# Determining the Dependence Structure of Multivariate Extremes

E. S. Simpson[1], J. L. Wadsworth[2], J. A. Tawn[2]

[1]*STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, LA1 4YR, U.K.*
[2]*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.*

### Abstract

In multivariate extreme value analysis, the nature of the extremal dependence between variables should be considered when selecting appropriate statistical models. Interest often lies with determining which subsets of variables can take their largest values simultaneously, while the others are of smaller order. Our approach to this problem exploits hidden regular variation properties on a collection of non-standard cones and provides a new set of indices that reveal aspects of the extremal dependence structure not available through existing measures of dependence. We derive theoretical properties of these indices, demonstrate their value through a series of examples, and develop methods of inference that also estimate the proportion of extremal mass associated with each cone. We apply the methods to UK river flows, estimating the probabilities of different subsets of sites being large simultaneously.

**Keywords:** asymptotic independence, extremal dependence structure, hidden regular variation, multivariate regular variation.

## 1   Introduction

When constructing models in multivariate extreme value analysis, we often need to exploit extremal dependence features. Consider the random vector $X = (X_1, \ldots, X_d)$, with $X_i \sim F_i$, as well as a subset of these variables $X_C = \{X_i : i \in C\}$, for some $C \in 2^D \setminus \emptyset$, i.e., $C$ lies in the power set of $D = \{1, \ldots, d\}$ without the empty set. For any $C$ with $|C| \geq 2$, extremal dependence within $X_C$ can be summarized by

$$\chi_C = \lim_{u \to 1} \mathrm{pr}\left\{F_i(X_i) > u : i \in C\right\} / (1 - u) \tag{1}$$

if the limit exists. In particular, if $\chi_C > 0$, the variables in $X_C$ are asymptotically dependent, i.e., can take their largest values simultaneously. If $\chi_C = 0$, the variables in $X_C$ cannot all take their largest values together, although it is possible that for some $\underline{C} \subset C$, $\chi_{\underline{C}} > 0$, see for example Hua and Joe (2011) or Wadsworth and Tawn (2013).

Many models for multivariate extremes are only applicable when data exhibit either full asymptotic dependence, entailing $\chi_C > 0$ for all $C \in 2^D \setminus \emptyset$ with $|C| \geq 2$, or full asymptotic independence, i.e., $\chi_{i,j} = 0$ for all $i < j$ (Heffernan and Tawn, 2004). However, often some $\chi_C$ are positive whilst others are zero, i.e., only certain subsets of the variables take their largest values simultaneously, while the other variables are of smaller order. The extremal dependence between variables can thus have a complicated structure, which should be exploited when modelling. In this paper, we present two methods for determining this structure.

The full extremal dependence structure is not completely captured by the $2^d - d - 1$ coefficients $\{\chi_C : C \in 2^D \setminus \emptyset, |C| \geq 2\}$ since we do not learn fully whether small values of some variables occur with large values of others, or whether individual variables can be extreme in isolation. This is revealed more clearly by decomposing the vector into radial and angular components, $(R, W)$, and examining their asymptotic structure. If the $X_i$ follow a common heavy-tailed marginal distribution, usually achieved via a transformation, these pseudo-polar coordinates are defined as $R = \|X\|_1$ and $W = X / \|X\|_2$, for arbitrary norms $\|\cdot\|_1$ and $\|\cdot\|_2$. We take both to be the $L_1$ norm, and assume that $X$ has standard Fréchet margins, so that $\mathrm{pr}(X_i < x) = \exp(-1/x)$ for $x > 0$ and $i = 1, \ldots, d$. As such, the radial and angular components are $R = \sum_{i=1}^d X_i$ and $W = X/R$, respectively, with $R > 0$ and $W \in \mathcal{S}_{d-1} = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i=1}^d w_i = 1\right\}$, the $(d-1)$-dimensional unit simplex. It
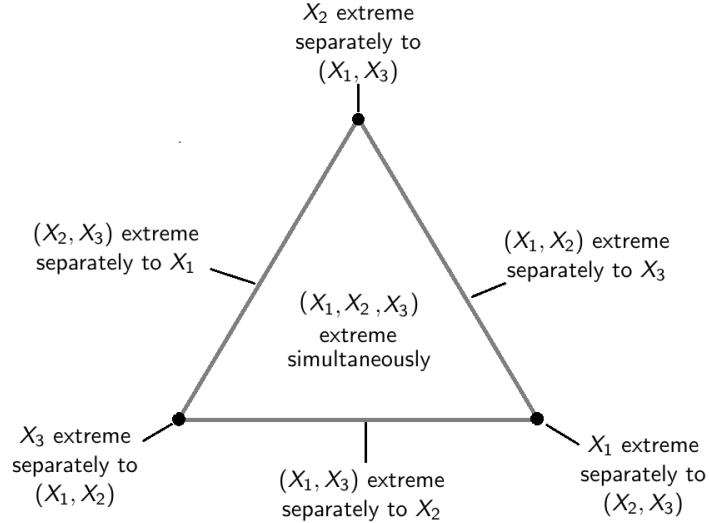
Figure 1: The simplex $\mathcal{S}_2$. Coordinates are transformed to the equilateral simplex.

follows that $\mathrm{pr}(R > r) \sim a r^{-1}$ as $r \to \infty$, for $a \geq 1$, so all the information about extreme events is contained in $W$, and in particular the distribution of $W$ conditioned on $R > r$ as $r \to \infty$. Under the assumption of multivariate regular variation (Resnick, 2007, Chapter 6),

$$\lim_{t \to \infty} \mathrm{pr}(R > tr, W \in B \mid R > t) = H(B)r^{-1}, \quad r \geq 1, \tag{2}$$

for $B$ a measurable subset of $\mathcal{S}_{d-1}$, where the limiting spectral measure $H$ satisfies

$$\int_{\mathcal{S}_{d-1}} w_i dH(w) = 1/d, \quad i = 1, \ldots, d. \tag{3}$$

As the radial component becomes large, the position of mass on $\mathcal{S}_{d-1}$ reveals the extremal dependence structure of $X$. We note the link between the dependence measure $\chi_C$ in (1), and the spectral measure $H$: if $\chi_C > 0$, then $H$ places mass on at least one region $\mathcal{S}_{d-1}^{\overline{C}} = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i \in \overline{C}} w_i = 1\right\}$, with $C \subseteq \overline{C} \subseteq D$. This underlines that the term asymptotic dependence is not so useful here, since it offers only partial insight into the structure. In what follows, we thus avoid this term where possible, talking instead about faces of the simplex on which $H$ places mass.

In the $d$-dimensional case, $\mathcal{S}_{d-1}$ can be partitioned into $2^d - 1$ faces, each of which could contain mass. Mass on each of these faces corresponds to a different subset of the variables $(X_1, \ldots, X_d)$ being the only ones taking their largest values concurrently. This is demonstrated in Fig. 1 for $d = 3$. For high, or even moderate, dimensions, there are many faces to consider, and the task of determining which faces truly contain mass, and therefore the extremal dependence structure of the variables, is not straightforward, as for a finite sample with continuous margins, points cannot lie exactly on the boundary of the simplex: no $W_i$ equal zero when $R < \infty$.

The multivariate regular variation assumption (2) can also be phrased in terms of measures on the cone $\mathbb{E} = [0, \infty]^d \setminus \{0\}$, see Section 2.1. Each face of $\mathcal{S}_{d-1}$ can be identified with a sub-cone of $\mathbb{E}$ for which one or more components are identically zero. Intuition and visualization are often simpler with $H$, but in the sequel we work with $\mathbb{E}$ and the sub-cones corresponding to faces of $\mathcal{S}_{d-1}$. Variants of our methods that directly use the radial-angular framework are presented in the 2019 Lancaster University PhD thesis of E. S. Simpson.

The problem of determining the extremal dependence structure of variables has been recently studied elsewhere in the literature. Under the assumption that the data are from an elliptical copula, Klüppelberg et al. (2015) use factor analysis on extreme correlations linked to the tail dependence function. Chautru (2015) introduces a non-parametric approach based on statistical learning, combining a principal component analysis algorithm with clustering techniques. A Bayesian clustering method is proposed by Vettori et al. (2018), based on the hierarchical dependence structure of the nested logistic distribution of Tawn (1990). Goix et al. (2016,

2017) propose a non-parametric simplex partitioning approach in which they condition on the radial variable being above some high threshold. They assume that there is mass on a particular face if the number of points in the corresponding region of the simplex is sufficiently large, leading to a sparse representation of the dependence structure. Chiapino and Sabourin (2017) propose an algorithm to group together nearby faces with extremal mass into feature clusters, by exploiting their graphical structure and a measure of extremal dependence. Finally, Chiapino et al. (2019) extend this approach by instead using the coefficient of tail dependence of Ledford and Tawn (1996).

In this paper we exploit additional, but commonly satisfied, hidden regular variation assumptions on non-standard sub-cones of $\mathbb{E}$, by introducing a new set of parameters that describes the dominant extremal dependence structure. We study properties of these parameters, their link to existing coefficients, and explore their values for a range of examples. Estimation of the parameters provides us with an asymptotically-motivated framework for determining the extremal dependence structure, as well as allowing us to estimate the proportion of mass associated with each set of variables. We propose two such inferential methods, both with computational complexity $O(dn \log n)$ if $d < n$, for $d$ representing the number of variables and $n$ the number of data points. This is the same complexity as the method of Goix et al. (2017).

## 2 Theoretical motivation

### 2.1 Multivariate regular variation

A function $\lambda : (0, \infty] \to (0, \infty]$ is said to be regularly varying at infinity, with index $\alpha \in \mathbb{R}$, if $\lambda(tx)/\lambda(t) \to x^\alpha$, as $t \to \infty$, for all $x > 0$. For such functions, we write $\lambda \in \mathrm{RV}_\alpha$. We can always express $\lambda(x) = L(x)x^\alpha$, with $L \in \mathrm{RV}_0$ termed a slowly varying function. A cone $G \subset \mathbb{R}^d$ is a set such that for any $x \in G$, $tx \in G$ for all $t > 0$. The assumption of multivariate regular variation on the cone $G$ means that there exists a scaling function $a(t) \to \infty$, and a positive measure $\mu$, such that

$$t \, \mathrm{pr}\left(X/a(t) \in \cdot\right) \to \mu(\cdot), \quad t \to \infty, \tag{4}$$

with vague convergence in the space of non-negative Radon measures on $G$ (Resnick 2007, Chapter 3). If we assume that the margins of $X$ are standard Fréchet or Pareto, we may take $a(t) = t$, and the limit measure $\mu$ is homogeneous of order $-1$. For the remainder of this section, we assume that $X$ has standard Fréchet marginal distributions.

### 2.2 Hidden regular variation

The concept of hidden regular variation was introduced by Resnick (2002), who formalized and extended the ideas of Ledford and Tawn (1996, 1997). Further work has been done by Maulik and Resnick (2004) and Mitra and Resnick (2011), for example, whilst Resnick (2007) provides a textbook treatment. Here, multivariate regular variation is assumed on some cone in $\mathbb{R}^d$. If there is also regular variation, but with a scaling function of smaller order, on some sub-cone, we have hidden regular variation on that sub-cone.

To our knowledge, the marginal case of this hidden regular variation framework is the only one previously exploited from a statistical perspective; from a theoretical viewpoint, Das et al. (2013) consider hidden regular variation on a series of non-standard cones, although these are mostly different from the ones we will consider. For $X_C = \{X_i : i \in C\}$, $x_C = \{x_i : i \in C\}$, Ledford and Tawn (1997) considered multivariate regular variation on the cone $\mathbb{E} = [0, \infty]^d \setminus \{0\}$ and hidden regular variation on

$$\mathbb{E}_C^* = \left\{ x \in \mathbb{E} : x_C \in (0, \infty]^{|C|}, \ x_{D \setminus C} \in [0, \infty]^{|D \setminus C|} \right\}, \tag{5}$$

with limit measures on $\mathbb{E}_C^*$ homogeneous of order $-1/\eta_C$, and the so-called coefficient of tail dependence $\eta_C$ taking values in $(0, 1]$. If $\mu(\mathbb{E}_C^*) > 0$, variables $X_C$ can take their largest values simultaneously. If we instead consider sub-cones of $\mathbb{E}$ of the form

$$\mathbb{E}_C = \left\{ x \in \mathbb{E} : x_C \in (0, \infty]^{|C|}, \ x_{D \setminus C} = \{0\}^{|D \setminus C|} \right\}, \tag{6}$$

where $\{0\}^m$ denotes an $m$-vector of zeros, then having $\mu(\mathbb{E}_C) > 0$ indicates that variables in $X_C$ can take their largest values simultaneously while variables in $X_{D\setminus C}$ are of smaller order. Our task is to determine the cones $\mathbb{E}_C$ on which $\mu$ places mass, equivalent to the problem of detecting where $H$ places mass on $\mathcal{S}_{d-1}$, thus revealing the extremal dependence structure of $X$. For simplicity, we assume that if $\mu$ places mass on $\mathbb{E}_C$, then for any measurable $B_C \subset \mathbb{E}_C$, $\mu(B_C) > 0$. More generally, it need only be true that there exists $B_C \subset \mathbb{E}_C$ such that $\mu(B_C) > 0$, however if the mass lies only in very restricted parts of $\mathbb{E}_C$, then the task of detecting which cones contain mass is naturally more difficult.

For a finite sample, mass will not occur on cones $\mathbb{E}_C$ with $|C| < d$; this is equivalent to all mass for $W$ being placed on the interior of the simplex in Fig. 1. One option is to truncate the variables to zero below some marginal threshold, to ensure mass on at least some of these cones at a finite level. Let us define

$$X^* = \begin{cases} 0, & X \leq -1/\log p, \\ X, & X > -1/\log p, \end{cases} \tag{7}$$

such that $p$ is the quantile at which we truncate. The variable $X^*$ has the same tail behaviour as $X$, but in general $\mathrm{pr}(X^*/t \in B_C) > 0$ for $B_C \subset \mathbb{E}_C$, and in this way we could define a hidden regular variation assumption on $\mathbb{E}_C$. Writing $B_C = \{x \in \mathbb{E} : x_C \in B \subset (0,\infty]^{|C|}, x_{D\setminus C} \in \{0\}^{|D\setminus C|}\}$, then $\mathrm{pr}(X^*/t \in B_C) = \mathrm{pr}(X_C/t \in B, X_{D\setminus C} \in [0, -1/\log p]^{|D\setminus C|})$, such that we consider the behaviour when the variables $X_C$ are growing at a common rate, but variables $X_{D\setminus C}$ have a fixed upper bound. However, the latter condition does not capture all possible behaviour that leads to variables $X_{D\setminus C}$ being of smaller order than $X_C$, and in general a more elaborate assumption is needed. We consider how we can allow $X_{D\setminus C}$ to be bounded above by a function that is growing, but at a potentially slower rate than $t$.

Define the set $(y, \infty]^C \times [0, z]^{D\setminus C} = \{x \in \mathbb{E} : x_i > y, i \in C; x_j \leq z, j \in D \setminus C\}$. Then under the regular variation assumption (4),

$$t\,\mathrm{pr}\left\{X/t \in (y, \infty]^C \times [0, z]^{D\setminus C}\right\} \to \mu\left((y, \infty]^C \times [0, z]^{D\setminus C}\right) \geq \mu\left((y, \infty]^C \times \{0\}^{D\setminus C}\right). \tag{8}$$

Therefore, if $\mu\left((y, \infty]^C \times \{0\}^{D\setminus C}\right) > 0$, and hence $\mu(\mathbb{E}_C) > 0$, this indicates that in (8) we may be able to consider $z = z_t \to 0$ at a suitable rate in $t$ and still observe

$$\lim_{t \to \infty} t\,\mathrm{pr}\left\{X/t \in (y, \infty]^C \times [0, z_t]^{D\setminus C}\right\} > 0. \tag{9}$$

A consequence of a positive limit in (9) is that $\mathrm{pr}\{X/t \in (y, \infty]^C \times [0, z_t]^{D\setminus C}\} \in \mathrm{RV}_{-1}$. When the limit in (9) is zero, then either $z_t \to 0$ too quickly — consider for example the case $z_t \equiv 0$ — or $\mu$ places no mass on $\mathbb{E}_C$. In these cases we focus on the rate of convergence to zero in (9). Taking $z_t = zt^{\delta-1}$ for $\delta \in [0,1]$, and rephrasing in terms of min and max projections, our main assumption is as follows.

**Assumption 1.** *Suppose we have regular variation on the cone $\mathbb{E} = [0, \infty]^d \setminus \{0\}$, so that equation (4) is satisfied with $\mu$ homogeneous of order $-1$. For all $C \subseteq D$, let $X_\wedge^C = \min_{i \in C} X_i$ and $X_\vee^{D\setminus C} = \max_{i \in D\setminus C} X_i$. We assume that for all $\delta \in [0,1]$,*

$$pr\left\{\left(X_\wedge^C/t, X_\vee^{D\setminus C}/t^\delta\right) \in (y, \infty] \times [0, z]\right\} \in \mathrm{RV}_{-1/\tau_C(\delta)}, \quad t \to \infty, \quad 0 < y, z < \infty, \tag{10}$$

*and that there exists $\delta^* < 1$ such that $\tau_C(\delta^*) = 1$ for all $C$ such that $\mu(\mathbb{E}_C) > 0$, and $\tau_C(\delta^*) < 1$ for all $C$ such that $\mu(\mathbb{E}_C) = 0$.*

We note that the probability in (10), and hence $\tau_C(\delta)$, is non-decreasing in $\delta$. The case $\delta = 0$ and $z = -1/\log p$ is identical to a regular variation assumption on the truncated variables $X^*$; allowing $\delta > 0$ produces a more diverse range of possibilities.

