ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Special issue on "Collection of Surveys on Tail Event Modeling"

Guest Editors:
Miguel de Carvalho
Anthony C. Davison
Jan Beirlant

Volume 10, No.1
March 2012

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 426 364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *300 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT included]

- Single issue …….. **€ 11**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)……. **€ 26**
- Annual subscription (No. 2, No. 3) …….. **€ 18**

FOREWORD

Modern methods of statistics of extremes have been widely applied in risk modeling, with more and more scientists concerned with extrapolating to the tails of a distribution, often beyond any existing data. Complex estimation and inference problems arise when assessing the probability of such rare events, and ingenious statistical methods have been developed based on various assumptions useful for extrapolation.

This special issue of *Revstat—Statistical Journal* presents an overview of such methods, by discussing recent developments in the statistical modeling of extremes and their applications to the analysis of risk. It covers a variety of topics, from methods for scalar extremes, often applied in a discrete time setting, to the infinite-dimensional setting, thus far mostly applied in the space domain. This special issue aims to provide a broad view of such topics, integrating modern advances with the historical perspective. This allows the authors to go to the origins of the concepts, methods and models of extremes, but in the light of the current state of the art. *A Collection of Surveys on Tail Event Modeling* tries, however, to offer more, as some challenges for future work are pinpointed by the authors. We hope these papers stimulate interaction between experts in the field of extremes and that they are useful for those entering this important field.

We thank Ivette Gomes, editor-in-chief of *Revstat—Statistical Journal*, for encouraging us to take on this challenge. On the behalf of the Editorial Board we would like to thank the authors for contributing to this special issue.

M. B. de Carvalho and A. C. Davison
Ecole Polytechnique Fédérale de Lausanne
Institute of Mathematics
Lausanne
Switzerland

J. Beirlant
University of Leuven
Katholieke Universiteit Leuven
Leuven
Belgium

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 1–31

AN OVERVIEW AND OPEN RESEARCH TOPICS IN STATISTICS OF UNIVARIATE EXTREMES

Authors: Jan Beirlant
– Catholic University of Leuven, Belgium
Jan.Beirlant@wis.kuleuven.be

Frederico Caeiro
– New University of Lisbon, F.C.T. and C.M.A., Portugal
fac@fct.unl.pt

M. Ivette Gomes
– University of Lisbon, F.C.U.L. (D.E.I.O.) and C.E.A.U.L., Portugal
ivette-gomes@fc.ul.pt

Abstract:

- This review paper focuses on statistical issues arising in modeling univariate extremes of a random sample. In the last three decades there has been a shift from the area of parametric statistics of extremes, based on probabilistic asymptotic results in extreme value theory, towards a semi-parametric approach, where the estimation of the right and/or left tail-weight is performed under a quite general framework. But new parametric models can still be of high interest for the analysis of extreme events, if associated with appropriate statistical inference methodologies. After a brief reference to Gumbel's classical block methodology and later improvements in the parametric framework, we present an overview of the developments on the estimation of parameters of extreme events and testing of extreme value conditions under a semi-parametric framework, and discuss a few challenging open research topics.

Key-Words:

- *extreme value index; parameters of extreme events; parametric and semi-parametric estimation and testing; statistics of univariate extremes.*

AMS Subject Classification:

- 62G32, 62E20.

1. INTRODUCTION, LIMITING RESULTS IN THE FIELD OF EXTREMES AND PARAMETRIC APPROACHES

We shall assume that we have a sample $(X_1, \ldots, X_n)$ of n *independent, identically distributed* (IID) or possibly *stationary, weakly dependent* random variables from an underlying *cumulative distribution function* (CDF), F, and shall use the notation $(X_{1,n} \leq \cdots \leq X_{n,n})$ for the sample of associated ascending *order statistics* (OSs). *Statistics of univariate extremes* (SUE) helps us to learn from disastrous or almost disastrous events, of high relevance in society and with a high societal impact. The domains of application of SUE are thus quite diverse. We mention the fields of hydrology, meteorology, geology, insurance, finance, structural engineering, telecommunications and biostatistics (see, for instance, and among others, Coles, 2001; Reiss & Thomas, 2001, 2007; Beirlant *et al.*, 2004, §1.3; Castillo *et al.*, 2005; Resnick, 2007). Although it is possible to find some historical papers with applications related to extreme events, the field dates back to Gumbel, in papers from 1935 on, summarized in his book (Gumbel, 1958). Gumbel develops statistical procedures essentially based on Gnedenko's (Gnedenko, 1943) *extremal types theorem* (ETT), one of the main limiting results in the field of *extreme value theory* (EVT), briefly summarized below.

1.1. Main limiting results in EVT

The main limiting results in EVT date back to the papers by Fréchet (1927), Fisher & Tippett (1928), von Mises (1936) and Gnedenko (1943). Gnedenko's ETT provides the possible limiting behaviour of the sequence of maximum or minimum values, linearly normalised, and an incomplete characterization, fully achieved in de Haan (1970), of the domains of attraction of the so-called *max-stable* (MS) or *min-stable* laws. Here, we shall always deal with the right-tail, $\overline{F}(x) := 1 - F(x)$, for large x, i.e., we shall deal with top OSs. But all results for maxima (top OSs) can be easily reformulated for minima (low OSs). Indeed, $X_{1,n} = -\max_{1 \leq i \leq n}(-X_i)$, and consequently, $\mathbb{P}(X_{1,n} \leq x) = 1 - \{1 - F(x)\}^n$. MS laws are defined as laws S such that the functional equation $S^n(\alpha_n x + \beta_n) = S(x)$, $n \geq 1$, holds for some $\alpha_n > 0$, $\beta_n \in \mathbb{R}$. More specifically, all possible non-degenerate weak limit distributions of the normalized partial maxima $X_{n,n}$, of IID random variables $X_1, \ldots, X_n$, are (generalized) *extreme value distributions* (EVDs), i.e., if there are normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$, and some non-degenerate CDF G such that, for all x,

$$\lim_{n \to \infty} \mathbb{P}\big\{(X_{n,n} - b_n)/a_n \leq x\big\} = G(x) \ , \tag{1.1}$$

we can redefine the constants in such a way that

$$G(x) \equiv G_\gamma(x) := \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & 1+\gamma x>0, \quad \text{if } \gamma \neq 0\,, \\ \exp\{-\exp(-x)\}, & x \in \mathbb{R}, \qquad\quad \text{if } \gamma = 0\,, \end{cases} \tag{1.2}$$

given here in the von Mises–Jenkinson form (von Mises, 1936; Jenkinson, 1955). If (1.1) holds, we then say that the CDF F which is underlying $X_1, X_2, \ldots$, is in the *max-domain of attraction* (MDA) of G_γ, in (1.2), and often use the notation $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$. The limiting CDFs G in (1.1) are then MS. They are indeed the unique MS laws. The real parameter γ, the primary parameter of interest in extreme value analysis (EVA), is called the *extreme value index* (EVI). The EVI, γ, governs the behaviour of the right-tail of F. The EVD, in (1.2), is often separated into the three types:

$$\begin{aligned} &\text{Type I (Gumbel)}: && \Lambda(x) = \exp\{-\exp(-x)\}, && x \in \mathbb{R}\,, \\ &\text{Type II (Fréchet)}: && \Phi_\alpha(x) = \exp(-x^{-\alpha}), && x \geq 0\,, \\ &\text{Type III (max-Weibull)}: && \Psi_\alpha(x) = \exp\{-(-x)^{\alpha}\} && x \leq 0\,. \end{aligned} \tag{1.3}$$

Indeed, with $\gamma = 0$, $\gamma = 1/\alpha > 0$ and $\gamma = -1/\alpha < 0$, respectively, we have $\Lambda(x) = G_0(x)$, $\Phi_\alpha(x) = G_{1/\alpha}\{\alpha(1-x)\}$ and $\Psi_\alpha(x) = G_{-1/\alpha}\{\alpha(x+1)\}$, with G_γ the EVD in (1.2). The Fréchet domain of attraction ($\gamma > 0$) contains heavy-tailed CDFs like the Pareto and the Student t-distributions, i.e., tails of a negative polynomial type and infinite right endpoint. Short-tailed CDFs, with finite right endpoint, like the beta CDFs, belong to the Weibull MDA ($\gamma < 0$). The Gumbel MDA ($\gamma = 0$), is relevant for many applied sciences, and contains a great variety of CDFs with an exponential tail, like the normal, the exponential and the gamma, but not necessarily with an infinite right endpoint. As an example of a CDF $F \in \mathcal{D}_{\mathcal{M}}(G_0)$, with a finite right endpoint x^F, we have the *exponential-type distribution*, $F(x) = K \exp\{-c/(x^F - x)\}$, for $x < x^F$, $c > 0$ and $K > 0$.

Apart from the ETT and the already mentioned EVD, in (1.2), it is also worth mentioning the *generalized Pareto distribution* (GPD), the limit distribution of scaled excesses over high thresholds (see the pioneering papers by Balkema & de Haan, 1974; Pickands, 1975), which can be written as

$$P_\gamma(x) = 1 + \ln G_\gamma(x) = \begin{cases} 1-(1+\gamma x)^{-1/\gamma}, & 1+\gamma x>0,\ x>0, \quad \text{if } \gamma \neq 0\,, \\ 1-\exp(-x), & x>0, \qquad\quad \text{if } \gamma = 0\,, \end{cases} \tag{1.4}$$

with G_γ given in (1.2), as well as the *multivariate* EVD, related with the limiting distribution of the k largest values $X_{n-i+1:n}$, $1 \leq i \leq k$, also called the *extremal process* (Dwass, 1964), with associated probability density function (PDF)

$$h_\gamma(x_1, \ldots, x_k) = g_\gamma(x_k) \prod_{j=1}^{k-1} \frac{g_\gamma(x_j)}{G_\gamma(x_j)} \qquad \text{if} \quad x_1 > \cdots > x_k\,, \tag{1.5}$$

where $g_\gamma(x) = dG_\gamma(x)/dx$, with $G_\gamma(x)$ given in (1.2).

1.2. Parametric approaches to SUE

Deciding upon the right tail-weight for the distribution underlying the sample data constitutes an important initial task in EVA. On the other hand, statistical inference about rare events is clearly linked to observations which are extreme in some sense. There are different ways to define such observations, leading to different approaches to SUE. We next briefly reference the most common parametric approaches to SUE. For further details on the topic, and pioneering papers on the subject, see Gomes *et al.* (2008a).

Block maxima (BM) method. With $(\lambda_n, \delta_n) \in \mathbb{R} \times \mathbb{R}^+$, a vector of unknown location and scale parameters, the ETT supports the approximation

$$P\big(X_{n,n} \le x\big) = F^n(x) \approx G_\gamma\big\{(x-\lambda_n)/\delta_n\big\} \,. \tag{1.6}$$

Gumbel was pioneer in the use of approximations of the type of the one provided in (1.6), but for any of the models in (1.3), suggesting the first model in SUE, usually called the BM model or the *annual maxima* model or the *extreme value* (EV) *univariate model* or merely *Gumbel's model*. The sample of size n is divided into k sub-samples of size r (usually associated to k years, with $n = r \times k$, r reasonably large). Next, the maximum of the r observations in each of the k sub-samples is considered, and one of the extremal models in (1.3), obviously with extra unknown location and scale parameters, is fitted to such a sample. Nowadays, whenever using this approach, still quite popular in environmental sciences, it is more common to fit to the data a univariate EVD, $G_\gamma\{(x-\lambda_r)/\delta_r\}$, with G_γ given in (1.2), $(\lambda_r, \delta_r, \gamma) \in (\mathbb{R}, \mathbb{R}^+, \mathbb{R})$ unknown location, scale and 'shape' parameters. All statistical inference is then related to EVDs.

The method of largest observations (LO). Although the BM-method has proved to be fruitful in the most diversified situations, several criticisms have been made on Gumbel's technique, and one of them is the fact that we are wasting information when using only observed maxima and not further top OSs, if available, because they surely contain useful information about the right-tail of the CDF underlying the data. To make inference on the right-tail weight of the underlying model, it seems sensible to consider a small number k of top OSs from the original data, and when the sample size n is large and k fixed, it is sensible to consider the *multivariate* EVD, with a standardized PDF given in (1.5). Again, unknown location and scale parameters, λ_n and δ_n, respectively, are considered and estimated on the basis of the k top OSs, out of n. This approach to SUE is the so-called LO method or *multivariate* EV *model*. It is now easier to increase the number k of observations, contrary to what happens in Gumbel's approach.

Multi-dimensional EV approaches. It is obviously feasible to combine the two aforementioned approaches to SUE. In each of the sub-samples asso-

ciated to Gumbel's classical approach, we can collect a few top OSs modelled through a *multivariate* EV *model*, and then consider the so-called *multidimensional* EV *model*. Under this approach, we have access to the multivariate sample, $(\underline{X}_1, \ldots, \underline{X}_k)$, where $\underline{X}_j = (X_{1j}, \ldots, X_{i_j j})$, $1 \le j \le k$, are *multivariate* EV *vectors*. The *multi-dimensional* EV *model* is indeed the multivariate EV model for the i_j top observations, $j = 1, \ldots, m$, in sub-samples of size m', with $m \times m' = n$. The choices $m = k$ $(m' = r)$ and $i_j = 1$ for $1 \le j \le k$ give the BM model. The choices $m' = n$ $(m = 1)$ and $i_1 = k$ give the LO model.

The peaks over threshold (POT) approaches. The *Paretian model* for the excesses, $X_j - u > 0$, $1 \le j \le k$, over a high threshold u, suitably chosen, is considered under this approach, in a certain sense parallel to the *multivariate* EV *model*, but where we restrict our attention only to observations that exceed a certain high *threshold* u, fitting the appropriate statistical model to the *excesses* over u. On the basis of the approximation $P\big(X - u \le x \mid X > u\big) \approx P_\gamma(x/\sigma)$, with $P_\gamma(x)$ given in (1.4), we come to the so-called *Paretian excesses model* or POT *model*. Statistical inference is then related to the GPD.

Bayesian approaches. The use of Bayesian methodology, within EVA, has recently become quite common. We mention only some recent papers, written after the monographs by Coles (2001) and Reiss & Thomas (2001), the ones by Bermudez & Amaral-Turkman (2003), Bottolo *et al.* (2003), Stephenson & Tawn (2004), on the use of reversible jump MCMC techniques for inference for the EVD and the GPD and Diebolt *et al.* (2005), on a quasi-conjugate Bayesian inference approach for the GPD with $\gamma > 0$, through the representation of a heavy-tailed GPD as a mixture of an exponential and a gamma distribution.

Statistical choice of EV models under parametric frameworks. The Gumbel type CDF, $\Lambda \equiv G_0$ or the exponential (E) type CDF, $E \equiv P_0$, with G_γ and P_γ given in (1.2) and (1.4), respectively, are favorites in SUE, essentially because of the simplicity of associated inference. Additionally, $\gamma = 0$ can be regarded as a change-point, and any separation between EV models, with Λ or E in a central position, turns out to be an important statistical problem. From a parametric point of view, empirical tests of $H_0 : \gamma = 0$ versus a sensible one-sided or two-sided alternative, either for the EVD or the GPD, date back to Jenkinson (1955) and Gumbel (1965). Next, we can find in the literature, different heuristic tests, among which we reference only one of the most recent (Brilhante, 2004). We can also find locally asymptotically normal tests (see Marohn, 2000, and Falk *et al.*, 2008, among others). The fitting of the GPD to data has been worked out in Castillo & Hadi (1997) and Chaouche & Bacro (2004). The problem of goodness-of-fit tests for the GPD has been studied by Choulakian & Stephens (2001) and Luceño (2006), again among others. Tests from large sample theory, like the likelihood ratio test have been dealt with by Hosking (1984) and Gomes (1989). Further details on this topic can be found in Gomes *et al.* (2007a), an enlarged version of Gomes *et al.* (2008a).

1.3. Scope of the paper

In the late 1970s, there was a move from a *parametric approach* based on the limiting models in EVT, towards a *semi-parametric approach*, where tail estimation is done under a quite general framework. In § 2 of this review paper we cover classical semi-parametric inference. Recently, essentially for heavy tails, i.e., for $\gamma > 0$, but also for a general $\gamma \in \mathbb{R}$, the accommodation of bias of the classical estimators of parameters of extreme events has been deeply considered in the literature. The topic of second-order reduced-bias (SORB) estimation still seems to open interesting perspectives in the field, and will be addressed in § 3. Finally, in § 4, we shall discuss some still challenging topics in SUE, providing some overall comments on the subject.

2. CLASSICAL SEMI-PARAMETRIC INFERENCE

Under these semi-parametric approaches, we work with the k top OSs associated to the n available observations or with the excesses over a high random threshold, assuming only that, for a certain γ, the model F underlying the data is in $\mathcal{D}_{\mathcal{M}}(G_\gamma)$ or in specific sub-domains of $\mathcal{D}_{\mathcal{M}}(G_\gamma)$, with G_γ provided in (1.2), γ being the key parameter of extreme events to be estimated, using a few large observations, and with suitable methodology. There is thus no fitting of a specific parametric model, dependent upon a location λ, a scale δ and a shape γ. We usually need to base the EVI-estimation on the k top OSs in the sample, with k intermediate, i.e., such that $k = k_n \to \infty$ and $k = o(n)$, i.e., $k/n \to 0$, as $n \to \infty$. Such estimators, together with semi-parametric estimators of location and scale (see, for instance, de Haan & Ferreira, 2006), can next be used to estimate extreme quantiles, return periods of high levels, upper tail probabilities and other parameters of extreme events. After introducing first and second-order conditions in § 2.1, § 2.2 describes several classical semi-parametric EVI-estimators. In § 2.3, we give results on the testing of the EV condition $F \in \mathcal{D}(G_\gamma)$, under a semi-parametric framework. Finally, in § 2.4, we outline the semi-parametric estimation of other parameters of extreme events.

2.1. First, second (and higher) order conditions

As mentioned above, in § 1, the full characterization of $\mathcal{D}_{\mathcal{M}}(G_\gamma)$ has been given in de Haan (1970), and can be also found in Falk *et al.* (2004) and de Haan & Ferreira (2006). Indeed, with U standing for a (reciprocal) quantile type function

associated with F and defined by $U(t) := \big(1/(1-F)\big)^{\leftarrow}(t) = F^{\leftarrow}(1-1/t) = \inf\big\{x\colon F(x) \geq 1-1/t\big\}$, the *extended regular variation* property,

$$F \in \mathcal{D}_{\mathcal{M}}\left(G_\gamma\right) \iff \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \begin{cases} \dfrac{x^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0\,, \\ \ln x & \text{if } \gamma = 0\,, \end{cases} \tag{2.1}$$

for every $x > 0$ and some positive measurable function a, is a well-known necessary and sufficient condition for $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$ (de Haan, 1984). Heavy-tailed models, i.e., models $F \in \mathcal{D}_{\mathcal{M}}^{+} := \mathcal{D}_{\mathcal{M}}(G_{\gamma>0})$, are quite important in many areas. We can then choose $a(t) = \gamma\, U(t)$ in (2.1), and $F \in \mathcal{D}_{\mathcal{M}}^{+}$ if and only if, for every $x > 0$, $\lim_{t\to\infty} U(tx)/U(t) = x^\gamma$, i.e., U is of regular variation with index γ, denoted $U \in RV_\gamma$. More generally, $F \in \mathcal{D}_{\mathcal{M}}^{+} \iff \overline{F} := 1-F \in RV_{-1/\gamma} \iff U \in RV_\gamma$. For full details on regular variation see Bingham *et al.* (1987).

Under a semi-parametric framework, apart from the first-order condition in (2.1), we often need to assume a second-order condition, specifying the rate of convergence in (2.1). It is then common to assume the existence of a function A^*, possibly not changing in sign and tending to zero as $t \to \infty$, such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{A^*(t)} = \frac{1}{\rho^*}\left(\frac{x^{\gamma+\rho^*}-1}{\gamma+\rho^*} - \frac{x^\gamma-1}{\gamma}\right), \qquad x > 0\,, \tag{2.2}$$

where $\rho^* \leq 0$ is a *second-order* parameter controlling the speed of convergence of maximum values, linearly normalized, towards the limit law in (1.2). Then $\lim_{t\to\infty} A^*(tx)/A^*(t) = x^{\rho^*}$, $x > 0$, i.e., $|A^*| \in RV_{\rho^*}$ (de Haan & Stadtmüller, 1996). For heavy tails, the second-order condition is usually written as

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}\,, \tag{2.3}$$

where $\rho \leq 0$ and $A(t) \to 0$ as $t \to \infty$. More precisely, $|A| \in RV_\rho$ according to Geluk & de Haan (1987). For the link between $(A^*(t), \rho^*)$ and $(A(t), \rho)$, see de Haan & Ferreira (2006) and Fraga Alves *et al.* (2007). Similarly, third-order conditions specify the rate of convergence either in (2.2) or in (2.3). For further details on the third-order condition for heavy tails, see Gomes *et al.* (2002a) and Fraga Alves *et al.* (2003a). For a general third-order framework, see Fraga Alves *et al.* (2003b, Appendix; 2006). Higher-order conditions can be similarly postulated, but restrict the chosen CDFs in $\mathcal{D}_{\mathcal{M}}(G_\gamma)$ more strictly.

2.2. Classical semi-parametric EVI-estimation

The most basic EVI-estimators that have motivated several other refined estimators, i.e., the *Hill* (H), *Pickands* (P), *moment* (M) and *peaks over random threshold-maximum likelihood* (PORT-ML) estimators, are described in § 2.2.1. Next, in § 2.2.2, we briefly discuss other classical EVI-estimators.

2.2.1. H, P, M and PORT-ML EVI-estimators

The H-estimator. For heavy tailed models, i.e., in $\mathcal{D}_{\mathcal{M}}^{+}$, a simple EVI-estimator has been proposed in Hill (1975). The H-estimator, denoted $\hat{\gamma}_{n,k}^{\mathrm{H}}$, is the average of the scaled log-spacings as well as of the log-excesses, given by

$$U_i := i\left(\ln \frac{X_{n-i+1,n}}{X_{n-i,n}}\right) \quad \text{and} \quad V_{ik} := \ln \frac{X_{n-i+1,n}}{X_{n-k,n}}, \quad 1 \le i \le k < n \ , \tag{2.4}$$

respectively. Its asymptotic properties have been thoroughly studied (see de Haan & Peng, 1998, and the review in Gomes *et al.*, 2008a).

The P-estimator. For a general EVI, $\gamma \in \mathbb{R}$, and considering as the basis of the estimation the k top OSs, we can write the P-estimator (Pickands, 1975) as

$$\hat{\gamma}_{n,k}^{\mathrm{P}} := \ln \left\{ \left(X_{n-[k/4]+1,n} - X_{n-[k/2]+1,n}\right) \Big/ \left(X_{n-[k/2]+1,n} - X_{n-k+1,n}\right) \right\} / \ln 2 \ ,$$

where $[x]$ denotes the integer part of x. Asymptotic properties of this estimator are provided in Dekkers & de Haan (1989).

The M-estimator. Dekkers *et al.* (1989) proposed the M-estimator, based on

$$M_{n,k}^{(j)} := \frac{1}{k} \sum_{i=1}^{k} \left(\ln X_{n-i+1,n} - \ln X_{n-k,n}\right)^j, \qquad j > 0 \ , \tag{2.5}$$

the j-moment of the log-excesses, $M_{n,k}^{(1)} \equiv \hat{\gamma}_{n,k}^{\mathrm{H}}$ being the H-estimator. The M-estimator is given by $\hat{\gamma}_{n,k}^{\mathrm{M}} := M_{n,k}^{(1)} + \frac{1}{2}\Big(1 - \big(M_{n,k}^{(2)}/[M_{n,k}^{(1)}]^2 - 1\big)^{-1}\Big)$.

The PORT-ML-estimator. Conditionally on $X_{n-k,n}$, with k intermediate, $D_{ik} := X_{n-i+1,n} - X_{n-k,n}$, $1 \le i \le k$, are approximately the k top OSs associated to a sample of size k from $GP_\gamma(\alpha x/\gamma)$, $\gamma, \alpha \in \mathbb{R}$, with $GP_\gamma(x)$ given in (1.4). The solution of the maximum-likelihood (ML) equations associated to the above mentioned set-up (Davison, 1984) gives rise to an explicit EVI-estimator, the PORT-ML EVI-estimator, named PORT after Araújo Santos *et al.* (2006), and given by $\hat{\gamma}_{n,k}^{\mathrm{PORT-ML}} := \frac{1}{k}\sum_{i=1}^{k} \ln(1 + \hat{\alpha}\, D_{ik})$, where $\hat{\alpha}$ is the implicit ML estimator of the unknown 'scale' parameter α. A comprehensive study of the asymptotic properties of this ML estimator has been undertaken in Drees *et al.* (2004). As recently shown by Zhou (2009, 2010), this estimator is valid for $\gamma > -1$.

2.2.2. Other 'classical' semi-parametric EVI-estimators

Kernel ($\mathcal{K}$) and QQ-estimators. A general class of estimators for a positive EVI are the $\mathcal{K}$-estimators proposed by Csörgő *et al.* (1985), given by $\hat{\gamma}^{\mathcal{K}}_{n,k} := \sum_{i=1}^{n} \mathcal{K}(i/k)\,(\ln X_{n-i+1,n} - \ln X_{n-k,n}) / \sum_{i=1}^{n} \mathcal{K}(i/k)$, where $\mathcal{K}(\cdot)$ is some non-negative, non-increasing kernel defined on $(0,\infty)$ and with unit integral. As an example, the H-estimator is a kernel estimator associated to the kernel $\mathcal{K}(t) = I_{]0,1]}(t)$, where $I_A(t)$ denotes the indicator function ($I_A(t) = 1$ if $t \in A$, and equal to 0 otherwise). Kernel estimators for a real EVI are considered in Groeneboom *et al.* (2003). The H-estimator can also be obtained from the Pareto QQ-plot, through the use of a naïve estimator of the slope in the ultimate right-end of the QQ-plot. More flexible regression methods can be applied to the highest k points of the Pareto QQ-plot; see Beirlant *et al.* (1996a,c), Schultze & Steinbach (1996), Kratz & Resnick (1996), Csörgő & Viharos (1998) and Oliveira *et al.* (2006). They are all $\mathcal{K}$-estimators.

Generalized P-estimators. The large asymptotic variance of the P-estimator has motivated different generalizations of the type $\hat{\gamma}^{\mathrm{P}(\theta)}_{n,k} := -\ln\left\{\left(X_{n-[\theta^2 k]+1,n} - X_{n-[\theta k]+1,n}\right) / \left(X_{n-[\theta k]+1,n} - X_{n-k,n}\right)\right\} / \ln\theta, \quad 0 < \theta < 1$. (Fraga Alves, 1992, 1995; Themido Pereira, 1993; Yun, 2002). Drees (1995) establishes the asymptotic normality of linear combinations of P-estimators, obtaining optimal weights that can be adaptively estimated from the data. Related work appears in Falk (1994). In Segers (2005), the P-estimator is generalized in a way that includes all of its previously known variants.

The generalized Hill (GH) estimator. The slope of a generalized quantile plot led Beirlant *et al.* (1996b) to the GH-estimator, valid for all $\gamma \in \mathbb{R}$, with the functional form, $\hat{\gamma}^{\mathrm{GH}}_{n,k} = \hat{\gamma}^{\mathrm{H}}_{n,k} + \frac{1}{k}\sum_{i=1}^{k}\left(\ln \hat{\gamma}^{\mathrm{H}}_{n,i} - \ln \hat{\gamma}^{\mathrm{H}}_{n,k}\right)$. Further study of this estimator has been performed in Beirlant *et al.* (2005).

The Mixed Moment (MM) estimator. Fraga Alves *et al.* (2009) introduced the so-called MM-estimator, involving not only the log-excesses but also another type of moment-statistics given by $\hat{\varphi}_{n,k} := \left(M^{(1)}_{n,k} - L^{(1)}_{n,k}\right) / \left(L^{(1)}_{n,k}\right)^2$, with $L^{(1)}_{n,k} := \frac{1}{k}\sum_{i=1}^{k}\left(1 - X_{n-k,n}/X_{n-i+1,n}\right)$, and where $M^{(1)}_{n,k}$ is defined in (2.5). The statistic $\hat{\varphi}_{n,k}$ can easily be transformed into what has been called the MM-estimator, valid for any $\gamma \in \mathbb{R}$, and given by $\hat{\gamma}^{\mathrm{MM}}_{n,k} := \left\{\hat{\varphi}_n(k) - 1\right\} / \left[1 + 2\min\{\hat{\varphi}_n(k) - 1, 0\}\right]$. This seems a promising alternative to the most popular EVI-estimators for $\gamma \in \mathbb{R}$.

Semi-parametric probability weighted moment (PWM) estimators. The PWM method is a generalization of the *method of moments*, introduced in Greenwood *et al.* (1979). For $\gamma < 1$ and for CDFs like the EVD, $EV_\gamma((x-\lambda)/\delta)$, with $EV_\gamma(x)$ given in (1.2), the *Pareto* d.f., $P_\gamma(x;\delta) = 1 - (x/\delta)^{-1/\gamma}$, $x > \delta$, and the GPD, $GP_\gamma(x/\delta)$, with $GP_\gamma(x)$ defined in (1.4), the PWM have simple expressions, which allow a simple parametric estimation of the EVI (see Hosking *et al.*, 1985; Hosking & Wallis, 1987; Diebolt *et al.*, 2007, 2008c).

On the basis of the GPD, de Haan and Ferreira (2006) considered, for $\gamma < 1$, the semi-parametric GPPWM EVI-estimator, with GPPWM standing for *generalized Pareto PWM*, given by $\hat{\gamma}_{n,k}^{\text{GPPWM}} := 1 - 2\,\hat{a}_1^\star(k)/\big(\hat{a}_0^\star(k) - 2\,\hat{a}_1^\star(k)\big)$, $1 \le k < n$, and $\hat{a}_s^\star(k) := \sum_{i=1}^k \big(\frac{i}{k}\big)^s \big(X_{n-i+1:n} - X_{n-k:n}\big)/k$, $s = 0, 1$. On the basis of the Pareto model, Caeiro & Gomes (2011) introduced the PPWM EVI-estimators, with PPWM standing for *Pareto PWM*, given by $\hat{\gamma}_{n,k}^{\text{PPWM}} := 1 - \hat{a}_1(k)/\big\{\hat{a}_0(k) - \hat{a}_1(k)\big\}$, where $\hat{a}_s(k) := \frac{1}{k+1} \sum_{i=1}^{k+1} \big(\frac{i}{k+1}\big)^s X_{n-i+1:n}$, $s = 0, 1$ with $1 \le k < n$.

Other estimators. Falk (1995a) proposed the location-invariant estimator, $\hat{\gamma}_{n,k} := \frac{1}{k} \sum_{i=1}^{k-1} \ln\big(X_{n,n} - X_{n-i,n}\big) \,/\, \big(X_{n,n} - X_{n-k,n}\big)$, as a complement of the PORT-ML estimator for $\gamma < -1/2$. Such an estimator has been improved, on the basis of an iterative procedure, in Hüsler & Müller (2005). The non-invariance for shifts of the H-estimator led Fraga Alves (2001) to the consideration for $k > k_0$, with k_0 appropriately chosen, of the location invariant Hill-type estimator $\hat{\gamma}_{n,k,k_0} := \frac{1}{k_0} \sum_{i=1}^{k} \ln\big(\big(X_{n-i+1,n} - X_{n-k,n}\big)/\big(X_{n-k_0+1,n} - X_{n-k,n}\big)\big)$. Beirlant *et al.* (1996b) consider a general class of estimators based on the mean, median and trimmed excess functions. Drees (1998) obtains asymptotic results for a general class of EVI-estimators, arbitrary smooth functionals of the empirical tail quantile function $Q_n(t) = X_{n-[k_n t],n}$, $t \in [0, 1]$. Such a class includes H, P and $\mathcal{K}$-estimators, among others. For further references see, e.g., §6.4 of Embrechts *et al.* (1997), Beirlant *et al.* (1996a;1998), Csörgő & Viharos (1998), §3 of de Haan & Ferreira (2006), and Ling *et al.* (2011).

2.2.3. Consistency and asymptotic normal behaviour of the estimators

Weak consistency of any of the aforementioned EVI-estimators is achieved in the sub-domain of $\mathcal{D}_{\mathcal{M}}(EV_\gamma)$ where they are valid, whenever (2.1) holds and k is intermediate. Under the validity of the second-order condition in (2.2), it is possible to guarantee their asymptotic normality. More precisely, with T denoting any of these EVI-estimators, and with $B(t)$ a bias function converging to zero as $t \to \infty$ and closely related with the $A^*(t)$ function in (2.3), it is possible to guarantee the existence of $\mathcal{C}_{_T} \subset \mathbb{R}$ and $(b_{_T}, \sigma_{_T}) \in \mathbb{R} \times \mathbb{R}^+$, such that

$$\hat{\gamma}_{n,k}^T \overset{d}{=} \gamma + \sigma_{_T} P_k^T/\sqrt{k} + b_{_T} B(n/k) + o_p\big\{B(n/k)\big\}\,, \tag{2.6}$$

with P_k^T an asymptotically standard normal random variable. Consequently, for values k such that $\sqrt{k}\; B(n/k) \to \lambda$, finite, as $n \to \infty$,

$$\sqrt{k}\,\big(\hat{\gamma}_{n,k}^T - \gamma\big) \xrightarrow[n\to\infty]{d} \text{Normal}(\lambda b_{_T}, \sigma_{_T}^2)\,.$$

The values $b_{_T}$ and $\sigma_{_T}^2$ are usually called the *asymptotic bias* and *asymptotic variance* of $\hat{\gamma}_{n,k}^T$ respectively. Details on the values of $(b_{_T}, \sigma_{_T})$ and the function B, in (2.6) are given in the aforementioned papers associated with the T-estimators.

2.3. Testing under a semi-parametric framework

Testing the hypothesis $H_0\colon F \in \mathcal{D}_{\mathcal{M}}(G_0)$ against $H_1\colon F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, $\gamma \neq 0$, or the corresponding one-sided alternatives, under a semi-parametric framework is obviously natural and sensible. In a broad sense, tests of this nature can already be found in papers prior to 2000 (see Gomes *et al.*, 2007a). Non-parametric tests appear in Jurečková & Picek (2001). But the testing of extreme value conditions can be dated back to Dietrich *et al.* (2002), who propose a test statistic to test whether the hypothesis $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$ is supported by the data, together with a simpler version devised to test whether $F \in \mathcal{D}_{\mathcal{M}}(G_{\gamma\geq 0})$. Further results of this last nature can be found in Drees *et al.* (2006) for testing $F \in \mathcal{D}_{\mathcal{M}}(G_{\gamma>-1/2})$. Tables of associated critical points are provided in Hüsler & Li (2006). Beirlant *et al.* (2006) tackle the goodness-of-fit problem for the class of heavy-tailed or Pareto-type distributions. For overviews of the subject see Hüsler & Peng (2008) and Neves & Fraga Alves (2008). See also Koning & Peng (2008) and Goegebeur & Guillou (2010).

2.4. Estimation of other parameters of extreme events

High quantiles of probability $1-p$, p small, or equivalently in financial frameworks the Value at Risk at a level p (VaR_p) are possibly the most important parameters of extreme events, functions of the EVI, as well as of location/scale parameters. In a semi-parametric context, the most usual estimators of a quantile $\chi_{1-p} := U(1/p)$, with p small, can be easily derived from (2.1), through the approximation $U(tx) \approx U(t) + a(t)\,(x^\gamma - 1)/\gamma$. The fact that $X_{n-k+1,n} \overset{p}{\sim} U(n/k)$ enables us to estimate χ_{1-p} on the basis of this approximation and appropriate estimates of γ and $a(n/k)$. For the simpler case of heavy tails, the approximation is $U(tx) \approx U(t)x^\gamma$, and we get $\hat{\chi}_{1-p,k} := X_{n-k:n}\big\{k/(np)\big\}^{\hat{\gamma}_k}$, where $\hat{\gamma}_k$ is any consistent semi-parametric EVI-estimator. This estimator is of the type introduced by Weissman (1978). Details on semi-parametric estimation of extremely high quantiles for $\gamma \in \mathbb{R}$, can be found in Dekkers & de Haan (1989), de Haan & Rootzén (1993) and more recently in Ferreira *et al.* (2003). Fraga Alves *et al.* (2009) also provide, jointly with the MM-estimator, accompanying shift and scale estimators that make high quantile estimation almost straightforward. Other approaches to high quantile estimation can be found in Matthys & Beirlant (2003). None of the above mentioned quantile estimators is equivariant. Araújo Santos *et al.* (2006) provide a class of semi-parametric VaR_p estimators which enjoy equivariance, the empirical counterpart of the theoretical linearity of a quantile χ_p, $\chi_p(\delta X + \lambda) = \delta\,\chi_p(X) + \lambda$, for any real λ and positive δ. This class of estimators is based on the PORT methodology, providing exact properties for risk

measures in finance: translation-equivariance and positive homogeneity. The estimation of the probability of exceedance of a fixed high level, has been dealt with by Dijk & de Haan (1992) and Ferreira (2002), among others. See also Guillou *et al.* (2010) and You *et al.* (2010). The estimation of the endpoint of an underlying CDF has been studied by Hall (1982), Csörgő & Mason (1989), and Aarssen & de Haan (1994), among others. Estimation of the mean of a heavy-tailed distribution has been undertaken by Peng (2001) and Johansson (2003). Estimation of the Weibull tail coefficient dates back to Girard (2004). See also Goegebeur *et al.* (2010a), among others. See also de Haan & Ferreira (2006).

3. SORB ESTIMATION

Most of the classical semi-parametric estimators of any parameter of extreme events have a strong bias for moderate up to large values of k, including the optimal k, in the sense of minimal *mean squared error* (MSE). Accommodation of bias of classical estimators of parameters of extreme events has been deeply considered in the recent literature. We mention the pioneering papers of Peng (1998), Beirlant *et al.* (1999), Feuerverger & Hall (1999) and Gomes *et al.* (2000), where the classical bias-variance trade-off always appears. Such a trade-off was removed with an appropriate estimation of the second-order parameters, as done in Caeiro *et al.* (2005) and Gomes *et al.* (2007b; 2008c), who introduced different types of minimum-variance reduced-bias (MVRB) EVI-estimators. Such estimators have an asymptotic variance equal to that of the Hill EVI-estimator but an asymptotic bias of smaller order, and thus beat the classical estimators for all k. In § 3.1 we deal with SORB semiparametric EVI-estimation and in § 3.2, we briefly describe the recent literature on SORB semi-parametric estimation of other parameters of extreme events.

3.1. SORB semi-parametric EVI-estimation

Let us consider any 'classical' semi-parametric EVI-estimator, $\hat{\gamma}_{n,k}$. Let us also assume that a distributional representation similar to the one in (2.6), with $(b_{_T}, \sigma_{_T})$ replaced by (b, σ), holds for $\hat{\gamma}_{n,k}$. For intermediate k, $\hat{\gamma}_{n,k}$ is consistent for EVI-estimation, and it is asymptotically normal if we further assume that $\sqrt{k}B(n/k) \to \lambda$, finite. Approximations for the variance and the squared-bias of $\hat{\gamma}_{n,k}$ are then σ^2/k and $b^2B^2(n/k)$ respectively. Consequently, these estimators exhibit the same peculiarities: a high variance for high thresholds $X_{n-k,n}$, i.e., for small k; a high bias for low thresholds, i.e., for large k; a small region of stability of the sample path (plot of the estimates versus k), making the adaptive choice of the threshold problematic on the basis of any sample path stability criterion; and

a very peaked MSE, making the choice of the value $k_0 := \arg\min_k \mathrm{MSE}\,(\hat{\gamma}_{n,k})$ difficult. These peculiarities have led researchers to consider the possibility of dealing with the bias term in an appropriate manner, building new estimators $\hat{\gamma}^R_{n,k}$, here called SORB EVI-estimators. In particular, for heavy tails, i.e., $\gamma > 0$, bias reduction is very important for the estimation of γ or of the Pareto index, $\alpha = 1/\gamma$, when the slowly varying part of the Pareto type model disappears at a very slow rate. We consider the following definition (Reiss & Thomas, 2007, §6).

Definition 3.1. Under the second-order condition in (2.2) and for intermediate k, the statistic $\hat{\gamma}^R_{n,k}$, a consistent EVI-estimator, based on the k top OSs in a sample from $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, is said to be a SORB semi-parametric EVI-estimator, if there exist $\sigma_{_R} > 0$ and an asymptotically standard normal random variable P^R_k, such that for a large class of models in $\mathcal{D}_{\mathcal{M}}(EV_\gamma)$, and with $B(\cdot)$ the function in (2.6),

$$\hat{\gamma}^R_{n,k} \stackrel{d}{=} \gamma + \sigma_{_R} P^R_k/\sqrt{k} + o_p\big\{B(n/k)\big\}\,. \tag{3.1}$$

Notice that for the SORB EVI-estimators, we no longer have a dominant component of bias of the order of $B(n/k)$, as in (2.6). Therefore,

$$\sqrt{k}\big(\hat{\gamma}^R_{n,k} - \gamma\big) \xrightarrow[n\to\infty]{d} \mathrm{Normal}\big(0, \sigma^2_{_R}\big)$$

not only when $\sqrt{k}B(n/k) \to 0$ (as for classical estimators), but also when $\sqrt{k}B(n/k) \to \lambda$, finite and non-null. Such a bias reduction provides usually a stable sample path for a wider region of k-values, a 'bath-shaped' MSE and a reduction of the MSE to the optimal level.

Such an approach has been carried out for heavy tails in different manners. The key ideas are either to find ways of getting rid of the dominant component $b\,B(n/k)$ of bias, in (2.6), or to go further into the second-order behaviour of the basic statistics used for the estimation of γ, like the log-excesses or the scaled log-spacings, in (2.4). We first mention some pre-2000 results about bias-corrected estimators in EVT. Such estimators date back to Gomes (1994b), Drees (1996) and Peng (1998), among others. Gomes uses the *generalized jackknife* (GJ) methodology in Gray & Schucany (1972), and Peng deals with linear combinations of appropriate EVI-estimators, in a spirit close to that associated to the GJ technique. Feuerverger & Hall (1999) discuss the question of the possible misspecification of the second-order parameter ρ at -1, a value that corresponds to many commonly used heavy-tailed models, like the Fréchet. Within the second-order framework, Beirlant *et al.* (1999) investigate the accommodation of bias in the scaled log-spacings and derive approximate 'ML' and 'least squares' SORB EVI-estimators. In §3.1.1, we provide details about the GJ EVI-estimation. In §3.1.2 we briefly review an approximate ML approach, together with the introduction of simple SORB EVI-estimators based on the scaled log-spacings or the log-excesses, in

(2.4). Second-order parameters are usually decisive for the bias reduction, and we deal with their estimation in §3.1.3. Finally in §3.1.4, we conclude with some remarks about further literature on SORB EVI-estimation, including the recent first steps on SORB EVI-estimation for a general $\gamma \in \mathbb{R}$.

3.1.1. A brief review of GJ estimators of a positive EVI

The pioneering SORB EVI-estimators are, in a certain sense, GJ estimators, i.e., affine combinations of well-known estimators of γ. For details on the GJ methodology, see Gray & Schucany (1972). Whenever we are dealing with semi-parametric EVI-estimators, or even estimators of other parameters of extreme events, we usually have information about their asymptotic bias. We can thus choose estimators with similar asymptotic properties, and build the associated GJ random variable or statistic. This methodology has been used in Gomes *et al.* (2000, 2002b), among others, and was revisited by Gomes *et al.* (2011c). Indeed, if the second-order condition in (2.3) holds, we can easily find two statistics $\hat{\gamma}_{n,k}^{(j)}$, such that (2.6) holds for both. The ratio between the dominant components of bias of $\hat{\gamma}_{n,k}^{(1)}$ and $\hat{\gamma}_{n,k}^{(2)}$ is $q = b_1/b_2 = q(\rho)$, and we get the GJ random variable,

$$\hat{\gamma}_{n,k}^{\mathrm{GJ}(\rho)} := \big(\hat{\gamma}_{n,k}^{(1)} - q(\rho)\,\hat{\gamma}_{n,k}^{(2)}\big)/\big\{1 - q(\rho)\big\}\,. \tag{3.2}$$

Then under the second-order condition in (2.3), a distributional representation of the type in (3.1) holds for $\hat{\gamma}_{n,k}^{\mathrm{GJ}(\rho)}$, with $\sigma_{GJ}^2 > \sigma_H^2 = \gamma^2$ and $\big(P_k^R, B(n/k)\big)$ replaced by $\big(P_k^{\mathrm{GJ}}, A(n/k)\big)$. The same result remains true for the GJ EVI-estimator, $\gamma_{n,k}^{\mathrm{GJ}(\hat{\rho})}$, provided that $\hat{\rho} - \rho = o_p(1)$ for all k on which we initially base the EVI-estimation. Then (Gomes & Martins, 2002), if $\sqrt{k}\;A(n/k) \to \lambda$, finite,

$$\sqrt{k}\big(\hat{\gamma}_{n,k}^{\mathrm{GJ}(\hat{\rho})} - \gamma\big) \xrightarrow[n\to\infty]{d} \mathrm{Normal}\big(0, \sigma_{GJ}^2\big)\,. \tag{3.3}$$

The result in (3.3), comes from the fact that, through the use of Taylor's expansion, we can write

$$\hat{\gamma}_{n,k}^{\mathrm{GJ}(\hat{\rho})} \overset{d}{=} \hat{\gamma}_{n,k}^{\mathrm{GJ}(\rho)}(k) + \big(\hat{\rho} - \rho\big)\Big[O_p\big(1/\sqrt{k}\big) + O_p\big\{A(n/k)\big\}\Big]\big\{1 + o_p(1)\big\}\,. \tag{3.4}$$

A closer look at (3.4) reveals that it does not seem convenient to compute $\hat{\rho}$ at the value k considered for the EVI-estimation. Indeed, if we do that, and since we have $\hat{\rho} - \rho = \hat{\rho}_k - \rho = O_p\big[1/\big\{\sqrt{k}\;A(n/k)\big\}\big]$ (see Fraga Alves *et al.*, 2003a), $(\hat{\rho} - \rho)\,A(n/k)$ is a term of the order of $1/\sqrt{k}$, and the asymptotic variance of the EVI-estimator will change. Gomes *et al.* (2000) have suggested the misspecification of ρ at $\rho = -1$, essentially due not only to the high bias and variance of the existing estimators of ρ at that time, but also to the idea of considering $\hat{\rho} = \hat{\rho}_k$. Nowadays, the use of any of the algorithms in Gomes & Pestana (2007a,b), among others, enables us to get the limiting result in (3.3), for k-values such that $\sqrt{k}\;A(n/k) \to \infty$, as $n \to \infty$.

3.1.2. Accommodation of bias in the scaled log-spacings and in the log-excesses: alternative SORB EVI-estimators

The ML EVI-estimation based on the scaled log-spacings. The accommodation of bias in the scaled log-spacings U_i in (2.4) has also been a source of inspiration for the building of SORB EVI-estimators. Under the second-order condition in (2.3), but for $\rho < 0$, i.e., working in Hall's class of Pareto-type models (Hall, 1982), with a right-tail function $\overline{F}(x) = Cx^{-1/\gamma}\big(1 + Dx^{\rho/\gamma} + o\big(x^{\rho/\gamma}\big)\big)$, as $x \to \infty$, $C > 0$, D real, $\rho < 0$, we can choose in (2.3),

$$A(t) = \alpha\, t^{\rho} =: \gamma\, \beta\, t^{\rho}\,, \qquad \beta \in \mathbb{R}, \quad \rho < 0\,, \tag{3.5}$$

where β can be regarded as a slowly varying function. Beirlant *et al.* (1999) provide the approximation

$$U_i \sim \big\{\gamma + A(n/k)\,(i/k)^{-\rho}\big\}\, E_i\,, \qquad 1 \le i \le k\,, \tag{3.6}$$

where E_i, $i \ge 1$, denotes a sequence of IID standard exponential random variables. Feuerverger and Hall (1999) consider the approximation

$$U_i \sim \gamma \exp\big(A(n/k)\,(i/k)^{-\rho}/\gamma\big)\, E_i = \gamma \exp\big(A(n/i)/\gamma\big)\, E_i\,, \quad 1 \le i \le k\,. \tag{3.7}$$

The approximation (3.6), or equivalently (3.7), has been made more precise in the asymptotic sense, in Beirlant *et al.* (2002). The use of the approximation in (3.7) and the joint maximization, in γ, β and ρ, of the approximate log-likelihood of the scaled log-spacings,

$$\log L\big(\gamma, \beta, \rho; U_i, 1 \le i \le k\big) \;=\; -k \log \gamma - \beta \sum_{i=1}^{k} (i/n)^{-\rho} - \frac{1}{\gamma} \sum_{i=1}^{k} e^{-\beta (i/n)^{-\rho}}\, U_i\,,$$

led Feuerverger and Hall to an explicit expression for $\hat{\gamma}$,

$$\hat{\gamma} \;=\; \hat{\gamma}_{n,k}^{\mathrm{FH}(\hat{\beta},\hat{\rho})} := \frac{1}{k} \sum_{i=1}^{k} e^{-\hat{\beta} (i/n)^{-\hat{\rho}}}\, U_i\,, \tag{3.8}$$

as a function of $\hat{\beta}$ and $\hat{\rho}$, where $\hat{\beta} = \hat{\beta}_{n,k}^{\mathrm{FH}(\hat{\rho})}$ and $\hat{\rho} = \hat{\rho}_{n,k}^{\mathrm{FH}}$ are both computed at the same k used for the EVI-estimation, and are numerically obtained through

$$(\hat{\beta}, \hat{\rho}) := \underset{(\beta,\rho)}{\arg\min} \left\{ \log\left(\frac{1}{k} \sum_{i=1}^{k} e^{-\beta (i/n)^{-\rho}}\, U_i \right) + \beta \left(\frac{1}{k} \sum_{i=1}^{k} (i/n)^{-\rho} \right) \right\}. \tag{3.9}$$

If k is intermediate and the second-order condition (2.3) hold, it is possible to state that if ρ is unknown as well as β, as usually happens, and they are both estimated through the above mentioned ML technique,

$$\sqrt{k}\, \left(\hat{\gamma}_{n,k}^{\mathrm{FH}(\hat{\beta},\hat{\rho})} - \gamma \right) \xrightarrow[n\to\infty]{d} \mathrm{Normal}\left(0, \sigma^2_{\mathrm{FH}} = \gamma^2 \left(\frac{1-\rho}{\rho} \right)^4 \right). \tag{3.10}$$

Again, even when $\sqrt{k}\,A(n/k) \to \lambda$, non-null, we have a null asymptotic bias for the reduced-bias EVI-estimator, but at the expenses of a larger asymptotic variance, ruled by $\sigma^2_{\rm FH} = \gamma^2\big\{(1-\rho)/\rho\big\}^4$. Note that the asymptotic variance is smaller, and given by $\gamma^2\big\{(1-\rho)/\rho\big\}^2$, if we assume ρ to be known.

A simplified maximum likelihood EVI-estimator based on the external estimation of ρ. The use of the first-order approximation, $e^x = 1+x$, as $x \to 0$, in the two ML equations that provided before $(\hat\beta, \hat\rho)$, led Gomes & Martins (2002) to an explicit estimator for β, given by

$$\hat\beta^{\rm GM(\hat\rho)}_{n,k} := \left(\frac{k}{n}\right)^{\hat\rho} \frac{\Big(\frac{1}{k}\sum_{i=1}^k \big(\frac{i}{k}\big)^{-\hat\rho}\Big)\hat C_0 - \hat C_1}{\Big(\frac{1}{k}\sum_{i=1}^k \big(\frac{i}{k}\big)^{-\hat\rho}\Big)\hat C_1 - \hat C_2}\,, \qquad \hat C_j = \frac{1}{k}\sum_{i=1}^k \Big(\frac{i}{k}\Big)^{-j\hat\rho} U_i\,, \tag{3.11}$$

and, on the basis of an appropriate consistent estimator $\hat\rho$ of ρ, they suggest the following approximate ML estimator for the EVI, γ,

$$\hat\gamma^{\rm GM(\hat\rho)}_{n,k} := \frac{1}{k}\sum_{i=1}^k U_i - \hat\beta^{\rm GM(\hat\rho)}_{n,k}\left(\frac{n}{k}\right)^{\hat\rho}\hat C_1\,. \tag{3.12}$$

The estimator in (3.12) is clearly a bias-corrected Hill estimator, i.e., the dominant component of the bias of the H-estimator, equal to $A(n/k)/(1-\rho) = \gamma\beta(n/k)^\rho/(1-\rho)$ is estimated through $\hat\beta^{\rm GM(\hat\rho)}_{n,k}(n/k)^{\hat\rho}\,\hat C_1$, and directly removed from the H-estimator, which can also be written as $\gamma^{\rm H}_{n,k} = \sum_{i=1}^k U_i/k$. Under the same conditions as before, the asymptotic variance of $\hat\gamma^{\rm GM(\hat\rho)}_{n,k}$ is $\sigma^2_{\rm GM} = \gamma^2(1-\rho)^2/\rho^2 < \sigma^2_{\rm FH}$, but still greater than $\sigma^2_{\rm H} = \gamma^2$.

