Multivariate extremes

Anne-Laure Fougères
Laboratoire de Statistique et Probabilités
INSA de Toulouse - Université Paul Sabatier[1]

1. Introduction.

A wide variety of situations concerned with extreme events has an inherent multivariate character, as pointed out by Coles & Tawn (1991). Let us consider for example the oceanographic context, and focus on the sea-level process. Such a variable can be divided into several physical components like mean-level, tide, surge and wave, which are driven by different physical phenomena (see for example Tawn (1992) for details). Moreover, extreme sea conditions leading to damages are usually a consequence of extreme values jointly in several components. The joint structure of the processes has therefore to be studied. Another type of dependence which can be of great interest is the temporal one: High sea levels can be all the more dangerous when they last for a long period of time. Therefore, a given variable observed at successive times is likely to contain crucial information. Other examples of applications have been listed recently by Kotz & Nadarajah (2000), concerning among others pollutant concentrations (Joe, Smith & Weissman, 1992), reservoir safety (Anderson & Nadarajah, 1993), or Dutch sea dikes safety (Bruun & Tawn, 1998; de Haan & de Ronde, 1998).

Historically, the first direction which has been explored concerning multivariate extreme events was the modeling of the asymptotic behaviour of componentwise maxima[2] of independent and identically distributed (i.i.d.) observations. Key early contributions to this domain of research are, among others, the papers of Tiago de Oliveira (1958), Sibuya (1960), de Haan & Resnick (1977), Deheuvels (1978) and Pickands (1981). The general structure of the multivariate extreme value distributions has been explored by de Haan & Resnick (1977). Useful representations in terms of max-stable distributions, regular variation functions, or point processes, have been established. Section 2 is devoted to the asymptotic model for componentwise maxima. The main results are sketched in Section 2.2, after a brief summary of the univariate extreme

[1] Dépt. GMM, INSA, 135 Avenue de Rangueil, 31077 Toulouse Cedex 04, France.

[2] Note that results on minima can immediately be deduced using the property that for any variable X, $\min X = -\max(-X)$. All what follows will just be written for maxima.

value context. Statistical inference developed in this setup will also be summarized. The limitations of this way of modeling multivariate extreme events will then be envisaged in Section 3, where we focus on asymptotically independent events. Recent alternatives introduced by Ledford & Tawn (1996, 1997) will be presented. Finally, the problem of how to measure extremal dependence is tackled, and some tools are reviewed.

2. Classical results on componentwise maxima.

2.1. Univariate extreme events: Summary.

The problem of how to model the tails of a univariate distribution has been widely studied, and presents a myriad of applications, as recently listed by Kotz & Nadarajah (2000, Sections 1.1, 1.9 and 2.8). The key assumption which underlies all the methods of modeling is the existence of a domain of attraction for the maxima, that is: If $X_1, \ldots, X_n$ are i.i.d. observations of a random phenomenon with distribution function (d.f.) F, then there exist two sequences $(a_n)_n$ and $(b_n)_n$, where $a_n > 0,\ \ b_n \in \mathbb{R}$, and a nondegenerate d.f. G such that

$$\lim_{n\to\infty} P\left\{ \frac{\max X_i - b_n}{a_n} \leq x \right\} = \lim_{n\to\infty} F^n(a_n\, x + b_n) = G(x). \tag{1}$$

The set of d.f. G such that (1) holds is now referred as the *Generalized Extreme Value (GEV) family*, introduced by von Mises (1954) and Jenkinson (1955). The d.f. G has the following parametric form:

$$G(x) \;=\; \exp\left\{ - \left(1 + \xi\, \frac{x-\mu}{\sigma}\right)_+^{-1/\xi} \right\},$$

where the notation a_+ stands for a if $a > 0$ and 0 otherwise, and where $\sigma > 0$ and $\mu, \xi \in \mathbb{R}$. This distribution function was originally called Fréchet, Weibull, or Gumbel distribution, depending on whether the shape parameter ξ is positive, negative or zero (as a limiting case). Fisher & Tippett (1928) exhibited these three types of distributions for G, see also Gnedenko (1943).

Under the fundamental hypothesis of the existence of a domain of attraction for the maxima, two main ways of modeling extreme events have emerged: Firstly, models for block maxima, based on the representation (1), where the asymptotic distribution of the (renormalized) maxima is considered as an approximation for the distribution of the maxima over a fixed (large enough) number of observations. Statistical inference in the GEV parametric family has been largely studied, see for example a review in Kotz & Nadarajah (2000), Sections 2.2 to 2.6. Secondly, threshold methods have been considered, based on the asymptotic form of the distribution of excesses over a given threshold: More precisely, two avenues have been exploited: the first one, referred to as the "peaks over threshold" (POT) method, is based on the generalized Pareto approximation. If the d.f. of a random variable (r.v.) X is in the domain of attraction of a GEV distribution with parameters (μ, σ, ξ), then the conditional distribution function of exceedances has the following property

$$P\,(X > u + x \mid X > u) \;\sim\; \left(1 + \xi\,\frac{x}{\tilde{\sigma}}\right)_+^{-1/\xi},\; u \to \infty$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$. This approach is due to Balkema & de Haan (1974) and Pickands (1975), and has been widely studied (see Leadbetter, Lindgren & Rootzén (1983), for example). Davison & Smith (1990) make a review of the statistical properties of this method, and focus also on the problem of the choice of the threshold. The second way of using threshold models is to approximate the point process associated to observations greater than u by a nonhomogeneous Poisson process with intensity measure of (x, ∞) given by $(1 + \xi\;(x - \mu)/\sigma\,)_+^{-1/\xi}$. This has been studied by Pickands (1971) and Smith (1989). Adapted methods have also been developed when extremes of dependent sequences are of interest, which is actually the usual case when considering notably environmental data. See for example Leadbetter, Lindgren & Rootzén (1983), Smith (1989), Davison & Smith (1990). Besides, the non stationary frame has been explored by Leadbetter, Lindgren & Rootzén (1983) and Hüsler (1986), among others. We refer to Coles (2001) for a review of practical methodologies when dealing with such data. For an introduction to univariate extreme value theory with applications to insurance and finance, see Embrechts, Klüppelberg & Mikosch (1997).

