

Statistical inference for multivariate extremes via a geometric approach

Jennifer Wadsworth*

Department of Mathematics and Statistics, Lancaster University, U.K.
and

Ryan Campbell

Department of Mathematics and Statistics, Lancaster University, U.K.

March 7, 2024

Abstract

A geometric representation for multivariate extremes, based on the shapes of scaled sample clouds in light-tailed margins and their so-called limit sets, has recently been shown to connect several existing extremal dependence concepts. However, these results are purely probabilistic, and the geometric approach itself has not been fully exploited for statistical inference. We outline a method for parametric estimation of the limit set shape, which includes a useful non/semi-parametric estimate as a pre-processing step. More fundamentally, our approach provides a new class of asymptotically-motivated statistical models for the tails of multivariate distributions, and such models can accommodate any combination of simultaneous or non-simultaneous extremes through appropriate parametric forms for the limit set shape. Extrapolation further into the tail of the distribution is possible via simulation from the fitted model. A simulation study confirms that our methodology is very competitive with existing approaches, and can successfully allow estimation of small probabilities in regions where other methods struggle. We apply the methodology to two environmental datasets, with diagnostics demonstrating a good fit.

Keywords: Extrapolation, limit set, multivariate extremes, tail dependence

*Corresponding author: j.wadsworth@lancaster.ac.uk

1 Introduction

1.1 Multivariate extreme value theory

Multivariate extreme value theory provides the basis for estimation of rare event probabilities that involve the effect of more than one variable. Applications are diverse and include estimating flood risk (Keef et al., 2013; Engelke and Hitz, 2020), extreme air pollution levels (Heffernan and Tawn, 2004; Vettori et al., 2019), structural design (Coles and Tawn, 1994), dietary risk assessment (Chautru, 2015) and financial risk assessment (Zhang and Huang, 2006; Hilal et al., 2014).

The study of multivariate extremes primarily began in the 1970s and 80s, with the theoretical study of multivariate regular variation (de Haan, 1970; de Haan and Resnick, 1977; Resnick, 1987). Multivariate regular variation is intrinsically tied up with the componentwise block maximum method for multivariate extremes. Suppose we have n independent replicates of a random vector $\mathbf{Y}_i \in \mathbb{R}^d$, $i = 1, \dots, n$; the componentwise maximum vector is

$$\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d}) = \left(\max_{1 \leq i \leq n} Y_{1,i}, \dots, \max_{1 \leq i \leq n} Y_{d,i} \right).$$

Univariate extreme value theory tells us that if, for each $j = 1, \dots, d$, there exists $a_{n,j} > 0, b_{n,j}$ such that $(M_{n,j} - b_{n,j})/a_{n,j}$ converges to a non-degenerate random variable, then the distribution of this limiting variable is generalized extreme value (Fisher and Tippett, 1928; Gnedenko, 1943), which is the only univariate *max-stable* distribution. A distribution is max-stable if it is invariant to the operation of taking (componentwise) block maxima, up to marginal location and scale changes. The additional condition for joint convergence of the entire vector $(\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n$ to a multivariate max-stable distribution is multivariate regular variation. Since this represents an assumption on the dependence structure it can be expressed in standardized margins: a common choice is to set $X_{P,j} = 1/[1 - F_j(Y_j)]$, where $X_{P,j}$ follows a standard Pareto distribution if $Y_j \sim F_j$ has a continuous distribution, else it is asymptotically Pareto. A common way to express the multivariate regular variation assumption is

$$\lim_{t \rightarrow \infty} \Pr(\mathbf{X}_P / \|\mathbf{X}_P\| \in B, \|\mathbf{X}_P\| > ts \mid \|\mathbf{X}_P\| > t) = s^{-1} H(B), \quad s > 1, \quad (1)$$

where $B \subset \mathcal{S}_{d-1} = \{\mathbf{v} \in [0, 1]^d : \|\mathbf{v}\| = 1\}$ is a measurable set with $H(\partial B) = 0$. Assumption (1) shows that large values of the “radial” component $\|\mathbf{X}_P\|$ become independent of the “angular” component $\mathbf{X}_P / \|\mathbf{X}_P\|$, which follows some probability distribution H on \mathcal{S}_{d-1} , commonly referred to as the *spectral measure*. The choice of norm $\|\cdot\|$ is arbitrary, see, e.g., Beirlant et al. (2004, Chap. 8), but the most common choice is the L_1 norm $\|\cdot\|_1$, so that $\mathcal{S}_{d-1} = \{\mathbf{v} \in [0, 1]^d : \sum_{j=1}^d v_j = 1\}$.

Statistical methodology for multivariate extremes followed shortly after this theoretical study, and focused initially on inference for data arising as componentwise block maxima through parametrized

forms of multivariate max-stable distributions (Tawn, 1990). This was soon followed by more direct exploitation of the multivariate regular variation assumption (1), whereby parametric models were proposed for the spectral measure H , and inference performed on these (Coles and Tawn, 1991).

The study of componentwise maxima is a natural multivariate extension of the univariate block maximum approach, and the associated multivariate regular variation dependence condition (1) widely applicable. However, it has been known for a long time that while examples not satisfying (1) are rare, the number of examples for which this assumption forms a *useful* basis for statistical inference is very much smaller. This is because, for many dependence structures, mass of the spectral measure H accumulates on one or more regions of the form

$$\mathbb{B}_C = \{\mathbf{v} \in \mathbb{S}_{d-1} : v_j > 0, j \in C; v_k = 0, k \notin C\}, \quad C \subset \{1, \dots, d\}. \quad (2)$$

When this is the case, joint extremes of the random vector \mathbf{Y} (or equivalently \mathbf{X}_P) may not always occur simultaneously; see, e.g., Goix et al. (2017) or Simpson et al. (2020) for a more detailed explanation. In practice, however, we never observe mass on such sets \mathbb{B}_C at finite levels. This is illustrated in Figure 1, which displays the distribution of $\mathbf{X}_P/\|\mathbf{X}_P\|_1$ when the associated radial variable $\|\mathbf{X}_P\|_1$ exceeds its 0.98 quantile for two examples. In the left panel, the true limiting spectral measure H places mass only on the points $\mathbb{B}_{\{1\}} = \{(1, 0)\}$ and $\mathbb{B}_{\{2\}} = \{(0, 1)\}$, yet at observable levels, the distribution of angles is relatively evenly spread over $\mathbb{B}_{\{1,2\}}$, represented by the interval $(0, 1)$. In the right panel, the limiting spectral measure places mass only on $\mathbb{B}_{\{3\}}$ and $\mathbb{B}_{\{1,2\}}$, but once again, at observable levels we see all values in $\mathbb{B}_{\{1,2,3\}}$. A consequence of this mismatch between finite-sample and limiting distribution is a common modelling assumption that H places all mass on $\mathbb{B}_{\{1,\dots,d\}} = \{\mathbf{v} \in \mathbb{S}_{d-1} : v_j > 0, j \in \{1, \dots, d\}\}$, leading to overestimation of the probability of joint extremes. Moreover, even if one successfully detects the location of mass of the limiting object H , this does not lead to a practical strategy for performing extrapolation beyond the observed values; to achieve this, more detailed information on the behaviour of \mathbf{X}_P before the limiting regime is required.

1.2 Geometric approach to multivariate extremes

The early study of multivariate regular variation was followed by a smaller body of work that examined the convergence of light-tailed multivariate sample clouds onto so-called limit sets (Davis et al., 1988; Kinoshita and Resnick, 1991). These ideas did not have a clear link with multivariate max-stable models and did not lead to the same proliferation of statistical methodology. More recently, several papers have revisited this geometric approach from a theoretical perspective (Balkema et al., 2010; Balkema and Nolde, 2010, 2012; Nolde, 2014; Nolde and Wadsworth, 2022) and in some cases shown how the shape of the limit set links to whether joint extremes of certain variables can occur.

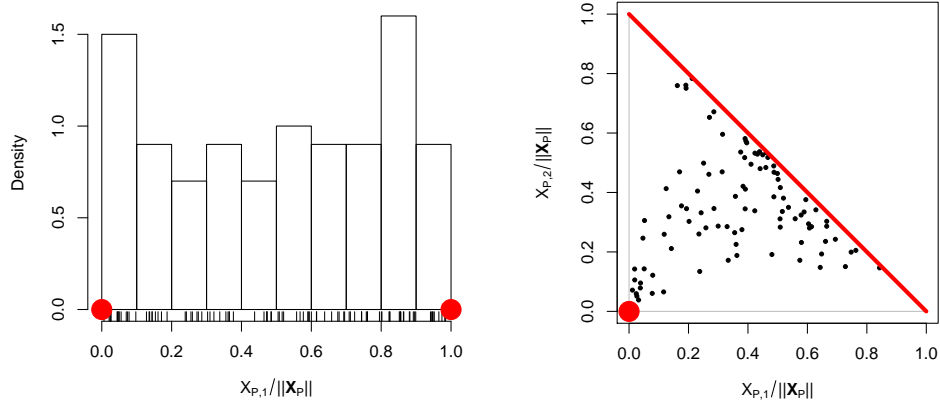


Figure 1: Illustration of the distribution of $\mathbf{X}_P/\|\mathbf{X}_P\|_1$, given that $\|\mathbf{X}_P\|_1$ is large, for two examples. Left panel: \mathbf{X}_P has a bivariate Gaussian dependence structure with correlation parameter $\rho = 0.8$. The ticks correspond to observed angles $X_{P,1}/\|\mathbf{X}_P\|_1$ associated to large radii, along with a histogram of their density. The position of mass of the theoretical limiting H is illustrated by large filled circles (red). Right panel: $(X_{P,1}, X_{P,2})$ have an inverted Clayton dependence structure while $X_{P,3}$ is simulated conditional upon the value of $X_{P,2}$ so that these variables have an inverted logistic dependence structure. Points correspond to the observed distribution of $(X_{P,1}, X_{P,2})/\|\mathbf{X}_P\|$ with large radii. The position of mass of the theoretical limiting H is illustrated by the large filled circle and thick solid line (red).

To make ideas more concrete consider n independent copies of a random vector \mathbf{X}_i , $i = 1, \dots, n$, with standard exponential margins; in practice, this will typically involve marginal transformation of the original vectors \mathbf{Y}_i . The *scaled n -point sample cloud* is defined as

$$N_n = \{\mathbf{X}_1/\log n, \dots, \mathbf{X}_n/\log n\},$$

and we assume that this converges onto a *limit set* $G = \{\mathbf{x} \in \mathbb{R}_+^d : g(\mathbf{x}) \leq 1\}$, where g is the 1-homogeneous *gauge function* of the limit set. This convergence is illustrated in Figure 2 for data with a logistic dependence structure, where the shape of the limit set can be seen to emerge in the scaled sample cloud as n becomes large. The precise sense of convergence of N_n onto G , and necessary and sufficient conditions for it, can be found in [Balkema et al. \(2010\)](#). Loosely, these conditions say that the expected number of points from N_n lying in sets that intersect with the limit set tends to infinity, whereas the expected number of points lying in sets that are disjoint from the limit set converges to zero. However, these specific conditions are rather unintuitive and make it difficult to determine the form of G for a given distribution, which led [Nolde \(2014\)](#) and [Nolde](#)

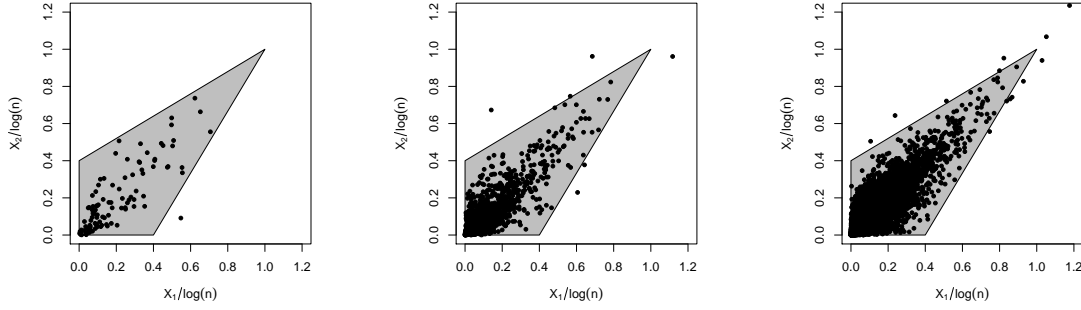


Figure 2: Illustration of the convergence of the scaled sample cloud N_n onto a limit set. From left to right, sample sizes are $n = 100, 1000, 10000$. The limit set G is depicted by the grey polygon.

and Wadsworth (2022) to consider alternative conditions in terms of the joint Lebesgue density of \mathbf{X} , when it exists. Denoting this joint density by $f_{\mathbf{X}}$, a sufficient condition for convergence of N_n onto G is

$$\lim_{t \rightarrow \infty} -\log f_{\mathbf{X}}(t\mathbf{x})/t = g(\mathbf{x}), \quad \mathbf{x} \in [0, \infty)^d, \quad (3)$$

for a continuous gauge function g . Given that many statistical models have tractable joint densities and continuous gauge functions, equation (3) provides a simple way to determine the form of g , and hence G , in several examples (Nolde and Wadsworth, 2022). Further illustrations of limit sets G are given for $d = 2, 3$ in Section A of the supplement.