Through the final line of Assumption 1, the indices $\tau_C(\delta)$ contain information on the limiting extremal dependence structure; the challenge is to find a suitable $\delta^*$, noting that if $\delta$ is too small we could have $\tau_C(\delta) < 1$ even when $\mu(\mathbb{E}_C) > 0$, but if $\delta$ is too large, some $\tau_C(\delta)$ could be close to one even when $\mu(\mathbb{E}_C) = 0$, making

the detection problem difficult in light of statistical uncertainty. These issues are discussed further below and in Section 3.

Overall, examining the regular variation properties in Assumption 1 leads to understanding of the sub-asymptotic behaviour of $\mu$ in relation to which cones $\mathbb{E}_C$ are charged with mass. This is analogous to determining the support of $H$ in (2). In the remainder of this section, we illustrate the utility and validity of our hidden regular variation assumption via examples, and discuss properties of $\tau_C(\delta)$. Theorems 1 and 2 clarify some links between $\tau_C(\delta)$ and $\eta_C$.

**Theorem 1.** *Assume regular variation, or hidden regular variation, on $\mathbb{E}_C^*$ defined in (5), such that $pr\left(X_\wedge^C > t\right) \in$ $\mathrm{RV}_{-1/\eta_C}$. Suppose further that for all $\overline{C} \subseteq D$ such that $\overline{C} \supseteq C$, Assumption 1 is satisfied, so that for $\delta = 1$, $pr\left(X_\wedge^{\overline{C}} > t, X_\vee^{D\backslash\overline{C}} \leq t\right) \in \mathrm{RV}_{-1/\tau_{\overline{C}}(1)}$. Then $\eta_C = \max_{\overline{C}:C\subseteq\overline{C}} \tau_{\overline{C}}(1)$.*

*Proof.* We have

$$\mathrm{pr}\left(X_\wedge^C > t\right) = \mathrm{pr}\left(X/t \in (1,\infty]^C \times [0,\infty]^{D\backslash C}\right);$$

$$\mathrm{pr}\left(X_\wedge^{\overline{C}} > t, X_\vee^{D\backslash\overline{C}} \leq t\right) = \mathrm{pr}\left(X/t \in (1,\infty]^{\overline{C}} \times [0,1]^{D\backslash\overline{C}}\right).$$

By the partition $(1,\infty]^C \times [0,\infty]^{D\backslash C} = \bigcup_{\overline{C}:C\subseteq\overline{C}}(1,\infty]^{\overline{C}} \times [0,1]^{D\backslash\overline{C}}$, we deduce that

$$\mathrm{pr}\left(X/t \in (1,\infty]^C \times [0,\infty]^{D\backslash C}\right) = \sum_{\overline{C}:C\subseteq\overline{C}} \mathrm{pr}\left(X/t \in (1,\infty]^{\overline{C}} \times [0,1]^{D\backslash\overline{C}}\right),$$

from which the result follows. □

We note $\tau_D(\delta)$ does not depend on $\delta$; we therefore denote it by $\tau_D$.

**Theorem 2.** *For all $C \in 2^D \setminus \emptyset$ with $|C| \geq 2$, assume regular variation or hidden regular variation on $\mathbb{E}_C^*$ with coefficient of tail dependence $\eta_C$, and suppose Assumption 1 holds. For any set $C$ with $|C| \geq 2$, assume that for any $\overline{C} \supseteq C$, $\eta_{\overline{C}'} < \eta_{\overline{C}}$ for all $\overline{C}' \supset \overline{C}$. Then $\tau_C(1) = \eta_C$, and $\tau_C(\delta) \leq \eta_C$ for all $\delta \in [0,1]$.*

*Proof.* Since $\mathbb{E}_D = \mathbb{E}_D^*$, we have $\eta_D = \tau_D$, and by Theorem 1, for any set $C_{d-1} \subset D$ with $|C_{d-1}| = d-1$,

$$\eta_{C_{d-1}} = \max\left\{\tau_{C_{d-1}}(1), \tau_D\right\} = \max\left\{\tau_{C_{d-1}}(1), \eta_D\right\}.$$

Since, by assumption, $\eta_{C_{d-1}} > \eta_D$, we have $\eta_{C_{d-1}} = \tau_{C_{d-1}}(1)$. Similarly, for any set $C_{d-2} \subset C_{d-1}$ with $|C_{d-2}| = d-2$,

$$\eta_{C_{d-2}} = \max\left\{\tau_{C_{d-2}}(1), \tau_{C_{d-1}}(1), \tau_D\right\} = \max\left\{\tau_{C_{d-2}}(1), \eta_{C_{d-1}}, \eta_D\right\}.$$

Again, since $\eta_{C_{d-2}} > \eta_{C_{d-1}} > \eta_D$ for all $C_{d-2} \subset C_{d-1}$, then $\eta_{C_{d-2}} = \tau_{C_{d-2}}(1)$. The result $\tau_C(1) = \eta_C$ follows by iteration for any set $C$ with $|C| \geq 2$. Since $\tau_C(\delta)$ is non-decreasing in $\delta$, $\tau_C(1) = \max_{\delta\in[0,1]} \tau_C(\delta)$, so $\tau_C(\delta) \leq \eta_C$ for all $\delta \in [0,1]$. □

In the Appendix and Supplementary Material, respectively, we calculate the value of $\tau_C(\delta)$, with $C \in 2^D \setminus \emptyset$, for a range of bivariate and multivariate copulas. For the bivariate case, we restrict our investigation to a subclass of bivariate extreme value distributions (Tawn, 1988) that covers all possible combinations of cones $\mathbb{E}_C$ charged with mass, focusing on the case where the spectral density is regularly varying at 0 and 1. For multivariate cases there are many more possibilities, so we study certain trivariate extreme value distributions (Tawn, 1990), which have $\chi_C > 0$ for at least one set $|C| \geq 2$, and two classes of copula having $\chi_C = 0$ for all $|C| \geq 2$. The results are summarized here.

The bivariate extreme value distribution in standard Fréchet margins has distribution function of the form $F(x,y) = \exp\{-V(x,y)\}$ for some exponent measure

$$V(x,y) = 2\int_0^1 \max\{w/x, (1-w)/y\}\, dH(w), \quad x,y > 0, \tag{11}$$

5

where $H$ denotes the spectral measure defined in equation (2) on the unit simplex $[0, 1]$. In the bivariate case, $\mathbb{E} = [0, \infty]^2 \setminus \{0\}$ can be partitioned into three natural cones: $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_{1,2}$. If $H(\{0\}) = \theta_2 \in [0, 1/2]$ and $H(\{1\}) = \theta_1 \in [0, 1/2]$, the distribution places mass $\theta_2, \theta_1, \theta_{1,2} = 1 - (\theta_1 + \theta_2)$ in the three cones. If $\theta_1 + \theta_2 = 1$, the variables are independent, and $\mu(\mathbb{E}_{1,2}) = 0$. In this case, all the limiting mass is placed on $\mathbb{E}_1$ and $\mathbb{E}_2$. Here, Assumption 1 holds for $C = \{1\}, \{2\}$ and $\{1, 2\}$ with $\tau_1(\delta) = \tau_2(\delta) = 1$ for all $\delta \in [0, 1]$, and $\tau_{1,2} = \eta_{1,2} = 1/2$.

When $\theta_1 + \theta_2 < 1$, $\mu(\mathbb{E}_{1,2}) > 0$ and $\tau_{1,2} = \eta_{1,2} = 1$, i.e., both variables can be simultaneously large. If $\theta_i > 0$, it follows that $\tau_i(\delta) = 1$ for $\delta \in [0, 1]$ and $i = 1, 2$, and there is mass on the corresponding cone $\mathbb{E}_i$. However, when $\theta_1 = \theta_2 = 0$, there is no mass on either of these cones, and additional conditions are required for (10) to hold. We suppose that $H$ is absolutely continuous on $(0, 1)$ with Lebesgue density $h(w) = dH(w)/dw$ satisfying $h(w) \sim c_1(1 - w)^{s_1}$ as $w \to 1$, and $h(w) \sim c_2 w^{s_2}$ as $w \to 0$, for $s_1, s_2 > -1$ and $c_1, c_2 > 0$. In the Appendix, we show that for $i = 1, 2$, $\tau_i(\delta) = \{(s_i + 2) - \delta(s_i + 1)\}^{-1}$. To illustrate this final case, consider the bivariate extreme value distribution with the logistic dependence structure (Tawn, 1988), with $V(x, y) = (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}$ and

$$h(w) = \frac{1}{2}\left(\alpha^{-1} - 1\right)\left\{w^{-1/\alpha} + (1 - w)^{-1/\alpha}\right\}^{\alpha - 2}\{w(1 - w)\}^{-1 - 1/\alpha}, \tag{12}$$

$0 < w < 1$, $\alpha \in (0, 1)$. For this model $s_1 = s_2 = -2 + 1/\alpha$, and so $\tau_1(\delta) = \tau_2(\delta) = \alpha/(1 + \alpha\delta - \delta)$ which increases from $\tau_i(\delta) = \alpha < 1$ at $\delta = 0$ to $\tau_i(\delta) = 1$ at $\delta = 1$.

| | $|C| = 1$ | $|C| = 2$ | $|C| = 3$ |
|---|---|---|---|
| (i) | $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$ | $\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 1/2$ | $\tau_{1,2,3} = 1/3$ |
| (ii) | $\tau_1(\delta) = \tau_2(\delta) = \frac{\alpha}{1+\alpha\delta-\delta}, \tau_3(\delta) = 1$ | $\tau_{1,2}(\delta) = 1, \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \frac{\alpha}{\alpha\delta+1+\alpha-\delta}$ | $\tau_{1,2,3} = 1/2$ |
| (iii) | $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = \frac{\alpha}{1+\alpha\delta-\delta}$ | $\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \frac{\alpha}{2+\alpha\delta-2\delta}$ | $\tau_{1,2,3} = 1$ |
| (iv) | $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$ | $\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 2^{-\alpha}$ | $\tau_{1,2,3} = 3^{-\alpha}$ |

Table 1: Values of $\tau_C(\delta)$ for some trivariate copula examples. For all logistic models, the dependence parameter $\alpha$ satisfies $0 < \alpha < 1$, with larger $\alpha$ values corresponding to weaker dependence. Case (i): independence; case (ii): independence and bivariate logistic; case (iii): trivariate logistic; case (iv): trivariate inverted logistic.

When $d = 3$, there are many more possibilities for combinations of cones $\mathbb{E}_C$ with mass. Table 1 gives $\tau_C(\delta)$ for four examples, in each case identifying $\tau_C(\delta)$ on cones $\mathbb{E}_C$ with $|C| = 1, 2, 3$. Cases (i)-(iii) in Table 1 are all special cases of the trivariate extreme value copula. Case (i) is the independence copula, which has limit mass on $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_3$. For the $d$-dimensional independence copula, $\tau_C(\delta) = |C|^{-1}$ for $|C| \leq d$, and does not depend on the value of $\delta$. Case (ii) is the copula corresponding to variables $(X_1, X_2)$ following a bivariate extreme value logistic distribution (12), independent of $X_3$. Here all the limit mass is placed on $\mathbb{E}_{1,2}$ and $\mathbb{E}_3$. Again, $\tau_C(\delta)$ differs between cones where $C$ is of different dimension. The trivariate extreme value logistic model, case (iii), places all extremal mass on $\mathbb{E}_{1,2,3}$, so that $\tau_{1,2,3} = 1$, and $\tau_C(\delta) < 1$ when $\delta < 1$, for all $C$ with $|C| < 3$. Since this is a symmetric model, $\tau_C(\delta)$ is the same on all three cones $\mathbb{E}_C$ with $|C| = 1$, and is also equal for each cone with $|C| = 2$.

Copula (iv) is the inverted extreme value copula (Ledford and Tawn, 1997), with a symmetric logistic dependence model. It places all limiting mass on cones $\mathbb{E}_C$ with $|C| = 1$, but unlike the independence copula, has sub-asymptotic dependence, reflected by the values of $\tau_C(\delta)$ for cones $\mathbb{E}_C$ with $|C| = 2, 3$, which are closer to one than in the independence case. The values of $\tau_C(\delta)$ do not depend on $\delta$ for this case.

The Gaussian copula with covariance matrix $\Sigma$ also exhibits asymptotic independence, with all limit mass on $\mathbb{E}_C$ with $|C| = 1$. We study the values of $\tau_C(\delta)$ for the trivariate case in the Supplementary Material. For sets $C = \{i\}$, $i = 1, 2, 3$, $\tau_C(\delta) = 1$ only if $\delta \geq \max\left(\rho_{ij}^2, \rho_{ik}^2\right)$, where $\rho_{ij} > 0$ is the Gaussian correlation parameter for variables $i$ and $j$; otherwise, $\tau_C(\delta) < 1$. We also know that $\tau_{1,2,3} = \eta_{1,2,3} = \left(1_3^T \Sigma^{-1} 1_3\right)^{-1}$, with $1_d \in \mathbb{R}^d$ denoting a vector of 1s. For $C = \{i, j\}$, $i < j$, under Assumption 1, Theorem 2 leads to $\tau_C(1) = \eta_C = \left(1_2^T \Sigma_{i,j}^{-1} 1_2\right)^{-1}$, with $\Sigma_{i,j}$ denoting the submatrix of $\Sigma$ corresponding to variables $i$ and $j$, provided the correlations satisfy $1 + \rho_C \neq \sum_{C':|C'|=2, C' \neq C} \rho_{C'}$; for $\delta < 1$, $\tau_C(\delta) \leq \eta_C$.

6

# 3 Methodology

## 3.1 Introduction to methodology

The coefficient $\tau_C(\delta)$ defined in Assumption 1 reveals whether the measure $\mu$ places mass on the cone $\mathbb{E}_C$. For $\mu(\mathbb{E}_C) > 0$, we assume there exists $\delta^* < 1$ such that $\tau_C(\delta^*) = 1$, but we could still have $\tau_C(\delta) < 1$ for values of $\delta < \delta^*$. For cones with $\mu(\mathbb{E}_C) = 0$, the detection problem becomes easier the further $\tau_C(\delta)$ is from 1, and since $\tau_C(\delta)$ is non-decreasing in $\delta$ it is ideal to take $\delta$ as small as possible. We therefore have a trade-off between choosing $\delta$ large enough that $\tau_C(\delta) = 1$ on cones $\mathbb{E}_C$ with extremal mass, but small enough that $\tau_C(\delta)$ is not close to 1 on those $\mathbb{E}_C$ without extremal mass. For the examples in Table 1, we could take $\delta = 0$, since the cones with $\mu(\mathbb{E}_C) > 0$ have $\tau_C(\delta) = 1$ for all $\delta \in [0, 1]$. However, the Gaussian case reveals that although $\mu(\mathbb{E}_i) > 0$, for $i = 1, \ldots, d$, we can have $\tau_i(0) < 1$, so it is necessary to take $\delta > 0$ for correct identification of cones with extremal mass.

We therefore introduce two approaches for determining the extremal dependence structure of a set of variables. In the first method we set $\delta = 0$, and apply a truncation to the variables $X$ by setting any values below some marginal threshold equal to zero. This transformation is analogous to the approach of Goix et al. (2017), who partition the non-negative orthant in a similar way, but we additionally exploit Assumption 1. In our second method, we consider $\delta > 0$ when exploiting the regular variation assumption. As well as aiming to determine the extremal dependence structure, both methods estimate the proportion of extremal mass associated with each cone $\mathbb{E}_C$.

## 3.2 Method 1: $\delta = 0$

We apply Assumption 1 with $\delta$ equal to zero by applying truncation (7) to variables $X$ for some choice of $p$. Recall that the cone $\mathbb{E}$ equals $\bigcup_{C \in 2^D \setminus \emptyset} \mathbb{E}_C$, with the components of the union disjoint and defined as in (6). We wish to partition $\mathbb{E}$ with approximations to $\mathbb{E}_C$, by creating regions where components indexed by $C$ are large and those not in $C$ are small. This is achieved via regions of the form

$$E_C = \left\{ x^* \in \mathbb{E} : x_C^* \in (-1/\log p, \infty]^{|C|}, \ x_{D \setminus C}^* \in \{0\}^{|D \setminus C|} \right\}. \tag{13}$$

Define the variable $Q = \min(X_i^* : X_i^* > 0, i = 1, \ldots, d)$, and recall that we denote $X_\wedge^C = \min_{i \in C} X_i$ and $X_\vee^{D \setminus C} = \max_{i \in D \setminus C} X_i$. Under Assumption 1, as $q \to \infty$,

$$\mathrm{pr}\left(Q > q \mid X^* \in E_C\right) \propto \mathrm{pr}\left(X_\wedge^C > q, X_\vee^{D \setminus C} < -1/\log p\right) \in \mathrm{RV}_{-1/\tau_C(0)},$$

so that $\mathrm{pr}(Q > q \mid X^* \in E_C) = L_C(q) q^{-1/\tau_C(0)}$ for some slowly varying function $L_C$. We now let $\tau_C = \tau_C(0)$, and assume that the model

$$\mathrm{pr}\left(Q > q \mid X^* \in E_C\right) = K_C q^{-1/\tau_C}, \quad q > u_C, \tag{14}$$

holds for a high threshold $u_C$, with $\tau_C \in (0, 1]$ and $K_C > 0$ for all $C \in 2^D \setminus \emptyset$. Here, the slowly varying function $L_C$ is replaced by the constant $K_C$ as a modelling assumption, removing the possibility of having $L_C(q) \to 0$ as $q \to \infty$.

