Biometrika (2020), **107**, 3, *pp.* 513–532
Printed in Great Britain
doi: 10.1093/biomet/asaa018
Advance Access publication 7 May 2020

Determining the dependence structure of multivariate extremes

BY E. S. SIMPSON, J. L. WADSWORTH AND J. A. TAWN

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.
e.simpson3@lancaster.ac.uk j.wadsworth@lancaster.ac.uk j.tawn@lancaster.ac.uk

SUMMARY

In multivariate extreme value analysis, the nature of the extremal dependence between variables should be considered when selecting appropriate statistical models. Interest often lies in 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 nonstandard 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 utility 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 river flows in the U.K., estimating the probabilities of different subsets of sites being large simultaneously.

Some key words: Asymptotic independence; Extremal dependence structure; Hidden regular variation; Multivariate regular variation.

1. INTRODUCTION

When constructing models in multivariate extreme value analysis, one often needs to exploit extremal dependence features. Consider the random vector $X = (X_1, \ldots, X_d)$, with $X_i \sim F_i$, and 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 $D = \{1, \ldots, d\}$ excluding the empty set. For any C with $|C| \geqslant 2$, extremal dependence within X_C can be summarized by

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

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 & Joe (2011) or Wadsworth & Tawn (2013).

Many models for multivariate extremes are applicable only when the data exhibit either full asymptotic dependence, entailing $\chi_C > 0$ for all $C \in 2^D \setminus \emptyset$ with $|C| \geqslant 2$, or full asymptotic independence, i.e., $\chi_{i,j} = 0$ for all $i < j$ (Heffernan & Tawn, 2004). However, it is often the case that some χ_C are positive while 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 $\{\chi_C : C \in 2^D \setminus \emptyset, |C| \geqslant 2\}$, since we do not learn fully whether small values of some variables occur with large values of others, or if 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 $\text{pr}(X_i < x) = \exp(-1/x)$ 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 $W \in \mathcal{S}_{d-1} = \{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i=1}^d w_i = 1\}$, the $(d-1)$-dimensional unit simplex. It follows that $\text{pr}(R > r) \sim ar^{-1}$ as $r \to \infty$ for $a \geqslant 1$, so all the information about extreme events is contained in W, in particular the distribution of W conditioned on $R > r$ as $r \to \infty$. Under the assumption of multivariate regular variation (Resnick, 2007, Ch. 6),

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

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

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

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 χ_C in (1) and the spectral measure H: if $\chi_C > 0$, then H places mass on at least one region $\mathcal{S}_{d-1}^{\bar{C}} = \{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i \in \bar{C}} w_i = 1\}$ with $C \subseteq \bar{C} \subseteq D$. This underlines the fact that the term asymptotic dependence is not so useful here, since it offers only partial insight into the structure. Therefore, in what follows, we 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 can 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 § 2.1. Each face of $\mathcal{S}_{d-1}$ can be identified with a subcone of $\mathbb{E}$ for which one or more components are identically zero. Intuition and visualization are often simpler with H, but in what follows we work with $\mathbb{E}$ and the subcones 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 recently been studied by other authors. Under the assumption that the data are from an elliptical copula, Klüppelberg et al. (2015) used factor analysis on extreme correlations linked to the tail dependence function. Chautru (2015) introduced a nonparametric approach based on statistical learning, combining a principal component analysis algorithm with clustering techniques. A Bayesian

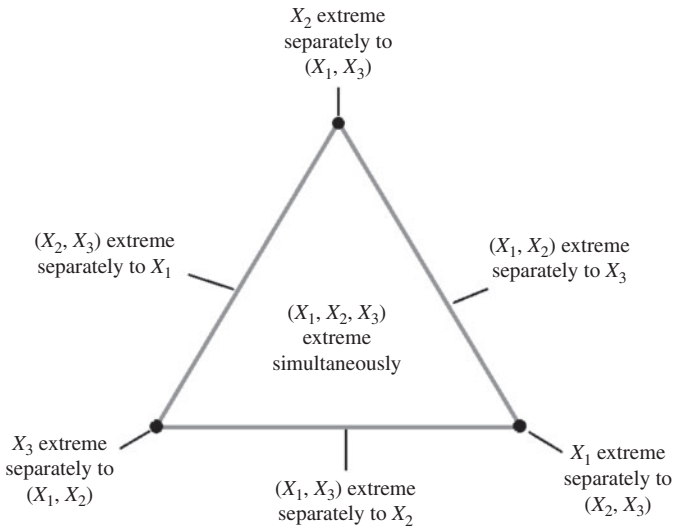

Fig. 1. The simplex $\mathcal{S}_2$; coordinates are transformed to the equilateral simplex.

clustering method was proposed by Vettori et al. (2020), based on the hierarchical dependence structure of the nested logistic distribution of Tawn (1990). Goix et al. (2016, 2017) developed a nonparametric 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 & Sabourin (2017) proposed 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) extended this approach by instead using the coefficient of tail dependence of Ledford & Tawn (1996).

