\boldsymbol{\sigma},\ \boldsymbol{\gamma},\ \mathcal{L}(\boldsymbol{U})]$.

If $m = 1$ and $\boldsymbol{A} = \boldsymbol{a} \in [0,\infty)^{1\times d}$, then Proposition 20 says that the law of $\boldsymbol{a}\boldsymbol{X}$ conditionally on $\boldsymbol{a}\boldsymbol{X} > 0$ is univariate GP with parameters $(\gamma, \boldsymbol{a}\boldsymbol{\sigma})$. Note that the gist of Proposition 20 does not depend on the way in which the GP is parametrized: the only condition is that all marginal shape parameters be the same.

Acknowledgements

We thank the Associate Editor and the two referees for careful reading and thoughtful comments, which helped us to improve the structure of the paper and to clarify certain points. H. Rootzén's research was supported by the Knut and Alice Wallenberg Foundation. J. Segers gratefully acknowledges funding by contract "Projet d'Actions de Recherche Concertées" No. 12/17–045 of the "Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy).

Appendix. Proofs

Proof of Proposition 2. For each $j \in \{1,\dots,d\}$, the lower endpoint, η_j, of G_j is given by $\eta_j = -\sigma_j/\gamma_j$ if $\gamma_j > 0$ and $\eta_j = -\infty$ if $\gamma_j \leqslant 0$. It follows that $x_j > \eta_j$ if and only if $\sigma_j + \gamma_j x_j > 0$. Furthermore, by (7), we have $-\ln G_j(0) = \tau_j \in (0,\infty)$, so that indeed $0 < G_j(0) < 1$, as required in the definition of $H = \mathrm{GP}(G)$.

By (6) and (8), we find, for such $\boldsymbol{x}$,

$$H(\boldsymbol{x}) = \frac{\ell\{-\ln G_1(x_1\wedge 0),\dots,-\ln G_d(x_d\wedge 0)\} - \ell\{-\ln G_1(x_1),\dots,-\ln G_d(x_d)\}}{\ell\{-\ln G_1(0),\dots,-\ln G_d(0)\}}.$$

Furthermore, by (7), we have

$$-\ln G_j(x_j) = \{1+\gamma_j(x_j-\mu_j)/\alpha_j\}^{-1/\gamma_j} = (1-\gamma_j\mu_j/\alpha_j)^{-1/\gamma_j}\,(1+\gamma_j x_j/\sigma_j)^{-1/\gamma_j}.$$

Combine these two equations to arrive at

$$H(\boldsymbol{x}) = \frac{\ell\{\boldsymbol{\tau}(1+\boldsymbol{\gamma}(\boldsymbol{x}\wedge\boldsymbol{0})/\boldsymbol{\sigma})^{-1/\boldsymbol{\gamma}}\} - \ell\{\boldsymbol{\tau}(1+\boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})^{-1/\boldsymbol{\gamma}}\}}{\ell(\boldsymbol{\tau})}.$$

Finally, use homogeneity of ℓ in (9) to arrive at (12). □

Proof of Proposition 3. Since $\boldsymbol{\pi} = \boldsymbol{\tau}/\ell(\boldsymbol{\tau})$, we have $\ell(\boldsymbol{\pi}) = 1$ by homogeneity of ℓ in (9). Let $\omega_j \in (0,\infty]$ denote the upper endpoint of G_j in Proposition 2; we have $\omega_j = \infty$ if $\gamma_j \geqslant 0$ and $\omega_j = \sigma_j/|\gamma_j|$ if $\gamma_j < 0$. In (12), fix $j \in \{1,\dots,d\}$ and $x_j \in [0,\omega_j)$ and let $x_k \to \omega_k$ for $k \in \{1,\dots,d\}\setminus\{j\}$. Then $(1+\gamma_k x_k/\sigma_k)^{-1/\gamma_k} \to 0$, so that, by $\ell(\boldsymbol{\pi}) = 1$ and (9), we find

$$H_j(x_j) = 1 - \pi_j(1+\gamma_j x_j/\sigma_j)^{-1/\gamma_j}, \quad x_j \in [0,\omega_j). \tag{A.1}$$

This yields both (13) and (14). Combine (12), (A.1) and $\ell(\boldsymbol{\pi}) = 1$ to arrive at (15). □

Proof of Proposition 4. Writing $\boldsymbol{Z} = \boldsymbol{\gamma}^{-1}\ln(1+\boldsymbol{\gamma}\boldsymbol{X}/\boldsymbol{\sigma})$, we have $\Pr(\boldsymbol{Z} \leqslant \boldsymbol{z}) = \Pr\{\boldsymbol{X} \leqslant \boldsymbol{\sigma}(e^{\boldsymbol{\gamma}\boldsymbol{z}}-1)/\boldsymbol{\gamma}\}$, which, by (12), yields (19). By (12) applied to $(\boldsymbol{\gamma},\boldsymbol{\sigma}) = (\boldsymbol{0},\boldsymbol{1})$, we find that the right-hand side of (19) is indeed the expression for the cdf of the $\mathrm{GP}(\boldsymbol{1},\boldsymbol{0},\boldsymbol{\pi},\ell)$ distribution. □