Model (14) may be fitted using a censored likelihood approach. Suppose that we observe $n_C$ values $q_1, \ldots q_{n_C}$ of $Q$ in $E_C$. The censored likelihood associated with $E_C$, is

$$L_C(K_C, \tau_C) = \prod_{j=1}^{n_C} \left(1 - K_C u_C^{-1/\tau_C}\right)^{\mathbb{1}_{\{q_j \leq u_C\}}} \left(\frac{K_C}{\tau_C} q_j^{-1-1/\tau_C}\right)^{\mathbb{1}_{\{q_j > u_C\}}}, \tag{15}$$

with $u_C$ a high threshold. Analytical maximization of (15) leads to closed-form estimates of $(K_C, \tau_C)$, with the latter corresponding to the Hill estimate (Hill, 1975). In particular,

$$\hat{\tau}_C = \left(\sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}}\right)^{-1} \sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}} \log\left(\frac{q_j}{u_C}\right), \quad \hat{K}_C = \left(\frac{\sum_{j=1}^{n_C} \mathbb{1}_{\{q_j > u_C\}}}{n_C}\right) u_C^{1/\hat{\tau}_C}.$$

This estimate of $\tau_C$ can exceed 1, so we prefer to use $\min(\hat{\tau}_C, 1)$, with an appropriate change to $\hat{K}_C$. The Hill estimator for $\tau_C$ is consistent if $u_C \to \infty$, $\sum_j \mathbb{1}_{\{q_j > u_C\}} \to \infty$ and $\sum_j \mathbb{1}_{\{q_j > u_C\}}/n_C \to 0$; the assumption of $L_C(q) \sim K_C > 0$ is not required for this. The second condition ensures that the number of points in $E_C$ with $Q > u_C$ goes to infinity, and since the expected number $n_C \mathrm{pr}(Q > u_C \mid X^* \in E_C) \sim n_C K_C u_C^{-1/\tau_C}$, this entails $u_C = o(n_C^{\tau_C})$.

The method of Goix et al. (2017) produces empirical estimates of $\mathrm{pr}(X \in E_C \mid R > r_0)$, for $R = \|X\|_\infty$, and some value of $r_0$ within the range of observed values. These estimates are then assumed to hold for all $R > r_0$ and are used to approximate the limit. If the conditional probability $\mathrm{pr}(X \in E_C \mid R > r)$ changes with $r > r_0$, Goix et al. estimate this as a positive constant or as zero. In contrast, our semiparametric method allows us to estimate $\mathrm{pr}(X^* \in E_C \mid Q > q)$, for all $q$ above a high threshold, via

$$\mathrm{pr}(X^* \in E_C \mid Q > q) = \frac{\mathrm{pr}(Q > q \mid X^* \in E_C)\mathrm{pr}(X^* \in E_C)}{\displaystyle\sum_{C' \in 2^D \setminus \emptyset} \mathrm{pr}(Q > q \mid X^* \in E_{C'})\mathrm{pr}(X^* \in E_{C'})}, \quad C \in 2^D \setminus \emptyset, \tag{16}$$

with $E_C$ as in (13). Our estimate of probability (16) varies continuously with $q$, with this variation being determined by the estimated values $\hat{\tau}_C$, for $C \in 2^D \setminus \emptyset$. In situations where sub-asymptotic dependence leads to many points in a region $E_C$, but $\mu(\mathbb{E}_C) = 0$ and $\hat{\tau}_C < 1$, this extrapolation can be helpful in obtaining a better approximation to the limit. The relative merits of these differences to the approach of Goix et al., which are common to Methods 1 and 2, are illustrated in Sections 4 and 5.

The right-hand side of equation (16) consists of two types of component. We estimate terms of the form $\mathrm{pr}(X^* \in E_C)$ empirically, and we estimate those of the form $\mathrm{pr}(Q > q \mid X^* \in E_C)$ as in (14) by replacing $K_C$ and $\tau_C$ by their estimates, and evaluating for some large choice of $q$, discussed in Section 5. This approach yields an estimate for the proportion of mass in each region. We denote the estimated vector of these proportions by $\hat{p} = (\hat{p}_C : C \in 2^D \setminus \emptyset)$. In order to obtain a sparse representation of the mass on the simplex, we follow Goix et al. (2016, 2017) and ignore any mass that has been detected which is considered to be negligible; see use of parameter $\pi$, below. A summary of our method is as follows.

First, transform the data to standard Fréchet margins, and for a choice of the tuning parameter $p$, apply transformation (7). Then assign each transformed observation to a region $E_C$ as in (13), removing any all-zero points. For each region $E_C$ containing more than $m$ points, fit model (14) for a choice of threshold $u_C$, and estimate $\mathrm{pr}(X^* \in E_C \mid Q > q)$ for a large value of $q$ by equation (16). Set $\mathrm{pr}(X^* \in E_C \mid Q > q) = 0$ in the remaining regions, denoting the resulting estimate by $\hat{p}_C$. Finally, if $\hat{p}_C < \pi$, for a choice of the tuning parameter $\pi$, set $\hat{p}_C$ to zero, renormalizing the resulting vector.

The parameter $m$ ensures there are enough points to estimate the parameters on each cone $\mathbb{E}_C$. In simulations, it was found not to have a significant effect on results, so we take $m = 1$.

## 3.3 Method 2: $\delta > 0$

An alternative to setting $\delta = 0$ and partitioning the positive orthant using regions $E_C$, is to consider $\delta > 0$ in the application of Assumption 1, specifically $\mathrm{pr}\left(X_\wedge^C > t, X_\vee^{D \setminus C} \le t^\delta\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}$. However, unlike with $\delta = 0$, this does not lead directly to a univariate structure variable with tail index $1/\tau_C(\delta)$. We instead consider $\mathrm{pr}\left\{X_\wedge^C > t, X_\vee^{D \setminus C} \le (X_\wedge^C)^\delta\right\} = \mathrm{pr}\left(X_\wedge^C > t, X \in \tilde{E}_C\right)$, with $\tilde{E}_C$ defined as

$$\tilde{E}_C = \left\{x \in \mathbb{E} : x_\vee^{D \setminus C} \le \left(x_\wedge^C\right)^\delta\right\}, \ |C| < d; \qquad \tilde{E}_D = \mathbb{E} \setminus \bigcup_{C \in 2^D \setminus \emptyset : |C| < d} \tilde{E}_C,$$

for each $C \in 2^D \setminus \emptyset$. We denote the corresponding tail index as $1/\tilde{\tau}_C(\delta)$, and assume that

$$\mathrm{pr}\left(X_\wedge^C > q, X \in \tilde{E}_C\right) \in \mathrm{RV}_{-1/\tilde{\tau}_C(\delta)}, \quad q \to \infty.$$

Analogously to equation (14) of Method 1, for each region $\tilde{E}_C$, we assume the model

$$\mathrm{pr}(X_\wedge^C > q \mid X \in \tilde{E}_C) = K_C q^{-1/\tilde{\tau}_C(\delta)}, \quad q > u_C, \tag{17}$$

8

for some large threshold $u_C$, where estimates of $K_C$ and $\tilde{\tau}_C(\delta)$ are again obtained by maximizing a censored likelihood. In the Supplementary Material, we examine estimates of $\tilde{\tau}_C(\delta)$, which we find to reasonably approximate the true values of $\tau_C(\delta)$. This indicates that the indices $\tilde{\tau}_C(\delta)$ provide useful information about $\tau_C(\delta)$. We note that the regions $\tilde{E}_C$ are not disjoint. Supposing we have observations $x_1, \ldots, x_n$, we obtain an empirical estimate of $\mathrm{pr}(X \in \tilde{E}_C)$ using

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\mathbb{1}_{\{x_j \in \tilde{E}_C\}}}{\sum_{C' \in 2^D \setminus \emptyset} \mathbb{1}_{\{x_j \in \tilde{E}_{C'}\}}}, \tag{18}$$

so that the contribution of each observation sums to one. Combining (17) and (18), we estimate

$$\mathrm{pr}\left(X_\wedge^C > q, X \in \tilde{E}_C\right) = \mathrm{pr}(X_\wedge^C > q \mid X \in \tilde{E}_C)\mathrm{pr}(X \in \tilde{E}_C), \qquad C \in 2^D \setminus \emptyset, \tag{19}$$

for some large $q$. To estimate the proportion of extremal mass associated with each cone $\mathbb{E}_C$, we consider probability (19) for a given $\tilde{E}_C$ divided by the sum over all such probabilities, corresponding to $\tilde{E}_{C'}, C' \in 2^D \setminus \emptyset$. As in Method 1, the result is evaluated at a high threshold $q$, and mass estimated to be below the threshold $\pi$ is removed.

# 4 Simulation study

## 4.1 Overview and metrics

We present simulations to demonstrate Methods 1 and 2, and compare them with the approach of Goix et al. (2017). Here, we consider a max-mixture distribution involving Gaussian and extreme value logistic distributions, described in equation (21). In the Supplementary Material we present results for a special case of this, the asymmetric logistic distribution (Tawn, 1990), that is used by Goix et al. (2017) to assess the performance of their methods. The key difference between these two distributions is that the Gaussian components in the max-mixture model lead to sub-asymptotic dependence, in contrast to independence, on certain sub-cones. Hence, distinguishing between cones $\mathbb{E}_C$ with and without limiting mass is a more difficult task for our max-mixture distribution. For the classes of model we consider, it is possible to calculate the proportion of extremal mass on the various cones analytically, allowing us to compare our estimates to the true distribution of mass using the Hellinger distance.

When incorporating a cut-off for sparse representation of the measure $\mu$, as mentioned in Section 3.2, the methods can be viewed as classification techniques. Plotting receiver operating characteristic curves is a common method for testing the efficacy of classifiers (Hastie et al., 2009). To obtain such curves, the false positive rate of a method is plotted against the true positive rate, as some parameter of the method varies. In our case, the false positive rate is the proportion of cones $\mathbb{E}_C$ incorrectly detected as having mass, while the true positive rate is the proportion of correctly detected cones. To obtain our curves, we vary the threshold, $\pi$, above which estimated mass is considered non-negligible. For $\pi = 0$, all cones will be included in the estimated dependence structure, leading to the true and false positive rates both being 1, while $\pi = 1$ includes none of the cones, so both equal 0. A perfect result for a given data set and method would be a false positive rate of 0 and true positive rate of 1: the closer the curve is to the point $(0, 1)$, the better the method. This is often quantified using the area under the curve, with values closer to 1 corresponding to better methods.

Let $p = \left(p_C; C \in 2^D \setminus \emptyset\right)$ denote the true proportion of mass on each cone, and denote its estimate by $\hat{p}$. The Hellinger distance between $p$ and $\hat{p}$,

$$\mathrm{HD}(p, \hat{p}) = \frac{1}{\sqrt{2}} \left\{ \sum_{C \in 2^D \setminus \emptyset} \left( p_C^{1/2} - \hat{p}_C^{1/2} \right)^2 \right\}^{1/2}, \tag{20}$$

is used to determine the precision of the estimated proportions. In particular, $\mathrm{HD}(p, \hat{p}) \in [0, 1]$, and equals 0 if and only if $p = \hat{p}$. The closer $\mathrm{HD}(p, \hat{p})$ is to 0, the better $p$ is estimated by $\hat{p}$. Errors on small proportions are penalized more heavily than errors on large proportions. A small positive mass on a region, estimated as zero, will incur a relatively heavy penalty.

## 4.2 Max-mixture distribution

Segers (2012) shows how to construct distributions that place extremal mass on different combinations of cones. Here, we take a different approach by considering max-mixture models with asymptotic and sub-asymptotic dependence in different cones. This can be achieved by using a mixture of extreme value logistic and multivariate Gaussian copulas, a particular example of which we consider here.

Let $Z_C = (Z_{i,C} : i \in C)$ be a $|C|$-dimensional random vector with standard Fréchet marginal distributions, and $\{Z_C : C \in 2^D \setminus \emptyset\}$ be independent random vectors. Define the vector $X = (X_1, \dots, X_d)$ with components

$$X_i = \max_{C \in 2^D \setminus \emptyset : i \in C} (\theta_{i,C} Z_{i,C}), \quad \theta_{i,C} \in [0,1], \quad \sum_{C \in 2^D \setminus \emptyset : i \in C} \theta_{i,C} = 1, \tag{21}$$

for $i = 1, \dots, d$. The constraints on $\theta_{i,C}$ ensure that $X$ also has standard Fréchet margins. The random vector $Z_C$ may exhibit asymptotic dependence, in which case mass will be placed on the cone $\mathbb{E}_C$, or it may exhibit asymptotic independence, in which case mass will be placed on the cones $\mathbb{E}_i$, $i \in C$.

Here, we consider one particular five-dimensional example. We define $Z_{1,2}$ and $Z_{4,5}$ to have bivariate Gaussian copulas with correlation parameter $\rho$, and $Z_{1,2,3}$, $Z_{3,4,5}$ and $Z_{1,2,3,4,5}$ to have three-dimensional and five-dimensional extreme value logistic copulas with dependence parameter $\alpha$. The bivariate Gaussian distribution is asymptotically independent with sub-asymptotic dependence, while the logistic distribution is asymptotically dependent for $\alpha \in (0,1)$. As such, the cones with mass resulting from this construction are $\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_4, \mathbb{E}_5, \mathbb{E}_{1,2,3}, \mathbb{E}_{3,4,5}$ and $\mathbb{E}_{1,2,3,4,5}$. The Gaussian components mean that cones $\mathbb{E}_{1,2}$ and $\mathbb{E}_{4,5}$ have no mass asymptotically, but the parameter $\rho$ controls the decay rate of the mass. We assign equal mass to each of the seven charged cones by setting

$$\theta_{1,2} = (5,5)/7, \quad \theta_{4,5} = (5,5)/7,$$
$$\theta_{1,2,3} = (1,1,3)/7, \quad \theta_{3,4,5} = (3,1,1)/7, \quad \theta_{1,2,3,4,5} = (1,1,1,1,1)/7.$$

In this model, the cones with mass are fixed, in contrast to the asymmetric logistic examples in the Supplementary Material, where following Goix et al. (2017), they are chosen at random over different simulation runs. Setting $\rho = 0$ in this max-mixture distribution gives an asymmetric logistic model.
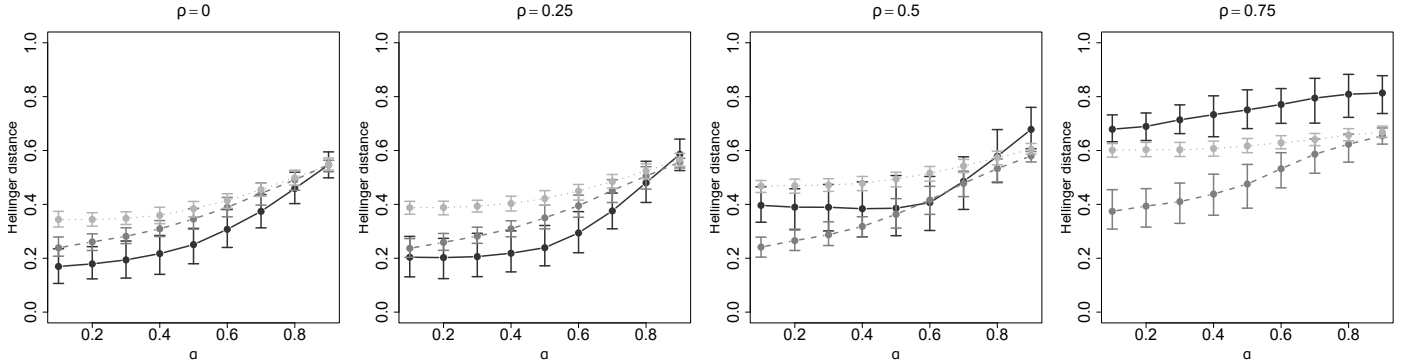


Figure 2: Mean Hellinger distance, 0.05 and 0.95 quantiles over 100 simulations. Method 1: solid lines; Method 2: dashed lines; Goix et al.: dotted lines.