The shape of G is important as a description of the extremal dependence of the underlying random vector. Limit sets exist for a much more general class of light-tailed marginal distributions than exponential, but we specialize to this case so that there is a clear correspondence between the shape of G and the dependence structure. In this case, the coordinatewise supremum of the limit set G is the point $(1, \dots, 1)$, since for independent copies of an exponential variable X_i , the random variable $\max_{1 \leq i \leq n} X_i / \log n$ converges in probability to 1. Gauge functions, and therefore limit sets, of lower dimensional margins indexed by $J \subset \{1, \dots, d\}$ can be found through the following minimization operation (Nolde and Wadsworth, 2022):

$$g_J(\mathbf{x}_J) = \min_{x_k \geq 0, k \notin J} g(\mathbf{x}),$$

where $\mathbf{x}_J = (x_j)_{j \in J}$. Exponential margins implies that for singleton sets $J = \{j\}$, $g_{\{j\}}(x_j) = x_j$.

Nolde and Wadsworth (2022) showed how G can be used to determine an array of extremal dependence measures which generally relate to representations of multivariate extremes that are more useful than multivariate regular variation when the spectral measure H places mass on one or more sets \mathbb{B}_C as in equation (2). These include expressions for determining the residual tail

dependence coefficient (Ledford and Tawn, 1997), key elements of the conditional extremes model (Heffernan and Tawn, 2004), the angular dependence function (Wadsworth and Tawn, 2013), and the dependence coefficients of Simpson et al. (2020), which can be used to help determine the sets \mathbb{B}_C on which H places mass.

Given the importance of the shape of G , a natural question that arises is how to estimate this from a sample of data. To date this question has been studied very little indeed; Jacob and Massé (1996) study estimation from a theoretical perspective but with no implementation. Very recently, Simpson and Tawn (2022) outlined an estimation approach in the bivariate case.

In this paper, we consider estimation of G as part of a wider new approach to the statistical analysis of extreme values. While G is an object of interest in itself, we direct our methodology more broadly at the question of statistical modelling and extrapolation for multivariate extreme values rather than focusing only on the descriptive aspects of extremal dependence that come from estimation of G . Our modelling approach allows in principle for any combination of joint extremes of sub-vectors of \mathbf{Y} (equivalently, H may place mass on any valid combination of sets \mathbb{B}_C), and permits extrapolation in all directions, i.e., into the joint tail where all variables are large, or into other regions of the multivariate tail where only some variables are large. Existing alternatives to methodology based on multivariate regular variation do not capture these possibilities in a coherent manner.

To illustrate the potential importance of being able to capture complex structure in extremes, consider the dataset of river flow measurements from Simpson et al. (2020) that will also be analyzed in Section 6.2. Their analysis showed that there are some events where all four rivers were extreme simultaneously, but that there were also extreme episodes involving single rivers, or groups of two or three rivers without the others. This might be explained physically by the weather patterns causing the extremes, and the relationships between catchments. While Simpson et al. (2020) introduced and estimated coefficients to help determine this structure, they did not provide any modelling approach that could account for it, as we do here.

Section 2 outlines our statistical model and assumptions. Section 3 details theoretical examples that demonstrate applicability of the method. We focus on details of statistical inference in Section 4, and use simulation to show that our approach is very competitive for estimation of extreme set probabilities in a wide range of scenarios in Section 5. Section 6 contains applications to oceanographic and fluvial datasets, and we conclude in Section 7.

2 Model and assumptions

Here and throughout the rest of the paper, we assume that we have a random vector \mathbf{X} with standard exponential margins and joint Lebesgue density denoted by $f_{\mathbf{X}}$. Marginal transformation can be applied as a standard step via estimation of each marginal distribution function. The assumption of a joint density is very common for statistical analysis, as it is required for most likelihood-based inference, for example. We further assume that the scaled sample cloud N_n converges onto a limit set G whose shape can either be described by a continuous gauge function g , or that we are only interested in the continuous part.

Assumption (3), which yields a sufficient condition for convergence of N_n onto G , can equivalently be expressed $f_{\mathbf{X}}(t\mathbf{x}) = \exp\{-tg(\mathbf{x})[1 + o(1)]\}$ for $g(\mathbf{x}) > 0$ as $t \rightarrow \infty$. The homogeneity of g suggests making the radial-angular transformation $R = \sum_{j=1}^d X_j$, $\mathbf{W} = \mathbf{X}/R$; such transformations are common in multivariate extremes, but normally on Pareto, rather than exponential, margins. The Jacobian of this transformation is r^{d-1} , which leads to joint density of (R, \mathbf{W}) : $f_{R, \mathbf{W}}(r, \mathbf{w}) = r^{d-1} f_{\mathbf{X}}(r\mathbf{w}) = r^{d-1} \exp\{-rg(\mathbf{w})[1 + o(1)]\}$, as $r \rightarrow \infty$. This in turn means that the conditional density of $R \mid \mathbf{W} = \mathbf{w}$ satisfies

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} \exp\{-rg(\mathbf{w})[1 + o(1)]\}, \quad r \rightarrow \infty. \quad (4)$$

If we were to ignore the $o(1)$ term, we recognize the form of the gamma kernel in equation (4), suggesting that when $R \mid \mathbf{W} = \mathbf{w}$ is large, its distribution could potentially be well approximated by a gamma distribution. Indeed, if the $o(1)$ term in the exponent is negligible, this suggests a truncated gamma approximation above a high threshold $r_0(\mathbf{w})$ of the conditional distribution $R \mid \mathbf{W} = \mathbf{w}$.

A valid concern is whether the $o(1)$ term in the exponent of (4) is really negligible. In Section 3 we detail several examples which in fact have the more helpful asymptotic form $f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} \exp[-rg(\mathbf{w})][1 + o(1)]$, i.e., with the $o(1)$ outside of the exponent, and give explicit rates for this term. Based on this latter asymptotic representation, we focus in this paper on the model

$$R \mid [\mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})] \sim \text{truncGamma}(\alpha, g(\mathbf{w})), \quad (5)$$

where $\alpha > 0$ is the gamma shape, and $g(\mathbf{w})$ is the gamma rate parameter. In most examples, the theoretical shape parameter is $\alpha = d$, but for modelling purposes the flexibility of an estimated shape is desirable. By parametrizing flexible forms for the gauge function $g(\mathbf{w}) = g(\mathbf{w}; \boldsymbol{\theta})$, we can use approximation (5) to estimate these parameters. Full details of our approach are given in Section 4, including diagnostic plots for assessing approximation (5).

3 Examples

In this section we consider a variety of examples. The convergence onto a limit set G holds very broadly, and in many examples the gauge function for this limit set in exponential margins can be recovered fully or partly from convergence (3). The form of the gauge function and limit set for several examples, including multivariate t_ν , light-tailed elliptical, skew-normal, generalized hyperbolic, certain mixture distributions and multivariate generalized Pareto forms has been derived in Balkema et al. (2010), Nolde (2014), Nolde and Wadsworth (2022), and Zhang et al. (2022), for example, although not always in exponential margins.

The validity and quality of the truncated gamma approximation in (5) to the conditional density in (4) depends on the $o(1)$ term. Since this lies in the exponent, it is not always guaranteed to be negligible. In this section, we explicitly calculate the density of $R \mid \mathbf{W} = \mathbf{w}$ for various theoretical examples, showing that most in fact have the form $f_{R \mid \mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} \exp[-rg(\mathbf{w})][1 + o(1)]$, as $r \rightarrow \infty$. The exception to this is the Gaussian dependence structure, for which we find $f_{R \mid \mathbf{W}}(r \mid \mathbf{w}) \propto r^{\alpha(\mathbf{w})-1} \exp[-rg(\mathbf{w})][1 + o(1)]$, as $r \rightarrow \infty$, i.e., the conditional gamma form is still applicable, but the shape parameter depends on the value of \mathbf{w} . Nonetheless, further investigations, described briefly below and in more detail in Section C of the supplement, show the assumption of a common shape in model (5) does not appear problematic in practice. This is also supported by our simulation study in Section 5. More generally, we will incorporate model checking of assumption (5) into our statistical analysis.

For each distribution, we provide the overall form of $f_{R \mid \mathbf{W}}(r \mid \mathbf{w})$, with further calculations given in Section B of the supplement. We denote the ordered values of the vector \mathbf{w} (and similarly \mathbf{x}) by $w_{(1)} \geq w_{(2)} \geq \dots \geq w_{(d)} > 0$, assuming the minimum to be positive. In the convergence rates given below, we assume a strict ordering $w_{(1)} > w_{(2)} > \dots > w_{(d)} > 0$; where this is not the case, following the derivations in the supplement, one usually observes improved rates, e.g., $O(e^{-r(w_{(d-2)}-w_{(d)})})$ replacing $O(e^{-r(w_{(d-1)}-w_{(d)})})$ if $w_{(d-2)} > w_{(d-1)} = w_{(d)}$.

Multivariate max-stable and generalized Pareto distributions Multivariate max-stable distributions are most readily expressed by their distribution functions. In exponential margins,

$$\Pr(\mathbf{X} \leq \mathbf{x}) = \exp\left(-V\left\{[-\log(1 - e^{-\mathbf{x}})]^{-1}\right\}\right),$$

where $V : \mathbb{R}_+^d \rightarrow \mathbb{R}_+$ is the homogeneous of order -1 *exponent function*, and operations are applied componentwise. The general asymptotic form of the density for a max-stable distribution in exponential margins is therefore

$$f_{\mathbf{X}}(t\mathbf{x}) = \exp\{-V[e^{t\mathbf{x}} + 1/2 + O(e^{-t\mathbf{x}})]\} \sum_{\pi \in \Pi} \prod_{s \in \pi} V_s[e^{t\mathbf{x}} + 1/2 + O(e^{-t\mathbf{x}})] \times e^{t \sum_{j=1}^d x_j} [1 + O(e^{-2tx_{(d)}})],$$

$t \rightarrow \infty$, where Π is the set of all partitions of $\{1, \dots, d\}$, and $V_s(\mathbf{z}) = \partial^{|s|} V(\mathbf{z}) / \prod_{j \in s} \partial z_j$.

We focus on the d -dimensional logistic distribution, for which $V(\mathbf{z}) = \left(\sum_{j=1}^d z_j^{-1/\gamma} \right)^\gamma$ with parameter $\gamma \in (0, 1]$. This distribution has gauge function $g(\mathbf{x}) = \sum_{j=1}^d x_j/\gamma + (1 - d/\gamma)x_{(d)}$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} [1 + O(e^{-r(w_{(d-1)} - w_{(d)})/\gamma}) + O(e^{-rw_{(d)}})], \quad r \rightarrow \infty.$$

The simpler form of the densities make calculations more straightforward for corresponding multivariate generalized Pareto distributions (MGPDs), which are related to max-stable distributions (Rootzén and Tajvidi, 2006; Rootzén et al., 2018). The support of MGPDs whose margins have unit scale and zero shape is contained in $\{\mathbf{x} \in \mathbb{R}^d : x_{(1)} > 0\}$. Densities for several models for which the spectral measure H places mass only on $\mathbb{B}_{\{1, \dots, d\}}$ are given in Kiriliouk et al. (2019); in such cases the dependence structure can be determined by focusing on large values of $\mathbf{x} > \mathbf{0}$. Further details are in the supplement, Section B.