Proof of Theorem 6. By (10), the function ℓ in (21) is indeed a stdf. Since $\max(\boldsymbol{S}) = 0$ almost surely, we also have $\ell(\boldsymbol{\pi}) = \mathrm{E}\{e^{\max(\boldsymbol{S})}\} = 1$. Clearly, $\max(\boldsymbol{S}+E) = E > 0$ almost surely. For $\boldsymbol{z} \in \mathbb{R}^d$ such that $\max(\boldsymbol{z}) > 0$, we have, since $e^{\max(\boldsymbol{S})} = 1$ almost surely,

$$\begin{aligned}\Pr(\boldsymbol{S}+E \leqslant \boldsymbol{z}) &= \Pr\{E \leqslant \min(\boldsymbol{z}-\boldsymbol{S})\} = 1 - \mathrm{E}[\min\{1, e^{\max(\boldsymbol{S}-\boldsymbol{z})}\}] = \mathrm{E}[\max\{0, 1-e^{\max(\boldsymbol{S}-\boldsymbol{z})}\}] \\ &= \mathrm{E}[\max\{0, e^{\max(\boldsymbol{S})} - e^{\max(\boldsymbol{S}-\boldsymbol{z})}\}] = \mathrm{E}[e^{\max\{\boldsymbol{S}-(\boldsymbol{z}\wedge\boldsymbol{0})\}} - e^{\max(\boldsymbol{S}-\boldsymbol{z})}].\end{aligned}$$

The last step can be seen through a case-by-case analysis. Now insert the expressions for $\boldsymbol{\pi}$ and ℓ in (20)–(21) into the right-hand side of (19) to see that $\Pr(\boldsymbol{S} + E \leqslant \boldsymbol{z}) = H(\boldsymbol{z})$ with $H = \mathrm{GP}(\mathbf{1}, \mathbf{0}, \boldsymbol{\pi}, \ell)$. □

Proof of Theorem 7. The uniqueness of the distribution of $\boldsymbol{S}$ follows from the fact that $\boldsymbol{S}$ can be recovered from $\boldsymbol{S} + E$ via $\boldsymbol{S} = \boldsymbol{S} + E - \max(\boldsymbol{S} + E)$, since $\max(\boldsymbol{S}) = 0$ by assumption. We show the existence of $\boldsymbol{S}$ with the required properties.

Let $\boldsymbol{V}$ be a random vector with values in $[0, \infty)^d$ such that $\mathrm{E}(V_j) = 1$ for all $j \in \{1, \ldots, d\}$ and such that (10) holds. Let $\boldsymbol{W} = \boldsymbol{\pi}\boldsymbol{V}$ and define $\boldsymbol{S}$ (or rather its distribution) through

$$\Pr(\boldsymbol{S} \in \cdot\,) = \frac{\mathrm{E}[\mathbf{1}\{\ln(\boldsymbol{W}/\max(\boldsymbol{W})) \in \cdot\,\}\, \max(\boldsymbol{W})]}{\mathrm{E}\{\max(\boldsymbol{W})\}}. \tag{A.2}$$

Here we put $\ln(0) = -\infty$. Although the probability of the event $\{\max(\boldsymbol{W}) = 0\}$ could be positive under the original distribution of $\boldsymbol{W}$, the probability is zero under the transformed probability measure $\max(\boldsymbol{w})\,[\mathrm{E}\{\max(\boldsymbol{W})\}]^{-1}\,\Pr(\boldsymbol{W} \in \mathrm{d}\boldsymbol{w})$. Eq. (A.2) can be written in terms of expectations of measurable functions g as

$$\mathrm{E}\{g(\boldsymbol{S})\} = \frac{\mathrm{E}[g(\ln\{\boldsymbol{W}/\max(\boldsymbol{W})\})\, \max(\boldsymbol{W})]}{\mathrm{E}\{\max(\boldsymbol{W})\}}. \tag{A.3}$$

The random vector $\boldsymbol{S}$ is a spectral random vector: set $g(\boldsymbol{s}) = \mathbf{1}\{\max(\boldsymbol{s}) = 0\}$ and $g(\boldsymbol{s}) = e^{s_j}$, respectively, in (A.3) to obtain

$$\Pr\{\max(\boldsymbol{S}) = 0\} = \frac{\mathrm{E}[\mathbf{1}[\max[\ln\{\boldsymbol{W}/\max(\boldsymbol{W})\}] = 0]\, \max(\boldsymbol{W})]}{\mathrm{E}\{\max(\boldsymbol{W})\}} = 1,$$
$$\mathrm{E}(e^{S_j}) = \frac{\mathrm{E}(W_j)}{\mathrm{E}\{\max(\boldsymbol{W})\}} = \frac{\pi_j}{\ell(\boldsymbol{\pi})} = \pi_j,$$

so that both $\Pr(S_j > -\infty) > 0$ and (20) hold. Eq. (21) follows from setting $g(\boldsymbol{s}) = \max\{(\boldsymbol{y}/\boldsymbol{\pi})e^{\boldsymbol{s}}\}$ in (A.3). □

Proof of Proposition 8. The statement that $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$ is a spectral random vector is trivial; note that the property that $\max(\boldsymbol{T}) > -\infty$ almost surely guarantees that $S_j = T_j - \max(\boldsymbol{T})$ is well-defined, even if $T_j = -\infty$ can occur with positive probability. To arrive at (24)–(25), just substitute $S_j = T_j - \max(\boldsymbol{T})$ into (20)–(21).