Each setting in the simulation study is repeated 100 times, taking samples of size 10,000. In Method 1, we set $p = 0.5$, $u_C$ to be the 0.75 quantile of observed $Q$ values in region $E_C$ for each $C \in 2^D \setminus \emptyset$, and the value of $q$ for which we estimate $\mathrm{pr}(X^* \in E_C \mid Q > q)$ to be the 0.9999 quantile of all observed $Q$ values. In Method 2, we set $\delta = 0.5$, each threshold $u_C$ to be the 0.85 quantile of observed $X_\wedge^C$ values in region $\tilde{E}_C$, and the extrapolation level $q$ to be the 0.9999 quantile of observed values of $X$. The parameters in the method of Goix et al. (2017) are chosen to be $(\epsilon, k) = (0.1, n^{1/2})$, using notation from that paper. When calculating the Hellinger distances, we used $\pi = 0.001$ as the value above which estimated mass is considered significant in all three methods. The tuning parameters are not optimized for individual data sets, but fixed at values that we

have found to work well across a range of settings. In Section 4.3, we discuss stability plots, which could be used as a guide as to which tuning parameter values may be sensible for a given set of data. In Section 5, we consider how the estimated extremal dependence structure changes as the tuning parameters vary for a particular data set, allowing us to further examine this mass stability and choose a reasonable value of $p$ in Method 1 or $\delta$ in Method 2.

| | $\rho = 0$ | | | $\rho = 0.25$ | | | $\rho = 0.5$ | | | $\rho = 0.75$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 0.25 | 0.5 | 0.75 | 0.25 | 0.5 | 0.75 | 0.25 | 0.5 | 0.75 | 0.25 | 0.5 | 0.75 |
| Goix et al. | 100 (0.0) | 100 (0.0) | 98.0 (1.1) | 99.7 (0.4) | 99.8 (0.4) | 96.3 (1.4) | 92.3 (0.6) | 91.9 (0.5) | 90.1 (1.2) | 91.0 (1.0) | 90.1 (1.7) | 87.6 (1.2) |
| Method 1 | 100 (0.0) | 100 (0.1) | 97.7 (1.4) | 100 (0.1) | 99.9 (0.3) | 96.7 (1.2) | 97.3 (1.6) | 96.3 (1.9) | 91.5 (1.9) | 92.9 (1.0) | 90.0 (0.9) | 87.5 (0.2) |
| Method 2 | 100 (0.0) | 99.2 (0.7) | 96.0 (1.6) | 100 (0.1) | 98.9 (0.8) | 94.6 (1.8) | 99.5 (0.6) | 97.5 (1.1) | 92.7 (1.7) | 94.4 (1.9) | 92.9 (1.9) | 89.1 (2.0) |

Table 2: Average area under the receiver operating characteristic curves, given as percentages, for 100 samples from a five-dimensional mixture of bivariate Gaussian and extreme value logistic distributions; the standard deviation of each result is given in brackets.
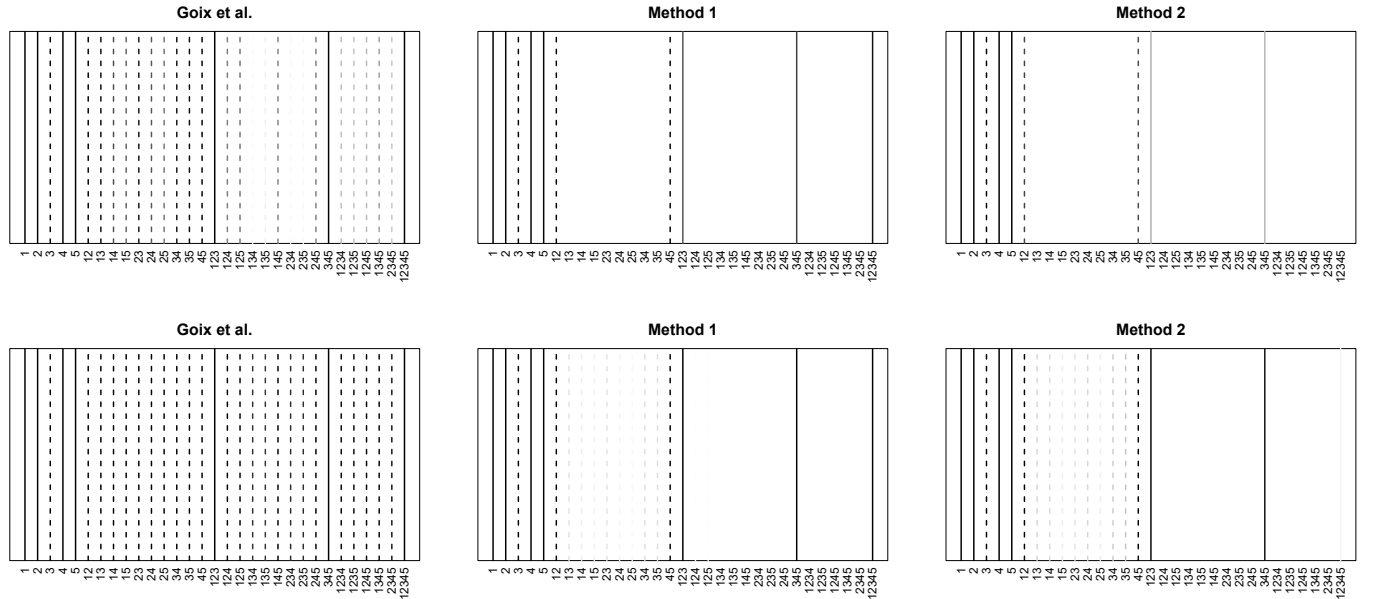


Figure 3: Plots to show the number of times each cone is assigned mass greater than $\pi = 0.01$ (top) and $\pi = 0.001$ (bottom), for $(\alpha, \rho) = (0.75, 0.5)$. Darker lines correspond to higher detection rates over 100 simulations. True cones with mass: solid lines; cones without mass; dashed lines.

In Fig. 2, we show the mean Hellinger distance achieved by each method for $\rho \in \{0, 0.25, 0.5, 0.75\}$ and $\alpha \in [0.1, 0.9]$. Results for the area under the receiver operating characteristic curves are provided in Table 2. The performance of all three methods deteriorates as the value of the correlation parameter $\rho$, or the dependence parameter $\alpha$, increases. In the former case this is due to the stronger sub-asymptotic dependence on cones without extremal mass; in the latter case, larger values of $\alpha$ in logistic component $Z_C$ lead to larger values of $\tau_{\underline{C}}(\delta)$ for $\underline{C} \subset C$, so it is harder to determine which cones $\mathbb{E}_C$ truly contain extremal mass. In terms of the Hellinger distance, Method 1 is the most successful for $\rho = 0, 0.25$, although its performance deteriorates when there is stronger correlation in the Gaussian components. Method 2 yields the best results for $\rho = 0.5, 0.75$. In terms of estimating the proportion of extremal mass associated with each cone $\mathbb{E}_C$, at least one of our proposed methods is always more successful than Goix et al. for this max-mixture model. The results in Table 2 reveal that all three methods are successful classifiers for low values of $\rho$ and $\alpha$. For $\alpha = 0.75$ and $\rho = 0, 0.25$, Method 1 and the approach of Goix et al. demonstrate similarly strong performance, while for $\rho = 0.5, 0.75$, Method 2 again provides the best results.

As a further comparison of the methods, in Fig. 3, we investigate how often each cone $\mathbb{E}_C$ is detected as

having mass above $\pi = 0.01, 0.001$ for the $(\alpha, \rho) = (0.75, 0.5)$ case. For $\pi = 0.001$, the approach of Goix et al. places mass on around three times as many cones as Methods 1 and 2, and over twice as many for the $\pi = 0.01$ case, so our methods provide sparser representations of the extremal mass that are both much closer to the truth. The reason for this difference is explained by the method of Goix et al. assuming there is extremal mass on a cone $\mathbb{E}_C$ if $\mathrm{pr}(X \in E_C \mid R > r_0) > \pi$, whereas we recognize that when $\hat{\tau}_C < 1$ or $\hat{\tau}_C(\delta) < 1$, non-limit mass can be on a cone at a finite threshold, but may progressively decrease to zero as the level of extremity of the vector variable is increased to infinity. When $\hat{\tau}_C = 1$ or $\hat{\tau}_C(\delta) = 1$, we estimate mass on cone $\mathbb{E}_C$ similarly to Goix et al. As a consequence, our approach integrates information over the entire tail to estimate which cones have limit mass, as opposed to Goix et al., who use information only at a single quantile. We also observe from Fig. 3 that Method 2 often fails to detect the cone corresponding to all five variables being large simultaneously, and places more mass on lower-dimensional cones, arising from the estimated values of $\tau_C(\delta)$. Method 1 also places mass on these lower-dimensional cones, but more often detects the true higher-dimensional cones with mass.

## 4.3 Stability plots

One way to decide on reasonable tuning parameter values for a given set of data is via a parameter stability plot. Here, we outline how to construct such a plot for an example using the max-mixture distribution of Section 4.2 with Method 2, where our aim is to obtain a sensible range of values for the tuning parameter $\delta$, by considering the region of $\delta$ where the number of cones determined as having mass is stable.
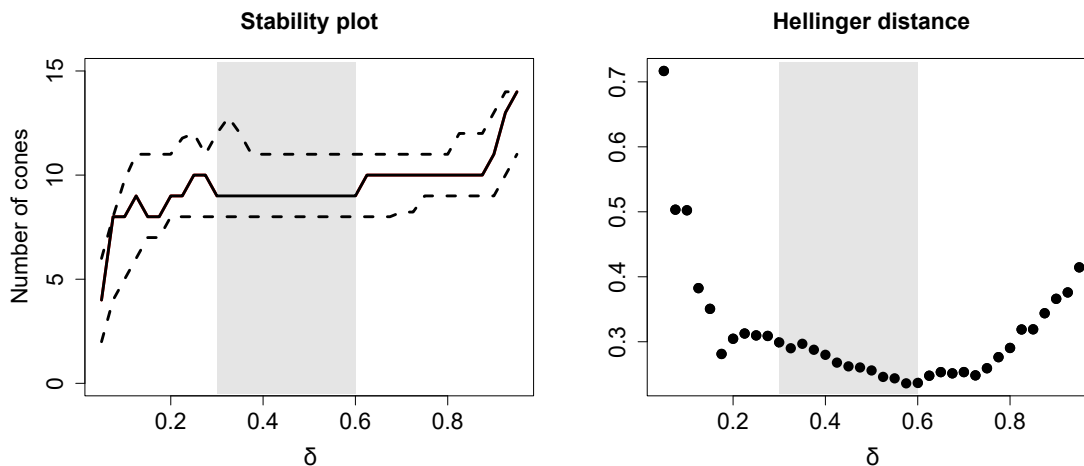


Figure 4: Stability plot (left) for Method 2, with dashed lines showing a 95% bootstrapped confidence interval for the number of cones $\mathbb{E}_C$ with mass, and a plot of the Hellinger distance (right) for each value of $\delta$. The shaded regions correspond to the stable range of tuning parameter values. Data were simulated from the max-mixture distribution of Section 4.2 with $n = 10,000$, $\alpha = 0.25$ and $\rho = 0.25$.

For $\delta \in \{0.05, 0.075, \ldots, 0.95\}$, we use Method 2 to estimate the proportion of extremal mass on each cone, and find the number of cones whose estimated mass is greater than $\pi = 0.001$ in each case. The remaining parameters are fixed as in Section 4.2. Figure 4 shows the estimates of the number of cones, with a 95% confidence interval constructed from 250 bootstrapped samples: these constitute our stability plot. Analogous plots can be created to choose $p$ in Method 1, or in each case to choose $\pi$. In practice, the choice of threshold $\pi$ should depend on the dimension of the data; this is not explored here.

The number of cones detected as having mass is most stable for values of $\delta$ between 0.3 and 0.6, indicated by the shaded regions in Fig. 4, suggesting values of $\delta$ in this range may be appropriate for this sample. The right-hand panel of Fig. 4 shows the Hellinger distance corresponding to the set of estimated proportions obtained for each value of $\delta$. For this particular sample, although values of $\delta$ within the stable range slightly overestimate the number of cones with mass, the smallest Hellinger distance occurs for a value of $\delta$ within

the stable range, and the Hellinger distance is reasonably consistent across these tuning parameter values. In practice, the true proportions on each cone $\mathbb{E}_C$ are unknown, so Hellinger plots cannot be constructed; the plot here supports the idea of using stability plots in choosing suitable tuning parameter values. There is no guarantee that stability plots will find the optimal tuning parameter values, but they do offer some insight into tuning parameter optimization. Consideration of the context of the problem may be useful in determining whether it is reasonable for extremal mass to be placed on particular combinations of cones, and such insight could facilitate the choice of different $p$ or $\delta$ values for different cones $\mathbb{E}_C$.

## 5   River flow data

We apply Methods 1 and 2 to daily mean river flow readings from 1980 to 2013, measured in cubic metres per second, at five gauging stations in the North West of England. These data are available from the Centre for Ecology and Hydrology at `nrfa.ceh.ac.uk` (Morris and Flavin, 1990, 1994). Estimates of the extremal dependence structure of the flows could be used to aid model selection, or one could carry out density estimation on each cone $\mathbb{E}_C$ to give an overall model. The locations of the five gauges are shown in Fig. 5; the labels assigned to each location will be used to describe the dependence structures estimated in this section. Figure 5 also illustrates the boundaries of the catchments associated with each gauge. These catchments demonstrate the areas from which surface water, usually as a result of precipitation, will drain to each gauge. The spatial dependence of river flow is studied by Keef et al. (2013) and Asadi et al. (2015). As high river flow is mainly caused by heavy rainfall, we may observe extreme river flow readings at several locations simultaneously if they are affected by the same extreme weather event. Gauges with adjacent or overlapping catchments are expected to take their largest values simultaneously, with stronger dependence between gauges that are closer together.



Figure 5: Locations of the river flow gauges, labelled A to E, and corresponding catchment boundaries.

Table 3 shows the percentage of the extremal mass assigned to each cone for tuning parameter values $p \in \{0.7, 0.725, \ldots, 0.975\}$ and $\delta \in \{0.2, 0.25, \ldots, 0.75\}$. We set $\pi = 0.01$ to be the threshold below which the proportion of mass is deemed negligible, and the extrapolation levels $q$ to be the 0.999 quantile of the observed $Q$ and $X$ values in Methods 1 and 2, respectively. Remaining parameters are fixed as in Section 4.2. By observing how the estimated dependence structure changes over a range of tuning parameter values, we aim to find a 'stable region' in which the results are most reliable. A further consideration is whether the tuning parameters give a feasible estimate of the extremal dependence structure. In particular, each variable should be represented on at least one cone, and moment constraint (3) should be taken into account. For Method 2, Table 3 indicates that for $\delta \geq 0.45$, the cone corresponding to location E is assigned more than 20% of the

extremal mass, which is not possible due to the moment constraint. Feasible stable regions are demonstrated by the shaded regions in Table 3. For Method 2, one could also look for a value of $\delta$ that give estimates of $\tau_C(\delta)$ satisfying $\max_{C:C \supseteq i} \hat{\tau}_C(\delta) = 1$, subject to estimation uncertainty, for every $i = 1, \ldots, d$.

| $p$ | B | C | D | E | AC | AD | BC | ABC | ACD | ADE | ABCD | ACDE | ABCDE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.700 | 2 | | | | | | | | | | 3 | | 95 |
| 0.725 | 2 | | | | | | | | | | 4 | | 93 |
| 0.750 | 1 | | | | | | | | | | 4 | | 95 |
| 0.775 | 1 | | | | | | | | | | 11 | | 88 |
| 0.800 | 7 | | | | | | | | | | 11 | | 82 |
| 0.825 | 10 | | 3 | | | | | | | | 16 | | 71 |
| 0.850 | 13 | | 7 | | | | | | | | 17 | | 63 |
| 0.875 | 12 | | 12 | | | | | | 1 | | 19 | | 55 |
| 0.900 | 8 | | 31 | | | | | | | | 15 | | 46 |
| 0.925 | 6 | | 57 | | | 2 | 2 | | | | 15 | | 19 |
| 0.950 | 5 | | 52 | | | 4 | 2 | | | | 16 | | 21 |
| 0.975 | 14 | 1 | 73 | 2 | 1 | | | | | | | 2 | 6 |

| $\delta$ | B | C | D | E | AC | AD | BC | ABC | ACD | ADE | ABCD | ACDE | ABCDE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.20 | | | | | | | | | | | 1 | | 99 |
| 0.25 | 4 | | | | | | | | | | 15 | | 81 |
| 0.30 | 3 | 1 | | | | | | 7 | | | 24 | | 65 |
| 0.35 | 8 | | 1 | | | | | 9 | | | 33 | | 49 |
| 0.40 | 14 | | 3 | | | | | 2 | | | 49 | | 30 |
| 0.45 | 27 | 1 | 26 | | | 1 | | 4 | | | 29 | | 12 |
| 0.50 | 22 | 1 | 49 | | | 1 | | 2 | | | 15 | 2 | 7 |
| 0.55 | 16 | | 52 | | 1 | 1 | | 3 | | | 18 | 4 | 6 |
| 0.60 | 13 | | 55 | | | | | 2 | 2 | | 19 | 5 | 4 |
| 0.65 | 16 | 3 | 48 | | 2 | 2 | | 3 | | | 17 | 6 | 3 |
| 0.70 | 16 | 2 | 4 | 45 | 2 | 3 | 3 | 4 | | | 13 | 6 | 3 |
| 0.75 | 19 | 3 | 4 | 40 | 1 | 2 | 3 | 5 | 4 | 1 | 11 | 4 | 2 |

Table 3: The percentage of mass assigned to each sub-cone for varying values of the tuning parameters in Method 1 (left) and Method 2 (right). The grey regions demonstrate the feasible stable ranges.

Focusing on tuning parameter values within each of the stable regions in Table 3, Method 1 suggests the dependence structure to be {B, E, ABCD, ABCDE}, while Method 2 suggests {B, E, ABC, ABCD, ABCDE}. All the cones detected by Method 2 are either also detected by Method 1, or are neighbours of cones detected by Method 1, showing there is some agreement between the methods. If we had used a higher threshold for the negligible mass, say $\pi = 0.1$, for tuning parameter values in the stable region, both methods would have detected the structure {B, ABCD, ABCDE}. We also investigated the behaviour of the methods using the 0.99 and 0.9999 quantiles for extrapolation level $q$. For both methods, the set of cones estimated as having mass was stable, but for the lower quantile Method 2 placed less mass on ABCDE, and there was more mass assigned to this cone at the higher quantile.

The subsets of locations detected as having simultaneously high river flows seem feasible when considering the geographic positions of the gauging stations. For instance, both methods suggest mass on ABCD; as station E lies towards the edge of the region under consideration, it is possible for weather events to affect only the other four locations. Both methods also suggest that locations B and E can experience high river flows individually; this seems reasonable as they lie at the edge of region we consider. The catchment of gauge C lies entirely within the catchment of gauge A. We observe that location C occurs with location A in the subsets of sites determined to take their largest values simultaneously, which may be a consequence of this nested structure.

| $p$ | A | B | C | D | AC | AD | BC | ABC | ACD | ABCD |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.700 | | | | | | | | | | 100 |
| 0.725 | | | | | | | | | | 100 |
| 0.750 | | | | | | | | | | 100 |
| 0.775 | | | | | | | | | | 100 |
| 0.800 | | | | | | | | | | 100 |
| 0.825 | | 3 | | | | | | | | 97 |
| 0.850 | | 3 | | | | | | | | 97 |
| 0.875 | | 4 | | | | | | | | 96 |
| 0.900 | | 2 | | | | | | | | 98 |
| 0.925 | | 5 | | | | | | 1 | | 94 |
| 0.950 | | 7 | | | | | 2 | 5 | 2 | 84 |
| 0.975 | 35 | 3 | | | 1 | 2 | | 7 | | 52 |

| $\delta$ | A | B | C | D | AC | AD | BC | ABC | ACD | ABCD |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.20 | | | | | | | | | | 100 |
| 0.25 | | | | | | | | | | 100 |
| 0.30 | | 2 | | | | | | | | 98 |
| 0.35 | | 4 | | | | | | | | 96 |
| 0.40 | | 9 | | | | | | | | 91 |
| 0.45 | | 23 | 1 | | | | | 2 | 1 | 72 |
| 0.50 | | 28 | 2 | | 1 | 1 | 1 | 6 | | 60 |
| 0.55 | | 19 | 1 | 4 | 2 | 2 | 3 | 14 | | 55 |
| 0.60 | | 22 | 1 | 4 | 5 | 2 | 4 | 17 | | 45 |
| 0.65 | | 26 | 2 | 5 | 6 | 3 | 5 | 18 | | 36 |
| 0.70 | | 26 | 3 | 6 | 5 | 4 | 6 | 20 | | 30 |
| 0.75 | 1 | 28 | 4 | 7 | 1 | 6 | 6 | 9 | 17 | 20 |

Table 4: Estimated percentage of extremal mass on to each sub-cone when considering locations A-D, for varying values of the tuning parameters in Method 1 (left) and Method 2 (right).