External estimation of second-order parameters and the weighted Hill (WH) EVI-estimator. In a trial to accommodate bias in the excesses over a high random threshold, Gomes *et al.* (2004b) were led, for heavy tails, to a weighted combination of the log-excesses V_{ik}, $1 \le i \le k < n$, also in (2.4), giving rise to the WH EVI-estimator in Gomes *et al.* (2008c),

$$\hat\gamma^{\rm WH}_{n,k,\hat\beta,\hat\rho} := \frac{1}{k}\sum_{i=1}^k p_{ik}(\hat\beta,\hat\rho)\,V_{ik}\,, \quad p_{ik}(\hat\beta,\hat\rho) := e^{\hat\beta\,(n/k)^{\hat\rho}\left((i/k)^{-\hat\rho}-1\right)/\left(\hat\rho\,\ln(i/k)\right)}, \tag{3.13}$$

where $(\hat\beta, \hat\rho)$ are suitable consistent estimators of second-order parameters (β, ρ). The key to the success of the WH-estimator lies in the estimation of β and ρ at a level k_1, such that $k = o(k_1)$, with k the number of top OSs used for the EVI-estimation. The level k_1 needs to be such that $(\hat\beta, \hat\rho)$ is consistent for the estimation of (β, ρ) and $\hat\rho - \rho = o_p(1/\ln n)$. For more details on the choice of k_1, see Gomes *et al.* (2008c), and more recently Caeiro *et al.* (2009). Compared to the SORB EVI-estimators available in the literature and published prior to 2005, this EVI-estimator is a MVRB EVI-estimator, in the sense that, in comparison with the Hill estimator, it keeps the same asymptotic variance $\sigma^2_{\rm WH} = \sigma^2_{\rm H} = \gamma^2$

and a smaller order asymptotic bias, outperforming the H-estimator for all k. Related work appears in Caeiro *et al.* (2005) and Gomes *et al.* (2007b). Gomes *et al.* (2007b) suggest the computation of the β-estimator $\hat{\beta}_{n,k}^{\mathrm{GM}(\hat{\rho})}$, used at (3.12), at the level k_1 used for the estimation of ρ. With the notation $\hat{\beta} := \hat{\beta}_{n,k_1}^{\mathrm{GM}(\hat{\rho})}$, they suggest thus the replacement of the estimator in (3.12) by

$$\hat{\gamma}_{n,k}^{\overline{\mathrm{M}}(\hat{\beta},\hat{\rho})} := \gamma_{n,k}^{\mathrm{H}} - \hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}} \hat{C}_1 \ , \tag{3.14}$$

where $\gamma_{n,k}^{\mathrm{H}}$ denotes the H-estimator, and $(\hat{\beta},\hat{\rho})$ are appropriate consistent estimators of the second-order parameters (β,ρ). With the same objectives, but with a simpler expression, we also mention the estimator (Caeiro *et al.*, 2005).

$$\hat{\gamma}_{n,k}^{\overline{\mathrm{H}}(\hat{\beta},\hat{\rho})} := \gamma_{n,k}^{\mathrm{H}}\big(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\big) \ . \tag{3.15}$$

The dominant component of the bias of the H-estimator is estimated in (3.15) through $\gamma_{n,k}^{\mathrm{H}}\,\hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})$, and directly removed from Hill's classical EVI-estimator. The appropriate estimation of β and ρ at a level k_1 of a higher order than the level k used for the EVI-estimation, enables, for a large diversity of heavy-tailed models, the reduction of bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator. Reiss & Thomas (2007), §6, and Gomes *et al.* (2008a) review this topic.

3.1.3. Second-order parameters estimation for heavy tails

The first estimator of the parameter ρ, in (2.3), with $A(\cdot)$ given in (3.5), but where β can possibly be any slowly varying function, appears in Hall & Welsh (1985). Peng (1998) claims that no good estimator for the second-order parameter ρ was then available in the literature, and considers a new ρ-estimator, alternative to the ones in Hall & Welsh (1985), Beirlant *et al.* (1996c) and Drees & Kaufmann (1998). Another estimator of ρ appears in Gomes *et al.* (2002a), and more recently, we mention the classes of ρ-estimators in Goegebeur *et al.* (2008; 2010b) and Ciuperca & Mercadier (2010). Here we choose particular members of the class of estimators of the second-order parameter ρ proposed by Fraga Alves *et al.* (2003a). Under appropriate general conditions, they are asymptotically normal estimators of ρ, if $\rho < 0$, which show highly stable sample paths as functions of k, the number of top OSs used, for a wide range of large k-values. Such a class of estimators, parameterised in a tuning real parameter $\tau \in \mathbb{R}$, is defined as

$$\hat{\rho}_{n,k}^{(\tau)} := -\left|3\big(T_{n,k}^{(\tau)}-1\big)/\big(T_{n,k}^{(\tau)}-3\big)\right|, \quad T_{n,k}^{(\tau)} := \frac{\big(M_{n,k}^{(1)}\big)^{\tau} - \big(M_{n,k}^{(2)}/2\big)^{\tau/2}}{\big(M_{n,k}^{(2)}/2\big)^{\tau/2} - \big(M_{n,k}^{(3)}/6\big)^{\tau/3}} \ , \tag{3.16}$$

with $M_{n,k}^{(j)}$ given in (2.5) and with the notation $a^{b\tau} = b \ln a$ whenever $\tau = 0$.

Gomes & Martins (2002) provide an explicit estimator for β, based on the scale log-spacings U_i, in (2.4), and already given in (3.11). An additional estimator of β, is provided in Caeiro & Gomes (2006). See also Gomes *et al.* (2010), for a β-estimator based on the log-excesses.

Algorithms for the estimation of second-order parameters can be found in Gomes & Pestana (2007a,b). The use of such algorithms, where the ρ-estimator is computed at $k_1 = [n^{1-\epsilon}]$, with ϵ small, say $\epsilon = 0.001$, enables us to guarantee that, for a large class of heavy-tailed models, as $n \to \infty$, $\left(\hat{\rho}^{(\tau)}_{n,k_1} - \rho\right) \ln n = o_p(1)$, a crucial property of the ρ-estimator, if we do not want to increase the asymptotic variance of the random variable, function of (β, ρ), underlying the SORB EVI-estimator. Such a crucial property can potentially be achieved if we compute $\hat{\rho}$ at its optimal level (see Caeiro *et al.*, 2009), but the adaptive choice of such a level is still an open research topic.

3.1.4. Additional Literature on SORB EVI-estimation

Other approaches to bias reduction, in the estimation of a positive EVI can be found in Gomes & Martins (2001, 2004), Caeiro & Gomes (2002), Gomes *et al.* (2004a; 2005a; 2005b; 2007c; 2011a), Canto e Castro & de Haan (2006), and Willems *et al.* (2007), among others. Recently, Cai *et al.* (2011) introduced the first SORB estimators for $\gamma \in \mathbb{R}$, based on the PWM methodology.

3.2. SORB semi-parametric estimation of other parameters of extreme events

Reduced bias quantile estimators have been studied in Matthys *et al.* (2004) and Gomes & Figueiredo (2006), who consider the classical SORB EVI-estimators. Gomes & Pestana (2007b) and Beirlant *et al.* (2008) incorporate the MVRB EVI-estimators in Caeiro *et al.* (2005) and Gomes *et al.* (2007b) in high quantile semi-parametric estimation. See also Diebolt *et al.* (2008b), Beirlant *et al.* (2009), Caeiro & Gomes (2009), Li *et al.* (2010). For a SORB estimation of the Weibull-tail coefficient, we mention Diebolt *et al.* (2008a). Finally, for SORB endpoint estimation, we mention Li & Peng (2009).

4. OVERALL COMMENTS AND FURTHER RESEARCH

We shall next discuss a few areas where a lot has been already done but further research is still welcome. In our opinion, SUE is still a lively topic of research. Important developments have appeared recently in the area of *spatial extremes*, where *parametric models* seem again to be quite relevant. In this case, and now that we have access to highly sophisticated computational techniques, a great variety of *parametric models* can further be considered. And in a semi-parametric framework, topics like *threshold selection*, *trends and change points* in the *tail behaviour*, and *clustering*, among others, are still challenging.

4.1. Rates of convergence and penultimate approximations

An important problem in EVT concerns the rate of convergence of $F^n(a_n\,x+b_n)$ towards $G_\gamma(x)$, in (1.2), or, equivalently, the search for estimates of the difference $d_n(F,G_\gamma,x):=F^n(a_n\,x+b_n)-G_\gamma(x)$. Indeed, as detailed in §1, parametric inference on the right-tail of F, usually unknown, is done on the basis of the identification of $F^n(a_n\,x+b_n)$ and of $G_\gamma(x)$. And the rate of convergence may or may not support use of the commonest models in SUE. As noted by Fisher & Tippett (1928), although the normal CDF $\Phi\in\mathcal{D}_{\mathcal{M}}(G_0)$, the convergence of $\Phi^n(a_n\,x+b_n)$ towards $G_0(x)$ is extremely slow. They then show, through the use of skewness and kurtosis coefficients as indicators of closeness, that $\Phi^n(x)$ is 'closer' to a suitable penultimate $G_{-1/\gamma_n}\{(x-\lambda_n)/\delta_n\}$, for $\gamma_n>0$, $\lambda_n\in\mathbb{R}$, $\delta_n>0$, than to the ultimate $G_0\{(x-b_n)/a_n\}$. Such an approximation is the so-called *penultimate approximation* and several penultimate models have been advanced by several authors. Dated overviews of the modern theory of rates of convergence in EVT, introduced in Anderson (1971), can be seen in Galambos (1984) and Gomes (1994a). More recently, Gomes & de Haan (1999) derived, for all $\gamma\in\mathbb{R}$, exact penultimate approximation rates with respect to the variational distance, under appropriate differentiability assumptions. Kaufmann (2000) proved, under weaker conditions, a result related to that in Gomes & de Haan (1999). This penultimate or pre-asymptotic behaviour has further been studied by Raoult & Worms (2003) and Diebolt & Guillou (2005), among others. Other type of penultimate approximations have been considered in Smith (1987b). Among them, we mention a penultimate parametric model of the type

$$PG_\gamma(x;r)\;=\;\exp\Big[-(1+\gamma\,x)^{-1/\gamma}\left\{1+r(1+\gamma\,x)^{-1/\gamma}\right\}\Big]. \tag{4.1}$$

These models surely deserve deeper statistical consideration. *Penultimate models* seem interesting alternatives to the classical models but have never been much used. Concomitantly, the convergence of the estimators can be very slow when $\rho=0$ or $\rho^*=0$, as happens with normal and loggamma distributions, important models in many areas, and alternative estimation procedures are still needed.

4.2. Max-semistable laws

We also mention the class of *max-semistable* (MSS) laws, introduced by Grienvich (1992a, 1992b), Pancheva (1992), and further studied in Canto e Castro *et al.* (2000) and in Temido & Canto e Castro (2003). Such a class is more general than the class of MS laws, given in (1.2). Indeed, the possible MSS laws are

$$G_{\gamma,\nu}(x) = \begin{cases} \exp\Big[-\nu\big\{\ln(1+\gamma\,x)\big\}\,(1+\gamma\,x)^{-1/\gamma}\Big], \quad 1+\gamma\,x>0\,, & \text{if } \gamma\neq 0\,, \\ \exp\big\{-\nu(x)\exp(-x)\big\}, \quad x\in\mathbb{R}\,, & \text{if } \gamma=0\,, \end{cases}$$

where $\nu(\cdot)$ is a positive, limited and periodic function. A unit ν-function enables us to get the MS laws in (1.2). Discrete models like the geometric and negative binomial, and some multimodal continuous models, are in $\mathcal{D}_{\mathcal{MSS}}$ but not in $\mathcal{D}_{\mathcal{M}}$. A recent survey of the topic can be found in Pancheva (2010). Generalized P-statistics have been used in Canto e Castro and Dias (2011), to develop methods of estimation in the MSS context. See also Canto e Castro *et al.* (2011). Such a diversity of models, if duly exploited from a statistical point of view, can surely provide fruitful topics of research, both in parametric and semi-parametric setups.

4.3. Invariance versus non-invariance

In *statistics of extremes* most of the methods of estimation are dependent on the log-excesses, and consequently, are *non-invariant* with respect to *shifts of the data*. But the invariance not only to changes in scale but also to changes in location of any EVI estimator is statistically appealing. Wouldn't be sensible to use the PORT methodology in Araújo Santos *et al.* (2006), and consider PORT EVI-estimators based on the transformed sample

$$X_i^* := X_i - X_{[np]+1,n}\,, \qquad 0<p<1, \quad 1\le i\le n\,? \tag{4.2}$$

A similar procedure was used by Fraga Alves *et al.* (2009), who also propose a class of EVI-estimators alternative to the MM-estimator, invariant to changes in location, and dependent on a similar *tuning parameter* p, $0<p<1$. Such estimators have the same functional expression as the original estimator, but the original observation X_i is replaced everywhere by X_i^*, in (4.2), $1\le i\le n$. A similar procedure has been used for the H and M EVI-estimators, and for quantile estimation in Araújo Santos *et al.* (2006). For PORT quantile estimation, see also Henriques-Rodrigues & Gomes (2009). The shift invariant versions, dependent on the tuning parameter p, have properties similar to those of the original estimator T, provided we keep to appropriate k-values and choose an appropriate *tuning parameter* p. For recent research on this topic see Gomes *et al.* (2011b), but more is needed.

4.4. Adaptive selection of sample fraction or threshold

A *threshold* is often set 'almost arbitrarily' (for instance at the 90% or the 95% sample quantile). However, the choice of the threshold, or equivalently of the number k of top OSs to be used is crucial for a reliable estimation of any parameter of extreme events. The topic has already been extensively studied for classical EVI-estimators, for which (2.6) holds. In Hall & Welsh (1985), Hall (1990), Beirlant *et al.* (1996c), Drees & Kaufmann (1998) and Danielsson *et al.* (2001), methods for the adaptive choice of k are proposed for the H-estimator, some of them involving the bootstrap technique. Gomes & Oliveira (2001) also uses the bootstrap methodology to provide an adaptive choice of the threshold, alternative to that in Danielsson *et al.* (2001), and easy to generalise to other semi-parametric estimators of parameters of extreme events. For a general γ and for the M-estimator and a generalized P-estimator, see Draisma *et al.* (1999). These authors also use the bootstrap. Beirlant *et al.* (2002) consider the exponential regression model (ERM) introduced in Beirlant *et al.* (1999), discuss applications of the ERM to the selection of the optimal sample fraction in EV estimation, and derive a connection between the new choice strategy in the paper and the diagnostic of Guillou & Hall (2001). Csörgő & Viharos (1998) provide a data-driven choice of k for kernel estimators. Apart from the papers by Drees & Kaufmann (1998) and Guillou and Hall (2001), where choice of the optimal sample fraction is based on bias stability, the other papers make the optimal choice minimizing the estimated MSE. Possible heuristic choices are provided in Gomes & Pestana (2007b), Gomes *et al.* (2008e) and Beirlant *et al.* (2011). The adaptive SORB estimation is still giving its first steps. We can however mention the recent papers by Gomes *et al.* (2011a,d). Is it sensible to use bootstrap computational intensive procedures for threshold selection or there will be simpler techniques possibly related with bias pattern? Is it possible to apply a similar methodology for the estimation of other parameters of extreme events?

4.5. Other possible topics of research in SUE

Testing whether $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma)$, for a certain γ, is a crucial topic, already dealt with in several articles mentioned in § 1.2 and 2.3. And what about testing second-order and third-order conditions? Change-point detection is also a challenging topic. And SUE for weakly dependent data, with all problems related with clustering of extreme values, merits further research. SUE for randomly censored data is another challenging topic. See Beirlant *et al.* (2007; 2010), Einmahl *et al.* (2008a) and Gomes & Neves (2011). Statistics of extremes in athletics and estimation of the endpoint is another of the relevant topics in SUE.

We mention the recent papers by Einmahl & Magnus (2008), Li & Peng (2009), Einmahl & Smets (2011), Henriques-Rodrigues *et al.* (2011) and Li *et al.* (2011). Recent models, like the *extreme value Birnbaum–Saunders* model in Ferreira *et al.* (2011), can also become relevant in the area of SUE. Moreover, the estimation of second and higher order parameters still deserves further attention, particularly due to the importance of such estimation in SORB estimators of parameters of extreme events.

ACKNOWLEDGMENTS

Research partially supported by National Funds through FCT — *Fundação para a Ciência e a Tecnologia*, projects PEst-OE/MAT/UI0006/2011 (CEAUL) and PEst-OE/MAT/UI0297/2011 (CMA/UNL) and PTDC/FEDER.

REFERENCES

Aarssen, K. & de Haan, L. (1994). On the maximal life span of humans, *Math. Pop. Stud.*, **4**, 259–281.

Anderson, C.W. (1971). *Contributions to the Asymptotic Theory of Extreme Values*, Ph.D. Thesis, University of London.

Araújo Santos, P.; Fraga Alves, M.I. & Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation, *Revstat*, **4**, 227–247.

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

Beirlant, J.; Teugels, J. & Vynckier, P. (1996a). *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, Belgium.

Beirlant, J.; Vynckier, P. & Teugels, J. (1996b). Excess functions and estimation of the extreme-value index, *Bernoulli*, **2**, 293–318.

Beirlant, J.; Vynckier, P. & Teugels, J. (1996c). Tail index estimation, Pareto quantile plots and regression, *J. Am. Statist. Assoc.*, **91**, 1659–1667.

Beirlant, J.; Teugels, J. & Vynckier, P. (1998). *Some thoughts on extreme values.* In "Probability Towards 2000" (L. Accardi and C.C. Heyde, Eds.), Lecture Notes in Statistics, **128**, 58–73, Springer, New York.

Beirlant, J.; Dierckx, G.; Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

Beirlant, J.; Dierckx, G.; Guillou, A. & Starica, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.

Beirlant, J.; Goegebeur, Y.; Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Chichester.

Beirlant, J.; Dierckx, G. & Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots, *Bernoulli*, **11**, 949–970.

Beirlant, J.; de Wet, T. & Goegebeur, Y. (2006). A goodness-of-fit for Pareto-type behaviour, *J. Comput. Appl. Math.*, **186**, 99–116.

Beirlant, J.; Guillou, A.; Dierckx, G. & Fils-Viletard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring, *Extremes*, **10**, 151–174.

Beirlant, J.; Figueiredo, F.; Gomes, M.I. & Vandewalle, B. (2008). Improved reduced bias tail index and quantile estimators, *J. Statist. Plann. Infer.*, **138**, 1851–1870.

Beirlant, J.; Joossens, E. & Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, *J. Statist. Plann. Infer.*, **139**, 2800–2815.

Beirlant, J.; Guillou, A. & Toulemonde, G. (2010). Peaks-over-threshold modeling under random censoring, *Comm. Statist. Th. Meth.*, **39**, 1158–1179.

Beirlant, J.; Boniphace, E. & Dierckx, G. (2011). Generalized sum plots, *Revstat*, **9**, 181–198.

Bermudez, P. de Zea & Amaral-Turkman, M.A. (2003). Bayesian approach to parameter estimation of the generalized Pareto distribution, *Test*, **12**, 259–277.

Bingham, N.; Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation*, Cambridge Univ. Press, Cambridge.

Bottolo, L.; Consonni, G.; Dellaportas, P. & Lijoi, A. (2003). Bayesian analysis of extreme values by mixture modeling, *Extremes*, **6**, 25–47.

Brilhante, M.F. (2004). Exponential versus generalized Pareto — a resistant and robust test, *Revstat*, **2**, 1–13.

Caeiro, F. & Gomes, M.I. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index, *Test*, **11**, 345–364.

Caeiro, C. & Gomes, M.I. (2006). A new class of estimators of a scale second order parameter, *Extremes*, **9**, 193–211.

Caeiro, F. & Gomes, M.I. (2009). Semi-parametric second-order reduced-bias high quantile estimation, *Test*, **18**, 392–413.

Caeiro, F. & Gomes, M.I. (2011). Semi-parametric tail inference through probability-weighted moments, *J. Statist. Plann. Infer.*, **141**, 937–950.

Caeiro, C.; Gomes, M.I. & Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *Revstat*, **3**, 113–136.

Caeiro, F.; Gomes, M.I. & Henriques-Rodrigues, L. (2009). Reduced-bias tail index estimators under a third order framework, *Comm. Statist. Th. Meth.*, **38**, 1019–1040.

Cai, J.; de Haan, L. & Zhou, C. (2011). *Bias correction in extreme value statistics with index around zero*, available at: `http://people.few.eur.nl/ldehaan/BiasCorrection11.pdf`

Canto e Castro, L. & Dias, S. (2011). Generalized Pickands estimators for the tail index parameter and max-semistability, *Extremes*, **14**, 429–449.

Canto e Castro, L. & de Haan, L. (2006). A class of distribution functions with less bias in extreme value estimation, *Statist. Prob. Lett.*, **76**, 1617–1624.

Canto e Castro, L.; de Haan, L. & Temido, M.G. (2000). Rarely observed maxima, *Th. Prob. Appl.*, **45**, 779–782.

Canto e Castro, L.; Dias, S. & Temido, M.G. (2011). Looking for max-semistability: a new test for the extreme value condition, *J. Statist. Plann. Infer.*, **141**, 3005–3020.

Castillo, E. & Hadi, A. (1997). Fitting the generalized Pareto distribution to data, *J. Am. Statist. Assoc.*, **92**, 1609–1620.

Castillo, E.; Hadi, A.; Balakrishnan, N. & Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, New Jersey.

Chaouche, A. & Bacro, J.-N. (2004). A statistical test procedure for the shape parameter of a generalized Pareto distribution, *Comput. Statist. Data Anal.*, **45**, 787–803.

Choulakian, V. & Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43**, 478–484.

Ciuperca, G. & Mercadier, C. (2010). Semi-parametric estimation for heavy tailed distributions, *Extremes*, **13**, 55–87.

Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*, Springer, New York.

Csörgő, S. & Mason, D.M. (1989). *Simple estimators of the endpoint of a distribution.* In "Extreme Value Theory, Proceedings Oberwolfach 1987" (J. Hüsler and R.-D. Reiss, Eds., 1989), 132–147, Springer, Berlin–Heidelberg.

Csörgő, S. & Viharos, L. (1998). *Estimating the tail index.* In "Asymptotic Methods in Probability and Statistics" (B. Szyszkowicz, Ed.), 833–881, North-Holland.

Csörgő, S.; Deheuvels, P. & Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *Ann. Statist.*, **13**, 1050–1077.

Danielsson, J.; de Haan, L.; Peng, L. & de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in the tail index estimation, *J. Mult. Anal.*, **76**, 226–248.

Davison, A.C. (1984). *Modelling excesses over high thresholds.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), 461–482, D. Reidel, Dordrecht.

Dekkers, A.L.M. & de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation, *Ann. Statist.*, **17**, 1795–1832.

Dekkers, A.; Einmahl, J. & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, **17**, 1833–1855.

Diebolt, J. & Guillou, A. (2005). Asymptotic behavior of regular estimators, *Revstat*, **3**, 19–44.

Diebolt, J.; El-Aroui, M.A.; Garrido, M. & Girard, S. (2005). Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling, *Extremes*, **8**, 57–78.

Diebolt, J.; Guillou, A. & Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *J. Statist Plann. Infer., 137*, 841–857.

Diebolt, J.; Gardes, L.; Girard, S. & Guillou, A. (2008a). Bias-reduced estimators of the Weibull tail-coefficient, *Test*, **17**, 311–331.

Diebolt, J.; Gardes, L.; Girard, S. & Guillou, A. (2008b). Bias-reduced extreme quantiles estimators of Weibull tail-distributions, *J. Statist. Plann. Infer.*, **138**, 1389–1401.

Diebolt, J.; Guillou, A.; Naveau, P. & Ribereau, P. (2008c). Improving probability-weighted moment methods for the generalized extreme value distribution, *Revstat*, **6**, 33–50.

Dietrich, D.; de Haan, L. & Hüsler, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.

Dijk, V. & de Haan, L. (1992). *On the estimation of the exceedance probability of a high level*. In "Order statistics and nonparametrics: theory and applications" (P.K. Sen and I.A. Salama, Eds.), 79–92, Elsevier, Amsterdam.

Draisma, G.; de Haan, L.; Peng, L. & Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes*, **2**, 367–404.

Drees, H. (1995). Refined Pickands estimator of the extreme value index, *Ann. Statist.*, **23**, 2059–2080.

Drees, H. (1996). Refined Pickands estimators with bias correction, *Comm. Statist. Th. Meth.*, **25**, 837–851.

Drees, H. (1998). A general class of estimators of the extreme value index, *J. Statist. Plann. Inf.*, **66**, 95–112.

Drees, H. & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. Application*, **75**, 149–172.

Drees, H.; Ferreira, A. & de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Appl. Prob.*, **14**, 1179–1201.

Drees, H.; de Haan, L. & Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions, *J. Stat. Plann. Infer.*, **136**, 3498–3538.

Dwass, M. (1964). Extremal processes, *Ann. Math. Statist.*, **35**, 1718–1725.

Einmahl, J. & Magnus, J.R. (2008). Records in athletics through extreme-value theory, *J. Am. Statist. Assoc.*, **103**, 1382–1391.

Einmahl, J. & Smeets, S.G.W.R. (2011). Ultimate 100-m world records through extreme-value theory, *Statist. Neerl.*, **65**, 32–42.

Einmahl, J.H.J.; Fils-Villetard, A. & Guillou, A. (2008a). Statistics of extremes under random censoring, *Bernoulli*, **14**, 207–227.

Embrechts, P.; Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin–Heidelberg.

Falk, M. (1994). Efficiency of a convex combination of Pickands estimator of the extreme value index, *J. Nonparam. Statist.*, **4**, 133–147.

Falk, M. (1995a). Some best parameter estimates for distributions with finite endpoint, *Statistics*, **27**, 115–125.

Falk, M.; Hüsler, J. & Reiss, R.-D. (1994; 2004). *Laws of Small Numbers: Extremes and Rare Events*, 1st edition; 2nd edition, Birkhäuser, Basel.

Falk, F.; Guillou, A. & Toulemonde, G. (2008). A LAN based Neyman smooth test for the generalized Pareto distribution, *J. Statist. Plann. Infer.*, **138**, 2867–2886.

Ferreira, A. (2002). Optimal asymptotic estimation of small exceedance probabilities, *J. Statist. Plann. Infer.*, **104**, 83–102.

Ferreira, A.; de Haan, L. & Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution, *Statistics*, **37**, 401–434.

Ferreira, M.; Gomes, M.I. & Leiva, V. (2011). On an extreme value version of the Birnbaum–Saunders distribution, *Revstat*. Accepted.

Feuerverger, A. & Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Ann. Statist.*, **27**, 760–781.

Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency of the largest or smallest member of a sample, *Proc. Cambridge Phil. Soc.*, **24**, 180–190.

Fraga Alves, M.I. (1992, 1995). Estimation of the tail parameter in the domain of attraction of an extremal distribution, *Extreme Value Theory and Applications* (Villeneuve d'Ascq, 1992) and *J. Statist. Plann. Infer.*, **45**(1–2), 143–173.

Fraga Alves, M.I. (2001). A location invariant Hill-type estimator, *Extremes*, **4**, 199–217.

Fraga Alves, M.I.; Gomes, M.I. & de Haan, L. (2003a). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Math.*, **60**, 193–213.

Fraga Alves, M.I.; de Haan, L. & Lin, T. (2003b). Estimation of the parameter controlling the speed of convergence in extreme value theory, *Math. Meth. Statist.*, **12**, 155–176.

Fraga Alves, M.I.; de Haan, L. & Lin, T. (2006). Third order extended regular variation, *Pub. l'Institut Math.*, **80**, 109–120.

Fraga Alves, M.I.; Gomes, M.I.; de Haan, L. & Neves, C. (2007). A note on second order condition in extremes: linking general and heavy tails conditions, *Revstat*, **5**, 285–305.

Fraga Alves, M.I.; Gomes, M.I.; de Haan, L. & Neves, C. (2009). The mixed moment estimator and location invariant alternatives, *Extremes*, **12**, 149–185.

Fréchet, M. (1927). Sur le loi de probabilité de l'écart maximum, *Ann. Soc. Polonaise Math.*, **6**, 93–116.

Galambos, J. (1984). *Rates of convergence in extreme value theory.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), 347–352, D. Reidel, Dordrecht.

Geluk, J. & de Haan, L. (1980). *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**, 423–453.

Girard, S. (2004). A Hill type estimate of the Weibull tail-coefficient, *Comm. Statist. Th. Meth.*, **33**, 205–234.

Goegebeur, Y. & Guillou, A. (2010). Goodness-of-fit testing for Weibull-type behavior, *J. Statist. Plann. Infer.*, **140**, 1417–1436.

Goegebeur, Y.; Beirlant, J. & de Wet, T. (2008). Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation, *Revstat*, **6**, 51–69.

Goegebeur, Y.; Beirlant, J. & de Wet, T. (2010a). Generalized kernel estimators for the Weibull-tail coefficient, *Comm. Statist. Th. Meth.*, **39**, 3695–3716.

Goegebeur, Y.; Beirlant, J. & de Wet, T. (2010b). Kernel estimators for the second order parameter in extreme value statistics, *J. Statist. Plann. Infer.*, **140**, 2632–2652.

Gomes, M.I. (1989). Generalized Gumbel and likelihood ratio test statistics in the multivariate GEV model, *Comput. Statist. Data Anal.*, **7**, 259–267.

Gomes, M.I. (1994a). *Penultimate behaviour of the extremes.* In "Extreme Value Theory and Applications" (J. Galambos *et al.*, Eds.), 403–418, Kluwer, The Netherlands.

Gomes, M.I. (1994b). *Metodologias Jackknife e bootstrap em estatística de extremos.* In "Actas II Congresso SPE" (Mendes-Lopes *et al.*, Eds.), 31–46 (in Portuguese).

Gomes, M.I. & Figueiredo, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation, *Test*, **15**, 375–396.

Gomes, M.I. & de Haan, L. (1999). Approximation by penultimate extreme value distributions, *Extremes*, **2**, 71–85.

Gomes, M.I. & Martins, M.J. (2001). Generalizations of the Hill estimator: asymptotic versus finite sample behaviour, *J. Statist. Planning Infer.*, **93**, 161–180.

Gomes, M.I. & Martins, M.J. (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter, *Extremes*, **5**, 5–31.

Gomes, M.I. & Martins, M.J. (2004). Bias reduction and explicit semi-parametric estimation of the tail index, *J. Statist. Plann. Infer.*, **124**, 361–378.

Gomes, M.I. & Neves, M.M. (2011). Estimation of the extreme value index for randomly censored data, *Biomet. Lett.*, **48**, 1–22.

Gomes, M.I. & Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction, *Extremes*, **4**, 331–358.

Gomes, M.I. & Pestana, D. (2007a). A simple second order reduced bias' tail index estimator, *J. Statist. Comput. Simul.*, **77**, 487–504.

Gomes, M.I. & Pestana, D. (2007b). A sturdy reduced bias extreme quantile (VaR) estimator, *J. Am. Statist. Assoc.*, **102**, 280–292.

Gomes, M.I.; Martins, M.J. & Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology, *Extremes*, **3**, 207–229.

Gomes, M.I.; de Haan, L. & Peng, L. (2002a). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour, *Extremes*, **5**, 387–414.

Gomes, M.I.; Martins, M.J. & Neves, M.M. (2002b). Generalized Jackknife semi-parametric estimators of the tail index, *Portugaliae Math.*, **59**, 393–408.

Gomes, M.I.; Caeiro, F. & Figueiredo, F. (2004a). Bias reduction of a extreme value index estimator trough an external estimation of the second order parameter, *Statistics*, **38**, 497–510.

Gomes, M.I.; de Haan, L. & Henriques-Rodrigues, L. (2004b). Tail index estimation through the accomodation of bias in the weighted log-excesses, *Notas e Comunicações CEAUL*, 14/2004.

Gomes, M.I.; Figueiredo, F. & Mendonça, S. (2005a). Asymptotically best linear unbiased tail estimators under a second order regular variation, *J. Statist. Plann. Infer.*, **134**, 409–433.

Gomes, M.I.; Miranda, C. & Pereira, H. (2005b). Revisiting the role of the Jackknife methodology in the estimation of a positive extreme value index, *Comm. Statist. Th. Meth.*, **34**, 1–20.

Gomes, M.I.; Canto e Castro, L.; Fraga Alves, M.I. & Pestana, D. (2007a). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions, *Notas e Comunicações CEAUL*, 16/2007. Available at: `http://www.ceaul.fc.ul.pt/notas.html?ano=2007`

Gomes, M.I.; Martins, M.J. & Neves, M. (2007b). Improving second order reduced bias extreme value index estimation, *Revstat*, **5**, 177–207.

Gomes, M.I.; Miranda, C. & Viseu, C. (2007c). Reduced bias extreme value index estimation and the Jackknife methodology, *Statist. Neerl.*, **61**, 243–270.

Gomes, M.I.; Canto e Castro, L.; Fraga Alves, M.I. & Pestana, D. (2008a). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions, *Extremes*, **11**, 3–34.

Gomes, M.I.; de Haan, L. & Henriques-Rodrigues, L. (2008c). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses, *J. R. Statist. Soc. B*, **70**, 31–52.

Gomes, M.I.; Henriques-Rodrigues, L.; Vandewalle, B. & Viseu, C. (2008e). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators, *J. Statist. Comput. Simul.*, **78**, 133–150.

Gomes, M.I.; Henriques-Rodrigues, L.; Pereira, H. & Pestana, D. (2010). Tail index and second order parameters' semi-parametric estimation based on the log-excesses, *J. Statist. Comput. Simul.*, **80**, 653–666.

Gomes, M.I.; Figueiredo, F. & Neves, M.M. (2011a). Adaptive estimation of heavy right tails: the bootstrap methodology in action, *Extremes*, in press. DOI: 10.1007/s10687-011-0146-6.

Gomes, M.I.; Henriques-Rodrigues, L. & Miranda, C. (2011b). Reduced-bias location-invariant extreme value index estimation: a simulation study, *Comm. Statist. Simul. Comput.*, **40**, 424–447.

Gomes, M.I.; Martins, M.J. & Neves, M.M. (2011c). Revisiting the role of the generalized JackKnife methodology in the field of extremes, *Notas e Comunicações CEAUL*, 20/2011. Submitted.

Gomes, M.I.; Mendonça, S. & Pestana, D. (2011d). Adaptive reduced-bias tail index and VaR estimation via the bootstrap methodology, *Comm. Statist. Th. Meth.*, **40**, 2946–2968.

Gray, H.L. & Schucany, W.R. (1972). *The Generalized Jackknife Statistic*, Dekker, New York.

Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Wat. Resour. Res.*, **15**, 1049–1054.

Grienvich, I.V. (1992a). Max-semistable laws corresponding to linear and power normalizations, *Th. Probab. Appl.*, **37**, 720–721.

Grienvich, I.V. (1992b). Domains of attraction of max-semistable laws under linear and power normalizations, *Th. Probab. Appl.*, **38**, 640–650.

Groeneboom, P.; Lopuha, H.P. & de Wolf, P.P. (2003). Kernel-type estimators for the extreme value index, *Ann. Statist.*, **31**, 1956–1995.

Guillou, A. & Hall, P. (2001). A diagnostic for selecting the threshold in extreme-value analysis, *J. R. Statist. Soc. B*, **63**, 293–305.

Guillou, A.; Naveau, P. & You, A. (2010). A folding method for extreme quantiles estimation, *Revstat*, **8**, 21–35.

Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

Gumbel, E.J. (1965). A quick estimation of the parameters in Fréchet's distribution, *Review Intern. Statist. Inst.*, **33**, 349–363.

de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amsterdam.

de Haan, L. (1984). *Slow variation and characterization of domains of attraction.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), 31–48, D. Reidel, Dordrecht, Holland.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

de Haan, L. & Peng, L. (1998). Comparison of tail index estimators, *Statist. Neerl.*, **52**, 60–70.

de Haan, L. & Rootzén, H. (1993). On the estimation of high quantiles, *J. Statist. Plann. Infer.*, **35**, 1–13.

de Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order, *J. Austral. Math. Soc. A*, **61**, 381–395.

Hall, P. (1982). On some simple estimates of an exponent of regular variation, *J. R. Statist. Soc. B*, **44**, 37–42.

Hall, P. (1990). Using the bootstrap to estimate mean-square error and selecting smoothing parameter in non-parametric problems, *J. Mult. Anal.*, **32**, 177–203.

Hall, P. & Welsh, A.H. (1985). Adaptative estimates of parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

Henriques-Rodrigues, L. & Gomes, M.I. (2009). High quantile estimation and the PORT methodology, *Revstat*, **7**, 245–264.

Henriques-Rodrigues, L.; Gomes, M.I. & Pestana, D. (2011). Statistics of extremes in athletics, *Revstat*, **9**, 127-153.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

Hosking, J.R.M. (1984). Testing whether the shape parameter is zero in the generalized extreme value distribution, *Biometrika*, **71**, 367–374.

Hosking, J.R.M. & Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

Hosking, J.R.M.; Wallis, J.R. & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, **27**, 251–261.

Hüsler, J. & Li, D. (2006). On testing extreme value conditions, *Extremes*, **9**, 69–86.

Hüsler, J. & Müller, S. (2005). Iterative estimation of the extreme value index, *Math. Comput. Appl. Probab.*, **7**, 139–148.

Hüsler, J. & Peng, L. (2008). Review of testing issues in extremes: in honor of Professor Laurens de Haan, *Extremes*, **11**, 99–111.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. R. Meteorol. Soc.*, **81**, 158–171.

Johansson, J. (2003). Estimating the mean of heavy-tailed distributions, *Extremes*, **6**, 91–109.

Jurečková, J. & Picek, J. (2001). A class of tests on the tail index, *Extremes*, **4**, 165–183.

Kaufmann, E. (2000). Penultimate approximations in extreme value theory, *Extremes*, **3**, 39–55.

Koning, A.J. & Peng, L. (2008). Goodness-of-fit tests for a heavy tailed distribution, *J. Statist. Plann. Infer.*, **138**, 3960–3981.

Kratz, M. & Resnick, S. (1996). The qq-estimator of the index of regular variation, *Stochastic Models*, **12**, 699–724.

Li, D. & Peng, L. (2009). Does bias reduction with external estimator of second order parameter work for endpoint? *J. Statist. Plann. Infer.*, **139**, 1937–1952.

Li, D.; Peng, L. & Yang, J. (2010). Bias reduction for high quantiles, *J. Statist. Plann. Infer.*, **140**, 2433–2441.

Li, D.; Peng, L. & Qi, Y. (2011). Empirical likelihood confidence intervals for the endpoint of a distribution function, *Test*, **20**, 353–366.

Ling, C.; Peng, Z. & Nadarajah, S. (2011). Location invariant Weiss–Hill estimator, *Extremes*, DOI 10.1007/s10687-011-0134-x.

Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, *Comput. Statist. Data Anal.*, **51**, 904–917.

Marohn, F. (2000). Testing extreme value models, *Extremes*, **3**, 363–384.

Matthys, G. & Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models, *Statist. Sinica*, **13**, 853–880.

Matthys, G.; Delafosse, M.; Guillou, A. & Beirlant, J. (2004). Estimating catastrophic quantile levels for heavy-tailed distributions, *Insur. Math. Econ.*, **34**, 517–537.

Mises, R. von (1936). La distribution de la plus grande de n valeurs, *Revue Math. Union Interbalcanique*, **1**, 141–160. Reprinted in *Selected Papers of Richard von Mises*, Amer. Math. Soc., **2** (1964), 271–294.

Neves, C. & Fraga Alves, M.I. (2008). Testing extreme value conditions — an overview and recent approaches, *Revstat*, **6**, 83–100.

Oliveira, O.; Gomes, M.I. & Fraga Alves, M.I. (2006). Improvements in the estimation of a heavy tail, *Revstat*, **4**, 81–109.

Pancheva, E. (1992). Multivariate max-semistable distributions, *Th. Probab. Appl.*, **37**, 731–732.

Pancheva, E. (2010). Max-semistability: a survey, *ProbStat Forum*, **3**, 11–24.

Peng, L. (1998). Asymptotically unbiased estimator for the extreme-value index, *Statist. Prob. Lett.*, **38**(2), 107–115.

Peng, L. (2001). Estimating the mean of a heavy-tailed distribution, *Statist. Prob. Lett.*, **52**, 255–264.

Pickands III, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

Raoult, J.P. & Worms, R. (2003). Rate of convergence for the generalized Pareto approximation of the excesses, *Adv. Appl. Probab.*, **35**, 1007–1027.

Reiss, R.-D. & Thomas, M. (2001; 2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 2nd edition; 3rd edition, Birkhäuser Verlag.

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

Schultze, J. & Steinbach, J. (1996). On least squares estimators of an exponential tail coefficient, *Statist. Decis.*, **14**, 353–372.

Segers, J. (2005). Generalized Pickands estimators for the extreme value index, *J. Statist. Plann. Infer.*, **28**, 381–396.

Smith, R.L. (1987b). *Approximations in extreme value theory*, Preprint, Univ. North-Carolina.

Stephenson, A. & Tawn, J. (2004). Bayesian inference for extremes: accounting for the three extremal types, *Extremes*, **7**, 291–307.

Temido, M.G. & Canto e Castro, L. (2003). Max-semistable laws in extremes of stationary random sequences, *Th. Prob. Appl.*, **47**, 365–374.

Themido Pereira, T. (1993). *Second order behavior of domains of attraction and the bias of generalized Pickands' estimator*. In "Extreme Value Theory and Applications III, Proc. Gaithersburg Conference" (J. Lechner, J. Galambos and E. Simiu, Eds.) (NIST special publ.).

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Am. Statist. Assoc.*, **73**, 812–815.

Willems, P.; Guillou, A. & Beirlant, J. (2007). Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying, *J. Hydrology*, **338**, 221–236.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index, *J. Mult. Anal.*, **100**, 794–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index, *J. Mult. Anal.*, **101**, 971–983.

You, A.; Schneider, U.; Guillou, A. & Naveau, P. (2010). Improving extreme quantile estimation by folding observations, *J. Statist. Plann. Infer.*, **140**, 1775–1787.

Yun, S. (2002). On a generalized Pickands estimator of the extreme value index, *J. Statist. Plann. Infer.*, **102**, 389–409.

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 33–60

A REVIEW OF EXTREME VALUE THRESHOLD ESTIMATION AND UNCERTAINTY QUANTIFICATION

Authors: Carl Scarrott
– Mathematics and Statistics Department, University of Canterbury, New Zealand
carl.scarrott@canterbury.ac.nz

Anna MacDonald
– Mathematics and Statistics Department, University of Canterbury, New Zealand
anna.macdonald@pg.canterbury.ac.nz

Abstract:

- The last decade has seen development of a plethora of approaches for threshold estimation in extreme value applications. From a statistical perspective, the threshold is loosely defined such that the population tail can be well approximated by an extreme value model (e.g., the generalised Pareto distribution), obtaining a balance between the bias due to the asymptotic tail approximation and parameter estimation uncertainty due to the inherent sparsity of threshold excess data. This paper reviews recent advances and some traditional approaches, focusing on those that provide quantification of the associated uncertainty on inferences (e.g., return level estimation).

Key-Words:

- *extreme value threshold selection; graphical diagnostics; mixture modelling; rule of thumb; threshold uncertainty.*

AMS Subject Classification:

- 62G32, 62G07, 62G30, 62E20.

1. INTRODUCTION

This paper reviews the key historical threshold estimation approaches for extreme value applications, and the latest developments. The focus is on approaches which provide not only threshold estimation but also uncertainty quantification for the threshold itself and subsequent inferences for quantities like return levels, though we also discuss some exceptions for key developments which provide threshold estimation, but not uncertainty quantification. There is a certain focus on recently developed mixture model type approaches, as these deal naturally with both estimation and formal uncertainty quantification. The aim is to be all encompassing, a near-impossible task, so we apologise for any omissions.

The classical asymptotically motivated model for excesses above a high threshold is the generalised Pareto distribution (GPD). Pickands (1975) and Balkema & de Haan (1974) showed that if there is a non-degenerate limiting distribution for appropriately linearly rescaled excesses of a sequence of independent and identically distributed observations $X_1, \ldots, X_n$ above a threshold u, then the limiting distribution will be a GPD. In applications, the GPD is used as a tail approximation to the population distribution from which a sample of excesses $x - u$ above some suitably high threshold u are observed. The GPD is parameterised by scale and shape parameters $\sigma_u > 0$ and ξ, and can equivalently be specified in terms of threshold excesses $x - u$ or, as here, exceedances $x > u$, as

$$G(x \mid u, \sigma_u, \xi) = \Pr\big(X < x \mid X > u\big) = \begin{cases} 1 - \left[1 + \xi\left(\dfrac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi}, & \xi \neq 0\,, \\ 1 - \exp\left[-\left(\dfrac{x-u}{\sigma_u}\right)\right]_+, & \xi = 0\,, \end{cases} \tag{1.1}$$

where $y_+ = \max(y, 0)$. When $\xi < 0$ there is an upper end point, so $u < x < u - \sigma_u/\xi$. Implicitly underlying the GPD is a third parameter required for estimation of quantities like return levels, the proportion of threshold excesses $\phi_u = \Pr(x > u)$, used to calculate the unconditional survival probability:

$$\Pr\big(X > x\big) = \phi_u\Big[1 - \Pr\big(X < x \mid X > u\big)\Big]\,. \tag{1.2}$$

This representation is often referred to as a Poisson-GPD, as it explicitly accounts for the Poisson rate of excess events. Smith (1989) and Davison & Smith (1990) consider statistical aspects of a Poisson point process (PPP) representation of the classical extreme value models, details of which are not provided for brevity. The principle benefits of the PPP representation is that it can be parameterised in terms of location μ, scale σ and shape ξ which are independent of the threshold, which can simplify extension to nonstationary or random effects type models, and the excess rate ϕ_u is incorporated as a function of these parameters.

The GPD satisfies a threshold stability property: for any higher threshold $v > u$, the subsequent excesses also follow a GPD with the same shape but shifted scale $\sigma_v = \sigma_u + \xi(v - u)$. The subscript u on σ_u makes the threshold dependence explicit, although in the limiting case $\xi = 0$ this disappears. The 'modified scale' reparameterisation $\sigma^* = \sigma_v - \xi v$ is constant above u, i.e., once the GPD provides an adequate tail approximation. The shape and scale parameter can be orthogonalised following Cox & Reid (1987) with the preferred form usually $(\tilde{\sigma}_u, \xi)$, where $\tilde{\sigma}_u = \sigma_u(1 + \xi)$, as the shape is often a key parameter of interest.

Traditionally, the threshold was chosen before fitting, giving the so-called *fixed threshold approach*. Threshold choice involves balancing bias and variance. The threshold must be sufficiently high to ensure that the asymptotics underlying the GPD approximation are reliable, thus reducing the bias. However, the reduced sample size for high thresholds increases the variance of the parameter estimates. Threshold choice is practically equivalent to estimation of the k^{th} upper order statistic $X_{(n-k+1)}$ from the ordered sequence $X_{(1)}, \ldots, X_{(n)}$, called the 'tail fraction' below. Formally, to ensure tail convergence, as $n \to \infty$ the order $k \to \infty$ but at a reduced rate so that $k/n \to 0$ (the so called intermediate sequence of order statistics of Leadbetter *et al.* (1983)), i.e., as the sample size grows, the quantile level of the threshold increases at a faster rate.

2. SUMMARY OF ESTIMATION APPROACHES

The classical fixed threshold modelling approach uses graphical diagnostics, essentially assessing aspects of the model fit, to make an *a priori* threshold choice. Some of the commonly used diagnostics and related statistics are described in §2.1. A benefit of this approach is that it requires practitioners to graphically inspect the data, comprehend their features and assess the model fit, when choosing the threshold. A key drawback with these approaches is they can require substantial expertise and can be rather subjective, as will be seen below. Further, application of this approach when there are many datasets (e.g., different stock returns series in finance applications) is time-consuming. In this situation, it is common for practitioners to assume a constant quantile level across all series, determined by some assessment of fit across all or a subset of the datasets. In some applications the threshold is pre-determined by physical considerations, e.g., government target level for pollution concentrations. Some simple rules of thumb for threshold selection are detailed in §3.

The drawback with fixed threshold approaches is that once the threshold has been chosen it is treated as fixed, so the associated subjectivity and/or uncertainty is ignored in subsequent inferences. Further, it is frequently observed in applications that there is more than one suitable threshold with different inferred

tail behaviours, which will be ignored when fixing the threshold. An informal approach to overcoming these problems is to evaluate the sensitivity of the inferences (e.g., parameters or quantiles) to different threshold choices. There has been a proliferation of new approaches to estimate the threshold more objectively and/or formally account for the threshold uncertainty. In §4 we summarise the large literature on tail fraction estimation including further graphical diagnostics, which mostly use asymptotic optimality-based arguments under various population distribution assumptions, and §5 outlines resampling based approaches which typically require weaker assumptions.

Direct comparison of the GPD likelihood for different thresholds is complicated by the varying sample sizes. Recently, various extreme value mixture models have been developed to overcome this problem. These mixture models typically approximate the entire distribution function, so have a fixed sample size for each threshold considered. Traditionally, the 'bulk' of the distribution below the threshold was ignored, as they were not supported by the tail asymptotics. Further, from a practical viewpoint the extreme and non-extreme events are often caused by different driving forces, so the latter will provide little information about the tails. These mixture models have a rather ad-hoc development, often motivated by their applications or by the underlying properties of the population distribution (e.g., bounds on the support, multi-modality). The guiding principle in their development is to choose a physically sensible model for the bulk distribution, for the application at hand, along with an appropriate tail/threshold model. The most widely applicable mixture models are carefully defined to ensure that the bulk and tail fits are not too influenced by each other. There are a range of mixture models with different assumptions for bulk, tail and threshold components, which have loosely been classified into parametric, semiparametric and nonparametric estimators for the bulk distribution in §6.1–6.3 below.

§7 describes approaches outside these general categories, e.g., robust estimation. A nice short review of a subset of the threshold estimation approaches outlined below is provided by de Zea Bermudez & Kotz (2010).

2.1. Graphical diagnostics

Coles (2001) outlines the common graphical diagnostics for threshold choice:

- Mean residual life (or mean excess) plot;
- Threshold stability plot(s);
- A suite of the usual distribution fit diagnostics (e.g., probability plots, quantile plots, return level plots, empirical and fitted density comparison).

The mean residual life (MRL) plot introduced by Davison & Smith (1990) uses the expectation of the GPD excesses, $E(X-u \mid X>u) = \sigma_u/(1-\xi)$, as a diagnostic, defined for $\xi < 1$ to ensure the mean exists. For any higher $v > u$ the expectation becomes

$$E\big(X-v \mid X>v\big) \;=\; \frac{\sigma_u + \xi v}{(1-\xi)}$$

which is linear in v with gradient $\xi/(1-\xi)$ and intercept $\sigma_u/(1-\xi)$. Examples of the behaviour of the MRL function for various distributions are given by Beirlant *et al.* (2004). Empirical estimates of the sample mean excesses are typically plotted against a range of thresholds, along with Wald type interval estimates, though bootstrap or similar estimates would generally be more appropriate for small tail samples. The threshold is chosen to be the lowest level where all the higher threshold based sample mean excesses are consistent with a straight line, once the sample uncertainty is accounted for. Coles (2001) acknowledges that the interpretation of such plots can be challenging.

Figure 1 gives an example of a MRL plot for the Fort Collins total daily precipitation data from the `extRemes` package in `R` (Gilleland *et al.*, 2010).

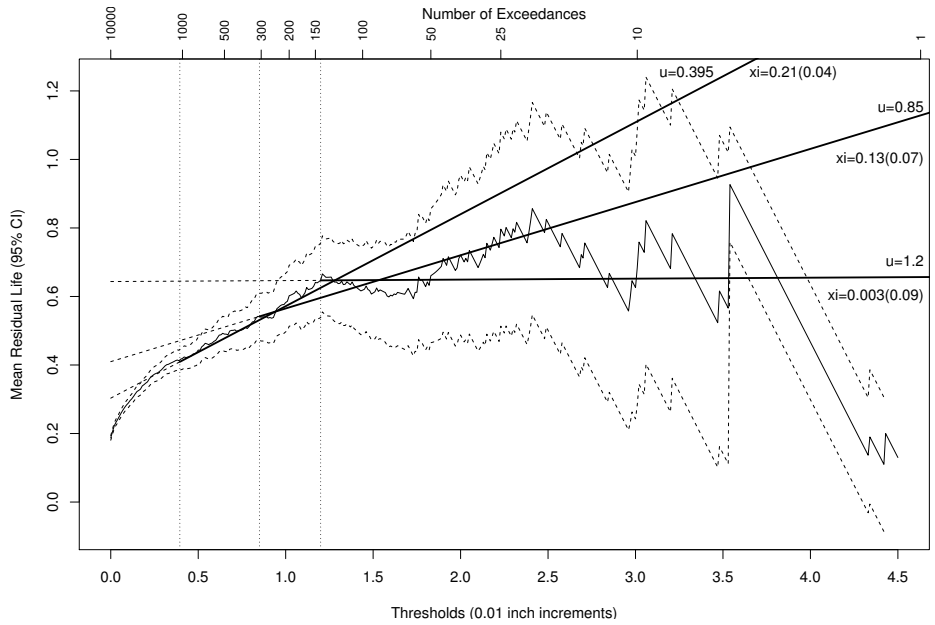


Figure 1: Mean residual life plot for the Fort Collins precipitation data. Solid jagged line is empirical MRL with approximate pointwise Wald 95% confidence intervals as dashed lines. The MRL implied by maximum likelihood (ML) parameter estimates for thresholds $u = 0.395$, 0.85 and 1.2 inches are the upper, middle and lower straight lines respectively. Vertical dashed lines mark these thresholds.

A simplistic analysis of the marginal distribution is considered here, ignoring the obvious seasonality and dependence structure, as we just want to demonstrate the diagnostic plots. The ML estimates of the parameters for three threshold choices provide three fitted MRL straight lines. The increasing variance of the MRL for high thresholds leads to wide confidence intervals, which must be taken into account when assessing the threshold choice.

The threshold of $u = 0.395$ inches suggested by Katz *et al.* (2002) gives over 1000 exceedances, and is justified by the empirical MRL becoming close to linear above this level, and below this level a curved MRL is observed indicating a bias due to the GPD asymptotics breaking down. However, above 1.2 inches strong deviations from the fitted MRL, shown by the upper straight line, are observed with higher thresholds exhibiting a lighter tail. The upper straight line is also close to the pointwise upper interval limit around 1.7 inches (where there is still a reasonable amount of data), and above 2.7 inches the line lies outside the intervals (although the sample sizes here are small so the interval estimates are less reliable). A very different picture arises when considering the highest threshold of $u = 1.2$, with 138 exceedances, which could be justified on the same grounds: above this threshold the MRL is consistent with a straight line (in fact a horizontal line indicating $\xi \approx 0$) and before this level the MRL could be considered to exhibit bias-related curvature. Both these thresholds are consistent with the general guidelines for threshold choice using the MRL plot, but they provide very different tail extrapolations ($\xi = 0.21$ and 0.003 respectively).

The threshold stability plots shown in Figure 2 also do not provide firm conclusions. At a threshold of $u = 0.395$ the shape parameter appears to reach a plateau, compared to lower thresholds which exhibit the bias-related curvature. However, as with the MRL plot, inconsistencies are observed between the estimated shape parameter at this level and higher thresholds around $u = 1.2$ and 2.3–2.4. The shape parameter reaches another plateau around 0.85–1.8 inches (above which the sample variation is too large to make useful inferences), and the shape parameter for the threshold $u = 0.85$ is essentially contained within all the confidence intervals for higher thresholds. The plateau above a threshold of $u = 1.2$ is also strongly indicated by the shape parameter threshold stability plot, despite the confidence interval being much wider at this level. The corresponding modified scale threshold stability plot exhibits some similar features, due to the negative dependence with the shape parameter, but is rather more challenging to interpret for high thresholds due to the large sample variability.