To obtain (26), just substitute $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$ into the expression for H in Theorem 6. □

Proof of Proposition 9. Eq. (27) implies that, for measurable functions g, we have

$$\mathrm{E}\{g(\boldsymbol{S})\} = \frac{\mathrm{E}[g\{\boldsymbol{U} - \max(\boldsymbol{U})\}\, e^{\max(\boldsymbol{U})}]}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}}, \tag{A.4}$$

in the sense that the expectation on the left-hand side is defined if and only if the one on the right-hand side is defined, in which case both sides of (A.4) are equal. The random vector $\boldsymbol{S}$ is indeed a spectral random vector:

$$\Pr\{\max(\boldsymbol{S}) = 0\} = \mathrm{E}[\mathbf{1}\{\max(\boldsymbol{U}) - \max(\boldsymbol{U}) = 0\}\, e^{\max(\boldsymbol{U})}]/\mathrm{E}\{e^{\max(\boldsymbol{U})}\} = 1,$$
$$\mathrm{E}(e^{S_j}) = \mathrm{E}\{e^{U_j - \max(\boldsymbol{U})}\, e^{\max(\boldsymbol{U})}\}/\mathrm{E}\{e^{\max(\boldsymbol{U})}\} = \mathrm{E}(e^{U_j})/\mathrm{E}\{e^{\max(\boldsymbol{U})}\} > 0.$$

Eqs. (28)–(29) follow from combining (20)–(21) and (A.4) with $g(\boldsymbol{S}) = e^{S_j}$ and $g(\boldsymbol{S}) = \max\{(\boldsymbol{y}/\boldsymbol{\pi})\, e^{\boldsymbol{S}}\}$, respectively.

To obtain (30), combine (19) with (28)–(29) and simplify, using the identities $\max\{\boldsymbol{U} - (\boldsymbol{z} \wedge \mathbf{0})\} = \max(\boldsymbol{U} - \boldsymbol{z}) \vee \max(\boldsymbol{U})$ and $a \vee b - a = b - b \wedge a$. □

Proof of Proposition 10. For $t \in (0, \infty)$ and for $\boldsymbol{x}$ such that $x_j + \sigma_j/\gamma_j > 0$, we have

$$\overline{F}_{\boldsymbol{R}}\{t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma})\} = \Pr\{\boldsymbol{R} \not\leqslant t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma})\} = \Pr\left\{\max_{j \in \{1,\ldots,d\}} \left(\frac{R_j}{x_j + \sigma_j/\gamma_j}\right)^{1/\gamma_j} > t\right\}$$
$$= \Pr\left[\max_{j \in \{1,\ldots,d\}} \{(1 + \gamma_j x_j/\sigma_j)^{-1/\gamma_j} e^{U_j}\} > t\right],$$

with $U_j = \gamma_j^{-1} \ln(\gamma_j R_j/\sigma_j)$ for all $j \in \{1, \ldots, d\}$. It follows that

$$\int_{t=0}^{\infty} \overline{F}_{\boldsymbol{R}}\{t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma})\}\, \mathrm{d}t = \mathrm{E}\left[\max_{j \in \{1,\ldots,d\}} \{(1 + \gamma_j x_j/\sigma_j)^{-1/\gamma_j} e^{U_j}\}\right].$$

Apply this identity three times to the right-hand side of (33) and compare the resulting expression with the right-hand side in (30) to see that H_R is indeed the cdf of the stated GP distribution. □

Proof of Lemma 11. Let $\boldsymbol{x} \in \mathbb{R}^d$ be such that $\boldsymbol{x} \not\leqslant \mathbf{0}$ and $\sigma_j + \gamma_j x_j > 0$ for all $j \in \{1, \ldots, d\}$. Let $\boldsymbol{z} = \boldsymbol{\gamma}^{-1} \ln(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})$. Then

$$h_{\boldsymbol{X}}(\boldsymbol{x}) = h_{\boldsymbol{Z}}(\boldsymbol{z}) \prod_{j=1}^{d} \frac{\mathrm{d}z_j}{\mathrm{d}x_j} = h_{\boldsymbol{Z}}\{\boldsymbol{\gamma}^{-1} \ln(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})\} \prod_{j=1}^{d} \frac{1}{\sigma_j + \gamma_j x_j}.$$ □

Proof of Theorem 12. Let $\boldsymbol{S} = \boldsymbol{T} - \max(\boldsymbol{T})$ and let E be a unit exponential random variable, independent of $\boldsymbol{T}$. By definition, H is the cdf of $\boldsymbol{Z} = \boldsymbol{S} + E = \boldsymbol{T} - \max(\boldsymbol{T}) + E$, so that

$$H(\boldsymbol{z}) = \Pr\{\boldsymbol{T} - \max(\boldsymbol{T}) + E \leqslant \boldsymbol{z}\} = \int_0^\infty \Pr\{\boldsymbol{T} - \max(\boldsymbol{T}) + y \leqslant \boldsymbol{z}\}\, e^{-y}\, \mathrm{d}y$$
$$= \int_{\mathbb{R}^d} \int_0^\infty \mathbf{1}\{\boldsymbol{t} - \max(\boldsymbol{t}) + y \leqslant \boldsymbol{z}\} f_{\boldsymbol{T}}(\boldsymbol{t})\, e^{-y}\, \mathrm{d}y\, \mathrm{d}\boldsymbol{t}.$$

In the inner integral, perform the substitution $\max(\boldsymbol{t}) - y = r$ to see that

$$H(\boldsymbol{z}) = \int_{\mathbb{R}^d} \int_{-\infty}^{\max(\boldsymbol{t})} \mathbf{1}\{\boldsymbol{t} - r \leqslant \boldsymbol{z}\} f_{\boldsymbol{T}}(\boldsymbol{t})\, e^{r - \max(\boldsymbol{t})}\, \mathrm{d}r\, \mathrm{d}\boldsymbol{t}.$$