For the MGPD associated to the negative logistic max-stable distribution (Galambos, 1975; Dombry et al., 2016), $g(\mathbf{x}) = (1 + d\gamma)x_{(1)} - \sum_{j=1}^d x_j\gamma$, $\gamma > 1$ and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} \left[1 + O\left(e^{r(w_{(2)} - w_{(1)})\gamma}\right) \right], \quad r \rightarrow \infty.$$

For the MGPD associated to the Dirichlet max-stable distribution (Coles and Tawn, 1991), $g(\mathbf{x}) = (1 + \sum_{j=1}^d \theta_j)x_{(1)} - \sum_{j=1}^d \theta_j x_j$, for all $\theta_j > 0$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} \left[1 + O\left(e^{r(w_{(2)} - w_{(1)})}\right) \right], \quad r \rightarrow \infty.$$

Inverted max-stable distributions Inverted max-stable distributions are derived by translating the joint lower tail of max-stable distributions to be the joint upper tail. This is achieved by applying a monotonically decreasing marginal transformation to a max-stable random vector. In exponential margins inverted max-stable distributions have density

$$f_{\mathbf{X}}(\mathbf{x}) = \exp[-l(\mathbf{x})] \sum_{\pi \in \Pi} \prod_{s \in \pi} l_s(\mathbf{x}),$$

where l is the *stable tail dependence function* of the corresponding max-stable distribution, obtained via $l(\mathbf{x}) = V(1/\mathbf{x})$, and $l_s(\mathbf{x}) = \partial^{|s|} l(\mathbf{x}) / \prod_{j \in s} \partial x_j$. The gauge function is always $g(\mathbf{x}) = l(\mathbf{x})$. Owing to the fact that $l_s(\mathbf{x})$ is homogeneous of order $1 - |s|$, we obtain

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} \exp[-rg(\mathbf{w})][1 + O(r^{-1})], \quad r \rightarrow \infty.$$

Multivariate Gaussian distribution We consider the multivariate Gaussian dependence structure with correlation matrix Σ . When one or more correlation parameters is negative then the

continuous convergence $-\log f_{\mathbf{X}}(t\mathbf{x})/t \rightarrow g(\mathbf{x})$ fails when components of \mathbf{x} are zero, because the gauge function of the limit set is not continuous away from the strictly positive orthant (Nolde and Wadsworth, 2022). Since we are considering $w_{(d)} > 0$ this is not an issue here, but we note that to fully capture negative association it is ideal to reformulate ideas in terms of Laplace rather than exponential margins; see Nolde and Wadsworth (2022) and Section 7. For Σ with non-negative entries, $g(\mathbf{x}) = (\mathbf{x}^{1/2})^\top \Sigma^{-1} \mathbf{x}^{1/2}$, where $\mathbf{x}^{1/2} = (x_1^{1/2}, \dots, x_d^{1/2})^\top$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{\alpha(\mathbf{w})-1} \exp[-rg(\mathbf{w})] \left[1 + O\left(\frac{(\log r)^2}{r}\right) \right], \quad \alpha(\mathbf{w}) = \frac{d}{2} + \frac{(\mathbf{w}^{1/2})^\top \Sigma^{-1} \mathbf{w}^{-1/2}}{2},$$

$r \rightarrow \infty$. In this case, the gamma shape parameter therefore depends on \mathbf{w} , and the region on which $\alpha(\mathbf{w}) > 0$ depends on the entries of Σ . We investigate this further in Section C of the supplement, showing that local estimates of α do not vary strongly with \mathbf{w} and may reasonably be assumed constant. We also show that results from our model are useful even in the (typically small) regions where $\alpha(\mathbf{w}) \leq 0$.

Multivariate t_ν distribution We consider the multivariate t distribution with ν degrees of freedom, focusing only on positive dependence; see the supplement for further comment. The gauge function is $g(\mathbf{x}) = (1 + d/\nu)x_{(1)} - \sum_{j=1}^d x_j/\nu$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} \left[1 + O(e^{r(w_{(2)}-w_{(1)})/\nu}) + O(e^{-2rw_{(d)}/\nu}) \right], \quad r \rightarrow \infty.$$

Clayton and inverted Clayton copulas We consider the Clayton and inverted Clayton copulas with parameter $\gamma > 0$. The Clayton copula has $g(\mathbf{x}) = \sum_{j=1}^d x_j$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} [1 + O(e^{-rw_{(d)}})], \quad r \rightarrow \infty.$$

The inverted Clayton copula has $g(\mathbf{x}) = (1 + d\gamma)x_{(1)} - \sum_{j=1}^d x_j\gamma$, and

$$f_{R|\mathbf{W}}(r \mid \mathbf{w}) \propto r^{d-1} e^{-rg(\mathbf{w})} [1 + O(e^{-r(w_{(2)}-w_{(1)})})], \quad r \rightarrow \infty.$$

In the supplement, we also calculate $f_{R|\mathbf{W}}(r \mid \mathbf{w})$ for a trivariate vine copula example.

4 Statistical inference

4.1 Calculating the threshold $r_0(\mathbf{w})$

To implement model (5), we firstly need to calculate $r_0(\mathbf{w})$, which represents a high threshold of the conditional distribution $R \mid \mathbf{W} = \mathbf{w}$. A natural approach to calculating this threshold is quantile regression, treating \mathbf{W} as the covariate. A similar approach has been taken in the context

of establishing covariate-dependent thresholds in univariate extreme value analysis (Northrop and Jonathan, 2011). When data are bivariate, so that $\mathbf{W} \in \mathcal{S}_1$ is equivalent to $W \in [0, 1]$, this approach is straightforward. However, standard parametric quantile regression requires a high degree of manual tuning to ensure that the model form captures the relation between R and \mathbf{W} well. We therefore suggest using additive quantile regression (Fasiolo et al., 2021) via the corresponding R package `qgam`.

When $\mathbf{W} \in \mathcal{S}_{d-1}$, $d > 2$, then both parametric and additive quantile regression become more difficult due to the specific support of \mathbf{W} on the simplex. A simple alternative is to calculate quantiles of $R \mid \mathbf{W} = \mathbf{w}$ from overlapping blocks of \mathbf{W} values, which is feasible for relatively low dimensions, but becomes more laborious as d grows. The top row of Figure 3 illustrates the concepts for $d = 2, 3$. In each case, $r_0(\mathbf{w})$ is calculated as the 0.95 quantile of $R \mid \mathbf{W} = \mathbf{w}$.

In the second row of Figure 3 we demonstrate that the threshold $r_0(\mathbf{w})$, suitably rescaled, can be viewed as a non-/semi-parametric estimate of g . The reason for this can roughly be explained by considering the case where the gamma approximation is exact. Let $\bar{F}(r \mid \mathbf{w})$ be the (gamma) survival function of $R \mid \mathbf{W} = \mathbf{w}$, then for quantile regression at level $\tau \in (0, 1)$, $\bar{F}[r_0(\mathbf{w}) \mid \mathbf{w}] = 1 - \tau$. We have

$$\begin{aligned} \bar{F}[r_0(\mathbf{w}) \mid \mathbf{w}] &= 1 - \int_0^{r_0(\mathbf{w})} \frac{g(\mathbf{w})^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-vg(\mathbf{w})} dv \\ &= 1 - \int_0^{r_0(\mathbf{w})g(\mathbf{w})} \frac{s^{\alpha-1}}{\Gamma(\alpha)} e^{-s} ds = 1 - \tau, \end{aligned} \quad (6)$$

using the change of variables $s = g(\mathbf{w})v$. Equation (6) is solved by taking $r_0(\mathbf{w}) = C_\tau/g(\mathbf{w})$, with C_τ the solution to the equation $\int_0^{C_\tau} \frac{s^{\alpha-1}}{\Gamma(\alpha)} e^{-s} ds = \tau$. Since the gamma approximation is only asymptotically valid, we have in practice that $r_0(\mathbf{w}) \approx C_\tau/g(\mathbf{w})$ for τ close to 1. To depict unit level sets of the gauge function g , we plot points $\mathbf{x} = \mathbf{v}/g(\mathbf{v})$, where \mathbf{v} is a sequence of points covering the simplex \mathcal{S}_{d-1} . Consequently, we can compare $r_0(\mathbf{w})$ to g by plotting points $\mathbf{x} = \mathbf{v}r_0(\mathbf{v})/C_\tau$. However, since the gamma approximation is not exact, we instead scale each margin so that the coordinatewise supremum exactly equals one, by plotting

$$\mathbf{x} = \{v_1 r_0(\mathbf{v}) / \max[v_1 r_0(\mathbf{v})], \dots, v_d r_0(\mathbf{v}) / \max[v_d r_0(\mathbf{v})]\}. \quad (7)$$

We will use the observation that links $r_0(\mathbf{w})$ and $g(\mathbf{w})$ later to assist with model checking, but note that, combined with extension of additive quantile regression to higher dimensions, this presents a very interesting avenue for future work.

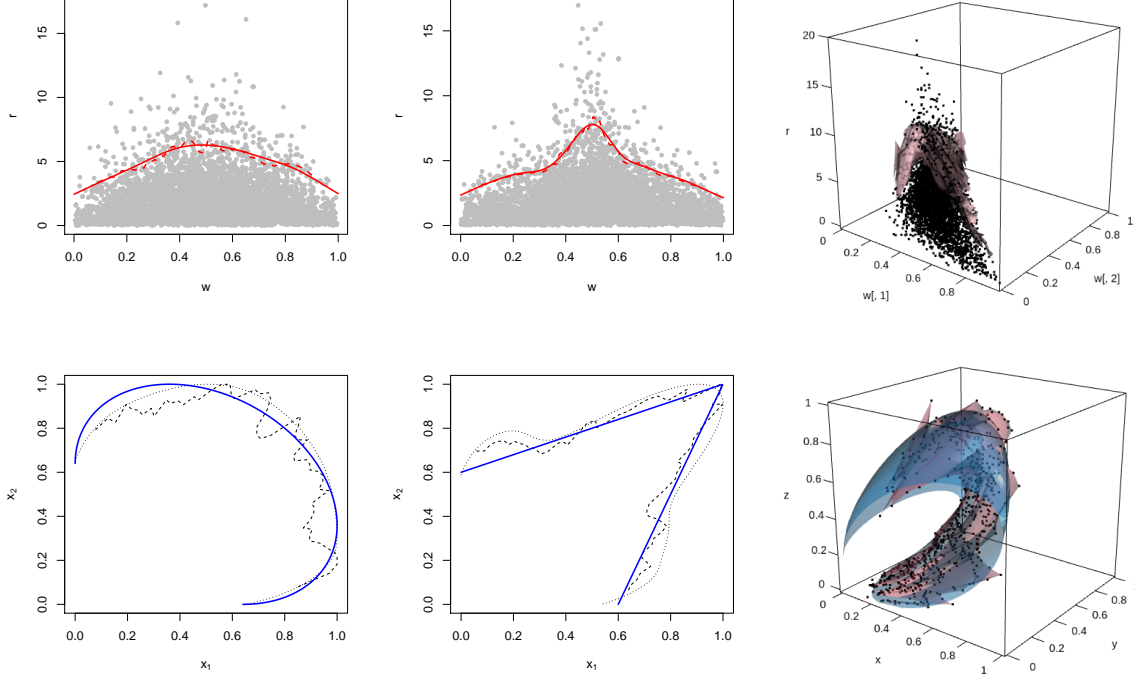


Figure 3: Top row: R against \mathbf{W} , with the estimated 0.95 quantile of $R \mid \mathbf{W} = \mathbf{w}$ in red. In the left and centre ($d = 2$) plots, solid lines represent the output from `qgam`, and dashed lines from rolling-windows quantiles. In the right ($d = 3$) plot, the surface is calculated through a rolling-windows technique. Bottom row: Plots of $\mathbf{x} = \mathbf{v}r_0(\mathbf{v})$, rescaled to lie in $[0, 1]^d$ as per equation (7). In the left and centre plots, dotted black lines represent the output from `qgam`, and dashed black lines from rolling-windows quantiles. The solid lines (blue) are the unit level sets of $g(\mathbf{x})$, with g the true gauge function. In the right plot, the jagged surface (red) comes from the rolling-windows technique, and the smooth surface (blue) is the unit level set of the true gauge function.

4.2 Likelihood

In order to fit model (5), we use likelihood-based inference. For n_0 independent observations of $R_i \mid [\mathbf{W}_i = \mathbf{w}_i, R_i > r_0(\mathbf{w}_i)]$, $i = 1, \dots, n_0$, we maximize the likelihood

$$L(\boldsymbol{\psi}) = \prod_{i=1}^{n_0} \frac{g(\mathbf{w}_i; \boldsymbol{\theta})^\alpha}{\Gamma(\alpha)} \frac{r_i^{\alpha-1} e^{-r_i g(\mathbf{w}_i; \boldsymbol{\theta})}}{\bar{F}[r_0(\mathbf{w}_i); \alpha, g(\mathbf{w}_i; \boldsymbol{\theta})]}, \quad (8)$$

where $\boldsymbol{\psi} = (\alpha, \boldsymbol{\theta})^\top$ and $\bar{F}[\cdot; \alpha, g(\mathbf{w}; \boldsymbol{\theta})]$ represents the gamma survival function with shape parameter α , and rate parameter $g(\mathbf{w}; \boldsymbol{\theta})$. Estimates of uncertainty in the maximum likelihood estimators may be obtained through the inverse Hessian matrix, subject to model validity and independence checks, or via the bootstrap. In practice, many datasets exhibit weak-to-moderate temporal dependence, so that while likelihood (8) may be used for parameter estimation (e.g., [Chandler and Bate, 2007](#)), block-bootstrap techniques will be preferable for estimation of uncertainty.

4.3 Gauge functions and model selection

4.3.1 Gauge functions from specific distributions

Key to a successful fit of model (5) via likelihood (8) are flexible parametrized forms of g that are able to capture a wide variety of limit set shapes. In Section 3, we detail various forms of gauge function that come from different underlying distributions, some of which are illustrated in Section A of the supplement. Further forms can also be found in [Nolde and Wadsworth \(2022\)](#). Any of these parametric forms could be fitted as a candidate model, and standard model-selection techniques, such as information criteria, used to establish a best choice; we will demonstrate this in our simulation study of Section 5.

A key attraction of our new approach to inference for multivariate extremes is the ability to capture the complex dependence structures that arise when different sub-groups of variables can potentially be co-extreme while the others are small. Under multivariate regular variation, this corresponds to the spectral measure H placing mass on sets \mathbb{B}_C as described in Section 1. In order to capture these scenarios, we consider the gauge function corresponding to the asymmetric logistic distribution ([Tawn, 1990](#)), which can place mass on any valid combination of sets \mathbb{B}_C . The full expression for this involves minimization over several components, and is given in Section D of the supplement. Figure 4 depicts some of the potential limit sets arising from this structure when $d = 3$.