To assess whether our methods are self-consistent across different dimensions, Table 4 shows similar results for locations A, B, C and D. We would expect the subsets of locations deemed to be simultaneously large to be the same as in Table 3 if we ignore location E. Considering the same tuning parameter values as for Table 3, we see that the extremal dependence structures are estimated to be {B, ABCD} for both methods. For Method 1, this is the set of cones we would expect based on the five-dimensional results. For Method 2, we would also expect to detect the cone labelled ABC, although this was only assigned a relatively small proportion of the mass in the five-dimensional case.

Tables 3 and 4 demonstrate the importance of tuning parameter selection in Methods 1 and 2. As $p$ or $\delta$

increase, we are more likely to detect mass on the one-dimensional cones, or cones corresponding to subsets of the variables with low cardinality. Likewise, for low values of $p$ or $\delta$, we assign more extremal mass to the cone representing all variables being simultaneously extreme. In practice, we should consider the feasibility of the detected dependence structures, as well as the stability of the regions determined to have extremal mass as $p$ or $\delta$ vary. Our methods could be used to impose structure in more complete models for multivariate extremes. Even if a handful of different options look plausible with some variation in $p$ or $\delta$, this is still a huge reduction over the full set of possibilities.

## Acknowledgement

## Supplementary Material

Supplementary Material includes calculations of $\tau_C(\delta)$ for copulas in Table 1, simulation results for estimates of $\tau_C(\delta)$ in Method 2, plots of simulation results for the max-mixture distribution of Section 4.2, and additional simulation results for the asymmetric logistic model.

## Appendix

### Calculation of $\tau_C(\delta)$ for a bivariate extreme value distribution

We determine the value of $\tau_C(\delta)$, defined in (10), by establishing the index of regular variation of

$$\mathrm{pr}\left(X_i > t, i \in C; X_j < t^\delta, j \in D \setminus C\right).$$

Here, we calculate $\tau_1(\delta)$, $\tau_2(\delta)$ and $\tau_{1,2}$ for a bivariate extreme value distribution, with distribution function given in (11). The exponent measure $V$ can be written as

$$V(x,y) = \frac{2}{y}\int_0^1 (1-w)dH(w) - \frac{2}{y}\int_{\frac{x}{x+y}}^1 (1-w)h(w)dw + \frac{2}{x}\int_{\frac{x}{x+y}}^1 wh(w)dw + \frac{2\theta_1}{x}.$$

To study $\tau_1(\delta)$, suppose that $h(w) \sim c_1(1-w)^{s_1}$ as $w \to 1$, for $s_1 > -1$. For $x \to \infty$ and $y = o(x)$, applying Karamata's theorem (Resnick, 2007, Theorem 2.1), we have

$$
\begin{aligned}
V(x,y) &= \frac{1}{y} - \frac{2c_1}{y(s_1+2)}\left(\frac{y}{x+y}\right)^{s_1+2}\{1+o(1)\} + \frac{2c_1}{x(s_1+1)}\left(\frac{y}{x+y}\right)^{s_1+1}\{1+o(1)\} + \frac{2\theta_1}{x} \\
&= \frac{1}{y} + 2c_1\left(\frac{y}{x+y}\right)^{s_1+1}\left\{\frac{1}{x(s_1+1)} - \frac{1}{(s_1+2)(x+y)}\right\}\{1+o(1)\} + \frac{2\theta_1}{x} \\
&= \frac{1}{y} + \frac{2c_1 y^{s_1+1} x^{-(s_1+2)}}{(s_1+1)(s_1+2)}\{1+o(1)\} + \frac{2\theta_1}{x}.
\end{aligned}
$$

By this result,

$$
\begin{aligned}
\mathrm{pr}\left(X_1 > t, X_2 < t^\delta\right) &= \mathrm{pr}\left(X_2 < t^\delta\right) - \mathrm{pr}\left(X_1 < t, X_2 < t^\delta\right) = \exp\left(-t^{-\delta}\right) - \exp\left\{-V\left(t, t^\delta\right)\right\} \\
&= \exp\left(-t^{-\delta}\right) - \exp\left\{-\frac{1}{t^\delta} - \frac{2c_1 t^{\delta(s_1+1)} t^{-(s_1+2)}}{(s_1+1)(s_1+2)}\{1+o(1)\} - \frac{2\theta_1}{t}\right\}
\end{aligned}
$$

$$
= \left\{ 1 - t^{-\delta} + o\left( t^{-\delta} \right) \right\}
$$

$$
\cdot \left( 1 - \left[ 1 - \frac{2c_1 t^{\delta(s_1+1)} t^{-(s_1+2)}}{(s_1+1)(s_1+2)} + o\left\{ t^{\delta(s_1+1)-(s_1+2)} \right\} \right] \left\{ 1 - 2\theta_1 t^{-1} + o\left( t^{-1} \right) \right\} \right)
$$

$$
= \left\{ 2\theta_1 t^{-1} + \frac{2c_1 t^{\delta(s_1+1)-(s_1+2)}}{(s_1+1)(s_1+2)} \right\} \left\{ 1 + o(1) \right\}.
$$

If $\theta_1 > 0$, i.e., the spectral measure places mass on $\{1\}$, we see that $\mathrm{pr}\left( X_1 > t, X_2 < t^\delta \right) \sim 2\theta_1 t^{-1}$ as $t \to \infty$, hence $\tau_1(\delta) = 1$ for all $\delta \in [0,1]$. If $\theta_1 = 0$, we have $\tau_1(\delta) = \{(s_1+2) - \delta(s_1+1)\}^{-1}$, which increases from $(s_1+2)^{-1}$ at $\delta = 0$ to 1 at $\delta = 1$. By similar calculations, if $h(w) \sim c_2 w^{s_2}$ as $w \to 0$ for $s_2 > -1$, we have $\tau_2(\delta) = 1$ if $\theta_2 > 0$, and $\tau_2(\delta) = \{(s_2+2) - \delta(s_2+1)\}^{-1}$ otherwise. Since $\tau_{1,2} = \eta_{1,2}$, we have $\tau_{1,2} = 1$ if $\theta_1 + \theta_2 < 1$, and $\tau_{1,2} = 1/2$ if $\theta_1 + \theta_2 = 1$.

# Supplementary Material

## A   Calculation of $\tau_C(\delta)$

### A.1   Overview

In the Appendix, we derived $\tau_1(\delta)$, $\tau_2(\delta)$ and $\tau_{1,2}$ for a particular subclass of bivariate extreme value distribution. Here, we present further calculations of $\tau_C(\delta)$ for several trivariate copula models.

In general, there are two cases to consider: $\delta = 0$ and $\delta > 0$. For many models, the asymptotic relations we study will differ by a constant in these two cases, while the tail index remains the same. For this reason, we focus on $\delta > 0$, and present $\delta = 0$ calculations separately only when the slowly varying function is no longer a constant, but instead varies with $t$.

### A.2   Independence copula

We begin by considering the case where all three variables $X_1, X_2, X_3$ are independent. To calculate the value of $\tau_C(\delta)$, we need to determine the index of regular variation of

$$\mathrm{pr}\left(X_i > t, i \in C; X_j < t^\delta, j \in D \setminus C\right).$$

In the independence case, this is equivalent to

$$\mathrm{pr}\left(X_i > t\right)^{|C|}\mathrm{pr}\left(X_i < t^\delta\right)^{|D\setminus C|} = \left(1 - e^{-1/t}\right)^{|C|}\left(e^{-1/t^\delta}\right)^{|D\setminus C|} \sim t^{-|C|},$$

so that $\tau_C(\delta) = 1/|C|$, which does not depend on the value of $\delta$. That is, $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$, $\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 1/2$, and $\tau_{1,2,3} = 1/3$.

### A.3   Trivariate logistic distribution

The trivariate extreme value logistic distribution belongs to the class of trivariate extreme value distributions. The exponent measure of the logistic distribution has the form

$$V(x, y, z) = \left(x^{-1/\alpha} + y^{-1/\alpha} + z^{-1/\alpha}\right)^\alpha, \tag{22}$$

for $\alpha \in (0, 1]$. Since $\alpha = 1$ corresponds to the independence case, we restrict our calculations to $\alpha \in (0, 1)$. This distribution exhibits asymptotic dependence, with all limiting mass on $\mathbb{E}_{1,2,3}$. Since $\tau_{1,2,3} = \eta_{1,2,3} = 1$, our interest lies with the values of $\tau_C(\delta)$ for $|C| = 1$ and $|C| = 2$, and we consider each of these in turn.

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. In a similar approach to the bivariate case, we calculate $\tau_1(\delta)$ by considering

$$\mathrm{pr}\left(X_1 > t, X_2 < t^\delta, X_3 < t^\delta\right) = \mathrm{pr}\left(X_2 < t^\delta, X_3 < t^\delta\right) - \mathrm{pr}\left(X_1 < t, X_2 < t^\delta, X_3 < t^\delta\right)$$

$$= \exp\left(-2^\alpha t^{-\delta}\right) - \exp\left[-2^\alpha t^{-\delta}\left\{1 + 2^{-1}t^{(\delta-1)/\alpha}\right\}^\alpha\right]$$

$$= 1 - 2^\alpha t^{-\delta} + O(t^{-2\delta}) - \exp\left(-2^\alpha t^{-\delta}\left[1 + 2^{-1}\alpha t^{(\delta-1)/\alpha} + O\left\{t^{2(\delta-1)/\alpha}\right\}\right]\right)$$

$$= 2^{\alpha-1}\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O(t^{-2\delta}) + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\} \sim 2^{\alpha-1}\alpha t^{(\delta-1-\alpha\delta)/\alpha},$$

yielding $\tau_1(\delta) = \alpha/(1 + \alpha\delta - \delta)$. By similar calculations, we have $\tau_2(\delta) = \tau_3(\delta) = \alpha/(1 + \alpha\delta - \delta)$, which increase from $\alpha < 1$ at $\delta = 0$ to 1 at $\delta = 1$.

$|C| = 2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We carry out a similar calculation to find the value of $\tau_{1,2}(\delta)$. Here, we have

$$\mathrm{pr}\left(X_1 > t, X_2 > t, X_3 < t^\delta\right)$$

$$= \mathrm{pr}\left(X_3 < t^\delta\right) - \mathrm{pr}\left(X_1 < t, X_3 < t^\delta\right) - \mathrm{pr}\left(X_2 < t, X_3 < t^\delta\right)$$

$$+ \operatorname{pr}\left(X_1 < t, X_2 < t, X_3 < t^\delta\right)$$

$$= \exp\left(-t^{-\delta}\right) - 2\exp\left\{-\left(t^{-1/\alpha} + t^{-\delta/\alpha}\right)^\alpha\right\} + \exp\left\{-\left(2t^{-1/\alpha} + t^{-\delta/\alpha}\right)^\alpha\right\}$$

$$= \exp\left(-t^{-\delta}\right)\left\{1 - 2\exp\left(-t^{-\delta}\left[\alpha t^{(\delta-1)/\alpha} + \frac{1}{2}\alpha(\alpha-1)t^{2(\delta-1)/\alpha} + O\left\{t^{3(\delta-1)/\alpha}\right\}\right]\right)\right.$$

$$\left. + \exp\left(-t^{-\delta}\left[2\alpha t^{(\delta-1)/\alpha} + 2\alpha(\alpha-1)t^{2(\delta-1)/\alpha} + O\left\{t^{3(\delta-1)/\alpha}\right\}\right]\right)\right\}$$

$$= \left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha(1-\alpha)t^{(2\delta-2-\alpha\delta)\alpha} + O\left\{t^{(3\delta-3-\alpha\delta)/\alpha}\right\}\right]$$

$$\sim \alpha(1-\alpha)\,t^{(2\delta-2-\alpha\delta)/\alpha}.$$

This implies that $\tau_{1,2}(\delta) = \alpha/(2 + \alpha\delta - 2\delta)$, which varies from $\alpha/2$ at $\delta = 0$ to $1$ at $\delta = 1$. Similarly, we have $\tau_{1,3}(\delta) = \tau_{2,3}(\delta) = \alpha/(2 + \alpha\delta - 2\delta)$.

These calculations reveal different indices of regular variation on cones $\mathbb{E}_C$ with $|C| = 1, 2$ in the trivariate logistic case.

## A.4   Trivariate distribution with extremal mass on one vertex and one edge

Now we consider a trivariate example where the extremal mass is placed on one cone $\mathbb{E}_C$ with $|C| = 1$, and another with $|C| = 2$. This can be achieved by taking $(X_1, X_2)$ to have a bivariate extreme value logistic distribution, and $X_3$ to be a standard Fréchet random variable independent of $(X_1, X_2)$. The exponent measure in this case has the form

$$V(x, y, z) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^\alpha + z^{-1}, \quad \alpha \in (0, 1).$$

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. We first consider the index of regular variation on the cone $\mathbb{E}_1$. Following a similar procedure to previously, and exploiting the independence of $(X_1, X_2)$ and $X_3$, we have

$$\operatorname{pr}\left(X_1 > t, X_2 < t^\delta, X_3 < t^\delta\right) = \operatorname{pr}\left(X_3 < t^\delta\right)\left\{\operatorname{pr}\left(X_2 < t^\delta\right) - \operatorname{pr}\left(X_1 < t, X_2 < t^\delta\right)\right\}$$

$$= \exp\left(-t^{-\delta}\right)\left(\exp\left(-t^{-\delta}\right) - \exp\left[-t^{-\delta}\left\{1 + t^{(\delta-1)/\alpha}\right\}^\alpha\right]\right)$$

$$= \exp\left(-2t^{-\delta}\right)\left\{1 - \exp\left(-t^{-\delta}\left[\alpha t^{(\delta-1)/\alpha} + O\left\{t^{2(\delta-1)/\alpha}\right\}\right]\right)\right\}$$

$$= \left\{1 - 2t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right] \sim \alpha t^{(\delta-1-\alpha\delta)/\alpha},$$

revealing that $\tau_1(\delta) = \alpha/(1 + \alpha\delta - \delta)$. By similar calculations, $\tau_2(\delta) = \alpha/(1 + \alpha\delta - \delta)$. For the cone $\mathbb{E}_3$, we have $\tau_3(\delta) = 1$, since

$$\operatorname{pr}\left(X_1 < t^\delta, X_2 < t^\delta, X_3 > t\right) = \operatorname{pr}\left(X_3 > t\right)\operatorname{pr}\left(X_1 < t^\delta, X_2 < t^\delta\right)$$

$$= \left(1 - e^{-1/t}\right)\exp\left\{-\left(t^{-\delta/\alpha} + t^{-\delta/\alpha}\right)^\alpha\right\} = \left(1 - e^{-1/t}\right)\exp\left(-2^\alpha t^{-\delta}\right)$$

$$= \left\{t^{-1} + O\left(t^{-2}\right)\right\}\left\{1 - 2^\alpha t^{-\delta} + O\left(t^{-2\delta}\right)\right\} \sim t^{-1}.$$

$|C| = 2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We begin by showing that $\tau_{1,2}(\delta) = 1$. We have

$$\operatorname{pr}\left(X_1 > t, X_2 > t, X_3 < t^\delta\right)$$

$$= \operatorname{pr}\left(X_3 < t^\delta\right)\left\{1 - \operatorname{pr}\left(X_1 < t\right) - \operatorname{pr}\left(X_2 < t\right) + \operatorname{pr}\left(X_1 < t, X_2 < t\right)\right\}$$

$$= \exp\left(-t^{-\delta}\right)\left\{1 - 2\exp\left(-1/t\right) + \exp\left(-2^\alpha t^{-1}\right)\right\}$$

$$= \left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left\{(2 - 2^\alpha)t^{-1} + O(t^{-2})\right\} \sim (2 - 2^\alpha)t^{-1}.$$

18

Next, we consider the cone $\mathbb{E}_{1,3}$. In this case, we have

$$\mathrm{pr}\left(X_1 > t, X_2 < t^\delta, X_3 > t\right) = \mathrm{pr}\left(X_3 > t\right)\left\{\mathrm{pr}\left(X_2 < t^\delta\right) - \mathrm{pr}\left(X_1 < t, X_2 < t^\delta\right)\right\}$$

$$= \{1 - \exp\left(-1/t\right)\}\left[\exp\left(-t^{-\delta}\right) - \exp\left\{-\left(t^{-1/\alpha} + t^{-\delta/\alpha}\right)^\alpha\right\}\right]$$

$$= \{t^{-1} + O\left(t^{-2}\right)\}\exp\left(-t^{-\delta}\right)\left(1 - \exp\left[-\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right]\right)$$

$$= \{t^{-1} + O\left(t^{-2}\right)\}\left\{1 - t^{-\delta} + O\left(t^{-2\delta}\right)\right\}\left[\alpha t^{(\delta-1-\alpha\delta)/\alpha} + O\left\{t^{(2\delta-2-\alpha\delta)/\alpha}\right\}\right]$$

$$\sim \alpha t^{(\delta-1-\alpha\delta-\alpha)/\alpha},$$

so we have $\tau_{1,3}(\delta) = \alpha/(\alpha\delta + 1 + \alpha - \delta)$. Again by symmetry, $\tau_{2,3}(\delta) = \alpha/(\alpha\delta + 1 + \alpha - \delta)$. These indices vary from $\alpha/(1 + \alpha)$ at $\delta = 0$ to $1/2$ at $\delta = 1$.