Next, apply Fubini's theorem and the substitutions $t_j - r = u_j$ for $j \in \{1, \ldots, d\}$ to see that

$$H(\boldsymbol{z}) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} \mathbf{1}\{r \leqslant \max(\boldsymbol{t}),\ \boldsymbol{t} - r \leqslant \boldsymbol{z}\} f_{\boldsymbol{T}}(\boldsymbol{t})\, e^{r - \max(\boldsymbol{t})}\, \mathrm{d}\boldsymbol{t}\, \mathrm{d}r$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}^d} \mathbf{1}\{\max(\boldsymbol{u}) \geqslant \mathbf{0},\ \boldsymbol{u} \leqslant \boldsymbol{z}\} f_{\boldsymbol{T}}(\boldsymbol{u} + r)\, e^{-\max(\boldsymbol{u})}\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}r$$
$$= \int_{\boldsymbol{u} \in (-\infty, \boldsymbol{z}]} \mathbf{1}\{\max(\boldsymbol{u}) \geqslant 0\}\, e^{-\max(\boldsymbol{u})} \int_{r \in \mathbb{R}} f_{\boldsymbol{T}}(\boldsymbol{u} + r)\, \mathrm{d}r\, \mathrm{d}\boldsymbol{u}.$$

We obtain that $H(\boldsymbol{z}) = \int_{(-\infty, \boldsymbol{z}]} h(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}$ with h given by (35). □

Proof of Theorem 13. By definition, the function H is the cdf of the random vector $\boldsymbol{S} + E$, with E a unit exponential random variable, independent of the random vector $\boldsymbol{S}$, the distribution of which is determined by the one of $\boldsymbol{U}$ through Eq. (27). By repeated applications of Fubini's theorem and by appropriate changes of variables, we find

$$H(\boldsymbol{z}) = \Pr\{\boldsymbol{S} + E \leqslant \boldsymbol{z}\} = \int_0^\infty \Pr\{\boldsymbol{S} + y \leqslant \boldsymbol{z}\}\, e^{-y}\, \mathrm{d}y = \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_0^\infty \mathrm{E}[\mathbf{1}\{\boldsymbol{U} - \max(\boldsymbol{U}) + y \leqslant \boldsymbol{z}\}\, e^{\max(\boldsymbol{U})}]\, e^{-y}\, \mathrm{d}y$$
$$= \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_0^\infty \int_{\mathbb{R}^d} \mathbf{1}\{\boldsymbol{u} - \max(\boldsymbol{u}) + y \leqslant \boldsymbol{z}\}\, e^{\max(\boldsymbol{u}) - y} f_{\boldsymbol{U}}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}y$$
$$= \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbf{1}\{\boldsymbol{u} - s \leqslant \boldsymbol{z},\ s < \max(\boldsymbol{u})\}\, e^s f_{\boldsymbol{U}}(\boldsymbol{u})\, \mathrm{d}s\, \mathrm{d}\boldsymbol{u}$$
$$= \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \mathbf{1}\{\boldsymbol{v} \leqslant \boldsymbol{z},\ \max(\boldsymbol{v}) > 0\}\, e^s f_{\boldsymbol{U}}(\boldsymbol{v} + s)\, \mathrm{d}s\, \mathrm{d}\boldsymbol{v}$$
$$= \frac{1}{\mathrm{E}\{e^{\max(\boldsymbol{U})}\}} \int_{\boldsymbol{v} \in (-\infty, \boldsymbol{z}]} \mathbf{1}\{\boldsymbol{v} \not\leqslant 0\} \int_{\mathbb{R}} e^s f_{\boldsymbol{U}}(\boldsymbol{v} + s)\, \mathrm{d}s\, \mathrm{d}\boldsymbol{v}.$$

We find that $H(\boldsymbol{z}) = \int_{(-\infty, \boldsymbol{z}]} h(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$ with h given by (36). □

Proof of Corollary 14. Let $\boldsymbol{U} = \boldsymbol{\gamma}^{-1} \ln(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})$. Let $\boldsymbol{u} \in \mathbb{R}^d$ and let $\boldsymbol{r} = (\boldsymbol{\sigma}/\boldsymbol{\gamma}) e^{\boldsymbol{\gamma}\boldsymbol{u}}$. The density function, $f_{\boldsymbol{U}}$, of $\boldsymbol{U}$ is given by

$$f_{\boldsymbol{U}}(\boldsymbol{u}) = f_{\boldsymbol{R}}(\boldsymbol{r}) \prod_{j=1}^d \frac{\mathrm{d}r_j}{\mathrm{d}u_j} = f_{\boldsymbol{R}}\{(\boldsymbol{\sigma}/\boldsymbol{\gamma}) e^{\boldsymbol{\gamma}\boldsymbol{u}}\} \prod_{j=1}^d \sigma_j e^{\gamma_j u_j}.$$

By Theorem 13, the density function, $h_{\boldsymbol{Z}}$, of $\boldsymbol{Z} \sim \mathrm{GP}_U[\mathbf{0}, \mathbf{1}, \mathcal{L}(\boldsymbol{U})]$ is then given by

$$h_{\boldsymbol{Z}}(\boldsymbol{z}) = \mathbf{1}(\boldsymbol{z} \not\leqslant \mathbf{0}) \frac{1}{\mathrm{E}[\max\{(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})^{1/\gamma}\}]} \int_0^\infty f_{\boldsymbol{R}}\{(\boldsymbol{\sigma}/\boldsymbol{\gamma})\, e^{\boldsymbol{\gamma}(\boldsymbol{z} + \ln t)}\} \prod_{j=1}^d \sigma_j\, e^{\gamma_j (z_j + \ln t)}\, \mathrm{d}t$$
$$= \mathbf{1}(\boldsymbol{z} \not\leqslant \mathbf{0}) \frac{\prod_{j=1}^d \sigma_j\, e^{\gamma_j z_j}}{\mathrm{E}[\max\{(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})^{1/\gamma}\}]} \int_0^\infty f_{\boldsymbol{R}}\{(\boldsymbol{\sigma}/\boldsymbol{\gamma})\, (t\, e^{\boldsymbol{z}})^{\boldsymbol{\gamma}}\}\, t^{\sum_{j=1}^d \gamma_j}\, \mathrm{d}t.$$