4.3.2 Additively mixing gauge functions

The gauge functions described in Section 3 provide a starting point for inference on model (5), but may not always be flexible enough to capture the structures of observed data. We now consider how

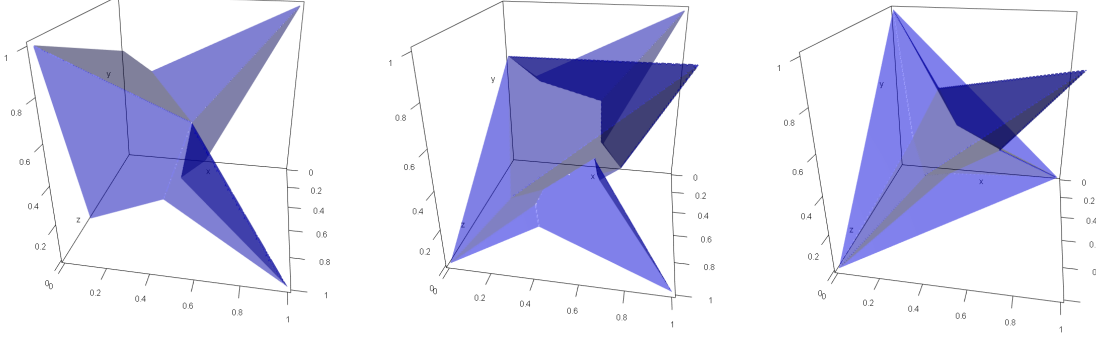


Figure 4: Example limit sets G (area between surface and planes $x_j = 0$) from the asymmetric logistic distribution. Left: mass of H on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{1,3\}}, \mathbb{B}_{\{2,3\}}$ with parameters $\gamma_{\{1,2\}} = 0.5$, $\gamma_{\{1,3\}} = 0.2$, $\gamma_{\{2,3\}} = 0.7$. Centre: mass on $\mathbb{B}_{\{3\}}, \mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{1,3\}}, \mathbb{B}_{\{1,2,3\}}$ with $\gamma_{\{1,2\}} = 0.5$, $\gamma_{\{1,3\}} = 0.5$, $\gamma_{\{1,2,3\}} = 0.7$. Right: mass on $\mathbb{B}_{\{1\}}, \mathbb{B}_{\{2\}}, \mathbb{B}_{\{3\}}, \mathbb{B}_{\{1,2,3\}}$ with $\gamma_{\{1,2,3\}} = 0.5$.

to mix gauge functions to generate more flexible models. As mentioned in Section 1, the limit sets G for data with exponential margins have coordinatewise supremum equal to $(1, \dots, 1)$; equivalently, the one-dimensional marginal gauge functions are $g_{\{j\}}(x_j) = x_j$. Each form of g given in Section 3 satisfies this constraint, and we require any scheme for mixing gauge functions to also satisfy this, since they will be applied to data in exponential margins.

A simple way to mix that retains the marginal condition $g_{\{j\}}(x_j) = x_j$ is via minimization: $g(\mathbf{x}) = \min[g^{[1]}(\mathbf{x}), \dots, g^{[m]}(\mathbf{x})]$, for $g^{[1]}, \dots, g^{[m]}$ each satisfying this marginal condition. The resulting gauge function is the one that would correspond to a mixture density $f_{\mathbf{X}}(\mathbf{x}) = \sum_{k=1}^m \pi_k f_{\mathbf{X}}^{[k]}(\mathbf{x})$ with $\sum_{k=1}^m \pi_k = 1$ and $\pi_k \in (0, 1)$ for each k ; the mixture weights do not appear in g because as $n \rightarrow \infty$ there would be infinitely many points in the sample cloud from each mixture component. However, such an approach has the effect of retaining the most protruding part of each limit set and may not yield the most realistic shapes; some examples are given in Section E of the supplement. Instead we focus on additive mixing, defining

$$\tilde{g}(\mathbf{x}) = a_1 g^{[1]}(\mathbf{x}) + \dots + a_{m-1} g^{[m-1]}(\mathbf{x}) + g^{[m]}(\mathbf{x}), \quad a_1, \dots, a_{m-1} > 0. \quad (9)$$

The resulting function is denoted by \tilde{g} as in general it will not satisfy the marginal condition, and will need to be rescaled to do so. Suppose that the coordinatewise supremum of the set $\tilde{G} = \{\mathbf{x} : \tilde{g}(\mathbf{x}) \leq 1\}$ is $\tilde{\mathbf{c}} = (\tilde{c}_1, \dots, \tilde{c}_m)$. Then the rescaled gauge function $g(\mathbf{x}) = \tilde{g}(\tilde{c}_1 x_1, \dots, \tilde{c}_d x_d)$ satisfies $g_{\{j\}}(x_j) = x_j$. Some examples of limit sets from additively mixed functions are depicted in Figure 5. Interestingly, we observe for $d = 2$ that this process is able to interpolate between limit sets for which $g(1, 1) = 1$ and have a “pointy” shape, to those with $g(1, 1) < 1$ and are described by Balkema and Nolde (2012) as “blunt”. The former arise for dependence structures representing

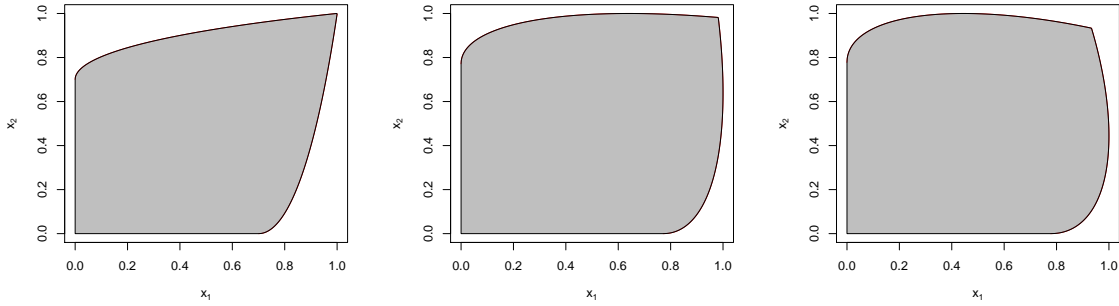


Figure 5: Examples of limit sets (grey shaded region) from additively mixed gauge functions according to equation (9), with subsequent rescaling. In each case the component gauge functions are $g^{[1]}(x_1, x_2) = [x_1 + x_2 - 2\rho(x_1x_2)^{1/2}]/(1 - \rho^2)$ with $\rho = 0.5$ and $g^{[2]}(x_1, x_2) = (x_1 + x_2)/\gamma + (1 - 2/\gamma)\min(x_1, x_2)$ with $\gamma = 0.5$; these gauges correspond to the Gaussian and logistic max-stable distributions, respectively. From left to right, the weights are $a_1 = 1, 2, 3$.

joint extremes (H places mass only on $\mathbb{B}_{\{1,2\}}$), while the latter arise for those representing separate extremes (H places mass only on $\mathbb{B}_{\{1\}}$ and $\mathbb{B}_{\{2\}}$). Figures in Section E of the the supplement also show that for $d = 3$ we retain the ability to move between “pointy” limit set shapes representing joint occurrence of extremes in some components and “blunt” shapes representing separate extremes. We focus in the figures only on the case $m = 2$, and leave theoretical study of this phenomenon for any m to future work.

Note that when using additive mixing, the component gauge functions $g^{[k]}(\mathbf{x})$ need not satisfy the marginal condition $g_{\{j\}}(x_j) = x_j$ due to the rescaling. This allows, for example, one to include the Gaussian gauge function $g(\mathbf{x}) = (\mathbf{x}^{1/2})^\top \Sigma^{-1} \mathbf{x}^{1/2}$ when Σ has negative entries, and increases the flexibility of this approach. In practice, we use numerical methods to find the vector $\tilde{\mathbf{c}}$ for rescaling, by finding the coordinatewise supremum of \tilde{G} on a grid.

4.4 Model checking

We propose checking the fitted model from likelihood (8) via probability-probability (PP) plots. The fitted distribution function (df) of the truncated gamma model is

$$\hat{F}_{\text{tg}}[r \mid \mathbf{w}, r_0(\mathbf{w})] := \Pr[R \leq r \mid \mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})] = 1 - \frac{\bar{F}[r; \hat{\alpha}, g(\mathbf{w}; \hat{\boldsymbol{\theta}})]}{\bar{F}[r_0(\mathbf{w}); \hat{\alpha}, g(\mathbf{w}; \hat{\boldsymbol{\theta}})]},$$

with \bar{F} as in likelihood (8), and $\hat{\alpha}, \hat{\boldsymbol{\theta}}$ representing the maximum likelihood estimates of the parameters. The PP plot for n_0 observations with $R_i > r_0(\mathbf{w}_i)$ is the set of points: $[i/(n_0 + 1), u_{(n_0 - i + 1)}]$, where $u_i = \hat{F}_{\text{tg}}[r_i; \mathbf{w}_i, r_0(\mathbf{w}_i)]$, and $u_{(1)} \geq u_{(2)} \geq \dots \geq u_{(n_0)}$ represent the ordered sample of u_i .

This diagnostic will be demonstrated in Section 6.

Comparison of the “empirical” estimate of the gauge function $\hat{g}(\mathbf{w}) \approx \hat{C}/r_0(\mathbf{w})$, as outlined in Section 4.1, provides another check on the form of the fitted model. As was seen in Section 4.1, while we do not expect perfect correspondence between $\hat{g}(\mathbf{w})$ and $g(\mathbf{w}; \hat{\theta})$, we can expect to see broad similarities in shape. Again we use this in Section 6.

4.5 Prediction

A key aspect of our proposed geometric framework for statistical inference is that we can use simulation from the fitted model to estimate probabilities of lying in extreme regions, enabling extrapolation outside the range of the observed data. Up to this point, we have focused on the conditional distribution of $R \mid [\mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})]$. In order to perform extrapolation and estimate multivariate tail probabilities, we need realizations of the distribution of \mathbf{X} in some suitably extreme region. Notationally it is helpful to introduce an alternative radial variable, $R' = R/r_0(\mathbf{W})$, so that $\mathbf{X} = R\mathbf{W} = R'r_0(\mathbf{W})\mathbf{W}$. Given a particular value of $\mathbf{W} = \mathbf{w}$, our extreme region to date has been $\{R > r_0(\mathbf{w})\}$. Now considering our extreme region across all (R, \mathbf{W}) values, this corresponds to $\{R' > 1\}$. In Figure 3, all points above the red line / surface in the top row are those with $\{R' > 1\}$.

We focus initially on simulating an arbitrary number of points satisfying the conditioning event $\{R' > 1\}$, and discuss below adaptations for simulating above higher thresholds. To get draws from the distribution of $\mathbf{X} \mid R' > 1$, we multiply simulations from two components:

- (i) Draw \mathbf{w}^* from the distribution of $\mathbf{W} \mid R' > 1$;
- (ii) Conditional upon \mathbf{w}^* , draw r^* from the distribution of $R \mid [\mathbf{W} = \mathbf{w}^*, R > r_0(\mathbf{w}^*)]$.

The sampled value is then $\mathbf{x}^* = r^*\mathbf{w}^*$. The second of these steps is a simple case of simulating from the fitted truncated gamma distribution, which can be done via the inverse probability integral transform. For the first of these, we may either resample from the empirical distribution of $\mathbf{W} \mid R' > 1$, or we could fit a parametric model to such samples and simulate from this. We opt for the former in this work, and note the latter as a potential line of future investigation. Figure 6 shows 5000 draws simulated from $\mathbf{X} \mid R' > 1$, based on a model fitted to 2500 data points.

To estimate the probability of lying in extreme sets, we exploit the simple equation

$$\Pr(\mathbf{X} \in B) = \Pr(\mathbf{X} \in B \mid R' > 1) \Pr(R' > 1), \quad (10)$$

for any set B lying entirely within the region $\{\mathbf{x} \in \mathbb{R}_+^d : \sum_{j=1}^d x_j > r_0(\mathbf{x}/\sum_{j=1}^d x_j)\}$; some examples are given in Figure 6. The first probability on the right-hand side of (10) can be estimated empirically from the simulated draws. The second probability may be estimated from the dataset

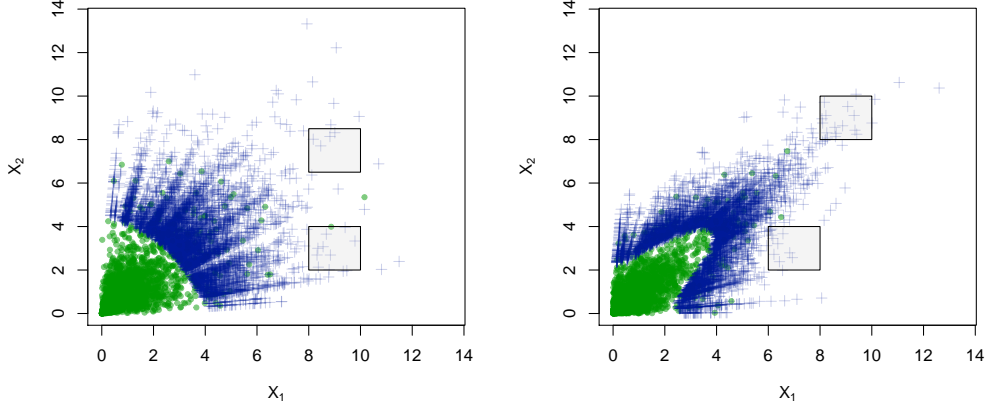


Figure 6: Example of 5000 points simulated from $\mathbf{X} \mid R' > 1$, depicted by crosses (blue), for the inverted logistic (left) and logistic (right) distributions. Models with the true gauge were fitted to threshold exceedances of 2500 data points, with all data depicted by dots (green). Light grey squares represent potential sets B in equation (10).

as the proportion of points R' exceeding 1. When quantile regression at level τ has been used to estimate $r_0(\mathbf{w})$, we expect the proportion of points above the threshold to be near $1 - \tau$.