This example demonstrates the substantial subjectivity in interpreting these diagnostic plots, and the resulting uncertainty. Similar challenges are seen with the River Nidd data, shown in Tancredi *et al.* (2006), and many other examples in the literature. These examples suggests that a more 'objective' threshold estimation approach is needed and that uncertainty must be accounted for.

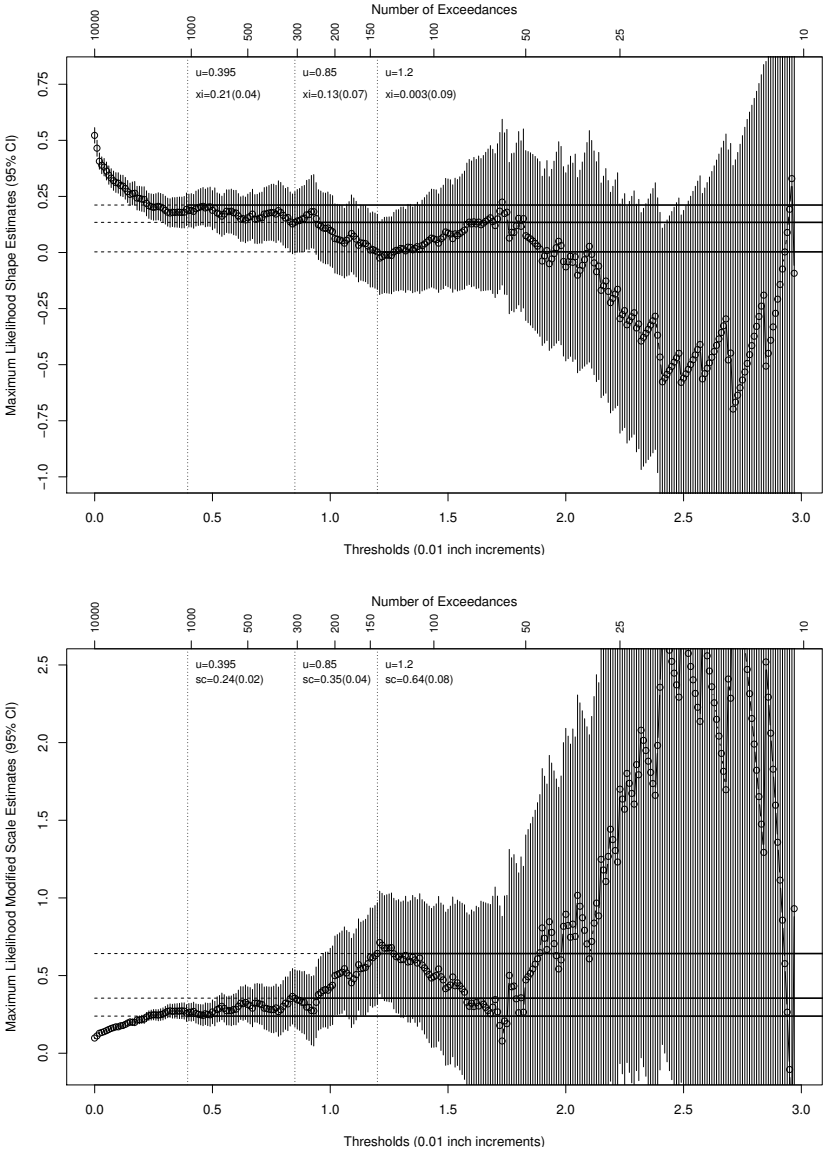


Figure 2: Threshold stability plots for (upper) shape and (lower) modified scale parameters respectively for the Fort Collins precipitation data. Circles are maximum likelihood estimates with vertical lines as approximate pointwise Wald 95% confidence intervals. Three thresholds $u = 0.395$, 0.85 and 1.2 inches are shown by vertical dashed lines with corresponding ML parameter estimates as horizontal lines.

In particular, multiple possible thresholds are indicated in both these examples, so our inference should take this into account. Of course, if the inferences are insensitive to the threshold choice then there is little value to be added from more sophisticated analysis, but in all other cases pre-fixing a single threshold in advance seems inappropriate.

There is a plethora of other diagnostic plots in the literature assessing other features of the model fit at different thresholds, e.g., assessing PPP assumptions or general model goodness of fit statistics considered by Davison & Smith (1990). Under the assumption of Pareto type tails ($\xi > 0$), Hill plots and their many variants, are commonly used, and are discussed in §4.

3. RULES OF THUMB

Leadbetter *et al.* (1983) showed that the threshold sequence (for different sample sizes n), for a population in the domain of attraction of a GPD, is a function of the properties of that distribution. So for a known distribution function F a closed form, or approximation, to the threshold sequence can be derived. For example, a unit exponential population has threshold sequence $u_n = \log(n)$. In the more relevant case of an unknown F there is no general form for the threshold sequence, although some properties of the sequence are known (e.g., order statistic convergence mentioned in §1). Many of the following results follow from fairly general classes of population distributions, such as the Hall (1982) class, which have certain specific tail properties.

Given the general order statistic convergence properties, various rules of thumb have been derived in the literature. Simple fixed quantile rules, like the upper 10% rule of DuMouchel (1983), are inappropriate from a theoretical viewpoint, though frequently used in practice. Ferreira *et al.* (2003), amongst others, use the square root rule $k = \sqrt{n}$ in their simulation study to deterministically specify the tail fraction, which satisfies the intermediate order statistic convergence property in §1, but the source of a formal derivation of this rule is unknown to us. Ho & Wan (2002) and Omran & McKenzie (1999) use the empirically driven rule $k = n^{2/3}/\log[\log(n)]$ proposed by Loretan & Philips (1994).

Reiss & Thomas (2007) heuristically justify choosing the lowest upper order statistic k to minimize

$$\frac{1}{k}\sum_{i=1}^{k} i^{\beta}\left|\hat{\xi}_i - \text{median}(\hat{\xi}_1, \ldots, \hat{\xi}_k)\right| \tag{3.1}$$

where $\hat{\xi}_i$ is an shape parameter estimator for the tail fraction above upper order statistic i and the tuning parameter satisfies $0 \leq \beta \leq 0.5$. Minimising (3.1) searches for the tail fraction where the distribution of these estimated shape parameters stabilises (downweighting the small tail fractions when $\beta \neq 0$) and k is chosen as an estimate of the location of the distribution. In practice, automated implementation is this approach is unreliable for small k (despite the weighting by i^{β}),

so a minimum value of k is usually used. Further, best practice would validate the selection using a graphical summary of (3.1), similar to the threshold stability plot above, in combination with other diagnostics. Reiss & Thomas (2007) point out there are many sensible extensions using alternative distance metrics or weighting schemes and potentially using smoothers when there is limited data. Neves & Alves (2004) investigated refinements to the choice of β.

4. PROBABILISTIC RESULTS

The Hill (1975) estimator is a classic tail index estimator for the Pareto type distribution ($\xi > 0$), which has a power law form with regularly varying tails,

$$1 - F(x) \approx x^{-\alpha} L(x) , \qquad \alpha > 0 , \tag{4.1}$$

where $L(x)$ is a slowly varying function, i.e.,

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1 , \qquad t > 0 ,$$

which allows flexibility in the lower tail but that ensures the power law behavior dominates the upper tail. Clearly, this model does not have such flexible upper tail behaviour as the GPD, but it is an important special case in many applications and since a wide range of techniques has been developed for both tail index and tail fraction estimation, it is worthy of a brief review. Let $X_{(1)}, \ldots, X_{(n)}$ represent the data in ascending order, where we assume that heavy-tailed but negative data are transformed to be positive. The Hill estimator for the tail index $\xi = \alpha^{-1}$ based on the $k+1$ upper order statistics,

$$H_k = \frac{1}{k} \sum_{i=1}^{k} \log X_{(n-i+1)} - \log X_{(n-k)} ,$$

is the ML estimator for relative excesses from a strict Pareto tail, i.e., with a constant for the slowly varying function in (4.1). Beirlant *et al.* (2004, §4.2) outline alternative derivations of H_k.

The results from the Hill estimator are critically dependent on the tail fraction chosen. The Hill plot, explored by Drees *et al.* (2000), is another graphical diagnostic for prior determination of the tail fraction, which plots the Hill estimator for a range of values of k against either k, the tail fraction, or the corresponding threshold. The value of k is chosen as the largest value (i.e., lowest threshold) such that the Hill estimator has stabilised. Despite its rather different formulation, the Hill plot is essentially the MRL plot of the log-transformed data,

so suffers from many of the same benefits and drawbacks, and has been referred to as the Hill horror plot by Resnick (1997). The Hill estimator can exhibit substantial bias if the slowly varying component decays slowly in the limit (Beirlant *et al.*, 2004). The Hill estimator is not shift invariant, although invariant estimators have been proposed by Fraga Alves (2001), Drees (1995, 1998) and Pereira (1994). The latter two issues are not of concern here, as we focus on threshold estimation.

The Hill plot for the Fort Collins data shown in Figure 3 provides a rather different picture of threshold choices than do Figures 1 and 2. The low threshold of $u = 0.395$ inches is suggested to be inappropriate due to the Hill function not having stabilised. The Hill function is rather more stable around the mid-range threshold of $u = 0.85$, but is unstable around the high threshold of $u = 1.2$ inches. So again, the Hill plot gives a different interpretation, thus redemonstrating the expertise required in using these diagnostic plots; see also Drees *et al.* (2000) and references therein. Drees *et al.* (2000) show that the 'altplot' of Resnick & Starica (1997), which is a simple modification to the Hill plot, by essentially using a log-scale on the k-axis, is beneficial for cases where the slowly varying function is non-constant but also state this is not a panacea for threshold determination.

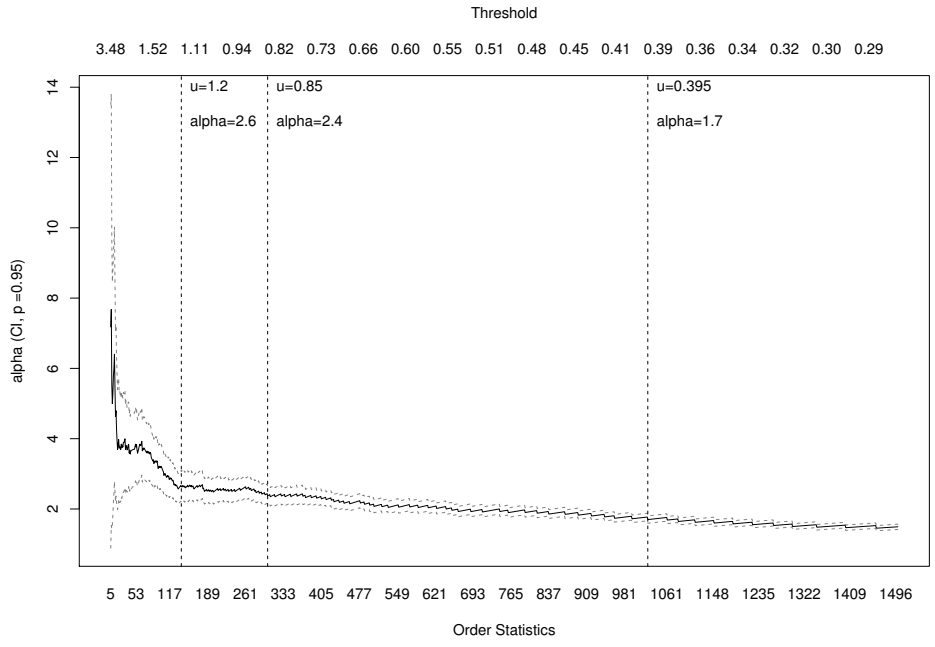


Figure 3: Hill plot of α of regularly varying tail model for the Fort Collins precipitation data as solid line with approximate 95% confidence intervals as dashed lines. The three thresholds $u = 0.395$, 0.85 and 1.2 used in Figure 1 are shown by the vertical dashed lines, with the corresponding Hill estimates.

Hill (1975) also suggested using the fact that the log spacings between the order statistics should be exponentially distributed under the Pareto tail assumption in choosing the tail fraction, by choosing the minimum k such that $\{\log(X_{(n-i+1)}) - \log(X_{(n-k)}) : i = 1, ..., k\}$ does not fail a exponentiality test (e.g., a Kolmogorov–Smirnov or similar test). However, in their work to find an optimal tail fraction Hall & Welsh (1985) showed that this tends to overestimate the tail fraction in large samples. Guillou & Hall (2001) extend the idea of Hill (1975) to derive a plug-in estimator by applying the hypothesis test on an accumulation of the log spacings. Goegebeur *et al.* (2008) further considered a kernel based goodness of fit statistic of the tail fit in the Pareto type tail case, extending Hill (1975), but taking advantage of the relationship between the specific kernel statistics and bias in the asymptotic mean square error of the Hill estimator. Pickands (1975) suggested choosing the tail fraction for the more general GPD case by minimising the distance between empirical and GPD distribution function estimators, with the latter using the Pickands' estimator of the parameters. This idea was extended by Gonzalo & Olmo (2004) using a weighted distance measure, which includes the Pickands' measure and the Kolmogorov–Smirnov statistic as special cases.

Various authors have investigated automated approaches to determining the tail fraction by, for example, minimising the mean square error of estimators of properties of the tail distribution, such as the tail index (Beirlant *et al.*, 1999), the quantiles (Ferreira *et al.*, 2003), or the tail probabilities (Hall & Weissman, 1997), for which optimal asymptotic results typically require second order assumptions in addition to (4.1), see Beirlant *et al.* (2004, §4.7). For example, the plug-in estimator for the tail fraction of Hall (1982) requires prior knowledge of the parameters of the restricted Hall class of tail behaviours. Dekkers (1993) considered an extension of the Hall (1982) estimator to the GPD case using an additional moment-based estimator incorporating the usual Hill estimator, allowing $\xi \leq 0$. Under assumed constraints on these parameters, Hall & Welsh (1985) were able to find a simple adaptive plug-in estimator for the tail fraction, but, the parameters are rather difficult to estimate. The approach of Feuerverger & Hall (1999) suffers similarly from the prescriptive form of the second order characteristics.

Even if the assumptions underlying these approaches are appropriate, their major drawback is that they do not account for the threshold uncertainty on subsequent inferences. Moreover the finite sample properties of such estimators are not well understood. The need to estimate parameters of the unknown population distribution F prior to determination of the optimal tail fraction (e.g., tail index or second order characteristics) has led to the development of bootstrap approaches or algorithmic estimators. As noted by Drees *et al.* (2000), both replace the problem of threshold determination with that of other characteristics — bootstrap sample sizes or stopping rules, respectively.

5. COMPUTATIONAL APPROACHES

Hall (1990) first proposed a resampling based method for estimation of the (asymptotically) optimal tail fraction, by minimising the mean square error (MSE) of the Hill estimator. This approach was applied to the Hall class of Pareto type tails, with a further restriction on the power law parameter of the slowly varying component. A sub-sampling approach is used for overcoming the substantial bias observed when using a standard bootstrap, but it requires an initial estimate of the bias. Gomes & Oliveira (2001) showed the entire procedure is rather sensitive to this estimate. The bias correction uses the relationship between the bootstrap on the full and sub-sample to quantify the bias.

Danielsson *et al.* (2001) overcome much of the restrictiveness of the plug-in estimator of Hall (1990) using a two stage bootstrap procedure for tail index estimation, by minimisation of the asymptotic MSE criterion, which includes automated sample fraction estimation. They avoid the need to know the second order parameter, which is consistently estimated as a consequence of their bootstrap procedure. Two bootstrap estimates of the tail fraction to minimise the MSE of an auxiliary statistic based on the Hill estimator are estimated using two different sample sizes of a particular form. These two estimates are then combined in a closed form to determine the asymptotically optimal tail fraction, and consequently the tail index. The only tuning parameters are the first stage bootstrap resample size and the number of bootstrap samples, the latter being essentially determined by computational capacity. The choice of tuning parameter for the bootstrap sample size at the first stage is in some sense automated by minimisation of a diagnostic criterion over a grid of suitable values. Further, Gomes & Oliveira (2001) have shown that the estimates are robust to the choice of resample size. As this approach is based around the Hill estimator it is restricted to positive shape parameter ($\xi > 0$), though Draisma *et al.* (1990) have considered extension to the GPD case.

Ferreira *et al.* (2003) developed a bootstrap procedure extending that of Danielsson *et al.* (2001), towards optimal estimation of high quantiles (or the upper end-point) by minimisation of the asymptotic MSE. Unlike the similar bootstrap procedure by Hall & Weissman (1997), which is designed for tail probability estimation, they do not require knowledge of the parameters or of second order properties of the distribution for appropriate choice of the bootstrap tuning parameters. They note that its performance is not satisfactory for sample sizes below 2000.

Beirlant *et al.* (1996) consider choosing the tail fraction to provide an optimal linear fit to the Pareto quantile plot (Beirlant *et al.*, 2004, §4.2), which uses the fact that the log transformed Pareto variables are exponential distributed.

They suggest using optimal weights in the regression which require knowledge of the second order parameter, but, estimation of these can be combined in an iterative procedure. However, as with all these approaches, after selection of the appropriate threshold there is no formal assessment of the uncertainty associated with the threshold choice, though some progress has been made by Caers & Dyck (1999) in extending the Hall (1990) approach to account for the tail fraction uncertainty in Pareto power law parameter.

Drees & Kaufmann (1998) provide a hierarchical algorithm to determine the optimal tail fraction for the Hall class. However, this procedure requires prior knowledge of the tail index and second order characteristics (power law of slowly varying function), for which they provide heuristically defined suggestions as part of their simulation study. Beirlant *et al.* (2004, §4.7) reviews comparative studies between this algorithmic, bootstrap based and adaptive Hill based estimators of the tail fraction. In general the restrictive assumptions underlying these approaches hinder their wide applicability.

6. MIXTURE MODELS

The main drawback of most of the aforementioned threshold estimation approaches is that they do not account for the uncertainty associated with the threshold choice. In the last decade, extreme value mixture models have been proposed which encapsulate the usual threshold model in combination with a component intended to capture (some or) all of the non-extreme distribution (henceforth called the 'bulk distribution'). The motivation for ignoring the non-extremal data in early statistical developments in extreme value modelling was mainly that:

1. Extreme and non-extreme events are often physically caused by different underlying processes, which implies there is little information in the bulk distribution for describing the tail behaviour;
2. The GPD is a flexible asymptotically justifiable model for the tail excesses, but classical probability models which combine flexible tail and bulk behaviours are hard to come by and are often application specific;
3. The information content in the sample is typically spread between the high density 'low information' bulk distribution and the low density but 'highly informative' tail observations. The balance between these in terms of influence on the parameters (or related quantities like return levels) is strongly dependent on the model and estimation method. Therefore, when including non-extreme data one has to be careful that the model/estimator provide sufficient relative importance to the tail versus the bulk fit.

The general principle with the mixture models is to combine the flexible threshold model (e.g., GPD), with a suitably flexible and/or application appropriate model of the bulk distribution. The threshold is either implicitly or explicitly defined as a parameter to be automatically estimated, and in most cases the uncertainty associated with the threshold choice can be accounted for naturally in the inferences. In this way, appropriate tail fits can be achieved whilst allowing automated threshold estimation and, provided the bulk distribution model is sufficiently flexible, the bulk and tail fit should not strongly influence each other.

The major drawback of such models is their ad-hoc heuristic definitions, the asymptotic properties of which are still little understood. They have also not had time to be well used in practice and currently there is no readily available software implementation to allow practitioners to gain wider experience. Arguably, the biggest danger with using these models is ensuring that the bulk and tail fits are fairly robust to each other. They cannot be fully disjoint, as they share information, at least about the threshold location, thus ruling out simple application of the EM algorithm. However, robustness of the tail fit to that of the bulk distribution is clearly of major concern. Some discussion of their robustness in this regard will be given, with some intuition, though a fuller comparative study is needed. Another concern with these models is their behaviour at the threshold (e.g., is the fitted density continuous?) and whether attempts to resolve such issues have been successful or have created other problems.

The available mixture models have been broadly classified by the type of bulk distribution models: fully parametric, semiparametric and nonparametric. A general description of the model and estimation approaches as proposed in their original papers will be given, followed by discussion and comparison of their properties and areas for future developments. Many of the common features of the mixture models are summarised in §6.1.

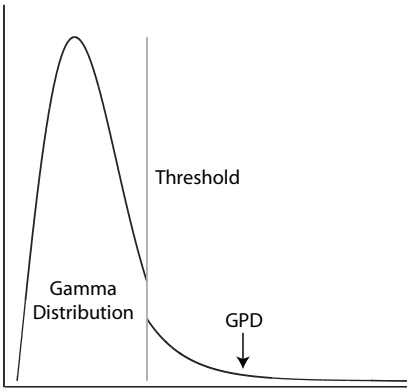
6.1. Parametric bulk models

One of the simplest extreme value mixture models is the spliced distribution consisting of two parametric components:

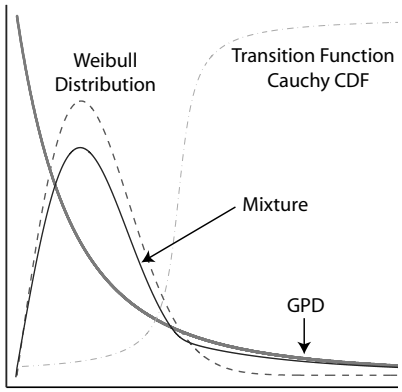
- A parametric model for the bulk distribution below the threshold — gamma in Behrens *et al.* (2004) and normal in Carreau & Bengio (2009a);
- A threshold tail model (GPD in these proposals) above the threshold.

These components are spliced together at the threshold, which is treated as a parameter to be estimated. Behrens *et al.* (2004) also mooted inclusion of other parametric, semiparametric and nonparametric possibilities below the threshold, but these were not developed further.

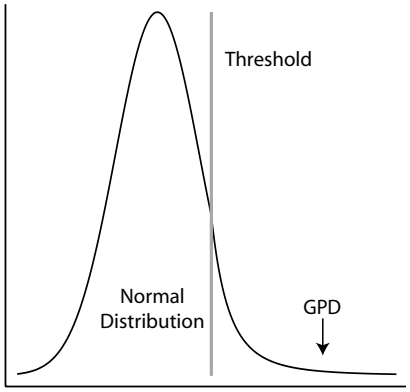
The normal spliced with GPD tail developed by Carreau & Bengio (2009a) shown in Figure 4.3, named the 'hybrid Pareto' model, was further developed to include constraints on the parameters to ensure continuity up to the first derivative of the density. However, due to its poor performance in practice, this model was extended to a mixture of hybrid Pareto distributions, and is discussed in §6.2.



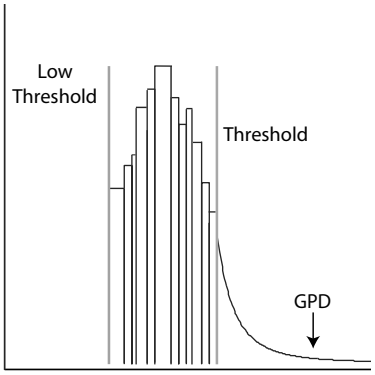
1. Behrens *et al.* (2004)



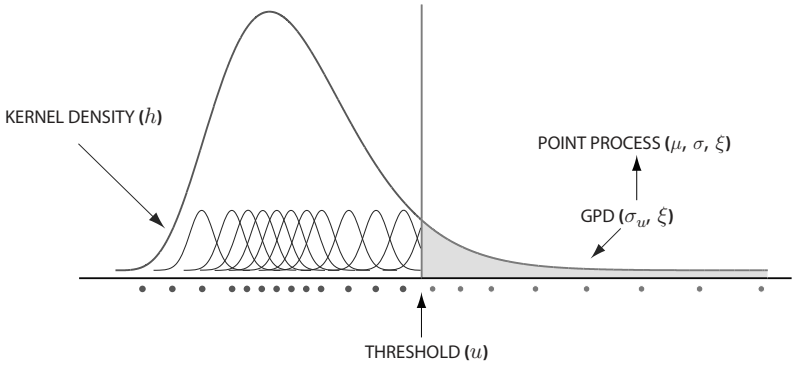
2. Frigessi *et al.* (2003)



3. Carreau & Bengio (2009a)



4. Tancredi *et al.* (2006)



5. MacDonald *et al.* (2011a)

Figure 4: Schematic representations of some mixture models in the literature.

Figure 4.1 illustrates the Behrens *et al.* (2004) model. The discontinuity at the threshold highlights the lack of continuity constraint. This model was specifically designed and evaluated for distributions with a notable discontinuity at the threshold, the relevance of which for applications is debatable. They also explicitly mention that in the case of a smooth transition at the threshold this type of model (and more importantly the estimation procedure) struggles. Some evaluation of performance in the latter situation was reported by Behrens *et al.* (2002). In the case of a strong discontinuity this would lead to a discernible kink in mean excess function at the threshold, so it would be easy to choose a threshold (with little uncertainty) using traditional graphical diagnostics. In the more realistic case of a smooth transition, the traditional diagnostics would also be harder to interpret.

Bayesian inference was used by Behrens *et al.* (2004) with sensible prior forms for the bulk, tail and threshold parameters. However, posterior sampling for their approach is rather inefficient, as they have not accounted for the GPD scale dependence on the threshold. Further, they treat the threshold and tail parameters as independent in the prior. These problems could be easily overcome by suitable transformation of the GPD scale parameter or use of the PPP representation of Smith (1989). The lack of independence of the threshold and GPD scale parameters also makes comprehension of posterior statistics challenging. Orthogonalisation of the GPD scale and shape parameters, or an adaptive posterior sampling scheme (Roberts & Rosenthal, 2009), may help with convergence. Cabras & Castellanos (2010) carried out a comparative study of this simple mixture model and a semiparametric alternative, discussed below.

The major benefits and drawbacks of this parametric approach compared to the semi- and nonparametric alternatives in §§6.2, 6.3, are common to other modelling situations, so are not discussed in detail. In short, if the parametric model is 'correct' then the parametric approach will usually provide the most efficient inferences, but it will suffer when the model is misspecified. There are currently no published results on the performance under model misspecification, and in particular the robustness of the tail fit. In general, the coarse split of information from the bulk and tail data will afford some robustness. However, if the bulk fit is poor, then this will influence the location of the threshold, which will impact the tail fit.

The most beneficial property of this mixture model approach is to provide an objective estimate of a suitable threshold, that provides the best fit to the data (according to the fitting metric used). However, the lack of a continuity (or higher order) constraint at the threshold causes a specific issue with this type of model and for similar approaches described below. The threshold adds an extra degree of freedom, which has a strong localised effect on the fitted distribution function; see Figure 6 of MacDonald *et al.* (2011a). Therefore, if the upper tail of the sample density has spurious peaks or troughs, due to natural sample variability, then

the threshold will often be drawn to those locations. This feature is frequently seen in the posterior threshold samples (e.g., modes where sharp changes in the sample density are observed) for this model and others below. However, this type of feature also causes kinks in the MRL and deviations in the threshold stability/Hill plots, so is actually capturing what practitioners would interpret from these traditional graphical diagnostics as well.

A further beneficial property of these mixture models when using Bayesian inference, or sampling based frequentist approaches, is that the entire parameter posterior distribution is available. It is frequently observed in real life applications from the traditional graphical diagnostics that multiple suitable thresholds are plausible, which is naturally accounted for in the inference for these mixture models. The multiple threshold choices show themselves as a bi- or multi-modal posteriors for the threshold, and sometimes other tail parameters; see Scarrott & MacDonald (2010), Figure 4 of Behrens *et al.* (2004) and Figure 2b of Tancredi *et al.* (2006) for example. Computational Bayesian inference approaches are particularly beneficial in this situation to avoid the optimisation challenges associated with likelihood inference when there are multiple modes.

Posterior predictive inference (PPI) is often appropriate in applications where the return levels for future events are of interest; see Coles & Powell (1996). For such mixture models, PPI has a secondary benefit. Even if the underlying density model is potentially discontinuous at the threshold, the PPI will integrate over all posterior threshold possibilities and often provides practically continuous density estimates. See further discussion and examples in do Nascimento *et al.* (2011) and MacDonald *et al.* (2011a).

The first approach in the literature which attempts to give a continuous transition between the bulk and tail models, but unfortunately can fail in applications, is provided by Frigessi *et al.* (2003); see Figure 4.2. Their model is appropriate when there is a lower bound on the support, if the upper tail is of interest (or vice versa). Without loss of generality we assume a zero lower bound. The GPD tail is defined over the whole range of support. The bulk distribution is also defined over the whole range of support, but is presumed to have a light upper tail (e.g., a Weibull tail). A dynamic weight function, such as the cumulative distribution function (cdf) of some smooth unimodal distribution, is then applied to these two components, with highest weight given to the bulk distribution function at low ranges in the support and high weight given to the GPD in the upper tail. A normalization constant ensures that the density has unit integral. The weight function allows the bulk model to dominate the lower tail and the GPD to dominate the upper tail (especially as the bulk model has a light upper tail), but permits a smooth transition. There is no explicit threshold, but the threshold could be estimated by the point at which the relative contribution of the weighted bulk model is sufficiently small compared to the weighted GPD

tail model. Frigessi *et al.* (2003) point out that if the transition function is the Heaviside function, then this model includes the Behrens *et al.* (2004) mixture model as a special case.

Frigessi *et al.* (2003) use the Weibull for the bulk model, the GPD for tail and the location-scale Cauchy cdf for the transition function, thus giving a six parameter model. ML estimation is used throughout. However, parameter identifiability can be challenging due to multiple modes, which cause problems for simple black box optimisation schemes. More sophisticated optimisation schemes (e.g., allowing multiple starting points) would likely be beneficial, or alternatively a Bayesian inference MCMC based sampling scheme would be easily implemented. Vrac & Naveau (2007) apply this mixture model to rainfall-runoff modelling.

The idea of a smooth transition is sensible, but two problems occur in practice (MacDonald, 2012). Consider the Cauchy cdf, which has a single parameter controlling the spread of the transition from the bulk to tail model. A quick but smooth transition (approximating the Heaviside function in the previous model) is achieved when the Cauchy scale is almost zero. Therefore, the supposed smooth transition can be lost in application, as the 'localised degree of freedom' of the threshold discussed on page 49 rears its ugly head again. Vrac & Naveau (2007) also noted a fast transition in some applications.

Another problem is the lack of robustness of the tail fit to that of the bulk. Although the weight function controls the relative contributions, because all three components of the model are defined over the whole range of support they all potentially contribute to the fit over entire range. In particular, the GPD reaches its pole at zero, so even though the weight applied to the GPD may be low close to the lower bound if it is non-zero then the relative contribution of the GPD to the lower tail fit can be high compared to the bulk model. Thus, the lower tail fit can impact on the upper tail fit, which is undesirable. Further, the fit in the upper tail is also impacted by the GPD and bulk models as, although at asymptotic levels the GPD will dominate, at sub-asymptotic levels of typically of interest in applications both the Weibull and GPD may contribute to the tail. Thus, the fit to the bulk can again affect the tail fit. This feature can be particularly problematic when there is an exponential or short upper tail.

As this model includes the Behrens *et al.* (2004) model as a special case, most of the benefits and drawbacks carry over. Of particular note is the threshold and GPD scale dependence, which complicates the inference.

Zhao *et al.* (2010) and Mendes & Lopes (2004) introduce a variant on the Behrens *et al.* (2004) mixture model. They propose a mixture with a normal distribution (as an example) for the bulk, with both tails represented by separate threshold models (a so called two-tail model). Neither consider the threshold dependence of the GPD scale parameter.

Zhao *et al.* (2010) apply their model to financial applications where both the gain and loss risks are of interest. The upper and lower thresholds are estimated in tandem with the other parameters in a Bayesian framework. Thus their approach, like that of Behrens *et al.* (2004), allows for automated threshold choice and uncertainty quantification. Zhao *et al.* (2011) and Zhao (2010) consider the testing of asymmetry of the gain and losses tails, by comparing the model fit between a general parameterisation of both tails and the two tails having a common shape parameter.

Prior to application of their two tail model Mendes & Lopes (2004) robustly standardise the data to make 'well defined' tails, though it is unclear what is meant by this phrase, as they apply the same linear rescaling to all the observations, thus not providing any separation of the bulk and tail data. A robust nonlinear transformation which pushes out the tail data and shrinks in the bulk observations would be needed to achieve this goal. After standardising the data Mendes & Lopes (2004) propose a rather lengthy estimation procedure:

1. Select a grid of candidate tail fractions for both tails, and estimate the corresponding empirical threshold (quantile) estimates);
2. Fit a standard normal or Student-t distribution to the robustified data and use it to find robust estimates of the two tail fractions;
3. Use L-moment estimators to robustly fit the GPD to both tails;
4. Calculate the likelihood for the combined two-tail mixture model with these robust estimates of the tail fractions, thresholds and GPD parameters;
5. The final selection of the tail fractions is that which maximises the likelihood.

This procedure provides an objective automated threshold estimate, but the threshold uncertainty is ignored and it is not obvious how to evaluate the overall uncertainty.

6.2. Semiparametric bulk models

Carreau & Bengio (2009a) propose the hybrid Pareto model depicted in Figure 4.3, which constrains the bulk and tail densities to have continuous zeroth and first derivatives to ensure a smooth transition at the threshold. The five parameters are thus reduced to three. Carreau & Bengio (2009a) chose to vary the mean and location of the normal and the GPD shape parameter, from which the threshold and GPD scale can be derived.

The fundamental idea underlying the hybrid Pareto to ensure smoothness at the threshold is sensible. Unfortunately, the model performs poorly in practice,

seemingly because these two constraints are rather strong, as noted by Carreau & Bengio (2009a). Exploration of the single constraint of continuity would confirm this. Carreau & Bengio (2009a) used a finite mixture of these hybrid Pareto densities as their final model to overcome this. This gives a distribution of thresholds across the mixture components. Further, the tail is approximated by multiple GPD's. Asymptotically the mixture component with the heaviest tail behaviour will dominate, but the sub-asymptotic behaviour is influenced by all the tail components. Carreau & Bengio (2009a) suggest the threshold for the dominant tail component could be used as a proxy for the threshold choice if this is of interest.

This model bypasses the threshold estimation problem, but it has been included because the final estimates in some sense encapsulate the uncertainty associated with the threshold estimation. Although Carreau & Bengio (2009a) call their mixture of hybrid Pareto's model a nonparametric model, it is categorised here as a semiparametric approach as the number of degrees of freedom used in estimation will generally be small, growing slowly with the sample size, compared to the nonparametric mixtures outlined below.

Maximum likelihood estimation is used throughout, including for the number of mixture components, although a neural network learning approach is implemented by Carreau & Bengio (2009b) and Carreau *et al.* (2009) in an extension to nonstationary and bivariate modelling situations. Carreau & Bengio (2009a) point out the the interrelation between the parameters means that the standard EM algorithm cannot be used, as also indicated by Behrens *et al.* (2004) and MacDonald *et al.* (2011a) for related mixture models.

The performance of the mixture of hybrid Pareto models in situations where the tail behaviour is exponential or short tailed needs exploration. In these cases, the normal bulk model in each component of the mixture could adequately approximate the tail, so the implied threshold would be rather variable and it is unclear how this would affect the inferences.

A major benefit of the mixture of hybrid Pareto's model compared to the other mixtures is the required continuity in the zeroth and first derivatives at the threshold. This will reduce the local degree of freedom problem of the threshold in the other mixtures mentioned above, but will introduce greater flexibility, so it is not clear whether the uncertainty associated with this will be reduced, particularly in the mixture of hybrid Paretos. A drawback with this mixture is the parameter dependence, which means there may be a lack of robustness of the tail fit to that of the bulk, as with the dynamically weighted mixture of Frigessi *et al.* (2003). However, the flexibility may override this lack of robustness, as with the fully nonparametric mixture models of Tancredi *et al.* (2006) and MacDonald *et al.* (2011a) outlined below. It would be interesting to investigate both of these issues in future research.

Cabras & Castellanos (2010) consider another semiparametric bulk model spliced with a GPD upper tail. The bulk distribution is approximated by an equi-spaced binning of the data, followed by fitting a Poisson log-link GLM to the counts, with a polynomial smoother for the mean parameter. A profile likelihood for all the observations is obtained by combining the GPD tail model for observations above the threshold and Poisson GLM density estimator below the threshold. However, the fitted density for the bulk distribution is assumed to be fixed, so the likelihood is maximised with respect to only the threshold and GPD parameters. Appropriate scalings are applied to ensure that the density integrates to unity. Bayesian inference is used, but the threshold dependence of the GPD scale parameter is ignored.

do Nascimento *et al.* (2011) extended the Behrens *et al.* (2004) model by defining the bulk distribution as a weighted mixture of gamma densities. They use Bayesian inference, conditional on a fixed number of the gamma components, which is chosen using a BIC- or DIC-based statistic. This ensures that specific parametric forms or constraints such as unimodality are not imposed, yielding a flexible model for the bulk distribution. This mixture model is unlike that of the hybrid model described by Carreau & Bengio (2009a), as it relies on a single GPD for tail estimate and thus requires that only one threshold be estimated. do Nascimento *et al.* (2011) also showed that the use of posterior predictive inference practically eliminates any discontinuity at the threshold, even though the individual posterior samples will likely exhibit a discontinuity.

The major benefit of these semiparametric mixtures over the parametric ones is that they provide reasonably flexible models for the bulk without using up as many degrees of freedom as the nonparametric alternatives discussed below. Thus they combine some of the benefits of both these approaches. Of course, it is not clear whether they provide sufficient flexibility to be robust to model misspecification, particularly robustness of the tail fit to the bulk.

6.3. Nonparametric bulk models

Tancredi *et al.* (2006) were the first to propose an extreme value mixture model combining a nonparametric estimator for the bulk distribution spliced with an extreme value tail model. They were also the first to overcome the dependence of the GPD scale parameter on the threshold by using the PP representation for the tail excesses. Their bulk model is the mixture of uniforms density estimator of Robert (1998) and Robert & Casella (2010), providing a piecewise linear approximation to the cdf below the threshold; see Figure 4.4. The nonparametric mixture is defined between the upper threshold and a lower threshold that is definitely too low (which could be the lower bound on the range of support).

A benefit of this approach is that it fixes the sample size being included in the inference, but potentially excludes data which is uninformative about the tail behaviour.

Bayesian inference using MCMC is implemented. A hierarchy determines whether the uniform widths should be the same or should be allowed to vary and to what degree. However, as there is an unknown number of uniform densities, the parameter space varies in dimension and a reversible jump algorithm is required (Robert & Casella, 2010), whose implementation can be challenging.

MacDonald *et al.* (2011a) use a kernel density estimator for the bulk distribution, spliced at the threshold with a PPP tail model shown in Figure 4.5, following a cruder version with a GPD tail model (Scarrott & MacDonald, 2010). Nonparametric kernel density estimators using symmetric kernels work well with populations with unbounded support, or at least a proper tail before the lower boundary. For populations with bounded support, a boundary-corrected kernel density estimator is used as an alternative by MacDonald *et al.* (2011b). The kernel density estimator assumes a particular kernel such as the normal density, which is centred at each datapoint, and is parameterised by a single bandwidth. Such kernel density estimators can approximate most smooth densities (Silverman, 1986). If the lower tail is heavy then MacDonald *et al.* (2011a) also showed that a mixture model, with both tails replaced by threshold models, can be used to provide flexibility and robustness of the fits between the two tails and the bulk.

MacDonald *et al.* (2011a) utilise standard cross-validation likelihood to choose the bandwidth, combined with the likelihood for PPP tail model, including the threshold, to give a full likelihood for all the observations. The combined likelihood is used in a Bayesian inference framework, with posterior predictive inference used for all the key quantities of interest.

The major benefit of these nonparametric approaches as compared to the parametric approaches is that the tail fit is robust to the bulk fit, as demonstrated by simulation in Tancredi *et al.* (2006). MacDonald *et al.* (2011b) and MacDonald (2012) use sensitivity curves to show the robustness of the tail fit to that of the bulk and vice-versa. The main drawback with the mixture of uniforms for the bulk distribution of Tancredi *et al.* (2006) is the computational complexity, noted by Thompson *et al.* (2009), and in particular the difficulty of ensuring convergence. The nonparametric kernel density estimator is computationally simpler, but the cross-validation likelihood can be burdensome for large samples, and the overall computing time can be reduced by other forms of subsampling or by using an alternative penalty function. Further, both nonparametric mixture models could be fit in two stages: (1) fit the nonparametric density estimator, followed by (2) put this in a combined likelihood with the tail model and carry out the tail inferences ignoring the uncertainty associated with the bulk parameter estimation, as in the profile likelihood approach of Cabras & Castellanos (2010).

The lack of sensitivity of the bulk parameter estimation to the tail parameters would mean that little would be lost while substantially reducing the computational burden.

7. OTHER APPROACHES

Dupuis (2000) proposed a robust procedure for GPD fitting, including statistics to guide threshold choice. The optimal bias robust estimation (OBRE) procedure weights each observation between 0 and 1 depending on how consistent they are with the GPD model, with high weights indicating a good fit. A test statistic (or p-value) for the weights under the null hypothesis that the GPD is the correct model is also provided, along with average weights from simulations under the fitted models for further guidance. She suggests trying out a sensible set of thresholds and choose the lowest threshold such that the weights are all 'sufficiently close to one'.

This procedure essentially replaces the threshold selection problem with that of selecting the tuning parameter which controls the cut-off of closeness to unity. Dupuis (2000) suggested that this new threshold choice is easier to automate. The OBRE also requires specification of the bound on the influence function, which balances efficiency and robustness of the estimator (too high a bound leads to a lack of robustness but higher efficiency as more information is utilised, and vice versa). Some guidance is required to provide a suitable choice for this bound, but this adds to concerns whether this this approach can be fully automated. The principle drawback with this approach is that after choosing the threshold, the OBRE procedure reduces to a fixed threshold approach.

Thompson *et al.* (2009) recently developed an automated procedure for threshold estimation and uncertainty quantification. They set a uniformly spaced grid of possible threshold values (between the median and 98% empirical quantile). For each potential threshold the GPD is fitted (using ML estimation) and the differences in the modified scale parameters for neighbouring thresholds is calculated. They assume asymptotic normality of these scale differences, and use the fact that they are centred around mean zero if the GPD is the correct model. They pool these differences and treat them as a sample of normal random variables. A forward selection procedure (increasing the threshold increment by increment) is then applied, until the Pearson χ^2 test under the assumption of normality and mean zero is not rejected. At this point the threshold is said to be consistent with a GPD. A simple bootstrap approach is used to quantify the threshold uncertainty on the final inferences.

de Zea Bermudez *et al.* (2001) use a Bayesian hierarchical model for tail estimation, essentially averaging over all possible values of tail fraction. The GPD is used to describe the tail excesses conditional on the tail fraction, with the prior distribution defined at the next level of the hierarchy. The tail fraction 'parameter' k at the next level in the hierarchy is described by a doubly truncated Poisson distribution with parameters for the mean and truncation points chosen to be relatively uninformative, but still ensuring a suitable tail fraction. The next hierarchy level describes the priors for the three tail fraction parameters, followed by their hyperparameters at the final level. They consider only the Gumbel case (exponential upper tail). In order to avoid the specification of the bulk model, as used in the extreme value mixture models outlined in §6, they use an approximate posterior, assuming that the contribution to the likelihood from the lower order statistics can be safely ignored and only the conditional contributions for the upper tail fraction contribute. Therefore, for each random effect for the tail fraction k in the hierarchy, there is a differing amount of data contributing to the approximate likelihood (and therefore posterior). The properties of such an estimator (e.g., asymptotics, relative weighting of data points) are unclear.

Cabras & Morales (2007) propose using a sequential outlier detection method to identify points above (below) some upper (lower) threshold which appear to differ in character from the tail(s) of some assumed known parametric model for the population distribution. The partial posterior predictive distribution is used to sequentially estimate the likelihood of observing the pairs of the most outlying (upper and lower) order statistics under the assumed population model, ignoring those that have already been dropped. This procedure continues until the probability of observing both in the pair is sufficiently high. The thresholds are then defined by the last upper and lower order statistics that were dropped. Of course, this requires specification of the outlyingness tolerance, which the user must specify *a priori*, though the authors claim this will be straightforward to elicit. In some ways this approach is related to the two-tail models considered by Zhao *et al.* (2010) and Mendes & Lopes (2004) discussed above. No threshold uncertainty quantification is applied.

ACKNOWLEDGMENTS

Anna MacDonald received support from both a New Zealand TEC Top Achievers Doctoral Scholarship under the Bright Futures Scheme and a Canterbury Scholarship from the University of Canterbury. We wish to thank the editors for the invitation to prepare this review. Much of this paper was written whilst Carl Scarrott was on sabbatical at Lancaster University, UK, whose support is acknowledged, particularly discussions with Jonathan Tawn and Jennifer Wadsworth.

REFERENCES

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

Behrens, C.N.; Lopes, H.F. & Gamerman, D. (2002). *Bayesian analysis of extreme events with threshold estimation*, Technical report, Laboratório de Estatística, Universidade Federal do Rio de Janeiro.

Behrens, C.N.; Lopes, H.F. & Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation, *Statist. Mod.*, **4**, 227–244.

Beirlant, J.; Dierckx, G.; Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

Beirlant, J.; Goegebeur, Y.; Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, London.

Beirlant, J.; Vynckier, P. & Teugels, J.L. (1996). Tail index estimation, Pareto quantile plots and regression diagnostics, *J. Am. Statist. Assoc.*, **91**, 1659–1667.

Cabras, S. & Castellanos, M.E. (2010). *An objective Bayesian approach for threshold estimation in the peaks over the threshold model*, Technical Report TR2010.10, Análisis de riesgo.

Cabras, S. & Morales, J. (2007). Extreme value analysis within a parametric outlier detection framework, *Appl. Stoch. Mod. Bus. Ind.*, **23**, 157–164.

Caers, J. & Dyck, J.V. (1999). Nonparametric tail estimation using a double bootstrap method, *Comp. Statist. Data Anal.*, **29**, 191–211.

Carreau, J. & Bengio, Y. (2009a). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case, *Extremes*, **12**, 53–76.

Carreau, J. & Bengio, Y. (2009b). A hybrid Pareto mixture for conditional asymmetric fat-tailed distributions, *IEEE Trans. Neur. Net.*, **20**, 1087–1101.

Carreau, J.; Naveau, P. & Sauquet, E. (2009). A statistical rainfall-runoff mixture model with heavy-tailed components, *Water Resourc. Res.*, **45**(W10437).

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

Coles, S.G. & Powell, E.A. (1996). Bayesian methods in extreme value modelling: A review and new developments, *Int. Statist. Rev.*, **64**, 119–136.

Cox, D. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *J. R. Statist. Soc. B*, **49**, 1–39.

Danielsson, J.; de Haan, L.; Peng, L. & de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation, *J. Mult. Anal.*, **76**, 226–248.

Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds (with discussion), *J. R. Statist. Soc. B*, **52**, 237–254.

de Zea Bermudez, P.; Amaral Turkman, M.A. & Turkman, K.F. (2001). A predictive approach to tail probability estimation, *Extremes*, **4**, 295–314.

de Zea Bermudez, P. & Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution – Part II, *J. Statist. Plan Inf.*, **140**, 1374–1388.

Dekkers, A.L.M. (1993). Optimal choice of sample fraction in extreme-value estimation, *J. Mult. Anal.*, **47**, 173–195.

do Nascimento, F.F.; Gamerman, D. & Lopes, H.F. (2011). A semiparametric Bayesian approach to extreme value estimation, *Statist. Comput.* pp. Early View: 10.1007/s11222–011–9270–z.

Draisma, G.; de Haan, L.; Peng, L. & Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes*, **2**, 367–404.

Drees, H. (1995). Refined Pickands estimators of the extremal index, *Ann. Statist.*, **23**, 2059–2080.

Drees, H. (1998). Optimal rates of convergence for estimates of the extreme value index, *Ann. Statist.*, **26**, 434–448.

Drees, H.; de Haan, L. & Resnick, S.I. (2000). How to make a Hill plot, *Ann. Statist.*, **28**, 254–274.

Drees, H. & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. Appl.*, **75**, 149–172.

DuMouchel, W.H. (1983). Estimating the stable index α in order to measure tail thickness: A critique, *Ann. Statist.*, **11**, 1019–1031.

Dupuis, D.J. (2000). Exceedances over high thresholds: A guide to threshold selection, *Extremes*, **1**, 251–261.

Ferreira, A.; de Haan, L. & Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution, *Statistics*, **37**, 401–434.

Feuerverger, A. & Hall, P.G. (1999). Estimating a tail exponent by modelling departure from a Pareto, *Ann. Statist.*, **27**, 760–781.

Fraga Alves, M.I. (2001). A location invariant Hill-type estimator, *Extremes*, **4**(3), 199–217.

Frigessi, A.; Haug, O. & Håvard, R. (2003). A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, **5**, 219–235.

Gilleland, E.; Katz, R.W. & Young, G. (2010). *extRemes: Extreme value toolkit.* R package version 1.62.

Goegebeur, Y.; Beirlant, J. & de Wet, T. (2008). Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation, *Revstat*, **6**, 51–69.

Gomes, M.I. & Oliveira, O. (2001). The bootstrap methodology in statistics of extremes: Choice of the optimal sample fraction, *Extremes*, **4**, 331–358.

Gonzalo, J. & Olmo, J. (2004). Which extreme values are really extreme? *J. Fin. Econ.*, **2**, 349–369.

Guillou, A. & Hall, P.G. (2001). A diagnostic for selecting the threshold in extreme value analysis, *J. R. Statist. Soc. B*, **63**, 293–305.

Hall, P.G. (1982). On some simple estimates of an exponent of regular variation, *J. R. Statist. Soc. B*, **44**, 37–42.

Hall, P.G. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *J. Mult. Anal.*, **32**, 177–203.

Hall, P.G. & Weissman, I. (1997). On the estimation of extreme tail probabilities, *Ann. Statist.*, **25**, 1311–1326.

Hall, P.G. & Welsh, A.H. (1985). Adaptive estimates of the parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

Ho, A.K.F. & Wan, A.T.K. (2002). Testing for covariance stationarity of stock returns in the presence of structural breaks: An intervention analysis, *Appl. Econ. Lett.*, **9**, 441–447.

Katz, R.W.; Parlange, M.B. & Naveau, P. (2002). Statistics of extremes in hydrology, *Adv. Water Res.*, **25**, 1287–1304.

Leadbetter, M.R.; Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series*, Springer, London.

Loretan, M. & Philips, P.C.B. (1994). Testing the covariance stationarity of heavy tailed time series: an overview of the theory with applications to several financial datasets, *J. R. Statist. Soc. D*, **1**, 211–248.

MacDonald, A. (2012). *Extreme value mixture modelling with medical and industrial applications*, PhD thesis, University of Canterbury, New Zealand.

MacDonald, A.; Scarrott, C.J.; Lee, D.; Darlow, B.; Reale, M. & Russell, G. (2011a). A flexible extreme value mixture model, *Comp. Statist. Data Anal.*, **55**, 2137–2157.

MacDonald, A.; Scarrott, C.J. & Lee, D.S. (2011b). Boundary correction, consistency and robustness of kernel densities using extreme value theory. Submitted. Available from: `http://www.math.canterbury.ac.nz/ c.scarrott`.

Mendes, B. & Lopes, H.F. (2004). Data driven estimates for mixtures, *Comp. Statist. Data Anal.*, **47**, 583–598.

Neves, C. & Alves, M.I.F. (2004). Reiss and Thomas' automatic selection of the number of extremes, *Comp. Statist. Data Anal.*, **47**, 689–704.

Omran, M.F. & McKenzie, E. (1999). Testing for covariance stationarity in the uk all-equity, *J. R. Statist. Soc. D*, **48**, 361–369.

Pereira, T.T. (1994). *Second order behaviour of domains of attraction and the bias of generalized Pickands' estimator*. In "Extreme Value Theory and Applications III", Proceedings of Gaithersburg Conference, 1993 (J. Galambos, L. Lechner and E. Simiu, Eds.), vol. 866, pp. 165–177.

Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

Reiss, R.-D. & Thomas, M. (2007). *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhauser, Boston.

Resnick, S.I. (1997). Heavy tail modeling and teletraffic data, *Ann. Statist.*, **25**, 1805–1869.

Resnick, S.I. & Starica, C. (1997). Smoothing the Hill estimator, *Adv. Appl. Prob.*, **29**, 271–293.

Robert, C.P. (1998). *Discretization and MCMC Convergence Assessment*, Lecture Notes in Statistics, vol. 135, Springer, New York.

Robert, C.P. & Casella, G. (2010). *Monte Carlo Statistical Methods*, Springer, New York.

Roberts, G.O. & Rosenthal, J.S. (2009). Examples of adaptive MCMC, *J. Comp. Graph. Statist.*, **18**, 349–367.

Scarrott, A. & MacDonald, C.J. (2010). Risk assessment for critical temperature exceedance in nuclear reactors, *J. Risk Rel.*, **224**.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman, London.

Smith, R.L. (1989). Extreme value analysis of environmental time series: An example based on ozone data (with discussion), *Statist. Sci.*, **4**, 367–393.

Tancredi, A.; Anderson, C.W. & O'Hagan, A. (2006). Accouting for threshold uncertainty in extreme value estimation, *Extremes*, **9**, 87–106.

Thompson, P.; Cai, Y.; Reeve, D. & Stander, J. (2009). Automated threshold selection methods for extreme wave analysis, *Coastal Engineering*, **56**, 1013–1021.

Vrac, M. & Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resourc. Res.*, **43**(W07402).

Zhao, X. (2010). *Extreme value modelling with application in finance and neonatal research*, PhD thesis, University of Canterbury, New Zealand.

Zhao, X.; Scarrott, C.J.; Oxley, L. & Reale, M. (2011). *Let the tails speak for themselves": Bayesian extreme value mixture modelling for estimating VaR*, Submitted. Available from: `http://www.math.canterbury.ac.nz/~c.scarrott`.

Zhao, X.; Scarrott, C.J.; Reale, M. & Oxley, L. (2010). Extreme value modelling for forecasting the market crisis, *Appl. Fin. Econ.*, **20**, 63–72.