Let $\boldsymbol{x}$ be such that $\boldsymbol{x} \not\leqslant \mathbf{0}$ and $\sigma_j + \gamma_j x_j > 0$ for all $j \in \{1, \ldots, d\}$. Let $\boldsymbol{z} = \boldsymbol{\gamma}^{-1} \ln(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})$. Then $\sigma_j\, e^{\gamma_j z_j} = \sigma_j + \gamma_j x_j$ and $(\sigma_j/\gamma_j)(t\, e^{z_j})^{\gamma_j} = (\sigma_j + \gamma_j x)\, t^{\gamma_j}/\gamma_j$. By Eq. (34), we obtain

$$h(\boldsymbol{x}) = h_{\boldsymbol{Z}}\{\boldsymbol{\gamma}^{-1} \ln(1 + \boldsymbol{\gamma}\boldsymbol{x}/\boldsymbol{\sigma})\} \prod_{j=1}^d \frac{1}{\sigma_j + \gamma_j x_j} = \frac{1}{\mathrm{E}[\max\{(\boldsymbol{\gamma}\boldsymbol{R}/\boldsymbol{\sigma})^{1/\gamma}\}]} \int_0^\infty f_{\boldsymbol{R}}\{t^{\boldsymbol{\gamma}}(\boldsymbol{x} + \boldsymbol{\sigma}/\boldsymbol{\gamma})\}\, t^{\sum_{j=1}^d \gamma_j}\, \mathrm{d}t.$$ □

Proof of Theorem 15. The cdf, H, of $\boldsymbol{Z}$ is given by

$$H(\boldsymbol{z}) = \int_0^\infty \Pr(\boldsymbol{S} + y \leqslant \boldsymbol{z})\, e^{-y}\, \mathrm{d}y = \int_0^\infty \left\{ \sum_{j=1}^d \int_{\mathcal{S}_j} \mathbf{1}(\boldsymbol{s} + y \leqslant \boldsymbol{z}) f_{\boldsymbol{S}}(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}_{-j} \right\} e^{-y}\, \mathrm{d}y,$$

where $\mathcal{S}_j = \{\boldsymbol{s} \in \mathbb{R}^d : s_j = \max(\boldsymbol{s}) = 0\}$, $\boldsymbol{s}_{-j} \in \mathcal{S}_j$, and $\mathrm{d}\boldsymbol{s}_{-j}$ is the $(d-1)$-dimensional Lebesgue measure on $\mathcal{S}_j$. Now let $\mathcal{U}_j = \mathcal{S}_j + (0, \infty) = \{\boldsymbol{u} \in \mathbb{R}^d : u_j = \max(\boldsymbol{u}) > 0\}$ and on $\mathcal{U}_j$ make the substitution $\boldsymbol{u}_j \equiv \boldsymbol{s}_{-j} + y$, consisting of $(\boldsymbol{u}_j)_j = y$ and $(\boldsymbol{u}_j)_i = s_i + y$ for $i \neq j$, so that $y = \max(\boldsymbol{u}_j) = (\boldsymbol{u}_j)_j$. It is easily verified that the determinant of the Jacobian of this transformation is equal to 1. Write $\mathrm{d}\boldsymbol{u}_j = \mathrm{d}y\, \mathrm{d}\boldsymbol{s}_{-j}$, the d-dimensional Lebesgue measure on $\mathcal{U}_j$. By Fubini's theorem,

$$\begin{aligned} H(\boldsymbol{z}) &= \sum_{j=1}^d \int_{\mathcal{S}_j} \int_0^\infty \mathbf{1}(\boldsymbol{s} + y \leqslant \boldsymbol{z}) f_{\boldsymbol{S}}(\boldsymbol{s})\, e^{-y}\, \mathrm{d}y\, \mathrm{d}\boldsymbol{s}_{-j} = \sum_{j=1}^d \int_{\mathcal{U}_j} \mathbf{1}(\boldsymbol{u}_j \leqslant \boldsymbol{z}) f_{\boldsymbol{S}}\{\boldsymbol{u}_j - \max(\boldsymbol{u}_j)\}\, e^{-\max(\boldsymbol{u}_j)}\, \mathrm{d}\boldsymbol{u}_j \\ &= \sum_{j=1}^d \int_{\mathcal{U}_j \cap (-\infty, \boldsymbol{z}]} f_{\boldsymbol{S}}\{\boldsymbol{u}_j - \max(\boldsymbol{u}_j)\}\, e^{-\max(\boldsymbol{u}_j)}\, \mathrm{d}\boldsymbol{u}_j = \int_{(-\infty, \boldsymbol{z}]} \mathbb{1}\{\max(\boldsymbol{u}) > 0\} f_{\boldsymbol{S}}\{\boldsymbol{u} - \max(\boldsymbol{u})\}\, e^{-\max(\boldsymbol{u})}\, \mathrm{d}\boldsymbol{u}, \end{aligned}$$

with $\mathrm{d}\boldsymbol{u}$ the d-dimensional Lebesgue measure on $\bigcup_{j=1}^d \mathcal{U}_j = \{\boldsymbol{u} \in \mathbb{R}^d : \max(\boldsymbol{u}) > 0\}$. □