In this paper we exploit additional, but commonly satisfied, hidden regular variation assumptions on nonstandard subcones of $\mathbb{E}$, by introducing a new set of parameters that describes the dominant extremal dependence structure. We study properties of these parameters and their links to existing coefficients, and we 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, which also allows 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 $d < n$, where d represents the number of variables and n the number of data points. This is the same complexity as that of 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 $\lambda(x)$ as $L(x)x^\alpha$, where $L \in \mathrm{RV}_0$ is called 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 exist a scaling function $a(t) \to \infty$ and a positive measure μ such that

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

as $t \to \infty$, with vague convergence in the space of nonnegative Radon measures on G (Resnick, 2007, Ch. 3). If we assume that the margins of X are standard Fréchet or Pareto, we may take $a(t) = t$, and the limit measure μ 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 & Tawn (1996, 1997). Further work has been done by Maulik & Resnick (2004) and Mitra & Resnick (2011), among others, while 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 subcone, we have hidden regular variation on that subcone.

To the best of our knowledge, the marginal case of this hidden regular variation framework is the only one that has previously been exploited from a statistical perspective; from a theoretical viewpoint, Das et al. (2013) considered hidden regular variation on a series of nonstandard cones, although these are mostly different from the ones we will consider. For $X_C = \{X_i : i \in C\}$ and $x_C = \{x_i : i \in C\}$, Ledford & 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 $\mathbb{E}_C^*$ that are homogeneous of order $-1/\eta_C$ and with the so-called coefficient of tail dependence η_C taking values in $(0, 1]$. If $\mu(\mathbb{E}_C^*) > 0$, the variables X_C can take their largest values simultaneously. If we instead consider subcones 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 μ places mass, which is 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 μ 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

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

so 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 can define a hidden regular variation assumption on $\mathbb{E}_C$. Writing $B_C = \{x \in \mathbb{E} : x_C \in B \subset (0, \infty]^{|C|},\ x_{D\setminus C} \in \{0\}^{|D\setminus C|}\}$, we have $\mathrm{pr}(X^*/t \in B_C) = \mathrm{pr}(X_C/t \in B,\ X_{D\setminus C} \in [0, -1/\log p]^{|D\setminus C|})$, so that we consider the

behaviour when the variables X_C are growing at a common rate, but the variables $X_{D\setminus C}$ have a fixed upper bound. However, the latter condition does not capture all possible behaviour that could lead 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 $(y,\infty]^C \times [0,z]^{D\setminus C} = \{x \in \mathbb{E} : x_i > y,\ i \in C;\ x_j \leqslant z,\ j \in D \setminus C\}$. Then under the regular variation assumption (4),

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

Therefore, if $\mu\{(y,\infty]^C \times \{0\}^{D\setminus C}\} > 0$ 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

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

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

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

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

The probability in (10), and hence $\tau_C(\delta)$, is nondecreasing in δ. The case of $\delta = 0$ 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 δ^*; if δ is too small we could have $\tau_C(\delta) < 1$ even when $\mu(\mathbb{E}_C) > 0$, but if δ is too large then some $\tau_C(\delta)$ could be close to 1 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 § 3.

Overall, examining the regular variation properties in Assumption 1 leads to understanding of the subasymptotic behaviour of μ 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 η_C.

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

Proof. We have

$$\mathrm{pr}\big(X_\wedge^C > t\big) = \mathrm{pr}\big\{X/t \in (1,\infty]^C \times [0,\infty]^{D\setminus C}\big\},$$
$$\mathrm{pr}\big(X_\wedge^{\bar{C}} > t, X_\vee^{D\setminus\bar{C}} \leqslant t\big) = \mathrm{pr}\big\{X/t \in (1,\infty]^{\bar{C}} \times [0,1]^{D\setminus\bar{C}}\big\}.$$

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

$$\mathrm{pr}\big\{X/t \in (1,\infty]^C \times [0,\infty]^{D\setminus C}\big\} = \sum_{\bar{C}:\, C\subseteq\bar{C}} \mathrm{pr}\big\{X/t \in (1,\infty]^{\bar{C}} \times [0,1]^{D\setminus\bar{C}}\big\},$$

from which the result follows. □

We remark that $\tau_D(\delta)$ does not depend on δ; we therefore denote it by τ_D.