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 61–82

MAX-STABLE MODELS FOR MULTIVARIATE EXTREMES

Author: Johan Segers
– ISBA, Université catholique de Louvain, Belgium
johan.segers@uclouvain.be

Abstract:

- Multivariate extreme-value analysis is concerned with the extremes in a multivariate random sample, that is, points of which at least some components have exceptionally large values. Mathematical theory suggests the use of max-stable models for univariate and multivariate extremes. A comprehensive account is given of the various ways in which max-stable models are described. Furthermore, a construction device is proposed for generating parametric families of max-stable distributions. Although the device is not new, its role as a model generator seems not yet to have been fully exploited.

Key-Words:

- *copula; domain of attraction; max-stable distribution; spectral measure; tail dependence.*

AMS Subject Classification:

- 60G70, 62G32.

1. INTRODUCTION

Multivariate extreme-value analysis is concerned with the extremes in a multivariate random sample, that is, points of which at least some components have exceptionally large values. Isolating a single component brings us back to univariate extreme-value theory. In this paper, the focus will rather be on the dependence between extremes in different components. The issue of temporal dependence will be ignored, so that the dependence will be understood as cross-sectional only.

Mathematical theory suggests the use of max-stable models for univariate and multivariate extremes. The univariate margins must be one of the classical extreme-value distributions, Fréchet, Gumbel, and extreme-value Weibull, unified in the generalized extreme-value distributions. For the dependence structure, however, matters are more complicated. A complete characterization in the multivariate case was given in de Haan & Resnick (1977), describing extremal dependence in terms of spectral measures on a subset of the unit sphere. Statistically, this formulation is not always the most practical one, and a large number of other concepts have been proposed.

The aim of this paper is to provide a comprehensive account of the various ways in which max-stable models are described (§2). Second, a construction device is proposed for generating parametric families of max-stable distributions (§3). The device is not new as it appears already for instance in the theory of regularly varying multivariate time series in Basrak & Segers (2009) or in the concept of a D-norm in Falk *et al.* (2010). Still, its role as a model generator seems not yet to have been fully appreciated.

Inference on multivariate extremes via max-stable models for joint tails is an exciting field that is still in development. Inference methods can be nonparametric or parametric, and in the latter case, they can be likelihood-based, frequentist as well as Bayesian, or based on other techniques such as the method of moments or minimum distance estimation.

Max-stable models have the drawback that they are too coarse to describe tails of multivariate distributions with asymptotic independence sufficiently accurately. More refined models exist and they are the topic of an extensive literature, originating from the seminal paper of Ledford & Tawn (1996).

2. FUNDAMENTALS

Max-stable distributions arise out of the study of the limit behaviour of vectors of componentwise maxima. Their dependence structure can be described via a number of dependence functions or via certain measures. All these objects are related and they highlight different aspects of the distribution of multivariate extremes.

2.1. Convergence of vectors of componentwise maxima

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$, for $i = 1, \ldots, n$, be independent and identically distributed random vectors in dimension d. For each component $j = 1, \ldots, d$, consider the sample maximum, and collect these maxima in a new random vector:

$$\boldsymbol{M}_n = (M_{n1}, \ldots, M_{nd}) , \qquad M_{nj} = \max(X_{1j}, \ldots, X_{nj}) .$$

Observe that most of the time, the point $\boldsymbol{M}_n$ does not belong to the sample, as the maxima in different components will typically occur at different time instances. Still, the analysis of the large-sample distribution of $\boldsymbol{M}_n$ is a natural starting point for multivariate extreme-value theory.

Weak convergence of a sequence of random vectors implies weak convergence of each of the components. As in univariate extreme-value theory, it is therefore reasonable to apply increasing affine transformations to each of the margins and consider the sequence of random vectors

$$\left(\frac{M_{nj} - b_{nj}}{a_{nj}} : j = 1, \ldots, d \right) , \qquad n = 1, 2, \ldots , \tag{2.1}$$

in terms of normalizing constants $a_{nj} > 0$ and b_{nj}. For each component $j = 1, \ldots, d$, the weak limit of $(M_{nj} - b_{nj})/a_{nj}$ as $n \to \infty$ must be a univariate max-stable distribution, and necessary and sufficient conditions on the marginal distribution F_j of the j^{th} component X_{ij} for such convergence to take place are well known, see for instance the monograph by de Haan & Ferreira (2006).

However, weak convergence of each of the d components in (2.1) is strictly weaker than joint convergence of the vector of normalized maxima. What is needed in addition is a condition on the dependence structure of the common joint distribution F of the random vectors $\boldsymbol{X}_i$. A convenient way to describe this dependence is via the copula C_1 of F, that is,

$$\Pr\big[\boldsymbol{X}_i \le \boldsymbol{x}\big] = F(\boldsymbol{x}) = C_1\big(F_1(x_1), \ldots, F_d(x_d)\big) . \tag{2.2}$$

Assuming the margins $F_1, \ldots, F_d$ are continuous, as we will do henceforth, the copula C_1 of the distribution function F in (2.2) is unique and can be obtained as the joint distribution function of the random vector $\big(F_1(X_{i1}), \ldots, F_d(X_{id})\big)$.

Weak convergence of a sequence of multivariate distribution functions to a limit with continuous margins is equivalent to weak convergence of the sequences of margins and of the sequence of copulas (Deheuvels, 1984; Galambos, 1987). The copula of the vector of component-wise maxima $\boldsymbol{M}_n$, and hence of any vector that is obtained by applying increasing transformations to each of its components, is given by

$$C_n(\boldsymbol{u}) \,=\, \Big\{C_1\big(u_1^{1/n}, \ldots, u_d^{1/n}\big)\Big\}^n \,. \tag{2.3}$$

This can be checked from the fact that the joint distribution function of $\boldsymbol{M}_n$ is F^n while its margins are F_j^n for $j = 1, \ldots, d$. Hence, in order for the normalized maxima in (2.1) to converge in distribution to a nondegenerate limit, besides marginal convergence, the sequence of copulas C_n must converge as well.

The copulas that can arise as weak limits of C_n as $n \to \infty$ are called extreme-value copulas, that is, a copula C is called an extreme-value copula if there exists a copula C_1 such that, as $n \to \infty$,

$$\lim_{n\to\infty} \Big\{C_1\big(u_1^{1/n}, \ldots, u_d^{1/n}\big)\Big\}^n \,=\, C(u_1, \ldots, u_d) \,. \tag{2.4}$$

Extreme-value copulas arise as the class of possible limit copulas of vectors $\boldsymbol{M}_n$ as $n \to \infty$. The copula C_1 is said to be in the domain of attraction of C. An extensive survey of the literature on extreme-value copulas is given in Gudendorf & Segers (2010).

The class of extreme-value copulas coincides with that of max-stable copulas, defined as follows. A copula C is max-stable if, for all $\boldsymbol{u} \in [0,1]^d$ and $k = 1, 2, \ldots,$

$$C(\boldsymbol{u}) \,=\, \Big\{C\big(u_1^{1/k}, \ldots, u_d^{1/k}\big)\Big\}^k \,.$$

In the setting of componentwise maxima of independent random samples, the previous identity means that the copula C_k of the random vector $\boldsymbol{M}_k$ is the same for every sample size k. Clearly, a max-stable copula is also an extreme-value copula, being in its own domain of attraction. Conversely, each extreme-value copula can be shown to be max-stable: in (2.4), partition the sample of size $n = mk$ in m blocks of size k and let m tend to infinity for fixed k. Since the limit must not depend on k, the max-stability relation follows.

In summary, we have found that nondegenerate limit distributions of vectors of appropriately normalized componentwise maxima have extreme-value margins and an extreme-value or max-stable copula. Specifically, if

$$\Pr\left[\bigcap_{j=1}^{d}\left\{\frac{M_{nj} - b_{nj}}{a_{nj}} \le x_j\right\}\right] \,\xrightarrow{w}\, G(x_1, \ldots, x_d) \,, \qquad n \to \infty \,,$$

then necessarily

$$G(x_1, \ldots, x_d) \;=\; C\big(G_1(x_1), \ldots, G_d(x_d)\big)$$

with extreme-value margins $G_1, \ldots, G_d$ and an extreme-value copula C. Convergence of margins and convergence of copulas being two isolated issues, we can ignore the former and rather focus on the latter. In fact, the way in which the components are normalized is immaterial, as long as the transformations applied to the components are increasing.

2.2. Dependence functions

Take logarithms and apply a linear expansion to see that (2.4) is equivalent to

$$\begin{aligned} \lim_{n\to\infty} n\Big\{1 - C_1\big(1 - n^{-1}x_1, \ldots, 1 - n^{-1}x_d\big)\Big\} \;&=\; -\log C\big(e^{-x_1}, \ldots, e^{-x_d}\big) \\ &=\; \ell(\boldsymbol{x}) \;, \qquad \boldsymbol{x} \in [0,\infty)^d \,. \end{aligned} \tag{2.5}$$

The limit ℓ is called the stable tail dependence function of C, going back to Huang (1992) and Drees & Huang (1998). The variable n tending to infinity along the positive integers can be replaced by a variable t tending to infinity along the positive reals.

The best known example is the Gumbel–Hougaard copula, for which $\ell_\theta(\boldsymbol{x}) = (x_1^\theta + \cdots + x_d^\theta)^{1/\theta}$ in terms of a parameter $\theta \in [1, \infty]$ (Gumbel, 1960; Hougaard, 1986). The function ℓ_θ happens to be the θ-norm of the vector $\boldsymbol{x}$. The fact that ℓ_θ is a norm is no coincidence: in a remarkable paper by Molchanov (2008), a characterization is given of all the norms that can give rise to stable tail dependence functions. In Falk *et al.* (2010), $\ell(\boldsymbol{x})$ is called the D-norm of $\boldsymbol{x}$, with D referring to the Pickands dependence function, see below.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ denote a generic random vector in the original sample. The expression on the left-hand side in (2.5) contains the rescaled probability

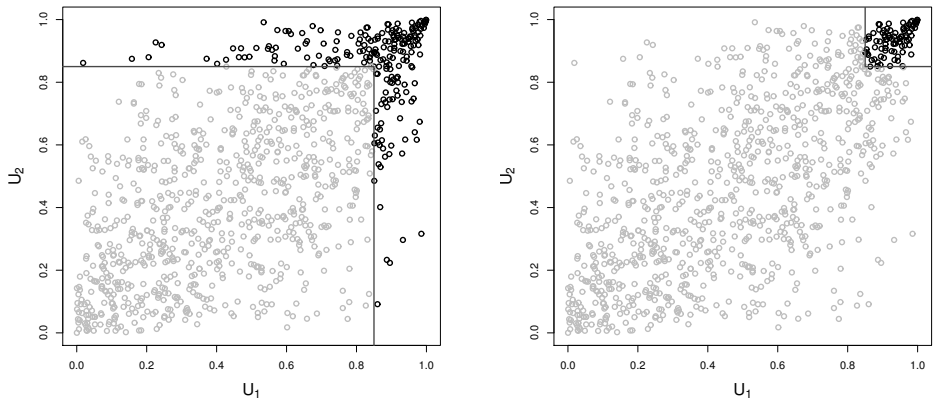
$$\begin{aligned} &1 - C_1\big(1 - x_1/n,\, \ldots,\, 1 - x_d/n\big) \;= \\ &\qquad =\; \Pr\Big[F_1(X_1) > 1 - x_1/n \quad \text{or} \quad \ldots \quad \text{or} \quad F_d(X_d) > 1 - x_d/n\Big] \,. \end{aligned} \tag{2.6}$$

This probability concerns the event that at least one among the d components $X_1, \ldots, X_d$ should exceed a high percentile of its own distribution. The copula domain-of-attraction condition (2.4), originally involving the vector of componentwise sample maxima, has been replaced by a condition on the upper tail of a single random vector. This is akin to the familiar peaks-over-threshold approach in univariate extreme-value theory.

The tail copula, R, of Schmidt & Stadtmüller (2006) arises if all d components are required to exceed a large percentile simultaneously:

$$\lim_{n\to\infty} n \Pr\Big[F_1(X_1) > 1 - x_1/n \quad \text{and} \quad \ldots \quad \text{and} \quad F_d(X_d) > 1 - x_d/n\Big] = \\ = R(\boldsymbol{x}) \,, \qquad \boldsymbol{x} \in [0,\infty)^d .$$

Clearly, the relation between the functions ℓ and R is governed by the inclusion-exclusion formula. In higher dimensions, ℓ is somewhat more convenient than R, as setting some components x_j in the definition of ℓ to zero allows one to retrieve the lower-dimensional margins of the extreme-value copula. This is not possible for the tail copula R, as setting even a single x_j to zero immediately yields $R(\boldsymbol{x}) = 0$. The difference between the two functions ℓ and R is depicted in Figure 1.



U_1 or U_2 are large $\qquad$ U_1 and U_2 are large

Figure 1: Left: the stable tail dependence function ℓ — at least one component should be large. Right: the tail copula R — all components must be large simultaneously. Here $U_j = F_j(X_j)$ for $j = 1, 2$.

The dependence functions ℓ and R are homogeneous:

$$\begin{aligned} \ell(a\,\boldsymbol{x}) &= \lim_{t\to\infty} t\Big\{1 - C_1\big(1 - x_1/(t/a), \ldots, 1 - x_d/(t/a)\big)\Big\} \\ &= \lim_{s\to\infty} a\,s\Big\{1 - C_1\big(1 - x_1/s, \ldots, 1 - x_d/s\big)\Big\} \\ &= a\,\ell(\boldsymbol{x}) \,, \qquad a > 0\,, \quad \boldsymbol{x} \in [0,\infty)^d \,, \end{aligned} \tag{2.7}$$

and similarly for R. It is therefore sufficient to consider the restriction of these functions to the unit simplex $\Delta_{d-1} = \big\{(w_1, \ldots, w_d) \in [0,1]^d \colon\, w_1 + \cdots + w_d = 1\big\}$.

The restriction of ℓ to Δ_{d-1} is called the Pickands dependence function, D, after Pickands (1981). By homogeneity,

$$\ell(\boldsymbol{x}) = (x_1 + \cdots + x_d)\, D(w_1, \ldots, w_d) \ , \qquad w_j = \frac{x_j}{x_1 + \cdots + x_d} \ .$$

Frequently, the variable $w_d = 1 - w_1 - \cdots - w_{d-1}$ is suppressed from the notation and D is written as a function of the arguments $(w_1, \ldots, w_{d-1})$ only.

The probability on the right-hand side of (2.6) involves the union of the events $\{F_j(X_j) > 1 - x_j/n\}$, each of which has probability x_j/n, provided $0 \le x_j \le n$. As a consequence, we have the elementary bounds

$$\begin{aligned} \max(x_1/n, \ldots, x_d/n) \ &\le \ \Pr\Big[F_1(X_1) > 1 - x_1/n \quad \text{or} \quad \ldots \quad \text{or} \quad F_d(X_d) > 1 - x_d/n\Big] \\ &\le \ x_1/n + \cdots + x_d/n \ . \end{aligned}$$

Multiplying by n and letting n tend to infinity, we obtain

$$\max(x_1, \ldots, x_d) \ \le \ \ell(x_1, \ldots, x_d) \ \le \ x_1 + \cdots + x_d \ , \qquad \boldsymbol{x} \in [0, \infty)^d \, . \tag{2.8}$$

By (2.5) and (2.8), an extreme-value copula C must satisfy

$$u_1 \cdots u_d \ \le \ C(u_1, \ldots, u_d) \ \le \ \max(u_1, \ldots, u_n) \ , \qquad \boldsymbol{u} \in [0, 1]^d \, . \tag{2.9}$$

The lower and upper bounds in the two previous displays can be attained, corresponding to the extreme cases of independence and perfect association. In particular, max-stable models are positive quadrant dependent. In fact, in Garralda-Guillem (2000), the stronger property is shown that bivariate extreme-value copulas are monotone regression dependent; see also Theorem 5.2.10 in Resnick (1987).

By (2.5), the copula C can be given in terms of the tail dependence function through

$$C(u_1, \ldots, u_d) \ = \ \exp\Big\{-\ell\big(-\log u_1, \ldots, -\log u_d\big)\Big\} \ , \qquad \boldsymbol{u} \in (0, 1]^d \, .$$

In extreme-value theory, it is often convenient to standardize to other distributions than the uniform $(0, 1)$ law. The three most common forms are the unit Fréchet distribution, the Gumbel distribution, and the reverse exponential distribution, yielding respectively

$$C\big(e^{-1/x_1}, \ldots, e^{-1/x_d}\big) \ = \ \exp\Big\{-\ell\big(1/x_1, \ldots, 1/x_d\big)\Big\} \, , \qquad \boldsymbol{x} \in (0, \infty)^d, \tag{2.10}$$

$$C\big(e^{-e^{-x_1}}, \ldots, e^{-e^{-x_d}}\big) \ = \ \exp\Big\{-\ell\big(e^{-x_1}, \ldots, e^{-x_d}\big)\Big\} \, , \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{2.11}$$

$$C\big(e^{x_1}, \ldots, e^{x_d}\big) \ = \ \exp\Big\{-\ell\big(-x_1, \ldots, -x_d\big)\Big\} \, , \qquad \boldsymbol{x} \in (-\infty, 0)^d. \tag{2.12}$$

When using unit Fréchet margins, the notation $V(x_1, \ldots, x_d) = \ell(1/x_1, \ldots, 1/x_d)$ is often employed too.

2.3. The intensity measure

The transformation of the components X_j to uniform $(0,1)$ random variables via the probability integral transform $F_j(X_j)$ has the disadvantage that all the action regarding the upper extremes is compressed to the neighbourhood of 1. Instead, for a univariate sequence $\xi_1, \xi_2, \ldots$ of independent and identically distributed random variables with common distribution function F_ξ, define the first exceedance time of the level x by

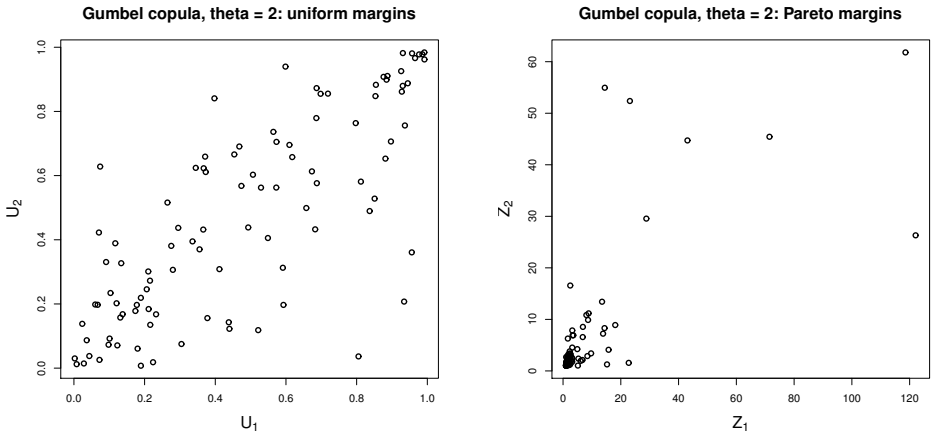
$$T(x) \;=\; \inf\Big\{ i = 1, 2, \ldots : \; \xi_i > x \Big\} .$$

If $F_\xi(x) < 1$, then $T(x)$ will be a geometric random variable with success probability equal to $1 - F_\xi(x)$. Its expectation,

$$\mathrm{E}[T(x)] \;=\; \frac{1}{1 - F_\xi(x)} \;,$$

is called the return time of the level x.

Now let us apply this return time transformation to each of the d components of the random vector $(X_1, \ldots, X_d)$. The return time of observation X_j is $Y_j = 1/\{1 - F_j(X_j)\}$. The law of Y_j is unit Pareto rather than uniform on $(0,1)$, as $\Pr\big[Y_j > y\big] = \Pr\big[F_j(X_j) > 1 - 1/y\big] = 1/y$ for $y \geq 1$. We find that values of X_j corresponding to high percentiles of F_j are mapped to large values of Y_j. As is evident from Figure 2, extremes are magnified.



$U_{ij} = F_j(X_{ij})$ $\qquad$ $Z_{ij} = 1/\big\{1 - F_j(X_{ij})\big\}$

Figure 2: Two views on a sample: uniform (left) versus Pareto (right).

Now suppose that the copula C_1 is in the domain of attraction of an extreme-value copula with stable tail dependence function ℓ. Equation (2.5) says

that the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ satisfies

$$\begin{aligned} t\Big\{1 - C_1\big(1 - x_1/t, \ldots, 1 - x_d/t\big)\Big\} &= t \Pr\left[\bigcup_{j=1}^{d} \big\{Y_j > t/x_j\big\}\right] \\ &= t \Pr\Big[\boldsymbol{Y}/t \in \big([0,\infty]^d \setminus [\boldsymbol{0}, 1/\boldsymbol{x}]\big)\Big] \\ &\to \ell(\boldsymbol{x}) , \qquad t \to \infty . \end{aligned}$$

It follows that on the space $\mathbb{E}_d = [0,\infty]^d \setminus \{\boldsymbol{0}\}$, there exists a measure μ such that

$$\mathrm{E}\left[\sum_{i=1}^{n} I\big(\boldsymbol{Y}_i/n \in \cdot\,\big)\right] = n \Pr\big[\boldsymbol{Y}/n \in \cdot\,\big] \xrightarrow{v} \mu(\,\cdot\,) , \qquad n \to \infty . \tag{2.13}$$

The limit takes place in the mode of vague convergence of measures, meaning that $\lim_{n\to\infty} n\,\mathrm{E}[f(\boldsymbol{Y}/n)] = \int_{\mathbb{E}_d} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$ for every bounded, continuous function f on $\mathbb{E}_d$ that vanishes in a neighbourhood of the origin. Intuitively, when n grows large, the vector $\boldsymbol{Y}/n$ is pulled towards the origin, in the neighbourhood of which the function f is zero. The intensity measure then only concerns the upper tail of the distribution of $\boldsymbol{Y}$. The first expression in the previous display shows that μ expresses the limiting average intensity in space of the normalized sample cloud $\big\{\boldsymbol{Y}_1/n, \ldots, \boldsymbol{Y}_n/n\big\}$. The stable tail dependence function acts as a distribution function for the intensity measure μ, as we have

$$\ell(\boldsymbol{x}) = \mu\big([0,\infty]^d \setminus [\boldsymbol{0}, 1/\boldsymbol{x}]\big) .$$

2.4. Extreme profiles

Assume that the vector of return times $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ with $Y_j = 1/\{1 - F_j(X_j)\}$ is large, that is, at least one of its components exceeds a high threshold. The relative sizes of the d components then inform us about the extremal dependence: are some components large simultaneously or is one specific component dominating all the other ones? Specifically, for $\boldsymbol{y} \in [0,\infty)^d \setminus \{0\}$ put

$$\begin{aligned} r(\boldsymbol{y}) &= y_1 + \cdots + y_d \;\in\; (0,\infty) , \\ \boldsymbol{w}(\boldsymbol{y}) &= \big(y_j/r(\boldsymbol{y}) \colon\, j = 1, \ldots, d\big) \;\in\; \Delta_{d-1} , \end{aligned}$$

to be thought of as the magnitude and the profile of the vector $\boldsymbol{y}$, respectively. The spectral measure H lives on the unit simplex and is defined by

$$H(B) = \mu\Big(\big\{\boldsymbol{y} \colon\, r(\boldsymbol{y}) > 1,\; \boldsymbol{w}(\boldsymbol{y}) \in B\big\}\Big) ,$$

for Borel subsets B of Δ_{d-1}.

The measure convergence in (2.13) implies that

$$\begin{aligned} &\mathrm{E}\left[\sum_{i=1}^{n} I\big\{r(\boldsymbol{Y}_i) > n,\ \boldsymbol{w}(\boldsymbol{Y}_i) \in \cdot\,\big\}\right] = \\ &\qquad = n\,\Pr\big[r(\boldsymbol{Y}) > n,\ \boldsymbol{w}(\boldsymbol{Y}) \in \cdot\,\big] \xrightarrow{w} H(\cdot)\,, \qquad n \to \infty\,, \end{aligned} \tag{2.14}$$

with vague convergence being replaced by weak convergence because the state space Δ_{d-1} is compact. In particular,

$$\Pr\big[\boldsymbol{w}(\boldsymbol{Y}) \in \cdot \mid r(\boldsymbol{Y}) > t\big] \xrightarrow{d} \frac{H(\cdot)}{H(\Delta_{d-1})} = Q(\cdot)\,, \qquad t \to \infty\,, \tag{2.15}$$

meaning that the conditional distribution of the profile $\boldsymbol{w}(\boldsymbol{Y})$ given that the magnitude $r(\boldsymbol{Y})$ is large converges in distribution to the normalized spectral measure Q.

The spectral measure H and the profile distribution $Q(\cdot)$ are alternative, equivalent ways of describing the extreme-value copula C. Indeed, homogeneity of ℓ in (2.7) implies homogeneity of μ:

$$\mu(a\,\cdot) = a^{-1}\mu(\cdot)\,, \qquad a > 0\,. \tag{2.16}$$

As a consequence, the intensity measure satisfies

$$\begin{aligned} \mu\Big(\big\{\boldsymbol{y}\colon\, r(\boldsymbol{y}) > z,\, \boldsymbol{w}(\boldsymbol{y}) \in B\big\}\Big) &= z^{-1}\mu\Big(\big\{\boldsymbol{y} \mid r > 1,\, \boldsymbol{w} \in B\big\}\Big) \\ &= z^{-1}H(B) \end{aligned} \tag{2.17}$$

for $z > 0$ and for Borel sets B of Δ_{d-1}. That is, when expressing a point $\boldsymbol{y}$ in the coordinates $(r, \boldsymbol{w})$, the intensity measure μ factorizes into a product measure on $(0,\infty) \times \Delta_{d-1}$ given by $r^{-2}\,\mathrm{d}r\,H(\mathrm{d}\boldsymbol{w})$. Equation (2.17) leads to

$$\int f(\boldsymbol{y})\,\mu(\mathrm{d}\boldsymbol{y}) = \int_{\Delta_{d-1}} \int_0^\infty f(r\,\boldsymbol{w})\,r^{-2}\,\mathrm{d}r\;H(\mathrm{d}\boldsymbol{w}) \tag{2.18}$$

for μ-integrable functions f, showing how to recover μ and thus ℓ and C from H. The special case where f is equal to the indicator function of the set $\big\{\boldsymbol{y}\colon \max_j x_j\,y_j > 1\big\}$ for some $\boldsymbol{x} \in [0,\infty)^d$ yields, after some computation,

$$\ell(\boldsymbol{x}) = \int_{\Delta_{d-1}} \max_{j=1,\ldots,d}(w_j\,x_j)\,H(\mathrm{d}\boldsymbol{w})\,. \tag{2.19}$$

Incidentally, this representation of ℓ implies that ℓ must be convex. By specializing the bounds in (2.8) to the unit vectors in $\mathbb{R}^d$, one finds that the spectral measure H must satisfy the constraints

$$1 = \ell(\boldsymbol{e}_j) = \int_{\Delta_{d-1}} w_j\,H(\mathrm{d}\boldsymbol{w})\,, \qquad j = 1,\ldots,d\,. \tag{2.20}$$

It follows that the total mass of H is equal to

$$H(\Delta_{d-1}) \;=\; \int (w_1 + \cdots + w_d)\, H(\mathrm{d}\boldsymbol{w}) \;=\; d \;.$$

Thanks to this property, it is possible to recover the spectral measure H from the profile distribution Q. From (2.20), it then follows that a random vector $\boldsymbol{W} = (W_1, \ldots, W_d)$ on Δ_{d-1} with law equal to Q must satisfy

$$\mathrm{E}_Q[W_j] \;=\; \int_{\Delta_{d-1}} w_j \, Q(\mathrm{d}\boldsymbol{w}) \;=\; 1/d \;, \qquad j = 1, \ldots, d \;. \tag{2.21}$$

In §3, we will see that any such law Q can appear as the profile distribution of a d-variate max-stable distribution.

In case of asymptotic independence, $\ell(\boldsymbol{x}) = x_1 + \cdots + x_d$, the profile distribution Q is equal to the discrete uniform distribution on the d vertices of Δ_{d-1}: asymptotically, only one component can be large at a time. In the case of asymptotic perfect dependence, $\ell(\boldsymbol{x}) = \max(x_1, \ldots, x_d)$, the profile distribution Q is degenerate at the center $(1/d, \ldots, 1/d)$ of Δ_{d-1}: all components are equally large. These two extreme cases are depicted in Figure 3.

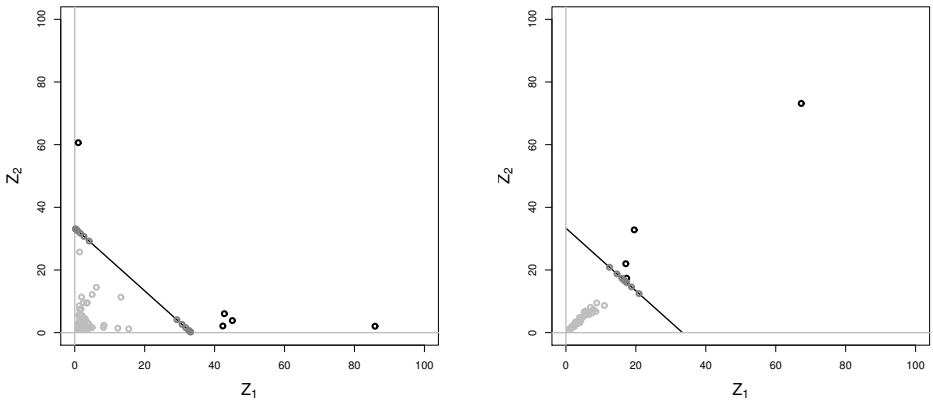


Figure 3: Asymptotic independence (left) versus asymptotic perfect dependence (right).

To show the ease with which coefficients related to extremal dependence can be computed, consider the random variable

$$N(t) \;=\; \sum_{j=1}^{d} I\big\{F_j(X_j) > 1 - 1/t\big\} \;,$$

counting the number of components that exceed a high percentile. The following dependence coefficients have natural interpretations.

- *Trouble is in the air:*

$$\lim_{t\to\infty} t\,\Pr\bigl[N(t)\geq 1\bigr] \;=\; \ell(1,\ldots,1) \;=\; \int_{\Delta_{d-1}} \max(w_1,\ldots,w_d)\,H(\mathrm{d}\boldsymbol{w}) \;.$$

- *Multiple failures:*

$$\lim_{t\to\infty} t\,\Pr\bigl[N(t)\geq k\bigr] \;=\; \int_{\Delta_{d-1}} w_{(d-k+1)}\,H(\mathrm{d}\boldsymbol{w}) \;,$$

  for $k=1,\ldots,d$, where $w_{(1)}\leq\cdots\leq w_{(d)}$ denote the order statistics of $(w_1,\ldots,w_d)$.

- *The sky is falling:*

$$\lim_{t\to\infty} t\,\Pr\bigl[N(t)=d\bigr] \;=\; R(1,\ldots,1) \;=\; \int_{\Delta_{d-1}} \min(w_1,\ldots,w_d)\,H(\mathrm{d}\boldsymbol{w}) \;.$$

- *System collapse — how bad will it get?*

$$\lim_{t\to\infty} \mathrm{E}\bigl[N(t)-k \mid N(t)\geq k\bigr] \;=\; \frac{\int \bigl(w_{(1)}+\cdots+w_{(d-k)}\bigr)\,H(\mathrm{d}\boldsymbol{w})}{\int w_{(d-k+1)}\,H(\mathrm{d}\boldsymbol{w})} \;,$$

  for $k=1,\ldots,d-1$.

3. CONSTRUCTING MODELS

There is a fairly large number of parametric max-stable models available; see for instance the overviews in Kotz & Nadarajah (2000) and Beirlant *et al.* (2004). In a search for flexible models in large dimensions, new families are still being constructed, as in Ballani & Schlather (2011), Boldi & Davison (2007), Cooley *et al.* (2010), and Fougères *et al.* (2009). In this section, a simple construction device will be proposed and illustrated.

From §2, we recall that max-stable models for extremal dependence can be represented either via the extreme-value copula C, the stable tail dependence function ℓ, the intensity measure μ, or the spectral measure H and its normalized version, the profile distribution Q. However, as these objects must satisfy certain constraints, construction of parametric models is not obvious, particularly in high dimensions. Even if flexible parametric forms can be found, interpretation of the model parameters may not be obvious. In addition, when working with the spectral measure or profile distribution, the passage to lower-dimensional margins can be awkward, as the conditioning events in (2.15) will be different according to which components are selected.

3.1. A construction device

Let Z be a unit Fréchet random variable, that is, $\Pr(Z \le z) = \exp(-1/z)$ for $z > 0$. Let $\boldsymbol{A} = (A_1, \ldots, A_d)$ be a random vector, independent of Z, such that $0 < \mathrm{E}[\max(A_j, 0)] < \infty$ for every $j \in \{1, \ldots, d\}$. Consider the random vector

$$\boldsymbol{X} = (X_1, \ldots, X_d) = (A_1 Z, \ldots, A_d Z) \; . \tag{3.1}$$

The max-stable attractor of the distribution of $\boldsymbol{X}$ can be explicitly calculated.

Lemma 3.1. *For $x = (x_1, \ldots, x_d) \in (0, \infty)^d$, we have*

$$\begin{aligned} \lim_{n \to \infty} \Pr\Big[X_1 \le n x_1, \ldots, X_d \le n x_d \Big]^n = \\ = \exp\Big\{ - \mathrm{E}\Big[\max\big(A_1/x_1, \ldots, A_d/x_d, 0 \big) \Big] \Big\} \; . \end{aligned} \tag{3.2}$$

Proof: Let $x \in (0, \infty)^d$. We have

$$\begin{aligned} \Pr\Big[X_1 \le x_1, \ldots, X_d \le x_d \Big] &= \Pr\Big[A_1 Z \le x_1, \ldots, A_d Z \le x_d \Big] \\ &= \Pr\Big[A_1/x_1 \le 1/Z, \ldots, A_d/x_d \le 1/Z \Big] \\ &= \Pr\Big[1/Z \ge \max\big(A_1/x_1, \ldots, A_d/x_d \big) \Big] \; . \end{aligned}$$

The distribution of $1/Z$ is unit exponential. Since A and Z are independent,

$$\Pr\Big[X_1 \le x_1, \ldots, X_d \le x_d \mid A_1, \ldots, A_d \Big] = \exp\Big\{ - \max\big(A_1/x_1, \ldots, A_d/x_d, 0 \big) \Big\} \; .$$

It follows that

$$\Pr\Big[X_1 \le x_1, \ldots, X_d \le x_d \Big] = \mathrm{E}\Big[\exp\Big\{ - \max\big(A_1/x_1, \ldots, A_d/x_d, 0 \big) \Big\} \Big] \; .$$

Let $A^{(1)}, A^{(2)}, \ldots$ be a sequence of independent and identically distributed copies of A. Fix positive integer n. We have

$$\begin{aligned} \Pr\Big[X_1 \le n x_1, \ldots, X_d \le n x_d \Big]^n = \\ &= \Big(\mathrm{E}\Big[\exp\Big\{ - \max\big(A_1/x_1, \ldots, A_d/x_d, 0 \big) \Big\} \Big] \Big)^n \\ &= \mathrm{E}\left[\exp\left\{ - \frac{1}{n} \sum_{i=1}^{n} \max\big(A_1^{(i)}/x_1, \ldots, A_d^{(i)}/x_d, 0 \big) \right\} \right] \; . \end{aligned}$$

Equation (3.2) now follows by the law of large numbers and the dominated convergence theorem. □

The margins of the limit distribution function, say G, in (3.2) are equal to $G_j(x_j) = \exp\{-\mathrm{E}[\max(A_j, 0)]/x_j\}$ for $x_j > 0$. Assume that $\mathrm{E}[\max(A_j, 0)] = 1$ for all $j = 1, \ldots, d$; this can always be achieved by rescaling the variables A_j if necessary. In that case, the margins of G are unit Fréchet. Comparing equations (3.2) and (2.10), we find that the stable tail dependence function of G is given by

$$\ell_{\boldsymbol{A}}(x_1, \ldots, x_d) \;=\; \mathrm{E}\Big[\max\big(x_1 A_1, \ldots, x_d A_d, 0\big)\Big]\,. \tag{3.3}$$

The spectral measure H corresponding to ℓ in (3.3) can be identified too. Write $A_j^+ = \max(A_j, 0)$ and put $R = A_1^+ + \cdots + A_d^+$. On the event $R > 0$, define $W_j = A_j^+/R$; on the event $R = 0$, the definition of W_j is immaterial — for definiteness, put $W_j = 1/d$ if $R = 0$. Note that $\boldsymbol{W}$ takes values in Δ_{d-1} and that $\mathrm{E}[R] = \sum_{j=1}^d \mathrm{E}[A_j^+] = d$. We have $A_j^+ = RW_j$ and thus

$$\begin{aligned}\ell_{\boldsymbol{A}}(x_1, \ldots, x_d) \;&=\; \mathrm{E}\Big[R\,\max\big(W_1 x_1, \ldots, W_d x_d\big)\Big] \\ &=\; \mathrm{E}\Big[\mathrm{E}\big[R \mid W_1, \ldots, W_d\big]\;\max\big(W_1 x_1, \ldots, W_d x_d\big)\Big]\,.\end{aligned}$$

Comparing this expression with (2.19), we find that the spectral measure H of $\ell_{\boldsymbol{A}}$ is given by

$$H(\mathrm{d}\boldsymbol{w}) \;=\; \mathrm{E}\big[R \mid \boldsymbol{W} = \boldsymbol{w}\big]\;\Pr\big[\boldsymbol{W} \in \mathrm{d}\boldsymbol{w}\big]\,, \tag{3.4}$$

that is, H is absolutely continuous with respect to the law of $\boldsymbol{W}$ with Radon–Nikodym derivative equal to $\mathrm{E}\big[R \mid \boldsymbol{W} = \boldsymbol{w}\big]$. Similarly, the profile distribution Q satisfies

$$Q(\mathrm{d}\boldsymbol{w}) \;=\; d^{-1}\,\mathrm{E}\big[R \mid \boldsymbol{W} = \boldsymbol{w}\big]\;\Pr\big[\boldsymbol{W} \in \mathrm{d}\boldsymbol{w}\big]\,. \tag{3.5}$$

Intuitively, this makes sense: profiles $\boldsymbol{W}$ that on average yield larger values of R will have a larger contribution to the joint extremes of $\boldsymbol{X}$.

Incidentally, this construction shows that any probability distribution Q on Δ_{d-1} satisfying (2.21) can appear as the profile distribution of a d-variate max-stable distribution. Indeed, let the random vector $\boldsymbol{W}$ on Δ_{d-1} have law Q and put $A_j = dW_j$ for all $j = 1, \ldots, d$. As $A_1 + \cdots + A_d = d$ by construction, the law of the random vector $\boldsymbol{X}$ in (3.1) is in the domain of attraction of a d-variate max-stable distribution with profile distribution equal to Q.

If the dimension d is large, realistic models of extremal dependence should allow for the possibility that only some but not all components of a random vector are large simultaneously. In terms of the spectral measure or the profile distribution, this is encoded by the lower-dimensional faces of the unit simplex. For a non-empty subset I of $\{1, \ldots, d\}$, let $\Delta_{d-1,I}$ denote the set of all $\boldsymbol{w}$ in Δ_{d-1} such that $w_j > 0$ if $j \in I$ and $w_j = 0$ otherwise. If the probability of the event $\big\{\min_{j\in I} A_j > 0 \geq \max_{j\in I^c} A_j\big\}$ is non-zero, then by (3.4), the spectral measure and the profile distribution will put positive mass on $\Delta_{d-1,I}$. The set I contains the indices of the components that are large.

3.2. Examples

The simplicity of (3.3) is appealing and the reader is invited to the apply the recipe in order to produce his or her own parametric models. In the remainder of the article, a number of well-known examples are worked out.

Example 3.1 (Independence)**.** Suppose that $\Pr\big[A_j > 0 \geq \max(A_i : i \neq j)\big] = p_j$ for $p_j > 0$ and $p_1 + \cdots + p_d = 1$. Then only component can be large at the time. After standardization ensuring that $\mathrm{E}[\max(A_j, 0)] = 1$ for all $j = 1, \ldots, d$, we find $\ell_{\boldsymbol{A}}(x_1, \ldots, x_d) = \mathrm{E}\big[\max(x_1 A_1, \ldots, x_d A_d, 0)\big] = x_1 + \cdots + x_d$, the stable tail dependence function of the independence copula.

Example 3.2 (Perfect dependence)**.** Suppose that $A_j = a_j B$ with probability one for all $j = 1, \ldots, d$, for some constants $a_j > 0$ and a random variable B such that $\mathrm{E}[\max(B, 0)]$ is positive and finite. Then the profile of an extreme vector is fixed. After standardization, $a_j = 1/\mathrm{E}[\max(B, 0)]$, the stable tail dependence function is that of perfect positive association, that is, $\ell_{\boldsymbol{A}}(x_1, \ldots, x_d) = \max(x_1, \ldots, x_d)$.

Example 3.3 (Discrete spectral measures)**.** Suppose that the distribution of $\boldsymbol{A}$ is discrete with a finite number of atoms. Specifically, suppose that $\Pr[\boldsymbol{A} = \boldsymbol{a}_k] = p_k$ for $k \in \{1, \ldots, m\}$, where $\boldsymbol{a}_k \in \mathbb{R}^d$ and $p_k \in (0, 1)$ such that $\sum_{k=1}^m p_k = 1$. Via standardization, ensure that $1 = \mathrm{E}[\max(A_j, 0)] = \sum_{k=1}^m p_k \, a_{kj}^+$ for all $j = 1, \ldots, d$. Put $r_k = a_{k1}^+ + \cdots + a_{kd}^+$ and write $\boldsymbol{a}_k^+ = r_k \, \boldsymbol{w}_k$: if $r_k > 0$, then $w_{kj} = a_{kj}^+ / r_k$, whereas if $r_k = 0$, then put $w_{kj} = 1/d$. It follows that

$$\begin{aligned} \ell_{\boldsymbol{A}}(x_1, \ldots, x_d) &= \sum_{k=1}^m p_k \max\big(a_{k1}^+ x_1, \ldots, a_{kd}^+ x_d\big) \\ &= \sum_{k=1}^m (p_k r_k) \max\big(w_{k1} x_1, \ldots, w_{kd} x_d\big) \, . \end{aligned}$$

We find that the spectral measure H and the profile distribution Q are discrete and are given by

$$H = \sum_{k=1}^m p_k r_k \, \delta_{\boldsymbol{w}_k} \, , \qquad Q = \sum_{k=1}^m d^{-1} p_k r_k \, \delta_{\boldsymbol{w}_k} \, ,$$

with $\delta_{\boldsymbol{w}}$ a unit point mass at $\boldsymbol{w}$. The probabilities p_k are tilted with the magnitudes r_k, giving higher prominence to profiles $\boldsymbol{w}_k$ that are associated to larger values of r_k.

Max-stable models with discrete spectral measures are called extreme-value factor models in Einmahl *et al.* (2011). Each of the m possible outcomes $\boldsymbol{a}_k$ results

in a different profile, according to the type of event or factor that triggered the extreme value.

Example 3.4 (Random indicators)**.** Let

$$\ell(x,y) \,=\, \mathrm{E}\big[\max(xA, yB, 0)\big]\,, \qquad (x,y)\in[0,\infty)^2\,,$$

in terms of random variables A and B such that $\mathrm{E}[\max(A,0)] = \mathrm{E}[\max(B,0)] = 1$. Let (I,J) be a pair of random indicators, independent of the pair (A,B), such that $\Pr[I=1] = p$, $\Pr[J=1] = q$ and $\Pr[I=J=1] = r$. In the definition of ℓ, replace the pair (A,B) by the pair $(p^{-1}IA, q^{-1}JB)$; we assume that $0 < p \le 1$ and $0 < q \le 1$. The new stable tail dependence function is equal to

$$\begin{aligned}\mathrm{E}\Big[\max\big(p^{-1}xIA,\, q^{-1}yJB,\, 0\big)\Big] \,&=\, \mathrm{E}\Big[\max\big(p^{-1}xA,\, q^{-1}yB\big)\Big]\,\Pr\big[I=J=1\big]\\ &\quad+\,\mathrm{E}\Big[\max\big(p^{-1}xA,\,0\big)\Big]\,\Pr\big[I=1,\,J=0\big]\\ &\quad+\,\mathrm{E}\Big[\max\big(q^{-1}yB,\,0\big)\Big]\,\Pr\big[I=0,\,J=1\big]\\ &=\,\ell\big(p^{-1}x, q^{-1}y\big)\,r\,+\,p^{-1}x\,(p-r)\,+\,q^{-1}y\,(q-r)\,.\end{aligned}$$

Writing $\alpha = r/p = \Pr\big[J=1 \mid I=1\big]$ and $\beta = r/q = \Pr\big[I=1 \mid J=1\big]$, we find

$$\ell_{\alpha,\beta}(x,y) \,=\, \ell(\alpha x, \beta y) + (1-\alpha)\,x + (1-\beta)\,y\,, \qquad (x,y)\in[0,\infty)^2\,. \tag{3.6}$$

The new tail copula is simply

$$\begin{aligned}R_{\alpha,\beta}(x,y) \,&=\, x + y - \ell_{\alpha,\beta}(x,y)\\ &=\,\alpha\,x + \beta\,y - \ell(\alpha x, \beta y) \,=\, R(\alpha x, \beta y)\,, \qquad (x,y)\in[0,\infty)^2\,.\end{aligned}$$

This is an asymmetric, two-parameter extension of the original model. Imposing the equality constraint $\alpha = \beta = \theta \in [0,1]$ yields the symmetric, one-parameter extension

$$\ell_\theta(x,y) \,=\, \theta\,\ell(x,y) + (1-\theta)\,(x+y)\,. \tag{3.7}$$

In higher dimensions, a vector of indicators $(I_1,\ldots,I_d)$ can serve to switch some components X_j 'on' or 'off'. The dependence structure in these indicators then yields an extremal dependence structure for the random vector X. Specifically, let $p_j = \Pr[I_j = 1]$; we assume $0 < p_j \le 1$. Let $\boldsymbol{A} = (A_1,\ldots,A_d)$ be a random vector independent of $(I_1,\ldots,I_d)$ and such that $\mathrm{E}[\max(A_j,0)] = 1$ for all $j = 1,\ldots,d$. Then we can define a stable tail dependence function via

$$\begin{aligned}\ell_{\boldsymbol{p}}(x_1,\ldots,x_d) \,&=\, \mathrm{E}\Big[\max\big(p_1^{-1}x_1I_1A_1,\,\ldots,\,p_d^{-1}x_dI_dA_d,\,0\big)\Big]\\ &=\,\sum_{\emptyset\neq c\subset\{1,\ldots,d\}} p(c)\;\mathrm{E}\Big[\max\big(p_j^{-1}x_j\,A_j:\;j\in c\big)\Big]\end{aligned} \tag{3.8}$$

where $p(c) = \Pr\big[\{j = 1, \ldots, d : I_j = 1\} = c\big]$. Keeping the law of $\boldsymbol{A}$ fixed, the parameter vector $\boldsymbol{p}$ is equal to a probability distribution $(p(c))_c$ on the non-empty subsets c of $\{1, \ldots, d\}$.

In this way, hierarchical structures can easily be built. For instance, in dimension $d = 4$, we can think of $(I_1, I_2, I_3, I_4) = (J, J, K, K)$. One can also think of logit-type models for the indicators.

Example 3.5 (Marshall–Olkin copula)**.** Applying the device in (3.6) to the function $\ell(x, y) = \max(x, y)$ yields the model

$$\ell_{\alpha,\beta}(x, y) \;=\; \max(\alpha x, \beta y) + (1 - \alpha)\, x + (1 - \beta)\, y \;=\; x + y - \min(\alpha x, \beta y) \;.$$

The extreme-value copula associated to $\ell_{\alpha,\beta}$ is the Marshall–Olkin copula

$$C_{\alpha,\beta}(u, v) \;=\; uv \, \min(u^{-\alpha}, v^{-\beta}) \;=\; \min(u^{1-\alpha} v,\, u v^{1-\beta}) \;, \qquad (u, v) \in [0, 1]^2 \,.$$

In higher dimensions, applying the device (3.8) to the function $\ell(\boldsymbol{x}) = \max(\boldsymbol{x})$, that is, $A_j = 1$ with probability one, we get the model

$$\ell_{\boldsymbol{p}}(x_1, \ldots, x_d) \;=\; \sum_{\emptyset \neq c \subset \{1,\ldots,d\}} p(c) \, \max\big(p_j^{-1} x_j : j \in c\big) \;,$$

the spectral measure of which is discrete. Another stochastic interpretation of this model is provided in Embrechts *et al.* (2003).

Example 3.6 (Dirichlet model)**.** Let $\alpha_1, \ldots, \alpha_d$ be positive parameters and let $Z_1, \ldots, Z_d$ be independent Gamma$(\alpha_j, 1)$ random variables, that is, the density of Z_j is

$$f_j(z) \;=\; \frac{1}{\Gamma(\alpha_j)} \; z^{\alpha_j - 1} e^{-z} \;, \qquad 0 < z < \infty \;.$$

Put $A_j = \alpha_j^{-1} Z_j$, a positive random variable with unit expectation. The stable tail dependence function associated with the random vector $\boldsymbol{A} = (A_1, \ldots, A_d)$ is

$$\begin{aligned} \ell_{\boldsymbol{A}}(x_1, \ldots, x_d) \;&=\; \mathrm{E}\Big[\max\big(x_1 A_1, \ldots, x_d A_d\big)\Big] \\ &=\; \mathrm{E}\Big[\max\big(\alpha_1^{-1} x_1 Z_1, \ldots, \alpha_d^{-1} x_d Z_d\big)\Big] \\ &=\; \mathrm{E}\Big[(Z_1 + \cdots + Z_d) \, \max\big(\alpha_1^{-1} x_1 V_1, \ldots, \alpha_d^{-1} x_d V_d\big)\Big] \,, \end{aligned}$$

where $V_j = Z_j / (Z_1 + \cdots + Z_d)$. It is well known that the random vector $(V_1, \ldots, V_d)$ is independent of $Z_1 + \cdots + Z_d$ and has a Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_d)$. We find

$$\begin{aligned} \ell_{\boldsymbol{A}}(\boldsymbol{x}) \;&=\; \mathrm{E}\Big[(\alpha_1 + \cdots + \alpha_d) \, \max\big(\alpha_1^{-1} x_1 V_1, \ldots, \alpha_d^{-1} x_d V_d\big)\Big] \\ &=\; \frac{\Gamma\big(\sum_{j=1}^d \alpha_j + 1\big)}{\prod_{j=1}^d \Gamma(\alpha_j)} \int_{\Delta_{d-1}} \max_{j=1,\ldots,d} \big(\alpha_j^{-1} x_j \, v_j\big) \, \prod_{j=1}^d v_j^{\alpha_j - 1} \, \mathrm{d}v_1 \cdots \mathrm{d}v_{d-1} \,. \end{aligned} \tag{3.9}$$

We recognize the Dirichlet model introduced in Coles & Tawn (1991).

Example 3.7 (Polynomial Pickands functions)**.** In the Dirichlet model (3.9), put $d = 2$ and $\alpha_1 = \alpha_2 = 1$, to obtain

$$\ell(x,y) \;=\; 2\int_0^1 \max\bigl\{x\,v,\; y(1-v)\bigr\}\,\mathrm{d}v \;=\; x + y - \frac{x\,y}{x+y}\;.$$

The Pickands dependence function associated to ℓ is $D(t) = \ell(1-t,t) = 1-t(1-t)$ for $t \in [0,1]$. Applying the transformation in (3.7) yields the mixture model

$$D(t) \;=\; 1 - \theta\,t\,(1-t)\;, \qquad t \in [0,1]\;,$$

in terms of a parameter $\theta \in [0,1]$, introduced in Tawn (1988). This is the only model in which D is a quadratic polynomial. Applying the transformation in (3.6) yields the rational model

$$\begin{aligned} D_{\alpha,\beta}(t) \;&=\; \ell\bigl(\alpha(1-t),\beta t\bigr) + (1-\alpha)\,(1-t) + (1-\beta)t \\ &=\; 1 - \frac{\alpha\,\beta\,t(1-t)}{\alpha(1-t)+\beta t}\;, \qquad t \in [0,1]\;, \end{aligned}$$

in terms of parameters $\alpha, \beta \in (0,1]$.

More generally, bivariate max-stable models of which the Pickands dependence function is a higher-order polynomial can be obtained via the transformation in (3.7) applied to the function $\ell(x,y) = \mathrm{E}[\max(xA, yB)]$ when A and B are random sums

$$A \;=\; E_1 + \cdots + E_J\;, \qquad B \;=\; F_1 + \cdots + F_K\;,$$

in terms of independent random variables $J, K, E_1, F_1, E_2, F_2, \ldots$ such that J and K are counting random variables having finite support and unit expectation and where $E_1, F_1, E_2, F_2, \ldots$ are unit exponential random variables. Polynomial Pickands dependence functions are studied in Guillotte & Perron (2012).