The fact we can simulate an arbitrary number of points from our model with the condition $\{R' > 1\}$ means that in principle we can extrapolate quite a way beyond the observed data. Nonetheless, such an approach may be computationally demanding for very extreme sets that require a large number of simulations. We consider now how to simulate given the condition $\{R' > k\}$, with $k > 1$; results will be illustrated in Section 5.

Simulation from the truncated gamma distribution of $R \mid [\mathbf{W} = \mathbf{w}, R > kr_0(\mathbf{w})]$ is again straightforward, but simulation from the distribution of angles $\mathbf{W} \mid R' > k$ is more challenging if k is sufficiently high that there are few or no empirical samples available. However, we have the relation

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{w} \mid R' > k) &= \frac{\int_k^\infty f_{R', \mathbf{W}}(r', \mathbf{w} \mid R' > 1) dr'}{\int_{\mathcal{S}_{d-1}} \int_k^\infty f_{R', \mathbf{W}}(r', \mathbf{v} \mid R' > 1) dr' d\mathbf{v}} \\ &= \frac{f_{\mathbf{W}}(\mathbf{w} \mid R' > 1) \int_k^\infty f_{R' \mid \mathbf{W}}(r' \mid \mathbf{w}, R' > 1) dr'}{\int_{\mathcal{S}_{d-1}} \int_k^\infty f_{R' \mid \mathbf{W}}(r' \mid \mathbf{v}, R' > 1) f_{\mathbf{W}}(\mathbf{v} \mid R' > 1) dr' d\mathbf{v}}, \end{aligned} \quad (11)$$

where $f_{\mathbf{U}}(\cdot \mid V > v)$ denotes the density of a random vector $\mathbf{U} \mid V > v$. Note that

$$\int_k^\infty f_{R' \mid \mathbf{W}}(r' \mid \mathbf{w}, R' > 1) dr' = \int_{kr_0(\mathbf{w})}^\infty f_{R \mid \mathbf{W}}(r \mid \mathbf{w}, R > r_0(\mathbf{w})) dr,$$

so that under the truncated gamma approximation (5) for $R \mid [\mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})]$, we have the

proportionality statement

$$f_{\mathbf{W}}(\mathbf{w} \mid R' > k) \propto f_{\mathbf{W}}(\mathbf{w} \mid R' > 1) \frac{\bar{F}[kr_0(\mathbf{w}); \alpha, g(\mathbf{w})]}{\bar{F}[r_0(\mathbf{w}); \alpha, g(\mathbf{w})]}. \quad (12)$$

The ratio of gamma survival functions in (12) can therefore be used as importance weights to derive an approximate sample from the distribution of $\mathbf{W} \mid R' > k$, using a sample from the distribution of $\mathbf{W} \mid R' > 1$.

Finally, to estimate $\Pr(R' > k)$, so that we can calculate extreme probabilities as in equation (10), note that the constant of proportionality in (12) is $\Pr(R' > k \mid R' > 1)$, from the denominator of equation (11). An estimate of this is therefore

$$\widehat{\Pr}(R' > k \mid R' > 1) = \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{\bar{F}[kr_0(\mathbf{w}_i); \alpha, g(\mathbf{w}_i)]}{\bar{F}[r_0(\mathbf{w}_i); \alpha, g(\mathbf{w}_i)]},$$

where $\mathbf{w}_i, i = 1, \dots, n_0$ are the angles corresponding to the values for which $R' > 1$. Lastly, $\widehat{\Pr}(R' > k) = \widehat{\Pr}(R' > k \mid R' > 1) \widehat{\Pr}(R' > 1)$, where $\widehat{\Pr}(R' > 1)$ is estimated empirically, as previously. We note that another alternative to this procedure is to fit the generalized Pareto distribution to $R' \mid R' > 1$ and use this fitted model to estimate $\Pr(R' > k \mid R' > 1)$. Our investigation into this found that both options perform similarly for relatively small k , but the generalized Pareto model introduces extra uncertainty for larger k , and so we stick to the first approach in Section 5.

In our experience we have found that estimates of $\Pr(\mathbf{X} \in B)$ are relatively insensitive to the precise choice of k , provided both that k is large enough to ensure that several sample points lie in B , and that $B \subset \{\mathbf{x} \in \mathbb{R}_+^d : \sum_{j=1}^d x_j > kr_0(\mathbf{x} / \sum_{j=1}^d x_j)\}$, as is required for the analogue of equation (10) to hold. The simplicity of checking this latter condition depends on the shape of B and of $r_0(\mathbf{w})$, but it is easy to check visually for $d = 2$, and it may crudely be checked by ensuring that $k < \sum_{j=1}^d \tilde{x}_{l,j} / r_0(\tilde{\mathbf{x}}_l / \sum_{j=1}^d \tilde{x}_{l,j})$ for a sample of points $\tilde{\mathbf{x}}_l, l = 1, \dots, m$, along the boundary of B . See the Supplement Section F for further discussion and summary of our investigation. We recommend taking an intermediate k that is slightly smaller than the maximum for which this series of m inequalities holds, to safeguard against the crudeness of this check.

4.6 Summary of inference and prediction procedures

For convenience, we briefly summarize the procedures for inference and prediction via the geometric framework:

1. Determine a high threshold $r_0(\mathbf{w})$ of the distribution of $R \mid \mathbf{W} = \mathbf{w}$ for all $\mathbf{w} \in \mathcal{S}_{d-1}$ using either additive quantile regression or a rolling-windows approach.

2. Select a set of candidate parametric gauge functions $g(\cdot; \boldsymbol{\theta})$ and for each one fit the truncated gamma likelihood (8) to the n_0 values of $R \mid [\mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})]$.
3. Compare model fits using selection criteria such as the Akaike or Bayesian information criterion.
4. Use diagnostics such as the PP plot and comparison with the empirically-estimated gauge function to confirm acceptable fit of the best model(s).
5. Letting $R' = R/r_0(\mathbf{W})$, simulate new realizations from the distribution of $\mathbf{X} \mid R' > 1$ by drawing from the empirical distribution of $\mathbf{W} \mid R' > 1$ and multiplying by draws of $R \mid [\mathbf{W} = \mathbf{w}, R > r_0(\mathbf{w})]$ from the fitted truncated gamma distribution. If required, adapt these steps to simulate from the distribution of $\mathbf{X} \mid R' > k$ with $k > 1$.
6. Estimate $\Pr(\mathbf{X} \in B)$ using equation (10), or suitable adaptation if $R' > k$.

5 Simulation study

We now demonstrate the performance of our methods against existing approaches for analyzing multivariate extremes. Our focus lies on estimation of probabilities $\Pr(\mathbf{X} \in B)$ for three sets B that lie in different parts of the region where \mathbf{X} may be considered extreme.

We begin with the bivariate case, which is well-established and understood, demonstrating that our methodology gives estimates with low bias in each situation, performing competitively with other methods across a range of scenarios. Specifically, we compare with estimation methodology based on multivariate regular variation (MRV), hidden regular variation (Ledford and Tawn, 1997) (HRV) and the conditional extreme value model (CE) of Heffernan and Tawn (2004). The simplest approach to implementing MRV methodology is to use the approximation $\Pr(\mathbf{X} \in v + B') \approx e^{-v} \Pr(\mathbf{X} \in B')$, where we take as the set of interest $B = v + B'$, and B' is extreme, but in the range of the data so can be estimated empirically. This is a nonparametric implementation, but parametric assumptions are possible as well. Specifically we can also assume that equation (1) holds at finite levels and choose a parametric form for the angular measure H . We adopt both techniques below. HRV is a refinement of MRV that allows for situations where the spectral measure H places no mass on $\mathbb{B}_{\{1, \dots, d\}}$. Implementation of this methodology relies on exploiting the relation $\Pr(\mathbf{X} \in v + B') \approx e^{-v/\eta} \Pr(\mathbf{X} \in B')$, where $\eta \in (0, 1]$ is the residual tail dependence coefficient; this is estimated using the Hill estimator (Ledford and Tawn, 1997). Parametric models based on HRV exist (Ramos and Ledford, 2009), but are generally poorly-justified since the so-called “hidden angular measure” is often not a finite measure over the unit simplex; we therefore do not consider

these here. Like MRV however, the asymptotics of HRV are suited only to extrapolating into regions where all variables are large simultaneously. Implementation of the CE methodology to model the distribution of $\mathbf{X} \mid X_j > u$, for large u , follows the original approach suggested in [Heffernan and Tawn \(2004\)](#), adapted to exponential margins. Probability estimation in this case is performed by simulating from the distribution of $\mathbf{X} \mid X_j > u', u' \geq u$. Following the bivariate case, we move on to the more difficult case of $d = 3$, and show that we can substantially outperform the CE model in this setting, which is the only other viable approach for providing an estimate of the probabilities of interest.

5.1 Dimension $d = 2$

For the bivariate case, we perform estimation based on 5000 datapoints simulated from four different dependence structures: (I) logistic distribution with parameter $\gamma = 0.4$; (II) Gaussian distribution with $\rho = 0.8$; (III) inverted logistic distribution with $\gamma = 0.7$; (IV) logistic distribution with $\gamma = 0.8$. Distributions (I) and (IV) represent moderately strong and weak logistic dependence structures, respectively. In [Section G](#) of the supplement we show examples of the four datasets, and three sets of interest $B_1 = (10, 12) \times (10, 12)$, $B_2 = (10, 12) \times (6, 8)$, and $B_3 = (10, 12) \times (2, 4)$.

In each case we fit model [\(5\)](#) to the data using four different gauge functions: those corresponding to the unique distributions (I)–(III), where the parameter is to be estimated from the data, and the function $g(\mathbf{x}; \theta) = \max[(x_1 - x_2)/\theta, (x_2 - x_1)/\theta, (x_1 + x_2)/(2 - \theta)]$. We select the model that yields the lowest value of the Akaike information criterion (AIC) for the prediction step, thereby avoiding using knowledge of the true data-generating process. Recall that before fitting model [\(5\)](#), we need to calculate a high threshold $r_0(\mathbf{w})$. In [Section 4.1](#), we described using either additive quantile regression or a rolling-windows quantile calculation for this. We used both techniques in the simulation study, setting $\tau = 0.95$, finding relatively little difference in the performance of the resulting inference, particularly in comparison to differences across extreme-value methodologies. Therefore, to keep presentation focused, we detail only the results where $r_0(\mathbf{w})$ was found using the simpler rolling-windows quantile method. Although our focus is on extreme probability estimation, we also display (non-)parametric estimates of G , obtained via $\hat{g}(\mathbf{w})$ and $g(\mathbf{w}; \hat{\theta})$, in [Section G](#).

For the parametric MRV approach, we employ a similar strategy to our geometric approach. After transforming to radial-angular coordinates $\|\mathbf{X}_P\|$ and $\mathbf{X}_P/\|\mathbf{X}_P\|_1$ from Pareto margins, we take all angles for which the corresponding radius exceeds the 0.95 quantile of radii, and fit a parametric form for the density of H via maximum likelihood. We choose between five parametric models for H using AIC, where the true logistic density for distributions (I) and (IV) is among the choices. Probabilities are estimated using numerical integration over B using the fitted model for

angles, combined with Pareto density for radii.

Figure 7 displays boxplots of the estimated probabilities for 200 repetitions across different methodologies:

- the geometric approach with different k (G1, G2);
- the conditional extremes model with different simulation thresholds (CE1, CE2);
- the nonparametric hidden regular variation approach (HRV), and
- nonparametric and parametric multivariate regular variation approaches (MRV, MRV2);

see the caption for further details. For distribution (I), all methods estimate $\Pr(\mathbf{X} \in B_1)$ with little bias; the smallest variance is attributed to the MRV approaches, which is as expected since we are looking at a distribution where H places mass on $\mathbb{B}_{\{1,2\}}$ and estimating a probability in the joint tail. The geometric approach and CE estimate $\Pr(\mathbf{X} \in B_2)$ relatively well, with the smallest variance attributable to the geometric approach based on $\mathbf{X} \mid R' > k$ for suitable $k > 1$. HRV and MRV start to exhibit some bias because B_2 lies outside the joint tail region. For $\Pr(\mathbf{X} \in B_3)$, all estimates based on the nonparametric HRV and MRV approaches are equal to zero. For the geometric approach, we are able to estimate this probability well when selecting a suitable k . Specifically, in each repetition, we select one of the largest values of k such that $B_3 \subset \{\mathbf{x} : (x_1 + x_2) > kr_0[\mathbf{x}/(x_1 + x_2)]\}$. This results in all probabilities having a non-zero estimate, compared to 0% for nonparametric HRV/MRV, and 4.5% for CE (at each of two thresholds). This probability can be estimated as non-zero by parametric MRV, but with a little bias. A boxplot of this case is included in the left panel of Figure 8.

Distribution (IV) also represents the case where H places mass on $\mathbb{B}_{\{1,2\}}$, yet interestingly, MRV gives biased estimates in for all probabilities. This is likely due to the practical rate of convergence to the limiting angular measure H being slower under this weaker dependence scenario. Indeed we see differing estimates from the two MRV approaches, which are based on different effective “thresholds” for defining extremes. MRV changes from appreciably over-estimating the probabilities $\Pr(\mathbf{X} \in B_1)$ and $\Pr(\mathbf{X} \in B_2)$ to hugely under-estimating $\Pr(\mathbf{X} \in B_3)$. The geometric approach suggests a small under-estimation of $\Pr(\mathbf{X} \in B_1)$ and $\Pr(\mathbf{X} \in B_2)$ and good performance for $\Pr(\mathbf{X} \in B_3)$. CE shifts from large under-estimation to over-estimation moving from $\Pr(\mathbf{X} \in B_1)$ to $\Pr(\mathbf{X} \in B_3)$.

For distributions (II) and (III), the geometric approach and CE exhibit quite similar performance, although CE has a smaller variance for estimates of $\Pr(\mathbf{X} \in B_2)$ under distribution (II), and of $\Pr(\mathbf{X} \in B_3)$ under distribution (III). MRV is not an appropriate method for these distributions and always performs badly; HRV is appropriate in the joint tail, where it exhibits similar performance to other methods for (II) and better performance for (III), while it leads to poor estimates in other regions. Additional boxplots in the right panel of Figure 8 display more detailed

information for the estimates of $\Pr(\mathbf{X} \in B_1)$ under distribution (III). As described for distribution (I), we also used a suitable $k > 1$ for estimating this probability. The geometric approach outperforms CE in this case. This is because, using an appropriate k , we are able to simulate points to generate non-zero estimates of the probabilities (93% and 100% of estimates are positive for the two thresholds shown). In contrast, only 45.5% and 46% of estimates are positive for CE.

5.2 Dimension $d = 3$

We again perform estimation based on 5000 data points from three different data structures: (I) asymmetric logistic distribution, for which the spectral measure H places mass on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{1,3\}}, \mathbb{B}_{\{2,3\}}$, with parameters $\gamma_{\{1,2\}} = \gamma_{\{1,3\}} = \gamma_{\{2,3\}} = 0.4$; (II) asymmetric logistic distribution where H places mass on $\mathbb{B}_{\{1\}}, \mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{2,3\}}$ and with parameters $\gamma_{\{1,2\}} = \gamma_{\{2,3\}} = 0.4$; (III) distribution constructed by taking an inverted Clayton copula with parameter 2 for (X_1, X_2) , with $X_3 \mid X_2 = x_2$ drawn from an inverted logistic dependence structure with parameter 0.5. Such a distribution is in the domain of attraction of a spectral measure H placing mass on $\mathbb{B}_{\{1,2\}}, \mathbb{B}_{\{3\}}$. In Section G of the supplement we display examples of the three datasets along with sets of interest $B_1 = (8, 10) \times (8, 10) \times (0.01, 3)$, $B_2 = (8, 10) \times (5, 7) \times (0.01, 3)$ and $(8, 10) \times (2, 4) \times (0.01, 3)$.

For the $d = 3$ case we consider only two methodologies: the geometric approach and CE, as HRV/MRV only perform well when considering sets B where all variables are of a similar magnitude, and the sets that we are considering are all small in x_3 . Moreover for MRV we require mass on $\mathbb{B}_{\{1,2,3\}}$ for good performance of this method.

For the geometric approach we fit model (5) to the data after identifying potential suitable forms for the gauge function g . For this initial step, we calculate the coefficients $\tau_C(\delta)$, and associated estimates of the probability of mass on \mathbb{B}_C as in Simpson et al. (2020), for $\delta = 0.4, 0.5, 0.6$ and $C \subseteq \{1, 2, 3\}$. These estimates help to identify potential faces of the simplex on which the limiting spectral measure H places mass, and hence a suitable structure for the form of the gauge function. Specifically, where they exist, the coefficients $\tau_C(\delta) \in (0, 1]$ should be equal to 1 if H places mass on \mathbb{B}_C , for all values of the tuning parameter $\delta \in [\delta^*, 1]$ and some $\delta^* \in [0, 1]$. However, because of difficulties in estimating these coefficients precisely in the presence of nuisance parameters, Simpson et al. (2020) use them as part of a broader strategy to estimate of the probability of mass on \mathbb{B}_C . If all estimates for the three values of δ suggest the same extremal dependence structure in terms of where H places mass, then a single model is fitted, where the gauge function corresponds to that of the asymmetric logistic distribution for the identified structure. Otherwise, up to three different models are fitted, and the model with the lowest AIC is selected. We note that, for distributions (I) and (II), this means that we have the potential to fit the correct model form to the data, subject

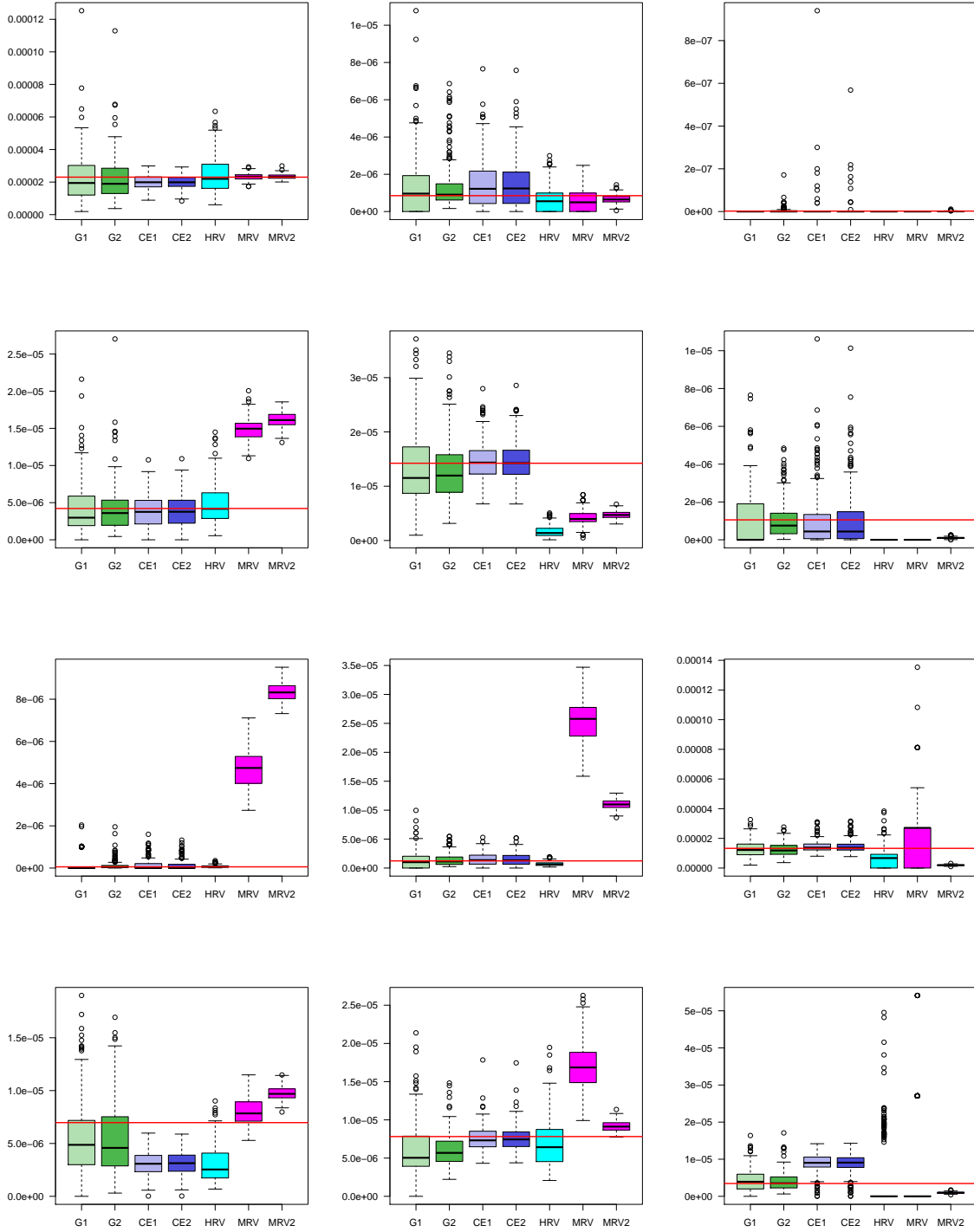


Figure 7: Boxplots of the 200 estimated probabilities for $d = 2$. From left to right columns represent $\Pr(\mathbf{X} \in B_1)$, $\Pr(\mathbf{X} \in B_2)$, $\Pr(\mathbf{X} \in B_3)$, respectively. From top to bottom, datasets are (I), (II), (III), (IV) respectively. Green boxplots, labelled G1, G2, give results from our geometric approach: G1 is calculated from $\mathbf{X} \mid R' > 1$; G2 is calculated from $\mathbf{X} \mid R' > k$, where k is determined as the maximum value such that all sets B_1, B_2, B_3 lie in the region $\{\mathbf{x} : x_1 + x_2 > kr_0(\mathbf{x}/(x_1 + x_2))\}$. Dark blue boxplots, labelled CE1, CE2 give results from the conditional extremes model: CE1 is calculated from $\mathbf{X} \mid X_1 > 6.9$; CE2 is calculated from $\mathbf{X} \mid X_1 > 10$. Turquoise boxplots, labelled HRV, give results from hidden regular variation²³ methodology; purple boxplots, labelled MRV and MRV2, represent nonparametric and parametric multivariate regular variation, respectively. True values of the probabilities are indicated by horizontal red lines.

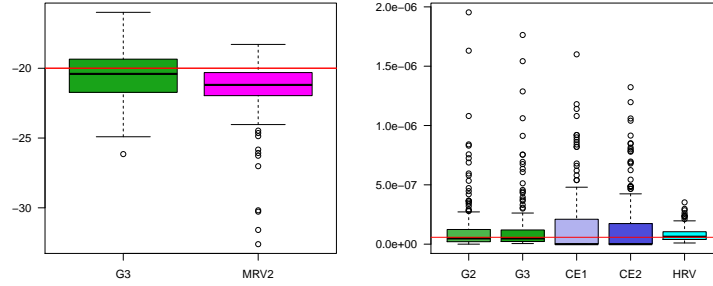


Figure 8: Left: boxplot of log estimates of $\Pr(\mathbf{X} \in B_3)$ for distribution (I), $d = 2$, using the geometric approach at a high threshold as described in the text (G3), and parametric MRV (MRV2). Right: estimates of $\Pr(\mathbf{X} \in B_1)$ for distribution (III), $d = 2$, using two different thresholds for the geometric approach, labelled G2, G3 (green). Estimates for two different thresholds from the conditional approach are labelled CE1, CE2 (dark blue), and from hidden regular variation, HRV (turquoise).

to its identification via the [Simpson et al. \(2020\)](#) methodology, although for distribution (III), we always have a misspecified model.

Figure 9 displays boxplots of the estimated probabilities across 200 repetitions using the two methods. In most cases, the geometric approach exhibits relatively low bias, particularly in comparison to CE, which is typically biased down for $\Pr(\mathbf{X} \in B_1)$ and up for $\Pr(\mathbf{X} \in B_2), \Pr(\mathbf{X} \in B_3)$. In conditional extreme value modelling, dependence structures are defined pairwise, so while any pair of variables (X_i, X_j) can theoretically have mass on $\mathbb{B}_{\{i,j\}}$ or $\mathbb{B}_{\{i\}}$ and $\mathbb{B}_{\{j\}}$, the methodology cannot usually capture more complex higher-order structures well. The structure of distribution (III) is the simplest, with only variables X_1, X_2 exhibiting simultaneous extremes, and CE is correspondingly more successful in this case. For $\Pr(\mathbf{X} \in B_3)$ and distributions (I) and (III), additional boxplots are provided in Section G of the supplement. These demonstrate that the geometric approach labelled G3 provides the best estimate in both cases, but underestimates the probability. In contrast we can see from Figure 9 that estimates of this probability for distribution (I) are biased strongly upwards for CE, while for distribution (III) only 5.5% of estimates for CE are positive at either threshold.