$|C| = 3$: $\tau_{1,2,3}$. Finally, we consider the cone $\mathbb{E}_{1,2,3}$, where

$$\mathrm{pr}\left(X_1 > t, X_2 > t, X_3 > t\right)$$

$$= \mathrm{pr}\left(X_3 > t\right)\left\{1 - \mathrm{pr}\left(X_1 < t\right) - \mathrm{pr}\left(X_2 < t\right) + \mathrm{pr}\left(X_1 < t, X_2 < t\right)\right\}$$

$$= \{1 - \exp\left(-1/t\right)\}\left\{1 - 2\exp\left(-1/t\right) + \exp\left(-2^\alpha t^{-1}\right)\right\}$$

$$= \{t^{-1} + O\left(t^{-2}\right)\}\left\{(2 - 2^\alpha)t^{-1} + O\left(t^{-2}\right)\right\} \sim (2 - 2^\alpha)t^{-2},$$

i.e., $\tau_{1,2,3} = 1/2$.

## A.5 Trivariate inverted logistic distribution

Next, we consider an inverted trivariate extreme value distribution, defined via its distribution function

$$\mathrm{pr}(X_1 < x, X_2 < y, X_3 < z)$$

$$= 1 - \left\{F_{X_1}(x') + F_{X_2}(y') + F_{X_3}(z')\right\}$$

$$+ \left\{F_{X_1,X_2}(x', y') + F_{X_1,X_3}(x', z') + F_{X_2,X_3}(y', z')\right\} - F_{X_1,X_2,X_3}(x', y', z'),$$

where $F_{X_1,X_2,X_3}$ denotes the corresponding trivariate extreme value distribution function; $F_{X_1,X_2}$, $F_{X_1,X_3}$ and $F_{X_2,X_3}$ are the corresponding bivariate distribution functions; $F_{X_1}$, $F_{X_2}$ and $F_{X_3}$ are the marginal distributions of $X_1$, $X_2$ and $X_3$; and noting that $-\log\left(1 - e^{-1/x}\right) = -\log\left\{x^{-1} + O(x^{-2})\right\}$,

$$x' = -\frac{1}{\log\left(1 - e^{-1/x}\right)} \sim \frac{1}{\log x},$$

as $x \to \infty$, with $y'$, $z'$ defined analogously. In the case of the trivariate inverted logistic distribution, which we focus on here, $F_{X_1,X_2,X_3}(x, y, z) = \exp\left\{-V(x, y, z)\right\}$, for $V$ defined as in (22). The inverted logistic distribution exhibits asymptotic independence, placing all extremal mass on the cones $\mathbb{E}_C$ with $|C| = 1$. We will show that $\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$ for this model, and then calculate $\tau_{1,2}(\delta), \tau_{1,3}(\delta), \tau_{2,3}(\delta)$ and $\tau_{1,2,3}$.

$|C| = 1$: $\tau_1(\delta)$, $\tau_2(\delta)$, $\tau_3(\delta)$. To begin, we focus on calculating $\tau_1(\delta)$ for $\delta > 0$, by considering

$$\mathrm{pr}\left(X_1 > t, X_2 < t^\delta, X_3 < t^\delta\right)$$

$$= \mathrm{pr}\left(X_1 > t\right) - \mathrm{pr}\left(X_1 > t, X_2 > t^\delta\right)$$

$$\quad - \mathrm{pr}\left(X_1 > t, X_3 > t^\delta\right) + \mathrm{pr}\left(X_1 > t, X_2 > t^\delta, X_3 > t^\delta\right)$$

$$= \left(1 - e^{-1/t}\right) - 2\exp\left(-\left[\{\log t + O\left(t^{-1}\right)\}^{1/\alpha} + \{\delta \log t + O\left(t^{-2\delta}\right)\}^{1/\alpha}\right]^\alpha\right)$$

$$\quad + \exp\left(-\left[\{\log t + O\left(t^{-1}\right)\}^{1/\alpha} + 2\{\delta \log t + O\left(t^{-2\delta}\right)\}^{1/\alpha}\right]^\alpha\right)$$

$$= \{t^{-1} + O\left(t^{-2}\right)\} - 2\exp\left[-\left(1 + \delta^{1/\alpha}\right)^\alpha \log t + O\left\{(t \log t)^{-1}\right\}\right]$$

19

$$+ \exp\left[ -\left(1 + 2\delta^{1/\alpha}\right)^\alpha \log t + O\left\{(t\log t)^{-1}\right\}\right]$$
$$= \left\{ t^{-1} - 2t^{-\left(1+\delta^{1/\alpha}\right)^\alpha} + t^{-\left(1+2\delta^{1/\alpha}\right)^\alpha}\right\}\{1 + o(1)\} \sim t^{-1},$$

so we have $\tau_1(\delta) = 1$ for $\delta > 0$.

For $\delta = 0$, we consider variables $X_1^*, X_2^*, X_3^*$ defined via truncation (7), and study

$$\mathrm{pr}\left(X_1^* > t, X_2^* = 0, X_3 = 0\right) = \mathrm{pr}\left(X_1 > t, X_2 < -1/\log p, X_3 < -1/\log p\right)$$
$$= \mathrm{pr}\left(X_1 > t\right) - \mathrm{pr}\left(X_1 > t, X_2 > -1/\log p\right) - \mathrm{pr}\left(X_1 > t, X_3 > -1/\log p\right)$$
$$\quad + \mathrm{pr}\left(X_1 > t, X_2 > -1/\log p, X_3 > -1/\log p\right)$$
$$= \left\{t^{-1} + O\left(t^{-2}\right)\right\} - 2\exp\left(-\left[\{\log t + O\left(t^{-1}\right)\}^{1/\alpha} + \{-\log(1-p)\}^{1/\alpha}\right]^\alpha\right)$$
$$\quad + \exp\left(-\left[\{\log t + O\left(t^{-1}\right)\}^{1/\alpha} + 2\{-\log(1-p)\}^{1/\alpha}\right]^\alpha\right)$$
$$= \left\{t^{-1} + O\left(t^{-2}\right)\right\} - 2\exp\left\{-\{\log t + O\left(t^{-1}\right)\}\left(1 + \left[\frac{-\log(1-p)}{\{\log t + O\left(t^{-1}\right)\}}\right]^{1/\alpha}\right)^\alpha\right\}$$
$$\quad + \exp\left\{-\{\log t + O\left(t^{-1}\right)\}\left(1 + 2\left[\frac{-\log(1-p)}{\{\log t + O\left(t^{-1}\right)\}}\right]^{1/\alpha}\right)^\alpha\right\}$$
$$= \left\{t^{-1} + O\left(t^{-2}\right)\right\}$$
$$\quad - 2\exp\left[-\{\log t + O\left(t^{-1}\right)\} - \alpha\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{1/\alpha-1}}\right.$$
$$\quad \left. -\frac{\alpha(\alpha-1)}{2}\frac{\{-\log(1-p)\}^{2/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\}\right]$$
$$\quad + \exp\left[-\{\log t + O\left(t^{-1}\right)\} - 2\alpha\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{1/\alpha-1}}\right.$$
$$\quad \left. - 2\alpha(\alpha-1)\frac{\{-\log(1-p)\}^{2/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\}\right]$$
$$= \left\{t^{-1} + O\left(t^{-2}\right)\right\}$$
$$\quad + t^{-1}\left(-2\exp\left[-\alpha\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{1/\alpha-1}}\right.\right.$$
$$\quad \left. +\frac{\alpha(1-\alpha)}{2}\frac{\{-\log(1-p)\}^{2/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\} + O\left(t^{-1}\right)\right]$$
$$\quad + \exp\left[-2\alpha\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{1/\alpha-1}}\right.$$
$$\quad \left.\left. + 2\alpha(1-\alpha)\frac{\{-\log(1-p)\}^{2/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{2/\alpha-1}} + o\left\{(\log t)^{1-2/\alpha}\right\} + O\left(t^{-1}\right)\right]\right)$$
$$\sim \frac{t^{-1}}{(\log t)^{2/\alpha-1}}\left[\alpha(1-\alpha)\{-\log(1-p)\}^{2/\alpha}\right],$$

as $t \to \infty$, which is regularly varying of order $-1$. As such, the index of regular variation is $\tau_1(\delta) = 1$ for $\delta = 0$. Combining these results, $\tau_1(\delta) = 1$ for all $\delta \in [0,1]$. By symmetric arguments, $\tau_2(\delta) = \tau_3(\delta) = 1$ for all $\delta \in [0,1]$.

$|C| = 2$: $\tau_{1,2}(\delta)$, $\tau_{1,3}(\delta)$, $\tau_{2,3}(\delta)$. We first consider the cone $\mathbb{E}_{1,2}$. For $\delta > 0$, we have

$$\mathrm{pr}\left(X_1 > t, X_2 > t, X_3 < t^\delta\right) = \mathrm{pr}\left(X_1 > t, X_2 > t\right) - \mathrm{pr}\left(X_1 > t, X_2 > t, X_3 > t^\delta\right)$$

$$= \exp\left[-2^\alpha\left\{-\log\left(1 - e^{-1/t}\right)\right\}\right]$$
$$- \exp\left(-\left[2\left\{-\log\left(1 - e^{-1/t}\right)\right\}^{1/\alpha} + \left\{-\log\left(1 - e^{-1/t^\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right)$$
$$= \left(1 - e^{-1/t}\right)^{2^\alpha} - \exp\left(-\left[2\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + \left\{\delta\log t + O\left(t^{-2\delta}\right)\right\}^{1/\alpha}\right]^\alpha\right)$$
$$= t^{-2^\alpha} + O\left(t^{-1-2^\alpha}\right) - \exp\left[-\left(2 + \delta^{1/\alpha}\right)^\alpha\log t + O\left\{(t\log t)^{-1}\right\}\right] \sim t^{-2^\alpha},$$

i.e., $\tau_{1,2}(\delta) = 2^{-\alpha}$, $\delta > 0$. For $\delta = 0$, using similar arguments as for the $|C| = 1$ case, we have

$$\mathrm{pr}\left(X_1^* > t, X_2^* > t, X_3^* = 0\right) = \mathrm{pr}\left(X_1 > t, X_2 > t, X_3 < -1/\log p\right)$$
$$= \mathrm{pr}\left(X_1 > t, X_2 > t\right) - \mathrm{pr}\left(X_1 > t, X_2 > t, X_3 > -1/\log p\right)$$
$$= \exp\left[-2^\alpha\left\{\log t + O\left(t^{-1}\right)\right\}\right]$$
$$- \exp\left(-\left[2\left\{\log t + O\left(t^{-1}\right)\right\}^{1/\alpha} + \left\{-\log(1-p)\right\}^{1/\alpha}\right]^\alpha\right)$$
$$= \exp\left[-2^\alpha\left\{\log t + O\left(t^{-1}\right)\right\}\right]$$
$$\cdot\left(1 - \exp\left[-\alpha 2^{\alpha-1}\frac{\{-\log(1-p)\}^{1/\alpha}}{\{\log t + O\left(t^{-1}\right)\}^{-1+1/\alpha}} + o\left\{(\log t)^{1-2/\alpha}\right\}\right]\right)$$
$$\sim \frac{t^{-2^\alpha}}{(\log t)^{-1+1/\alpha}}\alpha 2^{\alpha-1}\{-\log(1-p)\}^{1/\alpha},$$

as $t \to \infty$. As such, the index of regular variation is $\tau_{1,2}(\delta) = 2^{-\alpha}$ for all $\delta \in [0,1]$. By analogous arguments, we also have $\tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 2^{-\alpha}$, $\delta \in [0,1]$.

$|C| = 3$: $\tau_{1,2,3}$. To calculate the index of regular variation for cone $\mathbb{E}_{1,2,3}$, we consider

$$\mathrm{pr}\left(X_1 > t, X_2 > t, X_3 > t\right) = \exp\left\{\log\left(1 - e^{-1/t}\right)V(1,1,1)\right\}$$
$$= \left(1 - e^{-1/t}\right)^{V(1,1,1)} = \left\{1 - 1 + t^{-1} + O(t^{-2})\right\}^{3^\alpha} \sim t^{-3^\alpha},$$

so $\tau_{1,2,3} = 3^{-\alpha}$. This corresponds to the known value of $\eta_{1,2,3}$ for the trivariate inverted logistic distribution, as $\eta_{1,2,3} = V(1,1,1)^{-1} = 3^{-\alpha}$.

## A.6 Multivariate Gaussian distribution

The multivariate Gaussian provides a further example of a distribution which asymptotically places all mass on cones $\mathbb{E}_C$ with $|C| = 1$. In the case where $d = 3$, for a multivariate Gaussian distribution with covariance matrix $\Sigma$, Nolde (2014), for example, shows that

$$\eta_{1,2,3} = \left(1_3^T\Sigma^{-1}1_3\right)^{-1}; \quad \eta_{i,j} = \left(1_2^T\Sigma_{i,j}^{-1}1_2\right)^{-1}, \quad i < j \in \{1,2,3\},$$

where $\Sigma_{i,j}$ is the submatrix of $\Sigma$ corresponding to variables $i$ and $j$, and $1_d \in \mathbb{R}^d$ is a vector of 1s.

The covariance matrix $\Sigma$ may be written as

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} = \begin{bmatrix} \Sigma_{12} & B \\ B^T & 1 \end{bmatrix}, \quad \text{where } \Sigma_{12} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} \rho_{13} \\ \rho_{23} \end{bmatrix}.$$

We note that since $\Sigma$ and $\Sigma_{12}$ are covariance matrices, they must be positive definite, with $\det(\Sigma) = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} > 0$ and $\det(\Sigma_{12}) = 1 - \rho_{12}^2 > 0$. The inverse of $\Sigma$ is given by the block matrix

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{12}^{-1} + \Sigma_{12}^{-1}B(1 - B^T\Sigma_{12}^{-1}B)^{-1}B^T\Sigma_{12}^{-1} & -\Sigma_{12}^{-1}B(1 - B^T\Sigma_{12}^{-1}B)^{-1} \\ -(1 - B^T\Sigma_{12}^{-1}B)^{-1}B^T\Sigma_{12}^{-1} & (1 - B^T\Sigma_{12}^{-1}B)^{-1} \end{bmatrix},$$

so that

$$1_3^T \Sigma^{-1} 1_3 = 1_2^T \Sigma_{12}^{-1} 1_2 + (1 - B^T \Sigma_{12}^{-1} B)^{-1} \left(1 - 1_2^T \Sigma_{12}^{-1} B - B^T \Sigma_{12}^{-1} 1_2 + 1_2^T \Sigma_{12}^{-1} B B^T \Sigma_{12}^{-1} 1_2\right)$$

$$= 1_2^T \Sigma_{12}^{-1} 1_2 + \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}} \left(1 - 21_2^T \Sigma_{12}^{-1} B + 1_2^T \Sigma_{12}^{-1} B B^T \Sigma_{12}^{-1} 1_2\right)$$

$$= 1_2^T \Sigma_{12}^{-1} 1_2 + \frac{\det(\Sigma_{12})}{\det(\Sigma)} \left(1 - 1_2^T \Sigma_{12}^{-1} B\right)^2$$

$$= 1_2^T \Sigma_{12}^{-1} 1_2 + \frac{\det(\Sigma_{12})}{\det(\Sigma)} \left(1 - \frac{\rho_{13} + \rho_{23}}{1 + \rho_{12}}\right)^2 \geq 1_2^T \Sigma_{12}^{-1} 1_2,$$

with equality if and only if $1 + \rho_{12} = \rho_{13} + \rho_{23}$. By similar calculations,

$$1_3^T \Sigma^{-1} 1_3 \geq 1_2^T \Sigma_{i,j}^{-1} 1_2, \quad i < j \in \{1, 2, 3\},$$

with equality if and only if $1 + \rho_{ij} = \rho_{ik} + \rho_{jk}$, in which case $\eta_{1,2,3} = \eta_{i,j}$. Applying Theorem 2, for this trivariate case, if $1 + \rho_C \neq \sum_{C': |C'|=2, C' \neq C} \rho_{C'}$ for all $C \subset \{1, 2, 3\}$ with $|C| = 2$, then $\tau_C(1) = \eta_C$ for any set $C \subseteq \{1, 2, 3\}$ with $|C| \geq 2$. Since $\tau_C(\delta)$ is non-decreasing in $\delta$, $\tau_C(\delta) \leq \eta_C$ for $\delta \in [0, 1)$. We also know $\tau_{1,2,3} = \eta_{1,2,3}$.

In this case, calculation of the explicit formulas for $\delta < 1$, in a manner similar to the inverted logistic case, is complicated by the need to consider asymptotic approximations of Gaussian cumulative distribution functions beyond first order. As such, we do not attempt this here, but note that since $\tau_C(1) \leq \eta_C$, $|C| \geq 2$, we would be likely to estimate $\tau_C(\delta) < 1$ in practice.