Example 3.8 (Schlather model)**.** Let (S,T) be a bivariate normal random vector with standard normal margins and with correlation $\rho \in (-1,1)$. Put $A = \sqrt{2\pi}\,S$ and $B = \sqrt{2\pi}\,T$, so that $\mathrm{E}[\max(A,0)] = \mathrm{E}[\max(B,0)] = 1$. The stable tail dependence function corresponding to the random pair (A,B) is

$$\begin{aligned} \ell_\rho(x,y) \;&=\; \mathrm{E}\bigl[\max(xA, yB, 0)\bigr] \\ &=\; \int_{s=-\infty}^{0}\int_{t=0}^{\infty} \sqrt{2\pi}\,y t\;\frac{1}{2\pi}\,e^{-(s^2+2\rho st+t^2)/2}\,\mathrm{d}s\,\mathrm{d}t \\ &\quad + \int_{s=0}^{\infty}\int_{t=-\infty}^{0} \sqrt{2\pi}\,x s\;\frac{1}{2\pi}\,e^{-(s^2+2\rho st+t^2)/2}\,\mathrm{d}s\,\mathrm{d}t \\ &\quad + \int_{0}^{\infty}\int_{0}^{\infty} \sqrt{2\pi}\;\max(xs, yt)\;\frac{1}{2\pi}\,e^{-(s^2+2\rho st+t^2)/2}\,\mathrm{d}s\,\mathrm{d}t\;. \end{aligned}$$

After some tedious calculations, this can be shown to be

$$\ell_\rho(x,y) \,=\, \frac{1}{2}\,(x+y)\left(1+\sqrt{1-2\,(\rho+1)\,\frac{x\,y}{(x+y)^2}}\right),$$

see Schlather (2002). The Pickands dependence function corresponding to this model is

$$D_\rho(t) \,=\, \frac{1}{2}\left\{1+\sqrt{1-2\,(\rho+1)\,t\,(1-t)}\right\}, \qquad t\in[0,1]\,.$$

Example 3.9 (Hüsler–Reiss model)**.** Let (S,T) be a bivariate normal random vector with standard normal margins and with correlation $\rho\in(-1,1)$. Let $\sigma>0$ and put

$$A \,=\, \exp\bigl(\sigma S-\sigma^2/2\bigr)\,, \qquad B \,=\, \exp\bigl(\sigma T-\sigma^2/2\bigr)\,.$$

The pair (A,B) is lognormal with unit expectations, and it yields the stable tail dependence function

$$\ell(x,y) \,=\, \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\max\bigl\{x\,e^{\sigma s},\,y\,e^{\sigma t}\bigr\}\,e^{-\sigma^2/2}\,\frac{1}{2\pi}\,e^{-(s^2+2\rho st+t^2)/2}\,\mathrm{d}s\,\mathrm{d}t\,.$$

The double integral can be calculated in terms of the standard normal cumulative distribution function Φ, yielding

$$\ell_a(x,y) \,=\, x\,\Phi\left(\frac{a}{2}+\frac{1}{a}\,\log(x/y)\right)+\,y\,\Phi\left(\frac{a}{2}+\frac{1}{a}\,\log(y/x)\right)$$

with parameter $a=\sigma\sqrt{2\,(1-\rho)}$. This is the model introduced in Hüsler & Reiss (1989).

ACKNOWLEDGMENTS

The author wishes to thank the journal Editors for the invitation to write this paper and for a detailed proofreading of an earlier version of the manuscript. The paper originates from a lecture given at the workshop on Copula Models and Dependence at the *Centre de Recherches Mathématiques*, June 6–9, 2011, in Montréal. Organizers and participants of the workshop are gratefully acknowledged for their support and their encouraging discussions, from which the paper has strongly benefited.

The author's research was supported by IAP research network grant No. P6/03 of the Belgian government (Belgian Science Policy) and by contract No. 07/12/002 of the *Projet d'Actions de Recherche Concertées* of the *Communauté française de Belgique*, granted by the *Académie universitaire Louvain*.

REFERENCES

Ballani, F. & Schlather, M. (2011). A construction principle for multivariate extreme value distributions, *Biometrika*, **98**, 633–645.

Basrak, B. & Segers, J. (2009). Regularly varying multivariate time series, *Stoch. Proc. Appl.*, **119**, 1055–1080.

Beirlant, J.; Goegebeur, Y.; Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Chichester.

Boldi, M.-O. & Davison, A.C. (2007). A mixture model for multivariate extremes, *J. R. Statist. Soc. B*, **69**, 217–229.

Coles, S. & Tawn, J. (1991). Modelling extreme multivariate events, *J. R. Statist. Soc. B*, **53**, 377–392.

Cooley, D.; Davis, R. & Naveau, P. (2010). The pairwise beta distribution: a flexible parametric multivariate model for extremes, *J. Mult. Anal.*, **101**, 2103–2117.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer, New York.

de Haan, L. & Resnick, S. (1977). Limit theory for multidimensional sample extremes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **40**, 317–337.

Deheuvels, P. (1984). *Probabilistic aspects of multivariate extremes*. In "Statistical Extremes and Applications" (J.T. de Oliveira, Ed.), pp. 117–130, Reidel, Dordrecht.

Drees, H. & Huang, X. (1998). Best attainable rates of convergence for estimates of the stable tail dependence function, *J. Mult. Anal.*, **64**, 25–47.

Einmahl, J.H.J.; Krajina, A. & Segers, J. (2011). *An M-estimator for tail dependence in arbitrary dimensions*, Technical Report 2011/05, Université catholique de Louvain, ISBA. arXiv:1112.0905v1 [math.ST].

Embrechts, P.; Lindskog, F. & McNeil, A. (2003). *Modelling dependence with copulas and applications to risk management*. In "Handbook of Heavy Tailed Distributions in Finance" (S. Rachev, Ed.), Chapter 8, pp. 329–384, Elsevier.

Falk, M.; Hüsler, J. & Reiss, R.-D. (2010). *Laws of Small Numbers: Extremes and Rare Events* (3rd ed.), Birkhäuser, Basel.

Fougères, A.-L.; Nolan, J.P. & Rootzén, H. (2009). Models for dependent extremes using stable mixtures, *Scand. J. Statist.*, **36**, 42–59.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd ed.), Robert E. Krieger Publishing Co. Inc., Melbourne, FL.

Garralda-Guillem, A.I. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées, *C. R. Acad. Sci. Paris, Série I Math.*, **330**, 593–596.

Gudendorf, G. & Segers, J. (2010). *Extreme-value copulas*. In "Copula theory and its applications (Warsaw, 2009)" (P. Jaworski, F. Durante, W. Härdle and W. Rychlik, Eds.), Lecture Notes in Statistics – Proceedings, pp. 127–146, Springer-Verlag, Berlin.

Guillotte, S. & Perron, F. (2012). *Polynomial Pickands functions: characterisation and inference*. In preparation.

Gumbel, E.J. (1960). Bivariate exponential distributions, *J. Am. Statist. Assoc.*, **55**, 698–707.

Hougaard, P. (1986). A class of multivariate failure time distributions, *Biometrika*, **73**, 671–678.

Huang, X. (1992). *Statistics of bivariate extreme values*, Ph. D. thesis, Tinbergen Institute Research Series.

Hüsler, J. & Reiss, R. (1989). Maxima of normal random vectors: between independence and complete dependence, *Statist. Prob. Lett.*, **7**, 283–286.

Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

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

Molchanov, I. (2008). Convex geometry of max-stable distributions, *Extremes*, **11**, 235–259.

Pickands, J. (1981). *Multivariate extreme value distributions*. In "Proc. 43rd Int. Statist. Inst., Vol. 2 (Buenos Aires, 1981)", Volume 49, pp. 859–878, 894–902. With a discussion.

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

Schlather, M. (2002). Models for stationary max-stable random fields, *Extremes*, **5**, 33–44.

Schmidt, R. & Stadtmüller, U. (2006). Non-parametric estimation of tail dependence, *Scand. J. Statist.*, **33**, 307–335.

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

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 83–107

BIVARIATE EXTREME STATISTICS, II

Authors: Miguel de Carvalho
– Swiss Federal Institute of Technology,
Ecole Polytechnique Fédérale de Lausanne, Switzerland
miguel.carvalho@epfl.ch

Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, Portugal

Alexandra Ramos
– Universidade do Porto, Faculdade de Economia, Portugal
aramos@fep.up.pt

Abstract:

- We review the current state of statistical modeling of asymptotically independent data. Our discussion includes necessary and sufficient conditions for asymptotic independence, results on the asymptotic independence of statistics of interest, estimation and inference issues, joint tail modeling, and conditional approaches. For each of these topics we give an account of existing approaches and relevant methods for data analysis and applications.

Key-Words:

- *asymptotic independence; coefficient of tail dependence; conditional tail modeling; extremal dependence; hidden regular variation; joint tail modeling; order statistics; maximum; multivariate extremes; sums.*

AMS Subject Classification:

- 60G70, 62E20.

1. INTRODUCTION

The concept of *asymptotic independence* connects two central notions in probability and statistics: asymptotics and independence. Suppose that X and Y are identically distributed real-valued random variables, and that our interest is in assessing the probability of a joint tail event $(X > u, Y > u)$, where u denotes a high threshold. We say that (X, Y) is asymptotically independent, $X \overset{\text{a. ind.}}{\sim} Y$, if

$$\lim_{u\to\infty} \operatorname{pr}\big(X > u \mid Y > u\big) \;=\; \lim_{u\to\infty} \frac{\operatorname{pr}\big(X > u, Y > u\big)}{\operatorname{pr}\big(Y > u\big)} \;=\; 0 \;. \tag{1.1}$$

Intuitively, condition (1.1) implies that given that the decay of the joint distribution is faster than the marginals, it is unlikely that the largest values of X and Y happen simultaneously.[1] Whereas independence is unrealistic for many data applications, there has been a recent understanding that when modeling extremes, asymptotic independence is often found in real data. It may seem surprising that although the problem of testing asymptotic independence is an old goal in statistics (Gumbel & Goldstein, 1964), only recently there has been an understanding that classical models for multivariate extremes are unable to deal with it.

In this paper we review the current state of statistical modeling of asymptotically independent data. Our discussion includes a list of important topics, including necessary and sufficient conditions, results on the asymptotic independence of statistics of interest, estimation and inference issues, and joint tail modeling. We also provide our personal view on some directions we think could be of interest to be explored in the coming years. Our discussion is not exhaustive, and in particular there are many results of probabilistic interest, on asymptotic independence of other statistics not relevant to extreme value analyses, which are not discussed here.

The title of this paper is based on the seminal work of Sibuya (1960), entitled "Bivariate Extreme Statistics, I" which presents necessary and sufficient conditions for the asymptotic independence of the two largest extremes in a bivariate distribution. Sibuya mentions that a practical application should be "considered in a subsequent paper" which to our knowledge never appeared.

Other recent surveys on asymptotic independence include Resnick (2002) and Beirlant *et al.* (2004, §9). The former mostly explores connections with hidden regular variation and multivariate second order regular variation.

[1] To be precise, the tentative definition in (1.1) corresponds simply to a particular instance of the concept, i.e., asymptotic independence of the largest extremes in a bivariate distribution. Although this is the version of the concept to which we devote most of our attention, the concept of asymptotic independence is actually broader, and has also been studied for many other pairs of statistics, other than bivariate extremes, even in the field of extremes; we revisit some examples in §6.

2. ASYMPTOTIC INDEPENDENCE—CHARACTERIZATIONS

2.1. Necessary and sufficient conditions for asymptotic independence

Early developments on asymptotic independence of the two largest extremes in a bivariate distribution, were mostly devoted to obtaining necessary or sufficient characterizations for asymptotic independence (Finkelstein, 1953; Geffroy, 1958, 1959; Sibuya, 1960; Berman, 1961; Ikeda, 1963; Mikhailov, 1974; Galambos, 1975; de Haan & Resnick, 1977; Marshall & Olkin, 1983; Takahashi, 1994).

Geffroy (1958) showed that the condition

$$\lim_{x,y\to\infty} \frac{\overline{C}\big\{F_X(x), F_Y(y)\big\}}{1 - F_{X,Y}(x,y)} \;=\; 0 \tag{2.1}$$

is sufficient for asymptotic independence, where the operator

$$\begin{aligned} \overline{C}\big\{F_X(x), F_Y(y)\big\} \;&\equiv\; \mathrm{pr}\big(X > x, Y > y\big) \\ &=\; 1 + F_{X,Y}(x,y) - F_X(x) - F_Y(y)\,, \qquad (x,y) \in \mathbb{R}^2\,, \end{aligned} \tag{2.2}$$

maps a pair of marginal distribution functions to their joint tails. We prefer to state results using a copula, i.e., a function $C\colon [0,1]^2 \to [0,1]$, such that

$$C(p,q) \;=\; F_{X,Y}\big\{F_X^{-1}(p), F_Y^{-1}(q)\big\}\,, \qquad (p,q) \in [0,1]^2\,.$$

Here $F_\cdot^{-1}(\cdot) = \inf\big\{x\colon F_\cdot(x) \geq \cdot \in [0,1]\big\}$, and the uniqueness of the function C for a given pair of joint and marginal distributions follows by Sklar's theorem (Sklar, 1959). Geffroy's condition can then be rewritten as

$$\lim_{p,q\uparrow 1} \frac{\overline{C}(p,q)}{1 - C(p,q)} \;=\; \lim_{p,q\uparrow 1} \frac{1 + C(p,q) - p - q}{1 - C(p,q)} \;=\; 0\,. \tag{2.3}$$

Example 2.1. Examples of dependence structures obeying condition (2.3) can be found in Johnson & Kotz (1972, §41), and include any member of the Farlie–Gumbel–Morgenstern family of copulas

$$C_\alpha(p,q) \;=\; p\,q\big\{1 + \alpha\,(1-p)\,(1-q)\big\}\,, \qquad \alpha \in [-1,1]\,,$$

and the copulas of the bivariate exponential and bivariate logistic distributions (Gumbel, 1960, 1961), respectively given by

$$\begin{aligned} C_\theta(p,q) \;&=\; p + q - 1 + (1-p)\,(1-q)\,\exp\big\{-\theta \log(1-p)\log(1-q)\big\}\,, \qquad \theta \in [0,1]\,, \\ C(p,q) \;&=\; \frac{p\,q}{p + q - p\,q}\,, \qquad (p,q) \in [0,1]^2\,. \end{aligned}$$

Sibuya (1960) introduced a condition related to (2.1)

$$\lim_{q\uparrow 1} \frac{\overline{C}(q,q)}{1-q} = 0 \,, \tag{2.4}$$

and showed that this is necessary and sufficient for asymptotic independence. Condition (2.4) is simply a reformulation of (1.1) which describes the rate at which we start lacking observations in the joint tails, as we move towards higher quantiles. Sibuya used condition (2.4) to observe that bivariate normal distributed vectors with correlation $\rho < 1$ are asymptotically independent, and similar results are also inherited by light-tailed elliptical densities (Hult & Lindskog, 2002).

Often the question arises on whether it is too restrictive to study asymptotic independence only for the bivariate case. This question was answered long ago by Berman (1961), who showed that a d-dimensional random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)$, with a regularly varying joint tail (Bingham *et al.*, 1987), is asymptotically independent if, and only if,

$$Z_i \overset{\text{a. ind.}}{\sim} Z_j \,, \qquad i \neq j \,.$$

Asymptotic independence in a d-vector is thus equivalent to pairwise asymptotic independence.[2] This can also be shown to be equivalent to having the exponent measure put null mass on the interior of the first quadrant, and to concentrate on the positive coordinate axes, or equivalently to having all the mass of the spectral measure concentrated on 0 and 1; definitions of the spectral and exponent measures are given in Beirlant *et al.* (2004, §8), and a formal statement of this result can be found in Resnick (1987, Propositions 5.24–25). In theory, this allows us to restrict the analysis to the bivariate case, so we confine the exposition to this setting. Using the result of Berman (1961) we can also state a simple necessary and sufficient condition, analogous to (2.4), for asymptotic independence of $\mathbf{Z} = (Z_1, \ldots, Z_d)$, i.e.,

$$\lim_{q\uparrow 1} \sum_{i=1}^{d} \sum_{\substack{j=1 \\ (j\neq i)}}^{d} \frac{\overline{C}_{ij}(q,q)}{1-q} = 0 \,, \qquad \overline{C}_{ij}(p,q) \equiv 1 + C_{ij}(p,q) - p - q \,, \quad (p,q) \in [0,1]^2 \,,$$

with the obvious notations (Mikhailov, 1974, Theorem 2).

Example 2.2. Consider the copula of bivariate logistic distribution in Example 2.1. Sibuya's condition (2.4) follows directly:

$$\lim_{q\uparrow 1} \frac{\overline{C}(q,q)}{1-q} = \lim_{q\uparrow 1} \frac{2(q-1)^2}{2-q} = 0 \,.$$

[2]The pairwise structure is however insufficient to determine the higher order structure; e.g., in general not much can be inferred on $\mathrm{pr}\big(X > x, Y > y, Z > z\big)$, from the pairs.

The characterizations in (1.1) and (2.1) are population-based, but a limiting sample-based representation can also be given, using the random sample $\{(X_i, Y_i)\}_{i=1}^n$, so that asymptotic independence is equivalent to

$$\lim_{n\to\infty} C^n\big(p^{1/n}, q^{1/n}\big) = p\,q\ , \qquad (p,q)\in[0,1]^2\,. \tag{2.5}$$

In words: the copula of the distribution function of the sample maximum $M_n = \max\big\{(X_1,Y_1),\ldots,(X_n,Y_n)\big\}$, where the maximum are taken componentwise, converges to the product copula $C_\pi = p\,q$; equivalently we can say that the extreme-value copula, $\lim_{n\to\infty} C^n\big(p^{1/n}, q^{1/n}\big)$, is C_π, or that C is in the domain of attraction of C_π.

Srivastava (1967) and Mardia (1964) studied results on asymptotic independence on bivariate samples, but for other order statistics, rather than the maximum. Consider a random sample $\{(X_i, Y_i)\}_{i=1}^n$ and the order statistics $X_{1:n} \le \cdots \le X_{n:n}$ and $Y_{1:n} \le \cdots \le Y_{n:n}$. It can be shown that if $(X_{1:n}, Y_{1:n})$ is asymptotically independent, then

$$X_{i:n} \overset{\text{a. ind.}}{\sim} Y_{j:n}\ , \qquad i,j\in\{1,\ldots,n\}\ .$$

See Srivastava (1967, Theorem 3).

The last characterization of asymptotic independence we discuss is due to Takahashi (1994). According to Takahashi's criterion, asymptotic independence is equivalent to

$$\exists\,(a,b)\in(0,\infty)^2\colon\ \ \ell(a,b) \equiv \lim_{q\uparrow 1} \frac{1-C\big\{1-a(1-q),1-b(1-q)\big\}}{1-q} = a+b\,. \tag{2.6}$$

Example 2.3. A simple analytical example to verify Takahashi's criterion is given by taking the bivariate logistic copula and checking that $\ell(1,1) = 2$.

Remark 2.1. The function $\ell(a,b)$ is the so-called stable tail dependence function, and as shown in Beirlant *et al.* (2004, p. 286), condition (2.6) is equivalent to

$$\ell(a,b) = a+b\ , \qquad (a,b)\in[0,\infty)\ .$$

2.2. Notes and comments

Some of the results obtained in Finkelstein (1953) were 'rediscovered' in later papers. Some of these include results proved by Galambos (1975), who claims that Finkelstein (1953) advanced his results without giving formal proofs.

Tiago de Oliveira (1962/63) is also acknowledged for pioneering work in statistical modeling of asymptotic independence of bivariate extremes. Mikhailov (1974) and Galambos (1975) obtained a necessary and sufficient condition for d-dimensional asymptotic independence of arbitrary extremes; a related characterization can also be found in Marshall & Olkin (1983, Proposition 5.2)

Most of the characterizations discussed above are directly based on distribution functions and copulas, but it seems natural to infer asymptotic independence from contours of the joint density. Balkema & Nolde (2010) establish sufficient conditions for asymptotic independence, for some homothetic densities, i.e., densities whose level sets all have the same shape. In particular, they show that the components of continuously differentiable homothetic light-tailed distributions with convex levels sets are asymptotically independent; in their Corollary 2.1 Balkema and Nolde also show that asymptotic independence resists quite notable distortions in the joint distribution.

Measures of asymptotic dependence for further order statistics are studied in Ferreira & Ferreira (2012).

2.3. Dual measures of extremal dependence: $(\chi, \overline{\chi})$

Many measures of dependence, such as the Pearson correlation coefficient, Spearman rank correlation, and Kendall's tau, can be written as functions of copulae (Schweizer & Wolff, 1981, p. 879), and as we discuss below, measures of extremal dependence can also be conceptualized as functions of copulae.

To measure extremal dependence we first need to convert the data $(\mathcal{X}, \mathcal{Y})$ to a common scale. The rescaled variables (X, Y) are transformed to have unit Fréchet margins, i.e., $F_X(z) = F_Y(z) = \exp(-1/z)$, $z > 0$; this can be done with the mapping

$$(\mathcal{X}, \mathcal{Y}) \mapsto (X, Y) = -\Big(\big\{ \log F_{\mathcal{X}}(\mathcal{X}) \big\}^{-1}, \big\{ \log F_{\mathcal{Y}}(\mathcal{Y}) \big\}^{-1} \Big) . \tag{2.7}$$

Since the rescaled variables have the same marginal distribution, any remaining differences between distributions can only be due to dependence features (Embrechts *et al.*, 2002). A natural measure to assess the degree of dependence at an arbitrary high level $\tau < \infty$, is the bivariate tail dependence index

$$\chi = \lim_{u \to \infty} \mathrm{pr}\big(X > u \mid Y > u \big) = \lim_{q \uparrow 1} \mathrm{pr}\big\{ X > F_X^{-1}(q) \mid Y > F_Y^{-1}(q) \big\} . \tag{2.8}$$

This measure takes values in $[0, 1]$, and can be used to assess the degree of dependence that remains in the limit (Coles *et al.*, 1999; Poon *et al.*, 2003, 2004).

If dependence persists as $u \to \infty$, then $0 < \chi \leq 1$ and X and Y are said to be asymptotically dependent; otherwise, the degree of dependence vanishes in the limit, so that $\chi = 0$ and the variables are asymptotically independent. The measure χ can also be rewritten in terms of the limit of a function of the copula C, by noticing that

$$\chi = \lim_{q \uparrow 1} \chi(q)\,, \quad \chi(q) = 2 - \frac{\log C(q,q)}{\log q}\,, \qquad 0 < q < 1\,. \tag{2.9}$$

Thus, the function C 'couples' the joint distribution function and its corresponding marginals, and it also provides helpful information for modeling joint tail dependence. The function $\chi(q)$ can be understood as a quantile dependent measure of dependence, and the sign of $\chi(q)$ can be used to ascertain if the variables are positively or negatively associated at the quantile q. As a consequence of the Fréchet–Hoeffding bounds (Nelsen, 2006, §2.5), the level of dependence is bounded,

$$2 - \frac{\log(2\,q-1)_+}{\log q} \;\leq\; \chi(q) \;\leq\; 1\,, \qquad 0 < q < 1\,, \tag{2.10}$$

where $a_+ = \max(a, 0)$, $a \in \mathbb{R}$. Extremal dependence should be measured according to the dependence structure underlying the variables under analysis. If the variables are asymptotically dependent, the measure χ is appropriate for assessing the strength of dependence which links the variables at the extremes. If however the variables are asymptotically independent then $\chi = 0$, so that χ pools cases where although dependence may not prevail in the limit, it may persist for relatively large levels of the variables. To measure extremal dependence under asymptotic independence, Coles *et al.* (1999) introduced the measure

$$\overline{\chi} \;=\; \lim_{u \to \infty} \frac{2 \log \mathrm{pr}\big(X > u\big)}{\log \mathrm{pr}\big(X > u, Y > u\big)} - 1\,, \tag{2.11}$$

which takes values on the interval $(-1, 1]$. The interpretation of $\overline{\chi}$ is to a certain extent analogous to that of the Pearson correlation: values of $\overline{\chi} > 0$, $\overline{\chi} = 0$ and $\overline{\chi} < 0$, respectively correspond to positive association, exact independence and negative association in the extremes, and if the dependence structure is Gaussian then $\overline{\chi} = \rho$ (Sibuya, 1960). This benchmark case is particularly helpful for guiding how does the dependence in the tails, as measured by $\overline{\chi}$, compares with that arising from fitting a Gaussian dependence model.

Asymptotic dependence and asymptotic independence can also be characterized through $\overline{\chi}$. For asymptotically dependent variables, it holds that $\overline{\chi} = 1$, while for asymptotically independent variables $\overline{\chi}$ takes values in $(-1, 1)$. Hence χ and $\overline{\chi}$ can be seen as dual measures of joint tail dependence: if $\overline{\chi} = 1$ and $0 < \chi \leq 1$, the variables are asymptotically dependent, and χ assesses the degree of dependence within the class of asymptotically dependent distributions; if $-1 < \overline{\chi} < 1$ and $\chi = 0$, the variables are asymptotically independent, and

$\overline{\chi}$ assesses the degree of dependence within the class of asymptotically independent distributions. In a similar way to (2.9), the extremal measure $\overline{\chi}$ can also be written using copulas, viz.

$$\overline{\chi} = \lim_{q \uparrow 1} \overline{\chi}(q) \ , \qquad \overline{\chi}(q) = \frac{2 \log(1-q)}{\log \overline{C}(q,q)} \ . \tag{2.12}$$

Hence, the function C can provide helpful information for assessing dependence in extremes both under asymptotic dependence and asymptotic independence. The function $\overline{\chi}(q)$ has an analogous role to $\chi(q)$, in the case of asymptotic independence, and it can also be used as quantile dependent measure of dependence, with the following Fréchet–Hoeffding bounds:

$$\frac{2 \log(1-q)}{\log(1-2\,q)_+} - 1 \ \leq \ \overline{\chi}(q) \ \leq \ 1 \ , \qquad 0 < q < 1 \ . \tag{2.13}$$

For an inventory of the functional forms of the extremal measures $\overline{\chi}$ and χ, over several dependence models, see Heffernan (2000). We remark that the dual measures $(\chi, \overline{\chi})$ can be reparametrized as

$$(\chi, \overline{\chi}) \ = \ (2-\theta, 2\,\eta - 1) \ , \tag{2.14}$$

where $\theta = \lim_{q \uparrow 1} \log C(q,q) / \log q$ is the so-called extremal coefficient, and η is the coefficient of tail dependence to be discussed in §3–4.

3. ESTIMATION AND INFERENCE

3.1. Coefficient of tail dependence-based approaches

The coefficient of tail dependence η corresponds to the extreme value index of the variable $Z = \min\{X, Y\}$, which characterizes the joint tail behavior above a high threshold u (Ledford & Tawn, 1996). The formal details are described in §4, but the heuristic argument follows by the simple observation that

$$\mathrm{pr}\big(Z > u\big) \ = \ \mathrm{pr}\big(X > u, Y > u\big) \ ,$$

and hence we reduce a bivariate problem to a univariate one. This implies that we can use the order statistics of the $Z_i = \min\{X_i, Y_i\}$, $Z_{(1)} \leq \cdots \leq Z_{(n)}$, to estimate η by applying univariate estimation methods, such as the Hill estimator

$$\widehat{\eta}_k \ = \ \frac{1}{k} \sum_{i=1}^{k} \big\{ \log Z_{(n-k+i)} - \log Z_{(n-k)} \big\} \ .$$

By estimating η directly with univariate methods we are however underestimating its uncertainty, since we ignore the uncertainty from transforming the data to equal margins, say by using (2.7). The estimators of Peng (1999), Draisma *et al.* (2004), Beirlant & Vandewalle (2002), can be used to tackle this, and a review of these methods can be found in Beirlant *et al.* (2004, pp. 351–353).

3.2. Score-based tests

Tawn (1988) and Ledford & Tawn (1996) proposed score statistics for examining independence within the class of multivariate extreme value distributions. Ramos & Ledford (2005) proposed modified versions of such tests which solve the problem of slow rate of convergence of such tests, due to infinite variance of the scores. Consider the following partition of the outcome space $\mathbb{R}_+^2$, given by

$$R_{kl} \;=\; \Big\{(x,y)\colon\; k = I(x>u),\; l = I(y>u)\Big\}\,, \qquad k,l \in \{0,1\}\;,$$

where u denotes a high threshold and I denotes the indicator function. The approach of Ramos and Ledford is based on censoring the upper tail R_{11} for a high threshold u, so that, using the logistic dependence structure, the score functions at independence of Tawn (1988) and Ledford & Tawn (1996) are respectively given by

$$U_n^1 \;=\; \sum_{(X_i,Y_i)\notin R_{11}} \Delta_1(X_i,Y_i) + \Lambda\;, \qquad U_n^2 \;=\; \sum_{(X_i,Y_i)\notin R_{11}} \Delta_2(X_i,Y_i) + \Lambda\;,$$

where

$$\begin{aligned}
\Delta_1(X_i,Y_i) \;&=\; (1-X_i^{-1})\log X_i + (1-Y_i^{-1})\log Y_i \\
&\quad +\; (2-X_i^{-1}-Y_i^{-1})\log(X_i^{-1}+Y_i^{-1}) - (X_i^{-1}+Y_i^{-1})^{-1}\;, \\
\Delta_2(X_i,Y_i) \;&=\; I\big\{(X_i,Y_i)\in R_{kl}\big\}\; S_{kl}(X_i,Y_i)\;, \\
\Lambda \;&=\; \frac{2\,u^{-1}\log 2\exp(-2\,u^{-1})\,N}{2\exp(-u^{-1}) - \exp(-2\,u^{-1}) - 1}\;,
\end{aligned}$$

with N denoting the number of observations in region R_{11}, and

$$\begin{aligned}
S_{00}(x,y) \;&=\; -2\,u^{-1}\log 2\;, \\
S_{01}(x,y) \;&=\; -u^{-1}\log u + (1-y^{-1})\log y + (1-u^{-1}-y^{-1})\log(u^{-1}+y^{-1})\;, \\
S_{10}(x,y) \;&=\; -u^{-1}\log u + (1-x^{-1})\log x + (1-x^{-1}-u^{-1})\log(x^{-1}+u^{-1})\;, \\
S_{11}(x,y) \;&=\; (1-x^{-1})\log x + (1-y^{-1})\log y + (2-x^{-1}-y^{-1})\log(x^{-1}+y^{-1}) \\
&\quad -\; (x^{-1}+y^{-1})^{-1}\;.
\end{aligned}$$

The modified score functions U_n^1 and U_n^2 have zero expectation and finite second moments. The limit distributions under independence are then given as

$$-n^{-1/2}\,\frac{U_n^i}{\sigma_i} \;\xrightarrow{\;\mathrm{d}\;}\; N(0,1)\;, \qquad n\to\infty\,, \quad i=1,2\;,$$

where $\xrightarrow{\;\mathrm{d}\;}$ denotes convergence in distribution and σ_i denotes the variance of the corresponding modified score statistics; we remark that these score tests typically reject independence when evaluated on asymptotically independent data.

3.3. Falk–Michel test

Falk & Michel (2006) proposed tests for asymptotic independence based on the characterization

(3.1)

$$(X \overset{\text{a. ind.}}{\sim} Y) \equiv \left\{ \mathbb{F}_\delta(t) = \mathrm{pr}\Big(X^{-1} + Y^{-1} < \delta\, t \;\Big|\; X^{-1} + Y^{-1} < \delta \Big) \underset{\delta \to 0}{\longrightarrow} t^2, \;\; t \in [0,1] \right\}.$$

Alternatively, under asymptotic dependence we have pointwise convergence of $\mathbb{F}_\delta(t) \to t$, for $t \in [0,1]$, as $\delta \to 0$. Falk & Michel (2006) use condition (3.1) to test for asymptotic independence of (X, Y) using a battery of classical goodness-of-fit tests. An extension of their method can be found in Frick *et al.* (2007).

3.4. Gamma test

Zhang (2008) introduced the tail quotient correlation to assess extremal dependence between random variables. If u is a positive high threshold, and W and V are exceedance values over u of X and Y, then the tail quotient correlation coefficient is defined as

$$q_{u,n} = \frac{\max\big\{(u+W_i)/(u+V_i)\big\}_{i=1}^n + \max\big\{(u+V_i)/(u+W_i)\big\}_{i=1}^n - 2}{\max\big\{(u+W_i)/(u+V_i)\big\}_{i=1}^n \, \max\big\{(u+V_i)/(u+W_i)\big\}_{i=1}^n - 1} \,. \tag{3.2}$$

Asymptotically, $q_{u,n}$ can take values between zero and one. If both $\max\big\{(u+W_i)/(u+V_i)\big\}_{i=1}^n$ and $\max\big\{(u+V_i)/(u+W_i)\big\}_{i=1}^n$ are large, so that large values of both variables tend to occur one at a time, $q_{u,n}$ will be close to zero. If the two 'max' are close to one, then $q_{u,n}$ approaches one, and hence large values of both variables tend to occur together. There is a connection to the tail dependence index χ in (2.8): if χ is zero, then $q_{u,n}$ converges to zero almost surely. So if (X, Y) is asymptotically independent, $q_{u,n}$ is close to zero, although, in practice, the tail quotient correlation coefficient may never reach zero. This brings us to the hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \; (X, Y) \text{ is asymptotically independent}\,, \\ \mathrm{H}_1 &: \; (X, Y) \text{ is asymptotically dependent}\,. \end{aligned}$$

The Gamma test for asymptotic independence says that as $n \to \infty$,

$$n\, q_{u,n} \;\overset{\mathrm{d}}{\longrightarrow}\; \Gamma\big\{2, 1 - \exp(-1/u)\big\} \,.$$

A large value of $q_{u,n}$ is indicative of tail dependence and thus leads to a smaller p-value. If H_0 is rejected, we can use $q_{u,n}$ as measure of extremal dependence.

Although it might seem that the tail quotient correlation increases as u increases, this is not the case as an increase in u leads to a decrease in the scale parameter $1-\exp(-1/u)$, leading to a larger α-percentile.

The tail quotient correlation in (3.2) is an extension of another measure of dependence—the quotient correlation—which is defined as

$$q_n \;=\; \frac{\max\{Y_i/X_i\}_{i=1}^n + \max\{X_i/Y_i\}_{i=1}^n - 2}{\max\{Y_i/X_i\}_{i=1}^n \times \max\{X_i/Y_i\}_{i=1}^n - 1}\;. \tag{3.3}$$

Zhang *et al.* (2011) shows that (3.3) is asymptotically independent of the Pearson correlation ρ_n, meaning that q_n and ρ_n measure different degrees of association between random variables, in a large sample setting.

3.5. Madogram test

Bacro *et al.* (2010) propose to test for asymptotic independence using a madogram

$$W = \frac{1}{2}\left|F_X(X) - F_Y(Y)\right|,$$

which is a tool often used in geostatistics to capture spatial structures. The expected value and the variance of the madogram depend on the extremal coefficient as follows:

$$\mu_W \;=\; \frac{1}{2}\left(\frac{\theta-1}{\theta+1}\right), \qquad \sigma_W^2 \;=\; \frac{1}{6} - \mu_W^2 - \frac{1}{2}\int_0^1 \frac{\mathrm{d}t}{\big\{1+A(t)\big\}^2}\;,$$

where A is the Pickands' dependence function, which is related to the spectral measure H, as follows:

$$A(t) \;=\; 2\int_0^1 \max\big\{w(1-t),(1-w)t\big\}\;\mathrm{d}H(w)\;.$$

Hence testing for asymptotic independence ($\theta = 2$) is the same as testing if $\mu_W = 1/6$. Inference is made on the basis of the asymptotic result

$$n^{1/2}\left(\frac{\widehat{\mu}_W - 1/6}{\widehat{\sigma}_W}\right) \;\xrightarrow{\mathrm{d}}\; N(0,1)$$

where $\widehat{\mu}_W$ and $\widehat{\sigma}_W$ are consistent estimators of μ and σ.

3.6. Notes and comments

Other tests of independence between marginal extremes include a Cramér–von Mises-type statistic by Deheuvels & Martynov (1996), a dependence function

based test by Deheuvels (1980), a test based on the number of points below certain thresholds by Dorea & Miasaki (1993), the dependence function approaches of Capéraà *et al.* (1997). The behavior of Kendall's-τ as a measure of dependence within extremes has been also examined; see Capéraà *et al.* (2000) and Genest & Rivest (2001). An alternative likelihood-based approach that uses additional occurrence time information is given in Stephenson & Tawn (2005), and Ramos & Ledford (2009) propose likelihood ratio-based tests for asymptotic independence, asymmetry, and ray independence, resulting from a joint tail modeling approach which we describe in §4.2.

The huge literature on inference for asymptotic independence itself requires an entire survey. The criterion for selecting the methods presented above was mainly their simplicity, but many other methods exist which would also meet this criterion; see de Haan & de Ronde (1998), Husler & Li (2009), Tsai *et al.* (2011), among others.

4. JOINT TAIL MODELS

4.1. Joint tail specifications

We start by discussing three different regular variation-based specifications that provide the basis for the joint tail models to be discussed. The idea is to provide a chronological view on the different specifications considered on extremal dependence models that accommodate both asymptotic dependence and asymptotic independence. Most of the emphasis is placed on the Ramos–Ledford spectral model.

Let $(\mathcal{X}, \mathcal{Y})$ be a bivariate random variable with joint distribution function $F_{\mathcal{X},\mathcal{Y}}$ with margins $F_{\mathcal{X}}$ and $F_{\mathcal{Y}}$; we apply (2.7) to obtain a pair of unit Fréchet distributed random variables, X and Y. Ledford & Tawn (1996) proposed the following specification for the joint survival function:

$$\overline{F}_{X,Y}(x,x) = \mathrm{pr}\big(X > x, Y > x\big) = \frac{\ell(x)}{x^{1/\eta}} ,$$

where $\eta \in (0,1]$ is the coefficient of tail dependence and ℓ is a slowly varying function, i.e., $\lim_{x\to\infty} \ell(tx)/\ell(x) = 1$, for all $t > 0$.

Ledford & Tawn (1997, 1998) proposed the more flexible joint asymptotic expansion

$$\overline{F}_{X,Y}(x,y) = \mathrm{pr}\big(X > x, Y > x\big) = \frac{\mathcal{L}(x,y)}{x^{c_1}\, y^{c_2}} , \qquad c_1 + c_2 = \eta , \tag{4.1}$$

where $\mathcal{L}$ is a bivariate slowly varying function, i.e., there is a function g, the so-called limit function of $\mathcal{L}$, such that for all $x, y > 0$ and $c > 0$

$$g(x,y) \equiv \lim_{r\to\infty} \left\{ \frac{\mathcal{L}(r\,x, r\,y)}{\mathcal{L}(r,r)} \right\} , \qquad g(cx, cy) = g(x,y) \ . \tag{4.2}$$

The so-called ray dependence function is then defined as

$$g_*(w) \equiv g(x,y) \ , \qquad w = x/(x+y) \in [0,1] \ .$$

If $g_*(w)$ varies with w, we say that $\mathcal{L}(x,y)$ is ray dependent; if otherwise $g_*(w) = 1$, $w \in (0,1)$, we say that is ray independent.

Ramos & Ledford (2009) considered a particular case of specification (4.1) where $c_1 = c_2$, i.e.,

$$\overline{F}_{X,Y}(x,y) \ = \ \mathrm{pr}\big(X > x, Y > x\big) \ = \ \frac{\mathcal{L}(x,y)}{(x\,y)^{1/(2\eta)}} \ . \tag{4.3}$$

4.2. Ramos–Ledford spectral model

Ramos & Ledford (2009) base their analysis on the bivariate conditional random variable $(S,T) = \lim_{u\to\infty} \big\{ (X/u, Y/u) \colon (X > u, Y > u) \big\}$, for a high threshold u. The joint survivor function of the conditional random variable (S,T) is such that

$$\begin{aligned} \overline{F}_{ST}(s,t) \ &= \ \mathrm{pr}\big(S > s \,,\, T > t\big) \\ &= \ \lim_{u\to\infty} \frac{\mathrm{pr}\big(X > s\,u, Y > t\,u\big)}{\mathrm{pr}\big(X > u, Y > u\big)} \\ &= \ \eta \int_0^1 \left\{ \min\left(\frac{w}{s}, \frac{1-w}{t} \right) \right\}^{1/\eta} \mathrm{d}H_\eta(w) \ , \end{aligned} \tag{4.4}$$

where H_η is a non-negative measure on $[0,1]$ that should obey the normalization constraint

$$\int_0^{1/2} w^{1/\eta} \, \mathrm{d}H_\eta(w) + \int_{1/2}^1 (1-w)^{1/\eta} \, \mathrm{d}H_\eta(w) \ = \ \frac{1}{\eta} \ . \tag{4.5}$$

The measure H_η is analogous to the spectral measure H in classical models for multivariate extremes, which in turn must obey normalization and marginal moment constraints:

$$\int_0^1 \mathrm{d}H(w) = 1 \ , \qquad \int_0^1 w \, \mathrm{d}H(w) \ = \ \frac{1}{2} \ .$$

The two measures can be related: for example, if $\eta = 1$, $\mathrm{d}H_1(w) = \chi \times 2\,\mathrm{d}H(w)$ (Ramos & Ledford, 2009, p. 240), with $\chi = 2 - \int_0^1 \max(w, 1-w)\,\mathrm{d}H(w)$. The

measure H_η is a particular case of the hidden angular measure, which has been studied by Resnick (2002) and Maulik & Resnick (2004), but in these papers the normalization constraint (4.5) has been omitted.

Using the joint tail specification (4.3) we can also relate the joint survivor function of the conditional random variable (S, T) with the ray dependence function $g_\star$, as follows:

$$\overline{F}_{ST}(s,t) \,=\, \lim_{u\to\infty}\left\{\frac{\mathcal{L}(us,ut)}{\mathcal{L}(u,u)\,(st)^{1/(2\eta)}}\right\} \,=\, \frac{g(s,t)}{(st)^{1/(2\eta)}} \,=\, \frac{g_*\big\{s/(s+t)\big\}}{(st)^{1/(2\eta)}}\ .$$

Treating the limit in (4.4) as an approximation in the joint tail, we have that for a sufficiently large threshold u

$$\overline{F}_{X,Y}(x,y) \,\approx\, \overline{F}_{X,Y}(u,u)\,\overline{F}_{S,T}(x/u,\,y/u)\ , \qquad (x,y)\in(u,\infty)^2\,. \tag{4.6}$$

For an arbitrary $(\mathcal{X}, \mathcal{Y})$ with joint distribution function $F_{\mathcal{X},\mathcal{Y}}$, with margins $F_\mathcal{X}$ and $F_\mathcal{Y}$, we apply (2.7) to obtain a pair of unit Fréchet distributed random variables, X and Y. The joint survivor function of $(\mathcal{X}, \mathcal{Y})$ can then be modelled by

$$\overline{F}_{(\mathcal{X},\mathcal{Y})}(x,y) \,=\, \lambda\,\overline{F}_{ST}\left\{\frac{-1}{u\,\log F_\mathcal{X}(x)},\,\frac{-1}{u\,\log F_\mathcal{Y}(y)}\right\}\,, \quad (x,y)\in(u_1,\infty)\times(u_2,\infty)\ ,$$

where λ denotes the probability of falling in R_{11}. Ramos & Ledford (2009) also showed that for this approach to yield a complete joint tail characterization, the marginal tails of the survivor function of S and T must satisfy certain monotonicity conditions, implying that their marginal tails cannot be heavier than the unit Fréchet survivor function. These conditions guarantee that a given function $\overline{F}_{ST}$ can arise as a limit in equation (4.4).

Example 4.1. To exploit this in applications, Ramos & Ledford (2009) propose a parametric model—the η-asymmetric logistic model—which is a modified version of the asymmetric logistic dependence structure for classical bivariate extremes (Tawn, 1988), according to the model discussed above. The hidden angular density for this model is

$$h_\eta(w) \,=\, \frac{\eta-\alpha}{\alpha\,\eta^2 N_\rho}\left\{(\rho\,w)^{-1/\alpha}+\left(\frac{1-w}{\rho}\right)^{-1/\alpha}\right\}^{\alpha/\eta-2}\big\{w(1-w)\big\}^{-(1+1/\alpha)}\,, \qquad w\in[0,1]\ ,$$

where

$$N_\rho \,=\, \rho^{-1/\eta}+\rho^{1/\eta}-\big(\rho^{-1/\alpha}+\rho^{1/\alpha}\big)^{\alpha/\eta}\ , \qquad \eta,\alpha\in(0,1]\,, \quad \rho>0\ .$$

Hence using (4.4) we obtain

$$\overline{F}_{ST}(s,t) \,=\, N_\rho^{-1}\left[(\rho s)^{-1/\eta}+\left(\frac{t}{\rho}\right)^{-1/\eta}-\left\{(\rho s)^{-1/\alpha}+\left(\frac{t}{\rho}\right)^{-1/\alpha}\right\}^{\alpha/\eta}\right],$$

so that by (4.6) the joint survival model for (X, Y) is

$$\overline{F}_{X,Y}(x,y) \,=\, \overline{F}_{X,Y}(u,u) \times \frac{u^{1/\eta}}{N_\rho}\left[(\rho x)^{-1/\eta} + \left(\frac{y}{\rho}\right)^{-1/\eta} - \left\{(\rho x)^{-1/\alpha} + \left(\frac{y}{\rho}\right)^{-1/\alpha}\right\}^{\alpha/\eta}\right],$$

for $(x, y) \in [u, \infty)^2$.

4.3. Curse of dimensionality?

The model admits a d-dimensional generalization, where the hidden angular measure now needs to obey the normalization constraint

$$\int_{\Delta_d} \min\{w_1, \ldots, w_d\}^{1/\eta} \, \mathrm{d}H_\eta(\mathbf{w}) \,=\, 1/\eta \;, \tag{4.7}$$

where $\Delta_d = \{\mathbf{w} \in \mathbb{R}_+^d \colon \sum_{i=1}^d w_i = 1;\ \mathbf{w} = (w_1, \ldots, w_d)\}$. The corresponding constraints that the angular measure needs to obey are

$$\int_{\Delta_d} \mathbf{w} \, \mathrm{d}H(\mathbf{w}) = 1 \;, \qquad \int_{\Delta_d} \mathbf{w} \, \mathrm{d}H(\mathbf{w}) \,=\, d^{-1}\,\mathbf{1}_d \;, \tag{4.8}$$

Hence, whereas in classical models for multivariate extremes $d+1$ constraints need to be fulfilled, in the d-dimensional version of the Ramos–Ledford model only one constraint needs to be fulfilled.

A d-dimensional version of the η-asymmetric model discussed in Example 4.1 can be found in Ramos & Ledford (2011, p. 2221).

4.4. Notes and comments

Qin *et al.* (2008) discuss a device for obtaining further parametric specifications for the Ramos–Ledford model, using a construction similar to Coles & Tawn (1991). Whereas Coles & Tawn (1991) propose a method that transforms any positive measure on the simplex to satisfy the constraints (4.8), Qin *et al.* (2008) propose a method that transforms any positive measure on the simplex, to satisfy the Ramos–Ledford constraint (4.7). Qin *et al.* (2008) use their device to produce a Dirichlet model for the hidden angular density h_η. Ramos & Ledford (2011) give a point process representation that supplements the model discussed above.

Wadsworth & Tawn (2012a) propose a model based on a specification on which the axis along which the extrapolation is performed is 'tilted' by assuming that the marginals grow at different rates. They also obtain analogues of

the Pickands and exponent functions for this setting, and propose the so-called inverted multivariate extreme value distributions, which are models for asymptotic independence, having a one-to-one correspondence with multivariate extreme value distributions; any construction principle or model generator for a multivariate extreme value distributed X can thus be readily adapted to create a inverted multivariate extreme value distributed Y. The link between multivariate extreme value distributions and their inverted versions allows the use of approaches which are amenable to non/semi-parametric methods for a moderate number of dimensions, and it also convenient for parametric modeling of high-dimensional extremes; for example, the max-mixture $\max\{aX, (1-a)Y\}$, $a \in [0,1]$, can then be used as a hybrid model, and this principle is adapted for spatial modeling of extremes in Wadsworth & Tawn (2012b).

Maxima of moving maxima (M4) processes have been recently extended by Heffernan *et al.* (2007) to produce models for asymptotic independence.

5. CONDITIONAL TAIL MODELS

5.1. Conditional tail specification

The models discussed in §4 focused on the joint tails, but under asymptotic independence it may be restrictive to confine the analysis to such region. Heffernan & Tawn (2004) propose conditional tail models, where the focus is on events where at least one component of (X, Y) is extreme, where here we now assume Gumbel marginal distributions. We thus need to model the distribution of $X \,|\, Y$ when Y is large, and of $Y \,|\, X$ when X is large; for concreteness we focus on the latter. Analogously to the joint tail modeling, a limiting specification is also needed here: we assume that there exist norming functions $a(u)$ and $b(u) > 0$, such that

$$\lim_{u\to\infty} \mathrm{pr}\left\{ \frac{Y - b(u)}{a(u)} \leq e \mid X = u \right\} = G(e) \ . \tag{5.1}$$

To ensure that Y has no mass at ∞, G needs to satisfy

$$\lim_{z\to\infty} G(z) = 1 \ .$$

We define the auxiliary variable $\varepsilon = \{Y - b(u)\}/a(u)$, so that specification (5.1) can be rewritten as $\lim_{u\to\infty} \mathrm{pr}\big(\varepsilon \leq e \mid X = u\big) = G(e)$.

5.2. Heffernan–Tawn model

The starting point for modeling is the following approximation to specification (5.1), which holds for a high threshold u:

$$\mathrm{pr}\big(\varepsilon \le \epsilon \mid X = x\big) \;\approx\; G(\epsilon) \;=\; \mathrm{pr}\big(\varepsilon \le \epsilon\big)\;, \qquad x > u\;.$$

Hence, we have that $\varepsilon \sim G$ is (almost) independent of X, for u large. We restrict our attention to a simplified version of the model where (X, Y) are non-negatively dependent, so that the norming functions are $a(x) = \alpha\, x$ and $b(x) = x^{\beta}$, with $\alpha \in [0, 1]$, $\beta \in (-\infty, 1]$, and $x > u$. The model can be thus written as a regression model

$$\begin{aligned} Y &= a(X) + b(X)\varepsilon \\ &= \alpha X + X^{\beta}\varepsilon\;, \qquad X > u\;, \end{aligned} \tag{5.2}$$

where ε has mean μ_{ε} and standard error σ_{ε}. Since the distribution of ε is unspecified, the model is semiparametric, with the estimation targets of interest being α, β and G. The variable ε is analogous to a standardized residual in a classical regression context, but here μ_{ε} need not equal zero in general, so the conditional mean and standard errors of the responses Y are

$$\mu_{Y|X=x} \;=\; \alpha x + \mu_{\varepsilon} x^{\beta}\,, \qquad \sigma_{Y|X=x} \;=\; \sigma_{\varepsilon} x^{\beta}\;.$$

The interpretation for the α and β are the following: the larger the α the greater the degree of extremal dependence; the larger the β the greater the conditional variance of $Y \mid X = x$. Asymptotic dependence occurs when $(\alpha, \beta) = (1, 0)$, whereas asymptotic independence holds whenever $\alpha \in [0, 1)$, regardless of the value of $\beta \in (-\infty, 1)$. Inference is often made assuming normality of ε so that maximum likelihood methods can be used for the parametric part of the model, and the empirical distribution function is often used to estimate G. Estimation can thus be based on the $k = \sum_{i=1}^{n} I(x_i > u)$ conditional exceedances using the following two-stage method (Keef *et al.*, 2009a):

Step 1. Parametric block

$$(\widehat{\alpha}, \widehat{\beta}) \;=\; \arg\max_{(\alpha,\beta)} \;-\; \sum_{i=1}^{n} \left\{ \log(\sigma_{Y|X=x_i}) + \frac{1}{2}\left(\frac{y_i - \mu_{Y|X=x_i}}{\sigma_{Y|X=x_i}} \right)^2 \right\} I(x_i > u)\;.$$

Step 2. Nonparametric block

$$\widehat{G}(e) \;=\; \frac{1}{k} \sum_{i=1}^{n} I\Big(y_i \le e\, x_i^{\widehat{\beta}} + \widehat{\alpha}\, x_i \Big)\, I(x_i > u)\;.$$

As an alternative to Step 2 we can also obtain a kernel estimate as follows:

$$\widetilde{G}(e) \;=\; \frac{1}{k}\sum_{i=1}^{n} K\!\left(e - \frac{y_i - \widehat{\alpha}\,x_i}{x^{\widehat{\beta}}}\right) I(x_i > u)\;, \tag{5.3}$$

with K denoting a kernel and $h > 0$ its bandwidth. This procedure suffers however from a weakness common to all two-stage approaches: uncertainty is underestimated in the second step.

5.3. Notes and comments

Heffernan & Resnick (2007) provide a mathematical examination of a modified Heffernan–Tawn model and its connections with hidden regular variation. A version of the model able to cope with missing data can be found in Keef *et al.* (2009b). For applications see, for instance, Paulo *et al.* (2006), Keef *et al.* (2009a), and Hilal *et al.* (2011).

6. REMARKS ON THE ONE-SAMPLE FRAMEWORK

6.1. Asymptotic independence of order statistics

The expression "asymptotic independence" did not appear for the first time in the works of Geffroy (1958, 1959) and Sibuya (1960), in the context of statistics of extremes. The concept was motivated by a conjecture that Gumbel made on the joint limiting distribution of pairs of order statistics, in a one-sample framework:

> "In a previous article [1] the assumption was used that the m^{th} observation in ascending order (from the bottom) and the m^{th} observation in descending order (from the top) are independent variates, provided that the rank m is small compared to the sample size n." (Gumbel, 1946).

While asymptotic independence, as described in § 2, is a two-sample concept, asymptotic independence as first described by Gumbel is a one-sample concept. Although the expression "asymptotic independence" is not used in Gumbel's paper, the expression started to appear immediately thereafter (e.g. Homma, 1951).

Many papers that appeared after Gumbel (1946) focused on the analysis of asymptotic independence of sets of order statistics (Ikeda, 1963; Ikeda & Matsunawa, 1970; Falk & Kohne, 1986; Falk & Reiss, 1988).

6.2. Asymptotic independence of sum and maximum

Chow & Teugels (1978) studied the asymptotic joint limiting distribution of the standardized sum and maximum

$$(S_n^*, M_n^*) = \left(\frac{S_n - n\, b_n}{a_n}, \frac{M_n - d_n}{c_n} \right), \qquad S_n = \sum_{i=1}^{n} X_i \,, \qquad M_n = \max\{X_i\}_{i=1}^{n} \,,$$

for norming constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$. Their results, which only apply to the case where the X_i are independent and identically distributed, were later extended to stationary strong mixing sequences by Anderson & Turkman (1991, 1995), who showed that for such sequences, (S_n, M_n) is asymptotically independent, under fairly mild conditions; these results also allow us to characterize the joint limiting distribution of $(\overline{X}_n, M_n)$, with $\overline{X}_n = n^{-1} S_n$. Hsing (1995) extended these results further, and showed that for stationary strong mixing sequences, asymptotic normality of S_n is sufficient for the asymptotic independence of (S_n, M_n).

Assume that $E(X_i) = 0$ and $E(X_i^2) = 1$, so that the process of interest has autocorrelation $r_n = E(X_{i+n} X_i)$. Ho & Hsing (1996) obtained the asymptotic joint limiting distribution of (S_n, M_n) for stationary normal random variables under the condition

$$\lim_{n \to \infty} r_n \log n \,=\, r \,\in\, [0, \infty) \tag{6.1}$$

and showed that (S_n, M_n) is asymptotically independent only if $r = 0$. Related results can be found in Peng & Nadarajah (2003), who obtain the asymptotic joint distribution of (S_n, M_n) under a stronger dependence setting. Ho & McCormick (1999) and McCormick & Qi (2000) showed that $(M_n - \overline{X}_n, S_n)$ is asymptotically independent if

$$\lim_{n \to \infty} n^{-1} \log n \sum_{i=1}^{n} |r_i - r_n| \,=\, 0 \,. \tag{6.2}$$

James *et al.* (2007) study multivariate stationary Gaussian sequences, and show, under fairly mild conditions, that if the componentwise maximum has a limiting distribution, then (S_n^*, M_n^*) is asymptotically independent.