2.2. Multivariate extreme value distributions.

As mentioned in the Introduction, exploration of how to model multivariate extreme events began with the study of the limiting behaviour of componentwise maxima. All the theory developed is based, as in the univariate case, on the existence of a domain of attraction. Denote in bold-face elements $\mathbf{x} = (x_1, \ldots x_d)$ of $\mathbb{R}^d$. If $\mathbf{X}_i = (X_{i,1}, \ldots X_{i,d}), i = 1, \ldots, n$, are i.i.d. random vectors of dimension d with d.f. F, one assumes that there exist $\mathbb{R}^d$-sequences $(\mathbf{a}_n)_n$ and $(\mathbf{b}_n)_n$, where $a_{n,j} > 0$ and $b_{n,j} \in \mathbb{R}$ for all $j = 1, \ldots, d$, and a d.f. G with non-degenerate margins such that

$$P\left\{(\max_{i=1,\ldots,n} \mathbf{X}_i - \mathbf{b}_n)/\mathbf{a}_n \leq \mathbf{x}\right\} = F^n(\mathbf{a}_n\,\mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}), \tag{2}$$

when $n \to \infty$. The d.f. G is then called a *multivariate EV distribution function*, and one says that F is in the *(multivariate) domain of attraction* of G (for the maxima). Note in particular that the univariate margins of G are EV distributions.

Example 1 **(i)** Consider the multivariate normal d.f. $F_{\mathcal{N}}$, with all univariate margins equal to $\mathcal{N}(0,1)$, and with all its correlations less than 1 ($\mathbb{E}X_iX_j < 1$, for all $i, j = 1, \cdots, d$). Such a distribution is in the domain of attraction of the independence with univariate Gumbel margins (Sibuya, 1960). Indeed, one has that

$$F_{\mathcal{N}}^n(a_n\mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}) = \prod_{j=1}^{d} \exp\{-e^{-x_j}\}.$$

The norming constants are respectively equal to $a_n = (2\log n)^{-1/2}$ and $\mathbf{b}_n = b_n\mathbf{1}$, where $b_n = (2\log n)^{1/2} - 1/2(\log\log n + \log 4\pi)/(2\log n)^{1/2}$, and $\mathbf{1} = (1, \cdots, 1)$ (see for example Resnick (1987), Example 2).

(ii) Next consider the Archimedean d.f. F_ϕ, with all univariate margins uniformly distributed on $[0,1]$, introduced by Genest & MacKay (1986a, 1986b), and defined by

$$F_\phi(\mathbf{x}) = \phi^{-1}\left\{\sum_{j=1}^{d} \phi(x_j)\right\},$$

where ϕ is a function defined on $(0,1]$ such that $\phi(1) = 0$ and $(-1)^j \mathrm{d}^j\phi^{-1}(t)/\mathrm{d}t^j \geq 0$, for all $j = 1, \cdots, d$. Moreover, assume that $\phi(1 - 1/t)$ is a regularly varying function at infinity with index $-m$, for some $m \geq 1$. Recall that a function $\psi : (0,\infty) \to (0,\infty)$ is said to be *regularly varying at infinity with index* ρ, denoted $\psi \in RV_\rho$, if and only if $\lim_{t\to\infty} \psi(st)/\psi(t) = s^\rho$ for all $s > 0$ (e.g., Bingham, Goldie & Teugels , 1989). These distributions are in the domain of attraction of the logistic EV distribution (see further in Example 2), namely :

$$F_\phi^n(\mathbf{x}/n + \mathbf{1}) \to \exp\left[-\left\{\sum_{j=1}^{d}(-x_j)^m\right\}^{1/m}\right],$$

for all $\mathbf{x} < \mathbf{0}$, where $\mathbf{0} = (0, \cdots, 0)$. This last result is due to Genest & Rivest (1989).

Even if the parametric character of the univariate EV family of distributions is now lost in the multivariate context, as a subset of the max-infinite divisible distributions, a specific structure still remains. The results sketched here are essentially due to de Haan & Resnick (1977), and are for example presented in Galambos (1987), Resnick (1987, Chap. 5) or Kotz & Nadarajah (2000, Chap. 3).

Let us first assume for convenience that the univariate extreme value margins follow unit Fréchet distributions (with d.f. defined for all $y > 0$ by $\phi_1(y) = e^{-1/y}$). This standardization leads to a separation of the marginal behaviour and the dependence part of the distribution. There is no loss of generality in assuming specific margins, as stated in Proposition 5.10 by Resnick (1987). Note that in the case of unit Fréchet margins, normalization sequences $(a_{n,j})$ and $(b_{n,j})$ can be shown to be respectively equal to $a_{n,j} = n$ and $b_{n,j} = 0$, for all $j = 1, \ldots, d$.

The following characterizations of the multivariate EV distributions can then be obtained (see for example Proposition 5.11 of Resnick, 1987). The set E denotes here $E = [0, \infty]^d \setminus \{\mathbf{0}\}$. The symbol $\bigvee$ is used for supremum. The function $\mathbb{1}$ is defined by $\mathbb{1}_{z \in C} = 1$ if $z \in C$, and 0 otherwise, and $|| \cdot ||$ denotes any norm on $\mathbb{R}^d$.

Theorem 1 *The following assertions are equivalent:*

[C1] *G is a multivariate EV distribution with unit Fréchet margins.*

[C2] *There exists a finite measure S on $\mathcal{B} = \{\mathbf{y} \in E : ||\mathbf{y}|| = 1\}$ such that for each $\mathbf{x} = (x_1, \ldots, x_d) \in E$, one has that*

$$G(\mathbf{x}) = \exp\left\{ - \int_{\mathcal{B}} \bigvee_{j=1}^{d} \frac{w_j}{x_j} \, dS(\mathbf{w}) \right\},$$

with

$$\int_{\mathcal{B}} w_j \, dS(\mathbf{w}) = 1, \quad \text{for all } j = 1, \ldots, d. \tag{3}$$