THEOREM 2. *For all $C \in 2^D \setminus \emptyset$ with $|C| \geqslant 2$, assume regular variation or hidden regular variation on $\mathbb{E}_C^*$, with coefficient of tail dependence η_C, and suppose that Assumption 1 holds. For any set C with $|C| \geqslant 2$, assume that for any $\bar{C} \supseteq C$, $\eta_{\bar{C}'} < \eta_{\bar{C}}$ for all $\bar{C}' \supset \bar{C}$. Then $\tau_C(1) = \eta_C$, and $\tau_C(\delta) \leqslant \eta_C$ 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 $|C_{d-1}| = d-1$,

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

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 $|C_{d-2}| = d-2$,

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

Again, since $\eta_{C_{d-2}} > \eta_{C_{d-1}} > \eta_D$ for all $C_{d-2} \subset C_{d-1}$, we have $\eta_{C_{d-2}} = \tau_{C_{d-2}}(1)$. The result $\tau_C(1) = \eta_C$ follows by iteration for any set C with $|C| \geqslant 2$. Since $\tau_C(\delta)$ is nondecreasing in δ, $\tau_C(1) = \max_{\delta\in[0,1]} \tau_C(\delta)$, and so $\tau_C(\delta) \leqslant \eta_C$ 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. In 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 $|C| \geqslant 2$, and two classes of copula having $\chi_C = 0$ for all $|C| \geqslant 2$. 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

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

Table 1. *Values of $\tau_C(\delta)$ for some trivariate copula examples; for all logistic models the dependence parameter α satisfies $0 < \alpha < 1$, with larger α values corresponding to weaker dependence*

Case	$\lvert C\rvert = 1$	$\lvert C\rvert = 2$	$\lvert C\rvert = 3$
(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)	$\tau_1(\delta) = \tau_2(\delta) = \tau_3(\delta) = 1$	$\tau_{1,2}(\delta) = \tau_{1,3}(\delta) = \tau_{2,3}(\delta) = 2^{-\alpha}$	$\tau_{1,2,3} = 3^{-\alpha}$

(i) independence; (ii) independence and bivariate logistic; (iii) trivariate logistic; (iv) trivariate inverted logistic.

where H denotes the spectral measure defined in (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 θ_2, θ_1 and $\theta_{1,2} = 1 - (\theta_1 + \theta_2)$ 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., the two 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 $h(w) = \mathrm{d}H(w)/\mathrm{d}w$ 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), where $V(x, y) = (x^{-1/\alpha} + y^{-1/\alpha})^\alpha$ and

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

for $0 < w < 1$ and $\alpha \in (0, 1)$. 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 $\delta = 1$.

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 limiting mass on $\mathbb{E}_1$, $\mathbb{E}_2$ and $\mathbb{E}_3$. For the d-dimensional independence copula, $\tau_C(\delta) = |C|^{-1}$ for $|C| \leqslant d$ and does not depend on the value of δ. 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 limiting 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 & 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 it has subasymptotic dependence, reflected by the values of $\tau_C(\delta)$ for cones $\mathbb{E}_C$ with $|C| = 2, 3$, which are closer to 1 than in the independence case. The values of $\tau_C(\delta)$ do not depend on δ in this example.

The Gaussian copula with covariance matrix Σ also exhibits asymptotic independence, with all limiting 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 $\delta \geqslant \max(\rho_{ij}^2, \rho_{ik}^2)$, where $\rho_{ij} > 0$ is the Gaussian correlation parameter for variables i and j; otherwise, $\tau_C(\delta) < 1$. We also know that $\tau_{1,2,3} = \eta_{1,2,3} = (1_3^{\mathrm{T}} \Sigma^{-1} 1_3)^{-1}$, with $1_d \in \mathbb{R}^d$ denoting a vector of ones. For $C = \{i, j\}$ with $i < j$, under Assumption 1, Theorem 2 leads to $\tau_C(1) = \eta_C = (1_2^{\mathrm{T}} \Sigma_{i,j}^{-1} 1_2)^{-1}$, where $\Sigma_{i,j}$ denotes the submatrix of Σ corresponding to variables i and j, provided the correlations satisfy $1 + \rho_C \neq \sum_{C' : |C'|=2,\, C' \neq C} \rho_{C'}$; for $\delta < 1$ we have $\tau_C(\delta) \leqslant \eta_C$.