Hu *et al.* (2009) show that the point process of exceedances of a standardized Gaussian sequence converges to a Poisson process, and that this process is asymptotically independent of the partial sums; in addition, they obtain the asymptotic joint distribution for the extreme order statistics and the partial sums.

6.3. Notes and comments

Related results on the asymptotic independence of sum and maximum are also discussed in Tiago de Oliveira (1961). Condition (6.1) was introduced by Berman (1964) and Mittal & Ylvisaker (1975), who studied the asymptotic distribution of M_n in the cases of $r = 0$ and $r > 0$, respectively. Conditions (6.1), was introduced by McCormick (1980), who studied the asymptotic distribution of $M_n - \overline{X}_n$.

From the statistical point of view, fewer estimation and inference tools have been developed for asymptotic independence in the one-sample framework, in comparison with the two-sample case, and many developments have been made without any statistical applications being given, and mostly at the probabilistic level.

7. CONCLUSION

We have reviewed key themes for statistical modeling of asymptotically independent data, with a focus on bivariate extremes. The inventory of approaches is large, and there exists in the literature a wealth of different perspectives potentially useful for modeling risk. Statistical and probabilistic issues are discussed, providing a fresh view on the subject, by combining modern advances with a historical perspective, and tools of theoretical and applied interest.

ACKNOWLEDGMENTS

We are grateful to Vanda Inácio, Anthony Davison, Feridun Turkman, Ivette Gomes and Jennifer Wadsworth.

The research of the first author was partially supported by the *Fundação para a Ciência e a Tecnologia*, through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

Anderson, C.W. & Turkman, K.F. (1991). The joint limiting distribution of sums and maxima of stationary sequences, *J. Appl. Prob.*, **28**, 33–44.

Anderson, C.W. & Turkman, K.F. (1995). Sums and maxima of stationary sequences with heavy tailed distributions, *Sankhya*, **57**, 1–10.

Bacro, J.; Bel, L. & Lantuéjould, C. (2010). Testing the independence of maxima: from bivariate vectors to spatial extreme fields, *Extremes*, **13**, 155–175.

Balkema, G.A.A. & Nolde, N. (2010). Asymptotic independence for unimodal densities, *Adv. Appl. Prob.*, **42**, 411–432.

Beirlant, J.; Goegebeur, Y.; Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, New York.

Beirlant, J. & Vandewalle, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value statistics, *Statist. Prob. Lett.*, **60**, 265–278.

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

Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences, *Ann. Math. Statist.*, **35**, 502–516.

Bingham, N.H.; Goldie, C.M. & Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

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 attractors, *J. Mult. Anal.*, **72**, 30–49.

Chow, T. & Teugels, J. (1978). *The sum and the maximum of iid random variables.* In "Proc. 2nd Prague Symp. Asymp. Statist." (P. Mandl and M. Huskova, Eds.), 81–92, North-Holland, Amsterdam.

Coles, S.G.; Heffernan, J. & Tawn, J.A. (1999). Dependence measures for extreme value analyses, *Extremes*, **2**, 339–365.

Coles, S.G. & Tawn, J.A. (1991). Modelling extreme multivariate events, *J. R. Statist. Soc. B*, **53**, 377–392.

de Haan, L. & de Ronde, J. (1998). Sea and wind: multivariate extremes at work, *Extremes*, **1**, 7–45.

Deheuvels, P. (1980). Some applications of the dependence functions to statistical inference: nonparametric estimates of extreme value distributions, and a Kiefer type universal bound for the uniform test of independence, *Colloq. Math. Societ. János Bolyai. 32. Nonparam. Stat. Infer., Budapest (Hungary)*, 183–201.

Deheuvels, P. & Martynov, G. (1996). Cramér–Von Mises-type tests with applications to tests of independence for multivariate extreme value distributions, *Comm. Statist. Theory Meth.*, **25**, 871–908.

Dorea, C. & Miasaki, E. (1993). Asymptotic test for independence of extreme values, *Acta Math. Hung.*, **62**, 343–347.

Draisma, G.; Drees, H. & de Haan, L. (2004). Bivariate tail estimation: Dependence in asymptotic independence, *Bernoulli*, **10**, 251–280.

Embrechts, P.; McNeil, A. & Straumann, D. (2002). *Correlation and dependence in risk management: properties and pitfalls.* In "Risk Management: Value at Risk and Beyond" (M. Dempster, Ed.), Cambridge University Press, Cambridge, pp. 176–223.

Falk, M. & Kohne, W. (1986). On the rate at which the sample extremes become independent, *Ann. Prob.*, **14**, 1339–1346.

Falk, M. & Reiss, R.-D. (1988). Independence of order statistics, *Ann. Prob.*, **16**, 854–862.

Falk, M. & Michel, R. (2006). Testing for tail independence in extreme value models, *Ann. Inst. Statist. Math.*, **58**, 261–290.

Ferreira, H. & Ferreira, M. (2012). Tail dependence between order statistics, *J. Mult. Anal.*, **105**, 176–192.

Finkelstein, B.V. (1953). On the limiting distributions of the extreme terms of a variational series of a two-dimensional random quantity, *Doklady Akad. SSSR*, **91**, 209–211 (in Russian).

Frick, M.; Kaufmann, E. & Reiss, R. (2007). Testing the tail-dependence based on the radial component, *Extremes*, **10**, 109–128.

Galambos, J. (1975). Order statistics of samples from multivariate distributions, *J. Am. Statist. Assoc.*, **70**, 674–680.

Geffroy, J. (1958). Contribution à la théorie des valeurs extrêmes, *Publ. Inst. Statist. Univ. Paris*, **7**, 37–121.

Geffroy, J. (1959). Contribution à la théorie des valeurs extrêmes, *Publ. Inst. Statist. Univ. Paris*, **8**, 3-52.

Genest, C. & Rivest, L.-P. (2001). On the multivariate probability integral transform, *Statist. Probab. Lett.*, **53**, 391–399.

Gumbel, E.J. (1946). On the independence of the extremes in a sample, *Ann. Math. Statist.*, **17**, 78–81.

Gumbel, E.J. (1960). Bivariate exponential distributions, *J. Am. Statist. Assoc.*, **55**, 698–707.

Gumbel, E.J. (1961). Bivariate logistic distributions, *J. Am. Statist. Assoc.*, **56**, 335–349.

Gumbel, E.J. & Goldstein, N. (1964). Analysis of empirical bivariate extremal distributions, *J. Am. Statist. Assoc.*, **59**, 794–816.

Haan, L. & Resnick, S.I. (1977). Limit theory for multivariate sample extremes, *Prob. Theory Rel.*, **40**, 317–337.

Heffernan, J.E.; Tawn, J.A. & Zhang, A. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes, *Extremes*, **10**, 57–82.

Heffernan, J.E. & Resnick, S.I. (2007). Limit laws for random vectors with an extreme component, *Ann. Appl. Prob.*, **17**, 537–571.

Heffernan, J.E. (2000). A directory of coefficients of tail dependence, *Extremes*, **3**, 279–290.

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

Hilal, S.; Poon, S.H. & Tawn, J.A. (2011). Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX, *J. Bank. Financ.*, **35**, 2374–2387.

Ho, H. & Hsing, T. (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables, *J. Appl. Prob.*, **33**, 138–145.

Ho, H. & McCormick, W. (1999). Asymptotic distribution of sum and maximum for Gaussian processes, *J. Appl. Prob.*, **36**, 1031–1044.

Homma, T. (1951). On the asymptotic independence of order statistics, *Rep. Stat. Appl. Res. JUSE*, **1**, 1–8.

Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables, *Ann. Prob.*, **23**, 938–947.

Hu, A.; Peng, Z. & Qi, Y. (2009). Joint behavior of point process of exceedances and partial sum from a Gaussian sequence, *Metrika*, **70**, 279–295.

Hult, H. & Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions, *Adv. Appl. Prob.*, **34**, 587–608.

Husler, J. & Li, D. (2009). Testing asymptotic independence in bivariate extremes, *J. Statist. Plann. Infer.*, **139**, 990–998.

Ikeda, S. (1963). Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence, *Ann. Inst. Statist. Math.*, **15**, 87–116.

Ikeda, S. & Matsunawa, T. (1970). On asymptotic independence of order statistics, *Ann. Inst. Statist. Math.*, **22**, 435–449.

James, B.; James, K. & Qi, Y. (2007). Limit distribution of the sum and maximum from multivariate Gaussian sequences, *J. Mult. Anal.*, **98**, 517–532.

Johnson, N. & Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.

Keef, C.; Svensson, C. & Tawn, J.A. (2009a). Spatial dependence in extreme river flows and precipitation for Great Britain, *J. Hydrol.*, **378**, 240–252.

Keef, C.; Tawn, J.A. & Svensson, C. (2009b). Spatial risk assessment for extreme river flows, *J. R. Statist. Soc. C*, **58**, 601–618.

Ledford, A.W. & Tawn, J.A. (1998). Concomitant tail behaviour for extremes, *Adv. Appl. Prob.*, **30**, 197–215.

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.

Mardia, K. (1964). Asymptotic independence of bivariate extremes, *Calcutta Statist. Assoc. Bull.*, **13**, 172–178.

Marshall, A. & Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions, *Ann. Prob.*, **11**, 168–177.

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

McCormick, W. (1980). Weak convergence for the maxima of stationary Gaussian processes using random normalization, *Ann. Prob.*, **8**, 483–497.

McCormick, W. & Qi, Y. (2000). Asymptotic distribution for the sum and maximum of Gaussian processes, *J. Appl. Prob.*, **8**, 958–971.

Mikhailov, V. (1974). Asymptotic independence of vector components of multivariate extreme order statistics, *Theory Prob. Appl.*, **19**, 817.

Mittal, Y. & Ylvisaker, D. (1975). Limit distributions for the maxima of stationary gaussian processes, *Stoch. Proc. Appl.*, **3**, 1–18.

Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed., Springer, New York.

Paulo, M.J.; Van der Voet, H.; Wood, J.C.; Marion, G.R. & Van Klaveren, J.D. (2006). Analysis of multivariate extreme intakes of food chemicals, *Food Chem. Toxic.*, **44**, 994–1005.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes, *Statist. Prob. Lett.*, **43**, 399–409.

Peng, Z. & Nadarajah, S. (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence, *Theory Prob. Appl.*, **47**, 706–708.

Poon, S.; Rockinger, M. & Tawn, J.A. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications, *Rev. Financ. Stud.*, **17**, 581–610.

Poon, S.-H.; Rockinger, M. & Tawn, J.A. (2003). Modelling extreme-value dependence in international stock markets, *Statist. Sinica*, **13**, 929–953.

Qin, X.; Smith, R. & Ren, R. (2008). *Modelling multivariate extreme dependence.* In "Proc. Joint Statist. Meet. Am. Statist. Assoc.", pp. 3089–3096.

Ramos, A. & Ledford, A.W. (2005). Regular score tests of independence in multivariate extreme values, *Extremes*, **8**, 5–26.

Ramos, A. & Ledford, A.W. (2009). A new class of models for bivariate joint tails, *J. R. Statist. Soc. B*, **71**, 219–241.

Ramos, A. & Ledford, A.W. (2011). An alternative point process framework for modelling multivariate extreme values, *Comm. Statist. Theory Meth.*, **40**, 2205–2224.

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

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

Schweizer, B. & Wolff, E.F. (1981). On nonparametric measures of dependence for random variables, *Ann. Statist.*, **9**, 879–885.

Sibuya, M. (1960). Bivariate extreme statistics, I, *Ann. Inst. Statist. Math.*, **11**, 195–210.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Stat. Paris*, **8**, 229–131.

Srivastava, M. (1967). Asymptotic independence of certain statistics connected with the extreme order statistics in a bivariate population, *Sankhya*, **29**, 175–182.

Stephenson, A. & Tawn, J.A. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima, *Biometrika*, **92**, 213–227.

Takahashi, R. (1994). Asymptotic independence and perfect dependence of vector components of multivariate extreme statistics, *Statist. Prob. Lett.*, **19**, 19–26.

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

Tiago de Oliveira, J. (1961). The asymptotical independence of the sample mean and the extremes, *Rev. Fac. Cienc. Univ. Lisboa*, **8**, 299–310.

Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes; extensions, *Est. Mat., Estat. Econ.*, **7**, 165–295.

Tsai, Y.; Dupuis, D. & Murdoch, D. (2011). A robust test for asymptotic independence of bivariate extremes, *Statistics* (DOI:10.1080/02331888.2011.568118).

Wadsworth, J. & Tawn, J.A. (2012a). Dependence modelling for spatial extremes, *Biometrika* (to appear).

Wadsworth, J. & Tawn, J.A. (2012b). *A new representation for multivariate tail probabilities* (submitted)

Zhang, Z. (2008). Quotient correlation: A sample based alternative to Pearson's correlation, *Ann. Statist.*, **36**, 1007–1030.

Zhang, Z.; Qi, Y. & Ma, X. (2011). Asymptotic independence of correlation coefficients with application to testing hypothesis of independence, *Elect. J. Statist.*, **5**, 342–372.

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 109–133

MODELLING TIME SERIES EXTREMES

Authors: V. Chavez-Demoulin
– Faculty of Business and Economics, University of Lausanne,
1015 Lausanne, Switzerland
valerie.chavez@unil.ch

A.C. Davison
– Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT,
Station 8, 1015 Lausanne, Switzerland
anthony.davison@epfl.ch

Abstract:

- The need to model rare events of univariate time series has led to many recent advances in theory and methods. In this paper, we review telegraphically the literature on extremes of dependent time series and list some remaining challenges.

Key-Words:

- *Bayesian statistics; Box–Cox transformation; clustering; dependence; extremal index; extremogram; generalized extreme-value distribution; generalized Pareto distribution; Hill estimator; nonparametric smoothing; non-stationarity; regression; tail index.*

AMS Subject Classification:

- 62E20, 62F15, 62G05, 62G08, 62G32, 62M05, 62M10.

1. INTRODUCTION

Statistical analysis of the extremes of time series is a traditional staple of hydrology and insurance, but the last two decades have seen applications broaden to a huge variety of domains, from finance to atmospheric chemistry to climatology. The most common approaches for describing the extreme events of stationary data are the block maximum approach, which models the maxima of a set of contiguous blocks of observations using the generalized extreme-value (GEV) distribution, and the peaks-over-threshold approach, in which a Poisson process model is used for exceedances of a fixed high or low threshold level; often this entails fitting the generalized Pareto distribution (GPD) to the exceedances. The two approaches lead to different but closely related descriptions of the extremes, determined by the marginal distribution of the series and by its extremal dependence structure. Whereas the marginal features are well-understood from the study of independent and identically distributed (iid) variates, the rather less well-explored dependence features are the main focus of this paper. We review some related relevant theory and methods and attempt to list some aspects that seem to need further study. Throughout the paper, we discuss maximum or upper extremes, but minima or lower extremes can be handled by negating the data.

Temporal dependence is common in univariate extremes, which may display intrinsic dependence, due to autocorrelation, or dependence due to the effects of other variables, or both, and this demands an appropriate theoretical treatment. Short-range dependence leading to clusters of extremes often arises: for example, financial time series usually display volatility clustering, and river flow maxima often occur together following a major storm. The joint behavior of the observations within a cluster is determined by the short-range dependence structure and can be accommodated, though not fully described, within a general theory. Long-range dependence of extremes seems implausible in most contexts, genetic or genomic data being a possible exception. Large-scale variation due to trend, seasonality or regime changes is typically dealt with by appropriate modelling.

Below we first give an account of the effect of dependence on time series extremes, and discuss associated statistical methods. For completeness we then outline some relevant Bayesian methods, and then turn to dealing with regression and non-stationarity. The paper closes with a brief list of some open problems.

2. SHORT-RANGE DEPENDENCE

2.1. Effect of short-range dependence

The discussion below is based partly on Leadbetter *et al.* (1983), a standard reference to the literature on extremes of time series and random processes, and on Beirlant *et al.* (2004, Ch. 10), which provides a more recent summary; see also Coles (2001, Ch. 5). It is usual to study the effect of autocorrelation under a type of mixing condition that restricts the impact of dependence on extremes.

Definition 2.1. A strictly stationary sequence $\{X_i\}$, whose marginal distribution F has upper support point $x_F = \sup\{x : F(x) < 1\}$, is said to satisfy $D(u_n)$ if, for any integers $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$\Big| \mathrm{P}\Big\{X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n, X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\Big\} - \mathrm{P}\Big\{X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n\Big\}\, \mathrm{P}\Big\{X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\Big\}\Big| \le \alpha(n, l) ,$$

where $\alpha(n, l_n) \to 0$ for some sequences $l_n = o(n)$ and $u_n \to x_F$ as $n \to \infty$.

The $D(u_n)$ condition implies that rare events that are sufficiently separated are almost independent. 'Sufficient' separation here is relatively short-distance, since $l_n/n \to 0$ as $n \to \infty$. This allows one to establish the following result, which shows that if the $D(u_n)$ condition is satisfied, then the GEV limit arises for the maxima of dependent data, thereby justifying the use of the block maximum approach for most stationary time series.

Theorem 2.1. *Let $\{X_i\}$ be a stationary sequence for which there exist sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$ and a non-degenerate distribution H such that $M_n = \max\{X_1, \ldots, X_n\}$ satisfies*

$$\mathrm{P}\Big\{(M_n - b_n)/a_n \le z\Big\} \to H(z) , \qquad n \to \infty .$$

If $D(u_n)$ holds with $u_n = a_n z + b_n$ for each z for which $H(z) > 0$, then H is a GEV distribution.

Thus the effect of dependence must be felt in the local behavior of extremes, the commonest measure of which is the *extremal index*, θ. This lies in the interval $[0, 1]$, though $\theta > 0$ except in pathological cases. If the sequence $\{X_n\}$ is independent, then $\theta = 1$, but this is also the case for certain dependent series. The relation between maxima of a dependent sequence and of a corresponding independent sequence is summarised in the following theorem:

Theorem 2.2. *Let $\{X_i\}$ be a stationary process and let $\{\tilde{X}_i\}$ be independent variables with the same marginal distribution. Set $M_n = \max\{X_1, \ldots, X_n\}$ and $\tilde{M}_n = \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$. Under suitable regularity conditions,*

$$\mathrm{P}\Big\{(\tilde{M}_n - b_n)/a_n \le z\Big\} \to \tilde{H}(z) \ , \qquad n \to \infty \ ,$$

for sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$, where $\tilde{H}$ is a non-degenerate distribution function, if and only if

$$\mathrm{P}\Big\{(M_n - b_n)/a_n \le z\Big\} \to H(z) \ ,$$

where $H(z) = \tilde{H}^\theta(z)$ for some constant $\theta \in [0, 1]$.

Since the extremal types theorem implies that the only possible non-degenerate limit $\tilde{H}$ is the GEV distribution, with location, scale and shape parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$, say, it follows that H is also GEV, with parameters

$$\tilde{\mu} \ = \ \mu - \frac{\sigma}{\xi}\big(1 - \theta^{-\xi}\big) \ , \qquad \tilde{\sigma} = \sigma\,\theta^{\xi}, \ \ \tilde{\xi} = \xi \ ,$$

and the value of θ determines by how much $\tilde{M}_n$ is stochastically larger than M_n. As $\tilde{\xi} = \xi$, the upper tail behaviour of $\tilde{H}$ is qualitatively the same as that of H, regardless of θ. For example, when $\tilde{H}$ is Gumbel, then $\tilde{\xi} = \xi = 0$, and the parameters of the independent case are related to those of the stationary process by $\tilde{\mu} = \mu + \sigma \log\theta$ and $\tilde{\sigma} = \sigma$: H is also Gumbel with the same scale parameter but a smaller location parameter.

The extremal index can be defined in various ways, which are equivalent under mild conditions. One is

$$\theta^{-1} \ = \ \lim_{n\to\infty} \mathrm{E}\left\{\sum_{j=1}^{p_n} I\big(X_i > u_n \mid M_{p_n} > u_n\big)\right\}, \tag{2.1}$$

where $p_n = o(n) \to \infty$ and the *threshold sequence* $\{u_n\}$ is chosen to ensure that $n\{1 - F(u_n)\} \to \lambda \in (0, \infty)$. Thus θ^{-1} is the limiting mean cluster size based on a block of p_n consecutive observations, as p_n increases. Another is

$$\theta \ = \ \lim_{n\to\infty} \mathrm{P}\Big\{\max(X_2, \ldots, X_{p_n}) \le u_n \mid X_1 \ge u_n\Big\}, \tag{2.2}$$

so θ is the limiting probability that an exceedance over u_n is the last of a cluster of such exceedances. Asymptotically, therefore, extremes of the stationary sequence occur in clusters of mean size $1/\theta$. Since the suitably rescaled times of exceedances over u_n in an independent sequence would in the limit arise as a Poisson process of rate λ, and since u_n is the same as for the corresponding independent series, the mean time between clusters in dependent series must increase by a factor $1/\theta$, corresponding to clusters of exceedances arising as a Poisson process of rate $\lambda\theta$.

Hsing (1987) shows that the structure of these clusters is essentially arbitrary; see also Hsing *et al.* (1988).

A consequence of Theorem 2.2 is that if the extremal types theorem is applicable, then for a suitable choice of parameters we may write

$$\mathrm{P}\big(M_n \leq x\big) \approx H(x) \approx \tilde{H}(x)^{\theta} \approx F(x)^{n\theta} \ ,$$

and so that M_n is effectively the maximum of $n\theta$ equivalent independent observations. Thus for dependent data and a large probability p, the marginal quantiles for X_j will be estimated by

$$F^{-1}(p) \approx H^{-1}(p^{n\theta}) > H^{-1}(p^n) \ ,$$

so ignoring the clustering would lead to an underestimation of quantiles of F. When clustering occurs, the notion of return level is more complex. If $\theta = 1$, for instance, then the '100-year-event' will occur on average ten times in the next millennium, but has probability 0.368 of not appearing in the next 100 years, whereas if $\theta = 1/10$, then on average the event also occurs ten times in a millennium, but all ten events will tend to appear together, leading to a probability around 0.9 of not seeing any in the next 100 years. Such information may be highly relevant to structural design.

Robinson & Tawn (2000) discuss how sampling a time series at different frequencies will affect the values of θ, and derive bounds on their relationships.

The left panel of Figure 1 shows a realization of $X_j = \sum_{i=1}^{6} i\,|Z_{j-i}|$, where the Z_j are iid with a Cauchy distribution.Clusters manifest themselves as vertical strings formed by points corresponding to successive large values of X_i, driven by occasional huge values of Z_j. The corresponding plot for an iid sequence would show no clustering. The middle panel shows realizations of the sequence $X_j = Z_j + 2\,Z_{j+1}$, with the Z_j iid Cauchy variates. In this case Davis & Resnick (1985) show that the average cluster size is 3/2. The right panel shows the Cauchy sequence $X_j = \rho X_{j-1} + (1 - |\rho|)Z_j$ where $\rho \in (0, 1)$ and the Z_j are iid standard Cauchy variates, for $\rho = 0.8$; Chernick *et al.* (1991) show that the extremal index is $1 - \rho$, so in this case the mean cluster size is 5.

Examples such as these are instructive, but such models are not widely used in applications. It follows from Sibuya (1960) that linear Gaussian autoregressive-moving average models have $\theta = 1$, corresponding to asymptotically independent extremes, despite the clumping that may appear at lower levels, and this raises the question of how to model the extremes of such series. Davis & Mikosch (2008, 2009a) show that while both GARCH and stochastic volatility models display volatility clustering, only the former shows clustering of extremes, thus providing a means to distinguish these classes of financial time series.

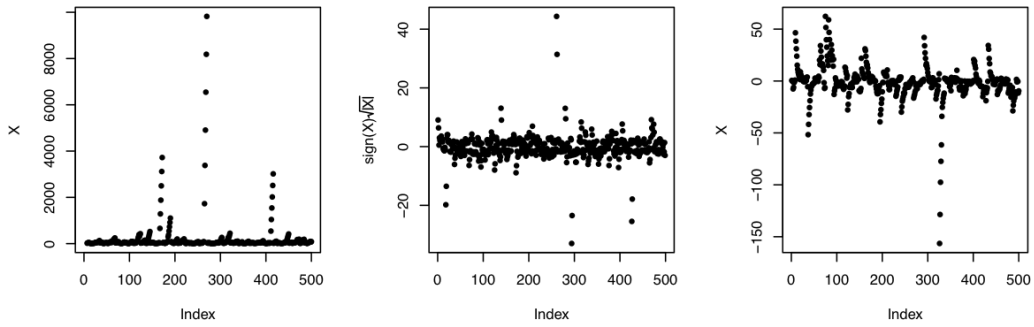


Figure 1: Clustering in realizations of some theoretical processes. Left panel: $X_i = \sum_{i=1}^{6} i\,|Z_{j-i}|$ where the Z_j are iid standard Cauchy. Middle panel: $X_i = Z_j + 2\,Z_{j+1}$ with the Z_j iid Cauchy; the data are transformed to $\mathrm{sign(X)}\sqrt{|\mathrm{X}|}$. Right panel: Cauchy AR(1) sequence $X_j = \rho\,X_{j-1} + Z_j$ with $\rho = 0.8$ and Z_j iid standard Cauchy.

Further conditions have been introduced to control local dependence of extremes, the best known of which is the following.

Definition 2.2. A strictly stationary sequence $\{X_n\}$ satisfies $D'(u_n)$ if

$$\limsup_{n\to\infty}\; n \sum_{j=2}^{[n/k]} \mathrm{P}\big\{X_1 > u_n,\, X_j > u_n\big\} \to 0\;, \qquad k = o(n)\,,\quad n \to \infty\;.$$

for some threshold sequence $\{u_n\}$ such that $n\{1 - F(u_n)\} \to \lambda \in (0, \infty)$.

This condition may be harder to satisfy than one might expect; Chernick (1981) gives an example of an autoregressive process with uniform margins that satisfies $D(u_n)$ but does not satisfy $D'(u_n)$.

It can be shown that a stationary process satisfying both $D(u_n)$ and $D'(u_n)$ has extremal index $\theta = 1$. Similar conditions have been introduced to ensure convergence of the point process of exceedances (Beirlant *et al.*, 2004, Ch. 10).

2.2. Statistics of cluster properties

Suppose that a sequence $\{X_i\}$ satisfies a suitable mixing condition, such as that in Definition 2.1, and call π the probability mass function of the size of a cluster of extreme values of mean size θ^{-1}. Suppose that we wish to estimate θ based on apparently stationary time series data of length n. The *blocks esti-*

mator of θ is computed using the empirical counterpart of (2.1), by selecting a value r, dividing the sample into $[n/r]$ disjoint contiguous blocks of length r, and then counting exceedances over a high threshold u in those blocks containing exceedances. The proportion of blocks with k exceedances estimates the probability $\pi(k)$ and the average number of exceedances per block having at least one exceedance estimates θ^{-1}. Likewise the *runs estimator* is the empirical counterpart of (2.2). Computations in Smith & Weissman (1994) suggest that the runs estimator has lower bias, and therefore is the preferable of the two. Ancona-Navarrete & Tawn (2000) compare the then-known estimators of the extremal index, using both nonparametric and parametric approaches.

In subsequent work Ferro & Segers (2003) proposed the *intervals estimator*, based on a limiting characterization of the rescaled inter-exceedance intervals: with probability θ an arbitrary exceedance is the last of a cluster, and then the time to the next exceedance has an exponential distribution with mean $1/\theta$; otherwise the next exceedance belongs to the same cluster, and occurs after a (rescaled) time 0. Thus the inter-exceedance distribution is $(1-\theta)\delta_0 + \theta\exp(\theta)$, where δ_0 and $\exp(\theta)$ represent a delta function with unit mass at 0 and the exponential distribution with mean $1/\theta$. The parameter θ can be estimated from the marginal inter-exceedance distribution in a variety of ways, of which the best seem to be due to Süveges (2007). The intervals estimator can be made automatic once the threshold has been chosen, and it also provides an automatic approach to declustering and thus to the estimation of cluster characteristics, including the cluster size distribution π. It can also be used to diagnose inappropriate thresholds (Süveges and Davison, 2010).

Laurini & Tawn (2003) suggest a two-thresholds approach, according to which a cluster starts with an exceedance of a higher threshold and ends either when the process drops below a lower threshold before another such exceedance, or after a sufficiently long period below the higher threshold. Although theoretical investigation of its properties is difficult, they establish numerically that their estimator is more stable than most of those above.

One reason to attempt declustering is that, as mentioned above, under the limiting model for threshold exceedances, the marginal distribution of an exceedance is the same as that of a cluster maximum; this is a consequence of length-biased sampling. Thus reliable estimates and uncertainty measures of the generalized Pareto distribution of exceedances may be obtained from the (essentially independent) cluster maxima; this is the basis of the *peaks over threshold* approach to modelling extremes. Its application requires reliable identification of cluster maxima, however, and Fawcett & Walshaw (2007, 2012) establish that the difficulty of this can lead to severe bias. This bias can be reduced by using all exceedances to estimate the GPD, though then the standard errors must be modified to allow for the dependence. Eastoe & Tawn (2012) suggest an alternative sub-asymptotic model for cluster maxima, with diagnostics of its appropriateness.

The threshold approach allows the modelling of cluster properties, for example using first-order Markov chains (Smith *et al.*, 1997; Bortot & Coles, 2003), which are estimated using a likelihood in which the extremal model is presumed to fit only those observations exceeding the threshold, with the others treated as censored. Standard bivariate extremal models can be used to generate suitable Markov chains, and so can near-independence models (Ledford & Tawn, 1997; Bortot & Tawn, 1998; Ramos & Ledford, 2009; de Carvalho & Ramos, 2012). Further papers on modelling dependence in clusters include Coles *et al.* (1994) and Fawcett & Walshaw (2006a,b). The use of self-exciting process models for clustering of extreme events in financial time series is described by Chavez-Demoulin *et al.* (2005) and Embrechts *et al.* (2011). Nonparametric estimation of cluster properties is discussed by Segers (2003).

2.3. Extremogram

The correlogram plays a central role in the exploratory analysis of time series, and attempts have been made to extend it to extremes, the goal being to try and estimate the limiting probabilities

$$\rho_h = \lim_{u\to\infty} \mathrm{P}\big(X_h > u \mid X_0 > u\big) \ ,$$

or, if $\rho_h = 0$, to attempt to distinguish different rates at which the convergence takes place. Under the assumptions that the marginal distribution of $\{X_i\}$ is unit Fréchet and that $\mathrm{P}(X_h > u \mid X_0 > u) \sim L_h(u)\, u^{1-1/\eta}$ for some slowly-varying function L_h and some $\eta_h \in (0, 1]$, Ledford & Tawn (2003) suggest plotting estimates of ρ_h and $\Lambda_h = 2\,\eta_h - 1$ as functions of h. If $\eta_h = 1$ and $L_h(u) \to \rho_h > 0$ as $u \to \infty$, then X_0 and X_h are asymptotically dependent, so the first of these plots, called an *extremogram* by Davis & Mikosch (2009b), provides an estimate of the extremal dependence at lag h. By contrast, if $\eta_h < 1$, then the limiting probability will equal zero, and the values of Λ_h better summarize the level of dependence among the asymptotically-independent extremes: $\Lambda_h > 0$ corresponds to positive extremal association, $\Lambda_h = 0$ to so-called near-independence, and $\Lambda_h < 0$ to negative extremal association. Natural estimators of ρ_h may be defined in terms of ratios of indicator functions for finite u, and their significance assessed by permuting the original series (Davis & Mikosch, 2009b), but the joint probability model corresponding to the equivalence above is needed to estimate Λ_h using maximum likelihood (Ledford & Tawn, 2003).

Figure 2, which shows the daily returns of Google from 19 August 2004 to 10 February 2012, displays the volatility clustering that is often seen in financial time series. This is supported by the upper panels of Figure 3, which show the correlograms for the returns themselves and for their absolute values; the correlogram for the values themselves shows little structure, while that of the ab-

solute values shows rather long-term volatility. The lower panels show estimates of ρ_h, with u taken at the 99% quantile of the absolute values of the log returns.

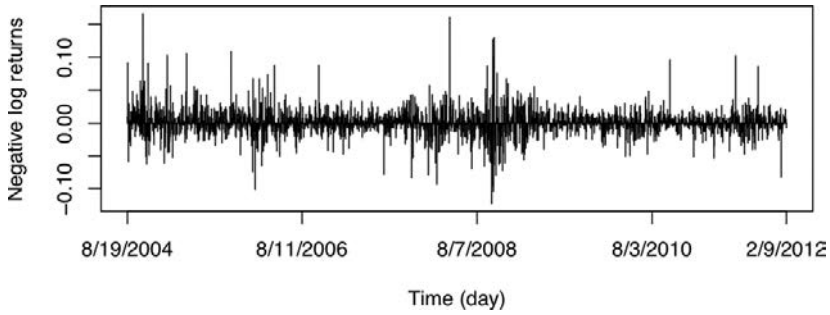


Figure 2: Google daily returns, from 19 August 2004 to 10 February 2012.

There is again little structure in the plot for the returns themselves, but that for their absolute values shows positive dependence of extremes over around 5 days. The computations of Davis & Mikosch (2008, 2009a) imply that a GARCH model would be preferred here, rather than a stochastic volatility model.

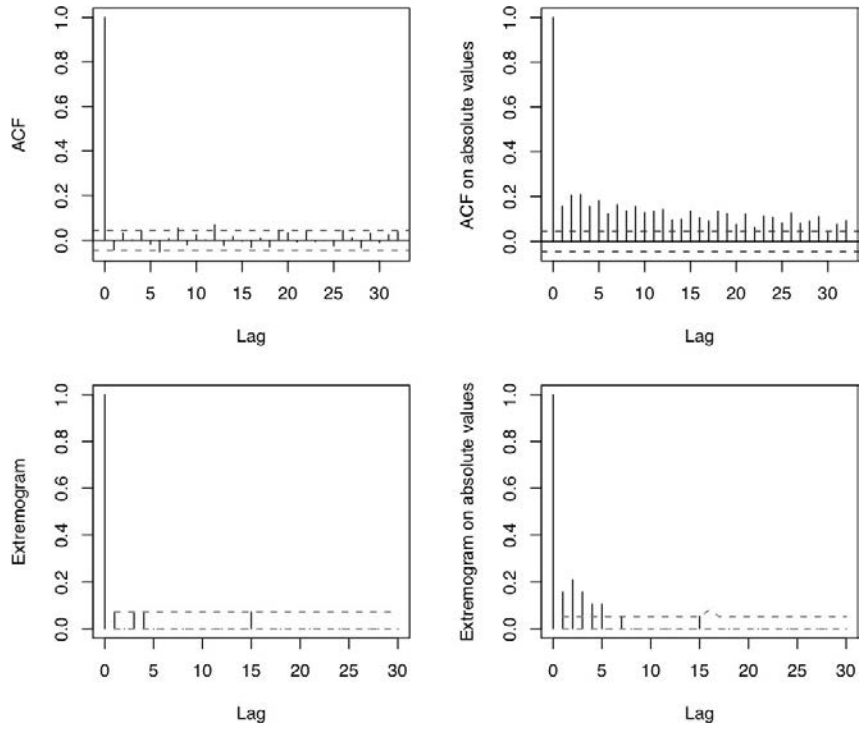


Figure 3: Correlogram (upper panels) and extremogram (lower panels) of the Google returns (left panels) and absolute returns (right panels), with 95% confidence bands for independence.

In the asymptotically dependent case (Davis & Mikosch, 2009b) extend the idea to broader sets of events $\mathcal{A}$ and $\mathcal{B}$ bounded away from zero, defining

$$\rho_{\mathcal{A},\mathcal{B}}(h) \,=\, \lim_{u\to\infty} \mathrm{P}\big(X_h^k \in u\mathcal{B} \mid X_0^k \in u\mathcal{A}\big)\,, \qquad h = 0, 1, 2, \dots\,,$$

if it exists, where X_h^k denotes $(X_h, \dots, X_{h+k})$ for some finite k, which yields ρ_h when $\mathcal{A} = \mathcal{B} = (1, \infty)$ and $k = 0$, but encompasses also events such as $\mathcal{A} = \{X_0 > u\}$, $\mathcal{B} = \{X_1 > u\} \cup \dots \cup \{X_k > u\}$, corresponding to at least one large positive value in the k time steps following a large positive value. The idea can be extended to multiple time series (Huser & Davison, 2012).

2.4. Hill's estimator

Let $\{X_i\}$ denote a sequence of random variables with common marginal distribution F, where $\bar{F} = 1 - F$ is regularly varying at ∞, i.e., there exists an $\alpha > 0$ such that

$$\bar{F}(tx)/\bar{F}(x) \to t^{-\alpha}\,, \qquad t > 0\,, \;\; x \to \infty\,,$$

or equivalently $\bar{F}(x) = x^{-\alpha}L(x)$, $x > 0$, for some slowly varying function $L(x)$. Given a sequence $X_1, \dots, X_n$ with j^{th}-largest value $X_{(j)}$, the Hill estimator

$$H_n \,=\, k^{-1} \sum_{j=1}^{k} \big(\log X_{(j)} - \log X_{(k+1)}\big)$$

is widely used to estimate α^{-1}. This estimator and its variants are widely used for independent heavy-tailed data, and it has been intensively studied. Beirlant *et al.* (2012) give a recent overview of its properties, and Beirlant *et al.* (2004, §10.6) discuss then-known results for dependent data; see also Drees (2003). When covariates are recorded simultaneously with the variable of interest, estimators of the tail index that depend on the covariates have been suggested by Beirlant & Goegebeur (2003), Wang & Tsai (2009) and Gardes *et al.* (2011).

3. BAYESIAN MODELLING

The use of Bayesian methods in statistics has grown vastly over the past two decades, owing to the development of computational tools for fitting complex models, and although Coles & Powell (1996) could write that 'there are only very few papers linking the themes of extreme value modelling and Bayesian inference', the situation has since greatly changed. From a practical viewpoint Bayesian methods have several advantages: they allow the insertion of prior information

leading to coherent inference; they may correspond to penalized estimators that result in stable inferences; and they provide a computationally straightforward way to 'borrow strength' across many related datasets through hierarchical modelling. As in other applications, the main difficulties are the logical status and appropriateness of the prior, and the computational burden, which can lead to too much effort being placed on programming and related matters, and too little on sensitivity analysis and other scientifically relevant aspects. In the study of sample extremes, appropriate prior information can be particularly valuable, because of the sparsity of rare events, but this implies that particular care is needed when choosing priors and assessing their effects. Moreover heavy tails may lead to problems with convergence of empirical estimates of posterior predictive distributions; similar problems arise with the bootstrap (Wood, 2000).

In the simplest setting of estimation based on independent annual maxima, it is straightforward to compute posterior distributions for the GEV parameters and quantities such as return levels, for example using the `R` package `evdbayes`. Very often the prior is chosen in a semi-automatic way, for example using a trivariate normal prior for the location, log-scale and shape parameters of the GEV distribution, but Coles & Tawn (1996) suggest that it will be easier for an expert to formulate prior beliefs in terms of its quantiles. They propose using independent gamma priors for differences of three quantiles, though clearly there are alternatives, such as placing beta priors on probabilities of exceeding certain levels. More general discussion of prior elicitation based on quantiles is given in Dey & Liu (2007). Quantiles may however be more strongly dependent *a priori* than are location and scale parameters, so that prior information on their dependence is needed, and this may be hard to elicit reliably. Ribereau *et al.* (2011) discuss the implications for estimation of parameter orthogonality for the GPD. As is often the case, weak prior information provides inferences that are essentially indistinguishable from those based on likelihood alone, whereas an informative prior may strongly influence extrapolation beyond the data, greatly reducing the associated uncertainty.

A common problem when fitting the GEV or GPD to small samples is absurd estimates of the shape parameter, owing to its large uncertainty. One way to deal with this is through robust estimation (Dupuis & Field, 1998; Dupuis & Morgenthaler, 2002; Dupuis, 2005), but another is through a penalty function corresponding to a prior. Martins & Stedinger (2000) suggest the use of maximum likelihood estimation modulated by a beta prior ensuring that $|\hat{\xi}| < 1/2$, and this does indeed produce improved estimators for the hydrological studies they consider, essentially by trading a small potential bias for a large variance reduction.

In more complex settings it is common to allow the parameters of extremal models to vary with space, time or some covariate. Examples are Coles & Casson (1998), Casson & Coles (1999), Fawcett & Walshaw (2006a), Cooley *et al.* (2006), Cooley *et al.* (2007) and Sang & Gelfand (2009). In such models the location and

log-scale parameters are commonly assumed to be sampled from an underlying Gaussian process, whose spatial structure allows both smooth local variation in these parameters and borrowing of strength across locations, leading to better estimates than would be provided using individual station data. Depending on the setting, it may be useful to constrain the parameters: very often the difficulty of estimating the shape parameter means that a common value is used, and sometimes the (scale parameter)/(location parameter) ratio is close to constant; if so, the complexity of the prior can be reduced; see §4.3. The simplest such models treat the data as independent, conditional on these processes, but more sophisticated models using copulas to allow spatial dependence beyond this have been suggested by Sang & Gelfand (2010) and Fuentes *et al.* (2012). Cooley *et al.* (2012) and Davison *et al.* (2012) give more extensive reviews of spatial extremes, including Bayesian modelling.

4. NON-STATIONARITY

4.1. Generalities

Stationary time series rarely arise in applications, where seasonality, trend, regime changes and dependence on external factors are the rule rather than the exception, and this must be taken into account when modelling extremes. There are broadly two strategies: first, to use the full dataset to detect and estimate non-stationarities, and then to apply methods for stationary extreme-value modelling to the resulting residuals; and, second, to fit an non-stationary extremal model to the original data. An example of the first strategy is McNeil & Frey (2000), who use the GPD to estimate conditional value-at-risk and expected shortfall in financial data after first removing volatility clustering by fitting a GARCH model. An example of the second strategy is Maraun *et al.* (2009), who fit the GEV with seasonally-varying parameters to monthly maxima of many parallel time series, in a study of seasonal variation in heavy precipitation across the United Kingdom.

A benefit of the first strategy, i.e., using the full dataset, is that any non-stationarities will be estimated much more precisely than would be the case based on the extremal data alone. If the extremes of the residuals of this fit are heterogeneous, however, then it will be necessary to model this directly. With daily temperature data, for example, residuals for summer maxima may have shorter tails than do those for winter maxima, so even if seasonal variation in the location and scale of the bulk of the data has been removed, non-stationarity persists in the extremes. Thus two models for non-stationarity are needed, one for the bulk of the data, and another for the extremes, and as in other two-stage fitting procedures, it may be awkward to combine their respective uncertainties.

Thus it is critical that the model for the bulk also removes non-stationarities in the extremes, so far as possible. One approach to this is described by Eastoe & Tawn (2009), who apply the Box–Cox transformation

$$\frac{Y_i^{\lambda(x_i)} - 1}{\lambda(x_i)} = \mu(x_i) + \sigma(x_i) Z_i$$

to the original time series $\{Y_i\}$, where the power transformation $\lambda(x_i)$ and the location and log scale parameters $\mu(x_i)$ and $\log \sigma(x_i)$ depend linearly on covariates x_i, which themselves vary with time. The residuals, which are estimates of the series $\{Z_i\}$, are modeled using a fixed threshold and a possibly time-varying GPD distribution. Eastoe & Tawn (2009) show that this approach can be appreciably more efficient than direct modelling of the extremes, even though the latter is typically simpler, at least when a fixed threshold is applied.

The main benefit of the alternative approach is its simplicity: a fixed threshold is applied, and its exceedance probability and the GPD parameters are modeled directly, without reference to the bulk of the data. A fixed threshold will often have a simple interpretation in terms of the underlying problem, making this strategy attractive despite the loss of statistical efficiency. However a time-varying threshold is preferable for more precise estimation of regression effects. It can be estimated using for example quantile regression (Northrop & Jonathan, 2011), trigonometric functions (Coles *et al.*, 1994) or by other approaches (e.g., de Carvalho *et al.*, 2012), though the difficulty of combining uncertainties from two separate models, one for the threshold and another for the extremes, again arises. An alternative that avoids modelling the threshold (Frossard, 2010; Chavez *et al.*, 2011; Frossard *et al.*, 2012) is to divide the data into homogeneous blocks, and then to base estimation on the largest r observations in each block, with parameters dependent on time and other covariates. In effect this takes the r^{th} largest observation in the block as the threshold, but includes its contribution to the likelihood, so there is just one model to be estimated; this will give results similar to the ideas in Smith (1989).

Using either strategy it is best to use the GEV parametrization of the extremal model, because the GPD parameters are not threshold-invariant. If the scale and shape parameters of the fitted GPD at threshold u are $\sigma_u(x)$ and $\xi(x)$, where x is a covariate, then at a higher threshold v they become $\sigma_v(x) = \sigma_u(x)(v-u) + \xi(x)$ and $\xi(x)$, so as the threshold changes the scale parameter varies with covariates in an unnatural way, unless $\xi(x) \approx 0$. Typically the covariates will enter the model linearly for the location, log scale and shape parameters, though other forms of dependence may be suggested in particular contexts.

The wide variety of possible ways in which covariates might enter the model makes likelihood estimation attractive: not only is it efficient when the model is well-chosen, but it can deal with censoring, rounding and related issues in a simple and unified way. Typically the clustering of rare events will be difficult to

model parametrically, however, and if the main goal is to model non-stationarity, it will be simpler to use an independence likelihood, which treats extreme observations as if they were independent, but then inflates standard errors to allow for unmodelled dependence (Chandler & Bate, 2007). As the limiting marginal distributions of cluster maxima and exceedances are the same, no bias should be incurred, provided the marginal model is correctly specified. The block bootstrap can also be used to assess uncertainty; it is typically applied to residuals, as in Chavez-Demoulin & Davison (2005).

4.2. Semi-parametric models

Non- or semi-parametric modelling, in which more flexible forms of dependence on covariates are used to supplement parametric forms, may be useful, particularly for exploratory analysis or for model-checking. There are two main approaches to this, based on local likelihood estimation and based on penalized regression, and we now briefly describe these. For purposes of exposition we suppose that the location parameter of the GEV distribution is to be modeled as a linear function of covariates x and as an unspecified function of a further covariate t, so that we take $\mu(x,t) = x^{\mathrm{T}}\beta + g(t)$, where g is a smooth nonlinear function to be estimated from the data. In principle it is straightforward to include several smooth terms depending on different covariates, or to include smooth formulations for the shape and scale parameters, though in practice limitations may be imposed by computational considerations or parametrization issues (Chavez-Demoulin & Davison, 2005). Hastie & Tibshirani (1990), Green & Silverman (1994), Fan & Gijbels (1996), Denison *et al.* (2002), Ruppert *et al.* (2003) and Wood (2006) give some entry points to the vast literature on nonparametric regression.

Local likelihood estimation involves polynomial expansion of $g(t)$ around a target value t_0 at which estimation is required, for example writing $g(t) \approx g(t_0) + (t - t_0)\, g'(t_0)$, and then estimating $g(t_0)$ by maximizing a locally-weighted likelihood, in which observations with t distant from t_0 are given less weight than those for which $t - t_0$ is small. The procedure is then repeated for a range of values of t_0, and the corresponding estimates of $g(t_0)$ are interpolated to form an estimate of $g(t)$. The relative weights given to the observations are determined by a bandwidth, a key parameter that can be varied to see the effects of different degrees of smoothing or chosen automatically, for example by cross-validation. The degree of smoothness is often expressed in terms of an equivalent degrees of freedom, which is a decreasing function of the bandwidth. The use of an odd-order polynomial reduces boundary bias, and thus typically a linear polynomial expansion is used. Davison & Ramesh (2000), Hall & Tajvidi (2000), Ramesh & Davison (2002), Butler *et al.* (2007) and Süveges (2007) have applied this in

different settings, including spatial extremal analysis for oceanography and time-varying estimation of the extremal index.

An alternative and in many ways more satisfactory approach is to replace the function $g(t)$ with a linear combination of suitable basis functions, $\alpha_0 + \alpha_1 t + B(t)\,\gamma$, where the columns of the matrix $B(t)$ are typically chosen to span a space orthogonal to that generated by the term $\alpha_0 + \alpha_1 t$. Spline or other basis functions with bounded support are generally used in order to limit the impact of outliers and non-local effects, to which polynomial fits are vulnerable. Spline modelling is underpinned by an elegant theory with links to optimal prediction of stochastic processes, has generally good computational properties, and suitable software is widely available. The number of basis functions may be fixed in advance, or may increase with sample size; in the latter case the penalized likelihood $\ell(\beta, \gamma, \sigma, \xi) - \lambda \gamma^{\mathrm{T}} K \gamma / 2$ is maximized, where the penalty depends on a positive definite matrix K that depends on the basis functions. The weight given to the penalty is determined by a positive quantity λ, with larger λ giving a strong penalization and thus a smoother fit, and conversely. Thus λ plays the same role as the bandwidth in local likelihood estimation, though an elegant link to random effects models may be used to choose λ by maximizing a marginal likelihood (Padoan & Wand, 2008). This approach fits readily into a general regression framework and has been thoroughly investigated (Ruppert *et al.*, 2003); it can also be easily applied using Bayesian computational tools (Laurini & Pauli, 2009). The penalized likelihood approach has been applied to various extremal models by Pauli & Coles (2001) and Chavez-Demoulin & Davison (2005). Yee & Stephenson (2007) place it in a general computational setting.

4.3. Examples

Figure 4 shows the superposed monthly maximum river flow at the station Muota-Ingenbohl, Switzerland, for the years 1923–2008. There is an exceptionally high value in August 2005, though it does not appear to be an outlier. The non-stationarity of the monthly maxima can be fitted by a nonparametric GEV with time dependent location parameter $\mu = \mu(m, t)$ where m is the month and t the year. We suppose that the scale parameter satisfies $\sigma(m) = c\,\mu(m, t)$, for some $c > 0$, and adapt the nonparametric smoothing approach of Chavez-Demoulin & Davison (2005) for peaks over thresholds to our GEV model, which can be written as

$$Z_{m,t} \sim \mathrm{GEV}\Big(\mu\big(\{m, \mathrm{df}_m\}; \{t, \mathrm{df}_t\}\big),\, c\mu\big(\{m, \mathrm{df}_m\}; \{t, \mathrm{df}_t\}\big),\, \xi\Big), \tag{4.1}$$

where df_m and df_t stand for "degrees of freedom" and control the smoothness of the fitted curves for months and years. Technical details for the peaks over threshold setting, including selection of the degrees of freedom and confidence interval calculation, are given in Chavez-Demoulin & Davison (2005).

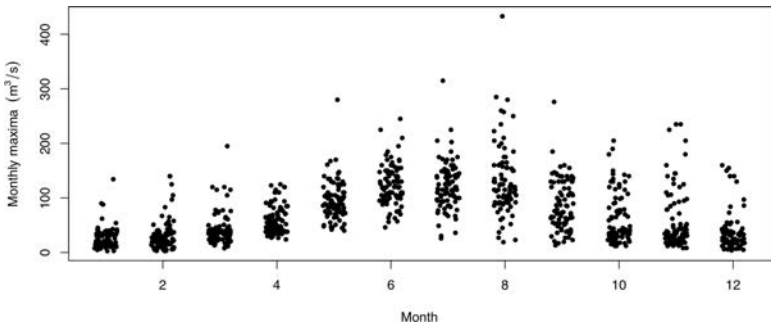


Figure 4: Monthly maximum river flow ($m^3 s^{-1}$), jittered, at Muota-Ingenbohl, Switzerland, for the years 1923–2008.

The estimated curves

$$\hat{\mu}\big(\{m, \hat{\mathrm{df}}_m\}; \{t, \hat{\mathrm{df}}_t\}\big)\,, \qquad \hat{\sigma}\big(\{m, \hat{\mathrm{df}}_m\}; \{t, \hat{\mathrm{df}}_t\}\big)$$

and estimated parameter $\hat{\xi}$ are shown in Figure 5. The constant c is estimated to be 0.64, with 95% confidence interval $[0.61, 0.66]$, so $\hat{\sigma}(m)$ has the same shape as

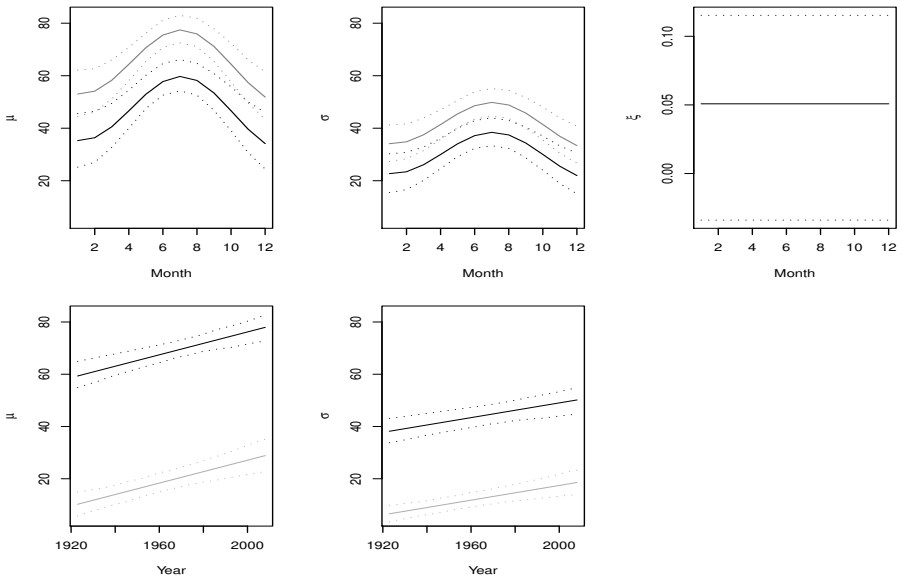


Figure 5: Muota-Ingenbohl data. The upper left panel shows the estimated location parameter $\hat{\mu}(m,t)$ over month for the year $t = 1924$ (black) and $t = 2005$ (red, upper), with 95% pointwise bootstrap confidence intervals (dots). The upper middle panel shows the estimated scale parameter $\hat{\sigma} = \hat{c}\hat{\mu}(m,t)$ for the year $t = 1924$ (black) and $t = 2005$ (red, upper). The upper right panel shows the estimated shape parameter $\hat{\xi}$. The lower left panel shows the estimated location parameter $\hat{\mu}(m,t)$ over year for July, $m = 7$, (black) and January, $m = 1$, (green, lower). The lower middle panel shows the estimated scale parameter $\hat{\sigma} = \hat{c}\hat{\mu}(m,t)$ over year for July, $m = 7$, (black) and January, $m = 1$, (green, lower).

the location parameter curve. The model selected using the AIC has $\hat{\mathrm{df}}_m = 2$ for the variable month and a linear trend ($\hat{\mathrm{df}}_t = 1$) for the year, with slope $0.22\,\mathrm{m}^3\mathrm{s}^{-1}$/year, giving an annual increase of both location and scale parameters.