To gain some insight into the remaining cases of $C = \{i\}$, $i = 1, 2, 3$, we consider the conditional extreme value model of Heffernan and Tawn (2004). Let $Y = \log X$, so that $Y$ has standard Gumbel marginal distributions, and all correlations be positive. Then conditioning on $Y_i$ gives

$$\mathrm{pr}\left(Y_j - \rho_{ij}^2 t \leq t^{1/2} z_j, Y_k - \rho_{ik}^2 t \leq t^{1/2} z_k, Y_i > t\right) \sim N(z_j, z_k)\mathrm{pr}(Y_i > t), \quad t \to \infty,$$

for $N(z_j, z_k)$ denoting the distribution function of a particular Gaussian distribution (Heffernan and Tawn, 2004, Section 8). For $z_j = z_k = 0$, this equates to

$$\mathrm{pr}\left(Y_j \leq \rho_{ij}^2 t, Y_k \leq \rho_{ik}^2 t, Y_i > t\right) \sim N(0, 0)\mathrm{pr}(Y_i > t), \quad t \to \infty. \tag{23}$$

In the trivariate case with Gumbel margins, equation (10) of Assumption 1 can be written as

$$\mathrm{pr}\left(Y_j \leq \delta t, Y_k \leq \delta t, Y_i > t\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}, \quad t \to \infty.$$

Considering equation (23) again, we see that if $\delta \geq \max\left(\rho_{ij}^2, \rho_{ik}^2\right)$, in place of the limiting Gaussian distribution, mass will occur at $(-\infty, -\infty)$ for variables $(Y_j, Y_k)$, which implies that $\tau_i(\delta) = 1$. Alternatively, if $\rho_{ij}^2 \leq \delta < \rho_{ik}^2$ or $\rho_{ik}^2 \leq \delta < \rho_{ij}^2$, we have mass at $(-\infty, \infty)$ or $(\infty, -\infty)$, respectively, and for $\delta < \min\left(\rho_{ij}^2, \rho_{ik}^2\right)$ mass occurs at $(\infty, \infty)$. In these cases, the left-hand side of equation (23) is $o\left\{\mathrm{pr}\left(Y_i > t\right)\right\}$, which is consistent with $\tau_i(\delta) < 1$.

# B  Simulation study

## B.1  Estimation of $\tau_C(\delta)$ in Method 2

In Method 2, introduced in Section 3.3 of the paper, we consider regions of the form

$$\tilde{E}_C = \left\{x \in \mathbb{E} : x_\vee^{D \setminus C} \leq \left(x_\wedge^C\right)^\delta\right\}, \quad |C| < d; \quad \tilde{E}_D = \mathbb{E} \setminus \bigcup_{C \in 2^D \setminus \emptyset : |C| < d} \tilde{E}_C,$$

for $C \in 2^D \setminus \emptyset$, and assume that $\mathrm{pr}\left(X_\wedge^C > q, X \in \tilde{E}_C\right) \in \mathrm{RV}_{-1/\tilde{\tau}_C(\delta)}$, for $X_\wedge^C = \min_{i \in C} X_i$. This is used as an alternative to considering $\mathrm{pr}\left(X_\wedge^C > t, X_\vee^{D \setminus C} \leq t^\delta\right) \in \mathrm{RV}_{-1/\tau_C(\delta)}$, for which there is no clear structure variable.

In Fig. 6, we demonstrate how well the parameter $\tau_C(\delta)$ is approximated using Method 2. We consider the trivariate logistic distribution, with theoretical $\tau_C(\delta)$ values given in case (iii) of Table 1. For $\alpha = 0.25$ and $\alpha = 0.5$, we take samples of size 100,000 from this distribution, and use Method 2 to estimate $\tilde{\tau}_1(\delta)$, $\tilde{\tau}_{1,2}(\delta)$ and $\tilde{\tau}_{1,2,3}$ for values of $\delta \in \{0.1, \ldots, 0.95\}$. The thresholds used correspond to the 0.985 quantile of observed $X_\wedge^C$ values in each region $\tilde{E}_C$. Each simulation is repeated 100 times, and the true $\tau_C(\delta)$ parameter values are shown in grey.

The results indicate that the estimator derived from considering $\mathrm{pr}\left\{X_\wedge^C > t, X_\vee^{D\setminus C} \le \left(X_\wedge^C\right)^\delta\right\}$ yields $\tau_C(\delta)$ as defined through $\mathrm{pr}\left(X_\wedge^C > t, X_\vee^{D\setminus C} \le t^\delta\right)$. We observe increased variability in the estimates of $\tilde{\tau}_1(\delta)$ and $\tilde{\tau}_{1,2}(\delta)$ for small values of $\delta$, and $\tilde{\tau}_{1,2,3}$ for large values of $\delta$, which is likely due to the limited data in the corresponding regions $\tilde{E}_C$ for these cases.
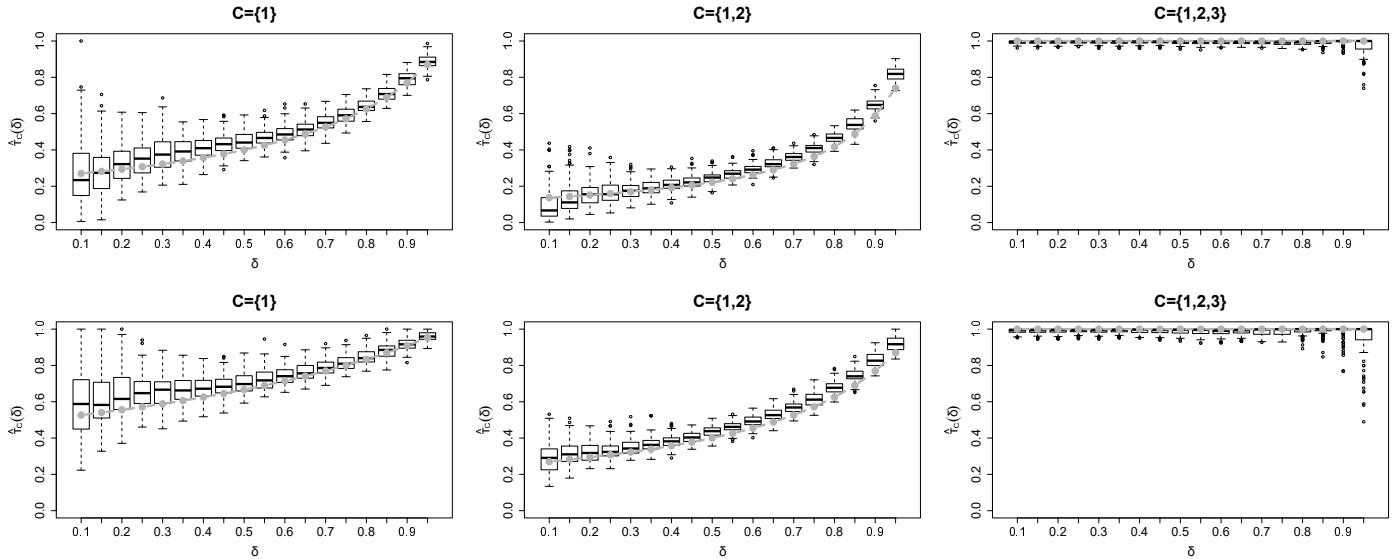


Figure 6: Estimates of $\tau_1(\delta)$, $\tau_{1,2}(\delta)$ and $\tau_{1,2,3}$ for data simulated from trivariate logistic distributions with $\alpha = 0.25$ (top) and $\alpha = 0.5$ (bottom).

## B.2 Area under the receiver operating characteristic curve results for the max-mixture distribution

In Table 2 of the main paper, we present the average area under the receiver operating characteristic curve for Method 1, Method 2 and the approach of Goix et al. applied to samples taken from a particular max-mixture distribution. Figure 7 provides boxplots of all the values obtained in these simulations.

## B.3 Asymmetric logistic distribution

We now present simulation results for the asymmetric logistic distribution, using the same metrics as for the max-mixture distribution in Section 4 of the paper. This model belongs to the class of multivariate extreme value distributions, and it is possible to calculate the proportion of extremal mass associated with each cone $\mathbb{E}_C$.

In standard Fréchet margins, the multivariate extreme value distribution function is of the form $\exp\{-V(x)\}$. Coles and Tawn (1991) show that the spectral density corresponding to cone $\mathbb{E}_C$ is $h_C(w_C) = -V^{\{C\}}(w_C)/d$, where

$$V^{\{C\}}(x_C) = \lim_{x_j \to 0: j \in D\setminus C} \left(\prod_{i \in C} \frac{\partial}{\partial x_i}\right) V(x), \qquad x_C = \{x_i : i \in C\},$$
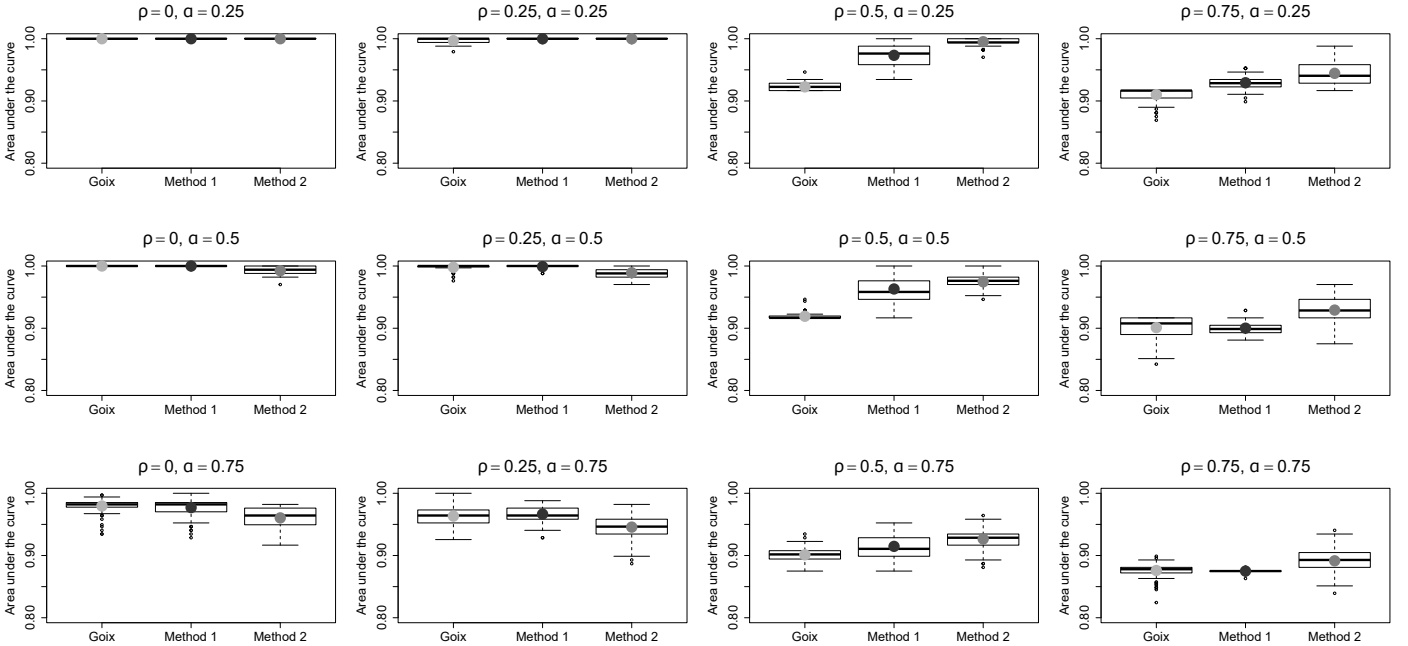
23

Figure 7: Area under the receiver operating characteristic curve results for 100 simulations from a five-dimensional max-mixture distribution.

for $w_C = x_C / r_C$ and $r_C = \sum_{i \in C} x_i$. Hence, for $\mathbb{B}_C = \{w \in \mathcal{S}_{d-1} : w_i \in (0,1], i \in C; \ w_j = 0, j \in D \setminus C\}$, the proportion of mass on corresponding cone $\mathbb{E}_C$ is

$$p_C = -\frac{1}{d} \int_{\mathbb{B}_C} V^{\{C\}}(w_C) \prod_{i \in C} dw_i. \tag{24}$$

For the asymmetric logistic model (Tawn, 1990), the exponent measure $V$ is defined as

$$V(x) = \sum_{C \in 2^D \setminus \emptyset} \left\{ \sum_{i \in C} (\theta_{i,C}/x_i)^{1/\alpha_C} \right\}^{\alpha_C}, \qquad \theta_{i,C} \in [0,1], \tag{25}$$

with $\theta_{i,C} = 0$ if $i \notin C$, $\sum_{C \in 2^D \setminus \emptyset} \theta_{i,C} = 1$ for all $i = 1, \ldots, d$ and $C \in 2^D \setminus \emptyset$, and dependence parameters $\alpha_C \in (0,1]$. In Proposition 1 of Section C, we show that for the $d$-dimensional asymmetric logistic model with all $\alpha_C \equiv \alpha$, the proportion of mass on cone $\mathbb{E}_C$ is

$$p_C^{(d)} = \sum_{i \in C} \theta_{i,C}/d, \qquad C \in 2^D \setminus \emptyset.$$

Using this new result, we can compare our estimated proportions to the truth using the Hellinger distance defined in equation (20) of the paper.

Following Goix et al. (2017), we simulate data from an asymmetric logistic distribution with $\alpha_C \equiv \alpha$, whose extremal mass is concentrated on $f$ randomly chosen sub-cones, ensuring that moment constraint (3) is satisfied. Suppose the sub-cones chosen correspond to subsets $F_1, \ldots, F_f \in 2^D \setminus \emptyset$. The conditions on the parameters of the asymmetric logistic distribution are satisfied by setting

$$\theta_{i,C} = |\{j : i \in F_j, j \in \{1, \ldots, f\}\}|^{-1}, \qquad C \in \{F_1, \ldots, F_f\}, \tag{26}$$

and $\theta_{i,C} = 0$ otherwise. We present results for dimensions $d = 5$ and $d = 10$. For $d = 5$, we simulate samples of size $n = 10,000$, and test both our methods when there are truly 5, 10 and 15 cones $\mathbb{E}_C$ with extremal mass.

For $d = 10$, we have $n = 100,000$ samples, and consider 10, 50 and 100 cones with extremal mass. We set the tuning parameters as in Section B.3 of the paper, and repeat each setting 100 times. In Table 5, we present the average area under the receiver operating characteristic curve for $\alpha \in \{0.25, 0.5, 0.75\}$. Boxplots of the full results obtained are provided in Figs. 8 and 9. In the asymmetric logistic model, the closer $\alpha_C$ is to 1, the larger the values of $\tau_{\underline{C}}(\delta)$ for any fixed $\delta$ and $\underline{C} \subset C$, as demonstrated by cases (ii) and (iii) in Table 1. Thus, the larger the value of $\alpha$ in our simulations, the more difficult it is to determine which cones $\mathbb{E}_C$ truly contain extremal mass.

| $(\alpha, f)$ | $(0.25, 5)$ | $(0.25, 10)$ | $(0.25, 15)$ | $(0.5, 5)$ | $(0.5, 10)$ | $(0.5, 15)$ | $(0.75, 5)$ | $(0.75, 10)$ | $(0.75, 15)$ |
|---|---|---|---|---|---|---|---|---|---|
| Goix et al. | 100 (0.1) | 99.7 (0.5) | 99.0 (1.3) | 100 (0.2) | 99.8 (0.5) | 99.4 (0.9) | 98.6 (1.7) | 91.1 (4.9) | 87.1 (7.1) |
| Method 1 | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.1) | 99.9 (0.3) | 92.5 (4.1) | 87.0 (5.5) | 85.6 (7.3) |
| Method 2 | 100 (0.0) | 100 (0.0) | 100 (0.0) | 99.7 (0.6) | 98.5 (1.8) | 96.8 (3.2) | 85.4 (4.3) | 84.4 (5.7) | 84.7 (7.5) |
| $(\alpha, f)$ | $(0.25, 10)$ | $(0.25, 50)$ | $(0.25, 100)$ | $(0.5, 10)$ | $(0.5, 50)$ | $(0.5, 100)$ | $(0.75, 10)$ | $(0.75, 50)$ | $(0.75, 100)$ |
| Goix et al. | 100 (0.0) | 100 (0.0) | 100 (0.1) | 100 (0.0) | 100 (0.0) | 99.8 (0.1) | 99.5 (0.2) | 98.8 (0.1) | 96.9 (0.5) |
| Method 1 | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 99.9 (0.0) | 99.6 (0.1) | 98.5 (0.4) | 96.0 (0.8) |
| Method 2 | 100 (0.0) | 100 (0.0) | 100 (0.0) | 100 (0.0) | 99.6 (0.1) | 99.2 (0.1) | 99.0 (0.1) | 98.8 (0.2) | 97.9 (0.4) |

Table 5: Average areas under the receiver operating characteristic curves, given as percentages, for 100 samples from five-dimensional (top) and ten-dimensional (bottom) asymmetric logistic distributions, with dependence parameter $\alpha$ and $\theta_{i,C}$ determined via (26). Standard deviations of these results are given in brackets.

The average areas under the receiver operating characteristic curves in Table 5 show that all three methods perform well when $\alpha = 0.25$ and $\alpha = 0.5$, for both $d = 5$ and $d = 10$, with Method 1 slightly outperforming the other two approaches. The results suggest that the method of Goix et al. (2017) is generally the most successful classifier when $\alpha = 0.75$, followed by Method 1, although the most substantial difference in results occurs for $(d, f, \alpha) = (10, 100, 0.75)$, where Method 2 is most successful; this is supported by the boxplots of results in Figs. 8 and 9. In principle, Method 1 should be better than Method 2 here, so greater assessment of tuning parameters may be required. It is possible that the method of Goix et al. is most successful for larger values of $\alpha$ since it is more difficult for Methods 1 and 2 to distinguish between regions where $\tau_C(\delta)$ does and does not equal 1 in these cases.