3. METHODOLOGY

3.1. *Introduction to our general approach*

The coefficient $\tau_C(\delta)$ defined in Assumption 1 reveals whether the measure μ 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 nondecreasing in δ, it is best to take δ as small as possible. We therefore have a trade-off between choosing δ large enough that $\tau_C(\delta) = 1$ on cones $\mathbb{E}_C$ with extremal mass and choosing it 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 to determining the extremal dependence structure of a set of variables. In the first method we take $\delta = 0$ and apply a truncation to the variables X by setting any values below some marginal threshold to zero. This transformation is analogous to the approach of Goix et al. (2017), who partitioned the nonnegative 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 seeking 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 use Assumption 1 with δ equal to zero by applying the 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 being 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

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

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

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

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

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

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) can 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

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

where u_C is a high threshold. Analytical maximization of (15) leads to closed-form estimates of (K_C, τ_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 τ_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 τ_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 $n_C \operatorname{pr}(Q > u_C \mid X^* \in E_C) \sim n_C K_C u_C^{-1/\tau_C}$, this entails $u_C = o(n_C^{\tau_C})$.

The method of Goix et al. (2017) produces empirical estimates of $\operatorname{pr}(X \in E_C \mid R > r_0)$ 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 $\operatorname{pr}(X \in E_C \mid R > r)$ changes with $r > r_0$, Goix et al. (2017) estimate this as a positive constant or zero. In contrast, our semiparametric method allows us to estimate $\operatorname{pr}(X^* \in E_C \mid Q > q)$ for all q above a high threshold, via

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

with E_C as in (13). Our estimate of the 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 subasymptotic 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 from the approach of Goix et al. (2017), which are common to our Methods 1 and 2, are illustrated in § 4 and § 5.

The right-hand side of (16) consists of two types of component. We estimate terms of the form $\operatorname{pr}(X^* \in E_C)$ empirically, and we estimate those of the form $\operatorname{pr}(Q > q \mid X^* \in E_C)$ as in (14) by replacing K_C and τ_C with their estimates and evaluating for some large choice of q, discussed in § 5. This approach yields an estimate for the proportion of mass in each region. We denote the estimated vector of these proportions by $\hat{p} = (\hat{p}_C : C \in 2^D \setminus \emptyset)$. 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 the use of the parameter π 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 the 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 $\text{pr}(X^* \in E_C \mid Q > q)$ for a large value of q by (16). Set $\text{pr}(X^* \in E_C \mid Q > q) = 0$ 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 π, set $\hat{p}_C$ to zero, renormalizing the resulting vector.

The parameter m ensures that 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 the 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 $\text{pr}(X_\wedge^C > t, X_\vee^{D\setminus C} \leqslant t^\delta) \in \text{RV}_{-1/\tau_C(\delta)}$. 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 $\text{pr}\{X_\wedge^C > t, X_\vee^{D\setminus C} \leqslant (X_\wedge^C)^\delta\} = \text{pr}(X_\wedge^C > t, X \in \tilde{E}_C)$, with $\tilde{E}_C$ defined as

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

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

$$\text{pr}(X_\wedge^C > q, X \in \tilde{E}_C) \in \text{RV}_{-1/\tilde{\tau}_C(\delta)}$$

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

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

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 $\tau_C(\delta)$. The regions $\tilde{E}_C$ are not disjoint. Supposing we have observations $x_1, \ldots, x_n$, we obtain an empirical estimate of $\text{pr}(X \in \tilde{E}_C)$ 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 1. Combining (17) and (18), we estimate

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

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

4. Simulation study

4.1. *Overview and metrics*

We conduct simulations to demonstrate our 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 (21). In the Supplementary Material we present results for a special case of this, the asymmetric logistic distribution (Tawn, 1990), which was 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 subasymptotic dependence, in contrast to independence, on certain subcones. 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 μ, as mentioned in § 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, π, above which estimated mass is considered nonnegligible. For $\pi = 0$, all cones will be included in the estimated dependence structure, leading to the true and false positive rates both being 1, whereas $\pi = 1$ includes none of the cones, so that both rates equal 0. A perfect result for a given dataset and method would be a false positive rate of 0 and a 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 $p = (p_C : C \in 2^D \setminus \emptyset)$ be 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 large penalty.

4.2. *Max-mixture distribution*

Segers (2012) shows how to construct distributions that place extremal mass on different combinations of cones. In this paper we take a different approach by considering max-mixture models with asymptotic and subasymptotic 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 let $\{Z_C : C \in 2^D \setminus \emptyset\}$ be independent random vectors. Define the vector

$X = (X_1, \ldots, X_d)$ with components

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

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 $\mathbb{E}_i$ for $i \in C$.

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 ρ, and define $Z_{1,2,3}$, $Z_{3,4,5}$ and $Z_{1,2,3,4,5}$ to have three-dimensional and five-dimensional extreme value logistic copulas with dependence parameter α. The bivariate Gaussian distribution is asymptotically independent with subasymptotic 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 the cones $\mathbb{E}_{1,2}$ and $\mathbb{E}_{4,5}$ have no mass asymptotically, but the parameter ρ 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.

Each setting in the simulation study is repeated 100 times, taking samples of size 10 000. In Method 1 we set $p = 0.5$, choose 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 take the value of q for which we estimate $\text{pr}(X^* \in E_C \mid Q > q)$ to be the 0.9999 quantile of all observed Q values. In Method 2 we set $\delta = 0.5$, take each threshold u_C to be the 0.85 quantile of observed $X_\wedge^C$ values in region $\tilde{E}_C$, and define 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 the 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 datasets, but rather are fixed at values that we have found to work well across a range of settings. In § 4.3 we discuss stability plots, which could be used as a guide to which tuning parameter values may be sensible for a given set of data. In § 5 we consider how the estimated extremal dependence structure changes as the tuning parameters vary for a particular dataset, allowing us to further examine this mass stability and choose a reasonable value of p in Method 1 or of δ in Method 2.

Figure 2 shows 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 curve are reported in Table 2. The performance of all three methods deteriorates as the value of the correlation parameter ρ, or of the dependence parameter α, increases. In the former case this is due to the stronger subasymptotic dependence on cones without extremal mass; in the latter case, larger values of α in the 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 $\rho = 0$ and 0.25, although its performance deteriorates when there is stronger correlation in the Gaussian components. Method 2 yields the best results for

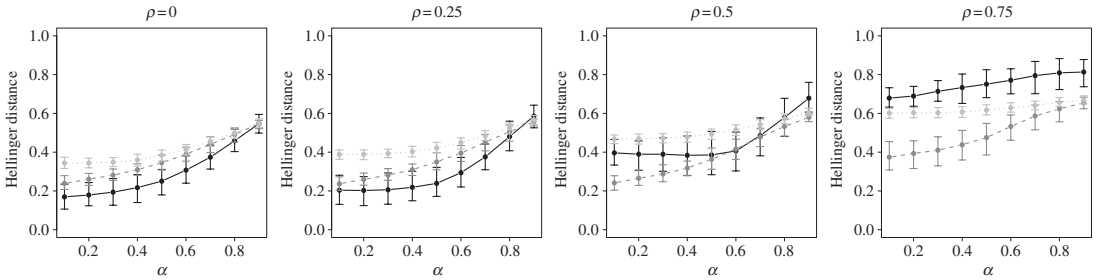

Fig. 2. Mean Hellinger distance together with 0.05 and 0.95 quantiles over 100 simulations, for Method 1 (solid), Method 2 (dashed) and the method of Goix et al. (2017) (dotted).

Table 2. *Average areas under the receiver operating characteristic curves: values are 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 parentheses*

		$\rho = 0$			$\rho = 0.25$	
α	0.25	0.5	0.75	0.25	0.5	0.75
Goix et al. (2017)	100 (0.0)	100 (0.0)	98.0 (1.1)	99.7 (0.4)	99.8 (0.4)	96.3 (1.4)
Method 1	100 (0.0)	100 (0.1)	97.7 (1.4)	100 (0.1)	99.9 (0.3)	96.7 (1.2)
Method 2	100 (0.0)	99.2 (0.7)	96.0 (1.6)	100 (0.1)	98.9 (0.8)	94.6 (1.8)
		$\rho = 0.5$			$\rho = 0.75$	
α	0.25	0.5	0.75	0.25	0.5	0.75
Goix et al. (2017)	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	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	99.5 (0.6)	97.5 (1.1)	92.7 (1.7)	94.4 (1.9)	92.9 (1.9)	89.1 (2.0)

$\rho = 0.5$ and 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 the method of Goix et al. (2017) for this max-mixture model. The results in Table 2 reveal that all three methods are successful classifiers for low values of ρ and α. For $\alpha = 0.75$ and $\rho = 0, 0.25$, Method 1 and the approach of Goix et al. (2017) demonstrate similarly strong performance, while for $\rho = 0.5, 0.75$ Method 2 again yields the best results.

As a further comparison of the methods, Fig. 3 shows how often each cone $\mathbb{E}_C$ is detected as having mass above $\pi = 0.01$ or 0.001 for the $(\alpha, \rho) = (0.75, 0.5)$ case. For $\pi = 0.001$, the approach of Goix et al. (2017) places mass on approximately three times as many cones as our Methods 1 and 2, and over twice as many in 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. (2017) assuming that there is extremal mass on a cone $\mathbb{E}_C$ if $\text{pr}(X \in E_C \mid R > r_0) > \pi$, whereas we recognize that when $\hat{\tau}_C < 1$ or $\hat{\tau}_C(\delta) < 1$, nonlimiting 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 the cone $\mathbb{E}_C$ similarly to Goix et al. (2017). As a consequence, our approach integrates information over the entire tail to estimate which cones have limiting mass, as opposed to the method of Goix et al. (2017), which uses information at only 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 it places more mass on lower-dimensional cones, arising from

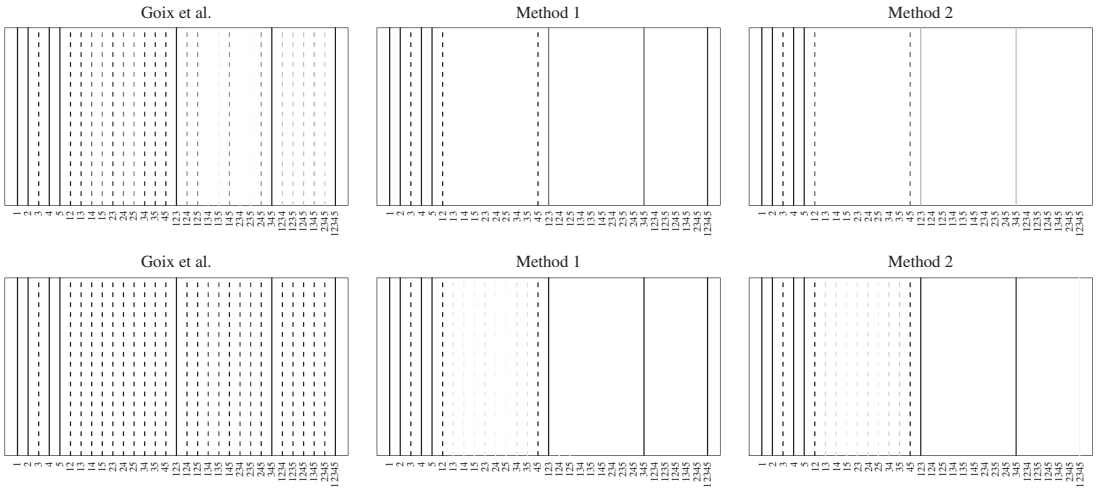

Fig. 3. Number of times each cone is assigned mass greater than $\pi = 0.01$ (top) or $\pi = 0.001$ (bottom), for $(\alpha, \rho) = (0.75, 0.5)$: true cones with mass (solid) and cones without mass (dashed); darker lines correspond to higher detection rates over 100 simulations.

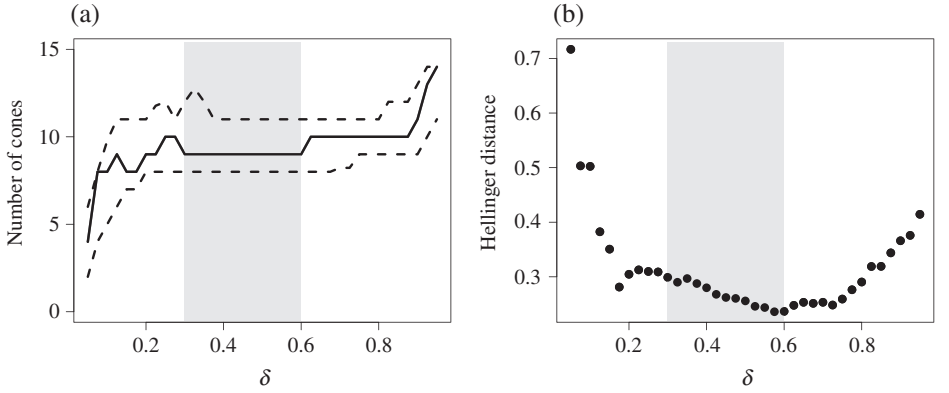

Fig. 4. (a) Stability plot for Method 2, with dashed lines delineating a 95% bootstrapped confidence interval for the number of cones $\mathbb{E}_C$ with mass; (b) plot of the Hellinger distance for each value of δ. In each panel the shaded region corresponds to the stable range of tuning parameter values. Data were simulated from the max-mixture distribution of § 4.2 with $n = 10\,000$, $\alpha = 0.25$ and $\rho = 0.25$.

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. In this subsection we outline how to construct such a plot for an example using the max-mixture distribution of § 4.2 with Method 2, where our aim is to obtain a sensible range of values for the tuning parameter δ, by considering the region of δ where the number of cones determined as having mass is stable.

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 we find the number of cones whose estimated mass is greater than $\pi = 0.001$ in each case. The remaining parameters are fixed as in § 4.2. Figure 4 shows the estimates of the number of cones, with 95% confidence intervals constructed from 250 bootstrapped samples;

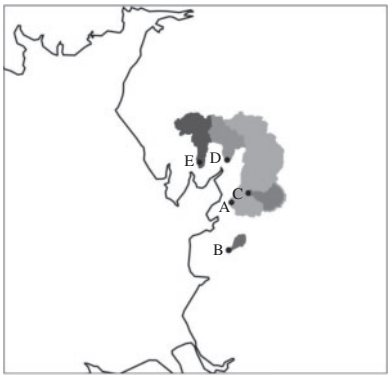

Fig. 5. Locations of the river flow gauges, labelled A to E; the grey-shaded regions represent the corresponding catchments, and the solid black line depicts the coast.

together these constitute the stability plot. Analogous plots can be created to choose p in Method 1, or to choose π in each case. In practice, the choice of the threshold π should depend on the dimension of the data, but this issue is not explored here.

The number of cones detected as having mass is most stable for values of δ between 0.3 and 0.6, indicated by the shaded regions in Fig. 4, suggesting that values of δ in this range may be appropriate for this sample. Figure 4(b) shows the Hellinger distance corresponding to the set of estimated proportions obtained for each value of δ. For this particular sample, although values of δ within the stable range slightly overestimate the number of cones with mass, the smallest Hellinger distance occurs for a value of δ 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; but Fig. 4 supports the idea of using stability plots to choose suitable tuning parameter values. There is no guarantee that stability plots will reveal 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 δ values for different cones $\mathbb{E}_C$.

5. River flow data

We apply Methods 1 and 2 to daily mean river flow readings, in cubic metres per second, measured at five gauging stations in the north-west of England from 1980 to 2013. These data are available from the Centre for Ecology and Hydrology at `nrfa.ceh.ac.uk` (Morris & 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 obtain an overall model. The locations of the five gauges are shown in Fig. 5; the labels assigned to the different locations will be used to describe the dependence structures estimated in this section. Figure 5 also shows the catchment associated with each gauge. These catchments represent the areas from which surface water, usually the result of precipitation, will drain to each gauge. The spatial dependence of river flow has been 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

Table 3. *Proportion (%) of mass assigned to each cone, for varying values of the tuning parameters in Method 1 (upper part) and Method 2 (lower part); the grey-shaded portions indicate the feasible stable ranges*

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							19	1	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
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

with adjacent or overlapping catchments are expected to take their largest values simultaneously, with stronger dependence between gauges that are closer together.

Table 3 shows the percentage of 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 we take the extrapolation level q to be the 0.999 quantile of the observed Q and X values in Methods 1 and 2, respectively. The remaining parameters are fixed as in § 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 the moment constraint (3) should be taken into account. For Method 2, Table 3 indicates that for $\delta \geqslant 0.45$, the cone corresponding to location E is assigned more than 20% of the extremal mass, which is not possible because of the moment constraint. Feasible stable regions are indicated by the shaded parts of Table 3. For Method 2, one could also look for a value of δ that gives estimates of $\tau_C(\delta)$ satisfying $\max_{C:\, C \supseteq i} \hat{\tau}_C(\delta) = 1$, subject to estimation uncertainty, for every $i = 1, \ldots, d$.

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 either are also detected by Method 1,

Table 4. *Proportion (%) of mass assigned to each cone for locations A, B, C and D, for varying values of the tuning parameters in Method 1 (upper part) and Method 2 (lower part); the grey-shaded portions indicate the feasible stable ranges*

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
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

or are neighbours of cones detected by Method 1, showing that 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 the 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 one considers 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 the region under consideration. 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 found to take their largest values simultaneously, which may be a consequence of this nested structure.

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 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} by 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 it was assigned only 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 δ increases, 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 δ, 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 δ varies. 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 δ, this is still a huge reduction compared with the full set of possibilities.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of the U.K. Engineering and Physical Sciences Research Council through the EP/L015692/1 STOR-i centre for doctoral training and fellowship EP/P002838/1. We acknowledge the National River Flow Archive Centre for Ecology and Hydrology for use of the river flow and catchment boundary data. We thank the referees and associate editor for their comments.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes calculations of $\tau_C(\delta)$ for the copulas in Table 1, simulation results for the estimates of $\tau_C(\delta)$ in Method 2, plots of simulation results for the max-mixture distribution of § 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

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

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

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

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) gives

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

By this result,

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

If $\theta_1 > 0$, i.e., the spectral measure places mass on $\{1\}$, we see that $\mathrm{pr}(X_1 > t, X_2 < t^\delta) \sim 2\theta_1 t^{-1}$ as $t \to \infty$, and 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 $\theta_1 + \theta_2 = 1$.

References

Asadi, P., Davison, A. C. & Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Statist.* **9**, 2023–50.

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electron. J. Statist.* **9**, 383–418.

Chiapino, M. & Sabourin, A. (2017). Feature clustering for extreme events analysis, with application to extreme stream-flow data. In *New Frontiers in Mining Complex Patterns*. Cham, Switzerland: Springer, pp. 132–47.

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

Das, B., Mitra, A. & Resnick, S. I. (2013). Living on the multidimensional edge: Seeking hidden risks using regular variation. *Adv. Appl. Prob.* **45**, 139–63.

Goix, N., Sabourin, A. & Clémençon, S. (2016). Sparse representation of multivariate extremes with applications to anomaly ranking. *J. Mach. Learn. Res. W&CP* **41**, 75–83. Proc. 19th Int. Conf. Artif. Intel. Statist. (AISTAT 2016), Cadiz, Spain.

Goix, N., Sabourin, A. & Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *J. Mult. Anal.* **161**, 12–31.

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

Heffernan, J. E. & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with Discussion). *J. R. Statist. Soc.* B **66**, 497–546.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–74.

Hua, L. & Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *J. Mult. Anal.* **102**, 1454–71.

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

Klüppelberg, C., Haug, S. & Kuhn, G. (2015). Copula structure analysis based on extreme dependence. *Statist. Interface* **8**, 93–107.

Ledford, A. W. & Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–87.

Ledford, A. W. & Tawn, J. A. (1997). Modelling dependence within joint tail regions. *J. R. Statist. Soc.* B **59**, 475–99.

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

Mitra, A. & Resnick, S. I. (2011). Hidden regular variation and detection of hidden risks. *Stoch. Mod.* **27**, 591–614.

Morris, D. G. & Flavin, R. W. (1990). A digital terrain model for hydrology. In *Proc. 4th Int. Sympos. Spatial Data Handling (July 23–27, Zurich)*, vol. 1, pp. 250–62.

Morris, D. G. & Flavin, R. W. (1994). *Sub-set of UK 50m by 50m Hydrological Digital Terrain Model Grids*. Wallingford: Institute of Hydrology, NERC.

Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5**, 303–36.

RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.
SEGERS, J. (2012). Max-stable models for multivariate extremes. *REVSTAT Statist. J.* **10**, 61–82.
TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75**, 397–415.
TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–53.
VETTORI, S., HUSER, R., SEGERS, J. & GENTON, M. G. (2020). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *J. Comp. Graph. Statist.* **29**, 174–90.
WADSWORTH, J. L. & TAWN, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli* **19**, 2689–714.

[*Received on* 28 *November* 2017. *Editorial decision on* 17 *October* 2019]