Proof of Proposition 16. This is straightforward from (19) and the representations of $(\boldsymbol{\pi}, \ell)$ in terms of $\mathcal{L}(\boldsymbol{S})$, $\mathcal{L}(\boldsymbol{T})$ and $\mathcal{L}(\boldsymbol{U})$. □

Proof of Proposition 17. Eq. (40) follows from (37) since $y_j = e^{-z_j}$ converges to ∞ if z_j tends to $-\infty$. Eq. (41) then follows from Eq. (40) by setting $\varepsilon = y_j^{-1}$ and using homogeneity from ℓ. Eq. (42) follows from Eq. (41), the fact that $\ell(\pi_j \boldsymbol{e}_j) = \pi_j$, and properties of directional derivatives. The first part of Eq. (43) follows by taking the limit as $z_j \to -\infty$ in (38) and applying the dominated convergence theorem together with the fact that $e^{\max(\boldsymbol{S})} = 1$ almost surely. The other two identities in (43) then follow from expressing the law of $\boldsymbol{S}$ in terms of those of $\boldsymbol{T}$ and $\boldsymbol{U}$, respectively. □

Proof of Proposition 18. For $\boldsymbol{x} \in \mathbb{R}^J$ such that $\max_{j \in J} x_j > 0$, we have,

$$\begin{aligned} \Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x} \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u}) &= \frac{\Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x}, \boldsymbol{X}_J \not\leqslant \boldsymbol{u})}{\Pr(\boldsymbol{X}_J \not\leqslant \boldsymbol{u})} = \frac{\Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x}) - \Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x}, \boldsymbol{X}_J \leqslant \boldsymbol{u})}{\Pr(\boldsymbol{X}_J \not\leqslant \boldsymbol{u})} \\ &= \frac{\Pr(\boldsymbol{X}_J \leqslant \boldsymbol{u} + \boldsymbol{x}) - \Pr\{\boldsymbol{X}_J \leqslant \boldsymbol{u} + (\boldsymbol{x} \wedge \boldsymbol{0})\}}{\Pr(\boldsymbol{X}_J \not\leqslant \boldsymbol{u})}. \end{aligned} \tag{A.5}$$

Since $\boldsymbol{u} \geqslant \boldsymbol{0}$, we have

$$\Pr(\boldsymbol{X}_J \not\leqslant \boldsymbol{u}) = \ell_J\{(\Pr[X_j > u_j])_{j \in J}\} = \ell_J\{\boldsymbol{\pi}_J (1 + \boldsymbol{\gamma}_J \boldsymbol{u}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}\}.$$

To see this, let $u_k \to \infty$ for $k \in \{1, \ldots, d\} \setminus J$ in (15). In addition, the J-margin of H is given by

$$\Pr[\boldsymbol{X}_J \leqslant \boldsymbol{v}] = \ell\{\boldsymbol{\pi}\,(1 + \boldsymbol{\gamma}(\boldsymbol{w} \wedge \boldsymbol{0})/\boldsymbol{\sigma})^{-1/\boldsymbol{\gamma}}\} - \ell_J\{\boldsymbol{\pi}_J\,(1 + \boldsymbol{\gamma}_J \boldsymbol{v}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}\}, \tag{A.6}$$

for $\boldsymbol{v} \in \mathbb{R}^J$ such that $\sigma_j + \gamma_j v_j > 0$ for all $j \in J$ and where $\boldsymbol{w} \in \mathbb{R}^d$ is defined by $w_j = v_j$ if $j \in J$ and $w_j = 0$ otherwise; this follows from (12). Substitute (A.6) for $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{x}$ and $\boldsymbol{v} = \boldsymbol{u} + (\boldsymbol{x} \wedge \boldsymbol{0})$ into (A.5). The numerator will have four instances of ℓ, two of which will cancel out because $\{\boldsymbol{u} + (\boldsymbol{x} \wedge \boldsymbol{0})\} \wedge \boldsymbol{0} = (\boldsymbol{u} + \boldsymbol{x}) \wedge \boldsymbol{0}$, a consequence of the assumption that $\boldsymbol{u} \geqslant \boldsymbol{0}$. It follows that

$$\Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x} \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u}) = \frac{\ell_J[\boldsymbol{\pi}_J\,(1 + \boldsymbol{\gamma}_J\{\boldsymbol{u} + (\boldsymbol{x} \wedge \boldsymbol{0})\}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}] - \ell_J[\boldsymbol{\pi}_J\,(1 + \boldsymbol{\gamma}_J\{\boldsymbol{u} + \boldsymbol{x}\}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}]}{\ell_J\{\boldsymbol{\pi}_J\,(1 + \boldsymbol{\gamma}_J \boldsymbol{u}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}\}}.$$

In addition, note that, for all $j \in J$,

$$\{1 + \gamma_j(u_j + y_j)/\sigma_j\}^{-1/\gamma_j} = (1 + \gamma_j u_j/\sigma_j)^{-1/\gamma_j}\{1 + \gamma_j y_j/(\sigma_j + \gamma_j u_j)\}^{-1/\gamma_j},$$

where y_j represents either x_j or $x_j \wedge 0$. Writing $\boldsymbol{\tau}_J = \boldsymbol{\pi}_J\,(1 + \boldsymbol{\gamma}_J \boldsymbol{u}/\boldsymbol{\sigma}_J)^{-1/\boldsymbol{\gamma}_J}$, we find

$$\Pr(\boldsymbol{X}_J - \boldsymbol{u} \leqslant \boldsymbol{x} \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u}) = \frac{\ell_J[\boldsymbol{\tau}_J\{1 + \boldsymbol{\gamma}_J(\boldsymbol{x} \wedge \boldsymbol{0})/(\boldsymbol{\sigma}_J + \boldsymbol{\gamma}_J \boldsymbol{u})\}^{-1/\boldsymbol{\gamma}_J}] - \ell_J[\boldsymbol{\tau}_J\{1 + \boldsymbol{\gamma}_J \boldsymbol{x}/(\boldsymbol{\sigma}_J + \boldsymbol{\gamma}_J \boldsymbol{u})\}^{-1/\boldsymbol{\gamma}_J}]}{\ell_J(\boldsymbol{\tau}_J)}.$$

Since $\tau_j = \Pr(X_j > u_j)$ and $\ell_J(\boldsymbol{\tau}_J) = \Pr(\boldsymbol{X}_J \not\leqslant \boldsymbol{u})$ and thus $\tau_j/\ell_J(\boldsymbol{\tau}_J) = \Pr(X_j > u_j \mid \boldsymbol{X}_J \not\leqslant \boldsymbol{u})$, we obtain (44). □

Proof of Proposition 19. Eq. (48) is the same as Eq. (15). Eq. (49) then follows from Eq. (48) and the inclusion–exclusion formula. □

Proof of Proposition 20. By definition, $\boldsymbol{S}$ is a spectral random vector (Definition 5) and by (18) we have

$$\boldsymbol{X} \stackrel{d}{=} \begin{cases} \sigma(e^{\gamma(\boldsymbol{S}+E)} - 1)/\gamma & \text{if } \gamma \neq 0, \\ \sigma(\boldsymbol{S} + E) & \text{if } \gamma = 0, \end{cases}$$

where E is a unit exponential random variable, independent of $\boldsymbol{S}$. Suppose $\gamma > 0$. Then

$$\boldsymbol{AX} \not\leqslant \boldsymbol{x} \Leftrightarrow \exists_{i\in\{1,\dots,m\}} \sum_{j=1}^{d} a_{i,j}\sigma_j \frac{e^{\gamma(S_j+E)} - 1}{\gamma} > x_i \Leftrightarrow \exists_{i\in\{1,\dots,m\}} e^{\gamma E} \sum_{j=1}^{d} a_{i,j}\sigma_j e^{\gamma S_j} - \sum_{j=1}^{d} a_{i,j}\sigma_j > \gamma x_i$$

$$\Leftrightarrow \exists_{i\in\{1,\dots,m\}} \left(\frac{\sum_{j=1}^{d} a_{i,j}\sigma_j e^{\gamma S_j}}{\sum_{j=1}^{d} a_{i,j}\sigma_j + \gamma x_i} \right)^{1/\gamma} > e^{-E}.$$

The random variable e^{-E} is independent of $\boldsymbol{S}$ and is uniformly distributed on the interval $[0, 1]$. We find

$$\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x}) = \mathrm{E}\left\{ 1 \wedge \max_{i\in\{1,\dots,m\}} \left(\frac{\sum_{j=1}^{d} a_{i,j}\sigma_j e^{\gamma S_j}}{\sum_{j=1}^{d} a_{i,j}\sigma_j + \gamma x_i} \right)^{1/\gamma} \right\}. \tag{A.7}$$

If $\gamma < 0$, then a similar argument yields the same expression, while if $\gamma = 0$, we can apply a similar reasoning to find that

$$\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x}) = \mathrm{E}\left\{ 1 \wedge \max_{i\in\{1,\dots,m\}} \exp\left(\frac{\sum_{j=1}^{d} a_{i,j}\sigma_j S_j}{\sum_{j=1}^{d} a_{i,j}\sigma_j} - \frac{x_i}{\sum_{j=1}^{d} a_{i,j}\sigma_j} \right) \right\}. \tag{A.8}$$

Since $\sum_{j=1}^{d} a_{i,j}\sigma_j = \boldsymbol{A}_i\boldsymbol{\sigma}$, the two expressions for $\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x})$ in (A.7)–(A.8) are equal to the one claimed in (50) with U_i given by (51).

Next we compute the conditional distribution of $\boldsymbol{AX}$ given that $\boldsymbol{AX} \not\leqslant \boldsymbol{0}$. For $\boldsymbol{x} \in \mathbb{R}^d$, we have, by a computation similar as the one leading to (A.5),

$$\Pr(\boldsymbol{AX} \leqslant \boldsymbol{x} \mid \boldsymbol{AX} \not\leqslant \boldsymbol{0}) = \frac{\Pr(\boldsymbol{AX} \leqslant \boldsymbol{x},\ \boldsymbol{AX} \not\leqslant \boldsymbol{0})}{\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{0})} = \frac{\Pr(\boldsymbol{AX} \leqslant \boldsymbol{x}) - \Pr(\boldsymbol{AX} \leqslant \boldsymbol{x} \wedge \boldsymbol{0})}{\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{0})}$$

$$= \frac{\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x} \wedge \boldsymbol{0}) - \Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x})}{\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{0})}.$$

For $\boldsymbol{x}$ such that $\boldsymbol{A}_i\boldsymbol{\sigma} + \gamma x_i > 0$ for all $i \in \{1, \dots, m\}$, the three probabilities can be worked out using (50). Regarding the denominator: since $U_i \leqslant 0$ almost surely, we find $\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{0}) = \mathrm{E}\{e^{\max(\boldsymbol{U})}\}$. Regarding the numerator: apply (50) twice, to $\boldsymbol{x} \wedge \boldsymbol{0}$ and to $\boldsymbol{x}$ itself. We find

$$\begin{aligned} &\Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x} \wedge \boldsymbol{0}) - \Pr(\boldsymbol{AX} \not\leqslant \boldsymbol{x}) \\ &= \mathrm{E}\left[1 \wedge \max_{i\in\{1,\dots,m\}} \{(1 + \gamma(x_i \wedge 0)/\boldsymbol{A}_i\boldsymbol{\sigma})^{-1/\gamma}\, e^{U_i}\} \right] - \mathrm{E}\left[1 \wedge \max_{i\in\{1,\dots,m\}} \{(1 + \gamma x_i/\boldsymbol{A}_i\boldsymbol{\sigma})^{-1/\gamma}\, e^{U_i}\} \right] \\ &= \mathrm{E}\left[\max_{i\in\{1,\dots,m\}} \{(1 + \gamma(x_i \wedge 0)/\boldsymbol{A}_i\boldsymbol{\sigma})^{-1/\gamma}\, e^{U_i}\} \right] - \mathrm{E}\left[\max_{i\in\{1,\dots,m\}} \{(1 + \gamma x_i/\boldsymbol{A}_i\boldsymbol{\sigma})^{-1/\gamma}\, e^{U_i}\} \right]. \end{aligned}$$

The reason we may omit the two instances of "$1 \wedge \dots$" is again because $U_i \leqslant 0$ almost surely. The identity can be confirmed by a case-by-case analysis.

Comparing the resulting expression for $\Pr(\boldsymbol{AX} \leqslant \boldsymbol{x} \mid \boldsymbol{AX} \not\leqslant \boldsymbol{0})$ with (30) confirms that the law of $\boldsymbol{AX}$ given that $\boldsymbol{AX} \not\leqslant \boldsymbol{0}$ is given by the $\mathrm{GP}_U[\boldsymbol{\gamma}, \boldsymbol{A\sigma}, \mathcal{L}(\boldsymbol{U})]$ distribution. □

References

[1] B. Basrak, J. Segers, Regularly varying multivariate time series, Stochastic Process. Appl. 119 (2009) 1055–1080.
[2] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes: Theory and Applications, Wiley, New Yor, 2004.
[3] L. de Haan, A. Ferreira, Extreme Value Theory: An Introduction, Springer, Berlin, 2006.
[4] C. Dombry, M. Ribatet, Functional regular variations, Pareto processes and peaks over threshold, Stat. Interface 8 (2015) 9–17.
[5] H. Drees, X. Huang, Best attainable rates of convergence for estimates of the stable tail dependence function, J. Multivariate Anal. 64 (1998) 25–47.
[6] J.H.J. Einmahl, L. de Haan, D. Li, Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition, Ann. Statist. 34 (2006) 1987–2014.
[7] J.H.J. Einmahl, A. Krajina, J. Segers, An M-estimator for tail dependence in arbitrary dimensions, Ann. Statist. 40 (2012) 1764–1793.
[8] M. Falk, J. Hüsler, R.-D. Reiß, Laws of Small Numbers: Extremes and Rare Events, Springer, Berlin, 2010.
[9] A. Ferreira, L. de Haan, The generalized Pareto process; with a view towards application and simulation, Bernoulli 20 (2014) 1717–1737.
[10] H. Joe, H. Li, A.K. Nikoloulopoulos, Tail dependence functions and vine copulas, J. Multivariate Anal. 101 (2010) 252–270.
[11] A. Kiriliouk, H. Rootzén, J. Segers, J.L. Wadsworth, Peaks over thresholds modelling with multivariate generalized Pareto distributions, arXiv:1612.01773, 2016.
[12] A.W. Marshall, I. Olkin, Domains of attraction of multivariate extreme value distributions, Ann. Probab. 11 (1983) 168–177.

[13] R.B. Nelsen, An Introduction To Copulas, second ed., Springer, New York, 2006.
[14] J. PickandsI II, Multivariate extreme value distributions, In: Proceedings of the 43rd Session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981) 49 859–878 1981 + 894–902.
[15] P. Ressel, Homogeneous distributions and a spectral representation of classical mean values and stable tail dependence functions, J. Multivariate Anal. 117 (2013) 246–256.
[16] H. Rootzén, J. Segers, J.L. Wadsworth, Multivariate peaks over thresholds models, Extremes, https://doi.org/10.1007/s10687-017-0294-4.
[17] H. Rootzén, N. Tajvidi, Multivariate generalized Pareto distributions, Bernoulli 12 (2006) 917–930.
[18] M. Schlather, J.A. Tawn, Inequalities for the extremal coefficients of multivariate extreme value distributions, Extremes 5 (2002) 87–102.
[19] F. Schmid, R. Schmidt, T. Blumentritt, S. Gaisser, M. Ruppert, Copula-based measures of multivariate association, in: P. Jaworski, F. Durante, W.K. Härdle, T. Rychlik (Eds.), Copula Theory and Its Applications, Berlin, Heidelberg, 2010, pp. 209–236.
[20] R. Schmidt, U. Stadtmüller, Non-parametric estimation of tail dependence, Scand. J. Stat. 33 (2006) 307–335.
[21] J. Segers, Max-stable models for multivariate extremes, REVSTAT–Stat. J. 10 (2012) 61–82.
[22] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.