[C3] *There exists a non homogeneous Poisson process $\sum_k \mathbb{1}_{(t_k, \mathbf{j}_k) \in \,\cdot}$ on $[0, \infty) \times E$ with intensity measure Λ defined, for $t > 0$ and $B \subset E$, by $\Lambda([0, t] \times B) = t\mu^*(B)$, where for all $A \subset \mathcal{B}$ and $r > 0$,*

$$\mu^* \left\{ \mathbf{y} \in E \ : \ ||\mathbf{y}|| > r \ ; \ \frac{\mathbf{y}}{||\mathbf{y}||} \in A \right\} = \frac{S(A)}{r}, \tag{4}$$

and S is a finite measure such that (3) *holds and*

$$G(\mathbf{x}) = P\left(\bigvee_{t_k \le 1} \mathbf{j}_k \le \mathbf{x}\right) = \exp\left(-\mu^*\{(\mathbf{0}, \mathbf{x}]^c\}\right).$$

Remark 1 Conditions (3) secure that the margins are all unit Fréchet distributed. Equation (4) shows that the measure μ^* composed with the application T^{-1} defined from $T : E \to (0,\infty] \times \mathcal{B}$, $\mathbf{y} \mapsto (||\mathbf{y}||, \mathbf{y}/||\mathbf{y}||)$, is a product measure of a simple function of the "radial" component and a measure S of the "angular" component. More precisely, one has $\mu^* \circ T^{-1}\{(r,\infty) \times A\} = S(A)/r$ for all $A \subset \mathcal{B}, r > 0$, so that

$$\mu^*\{(\mathbf{0}, \mathbf{x}]^c\} \;=\; \mu^* \circ T^{-1}(T\{(\mathbf{0}, \mathbf{x}]^c\}) \;=\; \int_{T\{(\mathbf{0},\mathbf{x}]^c\}} \frac{1}{r^2}\, dS(\mathbf{w})dr. \tag{5}$$

Moreover, writing $T((\mathbf{0}, \mathbf{x}]^c) = \{(r,w) \in (0,\infty) \times \mathcal{B} \;:\; rw \in (\mathbf{0}, \mathbf{x}]^c\}$

$$= \{(r,w) \in (0,\infty) \times \mathcal{B} \;:\; r > \bigwedge_{j=1}^{d} \frac{x_j}{w_j}\},$$

and using this last expression in (5) leads to

$$\mu^*\{(\mathbf{0}, \mathbf{x}]^c\} = \int_{\mathcal{B}} \bigvee_{j=1}^{d} \frac{w_j}{x_j}\, dS(\mathbf{w}).$$

The measure S is often called *spectral measure*, and μ^* is the *exponent measure*. Finally, without going into the proof, note that the key result leading to Theorem 1 is that the EV distributions coincide with the max-stable distributions. Assuming unit Fréchet margins, these distributions are of the form $G(\mathbf{x}) = \exp(-\mu^*\{(\mathbf{0}, \mathbf{x}]^c\})$, for all $\mathbf{x} \in E$, where μ^* is a measure satisfying the homogeneity property $t\mu^*(tB) = \mu^*(B)$, for all $t > 0$ and B a Borel set of E.

Example 2 **(i)** A particular and important case is the case of independence. It corresponds in the representation [C2] to a measure S which is concentrated on $\{e_i, i = 1, \cdots, d\}$, where $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ are the vectors of the canonical basis of $\mathbb{R}^d$ (see for example Corollary 5.25 of Resnick, 1987).

(ii) Several parametric families of bivariate and multivariate EV distributions have been proposed by Tawn (1988), Coles & Tawn (1991), Joe (1990) and Tawn (1990), among others. See Kotz & Nadarajah (2000) for a recent review of these existing parametric models. One of the most classical is the so-called *logistic model*, proposed by Gumbel (1960), and defined for $\mathbf{x} > \mathbf{0}$ by

$$G_\alpha(\mathbf{x}) = \exp\left\{ - \left(\sum_{j=1}^{d} x_j^{-\alpha} \right)^{1/\alpha} \right\}, \tag{6}$$

for some parameter $\alpha \geq 1$. The limit case $\alpha = 1$ corresponds to the independence between the variables. Different asymmetric generalizations of this family have been proposed, see for example the nested logistic model (Coles & Tawn, 1991).

In the bivariate case, the family of EV distributions can be represented in a different way. Indeed, Pickands (1981) has shown that a bivariate d.f. G is an EV d.f. with unit Fréchet margins if and only if

$$G(\mathbf{x}) = \exp\left\{ - \left(\frac{1}{x_1} + \frac{1}{x_2} \right) A\left(\frac{x_2}{x_1 + x_2} \right) \right\}, \tag{7}$$

where A is a convex function, $A : [0,1] \to [1/2, 1]$, such that $\max(t, 1-t) \leq A(t) \leq 1$ for all $0 \leq t \leq 1$. See also Sibuya (1960) and Tiago de Oliveira (1975, 1980) for other bivariate representations. The measure S defined in representation [C2] is thus related to the function A by the following

$$A(t) = \int_{\mathcal{B}} \max\{tw_1, (1-t)w_2\}\ dS(\mathbf{w}).$$

Hence, except for the margins, the d.f. G is characterized by a one-dimensional function A, referred as the "dependence function". Particular examples for A are $A(t) = 1$, for all $0 \leq t \leq 1$, which corresponds to the independence for G, or $A(t) = \max(t, 1-t)$ corresponding to total positive dependence. The logistic model defined in (6) corresponds to $A_\alpha(t) = \exp\{t^\alpha + (1-t)^\alpha\}^{1/\alpha}$, for $0 \leq t \leq 1$ and $\alpha \geq 1$.

At this stage, it is of practical importance to examine what the different characterizations yield when formulated from a d.f. F belonging to a specific domain of attraction. As before, one considers i.i.d. observations $\mathbf{X}_i = (X_{i,1}, \ldots X_{i,d}), i = 1, \ldots, n$, which are assumed to have unit Fréchet margins. In practice, when the margins have unknown distributions, one may for example transform the observations $X_{i,j}$, $(i = 1, \ldots, n$, and $j = 1, \ldots, d)$ into the pseudo-observations $Z_{i,j} = 1/\log\{n/(R_{i,j} - 1/2)\}$, where $R_{i,j}$ is the rank of $X_{i,j}$ among $X_{1,j}, \ldots, X_{n,j}$. Such a transformation, suggested in Joe, Smith & Weissman (1992), ensures the $Z_{i,j}$ to be in the univariate domain of attraction of a unit Fréchet distribution. Refer also to Mason & Huang's work for the estimation of the dependence function using ranks only (see Huang, 1992, or Drees & Huang, 1998). We use the same notations as in Theorem 1.

Theorem 2 *The following statements are equivalent:*

[D1] *The d.f. F of the $\mathbf{X}_i$'s ($i = 1, \ldots, n$) is in the domain of attraction of a multivariate EV distribution G with unit Fréchet margins.*

[D2] $\lim_{t\to\infty} \frac{-\log F(t\mathbf{x})}{-\log F(t\mathbf{1})} = \lim_{t\to\infty} \frac{1-F(t\mathbf{x})}{1-F(t\mathbf{1})} = \frac{-\log G(\mathbf{x})}{-\log G(\mathbf{1})} = \frac{\mu^*([\mathbf{0},\mathbf{x}]^c)}{\mu^*([\mathbf{0},\mathbf{1}]^c)}$, *where the measure μ^* is defined in Theorem 1.*

[D3] $\lim_{t\to\infty} t\, P\left\{||\mathbf{X}_i|| > t\ ;\ \frac{\mathbf{X}_i}{||\mathbf{X}_i||} \in A\right\} = S(A)$, *for each A Borel set of $\mathcal{B}$ and $i = 1, \ldots, n$, where S is defined in Theorem 1.*

[D4] *The point process associated with $\{\mathbf{X}_1/n, \ldots, \mathbf{X}_n/n\}$ converges weakly to a non homogeneous Poisson process on E with intensity measure μ^*.*

Remark 2 Note that expression [D3] is also equivalent to

$$[D3'] \qquad \lim_{t\to\infty} P\{\mathbf{X}_i\,/\,||\mathbf{X}_i|| \in A \mid ||\mathbf{X}_i|| > t\} = S(A)/S(\mathcal{B}).$$

Formulation [D3] clearly suggests a simple nonparametric way to estimate the measure S associated with an EV d.f. G from observations that are in the domain of attraction of G. If a sample $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is available from F, then, provided that t is a well chosen function of n which ensures convergence, a natural candidate to estimate S is deduced from the empirical measure of the $(||\mathbf{X}_i||/t, \mathbf{X}_i/||\mathbf{X}_i||)$'s, that is $1/n \sum_{i=1}^n \mathbb{1}\{(||\mathbf{X}_i||/t, \mathbf{X}_i/||\mathbf{X}_i||) \in \cdot\ \}$. More precisely, such a convergence is achieved as soon as $t = n/k_n$, where $(k_n)_n$ is a sequence of integers such that $k_n \to \infty$ and $n/k_n \to \infty$ when $n \to \infty$ (Resnick, 1986, Proposition 5.3). An estimator of S can therefore be obtained using the observations $\mathbf{X}_i$ such that $||\mathbf{X}_i|| > n/k_n$. From a practical point of view, it is usually more convenient to replace the condition $\{||\mathbf{X}|| > n/k_n\}$ by $\{||\mathbf{X}|| > ||\mathbf{X}||_{[k_n]}\}$, where $||\mathbf{X}||_{[k_n]}$ denotes the $(n - k_n + 1)$th order statistic of the $||\mathbf{X}||_i$'s. Both conditions are asymptotically equivalent (see for example Appendix 3 of Capéraà & Fougères, 2000); the second one offers the advantage of keeping a fixed number of observations for estimating S. This finally leads to the estimator S_n of S, defined for any Borel set A of $\mathcal{B}$, by :

$$S_n(A) = \frac{1}{k_n} \sum_{i=1}^{n} \mathbb{1} \left\{ ||\mathbf{X}_i|| > ||\mathbf{X}||_{[k_n]} \ , \ \frac{\mathbf{X}_i}{||\mathbf{X}_i||} \in A \right\}. \tag{8}$$

A variety of nonparametric estimation techniques have been developed from the useful representation [D3], referred as "multivariate threshold methods". The estimators proposed essentially differ from each other in terms of the choice of the norm and of a function closely related to the mapping T defined in Remark 1. The developments have been mostly formulated in the bivariate case. See for example de Haan (1985a), Joe, Smith & Weissman (1992), de Haan & Resnick (1993), Einmahl, de Haan & Huang (1993), Einmahl, de Haan & Sinha (1997), as well as Capéraà & Fougères (2000) for a small-sample study comparing several methods from the previously cited works. Further interesting references are de Haan & de Ronde (1998), de Haan & Sinha (1999), Abdous, Ghoudi & Khoudraji (1999) and Einmahl, de Haan & Piterbarg (2001). Note that no theoretical result has been obtained yet concerning an optimal choice for the threshold k_n in the multivariate setup. In practice, choosing k_n is a tricky problem, and usually several values are considered, for which a relative stability of the estimations is expected. Improving this empirical choice, Abdous & Ghoudi (2002) proposed recently an interesting and convenient procedure for an optimal threshold selection, based on a double kernel technique (Devroye, 1989). Abdous & Ghoudi also suggested a unifying approach which includes most of the estimates previously mentioned. Parametric approaches have been proposed by Coles & Tawn (1991, 1994) and Joe, Smith & Weissman (1992), among others. These approaches make use of the point process representation [D4] and of parametric families of multivariate EV distributions. Finally, some parametric models based on the representation [D2] have also been introduced by Ledford & Tawn (1996) and Smith, Tawn & Coles (1997). Even if traditionally "parametric and nonparametric schools" seem to confront each other, they present complementary advantages, and nonparametric estimations can notably be used as a starting point for inference, on which flexible parametric models can be built.

Remark 3 In the particular case where observations from an EV d.f. G are directly available, specific techniques which differ from the threshold methods have been developed in the bivariate case by Pickands (1981), Tawn (1988), Tiago de Oliveira (1989), Smith, Tawn & Yuen (1990), Deheuvels (1991), Coles & Tawn (1991), Capéraà, Fougères & Genest (1997), Hall & Tajvidi (2000), among others. In such a situation where no selection is needed above a sufficiently high threshold, estimation techniques are of course much more accurate.

Below we summarize for a simple case the practical estimation of the probability of an extreme event via multivariate EV models. Given a sample of random vectors $\mathbf{X}_i, i = 1, \ldots, n$, with d.f. F, consider the problem of estimating the probability $P(\mathbf{X} \in A)$, where A is an exceptional set in which none data have been observed. We assume that F is in the domain of attraction of a multivariate EV d.f. G, and again, for simplicity we deal with known margins and assume that they are unit Fréchet distributed. If A is the complement of a rectangle, $A = (\mathbf{0}, n\mathbf{u}]^c$, one may write for example, using [D2]:

$$P(\mathbf{X} \in A) = 1 - F(n\mathbf{u}) \approx -\frac{1}{n} \log G(\mathbf{u}) = \frac{1}{n}\mu^*\{(\mathbf{0}, \mathbf{u}]^c\} = \frac{1}{n} \int_{\mathcal{B}} \bigvee_{j=1}^{d} \frac{w_j}{u_j} dS(\mathbf{w}).$$

Making use of the empirical measure S_n defined in (8), an estimator of $P(\mathbf{X} \in A)$ is then given by

$$\frac{1}{nk_n} \sum_{i=1}^{n} \left(\bigvee_{j=1}^{d} \frac{X_{i,j}}{u_j ||X_{i,j}||} \right) 1\!\!1\{||\mathbf{X}_i|| > ||\mathbf{X}||_{[k_n]}\}.$$

This is one possible nonparametric way to make use of the multivariate EV model. The choice of the proportion of data used for the estimation of S is a delicate point in practice. Dealing with any form of extreme event A is of course not so straightforward, and needs care. We refer for example to de Haan & de Ronde (1998), or Bruun & Tawn (1998), for a complete application and evaluation of failure probabilities. Note that, even if from a theoretical point of view the methods based on multivariate EV models which have been developed in the literature are available in a d-dimensional context, the complexity linked to the solution of problems in practice increases rapidly with the dimension d. Most of the work done in the multivariate context concerns examples where $d = 2$ or 3.

An important point is that both parametric and nonparametric threshold estimation techniques present some problems in the particular case where the data are asymptotically independent, i.e. when their distribution is in the domain of attraction of the independence. This limit situation corresponds in the representation [C2] to the case where S is singular, concentrated on some boundary points of $\mathcal{B}$ (see Example 2). Hence parametric methods, as maximum likelihood estimation, face a problem of non regularity, and nonparametric methods also present less satisfying results, as shown via simulations by Capéraà & Fougères (2000) in the bivariate case. Moreover, EV models come from componentwise maxima $(\max_{i=1,\cdots,n} X_{i,1}, \cdots, \max_{i=1,\cdots,n} X_{i,d})$, which in practice typically do not correspond to any observation $\mathbf{X}_k$. In case of asymptotic dependence, the componentwise maxima however tend to occur jointly, so EV models are useful in such a situation. This is not the case anymore in case of asymptotic independence. Asymptotic independence seems however to be an important case in practice, as pointed out by Marshall & Olkin (1983) or de Haan & de Ronde (1998). Indeed, it actually corresponds to most of the classical families of distributions, as listed in Marshall & Olkin (1983) or Capéraà, Fougères & Genest (2000), among others. For example, see the wind and rain data considered by Anderson & Nadarajah (1993) and Ledford & Tawn (1996). Some alternatives and refined models have been proposed in this particular case by Ledford & Tawn (1996, 1997), and will be presented in the following section.

3. An alternative modeling approach.

According to Sibuya (1960), a bivariate pair of r.v. (X_1, X_2) with common marginal d.f. F_1 is said to have *asymptotically independent* components if and only if

$$\lim_{u \to x_{F_1}} P\left(X_1 > u \,|\, X_2 > u\right) = 0,$$

where $x_{F_1} = \sup\{x \in \mathbb{R} : F_1(x) < 1\}$. For the distribution of (X_1, X_2) this property is equivalent to being in the domain of attraction of the independence. This follows from the next result, which also states how multivariate asymptotic independence in general actually reduces to the bivariate case (Berman, 1961; see for example Resnick, 1987, Proposition 5.27).

Theorem 3 *Let $\{\mathbf{X}_n, n \geq 1\}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ $(d \geq 2)$ with d.f. F. Assume for simplicity that all the univariate margins are the same, with common d.f. F_1 belonging to the univariate domain of attraction of G_1. So one has, for some $a_n > 0, b_n \in \mathbb{R}$, the convergence $F_1^n(a_n x + b_n) \to G_1(x)$, as $n \to \infty$. The following assertions are then equivalent:*

(i) *The d.f. F is in the domain of attraction of the independence:*

$$F^n(a_n\mathbf{x} + b_n\mathbf{1}) = P\left(\bigvee_{i=1}^{n} \mathbf{X}_i \leq a_n\mathbf{x} + b_n\mathbf{1}\right) \to \prod_{j=1}^{d} G_1(x_j).$$

(ii) *For all* $1 \le k < \ell \le d$,

$$P\left(\bigvee_{i=1}^{n} X_{i,k} \le a_n x_k + b_n \ , \ \bigvee_{i=1}^{n} X_{i,\ell} \le a_n x_\ell + b_n\right) \to G_1(x_k) G_1(x_\ell).$$

(iii) *For all* $1 \le k < \ell \le d$, *and* x_k, x_ℓ *such that* $G_1(x_k), G_1(x_\ell) > 0$,

$$\lim_{n \to \infty} n\, P\left(X_{1,k} > a_n x_k + b_n \ , \ X_{1,\ell} > a_n x_\ell + b_n\right) = 0.$$

(iv) *For all* $1 \le k < \ell \le d$,

$$\lim_{t \to x_{F_1}} P(\, X_{1,k} > t \mid X_{1,\ell} > t\,) = 0.$$

For simplicity consider the bivariate case of asymptotic independence. Note that in this case, the probability mass of joint tails, that is of sets of the form $\{(X_1 - b_{n,1})/a_{n,1} > x_1 \ , \ (X_2 - b_{n,2})/a_{n,2} > x_2\}$, is of lower order than that for sets like $\{(X_1 - b_{n,1})/a_{n,1} > x_1$ **or** $(X_2 - b_{n,2})/a_{n,2} > x_2\}$. Therefore, models based on bivariate extreme value distributions do not provide any satisfying way to estimate such joint tails.

In order to fill this gap in the bivariate setup, Ledford & Tawn (1996, 1997) proposed joint tail models adapted to the asymptotic independence case. When stated with unit Fréchet margins, these models are essentially based on the model

$$P(Z_1 > z_1\, , \, Z_2 > z_2) \sim \frac{\mathcal{L}(z_1, z_2)}{z_1^{c_1} z_2^{c_2}}, \tag{9}$$

when $z_1, z_2 \to \infty$, where $c_1, c_2 > 0$ are such that $c_1 + c_2 \ge 1$, and $\mathcal{L}$ is a bivariate slowly varying function. Recall that a measurable function $\mathcal{L} : \mathbf{x} \in \mathbb{R}^d \mapsto \mathcal{L}(\mathbf{x}) > 0$ is called a *multivariate slowly varying function* if there exists a positive function λ satisfying $\lambda(t\mathbf{x}) = \lambda(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$, $t > 0$, and such that

$$\lim_{t \to \infty} \frac{\mathcal{L}(t\mathbf{x})}{\mathcal{L}(t\mathbf{1})} \ = \ \lambda(\mathbf{x}),$$

for all $\mathbf{x} \in \mathbb{R}^d$. See for example de Haan (1985b), Basrak, Davis & Mikosch (2000) or Mikosch (2001) for further properties. A measure of extremal dependence is then provided by the so-called *coefficient of tail dependence* $\eta = 1/(c_1+c_2) \in (0,1]$. Asymptotic dependence corresponds to $\eta = 1$ and $\mathcal{L}(r,r) \to 0$ as $r \to \infty$, whereas $\eta < 1$ implies asymptotic independence. Since because of (9) one has

$$P(T > u+t \mid T > u) \sim \frac{\mathcal{L}(u+t,u+t)\,(u+t)^{-1/\eta}}{\mathcal{L}(u,u)\,u^{-1/\eta}} \approx \left(1+\frac{t}{u}\right)^{-1/\eta},$$

Ledford & Tawn (1996) suggest to estimate η as the shape parameter of the GPD for $T = \min(Z_1, Z_2)$. Peng (1999) proposed another estimator of η, for which he obtained the asymptotic normality.

Note that (9) is ensured as soon as a second order condition is assumed for (Z_1, Z_2), which strengthens the multivariate domain of attraction condition (2). More precisely, one assumes the existence of a positive function ψ and a finite and non-zero function h such that

$$\lim_{t\to\infty} \frac{1}{\psi(t)} \left\{t\,P(Z_1 > tz_1 \text{ or } Z_2 > tz_2) + \log G(z_1,z_2)\right\} \;=\; h(z_1,z_2).$$

As mentioned by de Haan & de Ronde (1998), the function ψ is then necessarily a regularly varying function with non positive exponent ρ (that is $\psi(tu)/\psi(t) \to u^\rho$ as $t \to \infty$, for all $u > 0$). The relation between ρ and η is $\rho = 1 - 1/\eta$.

Ledford & Tawn (1997) and Bruun & Tawn (1998) implemented different submodels from (9), specifying further conditions concerning the form of $\mathcal{L}$. For example, Bruun & Tawn modelled the joint tails assuming that

$$P(Z_1 > z_1\,,\, Z_2 > z_2) = \frac{a_0 + \mathcal{L}_*(z_1,z_2,V)}{(z_1 z_2)^{1/2\eta}}, \tag{10}$$

for z_1, z_2 large enough, in terms of $\eta \in (0,1]$, and $a_0 \in \mathbb{R}^+$. The function $\mathcal{L}_*$ is defined by

$$\mathcal{L}_*(z_1,z_2,V) = \left(\frac{z_1}{z_2}\right)^{1/2} + \left(\frac{z_2}{z_1}\right)^{1/2} - (z_1 z_2)^{1/2} V(z_1,z_2),$$

where V is an exponent measure as in Theorem 1, $V(z_1,z_2) = \mu^*\{(\mathbf{0},\mathbf{z}]^c\}$. Bruun & Tawn used more specifically a parametric model for V. They compared, through Monte-Carlo simulations and for different degrees of dependence, the behaviour of three models, namely: the model (10), the bivariate extreme value model with d.f. $F(z_1,z_2) = \exp\{-V(z_1,z_2)\}$, and a model based on univariate extreme value models for the margins. In case of asymptotic independence, the model (10) really outperforms the bivariate extreme value model in terms of relative percentage error when estimating a specific 10^{-3} quantile (see Bruun & Tawn (1998), Section 4, for details). The bivariate extreme value model has actually the tendency to overestimate the probability of failure when asymptotic independence occurs.

4. Measuring extremal dependence

One of the main topics strongly connected with multivariate extreme events modeling is the problem of how to measure the dependence in the extreme observations. A first way to look at this problem is to use multivariate EV models and to measure the strength of the dependence in the limiting distribution. Representation [C2] clearly exhibits that all the dependence is contained in the measure S, but summaries of this information can be useful. Note in passing that the correlation coefficient is useless in the EV context, as on the one hand it is not invariant under transformations of the margins, and on the other hand several cases of univariate EV d.f. may not have a finite variance, as the Fréchet d.f., for example. Focusing on the EV model, several measures have been proposed, mainly in the bivariate case. Consider a pair (X_1, X_2) from an EV distribution, with common marginal d.f. F. Tiago de Oliveira (see de Haan, 1985a) proposed as an index of extremal dependence $\theta \in [1,2]$ defined by

$$P(\max(X_1, X_2) \leq x) = F^{\theta}(x). \tag{11}$$

Using Pickands' representation (7) gives another expression of θ, that is $\theta = 2A(1/2)$. Such a measure has been considered by Tawn (1988) for testing independence in the bivariate EV context (see also Capéraà, Fougères & Genest, 1997).

Alternative measures for bivariate EV distributions are Kendall's τ and Spearman's ρ. They are nonparametric measures of dependence, which have the following closed forms in terms of A (see e.g. Ghoudi, Khoudraji & Rivest, 1998):

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)}\, dA'(t), \quad \rho = 12 \int_0^1 \{A(t)+1\}^{-2}\, dt - 3.$$

The definition (11) can clearly be extended by considering $\theta \in [1, d]$ such that

$$P(\max(X_1, \cdots, X_d) \leq x) = F^{\theta}(x).$$

In a more general context than EV distributions, Buishand (1984) considered, for all r.v. X_1, X_2 with common d.f. F and joint d.f. H, the function θ_H defined by

$$P(\max(X_1, X_2) \leq x) = F^{\theta_H(x)}(x). \tag{12}$$

In the particular case of the EV distribution, such a function is constant, $\theta_H(x) = \theta$.

Another way to introduce the measure θ can be deduced from Joe (1993). Indeed, in the bivariate case he defines the *upper tail dependence parameter* λ for a d.f. H with univariate marginal d.f. F as

$$\lambda = \lim_{x \to x_F} \frac{\bar{H}(x,x)}{1 - F(x)} = \lim_{x \to x_F} P(X_1 > x | X_2 > x), \tag{13}$$

where (X_1, X_2) is a random pair from H, and $\bar{H}$ denotes the survival function defined by $\bar{H}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$. Asymptotic independence is therefore equivalent to $\lambda = 0$. The tail dependence parameter of a distribution is linked to the domain of attraction of this distribution. More precisely, if H is a d.f. belonging to the domain of attraction of an EV d.f. L, and if λ is the tail dependence parameter of H, then λ is also the tail dependence parameter of L (see Joe, 1997, Theorem 6.8). Moreover, as

$$\frac{\bar{H}(x,x)}{1 - F(x)} \sim 2 - \frac{\log H(x,x)}{\log F(x)}$$

in the neighbourhood of x_F, it follows from (12) and (13) that

$$\lambda = 2 - \lim_{x \to x_F} \frac{\log H(x,x)}{\log F(x)} = 2 - \lim_{x \to x_F} \theta_H(x).$$

Consider a bivariate d.f. H, with upper tail dependence parameter λ. Assume that H belongs to the domain of attraction of an EV d.f. L, with univariate margins G, and with index θ in (11). Then Joe's result above yields for θ_L constant:

$$\lambda = 2 - \lim_{x \to x_F} \theta_L(x) = 2 - \theta.$$

One should note that the upper tail dependence parameter just depends on the dependence structure, so it can be expressed in terms of the copula, as did actually Joe (1993). Recall that a *copula* is a multivariate d.f. with all its univariate margins uniformly distributed on $[0,1]$ (Sklar, 1959). Given a multivariate d.f. H with continuous univariate margins $F_1, \cdots, F_d$, there exists a unique function C_H associated to H such that

$$H(\mathbf{x}) = C_H(F_1(x_1), \cdots, F_d(x_d)).$$

This function C_H, called the copula of H, does not depend on the margins of H, and contains therefore all the information relative to the dependence between the different components $X_1, \cdots, X_d$. It is the d.f. of the vector $\mathbf{U} = (U_1, \cdots, U_d) = (F_1(X_1), \cdots, F_d(X_d))$. See for example Kimeldorf & Sampson (1975), Schweizer & Sklar (1983), or Nelsen (1998), for further details. The upper tail dependence parameter defined by (13) is equivalently defined by:

$$\lambda = \lim_{u \to 1} \frac{\bar{C}_H(u,u)}{1-u} = 2 - \lim_{u \to 1} \frac{\log C_H(u,u)}{\log u}. \tag{14}$$

In case of asymptotic independence ($\lambda = 0$), Ledford & Tawn (1996) proposed to measure the tail dependence via the coefficient $\eta = 1/(c_1 + c_2)$ defined from (9). Coles, Heffernan & Tawn (2000) make use of both measures λ and η (denoted χ and $(1+\bar{\chi})/2$ respectively) for diagnostics of asymptotic independence on different sets of data.

5. Conclusion

The aim of this paper was to present a review of the results obtained in the area of multivariate extreme events. The models developed in the literature are essentially based on the limiting behaviour of renormalized componentwise maxima. The structure of the family of limiting distributions is actually quite rich, and can be studied in terms of max-stable distributions, as well as via point process representations. Statistical inference methods deduced from this family of EV distributions have also been reviewed. The limits of the multivariate EV models have been pointed out, especially in the case of asymptotic independence. Alternative models have been summarized. Finally, some measures of extremal dependence have been presented. One should note that this domain of research is currently very active, and promising alternatives have recently been proposed, for example Heffernan & Tawn (2001).

References

Abdous, B. & Ghoudi, K. (2002). Nonparametric estimates of multivariate extreme dependence functions. *Personal communication.*

Abdous, B., Ghoudi, K., & Khoudraji, A. (1999). Nonparametric estimation of the limit dependence function of multivariate extremes. *Extremes*, **2**, 243-265.

Anderson, C. W. & Nadarajah, S. (1993). Environmental factors affecting reservoir safety. In *Statistics for the Environment.* Barnett, V. & Turkman, K. F., Editors, 163-182, Wiley.

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

Basrak, B., Davis, R. A., & Mikosch, T. (2000). A Characterization of Multivariate Regular Variation. *Annals of Applied Probability.*, to appear.

Berman, S. M. (1961). Convergence to bivariate limiting extreme value distributions. *Ann. Inst. Statist. Math.* **13**, 217-223.

Bingham, N. H. , Goldie, C. M., & Teugels, J. L. (1989). *Regular variation.* Cambridge University Press.

Bruun, J. T. & Tawn, J. A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *Appl. Statist.* **47**, 405-423.

Buishand, T. A. (1984). Bivariate extreme-value data and the station-year method. *J. Hydrol.* **69**, 77-95.

Capéraà, P. & Fougères, A.-L. (2000). Estimation of a bivariate extreme value distribution., *Extremes* **3**, 311-329.

Capéraà, P., Fougères, A.-L., & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**, 567-577.

Capéraà, P., Fougères, A.-L., & Genest, C. (2000). Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.* **72**, 30-49.

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

Coles, S. G., Heffernan, J. E., & Tawn, J. A. (2000). Dependence measures for extreme value analyses. *Extremes* **2**, 339-365.

Coles, S. G. & Tawn, J. A. (1991). Modelling multivariate extreme events. *J. R. Statist. Soc.* B **53**, 377-92.

Coles, S. G. & Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *Appl. Statist.* **43**, 1-48.

Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc.* B **52**, 393-442.

Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence aux types extrêmes. *Publ. Inst. Statist. Univ. Paris* **23**, 1-36.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme value distributions. *Statist. Probab. Lett.* **12**, 429-439.

Devroye, L. (1989). The double kernel method in density estimation. *Annales de l'Institut Poincaré* **25**, 533-580.

Drees, H., & Huang, X. (1998). Best attainable rates of convergence for estimators of the stable dependence function. *J. Multivariate Anal.* **64**, 25-47.

Einmahl, J., de Haan, L. & Huang, X. (1993). Estimating a multidimensional extreme value distribution. *J. Multivariate Anal.* **47**, 35-47.

Einmahl, J., de Haan, L. & Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stoch. Proc. Appl.* **70**, 143-171.

EINMAHL, J., DE HAAN, L. & PITERBARG, V. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.*, to appear.
EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.
FISHER, R.A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24**, 180-190.
GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd edition. Kreiger, Malabar, FL.
GENEST, C. and MACKAY, R. J. (1986a). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données, *Canad. J. Statist.* **14**, 145-159.
GENEST, C. and MACKAY, R. J. (1986b). The joy of copulas: Bivariate distributions with uniform marginals, *Amer. Statist.* **40**, 280-283.
GENEST, C. and RIVEST, L.-P. (1989). A characterization of Gumbel's family of extreme value distributions, *Statist. Probab. Lett.* **8**, 207-211.
GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.
GHOUDI, K., KHOUDRAJI, A., and RIVEST, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.* **26** , 187-197.
GUMBEL, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* **9**, 171-173.
DE HAAN, L. (1985a). Extremes in higher dimensions: The model and some statistics. In *Proc. 45th Session Inter. Statist. Inst.*, Amsterdam, paper 26.3. Amsterdam: International Statistical Institute.
DE HAAN, L. (1985b). Multivariate regular variation and applications in probability theory. In *Multivariate analysis, VI*, P. R. Krishnaiah ed., North-Holland, Amsterdam, 281-288.
DE HAAN, L. & RESNICK, S. I. (1977). Limit theory for multidimensional sample extremes. *Z. Wahr. verw. Geb.* **40**, 317-337.
DE HAAN, L. & RESNICK, S. I. (1993). Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stoch. Models* **9**, 275-309.
DE HAAN, L. & DE RONDE, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes* **1**, 7-45.
DE HAAN, L. & SINHA, A. K. (1999). Estimating the probability of a rare event. *Ann. Statist.* **27**, 732-759.
HALL, P. & TAJVIDI, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6**(5), 835-844.
HEFFERNAN, J. E. & TAWN, J. A. (2001). A conditional approach for multivariate extreme values. *Submitted.*
HUANG, X. (1992). *Statistics of bivariate extreme values.* Thesis, Erasmus University Rotterdam, Tinbergen Institute Research Series **22**.

HÜSLER, J. (1986). Extreme values of nonstationary random sequences. *J. Appl. Probab.* **23** (4), 937-950.
JENKINSON, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. Roy. Meteorol. Soc.* **87**, 158-171.
JOE, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* **9**, 75-81.
JOE, H. (1993). Parametric family of multivariate distributions with given margins. *J. Multivariate Anal.* **46**, 262-282.
JOE, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman and Hall, London.
JOE, H., SMITH, R. L. & WEISSMAN, I. (1992). Bivariate threshold models for extremes. *J. R. Statist. Soc.* B **54**, 171-183.
KIMELDORF, G. & SAMPSON, A. R. (1975). Uniform representations of bivariate distributions. *Comm. Statist.* **4**, 617-627.
KOTZ, S. & NADARAJAH, S. (2000). *Extreme value distributions. Theory and applications.* Imperial College Press.
LEADBETTER, M. R., LINDGREN, G. & ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Series*, Springer Verlag, New York.
LEDFORD, A. W. & TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169-187.
LEDFORD, A. W. & TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. R. Statist. Soc.* B **59**, 475-499.
MARSHALL, A. W. & OLKIN, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11**, 168-177.
MIKOSCH, T. (2001). Modeling dependence and tails for financial time series. *Semstat lecture.*
MISES, R., VON (1954). La distribution de la plus grande de n valeurs. *Selected papers*, Vol. II, p. 271-294. Providence, R. I.: Amer. Math. Soc.
PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.* **43**, 399-409.
NELSEN, R. B. (1998). *An introduction to copulas.* Springer Verlag.
PICKANDS, J. (1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Probab.* **8**, 745-756.
PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.
PICKANDS, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst.* 859-878.
RESNICK, S. I. (1986). Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66-138.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* New York: Springer.

SCHWEIZER, B., & SKLAR, A. (1983). *Probabilistic Metric Spaces.* New York: North-Holland.

SIBUYA, M. (1960). Bivariate extreme statistics. *Ann. Inst. Math. Statist.* **11**, 195-210.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* **8**, 229-231.

SMITH, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in Ground-Level Ozone, *Statist. Sc.* **4**, 367-393.

SMITH, R. L., TAWN, J. A. & YUEN, H. K. (1990). Statistics of multivariate extremes. *Inter. Statist. Rev.* **58**, 47-58.

SMITH, R. L., TAWN, J. A. & COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249-268.

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**, 245-253.

TAWN, J. A. (1992). Estimating probabilities of extreme sea-levels. *Appl. Statist.* **41**, 77-93.

TIAGO DE OLIVEIRA, J. (1958). Extremal distributions. *Rev. Fac. Sci. Lisboa, Ser. A* **7**, 215-227.

TIAGO DE OLIVEIRA, J. (1975). Bivariate extremes; extensions. In *Proceedings of the 40th Session of the International Statistical Institute. Bull. Inst. Internat. Statist.* **46**, no. 2, 241-254.

TIAGO DE OLIVEIRA, J. (1980). Bivariate extremes: foundations and statistics. In *Multivariate analysis, V, Proceedings of the 5th International Symposium, Univ. Pittsburgh*, North-Holland, Amsterdam-New York, 349-366.

TIAGO DE OLIVEIRA, J. (1989). Intrinsic estimation of the dependence structure for bivariate extremes. *Statist. Prob. Letters* **8**, 213-8.