Figure 6 shows the estimated 100-year return level curve against month for $t = 2005$ and the estimated 100-year return level curve against year for July. The points in the left panel are the largest monthly values for 2005; they show how unusual the August value that year was. Those in the right panel are July observations from 1923 to 2008, which have been used to estimate the GEV parameters. The 100-year return level slope evaluated in July has an annual increase of $0.53\,\mathrm{m}^3\mathrm{s}^{-1}$. The upper confidence interval bound was exceeded once, so the estimation appears realistic. The confidence limits are rather narrow, but there are 12 times more observations than appear in the panel.

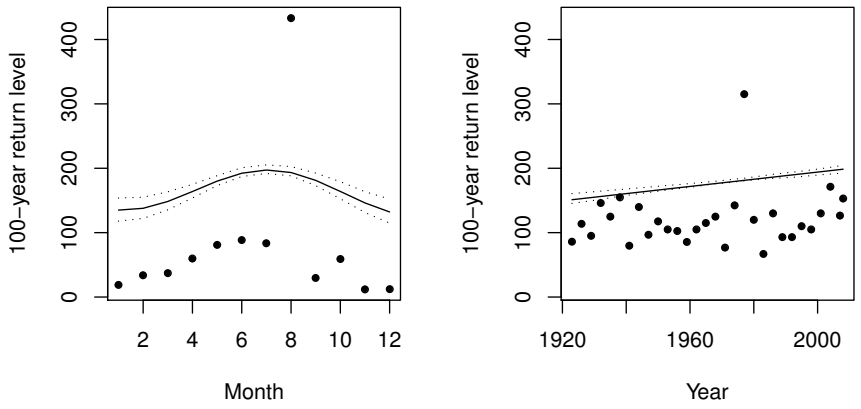


Figure 6: Muota-Ingenbohl data. Left panel: estimated 100-year return level curve against month for $t = 2005$. The dotted lines are pointwise 95% confidence intervals. The points are the largest monthly values of 2005. Right panel: Estimated 100-year return level curve against year for July, $m = 7$. The points are 30 observations during July from 1923 to 2008.

As a second example, we take data from the US National Oceanic and Atmospheric Administration (`http://www.ngdc.noaa.gov/`) on Japanese tsunamis from 1400–2011. The upper panel of Figure 7 shows the log maximum water height above sea level in meters (not to be confused with the elevation at the limit of inundation, called a run-up height) during a tsunami due to a preceding earthquake. The maximum water height of 85.4m appeared in 1771, due to a earthquake of magnitude 7.4 in Ryukyu Islands that led to around 13,500 deaths. The most recent events are the 54m water height in Sea of Japan that succeeded an earthquake of magnitude 7.7 in 1993, leading to 208 deaths, and the 2011 event in Honshu, with a preceding earthquake of magnitude 9, which led to 15,550 deaths. With such data there are obvious concerns about changes in

measurement and estimation of the earlier heights, and the increasing frequency of events is probably also due to improved record-keeping. With this in mind we focus on the amplitudes, using a GPD model for the water heights above a threshold of 0.6 m. The lower panel of Figure 7 shows the logarithm of the maximum water height above the sea level in meters against the logarithm of the earthquake magnitude preceding the tsunami for such events. We model the maximum tsunami water height as a function of the magnitude x of the preceding earthquake, plus a function of year t, giving

$$\beta_0 + \beta_1 m(x) + g(t) \;\approx\; \beta_0 + \beta_1 a_1(x) + \cdots + \beta_q a_q(x) + \gamma_1 b_1(t) + \cdots + \gamma_p b_p(t) \;,$$

in terms of suitable basis functions.

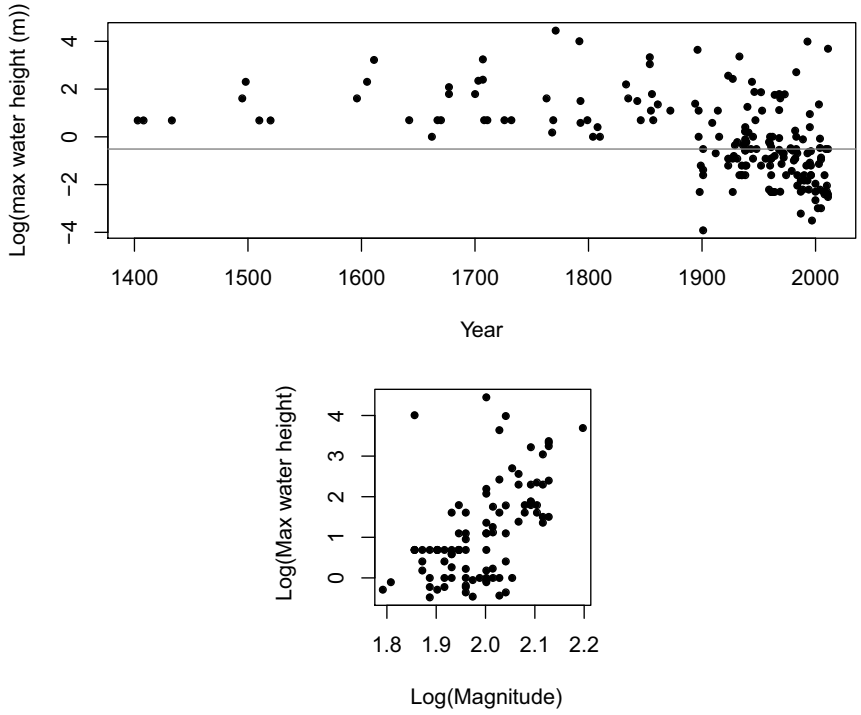


Figure 7: Tsunami data: The upper panel shows the logarithm of maximum water height above the sea level in meters for each tsunami from 1403 to 2011, the horizontal red line is the threshold 0.6 m in logarithm. The lower panel shows the logarithm of the maximum water height for each tsunami above the sea level in meters above a threshold of 0.6 m against the logarithm of the earthquake magnitude preceding the tsunami.

As pointed out by Yee & Stephenson (2007), nonparametric estimation of both scale and the shape parameters may be problematic in small datasets, owing to the difficulty in estimating the shape, and the non-orthogonality of these parameters. In this case the model selected among the various parametric and

nonparametric models, fitted using the approach of Chavez-Demoulin & Davison (2005), gives $\hat{\xi} = 0.47$ (0.12), linear dependence on earthquake magnitude and three degrees of freedom for the dependence on time; see the lower panels of Figure 8. This figure also shows the corresponding estimates for the GPD vector generalized additive model of Yee & Stephenson (2007) and the generalized linear mixed model representation for the extreme value spline model of Laurini & Pauli (2009). There is reassuringly little to choose between the fits. The approach of Laurini & Pauli (2009) is slowest but uncertainty on the equivalent degrees of freedom is accounted for, and this leads to slightly wider confidence intervals, whereas the Yee & Stephenson (2007) approach is overall somewhat less flexible in terms of the modelling possibilities.

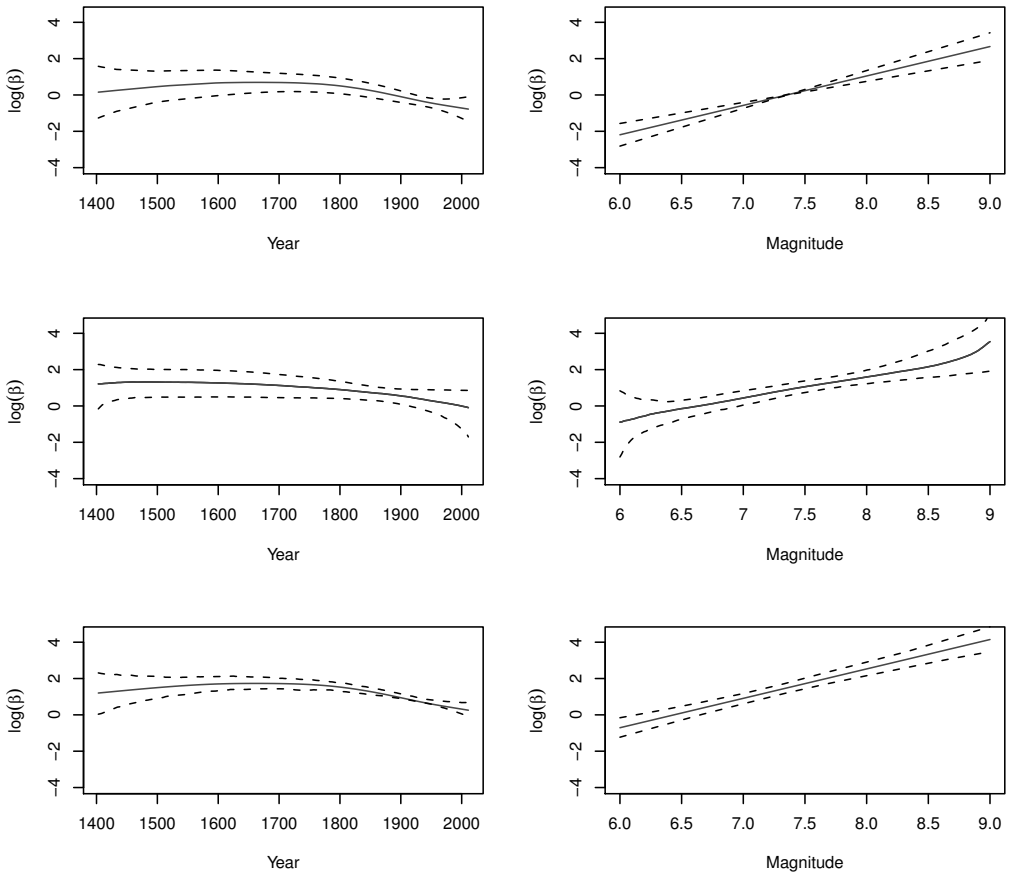


Figure 8: Tsunami data. Nonparametric estimation of the logarithm of the GPD scale parameter using methods of Yee & Stephenson (2007) (upper panels), Laurini & Pauli (2009) (middle panels) and Chavez-Demoulin & Davison (2005) (lower panels). The left panels show the estimated dependence on year and the right panels show the estimated dependence on earthquake magnitude, with 95% pointwise confidence intervals.

5. DISCUSSION

Although impressive progress has been made in modelling time series extremes over the past two decades, certain topics still require further investigation. One, an overarching theme in extreme-value statistics, is the relevance of asymptotic theory to applications. At the sub-asymptotic levels that can be observed in practice the limiting results provide approximations that may be poor in certain cases, and it is necessary to expand the theory. The resulting pre-asymptotic models often prove difficult to fit, however, and so care is needed when providing tools that are useful for practice. For example, it would be valuable to have available some broad classes of models for clusters, beyond first-order Markov chains and able to encompass both dependent and near-independent extremes, perhaps based on developments of Heffernan *et al.* (2007), Fougères *et al.* (2009) or Bortot & Gaetan (2011). One interesting class of models for multivariate series is the so-called multivariate maxima of moving maxima process (Zhang & Smith, 2010), and it would be valuable to further develop suitable inference procedures, for example along the lines suggested by Süveges and Davison (2012), and more broadly to assess whether such models are broadly adequate for use in applications; there is a close connection to extremal modelling with mixtures. A related topic of interest is further investigation of extremal properties of standard time series models, including the effect of discretisation of continuous-time processes. A potentially important advance would be the development of full likelihood inference for time series extremes, perhaps based on an EM algorithm or suitable Kalman filter. Absent this, it is tempting to use the independence likelihood (Chandler & Bate, 2007) or related approaches for estimating marginal properties of extremal time series, but inference for this could be further developed.

Analogues of the extremal index beyond time series are well-studied for asymptotically dependent data, but deserve fuller attention for near-independence models.

Various classical topics also merit further study. One is the choice of threshold for peaks over threshold analysis of dependent data, based on many related series that display seasonality; the methods reviewed by Scarrott & MacDonald (2012) are relevant. Others are extremal index estimation at sub-asymptotic levels, particularly in many series and detection of regime change — often confounded with long-range dependence in classical time series — in extremes.

ACKNOWLEDGMENTS

The work was supported by the Swiss National Science Foundation.
We thank the editors for helpful comments.

REFERENCES

Ancona-Navarrete, M.A. & Tawn, J.A. (2000). A comparison of methods for estimating the extremal index, *Extremes*, **3**, 5–38.

Beirlant, J.; Caeiro, F. & Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes, *Revstat*, **10**, 1–31.

Beirlant, J. & Goegebeur, Y. (2003). Regression with response distributions of Pareto-type, *Comp. Statist. Data Anal.*, **42**, 595–619.

Beirlant, J.; Goegebeur, Y.; Teugels, J. & Segers, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, New York.

Bortot, P. & Coles, S.G. (2003). Extremes of Markov chains with tail-switching potential, *J. R. Statist Soc. B*, **65**, 851–867.

Bortot, P. & Gaetan, C. (2011). *A latent process model for temporal extremes.* Submitted.

Bortot, P. & Tawn, J.A. (1998). Models for the extremes of Markov chains, *Biometrika*, **85**, 851–867.

Butler, A.; Heffernan, J.E.; Tawn, J.A. & Flather, R.A. (2007). Trend estimation in extremes of synthetic North Sea surges, *Appl. Statist.*, **56**, 395–414.

Casson, E. & Coles, S. (1999). Spatial regression models for extremes, *Extremes*, **1**, 449–468.

Chandler, R.E. & Bate, S. (2007). Inference for clustered data using the independence loglikelihood, *Biometrika*, **94**, 167–183.

Chavez, V.; Davison, A.C. & Frossard, L. (2011). Discussion of "Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights" by P.J. Northrop and P. Jonathan, *Environmetrics*, **22**, 810–812.

Chavez-Demoulin, V. & Davison, A.C. (2005). Generalized additive models for sample extremes, *Appl. Statist.*, **54**, 207–222.

Chavez-Demoulin, V.; Davison, A.C. & McNeil, A.J. (2005). A point process approach to value-at-risk estimation, *Quant. Finance*, **5**, 227–234.

Chernick, M.R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distributions, *Ann. Prob.*, **9**, 145–149.

Chernick, M.R.; Hsing, T. & McCormick, W. (1991). Calculating the extremal index of a class of stationary sequences, *Adv. Appl. Prob.*, **23**, 835–850.

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

Coles, S.G. & Casson, E. (1998). Extreme value modelling of hurricane wind speeds, *Struct. Safety*, **20**, 283–296.

Coles, S.G. & Powell, E.A. (1996). Bayesian methods in extreme value modelling: A review and new developments, *Int. Statist. Rev.*, **64**, 119–136.

Coles, S.G. & Tawn, J.A. (1996). A Bayesian analysis of extreme rainfall data, *Appl. Statist.*, **45**, 463–478.

Coles, S.G.; Tawn, J.A. & Smith, R.L. (1994). A seasonal Markov model for extremely low temperatures, *Environmetrics*, **5**, 221–239.

Cooley, D.; Cisewski, J.; Erhardt, R.J.; Jeon, S.; Mannshardt, E.; Omolo, B.O. & Sun, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects, *Revstat*, **10**, 135–165.

Cooley, D.; Naveau, P.; Jomelli, V.; Rabatel, A. & Grancher, D. (2006). A Bayesian hierarchical extreme value model for lichenometry, *Environmetrics*, **17**, 555–574.

Cooley, D.; Nychka, D. & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Statist. Assoc.*, **102**, 824–840.

Davis, R.A. & Mikosch, T. (2008). *Extreme value theory for GARCH models.* In "Handbook of Financial Time Series" (T.G. Anderson, R.A. Davis, J.-P. Kreiß and T. Mikosch, Eds.), pp. 186–200, Springer, Berlin.

Davis, R.A. & Mikosch, T. (2009a). *Extremes of stochastic volatility models.* In "Handbook of Financial Time Series" (T.G. Anderson, R.A. Davis, J.-P. Kreiß and T. Mikosch, Eds.), pp. 355–364, Springer, Berlin.

Davis, R.A. & Mikosch, T. (2009b). The extremogram: A correlogram for extreme events, *Bernoulli*, **15**, 977–1009.

Davis, R.A. & Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities, *Ann. Prob.*, **13**, 179–195.

Davison, A.C.; Padoan, S. & Ribatet, M. (2012). Statistical modelling of spatial extremes (with Discussion), *Statist. Science*, to appear.

Davison, A.C. & Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *J. R. Statist. Soc. B*, **62**, 191–208.

de Carvalho, M. & Ramos, A. (2012). Bivariate extreme statistics, II, *Revstat*, **10**, 83–107.

de Carvalho, M.B.; Turkman, K.F. & Rua, A. (2012). *Nonstationary extremes and the US business cycle*, submitted.

Denison, D.G.T.; Holmes, C.C.; Mallick, B.K. & Smith, A.F.M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*, Wiley, New York.

Dey, D.K. & Liu, J. (2007). A quantitative study of quantile based direct prior elicitation from expert opinion, *Bayesian Anal.*, **2**, 137–166.

Drees, H. (2003). Extreme quantile estimation for dependent data with applications to finance, *Bernoulli*, **9**, 617–657.

Dupuis, D.J. (2005). Ozone concentrations: A robust analysis of multivariate extremes, *Technometrics*, **47**, 191–201.

Dupuis, D.J. & Field, C.A. (1998). Robust estimation of extremes, *Canadian J. Statist.*, **26**, 199–215.

Dupuis, D.J. & Morgenthaler, S. (2002). Robust weighted likelihood estimators with an application to bivariate extreme value problems, *Canadian J. Statist.*, **30**, 17–36.

Eastoe, E.F. & Tawn, J.A. (2009). Modelling non-stationary extremes with application to surface level ozone, *Appl. Statist.*, **58**, 25–45.

Eastoe, E.F. & Tawn, J.A. (2012). Modelling the distribution for the cluster maxima of exceedances of sub-asymptotic thresholds, *Biometrika*, to appear.

Embrechts, P.; Liniger, T. & Lu, L. (2011). Multivariate Hawkes processes: an application to financial data, *J. Appl. Prob.*, **48A**, 367–378.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Fawcett, L. & Walshaw, D. (2006a). A hierarchical model for extreme wind speeds, *Appl. Statist.*, **55**, 631–646.

Fawcett, L. & Walshaw, D. (2006b). Markov chain models for extreme wind speeds, *Environmetrics*, **17**, 795–809.

Fawcett, L. & Walshaw, D. (2007). Improved estimation for temporally clustered extremes, *Environmetrics*, **18**, 173–188.

Fawcett, L. & Walshaw, D. (2012). Estimating return levels from serially dependent extremes, *Environmetrics*, to appear.

Ferro, C.A.T. & Segers, J. (2003). Inference for clusters of extreme values, *J. R. Statist. Soc. B*, **65**, 545–556.

Fougères, A.-L.; Nolan, J.P. & Rootzén, H. (2009). Models for dependent extremes using stable mixtures, *Scandinavian J. Statist.*, **36**, 42–59.

Frossard, L. (2010). *Modelling Minima of Ozone Data in the Southern Mid-Latitudes*, Master's thesis, Ecole Polytechnique Fédérale de Lausanne, Section de Mathématiques.

Frossard, L.; Rieder, H.E.; Ribatet, M.; Staehelin, J.; Maeder, J.A.; Rocco, S.D.; Davison, A.C. & Peter, T. (2012). On the relationship between total ozone and atmospheric dynamics and chemistry at mid-latitudes: Part 1 — Statistical models and spatial fingerprints of atmospheric dynamics and chemistry, *Atmosph. Chem. Physics Disc.*, submitted.

Fuentes, M.; Henry, J. & Reich, B. (2012). Nonparametric spatial models for extremes: Application to extreme temperature data, *Extremes*, to appear.

Gardes, L.; Guillou, A. & Schorgen, A. (2011). *Estimating the conditional tail index by integrating a kernel conditional quantile estimator*, submitted.

Green, P.J. & Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London.

Hall, P. & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statist. Science*, **15**, 153–167.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman & Hall, London.

Heffernan, J.E.; Tawn, J.A. & Zhang, Z. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes, *Extremes*, **10**, 57–82.

Hsing, T. (1987). On the characterization of certain point processes, *Stoch. Proc. Appl.*, **26**, 297–316.

Hsing, T.; Hüsler, J. & Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence, *Prob. Th. Rel. Fields*, **78**, 97–112.

Huser, R. & Davison, A.C. (2012). *Space-time modelling of extreme events*, submitted.

Laurini, F. & Pauli, F. (2009). Smoothing sample extremes: The mixed model approach, *Comp. Statist. Data Anal.*, **53**, 3842–3854.

Laurini, F. & Tawn, J.A. (2003). New estimators for the extremal index and other cluster characteristics, *Extremes*, **6**, 189–211.

Leadbetter, M.R.; Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Ledford, A.W. & Tawn, J.A. (1997). Modelling dependence within joint tail regions, *J. R. Statist. Soc. B*, **59**, 475–499.

Ledford, A.W. & Tawn, J.A. (2003). Diagnostics for dependence within time series extremes, *J. R. Statist. Soc. B*, **65**, 521–543.

Maraun, D.; Rust, H.W. & Osborn, T.J. (2009). The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics, *Int. J. Clim.*, **29**, 1731–1744.

Martins, E. & Stedinger, J. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Res. Research*, **36**, 737–744.

McNeil, A.J. & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *J. Empir. Finance*, **7**, 271–300.

Northrop, P.J. & Jonathan, P. (2011). Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion), *Environmetrics*, **22**, 799–809.

Padoan, S.A. & Wand, M.P. (2008). Mixed model-based additive models for sample extremes, *Statist. Prob. Lett.*, **78**, 2850–2858.

Pauli, F. & Coles, S.G. (2001). Penalized likelihood inference in extreme value analyses, *J. Appl. Statist.*, **28**, 547–560.

Ramesh, N.I. & Davison, A.C. (2002). Local models for exploratory analysis of hydrological extremes, *J. Hydrology*, **256/1–2**, 106–119.

Ramos, A. & Ledford, A.W. (2009). A new class of models for bivariate joint extremes, *J. R. Statist. Soc. B*, **71**, 219–241.

Ribereau, P.; Naveau, P. & Guillou, A. (2011). A note of caution when interpreting parameters of the distribution of excesses, *Adv. Water Res.*, **34**, 1215–1221.

Robinson, M.E. & Tawn, J.A. (2000). Extremal analysis of processes sampled at different frequencies, *J. R. Statist. Soc. B*, **62**, 117–135.

Ruppert, D.; Wand, M.P. & Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.

Sang, H. & Gelfand, A.E. (2009). Hierarchical modeling for extreme values observed over space and time, *Environ. Ecol. Statist.*, **16**, 407–426.

Sang, H. & Gelfand, A.E. (2010). Continuous spatial process models for spatial extreme values, *J. Agric. Biol. Environ. Statist.*, **15**, 49–65.

Scarrott, C. & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *Revstat*, **10**, 33–60.

Segers, J.J. (2003). Functionals of clusters of extremes, *Adv. Appl. Prob.*, **35**, 1028–1045.

Sibuya, M. (1960). Bivariate extreme statistics, I, *Ann. Inst. Statist. Math.*, **11**, 195–210.

Smith, R.L. (1989). Extreme value analysis of environmental time series: An example based on ozone data (with Discussion), *Statist. Science*, **4**, 367–393.

Smith, R.L.; Tawn, J.A. & Coles, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**, 249–268.

Smith, R.L. & Weissman, I. (1994). Estimating the extremal index, *J. R. Statist. Soc. B*, **56**, 515–528.

Süveges, M. (2007). Likelihood estimation of the extremal index, *Extremes*, **10**, 41–55.

Süveges, M. & Davison, A.C. (2010). Model misspecification in peaks over threshold analysis, *Ann. Appl. Statist.*, **4**, 203–221.

Süveges, M. & Davison, A.C. (2012). A case study of a "Dragon King": The 1999 Venezuelan catastrophe, *Euro. Physical J. E*, to appear.

Wang, H. & Tsai, C. (2009). Tail Index Regression, *J. Am. Statist. Assoc.*, **104**, 1233–1240.

Wood, A.T.A. (2000). Bootstrap relative errors and sub-exponential distributions, *Bernoulli*, **6**, 809–834.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, Boca Raton.

Yee, T.W. & Stephenson, A.G. (2007). Vector generalized linear and additive extreme value models, *Extremes*, **9**, 1–19.

Zhang, Z. & Smith, R.L. (2010). On the estimation and application of max-stable processes, *J. Statist. Plann. Infer.*, **140**, 1135–1153.

REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 135–165

A SURVEY OF SPATIAL EXTREMES: MEASURING SPATIAL DEPENDENCE AND MODELING SPATIAL EFFECTS

Authors: Daniel Cooley
– Department of Statistics, Colorado State University, Fort Collins, CO, USA (cooleyd@stat.colostate.edu)

Jessi Cisewski
– Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (cisewski@email.unc.edu)

Robert J. Erhardt
– Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (erhardt@email.unc.edu)

Soyoung Jeon
– Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (soyoung@email.unc.edu)

Elizabeth Mannshardt
– Department of Statistics, The Ohio State University, Columbus, OH, USA (elizabeth@stat.osu.edu)

Bernard Oguna Omolo
– Division of Mathematics and Computer Science, University of South Carolina – Upstate, Spartanburg, SC, USA (bomolo@uscupstate.edu)

Ying Sun
– Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC, USA (sunwards@samsi.info)

Abstract:

- We survey the current practice of analyzing spatial extreme data, which lies at the intersection of extreme value theory and geostatistics. Characterizations of multivariate max-stable distributions typically assume specific univariate marginal distributions, and their statistical applications generally require capturing the tail behavior of the margins and describing the tail dependence among the components. We review current methodology for spatial extremes analysis, discuss the extension of the finite-dimensional extremes framework to spatial processes, review spatial dependence metrics for extremes, survey current modeling practice for the task of modeling marginal distributions, and then examine max-stable process models and copula approaches for modeling residual spatial dependence after accounting for marginal effects.

Key-Words:

- *copula; extremal coefficient; hierarchical model; madogram; max-stable process; multivariate extreme value distribution.*

AMS Subject Classification:

- 62M30, 62H11, 62H20.

1. INTRODUCTION

Assessing the behavior of rare events such as the risk of flooding, potential crop damage from drought, or health effects of potential extreme air pollution events presents unique statistical challenges, and requires one to characterize the tail of the distribution of the quantity of interest. Since many quantities such as rainfall, temperature, or air pollution are measured at specifically-located monitors, spatial modeling is necessary. Applications such as these have motivated the development of methods and tools for analyzing and characterizing spatial extreme data. In this paper, we survey the current practice of spatial extremes. Recently, Davison *et al.* (2012) review spatial extremes methods via a case study of extreme precipitation in Switzerland. This survey can be viewed as complementary to Davison *et al.* (2012), and we aim to provide an entry point for the study of spatial extremes.

An analysis of spatial extreme data lies at the intersection of two branches of statistics: extreme value analysis and geostatistics. Below, we give brief introductions to each field. There are a number of books in each field which give comprehensive overviews. For extremes, references include de Haan & Ferreira (2006), Beirlant *et al.* (2004) and Coles (2001); geostatistics references include Schabenberger & Gotway (2005), Banerjee *et al.* (2004) and Cressie (1993). For many studies, spatial effects are often separated into large scale (i.e., regional) effects and small scale (i.e., local) effects.

In terms of a statistical model, regional spatial effects are often captured by characterizing how the marginal distribution varies over a study region. Local spatial effects are typically described by a dependence structure. Given data, the distinction between the local and regional effects is likely not obvious, and one can view the task of separating these two effects as analogous to the task of decomposing time series data into mean (trend and seasonal effects) and a stationary noise process described by a covariance structure (Brockwell & Davis, 2002, §1.3.3).

Spatial data are necessarily multivariate as they are recorded at multiple locations. Throughout, we will assume that we analyze only one quantity (e.g., rainfall) at multiple sites, although spatially analyzing multiple quantities is a possible extension of the work surveyed herein.

Spatial data are often modeled as a realization of a spatial process which is observed at a finite set of locations. We, too, will consider a spatial process, but we begin with a review of the theory and statistical practice for finite-dimensional extremes in order to introduce important concepts.

1.1. Multivariate extremes: theory and practice

We assume that the reader is familiar with univariate extreme value theory and its statistical application. For background on the univariate case see the monographs by Coles (2001), Beirlant *et al.* (2004), and de Haan & Ferreira (2006).

The notion of max-stability forms the foundation of extreme value theory. Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^{\mathrm{T}}$, $i = 1, \ldots, n$, be an i.i.d. sequence of d-dimensional continuous random vectors, and let $\boldsymbol{M}_n = (M_{n,1}, ..., M_{n,d})^{\mathrm{T}} = \left(\bigvee_{i=1}^n X_{i,1}, ..., \bigvee_{i=1}^n X_{i,d}\right)^{\mathrm{T}}$, where $\bigvee$ denotes the maximum. Assume there exist normalizing sequences $\{\boldsymbol{a}_n\}$ and $\{\boldsymbol{b}_n\}$ such that

$$\Pr\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{y}\right) \longrightarrow G(\boldsymbol{y}) \ , \qquad n \longrightarrow \infty$$

where the division is understood to be element-wise and G is non-degenerate. Then G belongs to the class of multivariate max-stable (equivalently, extreme value) distributions. We will denote by $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$ a max-stable random vector; that is, $\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \xrightarrow{\mathrm{d}} \boldsymbol{Y}$. The i.i.d. assumption can be relaxed and the multivariate max-stable distributions continue to serve as the possible limiting distributions, if certain mixing conditions are met (Leadbetter *et al.*, 1983).

Unlike in the univariate case, no fully parametric representation exists for the multivariate max-stable distributions. The univariate marginal distributions must be univariate max-stable and therefore can be described by the generalized extreme value (GEV) distribution:

$$\Pr\big(Y_j \leq y\big) \ = \ G_j(y) \ = \ \exp\left\{ - \left[\left(1 + \xi_j \, \frac{y - \mu_j}{\sigma_j} \right)_+^{-1/\xi_j} \right] \right\} , \tag{1.1}$$

for $j = 1, \ldots, d$. Here, μ_j, σ_j and ξ_j are the location, scale and shape parameters for the j^{th} component's marginal and $x_+ = \max(0, x)$. Accounting for each GEV marginal is a nuisance, and representations for multivariate max-stable distributions generally presuppose that the marginals have a common, convenient max-stable distribution. Throughout, we will assume $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^{\mathrm{T}}$ has a multivariate max-stable distribution with unit-Fréchet marginals: $\Pr(Z_j \leq z) = \exp(-z^{-1})$. Then

$$\Pr(\boldsymbol{Z} \leq \boldsymbol{z}) \ = \ G^*(\boldsymbol{z}) \ = \ \exp\big\{ -V(\boldsymbol{z}) \big\} \ , \tag{1.2}$$

$$V(\boldsymbol{z}) \ = \ d \int_{\Delta_d} \bigvee_{j=1}^{d} \frac{w_j}{z_j} \, H(\mathrm{d}\boldsymbol{w}) \ . \tag{1.3}$$

Here $\Delta_d = \big\{ \boldsymbol{w} \in \mathbb{R}_+^d \mid w_1 + \cdots + w_d = 1 \big\}$ is the $(d-1)$-dimensional simplex, and the angular (or spectral) measure H is a probability measure on Δ_d, which de-

termines the dependence structure of the random vector. Due to the common marginals, H obeys the moment conditions $\int_{\Delta_d} w_j \, H(d\boldsymbol{w}) = 1/d$ for $j = 1, \ldots, d$.

There is no loss of generality in assuming the multivariate max-stable distribution has unit-Fréchet margins, as Resnick (1987, Prop. 5.10) states that the domain-of-attraction condition is preserved under monotone transformations of the marginal distributions. If $\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n) \xrightarrow{\mathrm{d}} \boldsymbol{Y}$ which does not have unit-Fréchet marginals, from (1.1) one can define marginal transformations

$$T_j(x) = \left(1 + \xi_j \, \frac{x - \mu_j}{\sigma_j}\right)^{-\xi_j}, \qquad j = 1, \ldots, d \; ,$$

and define

$$G^*(z_1, \ldots, z_d) \; = \; G\big\{T_1^{\leftarrow}(z_1), \ldots, T_d^{\leftarrow}(z_d)\big\} \; ,$$

where $T_j^{\leftarrow}$ is the inverse function of T_j, for $j = 1, \ldots, d$. This approach of transforming to convenient marginals is similar to copula approaches, albeit with a marginal suggested by extreme value theory rather than Uniform [0,1].

The above asymptotic theory suggests the following general statistical methodology, referred to as the block maxima approach. Choose n to be a fixed block size which is large enough such that the asymptotic theory holds approximately, and assume a sequence of i.i.d. $\boldsymbol{X}_i$, $i = 1, \ldots, nm$, are observed, where m denotes the number of blocks. Define $\boldsymbol{M}_k = \big(\bigvee_{i=(k-1)n+1}^{kn} X_{i,1}, \ldots, \bigvee_{i=(k-1)n+1}^{kn} X_{i,d}\big)^{\mathrm{T}}$ for $k = 1, \ldots, m$ (note that the dependence on n in the notation $\boldsymbol{M}_k$ has been suppressed), and fit a multivariate max-stable distribution to the $\boldsymbol{M}_k$. It is important to note that $\boldsymbol{M}_k$ will not appear in the observation record unless the occurrence times of each element's block maximum coincide.

Using representation (1.2) to fit a multivariate max-stable distribution requires that the marginals be unit Fréchet. Although transforming the marginals is a simple theoretical procedure, in practice the marginal distributions must be estimated. Subsequently, utilizing (1.2) to perform a multivariate analysis of extremes involves two tasks: (1) estimating the marginals, and (2) characterizing the dependence via a model for $V(\boldsymbol{z})$ or $H(\boldsymbol{w})$. Tasks (1) and (2) seem sequential; however, we note that inference can be performed all-at-once either in the frequentist (Padoan *et al.*, 2010) or Bayesian (Ribatet *et al.*, 2011) settings.

For spatial extremes studies, the aforementioned regional and local spatial effects each can be associated with one of the above tasks. Most study regions are large enough that the marginal distribution of the studied quantity will vary over the region. Thus, in order to transform to a common marginal, one must first account for how the distribution's tail varies by location. The local spatial effect is related to the spatial extent of individual extreme events and the resulting dependence in the data due to multiple sites being affected by the same event. In terms of (1.2), this dependence is captured by $V(\boldsymbol{z})$ or $H(\boldsymbol{w})$. We will refer to

the dependence remaining after the marginal standardization as 'residual' dependence, as Sang & Gelfand (2010) termed the data after marginal transformation as 'standardized residuals'. There is a useful analogy here drawn from atmospheric science: these two types of spatial effects can be thought of as corresponding to 'climate' and 'weather' effects. Climate can be thought of as the distribution of weather (Guttorp & Xu, 2011), and climate varies with location. Performing the marginal transformation is akin to standardizing the climate across the study region. Weather events have a spatial extent which is best captured by a stochastic representation. For most applications, there is a difference in the scale of these two spatial effects. Climate varies on a larger (regional) spatial scale, and can be largely (but often not completely) characterized by covariates such as latitude and elevation. Weather spatial effects, particularly for extremes, are often more localized.

Although all data are finite-dimensional, a finite-dimensional framework can be inadequate for dealing with unobserved locations, and thus most classical spatial work assumes a stochastic process framework. Let $\mathcal{S}$ be a study region, and let $\boldsymbol{s}$ denote a location in a study region. For spatial applications, most often $\boldsymbol{s} \in \mathbb{R}^2$ and we will assume this throughout. We will assume $X_i(\boldsymbol{s})$ is a stochastic process, where it may be helpful to think of i as indexing the day of the observation. A fundamental construct for spatial extremes is the max-stable process, which is the infinite-dimensional analogue to a max-stable random vector. If for all $\boldsymbol{s} \in \mathcal{S}$ there exist normalizing sequences $a_n(\boldsymbol{s})$ and $b_n(\boldsymbol{s})$ such that

$$a_n^{-1}(\boldsymbol{s}) \left\{ \max_{i=1,\ldots,n} X_i(\boldsymbol{s}) - b_n(\boldsymbol{s}) \right\} \xrightarrow{\mathrm{d}} Y(\boldsymbol{s}) \ , \tag{1.4}$$

which has a non-degenerate distribution, then $Y(\boldsymbol{s})$ is a max-stable process. When the max-stable process has unit Fréchet margins, we will denote it by $Z(\boldsymbol{s})$. Given any finite set of locations $s_1, \ldots, s_d$, one can let $\boldsymbol{X}_i = (X_i(s_1), \ldots, X_i(s_d))^{\mathrm{T}}$ or $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_d))^{\mathrm{T}}$ and return to the finite-dimensional setting.

An alternative approach to analyzing block maxima is to instead select and analyze a subset of threshold exceedances. Although there has been work to develop methods for multivariate threshold exceedance data, most spatial extremes work to date has aimed at developing max-stable models and fitting block-maximum data. This survey will primarily focus on such work, but we briefly discuss ongoing work in developing methods for spatial threshold exceedance data in the discussion in §5.

1.2. Standard geostatistics

Let $X(\boldsymbol{s})$ be a stochastic process. The field of geostatistics provides a framework for exploring, modeling, and predicting or interpolating $X(\boldsymbol{s})$. Much of clas-

sical geostatistics tries to characterize $X(\boldsymbol{s})$ in terms of its mean and covariance function. One typically thinks of using the mean to represent large scale changes of $X(\boldsymbol{s})$ and the covariance function to capture the variability due to small- and micro-scale stochastic sources (Schabenberger & Gotway, 2005, p. 132).

A basic model can be formulated as

$$X(\boldsymbol{s}) = \alpha(\boldsymbol{s}) + e(\boldsymbol{s}) \ , \tag{1.5}$$

where $\alpha(\boldsymbol{s})$ is the (non-random) mean function and $e(\boldsymbol{s})$ is a zero-mean stochastic process. Often, a regression relation is assumed for the mean function: $\alpha(\boldsymbol{s}) = W(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta}$, where $W(\boldsymbol{s})$ is a vector of covariate information at location $\boldsymbol{s}$ and $\boldsymbol{\beta}$ is a vector of regression coefficients.

The process e tries to account for any behavior not captured by the mean function α. A simple geostatistical model may assume $e(\boldsymbol{s})$ is second-order stationary and isotropic. Stationarity implies that the covariance does not depend on location, i.e., $\mathrm{Cov}(e(\boldsymbol{s}), e(\boldsymbol{s}')) = \mathrm{Cov}(e(\boldsymbol{s}+\boldsymbol{h}), e(\boldsymbol{s}'+\boldsymbol{h}))$, while isotropy implies that covariance is a function of distance only, i.e., $\mathrm{Cov}(e(\boldsymbol{s}), e(\boldsymbol{s}+\boldsymbol{h})) = C(h)$, where $h = \|\boldsymbol{h}\|$. At times it is useful to further assume that $e(\boldsymbol{s})$ is a Gaussian process, which in turn implies $X(\boldsymbol{s})$ is a Gaussian process with mean $\alpha(\boldsymbol{s})$.

The second-order stationary and isotropic random field e is characterized by its covariance function $C(h)$ or equivalently its semivariogram

$$\gamma(h) \ = \ \frac{1}{2}\,\mathrm{Var}\Big[e(\boldsymbol{s}+\boldsymbol{h}) - e(\boldsymbol{s})\Big] \ = \ \frac{1}{2}\,\mathrm{E}\Big[\big\{e(\boldsymbol{s}+\boldsymbol{h}) - e(\boldsymbol{s})\big\}^2\Big] \ .$$

Since the covariance function or semivariogram must satisfy several requirements to be valid, models for $C(h)$ or $\gamma(h)$ are generally selected from parametric families (see Schabenberger & Gotway, 2005, § 4.2) known to meet these requirements.

Less restrictive than second-order stationary is intrinsic stationarity, which implies that a process has stationary increments, i.e., $e(\boldsymbol{s}) - e(\boldsymbol{s}+\boldsymbol{h}) \stackrel{d}{=} e(\boldsymbol{h}) - e(\boldsymbol{0})$. An intrinsically stationary process can be viewed as akin to a random walk in time series, which is stationary after differencing. An intrinsically stationary process does not have a covariance function $C(\boldsymbol{h})$ but does have a semivariogram $\gamma(\boldsymbol{h})$.

A geostatistical analysis often begins with an exploratory phase where dependence is investigated via an empirical covariogram or semivariogram. As best as one can, one must first account for large scale effects in the mean as "much damage can be done by applying semivariogram estimators … to data from non-stationary spatial processes" (Schabenberger & Gotway, 2005, p. 135). Assuming stationarity and isotropy, the traditional sample semivariogram is

$$\hat{\gamma}(h) \ = \ \frac{1}{2\,|\mathcal{N}(h)|} \sum_{(\boldsymbol{s},\boldsymbol{s}')\in\mathcal{N}(h)} \big\{e(\boldsymbol{s}) - e(\boldsymbol{s}')\big\}^2 \ , \tag{1.6}$$

where $\mathcal{N}(h)$ denotes the number of pairs $(\boldsymbol{s}, \boldsymbol{s}')$ separated by the distance h. Applying the empirical semivariogram to observations provides insight for semivariogram model selection. Having selected a parametric family $\gamma_{\boldsymbol{\phi}}(h)$ for the semivariogram function, one often proceeds to estimate the model parameters $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$.

A primary goal of many geostatistics analyses is spatial prediction / interpolation employing an estimated semivariogram, which is known as *kriging*. The point predictor from kriging corresponds to the best linear unbiased predictor (or the conditional expectation under a Gaussian assumption) of the value of $X(\boldsymbol{s}_0)$ at unobserved location $\boldsymbol{s}_0$ given observed values $X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)$. Prediction uncertainty is typically quantified in terms of mean-square prediction error.

There is an analogy between the two tasks described in §1.1 and the geostatistics model (1.5). If e is stationary and Gaussian, then the marginal distribution can only vary with $\alpha(\boldsymbol{s})$ which captures the regional spatial effects. After accounting for regional effects with α, the residual dependence in e is characterized by its semivariogram or covariance function.

There are important fundamental differences between geostatistics and spatial extremes. As it is based on first and second moments, geostatistics focuses on central tendencies, not on the distribution's tail. The Gaussian framework which is never far from a traditional geostatistics analysis is incorrect for data that are maxima, as the Gaussian distribution is not max-stable. Dependence in extremes is described via the exponent measure function $V(\boldsymbol{z})$ or angular measure $H(\boldsymbol{w})$ which cannot be linked to covariance. Finally, much of classical geostatistics is applied to situations where one has only one realization of the process $X(\boldsymbol{s})$, observed at multiple locations. To perform an extreme value analysis, it is necessary that multiple realizations $X_i(\boldsymbol{s})$ underlie the subset of extreme data which are eventually analyzed.

Much of the paper will study max-stable processes, and we need to extend the notion of stationarity and isotropy to these processes. Stationarity for max-stable processes is first-order, implying invariance of any finite-dimensional joint distribution to translation:

$$\Pr\Big(Y(\boldsymbol{s}_1) \le y_1, \ldots, Y(\boldsymbol{s}_d) \le y_d\Big) = \Pr\Big(Y(\boldsymbol{s}_1 + \boldsymbol{h}) \le y_1, \ldots, Y(\boldsymbol{s}_d + \boldsymbol{h}) \le y_d\Big) .$$

Isotropy will imply that all bivariate joint distributions are also invariant to rotation:

$$\Pr\Big(Y(\boldsymbol{s}) \le y_1, Y(\boldsymbol{s} + \boldsymbol{h}) \le y_2\Big) = \Pr\Big(Y(\boldsymbol{s}) \le y_1, Y(\boldsymbol{s} + \boldsymbol{h}') \le y_2\Big) ,$$

if $\|\boldsymbol{h}\| = \|\boldsymbol{h}'\|$. For simplicity, we will generally assume that a random field is stationary and isotropic.

The remainder of the paper is structured to follow the order of a possible spatial extremes analysis. We begin by reviewing tools which measure spatial dependence in §2. In §3 we survey methods for modeling the marginal tail behavior over a study region. In §4, we review two primary methods for modeling the residual dependence: max-stable process models and copula approaches. We conclude with a discussion which mentions work in development, challenges posed by applications, and open problems.

2. MEASURING SPATIAL DEPENDENCE

To completely characterize the dependence among the components of a max-stable random vector requires one to specify the angular measure $H(\boldsymbol{w})$ or the exponent measure function $V(\boldsymbol{z})$. Specification of $H(\boldsymbol{w})$ or $V(\boldsymbol{z})$ is arduous, especially as the dimension d grows, and representations are not easily compared. It is useful to have summary measures of tail dependence, and several metrics have been developed which aim to summarize the amount of tail dependence in one number.

A complication arises because tail dependence falls into two distinct categories: asymptotic dependence and asymptotic independence. Since the categories are distinct, summary measures for dependence have been developed for each category. We first focus on metrics for the asymptotic dependence case, with a particular interest in measuring dependence in terms of spatial distance. We then explain the notion of asymptotic independence (which does not imply complete independence) and briefly mention how the amount of dependence in the asymptotic independence case can be measured.

We note that the metrics all assume at least that random vectors or fields have a common marginal distribution. Like in geostatistics, one must first try to account for large-scale marginal effects before using these tools to assess (residual) dependence.

2.1. Tail dependence metrics for asymptotic dependence

There are many related metrics for quantifying tail dependence when the random vector exhibits asymptotic dependence. The list includes the metric d of Davis & Resnick (1989) and the metric χ of Coles *et al.* (1999). We focus on two metrics: the extremal coefficient (Smith, 1990; Schlather & Tawn, 2003) which is readily interpretable and the madogram (Cooley *et al.*, 2006; Naveau *et al.*, 2009) which has ties to the semivariogram.

2.1.1. Extremal coefficient

Let $\boldsymbol{Y}$ be a d-dimensional max-stable random variable with common margins. The d-dimensional extremal coefficient θ_d can be implicitly defined as

$$\Pr\Big(Y_1 \le y, \ldots, Y_d \le y\Big) \;=\; \Pr\left(\bigvee_{j=1}^{d} Y_j \le y\right) \;=\; \Pr^{\theta_d}\big(Y_1 \le y\big)\;,$$

for any y in the support of Y_1. Transforming the marginals of $\boldsymbol{Y}$ to obtain $\boldsymbol{Z}$, and due to the homogeneity property of V, we have

$$\Pr\Big(Z_1 \le z, \ldots, Z_d \le z\Big) \;=\; \exp\big\{-z^{-1}V(1,\ldots,1)\big\} \;\Rightarrow\; \theta_d = V(1,\ldots,1)\;. \tag{2.1}$$

The value θ_d can be thought of as the effective number of independent random variables in the d-dimensional random vector. The coefficient takes values between 1 and d, with a value of 1 corresponding to complete dependence among the locations, and a value of d corresponding to complete independence. The extremal coefficient is studied extensively in Schlather & Tawn (2003) and relations between extremal coefficients of different orders are given in Schlather & Tawn (2002).

Given replicates of a d-dimensional random vector $\boldsymbol{Z}$, Smith (1990) and Coles & Dixon (1999) propose an estimator of the extremal coefficient θ_d. As $\boldsymbol{Z}$ has unit Fréchet margins, $1/Z_j$ is unit exponential, and $1/\bigvee_{j=1}^{d} Z_j$ is exponential with mean $1/\theta_d$. Given i.i.d. replicates $\boldsymbol{Z}_k$, $k = 1, \ldots, m$, a simple estimator is

$$\hat{\theta}_d \;=\; \frac{m}{\sum_{k=1}^{m} 1/\bigvee_{j=1}^{d}(z_{k,j})}\;, \tag{2.2}$$

where $z_{k,j}$ is the j^{th} component of the observation $\boldsymbol{z}_k$.

Although higher-order extremal coefficients are sometimes useful (see Erhardt & Smith, 2011), θ_2 is most widely used as it conveys the amount of dependence between a pair of components. Bivariate dependence metrics are especially useful in spatial studies as one generally wants to link the level of dependence to spatial distance. Let $Z(\boldsymbol{s})$ be a stationary and isotropic max-stable random field with unit-Fréchet margins. It is possible to extend (2.1) to be a distance-based dependence metric:

$$\theta(h) \;=\; \Pr\Big(Z(\boldsymbol{s}) \le z,\, Z(\boldsymbol{s}+\boldsymbol{h}) \le z\Big)\;.$$

One could extend (2.2) to construct a distance-based estimator for $\theta(h)$; however, to our knowledge, distance-based dependence summary measures have been primarily estimated via the madogram (for instance, see the `SpatialExtremes` package in `R`, Ribatet, 2011).

2.1.2. Madogram

The madogram, a first-order semivariogram, has its roots in geostatistics and its properties were studied by Matheron (1987). Since the madogram requires the first-moment to be finite which is not always the case in extremes studies, Cooley *et al.* (2006) proposed the F-madogram, which first transforms the random variable by applying the cdf and is finite for any distribution. If $Y(\boldsymbol{s})$ is a stationary and isotropic max-stable process with marginal distribution G, then the F-madogram is:

$$\nu(h) = \frac{1}{2}\,\mathbf{E}\Big|G\big\{Y(\boldsymbol{s})\big\} - G\big\{Y(\boldsymbol{s}+\boldsymbol{h})\big\}\Big| \tag{2.3}$$

(for consistency with above, we have used G, rather than F, to denote a max-stable marginal distribution). The F-madogram's values range from 0 to 1/6, which corresponds to complete dependence and independence respectively. The F-madogram is related to the other extremal dependence metrics, as Cooley *et al.* (2006) show that

$$\theta(h) = \frac{1+2\,\nu(h)}{1-2\,\nu(h)}\ . \tag{2.4}$$

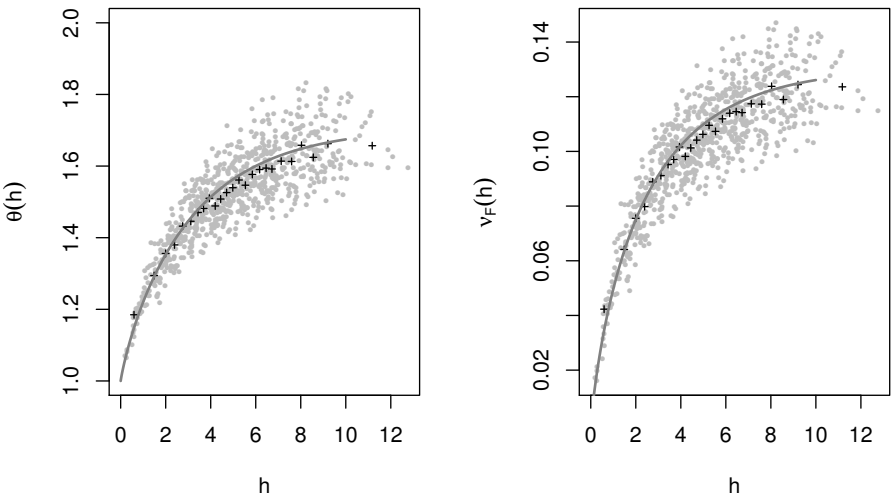


Figure 1: Extremal coefficient (left panel) and the F-Madogram (right panel) with unit Gumbel margins for the Schlather model with Whittle–Matérn correlation functions. The red lines are the theoretical extremal coefficient and F-madogram, gray points are pairwise estimates, and black crosses are binned estimates.

An advantage of the F-madogram is that its definition (2.3) suggests an estimator. Let $y_k(\boldsymbol{s})$, $k = 1, \ldots, m$, be i.i.d. replicates of a max-stable process with marginal distribution G which is observed at a finite set of locations. Then the sample F-madogram is given by

$$\hat{\nu}(h) \;=\; \frac{1}{m} \sum_{k=1}^{m} \frac{1}{2\,|\mathcal{N}(h)|} \sum_{(\boldsymbol{s},\boldsymbol{s}')\,\in\,\mathcal{N}(h)} \left| G\big\{y_k(\boldsymbol{s})\big\} - G\big\{y_k(\boldsymbol{s}')\big\} \right| , \tag{2.5}$$

analogous to the semivariogram estimator (1.6). An estimator $\hat{\theta}(h)$ can be obtained via a plug-in estimator from (2.4). Equation (2.5) assumes that the marginal distribution G is known. Naveau *et al.* (2009) discuss estimation of the madogram when the marginal distribution is not known and further define the λ-madogram, which is related to the Pickands' dependence function and which extends the notion of the madogram to completely describe the bivariate dependence structure.

2.2. Asymptotic independence

Two components X_1 and X_2 from a random vector $\boldsymbol{X}$ with common marginals are *asymptotically independent* if

$$\lim_{x \to x^*} \Pr\Big(X_2 > x \mid X_1 > x\Big) = 0 \;, \tag{2.6}$$

where x^* is the upper endpoint of the common marginal distribution. The components are *asymptotically dependent* if the limit in (2.6) is a non-zero constant. If asymptotically independent, then (X_1, X_2) lie in the domain of attraction of max-stable (Y_1, Y_2), whose angular measure $H(\boldsymbol{w})$ has a mass of $1/2$ at $\boldsymbol{w} = (0, 1)$ and at $\boldsymbol{w} = (1, 0)$, and 0 elsewhere. That is, the bivariate angular measure has mass only on the axes. The extremal coefficient of (Y_1, Y_2) is 2, correspondingly.

Asymptotic independence does not imply (complete) independence and, in fact, can mask relevant dependencies. The canonical example is that if (X_1, X_2) is standard bivariate normal with fixed correlation $\rho < 1$, then it can be shown (Sibuya, 1960) that the variables are asymptotically independent. Summarizing tail dependence with the extremal coefficient or madogram, or modeling tail dependence via $H(\boldsymbol{w})$ or $V(\boldsymbol{z})$, will ignore any dependence in an asymptotically independent couple.