Figure 10 shows the average Hellinger distance for $\alpha \in [0.1, 0.9]$ for each of the cases described above. For the most sparse cases, $(d, f) = (5, 5)$ and $(d, f) = (10, 10)$, Method 1 performs significantly better than the other two approaches overall. For the less sparse cases, $(d, f) = (5, 15)$ and $(d, f) = (10, 100)$, the three methods give similar results in terms of the Hellinger distance for $\alpha \leq 0.5$, but the method of Goix et al. is most successful for larger $\alpha$ values. When the extreme values are concentrated on fewer cones $\mathbb{E}_C$, it may be easier to estimate true values of $\tau_C(\delta) = 1$ using Method 1 than in less sparse examples. For the asymmetric logistic distribution, Method 1 often performs better in terms of estimating the proportion of extremal mass on each cone $\mathbb{E}_C$, while the method of Goix et al. (2017) is better at classification, although this method does tend to place mass on too many cones $\mathbb{E}_C$, as shown in Fig. 3 of the paper.

# C  Calculating the mass on each cone $\mathbb{E}_C$ for an asymmetric logistic model

**Proposition 1.** *For the $d$-dimensional asymmetric logistic model with exponent measure (25) and $\alpha_C \equiv \alpha \in (0, 1)$ for all $C \in 2^D \setminus \emptyset$,*

$$dp_C^{(d)} = \sum_{i \in C} \theta_{i,C},$$

*where $p_C^{(d)}$ denotes the proportion of mass on cone $\mathbb{E}_C$.*

*Proof.* Consider the exponent measure of the asymmetric logistic model $V(x)$ as a sum of functions $V_C(x_C)$, for $C \in 2^D \setminus \emptyset$, i.e.,

$$V(x) = \sum_{C \in 2^D \setminus \emptyset} V_C(x_C), \quad V_C(x_C) = \left\{ \sum_{i \in C} \left( \frac{\theta_{i,C}}{x_i} \right)^{1/\alpha} \right\}^{\alpha}.$$
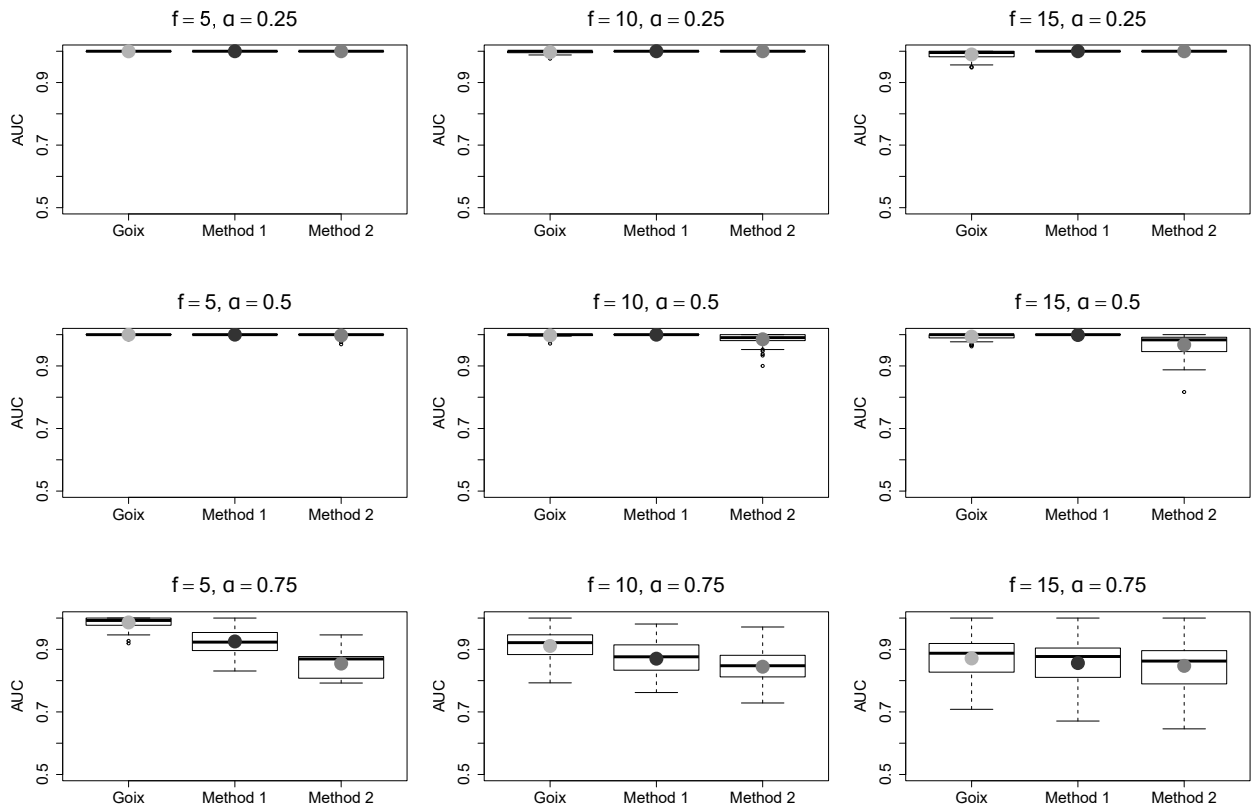
Figure 8: Areas under the receiver operating characteristic curves for 100 simulations from a five-dimensional asymmetric logistic distribution.
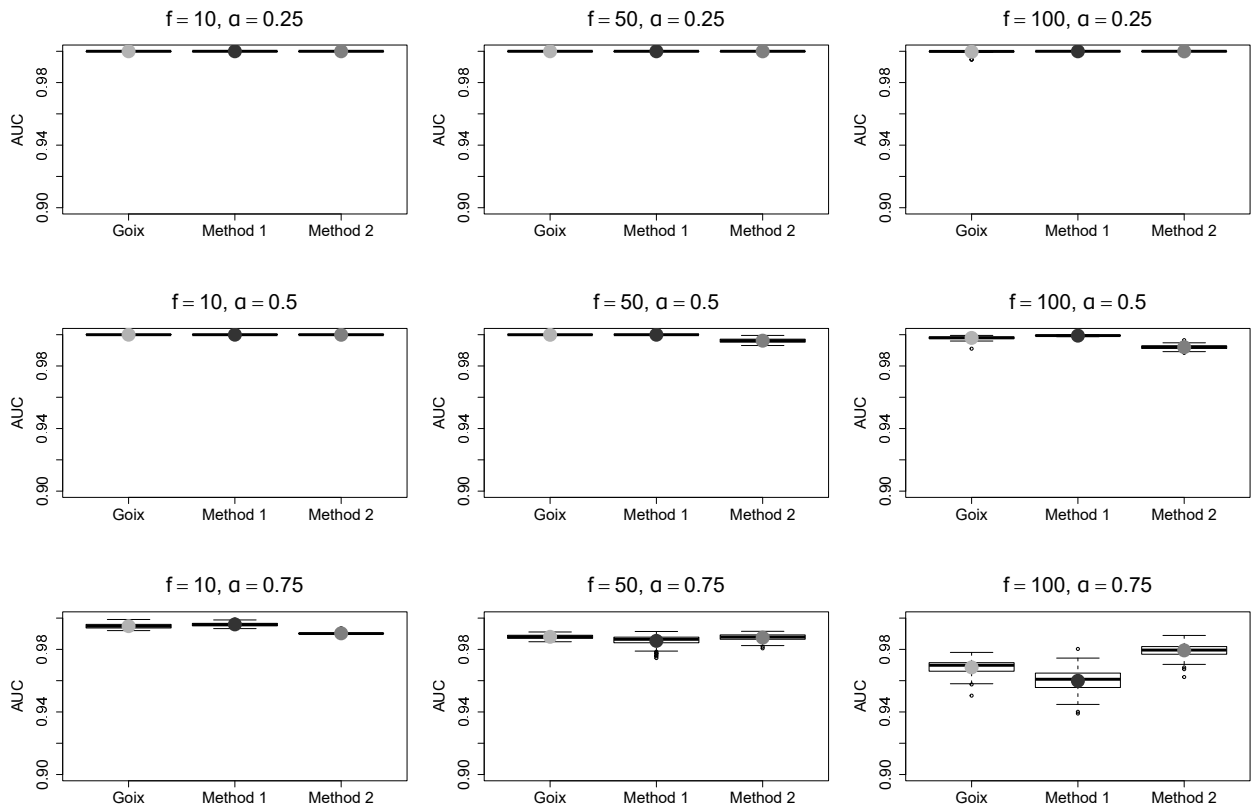


Figure 9: Areas under the receiver operating characteristic curves for 100 simulations from a ten-dimensional asymmetric logistic distribution.
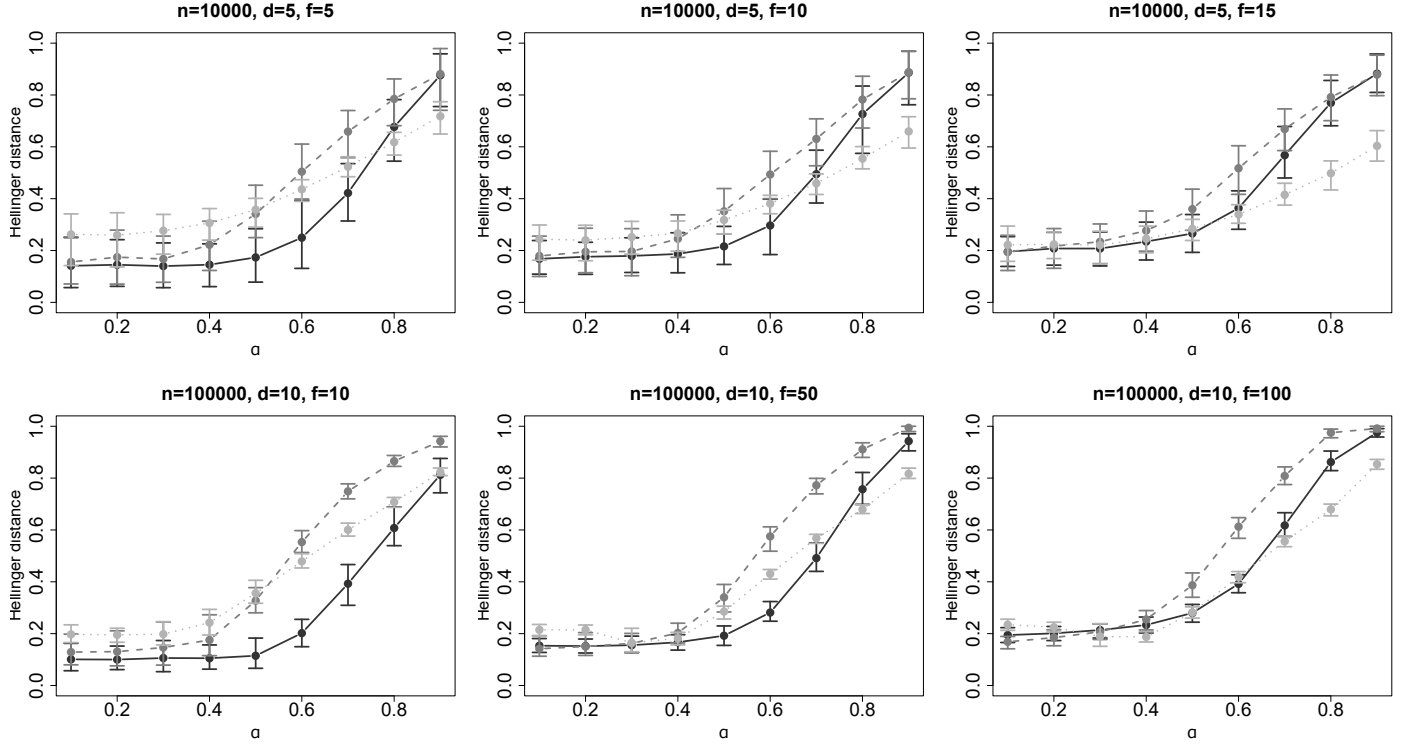
Figure 10: Mean Hellinger distance, 0.05 and 0.95 quantiles over 100 simulations. Method 1: solid lines; Method 2: dashed lines; Goix et al.: dotted lines.

Then for any dimension $d \geq |C|$,

$$V^{\{C\}}(x_C) = \left(\prod_{i \in C} \frac{\partial}{\partial x_i}\right) V_C(x_C) = \left\{\prod_{i=0}^{|C|-1} -\left(\frac{\alpha - i}{\alpha}\right)\right\} \left(\prod_{i \in C} \frac{\theta_{i,C}^{1/\alpha}}{x_{C,i}^{1+1/\alpha}}\right) \left\{\sum_{i \in C} \left(\frac{\theta_{i,C}}{x_{C,i}}\right)^{1/\alpha}\right\}^{\alpha - |C|},$$

since for $\overline{C} \supset C$, $\displaystyle\lim_{x_i \to 0: i \in \overline{C} \backslash C} \left(\prod_{j \in C} \frac{\partial}{\partial x_j}\right) V_{\overline{C}}(x_{\overline{C}}) = 0$. Hence, by result (24),

$$dp_C^{(d)} = -\int_{\mathbb{B}_C} V^{\{C\}}(w_C) \prod_{i \in C} dw_i, \tag{27}$$

which we note does not depend on $d$. We claim that

$$-\int_{\mathbb{B}_C} V^{\{C\}}(w_C) \prod_{i \in C} dw_i = \sum_{i \in C} \theta_{i,C}. \tag{28}$$

First consider $|C| = 1$, i.e. $\mathbb{B}_C = \{w : w_i = 1\}$ for $C = \{i\}$. Here,

$$V^{\{i\}}(x_i) = \frac{\partial}{\partial x_i} V_i(x_i) = -\frac{\theta_{i,i}}{x_i^2},$$

so

$$dp_i^{(d)} = \left.\frac{\theta_{i,i}}{w_i^2}\right|_{w_i=1} = \theta_{i,i}, \qquad i = 1, \ldots, d.$$

Now consider $|C| = 2$. We have

$$V^{\{i,j\}}(x_i, x_j) = \left(\frac{\alpha - 1}{\alpha}\right) \frac{\{(1 - \theta_{i,i})(1 - \theta_{j,j})\}^{1/\alpha}}{(x_i x_j)^{1+1/\alpha}} \left\{\left(\frac{\theta_{i,ij}}{x_i}\right)^{1/\alpha} + \left(\frac{\theta_{j,ij}}{x_j}\right)^{1/\alpha}\right\}^{\alpha - 2},$$

27

so

$$h_{i,j}(w_i) = \left(\frac{1-\alpha}{\alpha}\right)\frac{(\theta_{i,ij}\theta_{j,ij})^{1/\alpha}}{\{w_i(1-w_i)\}^{1+1/\alpha}}\left\{\left(\frac{\theta_{i,ij}}{w_i}\right)^{1/\alpha} + \left(\frac{\theta_{j,ij}}{1-w_i}\right)^{1/\alpha}\right\}^{\alpha-2},$$

and

$$dp_{i,j}^{(d)} = \int_0^1 h_{i,j}(w_i)dw_i. \tag{29}$$

However, taking $d = 2$, we know $2p_{1,2}^{(2)} + 2p_1^{(2)} + 2p_2^{(2)} = 2$, so $dp_{1,2}^{(d)} = \theta_{1,12} + \theta_{2,12}$, and similarly, by (29), $\int_0^1 h_{i,j}(w_i)dw_i = dp_{i,j}^{(d)} = \theta_{i,ij} + \theta_{j,ij}$. So, (27) holds for $|C| = 1$ and $|C| = 2$, and we suppose it holds for $|C| \le k$, i.e. $dp_C^{(d)} = -\int_{\mathbb{B}_C} V^{\{C\}}(w_C)\prod_{i\in C} dw_i = \sum_{i\in C}\theta_{i,C}$.

Since (27) does not depend on $d$, take $d = k+1$. So for all $\underline{C}$ with $|\underline{C}| \le k$, $(k+1)p_{\underline{C}}^{(k+1)} = \sum_{i\in\underline{C}}\theta_{i,\underline{C}}$. Now take $C = \{1,\ldots,k+1\}$. Then,

$$\begin{aligned}
(k+1)p_C^{(k+1)} &= (k+1) - \sum_{\underline{C}\subset C}\sum_{i\in\underline{C}}\theta_{i,\underline{C}} = (k+1) - \sum_{i\in C}\sum_{\underline{C}\subset C}\theta_{i,\underline{C}} \\
&= \sum_{i\in C}\left(1 - \sum_{\underline{C}\subset C}\theta_{i,\underline{C}}\right) \\
&= \sum_{i\in C}\theta_{i,C}.
\end{aligned}$$

As such, (28) holds by induction. $\qquad\square$

# References

Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050.

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, 9(1):383–418.

Chiapino, M. and Sabourin, A. (2017). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *New Frontiers in Mining Complex Patterns*, pages 132–147. Springer, Cham.

Chiapino, M., Sabourin, A., and Segers, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes*, 22(2):193–222.

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

Das, B., Mitra, A., and Resnick, S. I. (2013). Living on the multidimensional edge: seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(1):139–163.

Goix, N., Sabourin, A., and Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pages 75–83, Cadiz, Spain. PMLR.

Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York, 2nd edition.

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.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

Hua, L. and Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, 102(10):1454–1471.

Keef, C., Tawn, J. A., and Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21.

Klüppelberg, C., Haug, S., and Kuhn, G. (2015). Copula structure analysis based on extreme dependence. *Statistics and its Interface*, 8(1):93–107.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

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.

Maulik, K. and Resnick, S. I. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Mitra, A. and Resnick, S. I. (2011). Hidden regular variation and detection of hidden risks. *Stochastic Models*, 27(4):591–614.

Morris, D. G. and Flavin, R. W. (1990). A digital terrain model for hydrology. In *Proceedings of the 4th International Symposium on Spatial Data Handling*, volume 1, pages 250–262, Zurich.

Morris, D. G. and Flavin, R. W. (1994). Sub-set of UK 50m by 50m hydrological digital terrain model grids. *NERC, Institute of Hydrology, Wallingford*.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag, New York.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT - Statistical Journal*, 10(1):61–82.

Simpson, E. S. (2019). *Modelling Dependence in Multivariate Extremes*. PhD thesis, Lancaster University.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Vettori, S., Huser, R., Segers, J., and Genton, M. G. (2018). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *arXiv:1705.10488*.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.