6 Data analyses

We use our new modelling approach to analyze two multivariate environmental datasets. The first is wave data from Newlyn, UK, included because of its extensive previous analysis in the literature.

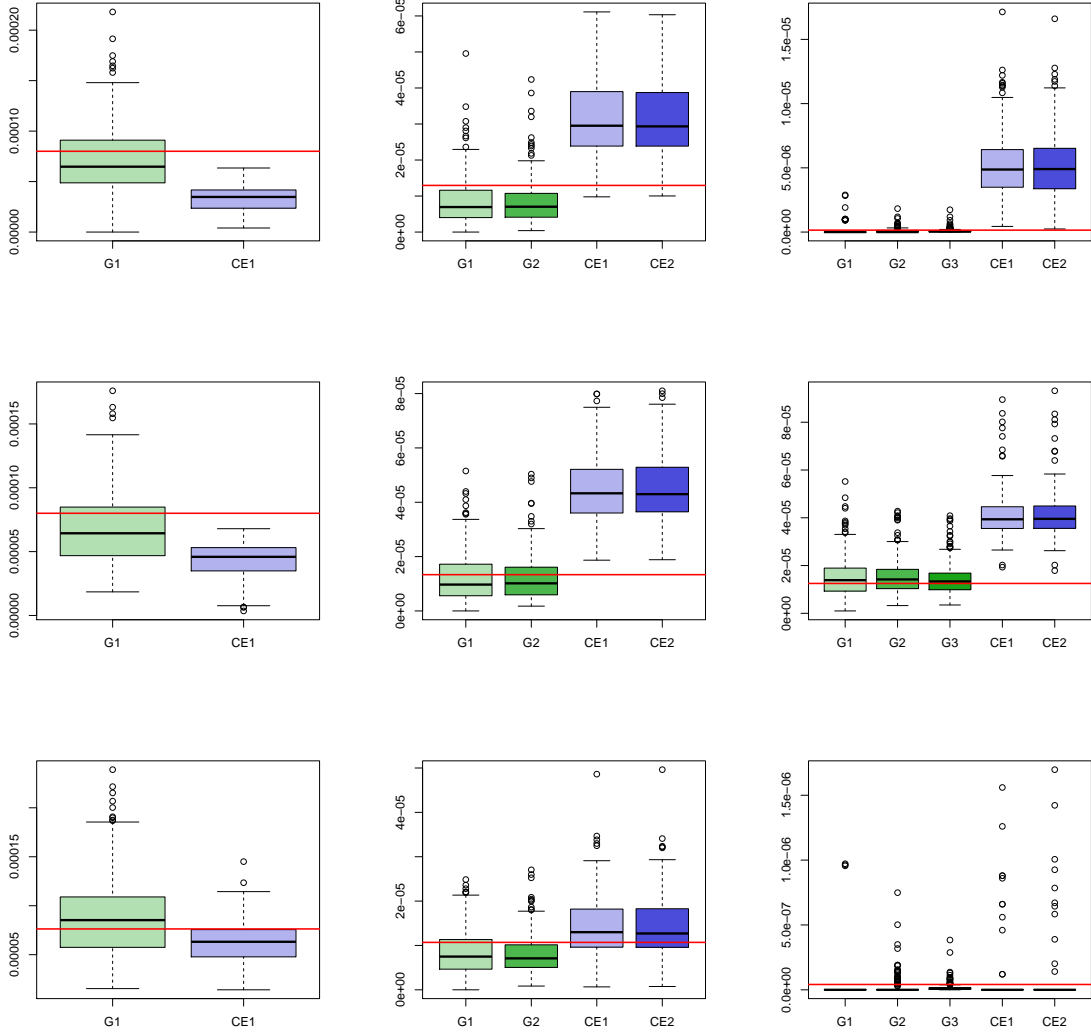


Figure 9: Boxplots of the 200 estimated probabilities for $d = 3$. From left to right columns represent $\Pr(\mathbf{X} \in B_1), \Pr(\mathbf{X} \in B_2), \Pr(\mathbf{X} \in B_3)$, respectively. From top to bottom, datasets are (I), (II), (III), respectively. Green boxplots, labelled G1, G2, G3 give results from the geometric approach: G1 is calculated from $\mathbf{X} \mid R' > 1$; G2 and G3 are calculated from $\mathbf{X} \mid R' > k_j$, $j = 1, 2$, where k_j is determined as a large value such that the sets B_2 or B_3 lie in the region $\{\mathbf{x} : x_1 + x_2 + x_3 > k_j r_0[\mathbf{x}/(x_1 + x_2 + x_3)]\}$. Dark blue boxplots, labelled CE1, CE2 give results from the conditional extremes model: CE1 is calculated from $\mathbf{X} \mid X_1 > 6.9$; CE2 is calculated from $\mathbf{X} \mid X_1 > 8$. True values of the probabilities are indicated by horizontal red lines.

The second is a set of river flow data from [Simpson et al. \(2020\)](#).

6.1 Newlyn wave data

This dataset of 2894 measurements of wave height (metres), surge (metres) and period (seconds), denoted here as (X_H, X_S, X_P) , was originally analyzed in [Coles and Tawn \(1994\)](#) using a model that assumed multivariate regular variation with all mass of the spectral measure on $\mathbb{B}_{\{H,S,P\}}$. The full trivariate dataset has subsequently been analyzed in [Bortot et al. \(2000\)](#), who assumed a censored multivariate Gaussian model, and [Coles and Pauli \(2002\)](#), whose model was able to accommodate the situation where the spectral measure places mass on some faces of the simplex, but was otherwise quite restrictive.

The first step is to transform each marginal to exponential, which is done using a semi-parametric estimate of the distribution function for each variable X_j :

$$\hat{F}_j(x) = \begin{cases} \tilde{F}_j(x), & x \leq u_j, \\ 1 - \phi_{u,j} [1 + \xi_j(x - u_j)/\sigma_j]_+^{-1/\xi_j}, & x > u, \end{cases} \quad (13)$$

where \tilde{F}_j is the empirical df, u_j is a high threshold, $\phi_{u,j} = \Pr(X_j > u_j)$, and the form above u_j is the generalized Pareto distribution with scale $\sigma_j > 0$ and shape ξ_j . We take the thresholds u_H, u_S and u_P to be the 95% quantiles of the respective distributions.

To get an initial idea of the extremal dependence structure, we use the [Simpson et al. \(2020\)](#) methodology and calculate $\tau_C(\delta)$ for a range of values of δ . These estimates suggest that the spectral measure places mass on the faces $\mathbb{B}_{\{H\}}, \mathbb{B}_{\{S\}}, \mathbb{B}_{\{P\}}, \mathbb{B}_{\{H,S\}}$, which fits with the assessment in [Bortot et al. \(2000\)](#) and [Coles and Pauli \(2002\)](#).

To calculate the threshold $r_0(\mathbf{w})$, we use the rolling-windows procedure described in [Section 4.1](#), with $\tau = 0.95$. We then fit model (5) with three forms for g : (i) the asymmetric logistic gauge function with the structure given by $\tau_C(\delta)$, (ii) gauge corresponding to the Gaussian distribution, and (iii) an additive mixture of the Gaussian and asymmetric logistic gauges, as described in [Section 4.3.2](#). The respective AIC values are 374.9, 365.5 and 369.5.

In spite of the structure suggested by the estimated $\tau_C(\delta)$ values, the AIC indicates a preference for the Gaussian gauge function. The maximum likelihood estimates are $(\hat{\alpha}, \hat{\theta}_{HP}, \hat{\theta}_{HS}, \hat{\theta}_{PS}) = (0.79, 0.70, 0.65, 0.30)$, where θ_{jk} are the Gaussian correlation parameters in the gauge function. The data have been filtered to give approximate temporal independence, so we estimate Hessian-based standard errors as $(0.74, 0.12, 0.12, 0.18)$. [Figure 10](#) displays the PP plot for this fit as described in [Section 4.4](#), as well as the same plot transformed onto the exponential scale to emphasize the upper tail, indicating no lack of fit. We also compare the empirical gauge $\hat{g}(\mathbf{w})$ and the fitted Gaussian

gauge function $g(\mathbf{w}; \hat{\boldsymbol{\theta}})$ in the right panel of Figure 10. The empirical gauge is relatively “jagged” and variable due to the manner of its calculation, but there is broad correspondence between its overall shape and that of the fitted gauge. Interestingly, the fit of the asymmetric logistic gauge returns a parameter estimate of $\hat{\gamma}_{HS} = 1$, which is on the boundary of the parameter space. This could be indicative of wave height and surge not displaying exceptionally strong dependence, but also because there are restrictions on the shape of the limit set arising from the asymmetric logistic distribution, and this parameter estimate provides the best overall fit to all data simultaneously.

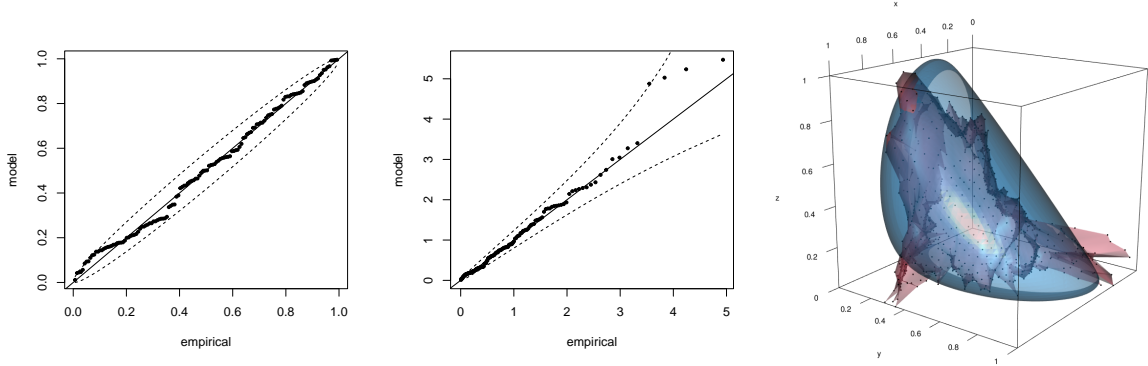


Figure 10: Left and centre: PP and exponential QQ plots for the fitted truncated gamma model with the Gaussian gauge function. Right: unit level set of the empirical gauge function (jagged, red) and fitted Gaussian gauge function (smooth, blue) for the Newlyn wave data.

As a further diagnostic, we compare empirical and model-based estimates of the sub-asymptotic joint *tail dependence coefficient*. For $X_j \sim F_j$, this is defined by

$$\chi_C(u) = \frac{1}{1-u} \Pr [F_j(X_j) > u, \forall j \in C], \quad u \in (0, 1), \quad C \subseteq \{H, S, P\}. \quad (14)$$

The empirical estimator of $\chi_C(u)$ is obtained by replacing each distribution function and joint probability with its empirical counterpart, while the model-based estimate is calculated using simulation from the fitted model as described in Section 4.5, and suitable sets B . In Figure 11 we consider $\chi_{HSP}(u)$ and $\chi_{HS}(u)$, meaning $B = (-\log(1-u), \infty)^3$ and $B = (-\log(1-u), \infty)^2 \times (0, \infty)$, respectively. The range over which the model-based tail dependence coefficients can be calculated depends on the values of \mathbf{X} constituting the extreme region $\{R' > 1\}$. There is good agreement with the empirical estimates, with the model-based estimates allowing extrapolation beyond the range of the data.

Finally we consider analysis of the structure variable outlined in Coles and Tawn (1994). They

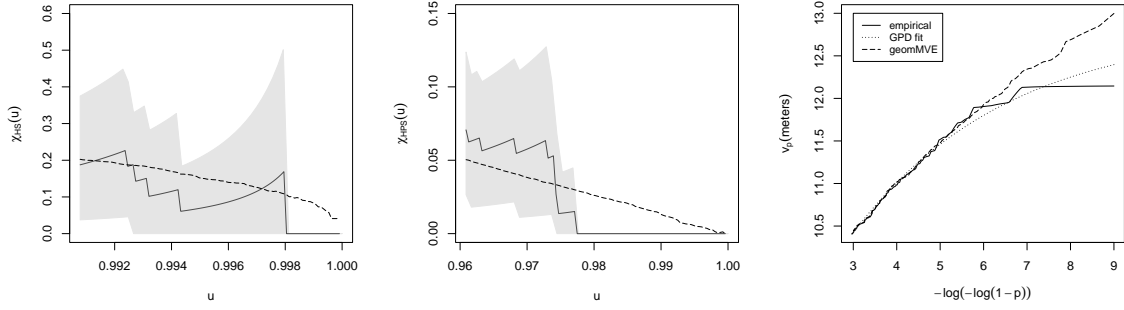


Figure 11: Left: Estimates of $\chi_{HS}(u)$, $u \in (0.99, 1)$; centre: estimates of $\chi_{HSP}(u)$, $u \in (0.965, 1)$. Black solid lines represent empirical estimates, grey shaded regions 95% pointwise confidence intervals, and dashed lines the model-based estimate for the Newlyn wave dataset. Right: plot of quantiles v_p of the structure variable V , representing sea wall height.

introduce the overtopping discharge rate $Q(v; \mathbf{X}_{HSP})$ for a sea wall of height v as

$$Q(v; \mathbf{X}_{HSP}) = a_1 X_S X_P \exp \left[-\frac{a_2 (v - X_S - l)}{X_P X_H^{*1/2}} \right], \quad X_H^* = X_H \left\{ 1 - \exp \left[-\frac{(l + X_S)^2}{2X_H^2} \right] \right\}^{1/2}.$$

The value X_H^* is introduced to approximate the actual off-shore wave height, since measurements are taken on-shore. The goal is to estimate the sea wall height v_p (in metres) for which the overtopping discharge rate is expected to exceed $0.002 \text{ m}^3 \text{ s}^{-1}$ per metre of sea wall with probability p . That is, setting $V = Q^{-1}(0.002; \mathbf{X}_{HSP})$, we solve $\Pr(V > v_p) = p$ for v_p using realizations of V generated through simulation and reverse marginal transformation. Specifically, we generate new realizations of \mathbf{X}_{HSP} , and hence V , in the tail region of our model by simulating on exponential margins and inverting equation (13). Outside of the tail region, we use the empirical distribution of V . As in Bortot et al. (2000), we fix $a_1 = 0.25$, $a_2 = 26$, and $l = 4.3$. The right panel of Figure 11 displays the obtained values v_p , with empirical quantiles and those calculated from fitting the generalized Pareto distribution directly to the tail of V (the so-called “structure variable approach”) for comparison. For very small p , the return levels obtained from the geometric model are larger than those from the generalized Pareto fit. They are comparable to those obtained in Bortot et al. (2000), but much lower than those in Coles and Tawn (1994), whose model incorrectly assumes that the spectral measure places mass on $\mathbb{B}_{\{H,S,P\}}$.

6.2 River flow data

We now apply our modelling approach on 12,327 measurements of daily mean river flow (m^3/s) from four gauging stations in the north west of England. The data were previously explored in

Simpson et al. (2020), where focus lay on determining the support of the spectral measure, but not subsequent modelling of the variables, due to lack of suitable models that could account for complex structures. We opt to consider four out of the five locations initially used in order to keep the number of parameters reasonable; further discussion on dimensionality can be found in Section 7. The four stations, labelled 1, 2, 3, 4, correspond to those labelled A, B, C, D in Simpson et al. (2020).

Margins are standardized using equation (13). We then use the Simpson et al. (2020) methodology, which suggests that the spectral measure may place mass on the faces $\mathbb{B}_{\{2\}}, \mathbb{B}_{\{4\}}, \mathbb{B}_{\{1,4\}}, \mathbb{B}_{\{1,3,4\}}$, and $\mathbb{B}_{\{1,2,3,4\}}$ of the simplex \mathcal{S}_3 . We fit the model with the corresponding asymmetric logistic gauge function, a Gaussian gauge function, and an additive mixture of the two. The AIC values are 2666, 2601 and 2609, respectively. Once again, the model with the Gaussian gauge is preferred, in apparent conflict with the estimated structure of the spectral measure, though we note this is also subject to uncertainty. Parameter estimates and approximate standard errors are given in Table 1. To account for temporal dependence of river flows, standard errors are found via use of a block bootstrap on the original data series, with block length 20. The asymmetric logistic gauge, while able to capture the structure of different groups of variables being co-extreme, appears too inflexible to capture other aspects of the dependence. The additively mixed model is an attempt to alleviate this problem, but leads to a large number of parameters without a sufficient improvement in fit to compensate for them.

Table 1: Parameter estimates and approximate block bootstrap-based standard errors for the river flow data. Parameter θ_{jk} represents the Gaussian gauge correlation parameter between sites j, k .

Parameter	α	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
Estimate	2.46	0.83	0.90	0.80	0.90	0.57	0.62
Standard error	0.62	0.11	0.14	0.14	0.14	0.16	0.14

Figure 12 displays coefficients $\chi_{123}(u)$, $\chi_{134}(u)$ and $\chi_{1234}(u)$, defined analogously to (14). If H places mass on $\mathbb{B}_{\{1,2,3,4\}}$, then each of these coefficients has a positive limit as $u \rightarrow 1$, but at observable levels, the model-based estimates from the Gaussian gauge all represent a good fit to the data. Plots of $\chi_C(u)$ for the remaining groups of variables are given in Section H of the supplement, along with the PP plot, showing no lack of fit. In the limit as $u \rightarrow 1$, estimates of $\chi_C(u)$ from the geometric model with Gaussian gauge will all be zero. However, the inference that H places mass on $\mathbb{B}_{\{1,2,3,4\}}$, and other faces, is subject to uncertainty. From the plots in Figure 12, it is difficult to determine whether the limits of $\chi_{123}(u)$, $\chi_{134}(u)$ and $\chi_{1234}(u)$ as $u \rightarrow 1$ are indeed positive or zero, and as a consequence whether a gauge function that reflects $H(\mathbb{B}_{\{1,2,3,4\}}) > 0$ is truly preferable.

Nonetheless, this framework offers the chance to test these models and assumptions in a way that was not previously possible.

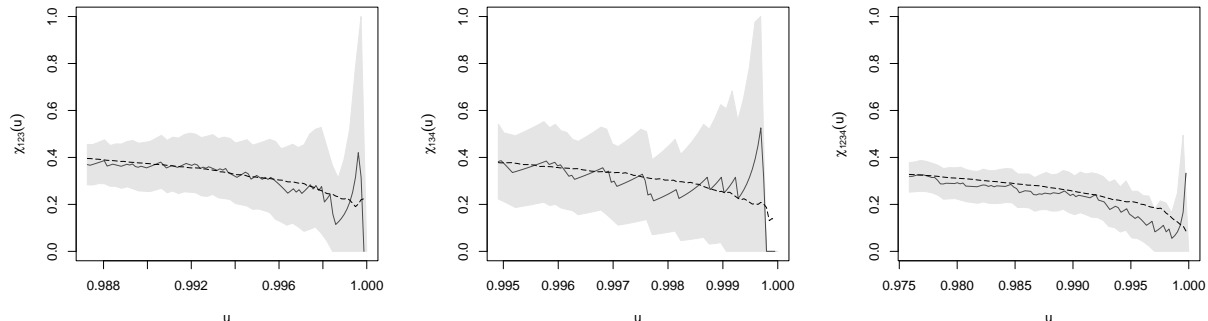


Figure 12: Empirical (solid black) estimates of $\chi_C(u)$ with 95% pointwise confidence interval (grey shaded region), and model-based estimate (dashed black) for $C = \{1, 2, 3\}$, $\{1, 3, 4\}$ and $\{1, 2, 3, 4\}$ (left to right).

7 Discussion

We have presented a new approach to multivariate extreme value modelling, based on estimation of the shape of the limit set of a sample cloud of data points in light-tailed margins. The methodology allows for modelling datasets with complicated extremal dependence structures, whereby different groups of variables may be co-extreme, as well as extrapolation into parts of the multivariate tail where only some variables are large.

By offering models for complex dependence structures with non-simultaneous extremes, this approach paves the way for more useful higher dimensional extreme value modelling. Recent literature on multivariate extremes that is targeted at higher dimensions typically involves making strong simplifying assumptions on the dependence structure. For example, the extremal graphical models outlined in [Engelke and Hitz \(2020\)](#) require an assumption that the spectral measure H places all mass on $\mathbb{B}_{\{1, \dots, d\}}$.

In this work, we demonstrated the methods up to dimension $d = 4$. The main challenges for routine application of the methods for d larger than 3 or 4 lie in calculation of the threshold function $r_0(\mathbf{w})$, and specification of flexible gauge functions. The former could potentially be addressed by adapting the additive quantile regression approach of [Fasiolo et al. \(2021\)](#) to incorporate basis functions whose support is the simplex \mathcal{S}_{d-1} . Addressing the latter challenge requires ways to build flexible and parsimonious gauge functions, which is a topic of current work. In particular, we note

that models fitted in Section 6 had the ability to capture the complex structures suggested by the Simpson et al. (2020) methodology, but the best fits were obtained through models that were more flexible in other aspects. This led to the conclusion that the model with the Gaussian gauge function was preferred for both datasets, which is likely a consequence of being able to capture a range of strengths of dependences across different groups of variables; in contrast, the asymmetric logistic gauge function treats groups of variables that do not exhibit simultaneous extremes as effectively independent. We note also that estimates of the faces \mathbb{B}_C on which H places mass are themselves subject to uncertainty, which is not easily quantifiable thanks to the requirement to select tuning parameters. Conflicts between the estimated structure and the selected gauge function may therefore not be too concerning, provided the diagnostics for the model are adequate.

A further challenge with our methodology for dimensions $d \geq 5$ is the use of the empirical distribution for the angles \mathbf{W} . We anticipate that considering (semi-)parametric forms for this distribution will be needed as part of adapting the methods to higher dimensions.

We have focused here primarily on positive dependence as it is common in many datasets and simplifies the presentation. For datasets exhibiting any form of negative dependence, the limit set shapes are more descriptive in Laplace, rather than exponential, margins. For example, we mentioned for the multivariate Gaussian case that the continuous convergence to $g(\mathbf{x})$ fails when some component of \mathbf{x} is zero; this is not an issue in Laplace margins, where the limit set lies in the region $[-1, 1]^d$ rather than $[0, 1]^d$, and similarly for the t_ν distribution. Moving from the positive quadrant to \mathbb{R}^d requires defining the angles \mathbf{W} differently, but otherwise a similar approach could be applied, and represents a natural next step in developing this methodology.

Acknowledgements

We thank two reviewers and an associate editor for helpful comments that have improved the manuscript.

Funding

We are grateful for EPSRC grant EP/X010449/1 supporting Jennifer Wadsworth, and EPSRC DTP funding EP/W523811/1 supporting Ryan Campbell.

Conflict of interest

The authors have no conflict of interest to declare.

Data availability

Data for the analyses in Section 6 are available as supplementary material. Acknowledgement: Data from the UK National River Flow Archive (Section 6.2).

Computer code

Code for the analyses in Section 6 is available as supplementary material. An R package `geometricMVE` for implementing the methodology presented in the article is available as supplementary material and at <http://www.lancaster.ac.uk/~wadswojl/geometricMVE.html>. Interactive versions of 3d plots are available at the same URL.

References

- Balkema, A., Embrechts, P., and Nolde, N. (2010). Meta densities and the shape of their sample clouds. *Journal of Multivariate Analysis*, 101:1738–1754.
- Balkema, A. and Nolde, N. (2012). Asymptotic dependence for homothetic light-tailed densities. *Advances in Applied Probability*, 44:506–527.
- Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.
- Bortot, P., Coles, S., and Tawn, J. (2000). The multivariate Gaussian tail model: An application to oceanographic data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 49(1):31–049.
- Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.
- Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, 9:383–418.
- Coles, S. and Pauli, F. (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89(1):183–196.
- Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 53(2):377–392.
- Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 43(1):1–31.
- Davis, R. A., Mulrow, E., and Resnick, S. I. (1988). Almost sure limit sets of random samples in \mathbb{R}^d . *Advances in Applied Probability*, 20(3):573–599.
- de Haan, L. (1970). *On Regular Variation and Its Application to Weak Convergence of Sample Extremes*. CWI Tract 32, Amsterdam.
- de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337.

- Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317.
- Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):871–932.
- Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., and Goude, Y. (2021). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, 116(535):1402–1412.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190.
- Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70(351a):674–680.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44:423–453.
- Goix, N., Sabourin, A., and Cléménçon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3):497–546.
- Hilal, S., Poon, S.-H., and Tawn, J. (2014). Portfolio risk assessment using multivariate extreme value methods. *Extremes*, 17:531–556.
- Jacob, P. and Massé, B. (1996). Asymptotic behavior of samples from general multivariate distributions. *Journal of Statistical Planning and Inference*, 52(2):183–196.
- Keef, C., Tawn, J. A., and Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21.
- Kinoshita, K. and Resnick, S. I. (1991). Convergence of scaled random samples in \mathbb{R}^d . *Annals of Probability*, 19(4):1640–1663.
- Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019). Peaks over thresholds modelling with multivariate generalized Pareto distributions. *Technometrics*, 61(1):123–135.

- Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 59(2):475–499.
- Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.
- Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717.
- Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809.
- Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(1):219–241.
- Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.
- Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.
- Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.
- Simpson, E. S. and Tawn, J. A. (2022). Estimating the limiting shape of bivariate scaled sample clouds for self-consistent inference of extremal dependence properties. *arXiv:2207.02626*.
- Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.
- Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.
- Vettori, S., Huser, R., and Genton, M. G. (2019). Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3):831–841.
- Wadsworth, J. and Tawn, J. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.
- Zhang, Z. and Huang, J. (2006). Extremal financial risk models and portfolio evaluation. *Computational Statistics & Data Analysis*, 51(4):2313–2338.
- Zhang, Z., Huser, R., Opitz, T., and Wadsworth, J. (2022). Modeling spatial extremes using normal mean-variance mixtures. *Extremes*, 25(2):175–197.