Ledford and Tawn (1996) identified the problems arising with the existing modeling methodology in the case of asymptotic independence and proposed a new parameter η, the coefficient of tail dependence. If (Z_1, Z_2) has unit Fréchet

marginals, Ledford and Tawn (1996) assume a joint survival function $\bar{F}$ satisfying

$$\bar{F}(z,z) \;=\; \Pr\Big(Z_1 > z,\, Z_2 > z\Big) \;\sim\; \mathcal{L}(z)\, z^{-1/\eta}\;, \qquad z \to \infty\;,$$

where $\mathcal{L}$ is a slowly varying function ($\mathcal{L}(tz)/\mathcal{L}(z) \to 1$ as $z \to \infty$) and $\eta \in (0,1]$.[1] The coefficient of tail dependence η is used to quantify the tail dependence in the asymptotic independent setting; $\eta = 1$ implies asymptotic dependence and $\eta < 1$ measures the degree of dependence under asymptotic independence. Ledford and Tawn (1996) spawned much subsequent work including other estimators for η (Draisma *et al.*, 2004; Peng, 1999), development of other dependence measures in the asymptotic independence setting (Coles *et al.*, 1999), and development of models in the case of asymptotic independence (Ledford & Tawn, 1997; Heffernan & Tawn, 2004; Ramos & Ledford, 2009).

Models for and measures of tail dependence for spatial extremes have thus far been limited to the asymptotically dependent case. However, an understanding of the concept of asymptotic independence is essential to fully understanding the limitations of summary dependence metrics such as $\theta(h)$ and for understanding the models for residual dependence presented in §4.

3. CAPTURING SPATIAL STRUCTURE IN MARGINAL BEHAVIOR

The representation of multivariate max-stable distributions (1.2) as well as the spatial dependence models in §§4.1, 4.2 presuppose that the univariate marginal distributions are known and common at all locations in the study region. In most applications, the study regions are large enough that the assumption of a common marginal distribution is unrealistic and the marginal distribution is not known. Therefore, it becomes essential to model the marginal distribution at all locations within the study region.

A simplistic approach would be to individually estimate the marginal distributions at each location. This approach has been used in (non-spatial) multivariate applications (e.g., Heffernan & Tawn, 2004; Cooley *et al.*, 2010). Such an approach is less-than-ideal for spatial applications for a number of reasons. First, a goal of many spatial projects is to make inference at locations where there are no data, i.e., to perform spatial interpolation. Constructing unconnected models at each location does not allow one to readily interpolate. Second, there is a desire to borrow strength across locations when estimating marginal parameters. Many spatial data have a temporal record of several decades. Such a data record

[1] We use F to denote the cdf as Ledford and Tawn (1996) do not require that (Z_1, Z_2) be max-stable.

is more-than-enough to pin down the central tendencies of the marginal distribution, but large uncertainties remain about tail quantities. It is well-documented that tail quantities, and in particular point estimates for the tail index ξ, can vary wildly over the spatial region when estimated individually (e.g., Cooley & Sain (2010)). Methods which borrow strength across locations 'trade space for time' and help to reduce uncertainties.

Below we briefly detail several methods used to capture spatial structure when describing the tail of the marginal distribution before focusing on the recent approach of hierarchical modeling.

3.1. Methods for marginal parameter estimation

As mentioned in §1.2, regression approaches are frequently used to model the mean process $\alpha(\boldsymbol{s})$ of a geostatistical model. Similarly, regression approaches have been used to model the parameters of the extreme value distributions. When modeling annual maximum precipitation in the southeast United States, Padoan *et al.* (2010) select a model in which the location parameter $\mu(\boldsymbol{s})$ and scale parameter $\sigma(\boldsymbol{s})$ are linear functions of latitude and elevation, and the shape parameter $\xi(\boldsymbol{s})$ is constant over the study region. Padoan *et al.* (2010) go on to model the residual dependence via a max-stable process (see §4.1). Mannshardt-Shamseldin *et al.* (2010) use a spatial regression approach which downscales extreme precipitation as generated by climate models to that observed at weather stations. Generally, one wishes to employ covariates which are known at all locations $\boldsymbol{s} \in \mathcal{S}$ (both observed and unobserved), and often spatial coordinates are the only such covariates. In geostatistics, regression models on spatial coordinates are known as trend surface models (Schabenberger & Gotway, 2005, §5.3.1). However simple regression models on available covariates are sometimes unable to fully capture complex spatial behavior, and Ribatet *et al.* (2011) found this to be the case in a study of extreme precipitation in Switzerland. Regional frequency analysis (RFA) (Hosking & Wallis, 1997) is a term for a statistical procedure which explicitly borrows strength across locations and which in turn characterizes the spatial nature of the tail of the marginal distribution. RFA has a long history in hydrology and traces its roots to the index flood procedure of Dalrymple (1960). RFA is a multi-step procedure which pools data over predefined subregions of $\mathcal{S}$ determined to be 'homogeneous'. Hosking & Wallis (1997) advocate an estimation method based on L-moments for RFA. A recent effort to update the precipitation-frequency atlases for the United States (Bonnin *et al.*, 2004a,b) employs an L-moment-based RFA coupled with an interpolation method based on the PRISM method (Daly *et al.*, 1994, 2002). A possible disadvantage of RFA is that it does not construct an explicit spatial model for the marginal parameters. To our knowledge, no RFA-based work has tried to account for residual dependence in the data after accounting for marginal effects.

3.2. Hierarchical modeling

Hierarchical (or multi-level) models have been extensively used in describing the relationship between observations and the complex processes that generate them. For many hierarchical models, the data collected is not well-suited for modeling within the usual Gaussian framework of geostatistics. An early example of spatial hierarchical modeling is Diggle *et al.* (1998), who model two sets of spatial data: the first data set is assumed to be Poisson distributed, the second set is assumed to be binomial. To explain the spatial variation of the data, hierarchical models typically assume that the behavior of the data over the study region is driven by an unobserved or latent process. For example, Diggle *et al.* (1998) modeled both the Poisson intensity and a risk level associated with the binomial with a Gaussian process which varied over the respective study regions. Treating the parameters of the observations' marginal distribution as spatial random effects is especially useful when the spatial behavior is too complex to capture with fixed effects. Much of the early work in spatial hierarchical modeling was done in the field of epidemiology, and the book of Banerjee *et al.* (2004) has several examples.

Hierarchical models are often devised within a Bayesian framework, and typically have three levels: (i) the data level, (ii) the process level, and (iii) the prior level. To describe the set-up of the Bayesian hierarchical model (BHM), let the vector of parameters, $\boldsymbol{\psi}$, be defined as $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$, where $\boldsymbol{\psi}_1$ are the parameters in the data level, and $\boldsymbol{\psi}_2$ are the parameters in the process level. Then, the posterior distribution of $\boldsymbol{\psi}$ given the data $\mathbf{y}$, $\pi(\boldsymbol{\psi} \mid \mathbf{y})$, is given by

$$\pi(\boldsymbol{\psi} \mid \mathbf{y}) \propto \pi(\mathbf{y} \mid \boldsymbol{\psi}_1)\, \pi(\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2)\, \pi(\boldsymbol{\psi}_2)\ . \tag{3.1}$$

Here $\pi(\mathbf{y} \mid \boldsymbol{\psi}_1)$ defines the likelihood function, $\pi(\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2)$ describes the distribution of the process and $\pi(\boldsymbol{\psi}_2)$ the (hyper) priors. When applied to extremes data, the likelihood is based on an extreme-value distribution (EVD). Spatial modeling at the process level is designed to borrow strength across locations and flexibly capture spatial variation, showing how the marginal parameters of the EVD vary over the study region.

In a spatial hierarchical model, the likelihood must account for the fact that the data are observed at multiple locations. A simple approach is to assume that the data at different locations are conditionally independent given the parameters $\boldsymbol{\psi}_1$, which themselves are spatially dependent from the process level of the model. With this assumption, the likelihood becomes a product of the individual likelihoods at each location. This conditional independence assumption is widely made in hierarchical modeling, and is quite sensible in most epidemiological settings where disease counts at different locations can be assumed to be independent once the latent risk level is accounted for. However, the conditional independence assumption is questionable when modeling weather data because

individual events can affect more than one location. A hierarchical model which assumes conditional independence in the likelihood ignores any residual dependence which remains after accounting for marginal effects.

Despite the aforementioned limitation of assuming conditional independence, there are several applications of BHMs in the spatial extremes literature whose primary aim is to characterize the marginal tail behavior and which make this assumption. Among the earliest work constructing BHM's for extremes is that of Casson & Coles (1999) who use a GEV-based model and study hurricane wind speed data. Fawcett & Walshaw (2006) apply a BHM to extreme wind speed data. Cooley *et al.* (2007) use a hierarchical approach to model extreme precipitation data from weather stations and then interpolate the marginal distribution over the study region to produce return level maps. Sang & Gelfand (2009) modeled annual maximum rainfall over a region of South Africa, using a spatial autoregressive model in the process level of their hierarchy, rather than a geostatistical model. Both Cooley & Sain (2010) and Schliep *et al.* (2010) use a BHM to characterize extreme precipitation as generated by climate models over spatial regions with thousands of locations. Mendes *et al.* (2010) apply BHMs to spatial extremes of large wildfire sizes.

Recent hierarchical work in extremes has sought to move beyond the conditional independence assumption and capture both regional spatial effects via a process-level spatial model on marginal parameters and local spatial effects via a dependence structure within the likelihood at the data level. Sang & Gelfand (2010) and Fuentes *et al.* (2011) respectively use a Gaussian copula model and a Dirichlet process model to try and capture residual dependence within the structure of a hierarchical model. Ribatet *et al.* (2011) use max-stable process models to formulate a likelihood designed to capture residual dependence. We discuss these approaches in depth in § 4.

Inference for BHMs is done by sampling, which is often complicated by the fact that the full conditional distributions for the parameters often do not exist in closed form. Markov Chain Monte Carlo (MCMC) methods can be used to approximate the posterior distributions. One of the most popular approaches is the Gibbs sampling method (Gelfand & Smith, 1990), which is accommodated by the conditional relationships in (3.1). As conjugate priors for EVD's are not known, BHMs for extremes require a Metropolis–Hastings step to be included within each iteration of the Gibbs sampler.

Both hierarchical approaches which assume conditional independence and other methods such as RFA ignore residual dependence which arises due to the spatial extent of individual events. In instances when interest is primarily on the marginal behavior, it may be appropriate to ignore the residual dependence. There is a long history in hydrology of producing return level maps; that is, a map which depicts a high quantile (e.g., the quantile associated with a 100-year return

level) at any site within a study region. The aforementioned projects by Bonnin *et al.* (2004a,b) produced such maps and NOAA's Hydrometeorological Design Studies Center maintains a website[2] which provides point-located return level estimates. Likewise, projects such as Cooley & Sain (2010), Sang & Gelfand (2009) and Cooley *et al.* (2007) aimed only to estimate pointwise return levels. It is imperative that it is recognized that such projects cannot be used to quantify the aggregated effects of a large event which occurs across multiple locations, nor can they be used to produce realistic simulated data (Davison *et al.*, 2012, Figure 4).

4. MODELS FOR SPATIAL DEPENDENCE

In this section, we consider two approaches for capturing the residual dependence assuming that the marginal effects have been accounted for. The first approach is to use a max-stable process model (§1.1), which will assume that the marginals have been transformed to be unit Fréchet. The second is a copula approach, a popular recent approach for modeling multivariate data which assumes the marginals are Uniform [0,1].

4.1. Max-stable processes

The definition of a max-stable process, Equation (1.4), as the infinite dimensional generalization of a multivariate max-stable distribution gives a well-defined class of models, but it does not suggest how to construct such processes. A conceptual construction of max-stable processes was first given with a spectral representation of extremal processes by de Haan (1984) and de Haan & Ferreira (2006). A general representation of max-stable processes can be described by two components, a stochastic process $\{X(\boldsymbol{s})\}$ and a Poisson process Π with intensity $\mathrm{d}\zeta/\zeta^2$ on $(0,\infty)$. Let $\{X_i(\boldsymbol{s})\}_{i\in\mathbb{N}}$ be independent realizations of a process $X(\boldsymbol{s})$ with $\boldsymbol{E}[X(\boldsymbol{s})] = 1$, and let $\zeta_i \in \Pi$ be points of the Poisson process. Then

$$Z(\boldsymbol{s}) = \max_{i\geq 1} \zeta_i \, X_i(\boldsymbol{s}) \ , \qquad \boldsymbol{s} \in \mathcal{S} \ ,$$

is a max-stable process with unit-Fréchet margins and the distribution function is determined by

$$\Pr\Big(Z(\boldsymbol{s}) \leq z(\boldsymbol{s}),\, \boldsymbol{s} \in \mathcal{S}\Big) \;=\; \exp\left(-E\left[\sup_{\boldsymbol{s}\in\mathcal{S}} \left\{\frac{X(\boldsymbol{s})}{z(\boldsymbol{s})}\right\}\right]\right) ,$$

where minus the exponent is the infinite-dimensional analogue to V; see Equation (1.3). Different choices of the process $X_i(\boldsymbol{s})$ lead to different classes of max-stable processes.

[2] `http://hdsc.nws.noaa.gov/hdsc/pfds/index.htm`

Although Gaussian distributions and processes are not well-suited for modeling extremes, they can be used within the above max-stable construction to produce appropriate models. This was first proposed by Smith (1990). Let $\{(\zeta_i, \boldsymbol{s}_i);\, i \geq 1\}$ denote the points of a Poisson process on $(0,\infty) \times \mathcal{S}$ with intensity $\zeta^{-2}\, d\zeta\, d\boldsymbol{s}$. Let $\{f(x)\}$ on $\mathbb{R}^d$ denote a non-negative function such that $\int f(x)\, dx = 1$ and define

$$Z(\boldsymbol{s}) = \max_i \zeta_i\, f(\boldsymbol{s} - \boldsymbol{s}_i)\ .$$

Then $Z(\cdot)$ is a max-stable process with unit-Fréchet margins. Smith also introduced the so-called *rainfall-storms* interpretation: think of $f(\cdot)$ as the shape of a storm centered at point $\boldsymbol{s}_i$, and think of the overall magnitude of storm as ζ_i. Then the max-stable process $Z(\cdot)$ is the pointwise maximum rainfall (taken over all storms) for each location in $\mathcal{S}$. If $f(\cdot)$ is a multivariate normal density with covariance parameter Σ, then the process $Z(\cdot)$ is called a *Gaussian extreme value process* and the joint distribution at two sites is given by

$$\begin{aligned}\Pr\Big\{Z(\boldsymbol{s}_1) \leq z_1,\, Z(\boldsymbol{s}_2) \leq z_2\Big\} &= \\ &= \exp\left\{-\frac{1}{z_1}\,\Phi\!\left(\frac{a}{2} + \frac{1}{a}\,\log\frac{z_2}{z_1}\right) - \frac{1}{z_2}\,\Phi\!\left(\frac{a}{2} + \frac{1}{a}\,\log\frac{z_1}{z_2}\right)\right\},\end{aligned}$$

where $a = \sqrt{(\boldsymbol{s}_1 - \boldsymbol{s}_2)^{\mathrm{T}}\, \Sigma^{-1} (\boldsymbol{s}_1 - \boldsymbol{s}_2)}$ and Φ is the standard normal cumulative distribution function. The dependence parameter a represents a transformed distance between two sites and the limits $a \to 0$ and $a \to \infty$ correspond to perfect dependence and independence, respectively. Thus the most common Smith model takes $X_i(s)$ to be a multivariate density function. Figure 2 shows a realization.

Schlather (2002) suggested a more flexible class of max-stable processes by taking $X_i(\boldsymbol{s})$ to be any stationary Gaussian random field with finite expectation. A stationary max-stable process with unit-Fréchet margins can be obtained by

$$Z(\boldsymbol{s}) = \max_i \zeta_i\, \max\big\{0, X_i(\boldsymbol{s})\big\}$$

where $\mu = E\big\{\max(0, X_i(\boldsymbol{s}))\big\} < \infty$ and $\{\zeta_i\}$ denotes the points of a Poisson process on $(0,\infty)$ with intensity measure $\mu^{-1}\zeta^{-2}\, d\zeta$. The max-stable process also allows a useful interpretation of spatial storm events. On taking a stationary random process $X_i(\boldsymbol{s})$, the spatial rainfall events have the same dependence structure but the realizations will vary stochastically, not the deterministic form $f(\cdot)$ such as Smith's model. If the random process is specified for a stationary isotropic Gaussian random field $X_i(\cdot)$ with unit variance, correlation $\rho(\cdot)$ and $\mu^{-1} = \sqrt{2\,\pi}$, then the process $Z(\boldsymbol{s})$ is called an *extremal Gaussian process* and the bivariate marginal distributions are given by

$$\begin{aligned}\Pr\Big\{Z(\boldsymbol{s}_1) \leq z_1,\, Z(\boldsymbol{s}_2) \leq z_2\Big\} &= \\ &= \exp\left\{-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2\big(\rho(h)+1\big)\,\frac{z_1\, z_2}{(z_1+z_2)^2}}\,\right)\right\}\end{aligned}$$

where h is the Euclidean distance between station $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. The correlation is chosen from one of the valid families of correlations for Gaussian processes. Figure 2 shows a realization of an extremal Gaussian process.

One drawback to the Schlather model is that it cannot attain independence of extremes, because the extremal coefficient $\theta(h) = 1+\left[\frac{1-\rho(h)}{2}\right]^{1/2}$ takes the value in the interval $[1, 1.838]$ (and not the usual range of [1,2]) as the distance $h \to \infty$. To overcome this, the process $X_i(\boldsymbol{s})$ can be restricted to a random set $\mathcal{B}$, i.e.,

$$Z(\boldsymbol{s}) \;=\; \max_i \zeta_i \, X_i(\boldsymbol{s}) \, I_{\mathcal{B}_i}(\boldsymbol{s} - \boldsymbol{s}_i)$$

where $I_{\mathcal{B}}$ is the indicator function of a compact random set $\mathcal{B} \subset \mathcal{S}$ and $\boldsymbol{s}_i$ are the points of a Poisson process on $\mathcal{S}$. If X_i is a Gaussian process, the bivariate marginal distribution is

$$\exp\left\{-\left(\frac{1}{z_1}+\frac{1}{z_2}\right)\left[1-\frac{\alpha(h)}{2}\left(1-\sqrt{1-2\big(\rho(h)+1\big)\,\frac{z_1\,z_2}{(z_1+z_2)^2}}\,\right)\right]\right\}$$

where $\alpha(h) = E\big\{|\mathcal{B}\cap(h+\mathcal{B})|\big\}/E(|\mathcal{B}|) \in [0,1]$. The extremal coefficient takes on any value in the interval $[1, 2]$ and thus independent extremes are available. One possible choice for the set $\mathcal{B}$ is a disc of radius r, meaning $\alpha(h) \doteq \{1-|h|/(2r)\}_+$, which equals 0 when $|h| > 2\,r$. One could consider $\mathcal{B}$ as a random set, which means that the radius of the disk is random and all elliptical sets having the same major axis are permissible as a generalization for $\mathcal{B}$ (Davison & Gholamrezaee, 2012).

Kabluchko *et al.* (2009) proposed an alternative specification for the $X_i(\cdot)$ process, which requires weaker assumptions than second-order stationarity. Let $X_i(\boldsymbol{s}) = \exp\big\{e_i(\boldsymbol{s}) - \frac{1}{2}\,\sigma^2(\boldsymbol{s})\big\}$ where $e_i(\boldsymbol{s})$ is a Gaussian process with stationary increments and $\sigma^2(\boldsymbol{s}) = \mathrm{Var}\{e(\boldsymbol{s})\}$. Then the process, known as the *Brown–Resnick process*, can be a very general class of max-stable processes which allows the use of semivariogram from standard geostatistics. The bivariate cdf transformed to

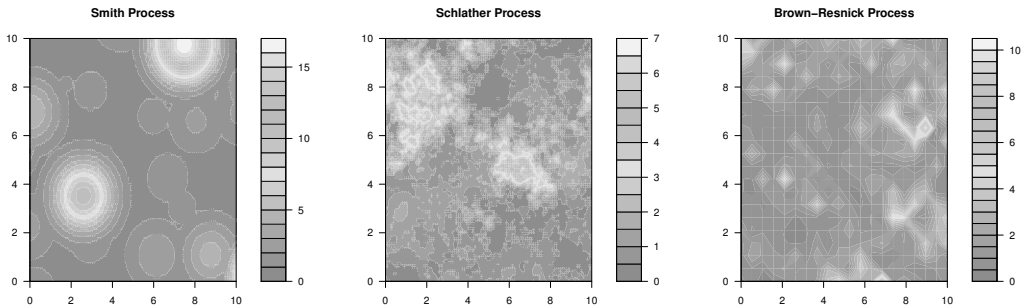


Figure 2: Realization of Gaussian extreme value process (left), extremal Gaussian process (centre), and Brown–Resnick process (right) from SpatialExtremes R package (Ribatet, 2011; R Development Core Team, 2011).

unit Fréchet margins is the same as the so-called Smith model but instead of the parameter a^2 we have $\gamma(\cdot)$, the semivariogram of $e(\cdot)$, i.e.

$$\Pr\Big\{ Z(\boldsymbol{s}_1) \le z_1, Z(\boldsymbol{s}_2) \le z_2 \Big\} = \\ = \exp\left\{ -\frac{1}{z_1}\,\Phi\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\,\log\frac{z_2}{z_1} \right) - \frac{1}{z_2}\,\Phi\left(\frac{\sqrt{\gamma(h)}}{2} + \frac{1}{\sqrt{\gamma(h)}}\,\log\frac{z_1}{z_2} \right) \right\},$$

where Φ is the standard normal distribution function and h is the Euclidean distance between location $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Realizations of the Gaussian extreme value process, the extremal Gaussian process, and the Brown–Resnick process are shown in Figure 2. The left panel shows a simulation from the Gaussian extreme value process with covariance matrix ($\sigma_{11} = \sigma_{22} = 9/8$ and $\sigma_{12} = 0$), the central panel shows shows one from the extremal Gaussian process with the Whittle–Matérn correlation function (nugget $= 0$, range $= 3$ and smooth $= 1$), and the right panel shows one from the Brown–Resnick process with parameter (range $= 3$, smooth $= 0.5$).

4.1.1. Inference for max-stable processes: composite likelihood approach

A barrier to fitting max-stable processes to data is that closed-form expressions for the joint likelihood can only be written out in low dimensional settings. The likelihood for the Smith model in $\mathbb{R}^2$ can be written out for dimension $d \le 3$ (Genton *et al.*, 2011), but the likelihood for all other max-stable processes can only be written for dimension $d \le 2$ (and we write this section for the more general case). This means that if the data are observed at $d > 2$ locations in space, the joint likelihood cannot be written in closed form. Padoan *et al.* (2010) proceeded with a likelihood-based approach to fitting max-stable processes by substituting a composite likelihood for the unavailable joint likelihood. We first introduce composite likelihoods, then show the connection to max-stable processes.

If $f(z;\psi)$ is a statistical model and we have a set of marginal or conditional events $\{\mathcal{A}_j \subseteq \mathcal{F}, j = 1, \ldots, J\}$ where $\mathcal{F}$ is a sigma algebra on $\mathcal{Z}$ and each event A_j has associated likelihood $\mathcal{L}_j(\psi; z) \propto f(z \in \mathcal{A}_j; \psi)$, then a composite log-likelihood is a weighted sum of log-likelihoods for each event (Lindsay, 1988; Varin, 2008):

$$\ell_C(\psi; z) = \sum_j w_j \cdot \log f\big(z \in \mathcal{A}_j; \psi\big) \,,$$

where w_j is a weight function on j^{th} event. If the weights are all equal they may be ignored, though non-equal weights may be used to improve the statistical performance in certain cases. One example of a composite log-likelihood is the

pairwise log-likelihood, defined (in a spatial application) as

$$\ell_C(\psi; z) = \sum_{k=1}^{m} \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} \log f\big(z_k(\boldsymbol{s}_j), z_k(\boldsymbol{s}_{j'}); \psi\big) ,$$

where each term $f\big(z_k(\boldsymbol{s}_j), z_k(\boldsymbol{s}_{j'}); \psi\big)$ is a bivariate marginal density function based on observations at locations j and j', and ψ is a spatial dependence parameter. The two inner summations sum over all unique pairs, while the outer sums over the m i.i.d. replicates. Similar to the full likelihood function, the parameter which maximizes a composite log likelihood can be found, and is termed a *maximum composite likelihood estimate*, or MCLE. When $m \to \infty$, the maximum composite likelihood estimator is consistent and asymptotically normal (Lindsay, 1988; Cox & Reid, 2004), with

$$\hat{\psi}_{\text{MCLE}} \sim \mathcal{N}(\psi, \tilde{I}^{-1}) , \qquad \tilde{I} = H(\psi)\, J^{-1}(\psi)\, H(\psi) , \tag{4.1}$$

where $H(\psi) = \mathrm{E}(-H_\psi\, \ell_C(\psi; Z))$ is the expected information matrix, $J(\psi) = V(D_\psi\, \ell_C(\psi; Z))$ is the covariance of the score, H_ψ is the Hessian matrix, D_ψ is the gradient vector, and V is the covariance matrix. When one has the full likelihood, $H(\psi) = J(\psi)$, but in the composite likelihood setting these matrices are not equal.

Padoan *et al.* (2010) used the composite likelihood to model the joint spatial dependence of extremes and accounted for regional effects with a regression model on the GEV parameters. This approach is implemented in the `R` package `SpatialExtremes` (Ribatet, 2011). The maximum composite likelihood estimator $\hat{\psi}_{\text{MCLE}}$ is found numerically. The variance is estimated using

$$\hat{H}(\hat{\psi}_{\text{MCLE}}) = -\sum_{k=1}^{m} \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} H_\psi \log f\big(z_k(\boldsymbol{s}_j), z_k(\boldsymbol{s}_{j'}); \hat{\psi}_{\text{MCLE}}\big)$$

and

$$\begin{aligned} \hat{J}(\hat{\psi}_{\text{MCLE}}) = \sum_{k=1}^{m} & \left\{ \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} D_\psi \log f\big(z_k(\boldsymbol{s}_j), z_k(\boldsymbol{s}_{j'}); \hat{\psi}_{\text{MCLE}}\big) \right\} \\ & \times \left\{ \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} D_\psi \log f\big(z_k(\boldsymbol{s}_j), z_k(\boldsymbol{s}_{j'}); \hat{\psi}_{\text{MCLE}}\big) \right\}^{\mathrm{T}} . \end{aligned}$$

The dependence parameter ψ is generic and stands in for the matrix Σ in the Smith model, the Gaussian correlation function $\rho(h; \psi)$ in the Schlather model, and the semivariogram $\gamma(h; \psi)$ in the Brown–Resnick model. For each of these models the target parameter appears in the corresponding bivariate density functions, and thus also in the pairwise log-likelihood.

Recently, there has been work which begins to explore the use of composite likelihood methods within Bayesian inference, and much of this work has been

driven by interest in spatial extremes. Both Pauli *et al.* (2011) and Ribatet *et al.* (2011) seek to employ a pairwise likelihood, rather than the unattainable true likelihood, to obtain a posterior distribution. The pairwise likelihood does not accurately represent the information in the data, as it repeatedly uses each observation when pairing it with others. Pauli *et al.* (2011) adjust the pairwise likelihood so that the first moment of the log-likelihood ratio corresponds to that of the asymptotic χ^2-distribution. Ribatet *et al.* (2011) suggest an adjustment which ensures that the curvature of the likelihood surface agrees with the asymptotic covariance matrix given in Equation (4.1). Ribatet *et al.* (2011) apply the likelihood within a spatial hierarchical model to study extreme precipitation in Switzerland.

Erhardt & Smith (2011) used approximate Bayesian computing (ABC) to obtain an approximate posterior distribution for the max-stable process dependence parameters ψ. ABC methods have been successfully implemented for problems where the joint likelihood function is intractable, but simulations are possible (Beaumont *et al.*, 2002; Sisson & Fan, 2010). Given observed data Z and prior $\pi(\psi)$, the simplest ABC algorithm is: (1) Draw $\psi' \sim \pi(\psi)$; (2) Simulate a new dataset Z' conditional on ψ'; (3) If $d(S(Z), S(Z')) \leq \epsilon$ for some summary statistic S, distance function d, and threshold ϵ, then accept ψ'; otherwise, reject. The method produces an i.i.d. sample from $\pi\big[\psi \mid d\{S(Z), S(Z')\} \leq \epsilon\big]$, which in the limit as $\epsilon \to 0$ equals $\pi(\psi \mid S(Z))$. Further, if $S(Z)$ were a sufficient statistic, this would be the exact posterior. In practice, computational costs often force concessions like in-sufficient statistics S and a non-zero threshold ϵ. Erhardt & Smith (2011) used tripletwise extremal coefficients in the construction of a summary statistic S, and then showed that the resulting ABC implementation can outperform the composite likelihood approach when estimating the spatial dependence of a max-stable process, though at an appreciably higher computational cost.

4.1.2. Spatial prediction/interpolation for max-stable processes

Kriging is a central focus of geostatistics and there has been recent work to perform spatial prediction for max-stable processes (§1.2). Wang & Stoev (2011) and Dombry *et al.* (2011) have proposed computational solutions for the prediction problem for max-stable processes.

Max-linear models are a subclass of the multivariate max-stable distributions. Let Y_i, $i = 1, \ldots, n$, be i.i.d. unit Fréchet random variables. Assume $\boldsymbol{Z} = \big(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_d)\big)^{\mathrm{T}}$ is a max-linear combination of the Y_i's; that is,

$$Z(\boldsymbol{s}_j) = \bigvee_{i=1}^{n} c_{j,i} Y_i \ , \qquad j = 1, \ldots, d \ ,$$

where $c_{j,i}$ are non-negative constants. $\boldsymbol{Z}$ is multivariate max-stable, and further if $\sum_{i=1}^{n} c_{j,i} = 1$ for all j, then $\boldsymbol{Z}$ has unit-Fréchet marginals. Any max-linear model with finite n will have a discrete angular measure; however, max-linear models form a dense subclass of multivariate max-stable random vectors as $n \to \infty$ (Zhang & Smith, 2004). Wang & Stoev (2011) propose an algorithm for efficient and exact sampling from the conditional distributions of a spectrally discrete max-stable random field. The main idea is to first generate samples from the regular conditional probability distribution of $\boldsymbol{Y} \mid \boldsymbol{Z} = \boldsymbol{z}$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $\boldsymbol{Z}$ is the vector of values at observed locations. Then, the conditional distribution of $Z(\boldsymbol{s}_0) = \bigvee_{i=1}^{n} c_{0,i} Y_i$ can be easily obtained for any given $c_{0,i}$. The performance of the algorithm was illustrated over the discretized Smith model for spatial extremes.

Dombry *et al.* (2011) also take a computational approach to spatial prediction. Specifically working with the Brown–Resnick process, Dombry *et al.* (2011) establish a link between the conditional distribution of this process and the multivariate log-normal distribution. Like Wang & Stoev (2011), the computational method of Dombry *et al.* (2011) considers different hitting-scenarios; that is, the possible combinations of individual events which could yield the observed maxima. Dombry *et al.* (2011) illustrate their method on precipitation data from Switzerland.

4.2. Copula approaches

Copulas provide another framework for representing the dependence structure of a multivariate distribution with known marginals. Copulas are multivariate distributions with standard uniform marginal distributions, and they characterize the dependence structure of a multivariate distribution from univariate marginal distributions by defining a joining mechanism (Nelsen, 2006; Joe, 1997). Given a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$ with corresponding marginal cdfs F_i for $j = 1, \ldots, d$ and joint distribution function F, a copula is a function $C\colon [0,1]^d \longrightarrow [0,1]$ such that

$$F(\boldsymbol{Y}) = C\big(F_1(Y_1), \ldots, F_d(Y_d)\big) \,. \tag{4.2}$$

If the marginal cdfs of $\boldsymbol{Y}$ are all continuous, then the copula function C is uniquely defined by (4.2). Conversely, for a copula C and continuous margins $F_1, \ldots, F_d$, the copula C corresponds to the distribution of $F_1(Y_1), \ldots, F_d(Y_d)$, i.e.,

$$C\big(u_1, \ldots, u_d\big) = F\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big) \,. \tag{4.3}$$

This follows from the multi-dimensional analog of Sklar's theorem (Sklar, 1959), which proves the existence of a copula in the bivariate case and uniqueness when the marginals are continuous (Nelsen, 2006).

The copula framework has appeal for modeling multivariate extremes, especially since the list of existing parametric subfamilies of the multivariate extreme value distribution (MEVDs) is limited. When working with multivariate block maximum data, extreme value theory suggests that each marginal should be approximately GEV distributed. Equation (4.3) suggests one can combine knowledge of the marginal distributions with a copula model to obtain a valid cdf. Further, (4.2) says that one can obtain a copula model from any multivariate distribution. While this approach allows one great flexibility to create multivariate distributions with GEV marginals, these distributions will not correspond to a MEVD as characterized in §1.1 unless one uses an extremal copula model (Joe, 1997), which essentially correspond to the existing parametric MEVD subfamilies. Use of (nonextremal) copula models to describe extremes has been controversial, and Mikosch (2006) and the associated discussion details much of the argument.

For spatial data which are typically observed at many locations, one would need a copula which can handle high dimensions, and further, whose pairwise dependence can be linked to distance. Two obvious choices are to use the multivariate Gaussian or multivariate Student t distributions to generate a copula. Given a d-dimensional Gaussian random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathrm{T}}$, the Gaussian copula is defined as

$$C_\Sigma\big(u_1, \ldots, u_d\big) \;=\; \Phi_\Sigma\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big)\;, \tag{4.4}$$

where $u_i \in [0,1]$ for $i = 1, \ldots, d$, Φ is the cdf for a standard normal distribution, and Φ_Σ is the joint cdf of the multivariate normal distribution with covariance matrix Σ. The multivariate Student t distribution copula is defined similarly, except the marginal and joint distributions displayed in (4.4) are replaced by marginal and joint (with scale matrix Σ) Student t distributions , i.e., $C_\Sigma(u_1, \ldots, u_d) = T_{\nu;\Sigma}\big(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d)\big)$, where $T_{\nu;\Sigma}$ and T_ν are the joint and marginal t distributions with ν degrees of freedom respectively.

The simplicity of the Gaussian and Student t copulas make them appealing, but they have some undesirable properties from the extreme value theory viewpoint. Because neither copula is extremal, the resulting cdf will not be max-stable. Additionally, the Gaussian copula is asymptotically independent, so it may underestimate joint tail probabilities if the data are actually asymptotically dependent.

Hüsler & Reiss (1989) devised an alternative approach which used the bivariate Gaussian distribution to create an extremal copula (equivalently a MEVD), and this general approach can be used with either the multivariate Gaussian distribution or multivariate t distribution as described by Davison *et al.* (2012). The approach is essentially the same as that which leads to the Smith (1990) max-stable process model, albeit in the multivariate rather than process setting. Thus, the drawback of the extremal Gauss or extremal t copulas is that the full multivariate distribution cannot be written in closed form, and composite like-

lihood methods must be employed. Davison *et al.* (2012) fit the Gaussian and t copulas as well as the extremal Gauss and t copulas to data which are annual maxima and conclude that the extremal copulas have improved ability to capture the dependence found in this data. Padoan (2011) discusses some further copula models potentially useful in spatial extremes.

In a desire to move away from the conditional independence assumption often made in hierarchical modeling, Sang & Gelfand (2010) use a Gaussian copula to create a likelihood function in a model structured as (3.1). Suppose that $Y(\boldsymbol{s})$ is an extremal spatial process at location $\boldsymbol{s}$, e.g., $Y(\boldsymbol{s})$ is the annual maximum of daily rainfall measurements at location $\boldsymbol{s}$. Then $Y(\boldsymbol{s}) \sim \text{GEV}\big(\mu(\boldsymbol{s}), \sigma(\boldsymbol{s}), \xi(\boldsymbol{s})\big)$. Furthermore, $Y(\boldsymbol{s})$ can be represented as

$$Y(\boldsymbol{s}) \;=\; \mu(\boldsymbol{s}) + \frac{\sigma(\boldsymbol{s})}{\xi(\boldsymbol{s})}\left(Z(\boldsymbol{s})^{\xi(\boldsymbol{s})} - 1\right) \tag{4.5}$$

where $Z(\boldsymbol{s})$ is unit-Fréchet. Focusing on the underlying unit-Fréchet process defined at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, Sang & Gelfand (2010) induced a dependence structure on the $Z(\boldsymbol{s}_i)$ for $i = 1, \ldots, d$ using the Gaussian copula of (4.4). Suppose that $\mathbf{X}(\boldsymbol{s}) = \big(X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)\big)$ is a centered spatial Gaussian process with a correlation structure defined by the function $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \psi)$ for $i, j = 1, \ldots, d$. The unit-Fréchet process $Z(\boldsymbol{s})$ is defined in terms of a transformed spatial Gaussian process $\mathbf{X}(\boldsymbol{s})$ by

$$\mathbf{Z} = \Big(G^{*-1}\Phi\big(X(\boldsymbol{s}_1)\big),\, \ldots,\, G^{*-1}\big(\Phi(X(\boldsymbol{s}_d))\big)\Big)\,,$$

where G^* is the unit-Fréchet distribution function. Given the spatial Gaussian process $X(\boldsymbol{s})$, the corresponding copula is defined as $C_X(u_1, \ldots, u_d) = F_{X,\Sigma_\rho}\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big)$ where $u_1, \ldots, u_d \in [0, 1]$, F_{X,Σ_ρ} is the multivariate Gaussian distribution function with a covariance matrix Σ_ρ defined by some correlation function $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\psi})$, $i, j = 1, \ldots, d$. Then the multivariate joint distribution of $\mathbf{Z}$ is

$$F_Z(z_1, \ldots, z_d) \;=\; C_X\big(G^*(z_1), \ldots, G^*(z_d)\big) \;=\; F_X\big(\Phi^{-1}G^*(z_1), \ldots, \Phi^{-1}G^*(z_d)\big)\,,$$

where $z_j \in \mathbb{R}$, $j = 1, \ldots, d$. Sang & Gelfand (2010) considered this the first stage of a hierarchical model, where the second stage deals with the characterizations of $\mu(\boldsymbol{s})$, $\sigma(\boldsymbol{s})$, and $\xi(\boldsymbol{s})$ found in equation (4.5). While one could be critical of the Gaussian copula that Sang & Gelfand (2010) employ, the approach does attempt to account for the residual spatial dependence which remains in the data after accounting for marginal effects, while a conditional independence assumption ignores this dependence entirely.

An alternative copula approach for spatial extremes data is proposed by Fuentes *et al.* (2011), who use a Dirichlet process (DP) copula model which is a Bayesian, nonparametric generalization of the spatial Gaussian copula model.

The DP copula model is flexible and capable of dealing with high dimensional data, but when combined with GEV marginal distributions does not yield a MEVD. The DP prior is a special case of the stick-breaking prior (Sethuraman, 1994). If a distribution function is such that

$$F \overset{d}{=} \sum_{i=1}^{m} p_i \, \delta_{\psi_i} \ , \tag{4.6}$$

where δ_x denotes a Dirac measure at x, $p_1 = V_1$, $p_i = V_i \prod_{j<i}(1 - V_j)$ with $V_i \overset{\text{iid}}{\sim} \text{Beta}(1, \nu)$, $\nu > 0$ and $\psi_1, \ldots, \psi_m \overset{\text{iid}}{\sim} H_0$, which is a centering distribution. In the spatial setting, the spatial DP copula introduces a latent process $\boldsymbol{X}$ such that the joint density of $\boldsymbol{X} = \big(X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)\big)$ is almost surely of a countable mixture of normals,

$$f\big(\boldsymbol{X} \,|\, H^d, \tau^2\big) \ = \ \sum_{i=1}^{\infty} p_i \, \phi_d\big(\boldsymbol{X} \,|\, \boldsymbol{\psi}_i, \tau^2 \boldsymbol{I}_d\big) \ , \tag{4.7}$$

where $\phi_d(\,\cdot \,|\, \boldsymbol{\mu}, \Sigma)$ denotes the d-dimensional multivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix Σ, p_i has the same distribution as in (4.6), and $\boldsymbol{\psi}_i = \big(\psi_i(\boldsymbol{s}_1), \ldots, \psi_i(\boldsymbol{s}_d)\big)$ with $\boldsymbol{\psi}_i \,|\, H^d \overset{\text{iid}}{\sim} H^d$ and $H^d \overset{d}{=} DP(\nu H_0^d)$, where $H_0^d \overset{d}{=} \phi_d(\,\cdot\,|\mathbf{0}, \Sigma)$. Let $F_{\boldsymbol{X}}$ be the cdf associated with the density in (4.7). Then the multivariate joint distribution of $\boldsymbol{Y}$ given by the copula

$$F(\boldsymbol{Y}) \ = \ C_{\boldsymbol{X}}\big(G^*(y_1), \ldots, G^*(y_d)\big) \ = \ F_{\boldsymbol{X}}\big(H_{\boldsymbol{s}_1}^{-1} G^*(y_1), \ldots, H_{\boldsymbol{s}_d}^{-1} G^*(y_d)\big) \ ,$$

where $H_{\boldsymbol{s}}$ is the cdf corresponding to the density in Equation (4.7) for $Z(\boldsymbol{s})$. As with the spatial Gaussian copula in Sang & Gelfand (2010), the unit-Fréchet spatial process is the transformed process defined as $Y(\boldsymbol{s}) = G^{*-1} H_{\boldsymbol{s}}(Z(\boldsymbol{s}))$ but with space-dependent parameters. The joint distribution function given by a spatial Gaussian copula can be expressed explicitly for any given set of locations, whereas the DP copula represents the joint distribution implicitly in a Bayesian framework.

5. DISCUSSION

The study of spatial extremes is young, as evidenced by the fact that the majority of the work referenced in this survey article has been done in the last 10 years, and nearly all of it in the last 20 years. The study has a well-developed foundation in probability theory, and the field has made great strides in developing realistic models based on this theory.

We anticipate further development of methods and models for spatial extremes, and we imagine this work will follow similar themes to recent work

in geostatistics. It remains a challenge to fit max-stable process models to data recorded at many locations, for example climate model output which has thousands of locations. Interest in large data sets has spawned geostatistics work in Gaussian Markov random fields (Rue & Held, 2005; Lindgren *et al.*, 2011), fixed-rank kriging (Cressie & Johannesson, 2008), and predictive processes (Banerjee *et al.*, 2008). Another interest in geostatistics has been space-time modeling (Cressie & Wikle, 2011), an area which spatial extremes modeling is only beginning to address (Huser & Davison, 2012). Likewise, models with nonstationary dependence structure (e.g., Sampson & Guttorp, 1992), models which can handle spatial misalignment (e.g., Berrocal *et al.*, 2010), and models for multivariate spatial data (e.g., Wackernagel, 2003) have been of interest in spatial statistics for some time, but these topics are thus far unexplored for spatial extremes.

This article has largely focused on modeling data which are block maxima, as the spatial extremes literature to date has concentrated on such data. There is work in progress (Jeon, 2012) to extend max-stable process models to appropriately model threshold exceedance data. Spatial threshold exceedance modeling would address the first issue that when restricting one's attention to block maximum data, one is likely to discard other extreme data which could be useful in describing the spatial extent of extreme events. Although from the classical extreme value theory point-of-view it is natural to consider vectors of componentwise maxima (see §1.1), practitioners can view these vectors as artificial since they likely do not appear in the data record.

Finally, we would be remiss in this survey if we did not mention the available software for analyzing spatial extremes data. The `R` package `SpatialExtremes` (Ribatet, 2011) can be used both to estimate spatial dependence and to fit max-stable process models. The `RandomFields` package (Schlather, 2011) is useful for simulating max-stable processes. The conditional simulation method of Wang & Stoev (2011) can be found in the package `maxLinear` (Wang, 2010).

ACKNOWLEDGMENTS

This work was supported by the 2011–12 Program on Uncertainty Quantification at the Statistical and Applied Mathematical Sciences Institute. Additionally, D. Cooley was partially supported by National Science Foundation grant DMS-0905315.

REFERENCES

Banerjee, S.; Carlin, B. & Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman, Boca Raton.

Banerjee, S.; Gelfand, A.; Finley, A. & Sang, H. (2008). Gaussian predictive process models for large spatial data sets, *J. R. Statist. Soc. B*, **70**, 825–848.

Beaumont, M.; Zhang, W. & Balding, D. (2002). Approximate Bayesian computing in population genetics, *Genetics*, **162**, 2025–2035.

Beirlant, J.; Goegebeur, Y.; Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, New York.

Berrocal, V.; Gelfand, A. & Holland, D. (2010). A bivariate space-time downscaler under space and time misalignment, *Ann. Appl. Statist.*, **4**, 1942–1975.

Bonnin, G.M.; Todd, D.; Lin, B.; Parzybok, T.; Yekta, M. & Riley, D. (2004a). *NOAA Atlas 14, Precip. Freq. Atlas United States*, Vol. 1, U.S. Dep. Comm., Nat. Oceanic Atmos. Administ, Nat. Weather Serv., Silver Spring, Maryland.

Bonnin, G.M.; Todd, D.; Lin, B.; Parzybok, T.; Yekta, M. & Riley, D. (2004b). *NOAA Atlas 14, Precip. Freq. Atlas United States*, Vol. 2, U.S. Dep. Comm., Nat. Oceanic Atmos. Administ, Nat. Weather Serv., Silver Spring, Maryland.

Brockwell, P. & Davis, R.A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.), Springer, New York.

Casson, E. & Coles, S.G. (1999). Spatial regression models for extremes, *Extremes*, **1**, 449–468.

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

Coles, S.G. & Dixon, M.J. (1999). Likelihood-based inference for extreme value models, *Extremes*, **2**, 5–23.

Coles, S.G.; Heffernan, J. & Tawn, J.A. (1999). Dependence measures for extreme value analysis, *Extremes*, **2**, 339–365.

Cooley, D.; Davis, R.A. & Naveau, P. (2010). The pairwise beta: A flexible parametric multivariate model for extremes, *J. Mult. Anal.*, **101**, 2103–2117.

Cooley, D.; Naveau, P. & Poncet, P. (2006). Variograms for spatial max-stable random fields, *Lect. Notes Statist.*, **187**, 373–390, Springer, New York.

Cooley, D.; Nychka, D. & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Statist. Assoc.*, **102**, 824–840.

Cooley, D. & Sain, S.R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model, *J. Agric. Bio. Environ. Statist.*, **15**, 381–402.

Cox, D.R. & Reid, N. (2004). A note on pseudo-likelihood constructed from marginal densities, *Biometrika*, **91**, 729–737.

Cressie, N.A.C. & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets, *J. R. Statist. Soc. B*, **70**, 209–226.

Cressie, N.A.C. & Wikle, C. (2011). *Statistics for Spatio-Temporal Data*, Wiley, New York.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Wiley, New York.

Dalrymple, T. (1960). *Flood frequency analyses*, Water supply paper 1543-a, U.S. Geological Survey, Reston, VA.

Daly, C.; Gibson, W.P.; Taylor, G.H.; Johnson, G.L. & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate, *Clim. Res.*, **23**, 99–113.

Daly, C.; Neilson, R.P. & Phillips, D.L. (1994). A statistical topographic model for mapping climatological precipitation in mountainous terrain, *J. Appl. Meteo.*, **33**, 140–158.

Davis, R.A. & Resnick, S.I. (1989). Basic properties and prediction of max-ARMA processes, *Adv. Appl. Prob.*, **21**, 781–803.

Davison, A.C. & Gholamrezaee, M.M. (2012). Geostatistics of extremes, *Proc. R. Soc. Lond. A*, **468**, 581–608.

Davison, A.C.; Padoan, S. & Ribatet, M. (2012). Statistical modeling of spatial extremes (with Discussion), *Statist. Sci.*, to appear.

de Haan, L. (1984). A spectral representation for max-stable processes, *Ann. Prob.*, **12**, 1194–2004.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory*, Springer, New York.

Diggle, P.J.; Tawn, J.A. & Moyeed, R.A. (1998). Model-based geostatistics (with Discussion), *Appl. Statist.*, **47**, 299–350.

Dombry, C.; Éyi-Minko, F. & Ribatet, M. (2011). *Conditional simulations of Brown–Resnick processes*, submitted.

Draisma, G.; Drees, H.; Ferreira, A. & de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli*, **10**, 251–280.

Erhardt, R.J. & Smith, R.L. (2011). Approximate Bayesian computing for spatial extremes, *Comp. Statist. Data Anal.*, to appear.

Fawcett, L. & Walshaw, D. (2006). A hierarchical model for extreme wind speeds, *J. R. Statist. Soc. C*, **15**, 631–646.

Fuentes, M.; Henry, J. & Reich, B. (2011). Nonparametric spatial models for extremes: Application to extreme temperature data, *Extremes*, to appear.

Gelfand, A. & Smith, F. (1990). Sampling-based approaches to calculating marginal densities, *J. Am. Statist. Assoc.*, **85**, 398–409.

Genton, M.G.; Ma, Y. & Sang, H. (2011). On the likelihood function of Gaussian max-stable processes, *Biometrika*, **98**, 481–488.

Guttorp, P. & Xu, J. (2011). Climate change, trends in extremes, and model assessment for a long temperature time series from Sweden, *Environmetrics*, **22**, 456–463.

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

Hosking, J.R.M. & Wallis, J.R. (1997). *Regional Frequency Analysis: An approach based on L-Moments*, Cambridge University Press, Cambridge.

Huser, R. & Davison, A.C. (2012). *Space-time modeling of extreme events*, submitted.

Hüsler, J. & Reiss, R. (1989). Maxima of normal random vectors: between independence and complete dependence, *Statist. Prob. Lett.*, **7**, 283–286.

Jeon, S. (2012). *Max-stable Processes for Threshold Exceedances in Spatial Extremes*, Ph. D. thesis, University of North Carolina at Chapel Hill.

Joe, H. (1997). *Multivariate Models and Dependence Concepts* (1st ed.), Chapman, New York.

Kabluchko, Z.; Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions, *Ann. Prob.*, **37**, 2042–2065.

Leadbetter, M.R.; Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, 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.

Lindgren, F.; Rue, H. & Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Statist. Soc. B*, **73**, 423–498.

Lindsay, B. (1988). Composite likelihood methods, *Contemp. Math.*, **80**, 221–239.

Mannshardt-Shamseldin, E.C.; Smith, R.L.; Sain, S.R.; Mearns, L.O. & Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data, *Ann. Appl. Statist.*, **4**, 484–502.

Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif? *Sciences Terre, Série Inform. Géo.*, **26**, 51–66.

Mendes, J.M.; de Zea Bermudez, P.C. & Pereira, J. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes, *Environ. Ecol. Statist.*, **17**, 1–28.

Mikosch, T. (2006). Copulas: Tales and facts, *Extremes*, **9**, 3–20.

Naveau, P.; Guillou, A.; Cooley, D. & Diebolt, J. (2009). Modeling pairwise dependence of maxima in space, *Biometrika*, **96**, 1–17.

Nelsen, R.B. (2006). *An Introduction to Copulas* (2nd ed.), Springer, New York.

Padoan, S.A. (2011). Multivariate extreme models based on underlying skew-t and skew-normal distributions, *J. Mult. Anal.*, **102**, 977–991.

Padoan, S.A.; Ribatet, M. & Sisson, S.A. (2010). Likelihood-based inference for max-stable processes, *J. Am. Statist. Assoc.*, **105**, 263–277.

Pauli, F.; Racugno, W. & Ventura, L. (2011). Bayesian composite marginal likelihoods, *Statist. Sinica*, **21**, 149–164.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes, *Statist. Prob. Lett.*, **43**, 399–409.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ramos, A. & Ledford, A.W. (2009). A new class of models for bivariate joint tails, *J. R. Statist. Soc. B*, **71**, 219–241.

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

Ribatet, M. (2011). *SpatialExtremes: Modelling Spatial Extremes*, R package version 1.8-1.

Ribatet, M.; Cooley, D. & Davison, A.C. (2011). Bayesian inference for composite likelihood models and an application to spatial extremes, *Statist. Sinica* (doi:10.5705/ss.2009.248).

Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman, Boca Raton.

Sampson, P. & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *J. Am. Statist. Assoc.*, **47**, 108–119.

Sang, H. & Gelfand, A.E. (2009). Hierarchical modeling for extreme values observed over space and time, *Environ. Ecol. Statist.*, **16**, 407–426.

Sang, H. & Gelfand, A.E. (2010). Continuous spatial process models for spatial extreme values, *J. Agric. Bio. Environ. Statist.*, **15**, 49–65.

Schabenberger, O. & Gotway, C.A. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman, Boca Raton.

Schlather, M. (2002). Models for stationary max-stable random fields, *Extremes*, **5**, 33–44.

Schlather, M. (2011). *RandomFields: Simulation and Analysis of Random Fields*, R package version 2.0.53.

Schlather, M. & Tawn, J.A. (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions, *Extremes*, **5**, 87–102.

Schlather, M. & Tawn, J.A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, **90**, 139–156.

Schliep, E.; Cooley, D.; Sain, S. & Hoeting, J. (2010). A comparison study of extreme precipitation from six regional climate models via spatial hierarchical modeling, *Extremes*, **13**, 219–239.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statist. Sinica*, **4**, 639–650.

Sibuya, M. (1960). Bivariate extremal distribution, *Ann. Inst. Statist. Math.*, **11**, 195–210.

Sisson, S. & Fan, Y. (2010). *Likelihood-free Markov chain Monte Carlo*. In "Handbook of Markov Chain Monte Carlo" (S. Brooks, A. Gelman, G. Jones and X. Meng, Eds.), pp. 319–341, Chapman, Boca Raton.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

Smith, R.L. (1990). *Max-stable processes and spatial extremes*, unpublished manuscript.

Varin, C. (2008). On composite marginal likelihoods, *Adv. Statist. Anal.*, **92**, 1–28.

Wackernagel, H. (2003). *Multivariate Geostatistics, An Introduction with Applications* (3rd ed.), Springer, Berlin.

Wang, Y. (2010). *maxLinear: Conditional Sampling for Max-Linear Models*, R package version 1.0.

Wang, Y. & Stoev, S.A. (2011). Conditional sampling for spectrally discrete max-stable random fields, *Adv. Appl. Prob.*, **43**, 461–483.

Zhang, Z. & Smith, R.L. (2004). The behavior of multivariate maxima of moving maxima processes, *J. Appl. Prob.*, **41**, 1113–1123.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted / indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright