Review in Advance first posted online on February 26, 2015. (Changes may still occur before final publication online and in print.)

Statistics of Extremes

A.C. Davison[1] and R. Huser[2]

[1]Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; email: Anthony.Davison@epfl.ch

[2]King Abdullah University of Science and Technology, Thuwal 23995-6900, Saudi Arabia; email: raphael.huser@kaust.edu.sa

Annu. Rev. Stat. Appl. 2015. 2:9.1–9.33

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

This article's doi: 10.1146/annurev-statistics-010814-020133

Keywords

asymptotic dependence, asymptotic independence, extrapolation, generalized extreme value distribution, generalized Pareto distribution, max-stability, Pareto process, peaks over thresholds, Poisson process

Abstract

Statistics of extremes concerns inference for rare events. Often the events have never yet been observed, and their probabilities must therefore be estimated by extrapolation of tail models fitted to available data. Because data concerning the event of interest may be very limited, efficient methods of inference play an important role. This article reviews this domain, emphasizing current research topics. We first sketch the classical theory of extremes for maxima and threshold exceedances of stationary series. We then review multivariate theory, distinguishing asymptotic independence and dependence models, followed by a description of models for spatial and spatiotemporal extreme events. Finally, we discuss inference and describe two applications. Animations illustrate some of the main ideas.

1. INTRODUCTION

Recent years have seen an upsurge of research on the modeling of extreme events, largely driven by impetus from two major areas of application: environmental science and finance. Environmental scientists and related stakeholders such as engineers and insurers have realized that in an evolving climate, changes in the sizes and frequencies of rare events, rather than changes in the averages, may be what lead to the most devastating losses of life and the greatest damage to infrastructure. The ongoing troubles in financial markets worldwide have underscored the centrality of appropriate risk modeling in complex systems, where correct assessment of dependence among extreme events may be critical to gauging how investors, market regulators, and other actors should behave. In these and other applications of the statistics of extremes, the crucial element is extrapolation—not merely forecasting future means and variances, but estimating the probabilities of events beyond our previous experience. This extrapolation entails not only strong assumptions concerning regularity in distribution tails, which allow it to be well founded, but also an act of faith that the phenomenon under consideration does not undergo a phase change beyond the realm of the observed data that renders forecasting useless. The idea that many extreme events are intrinsically unforecastable "black swans" has been popularized by Taleb (2007), although unfortunately he does not mention the existence of a well-developed theory of stochastic extremes that is widely used to estimate risk in practical situations. The centrality of extrapolation in extremes has two important implications: Inferences will typically be more than usually uncertain, and any well-founded knowledge from external sources will be particularly valuable.

The identification of limiting distributions for maxima is attributable to Fisher & Tippett (1928), whose work was further developed by von Mises (1936), Gnedenko (1943), and many others. Slightly earlier, Fréchet (1927) had identified a functional equation, which he called the stability postulate, that provides a mathematical basis for extrapolation and thus lies at the heart of the classical theory of extremes. His stability postulate is now referred to as max-stability (see Equation 2). An early advocate for the application of extreme value theory to engineering problems was E.J. Gumbel, whose 1958 book remained essentially the only text on the topic until the appearance of one by Galambos (1978). Since then, the number of books on extreme value theory has multiplied: Resnick (1987) presents an important early treatment of probability models for extremes; Embrechts et al. (1997) provide a careful account oriented toward applications in finance and insurance; Kotz & Nadarajah (2000) survey the literature up to the start of the twenty-first century; Coles (2001) offers an accessible account of modeling statistical extremes that has done much to bring these ideas to a wide readership; Beirlant et al. (2004) provide a broader description of both probabilistic and statistical aspects; and de Haan & Ferreira (2006) and Resnick (2006) provide detailed mathematical treatments from different perspectives. More recent work is summarized in collections of papers that appear in the March 2012 issue of *REVSTAT Statistical Journal* and the 2013 volume of the *Journal de la Société Française de Statistique*.

2. BASIC PROBABILITY MODELS

2.1. Maxima

In deriving probability models for extremes, it is natural to argue by analogy with the central limit theorem and to ask what distributions may arise as the limit, not for the average, but for the maximum Z_n of independent and identically distributed random variables $Y_1, \ldots, Y_n$, as $n \to \infty$. Fisher & Tippett (1928) showed that if sequences of real numbers $\{b_n\}$ and $\{a_n\} > 0$ can be chosen such that the linearly rescaled maximum $Z_n^* = [\max(Y_1, \ldots, Y_n) - b_n]/a_n$ has a nondegenerate

limiting distribution, then this distribution must have one of three forms, which may be merged to yield the generalized extreme value distribution (GEV),

$$G(z) = \begin{cases} \exp\left\{-[1+\xi(z-\mu)/\sigma]_+^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left[-(z-\mu)/\sigma\right]\right\}, & \xi = 0, \end{cases} \quad (1)$$

where $a_+ = \max(a, 0)$. The quantities μ and $\sigma > 0$ in Equation 1 are location and scale parameters, respectively, and the shape parameter ξ determines the weight of the upper tail of the density. The three classical extreme-value distributions are as follows: (*a*) the Gumbel form ($\xi = 0$); (*b*) the Fréchet form ($\xi > 0$), which has a lower limit at $\mu - \sigma/\xi$; and (*c*) the Weibull (or more correctly, reverse Weibull) form ($\xi < 0$), which has an upper limit at $\mu - \sigma/\xi$. Fisher & Tippett (1928) derived these cases separately and called them Types I, II, and III, respectively; their limiting result is now known as the Extremal Types Theorem. The k th moment of G exists only if $\xi < 1/k$, and the probabilities of extremely large values increase as ξ increases. In practice, it is rare to find $|\xi| > 1$, although this situation can arise in financial applications; estimates of ξ typically lie in the interval $(-1/2, 1/2)$.

The distribution in Equation 2 satisfies Fréchet's max-stability property: For any $n \in \mathbb{N}$, there exist real numbers $a_n > 0$ and b_n such that

$$G^n(b_n + a_n z) = G(z), \quad z \in \mathbb{R}. \quad (2)$$

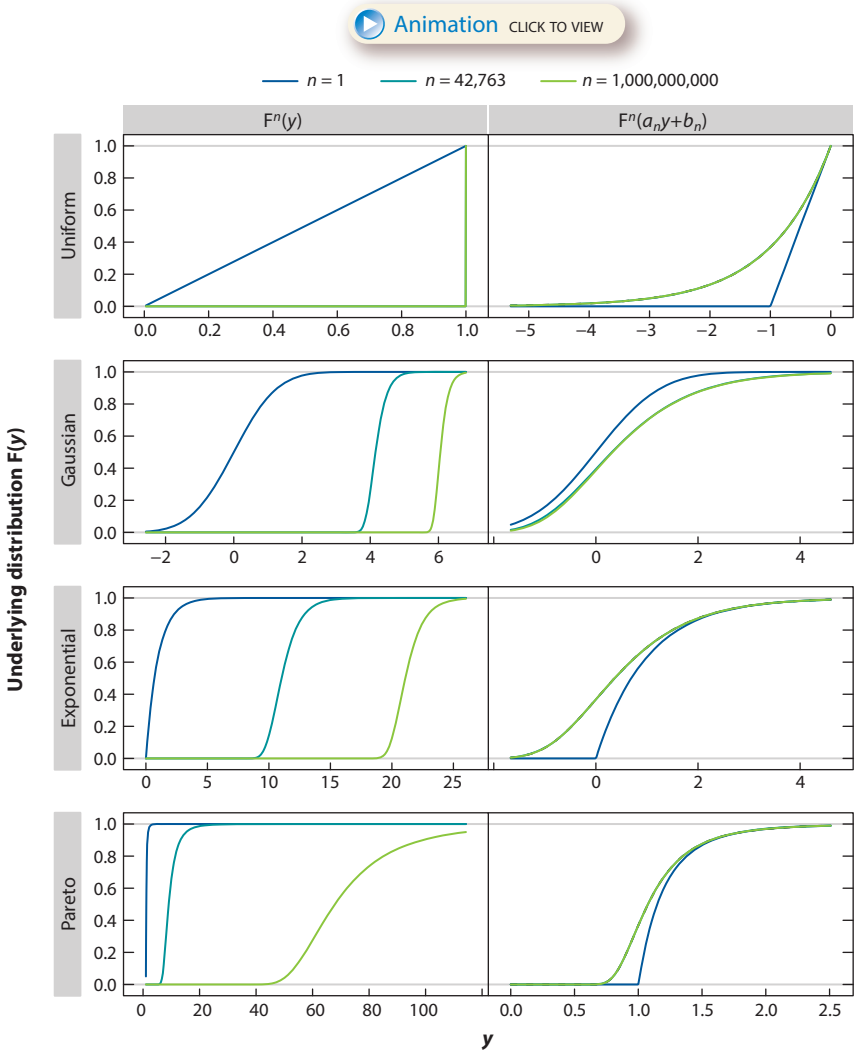
This necessary condition for a limiting distribution for maxima is satisfied only by the GEV, which therefore has strong mathematical support as a suitable distribution for fitting to maxima for scalar random variables. Minima are treated simply by noting that $\min(Y_1, \ldots, Y_n) = -\max(-Y_1, \ldots, -Y_n)$: The results for maxima are simply applied to a transformed version of the original variables, followed by a back-transformation.

The discussion above raises several questions, including the following:

1. Which underlying distributions F for $Y_1, \ldots, Y_n$ yield which members of Equation 1?
2. Given a particular F, how fast is convergence to Equation 1?
3. What happens if the Y_j are dependent?

The first of these questions concerns so-called max-domains of attraction, the sets MDA_ξ of distributions F yielding a particular limiting value of the shape parameter ξ. This question was addressed by von Mises (1936) and particularly by Gnedenko (1943), who provided necessary and sufficient conditions for $F \in \mathrm{MDA}_\xi$ in terms of the upper tail properties of F (de Haan & Ferreira 2006, chapter 1). For example, the Gaussian and exponential distributions belong to MDA_0 and thus yield Gumbel limits for maxima, whereas the uniform distribution is an element of MDA_{-1}. Although many common continuous distributions lie in MDA_ξ for some ξ, some upper tails are so heavy that linear renormalization is inadequate to produce a limit; an example is the log-Pareto distribution, which has survivor function $1/\log y$. In this case, log transformation of the original data produces a limit in MDA_1, suggesting that for practical purposes G might be enlarged. Enlarging G, however, raises two issues: The enlarged distribution would not satisfy Equation 2, and the uncertainty of the fitted model, often uncomfortably large, will increase further. The existence of limits for discrete distributions, or the lack of them, depends on the precise setting (Anderson et al. 1997, Nadarajah & Mitov 2002); it hinges on whether the underlying variables may be treated as approximately continuous or whether the discreteness is dominant. Animation 1 illustrates the convergence of maxima of uniform, exponential, Gaussian, and Pareto variables.

The second question is related to the quality of the approximation provided by the limit (Equation 1) in practice and was considered by Fisher & Tippett (1928), who showed that convergence to the Gumbel limit for underlying Gaussian variables is painfully slow (see Animation 1).



Animation 1

Illustration of the Extremal Types Theorem. For increasing values of n, the left panels display the distribution of the maximum Z_n of n independent uniform (*top row*), standard Gaussian (*second row*), unit exponential (*third row*), and 0.2-Pareto (*bottom row*), i.e., $F(y) = 1 - y^{-0.2}$, $y > 1$, random variables. The right panels display the distribution of $a_n^{-1}(Z_n - b_n)$ for appropriate sequences $a_n > 0$ and b_n. In each case, Z_n is asymptotically degenerate, whereas $a_n^{-1}(Z_n - b_n)$ is not.

Although the quality of approximation is important for extrapolation, the underlying distribution F is very rarely known in applications, and one must invariably assess goodness of fit empirically. As Equation 1 is rather flexible and the number of maxima is rarely large enough to give powerful tests, a lack of fit is often difficult to detect. Moreover, even if the limiting class is known, another member of the distribution given by Equation 1 may provide a better approximation to the distribution of $\max(Y_1, \ldots, Y_n)$ for finite n. For example, Fisher & Tippett (1928) showed that taking $\xi \equiv \xi(n) < 0$, and thus using the Weibull form, provides a better approximation for maxima of

finite samples from the Gaussian distribution than does the limiting Gumbel distribution. This has stimulated a large literature on so-called penultimate approximations (e.g., Anderson 1971, Gomes & de Haan 1999, Kaufmann 2000), through which maxima pass en route to their ultimate destination (Equation 1). In most cases, the penultimate approximation has $\xi \neq 0$ even if the limiting distribution is of the Gumbel form. From a practical point of view, therefore, using Equation 1 has advantages over fitting Type I, II, or III separately, as practiced in the older literature.

The third question has been investigated extensively (Leadbetter et al. 1983, Beirlant et al. 2004, chapter 10): Under mild conditions, frequently plausible in applications, the maxima of contiguous blocks of observations from stationary processes lie in the same MDA_ξ as would independent data with the same marginal distribution. In fact, the limiting distribution is of the form $G^\theta(z)$, where $\theta \in (0, 1]$. The quantity θ, called the extremal index, is discussed in more detail below. The form of the limiting distribution implies that dependence changes only the location and scale parameters, so in realistic settings, the same limiting forms arise for dependent and for independent data.

The upshot is that Equation 1 provides the natural parametric class of probability distributions for maxima. It is typically applied to a stationary time series by dividing the data into M blocks of N contiguous observations, the maxima of which are then modeled using Equation 1. As environmental data often show seasonality, consideration of annual maxima is natural; doing so corresponds to taking $N = 365$ for daily data. Even if the underlying data are neither independent nor identically distributed, the series might be considered approximately stationary during the period in which maxima occur, although the resulting effective block size may then be much smaller than 365. If the series appears stationary, then N must be chosen: For a fixed number of observations, there is a familiar bias–variance trade-off, as taking large N should reduce bias by improving the approximation given by Equation 1. Taking large N will however give fewer block maxima, thus increasing the variance, and the converse is true if N is taken to be small. We discuss further statistical aspects of this trade-off in Section 5.

2.2. Exceedances

In many applications, the use of maxima for inference on extremes has been supplanted by the use of threshold exceedances, for essentially two reasons. First, taking maxima seems to waste information because other extreme observations, such as the second-largest values in the blocks, should also be informative about extremes. Second, detailed modeling of extremal events is often needed: A succession of large values may pose more risk than a single large value does, but an analysis of block maxima does not allow one to model clusters of rare events.

In the independent case, the connection to maxima is simple: The distribution of the rescaled maximum $Z_n^* = [\max(Y_1, \ldots, Y_n) - b_n]/a_n$ of a random sample $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F$ is

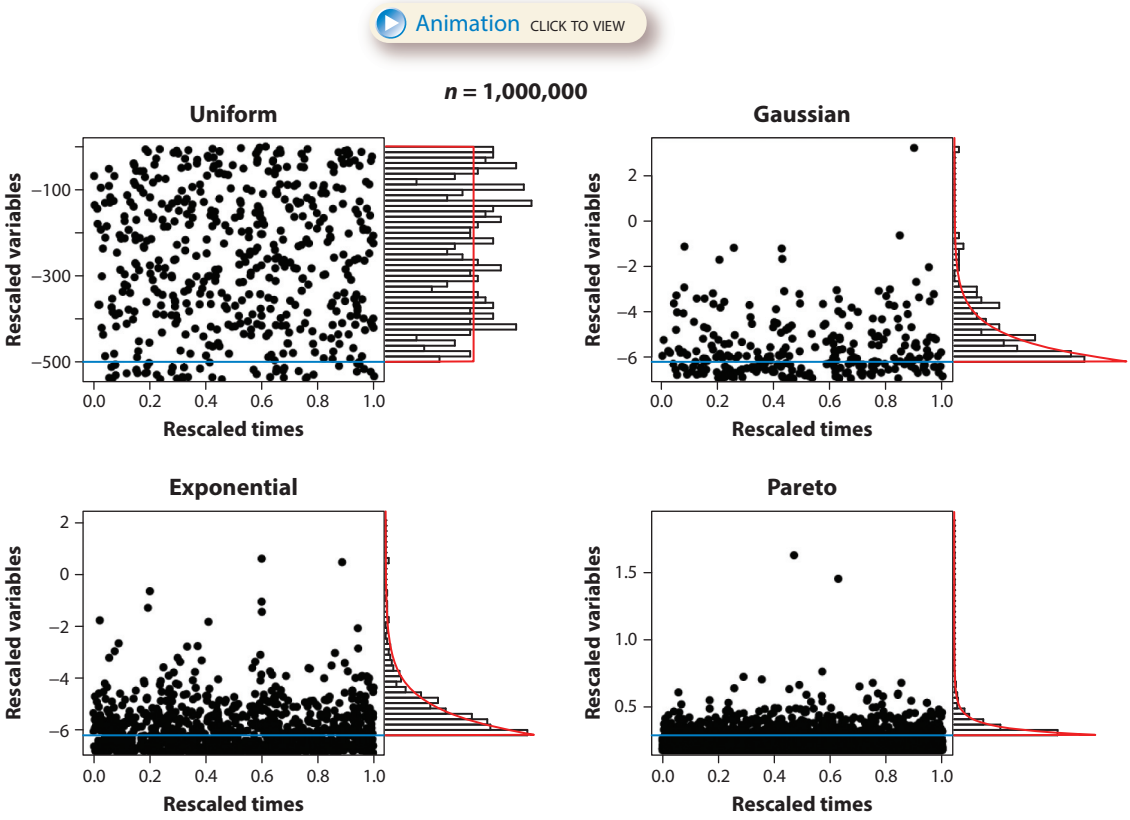
$$\Pr(Z_n^* \leq z) = F^n(b_n + a_n z).$$

As $\log u \doteq u - 1$ when $u \to 1$, convergence of this probability to the $G(z)$ given in Equation 1 as $n \to \infty$ occurs if and only if

$$\begin{aligned} n[1 - F(b_n + a_n z)] &\doteq -n \log F(b_n + a_n z) \\ &\to -\log G(z) \\ &= [1 + \xi(z - \mu)/\sigma]_+^{-1/\xi}. \end{aligned} \tag{3}$$

Now the left-hand side of Equation 3 equals

$$\mathrm{E}\left[\sum_{j=1}^{n} I(Y_j \geq b_n + a_n z)\right],$$



Animation 2

Illustration of the point process of exceedances and the convergence to the GPD. For increasing values of n, the plots display the point process of rescaled times and rescaled variables, namely $(j/(n+1), (Y_j - b_n)/a_n)$, for data simulated from the uniform (*top left*), standard Gaussian (*top right*), unit exponential (*bottom left*), and 0.2-Pareto distributions. The side plots are histograms of the exceedances over the threshold u (*horizontal blue line*), i.e., $a_n^{-1}(Y_j - b_n) | a_n^{-1}(Y_j - b_n) > u$. The solid red curves are the corresponding asymptotic GPD densities.

where $I(\cdot)$ denotes the indicator function, and the Poisson limit for the binomial distribution implies that the rescaled variates $\{(Y_j - b_n)/a_n : j = 1, \ldots, n\}$ converge to a Poisson process on $\mathbb{R}$ with mean measure $-\log G(z)$ (see Animation 2). Similarly, the two-dimensional point process $\{(j/(n+1), (Y_j - b_n)/a_n) : j = 1, \ldots, n\}$ converges to a Poisson process on $[0,1] \times \mathbb{R}$ with the following measure:

$$\Lambda\{[t_1, t_2] \times [z, \infty)\} = (t_2 - t_1)\{1 + \xi(z - \mu)/\sigma\}_+^{-1/\xi}, \quad 0 \le t_1 < t_2 \le 1, z \in \mathbb{R}. \tag{4}$$

This argument implies that events in the limiting process that exceed a threshold u occur according to a Poisson process of rate $\zeta_u = [1 + \xi(u - \mu)/\sigma]_+^{-1/\xi}$ and that

$$\Pr(Y > u + y | Y > u) = \frac{[1 + \xi(y + u - \mu)/\sigma]_+^{-1/\xi}}{[1 + \xi(u - \mu)/\sigma]_+^{-1/\xi}} = (1 + \xi y/\tau_u)_+^{-1/\xi}, \tag{5}$$

say, where $\tau_u = \sigma + \xi(u - \mu)$ must be positive for exceedances of u to arise. The rightmost expression in Equation 5 is the survivor function of the generalized Pareto distribution (GPD) $H(y)$ (Davison & Smith 1990); we see that $\Pr(Y \le u + y) \doteq 1 - \zeta_u(1 + \xi y/\tau_u)_+^{-1/\xi}$ for exceedances

of u and $\Pr(Y \leq u) = 1 - \zeta_u$. Thus the limit provides a distributional approximation for large but finite Y. Although the parameterization (ζ_u, τ_u, ξ) in terms of exceedances and their conditional distributions is easy to interpret, it is often preferable to use the original parameters (μ, σ, ξ), which do not depend on the choice of u. The joint distribution of the largest few order statistics is readily derived from the point process formulation and may be useful for inference.

If the underlying series is dependent, then under mild conditions, the limiting process under which exceedances appear is compound Poisson (Hsing 1987, Hsing et al. 1988, Leadbetter 1991). Clusters of exceedances of u then appear on the time axis according to a Poisson process of rate $\theta\zeta_u$, where $\theta \in (0, 1]$ is the extremal index and the mean cluster size is θ^{-1}. Thus, the mean number of exceedances per unit of time does not change, but because exceedances occur in clusters, the clusters themselves necessarily tend to be further apart.

2.3. Clustering and the Extremal Index

The extremal index of a stationary process summarizes the degree of clustering of its extremes, and it may be defined in several different ways that are equivalent under mild conditions. Let us call $\{u_n\}$ a threshold sequence if $n[1-F(u_n)] \to \zeta > 0$ as $n \to \infty$. The blocks approach defines the extremal index as the reciprocal of the limiting expected number of exceedances of a threshold sequence $\{u_n\}$ in a block of length $r_n = o(n)$. That is, the extremal index is defined as the reciprocal of

$$\lim_{n\to\infty} \sum_{j=1}^{r_n} \mathrm{E}[I(Y_j > u_n) | Z_{r_n} > u_n] = \lim_{n\to\infty} \frac{r_n[1 - F(u_n)]}{\Pr(Z_{r_n} > u_n)}, \tag{6}$$

where $r_n \to \infty$ as $n \to \infty$. A second natural definition uses the subasymptotic extremal index, namely, the probability that an arbitrary exceedance of u is followed by a run of $r-1$ values below u,

$$\theta(u, r) = \Pr(Y_2 \leq u, \ldots, Y_r \leq u | Y_1 > u). \tag{7}$$

Then, the runs definition of the extremal index is defined in terms of a threshold sequence $\{u_n\}$, and a sequence of run lengths $r_n = o(n)$ is defined as the limit

$$\theta = \lim_{n\to\infty} \theta(u_n, r_n).$$

As mentioned above, this limit coincides with that based on the blocks definition under mild conditions. The limiting cluster size distribution may then be defined as follows:

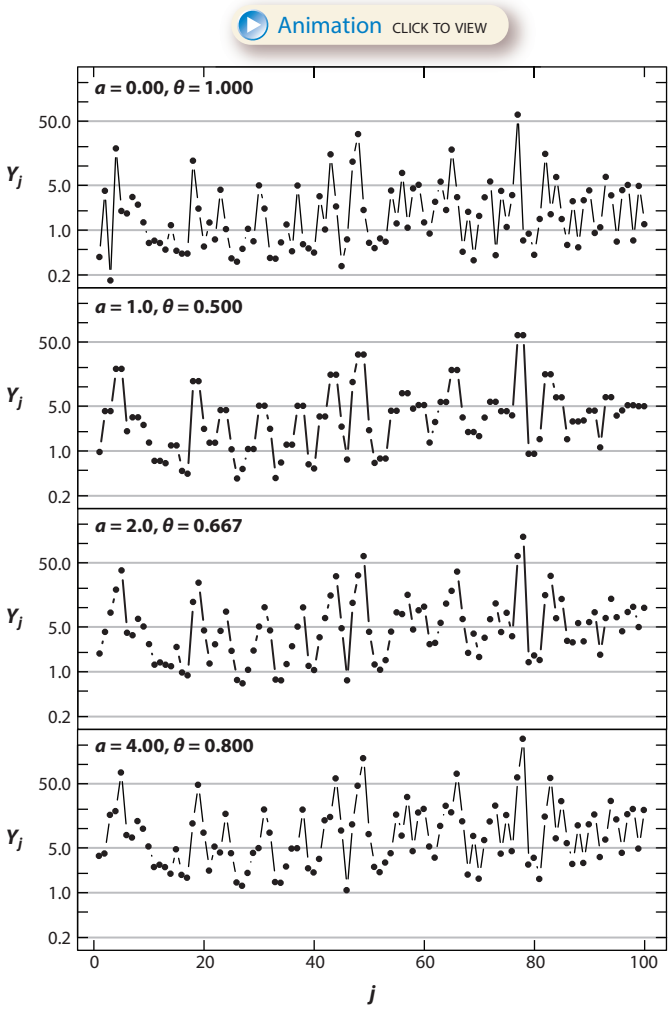
$$\pi(i) = \lim_{n\to\infty} \Pr\left[\sum_{j=1}^{r_n} I(Y_j > u_n) = i \,\middle|\, Y_1 > u_n \right], \; i = 1, 2, \ldots.$$

Animation 3 illustrates the formation of clusters using a model for which $\pi(1) = 1 - a$, $\pi(2) = a$, $\pi(i) = 0$ $(i = 3, 4, \ldots)$, and $\theta = (1 + a)^{-1}$ for $0 < a \leq 1$.

A striking consequence of the discussion above is that cluster maxima and arbitrary exceedances have the same GPD (Equation 5). To see this, let

$$\theta(u_n) = \frac{\Pr(Z_{r_n} > u_n)}{r_n[1 - F(u_n)]} = \frac{\Pr(Z_{r_n} > u_n)}{r_n \Pr(Y_1 > u_n)},$$

and note that this expression tends toward θ for any threshold sequence. Let $y > 0$, and define another threshold sequence $u'_n = u_n + y a_n$ where $\{a_n\}$ denotes the renormalizing sequence needed



Animation 3

Illustration of extremal clustering for data simulated from an ARMAX(a) process with $a \geq 0$, i.e., $Y_j = \max(aY_{j-1}, Z_j)$, $j = 1, 2, \ldots$, where the Z_j are independent unit Frechet random variables. The extremal index for this process is $\theta = \max(1, a)/(1 + a)$, so extreme events form clusters with mean size 2 when $a = 1$. The value of a increases from $a = 0$ to $a = 4$ throughout the animation.

for convergence of maxima. Then,

$$\begin{aligned}\Pr\left(\frac{Z_{r_n} - u_n}{a_n} > y \,\middle|\, Z_{r_n} > u_n\right) &= \frac{\Pr(Z_{r_n} > u_n')}{\Pr(Z_{r_n} > u_n)} \\ &= \frac{\theta(u_n')}{\theta(u_n)} \frac{\Pr(Y_1 > u_n')}{\Pr(Y_1 > u_n)} \\ &= \frac{\theta(u_n')}{\theta(u_n)} \Pr\left(\frac{Y_1 - u_n}{a_n} > y \,\middle|\, Y_1 > u_n\right),\end{aligned}$$

and because the ratio $\theta(u_n')/\theta(u_n) \to 1$ as $n \to \infty$, we see from the stationarity of the process that the limiting distribution of a cluster maximum that exceeds the threshold is the same as that of

an arbitrary observation that exceeds the threshold. One implication of this result is that we may estimate the generalized Pareto parameters using either all of the exceedances or only the cluster maxima. The former approach appears straightforward, but we must allow for dependence among the exceedances in assessing uncertainty (Fawcett & Walshaw (2007). It is not necessary to make this allowance when analyzing the cluster maxima, as they should be independent. These maxima must be identified, however, and incorrect identification may introduce serious bias (Fawcett & Walshaw 2012). A second implication is that at subasymptotic levels, the GPDs H_Z of cluster maxima and H_Y of arbitrary exceedances over a threshold u approximately satisfy the equation $1 - H_Z(y) = [\theta(y, r)/\theta(u, r)\} \times \{1 - H_Y(y)]$, where we have used Equation 7 in place of the subasymptotic block extremal index. We return to this point in Section 5.3

3. MULTIVARIATE PROBABILITY MODELS

3.1. General Framework

Section 2 described probability models appropriate for extremes of a single time series of scalar quantities. In practice, however, the joint modeling of extremes from a D-dimensional time series $\boldsymbol{Y}_j = (Y_{j,1}, \ldots, Y_{j,D})$ $(j = 1, 2, \ldots)$ may be of interest for several reasons. First, one may want to have a qualitative description of both the structure and degree of extremal dependence between two or more series of observations. This is especially important for risk assessment. In hydrology, for example, extreme rainfall events that occur simultaneously over a whole catchment may greatly increase the overall risk of flooding (Thibaud et al. 2013). If several stock markets have huge losses on the same day, the risk of a global financial collapse increases. In aviation safety analysis, a combination of several poor performance factors may increase the probability of a serious accident (Einmahl et al. 2009). Thus, if extremal dependence is not properly treated, one might misestimate the associated risk. Second, the use of a multivariate model permits us to treat data in a general and coherent way, and the interpretation of the results is sometimes easier in a multivariate framework. Third, joint modeling of extremal dependence may allow us to borrow strength from related time series in order to better estimate marginal parameters. In spatial statistics, this practice is often called the trade-off between space and time.

Although appealing, the joint modeling of extremes is difficult in several respects. First, there is no obvious way to order multivariate observations, so the definition of an extreme multivariate observation is not as clear as that for univariate data. Second, as explained below, the class of multivariate extreme value distributions is nonparametric; unlike the univariate case, it cannot be characterized by a finite number of parameters. Third, bivariate extreme value theory is fairly well understood and developed, but few flexible parametric models for higher dimensions exist. Finally, the curse of dimensionality cannot be avoided: Modeling, fitting, simulation, and model checking are more demanding and computationally intensive for large D.

By analogy with the univariate case, it is natural to study the class of limiting distributions for componentwise maxima $\boldsymbol{Z}_n = (Z_{n,1}, \ldots, Z_{n,D})$, which stem from a stream of independent random vectors $\boldsymbol{Y}_j = (Y_{j,1}, \ldots, Y_{j,D})$ $(j = 1, 2, \ldots)$. One might question this approach because the vector of componentwise maxima may not correspond to a real observation, although each of its components does. However, as explained below, there is a close link between the joint distribution of $\boldsymbol{Z}_n$ and the joint upper tail of the original observations $\boldsymbol{Y}|!_j$. Clearly, if there exist sequences $\{a_{n,d}\} > 0$ and $\{b_{n,d}\}$ $(d = 1, \ldots, D)$ such that the vector of renormalized componentwise maxima $\boldsymbol{Z}_n^* = (a_{n,1}^{-1}(Z_{n,1} - b_{n,1}), \ldots, a_{n,D}^{-1}(Z_{n,D} - b_{n,D}))$ has a limiting distribution $G(z_1, \ldots, z_D)$ with nondegenerate marginals, these marginals must belong to the GEV family (recall Equation 1). The characterization of the possible dependence structures is much more complex, and it cannot be

summarized by a parametric family. Without loss of generality, we shall assume from now on that the original observations have unit Frechet marginals, i.e., $\Pr(Y_{j,d} \leq y) = \Pr(Z^*_{n,d} \leq y) = \exp(-1/y)$, $y > 0$, where the normalizing sequences are $a_{n,d} = n$, $b_{n,d} = 0$ $(d = 1, \ldots, D)$. Then, the limiting distribution for $\boldsymbol{Z}^*_n$, called a multivariate extreme-value distribution, may be expressed as

$$G(z_1, \ldots, z_D) = \exp[-V(z_1, \ldots, z_D)], z_1, \ldots, z_D > 0, \tag{8}$$

where the function $V(z_1, \ldots, z_D)$ is called the underlying exponent measure. A valid exponent measure can be provided by any positive function $V(z_1, \ldots, z_D)$ that (*a*) is homogeneous of order -1 [i.e., for which $lV(lz_1, \ldots, lz_D) = V(z_1, \ldots, z_D)$ for any $l > 0$ and $z_1, \ldots, z_D > 0$] and (*b*) satisfies the marginal constraints $V(z, \infty, \ldots, \infty) = 1/z$ for any permutation of the D arguments. The homogeneity property stems from the max-stability of the limiting distribution (Equation 8), which must satisfy the vector analog of Equation 2, whereas the marginal constraints ensure that $G(z_1, \ldots, z_D)$ has unit Frechet marginal distributions. Clearly, the exponent measure provides information about the extremal dependence structure of the original variables $\boldsymbol{Y}_j$: When $V(z_1, \ldots, z_D) = 1/z_1 + \cdots + 1/z_D$, the variables are asymptotically independent, and when $V(z_1, \ldots, z_D) = \max(1/z_1, \ldots, 1/z_D)$, they are completely dependent. Furthermore, by noting that for large values of $z_1, \ldots, z_D$,

$$G(z_1, \ldots, z_D) \approx 1 - V(z_1, \ldots, z_D), \tag{9}$$

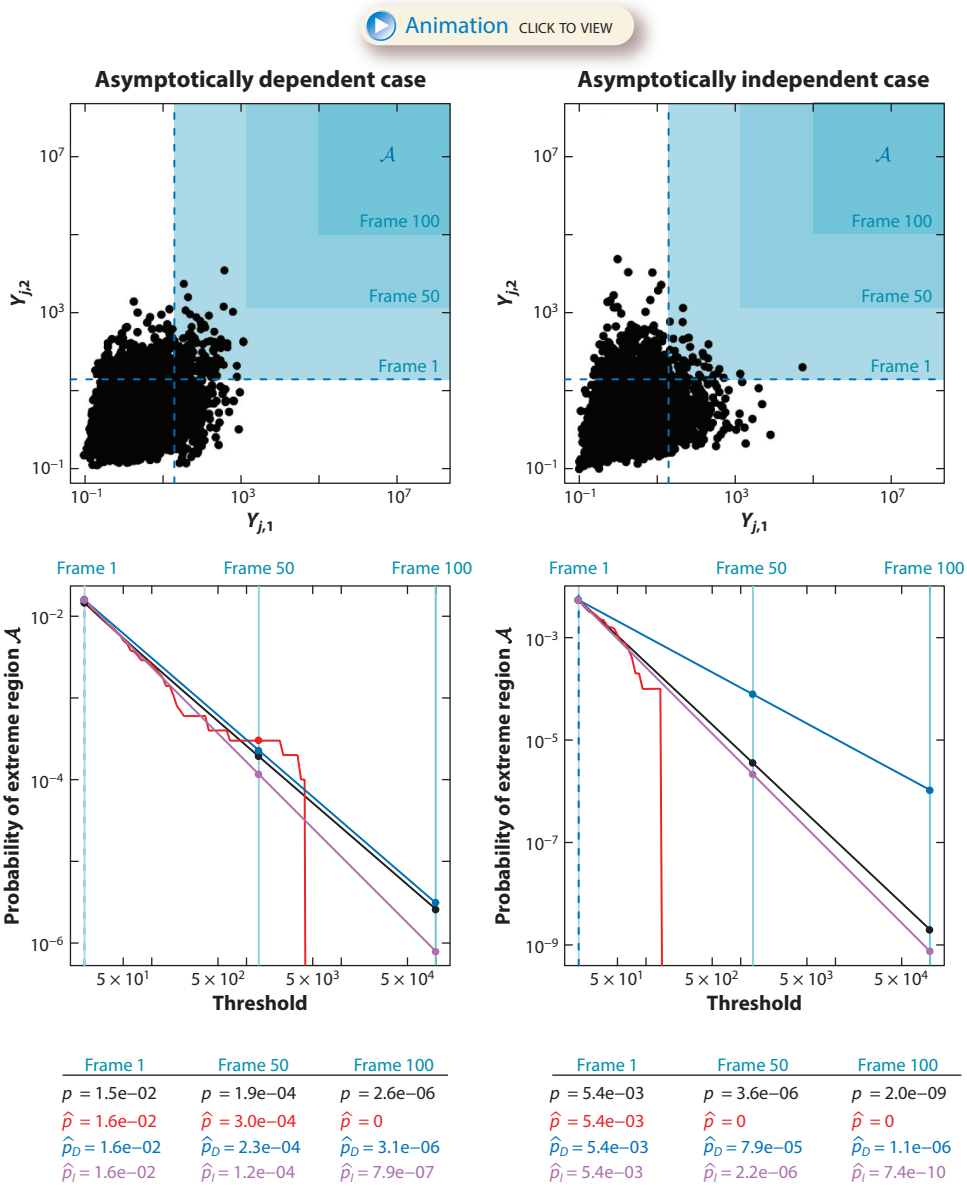
the exponent measure can be interpreted as the approximate probability that $\boldsymbol{Z}^*_n$ lies in $\mathbb{R}^D_+ \setminus \times^D_{d=1} [0, z_d]$, that is, the probability that at least one of the individual maxima $Z_{n,d}$ exceeds its threshold nz_d. Moreover, if $\boldsymbol{Z} = (Z_1, \ldots, Z_D)$ is distributed according to Equation 8, more insight can be obtained by considering the variables

$$R = \sum_{d=1}^{D} Z_d, \; W_1 = \frac{Z_R}{1}, \ldots, W_D = \frac{Z_D}{R}, \tag{10}$$

where $\sum_{d=1}^{D} W_d = 1$. The variable R represents the overall size of the vector $\boldsymbol{Z}$, and W_d may be interpreted as the relative importance of the dth variable. For large values of $Z_1, \ldots, Z_D$ and therefore large R, the approximation given by Equation 9 holds, and the associated probability density function, if it exists, may therefore be expressed in terms of the pseudopolar coordinates. This expression is as follows:

$$\begin{aligned} g(r, w_1, \ldots, w_{D-1}) &\approx -V_{1:D}(rw_1, \ldots, rw_D) \times |\text{Jacobian}| \\ &= -\frac{1}{r^2} V_{1:D}(w_1, \ldots, w_D), \end{aligned} \tag{11}$$

where $V_{1:D}$ denotes the derivative of V with respect to all of its components, and Equation 11 follows from the facts that the Dth derivative of the of the exponent measure is homogeneous of order $-(D + 1)$ and the Jacobian of the transformation to the new coordinates equals r^{D-1}. Hence, we see from Equation 11 that the radial and angular parts of $\boldsymbol{Z}$ are asymptotically independent as $r \to \infty$. Loosely speaking, this means that the global extremeness of several variables, summarized by R, is not influenced by their relative importances, summarized by the W_d, and vice versa. Moreover, the right tail of the pseudoradius R decays as $O(r^{-2})$, $r \to \infty$, i.e., as a standard Pareto density, and the asymptotic joint density $q(w_1, \ldots, w_D)$ of the pseudoangles $(W_1, \ldots, W_D)$ is proportional to $-V_{1:D}(w_1, \ldots, w_D)$. This fact may be used to estimate the probability of extremely rare events that have not yet been observed: An estimate of $q(w_1, \ldots, w_D)$ can be obtained from the available data, and the regularity of R allows extrapolation to the upper tail of $\boldsymbol{Z}$ (see the illustration in Animation 4).



Frame 1	Frame 50	Frame 100
p = 1.5e−02	p = 1.9e−04	p = 2.6e−06
$\hat{p}$ = 1.6e−02	$\hat{p}$ = 3.0e−04	$\hat{p}$ = 0
$\hat{p}_D$ = 1.6e−02	$\hat{p}_D$ = 2.3e−04	$\hat{p}_D$ = 3.1e−06
$\hat{p}_I$ = 1.6e−02	$\hat{p}_I$ = 1.2e−04	$\hat{p}_I$ = 7.9e−07

Frame 1	Frame 50	Frame 100
p = 5.4e−03	p = 3.6e−06	p = 2.0e−09
$\hat{p}$ = 5.4e−03	$\hat{p}$ = 0	$\hat{p}$ = 0
$\hat{p}_D$ = 5.4e−03	$\hat{p}_D$ = 7.9e−05	$\hat{p}_D$ = 1.1e−06
$\hat{p}_I$ = 5.4e−03	$\hat{p}_I$ = 2.2e−06	$\hat{p}_I$ = 7.4e−10

Animation 4

Illustration of the estimation of low-probability events. The upper panels display asymptotically dependent (*left*) and asymptotically independent (*right*) data, along with a target extreme region A. The numbers correspond to the true probability p that a point lies in A (*black*), its naive empirical estimate $\hat{p}$ (*red*), its estimate $\hat{p}_D$ under asymptotic dependence using Equation 12 (*blue*) and its estimate $\hat{p}_I$ under asymptotic independence using Equation 23 (*purple*). For $\hat{p}_D$ and $\hat{p}_I$, extrapolation is based on the empirical estimate at the 0.95 level (*dashed blue lines*), and $\hat{p}_I$ uses an estimate $\hat{\eta}$ of the coefficient of tail dependence proposed in an article by Ledford & Tawn (1996). The bottom panels show these probabilities as a function of the threshold (i.e., the x-coordinate of the lower left corner of A).

From Equation 11, one can deduce that for suitable extreme sets $\mathcal{A}$,

$$\Pr(\boldsymbol{Z} \in v\mathcal{A}) \approx v^{-1} \Pr(\boldsymbol{Z} \in \mathcal{A}), \quad v \geq 0, \tag{12}$$

and consequently, for suitable extreme sets $\mathcal{A}'$, $\Pr[\log(\boldsymbol{Z}) \in v' + \mathcal{A}'] \approx \exp(-v') \Pr[\log(\boldsymbol{Z}) \in \mathcal{A}']$. Thus, extrapolation with unit Fréchet margins is performed along rays emanating from the origin, whereas extrapolation with Gumbel margins, obtained by logarithmic marginal transformation, is performed along lines parallel to the diagonal. Wadsworth & Tawn (2013) develop characterizations for multivariate extremes that allow extrapolation along other trajectories.

The decomposition into pseudopolar coordinates yields a useful spectral representation of the exponent measure, namely,

$$V(z_1, \ldots, z_D) = D \int_{\mathbb{S}_D} \max\left(\frac{w_1}{z_1}, \ldots, \frac{w_D}{z_D}\right) \mathrm{d}Q(w_1, \ldots, w_D). \tag{13}$$

In this representation, Q is a probability measure on the $(D-1)$ -dimensional simplex $\mathbb{S}_D = \{(w_1, \ldots, w_D) : \sum_{d=1}^{D} w_d = 1\}$, which satisfies the mean constraints

$$\int_{\mathbb{S}_D} w_d \mathrm{d}Q(w_1, \ldots, w_D) = D^{-1}, \quad d = 1, \ldots, D. \tag{14}$$

If Q has a probability density function, then this equals $q(w_1, \ldots, w_D) = -D^{-1} V_{1:D}(w_1, \ldots, w_D)$. Asymptotic independence is attained when the measure Q has mass D^{-1} on each of the vertices of $\mathbb{S}_D$, whereas complete dependence corresponds to placing all the mass at the barycenter $(D^{-1}, \ldots, D^{-1})$. Animation 5 illustrates that the empirical density estimator of the pseudoangles for bivariate Gaussian data converges to point masses at the boundaries, i.e., $\mathrm{d}Q(0, 1) = \mathrm{d}Q(1, 0) = 1/2$ and $\mathrm{d}Q(w, 1-w) = 0$ for $0 < w < 1$. In contrast, the pseudoangles for bivariate Student t data are more evenly distributed above large thresholds and have a limiting continuous density $q(w, 1-w)$, $w \in [0, 1]$, which has a positive mass in the interior of the interval $[0, 1]$.

As an alternative to Equation 13, Pickands (1981) proposed another characterization of the exponent measure:

$$V(z_1, \ldots, z_D) = \sum_{d=1}^{D} \frac{1}{z_d} A\left(\frac{z_1}{z_1 + \cdots + z_D}, \cdots, \frac{z_D}{z_1 + \cdots + z_D}\right), \tag{15}$$

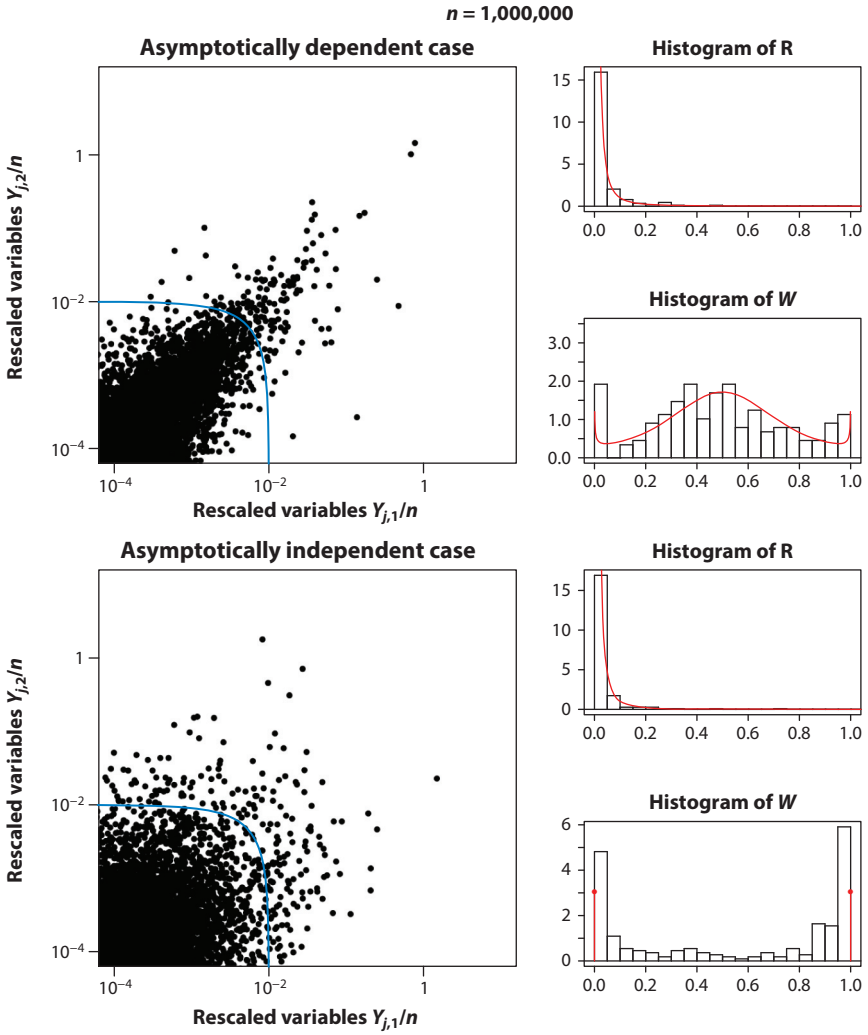
where $A(w_1, \ldots, w_D)$ is called the Pickands dependence function. In the bivariate case, the function $\tilde{A}(w) \equiv A(w, 1-w)$ is convex, lies within the set $\max(w, 1-w) \leq \tilde{A}(w) \leq 1$ (the lower boundary corresponds to perfect dependence, and the upper boundary corresponds to asymptotic independence), and satisfies $\tilde{A}(0) = \tilde{A}(1) = 1$. Similar conditions apply in the multivariate framework. Clearly, there are one-to-one mappings between the family of multivariate extreme-value distributions (Equation 8), the class of exponent measures V, the class of probability measures Q and the class of Pickands dependence functions A.

Numerous parametric models for bivariate extremes have been proposed (Kotz & Nadarajah 2000, chapter 3), but the literature on higher dimensions is much sparser, owing to the difficulty of formulating general classes of models that satisfy the mean constraints. Some recent contributions are those from Ballani & Schlather (2011) and Cooley et al. (2010).

3.2. Point Process Formulation

Similar to the case of univariate extremes, a closely related point process characterization may be formulated in the multivariate case. Under the assumptions leading to Equation 8, as $n \to \infty$, the

Animation CLICK TO VIEW



Animation 5

Illustration of the point process of exceedances in the bivariate framework for increasing n. The upper left panel displays bivariate Student t data with 2 degrees of freedom and standard Pareto marginals, rescaled by n. These data are asymptotically dependent. The right-hand panels show the corresponding histograms of the pseudoradius R and pseudoangle W for points lying above the threshold defined by $R > 10^{-2}$ (*solid blue line*), with theoretical limiting densities superimposed in red. The bottom plots illustrate the Gaussian case, which is asymptotically independent.

point process of rescaled variables

$$\left\{\left(\frac{Y_{j,1}}{n}, \cdots, \frac{Y_{j,D}}{n}\right) : j = 1, \ldots, n\right\} \tag{16}$$

converges weakly on the cone $\mathbb{R}_+^D \backslash \{(0, \ldots, 0)\}$ to a Poisson process with mean measure

$$\Lambda(\mathbb{R}_+^D \backslash [0, y_1] \times \cdots \times [0, y_D]) = V(y_1, \ldots, y_D). \tag{17}$$

If the corresponding intensity function exists, it equals

$$\lambda(y_1, \ldots, y_D) = -V_{1:D}(y_1, \ldots, y_D) = \frac{D}{r^2} q(w_1, \ldots, w_D). \tag{18}$$

When the margins are not unit Frechet, the normalization in Equation 16 instead uses the sequences $\{a_{n,d}\}$ and $\{b_{n,d}\}$ that appear in Equation 8, and the quantities in Equations 17 and 18 must have suitable marginal transformations applied. In Animation 5, the point process (Equation 16) is plotted for bivariate Gaussian and Student t data that have been transformed to have standard Pareto margins, that is, to lie in the maximum domain of attraction of the unit Fréchet distribution. The effect of the normalization is that although the points $n^{-1}\boldsymbol{Y}_j$ shrink toward the origin as n increases, the distribution of points in any region bounded away from the origin stabilizes. Expression 17 implies that extreme events defined as the vectors $\boldsymbol{Y}_j$ for which at least one component is large tend to occur independently, and the number of those events within some extreme region $\mathcal{A} \subset \mathbb{R}_+^D$ has a Poisson distribution with mean $\Lambda(\mathcal{A})$. This result is consistent with Equation 8, as

$$\begin{aligned} \Pr(Z_{n,1}^* \le z_1, \ldots, Z_{n,D}^* \le z_D) &= \Pr(\text{No point } n^{-1}\boldsymbol{Y}_j \text{ in } \mathbb{R}_+^D \backslash \times_{d=1}^D [0, z_d]) \\ &\to \exp[-V(z_1, \ldots, z_D)], \, n \to \infty. \end{aligned} \tag{19}$$

Equation 19 in this case is implied by the limiting Poisson process of the rescaled variables $n^{-1}\boldsymbol{Y}_j$ as $n \to \infty$. The convergence of measures that underpins the proof of Equation 17 justifies the following large sample approximation:

$$\begin{aligned} F(ny_1, \ldots, ny_D) &= \{\Pr(Z_{n,1} \le ny_1, \ldots, Z_{n,D} \le ny_D)\}^{1/n} \\ &\approx \exp[-V(y_1, \ldots, y_D)]^{1/n}, \quad n \to \infty, \\ &= \exp[-V(ny_1, \ldots, ny_D)]. \end{aligned} \tag{20}$$

This equation suggests that the max-stable model approximates the original distribution F in its joint upper-tail region. Hence, the limiting dependence structure of componentwise maxima is essentially the same as that of large original observations. Moreover, if we extend Equation 5 to the multivariate framework, the point process convergence also implies that the joint distribution of $\boldsymbol{Y}_j$, conditional on there being at least one extreme component, can be approximated using a multivariate GPD (Rootzén & Tajvidi 2006).

3.3. Asymptotic Dependence and Independence

Animation 5 illustrates an awkward truth. Several common multivariate distributions, including the multivariate Gaussian, show ever-decreasing dependence at higher levels—termed asymptotic independence. Asymptotic dependence, illustrated by the Student t distribution in Animation 5 and described in Section 3.1, means that the sizes and shapes of extreme events are independent at high levels (see Equation 11 and Animation 4). Thus, the ability to identify situations of asymptotic independence is a useful one.

Coles et al. (1999) noted that a vector (Y_1, Y_2) with common unit Fréchet margins is asymptotically dependent if

$$\Pr(Y_1 > y | Y_2 > y) \to \chi > 0, \quad y \to \infty, \tag{21}$$

whereas the variables are asymptotically independent when $\chi = 0$. If the variables are asymptotically dependent, then $\chi = 2 - V(1, 1)$, making a link to the exponent measure defined in Equation 8. When $\chi = 0$, the rate of convergence to zero is crucial because of its implications for extrapolation to the joint upper tail of the variables (see Equation 23). This led Ledford & Tawn (1996) to propose an asymptotic independence model capable of detecting and measuring the speed of the decay toward independence at high levels via the coefficient of tail dependence, η. They assumed that the conditional probability in Equation 21 may be written as follows:

$$\Pr(Y_1 > y | Y_2 > y) \sim \mathcal{L}(y) y^{1-1/\eta}, \quad y \to \infty, \tag{22}$$

where $\eta \in (0, 1]$, and $\mathcal{L}(y)$ is a slowly varying function. That is, $\mathcal{L}(vy)/\mathcal{L}(y) \to 1$ for any $v > 0$ as $y \to \infty$, making the variables asymptotically dependent if $\eta = 1$ and $\mathcal{L}(y) \not\to 0$ as $y \to \infty$, and asymptotically independent otherwise. Under this mild assumption, the extrapolation formula in Equation 12 is replaced by

$$\Pr(\boldsymbol{Z} \in v\mathcal{A}) \approx v^{-1/\eta} \Pr(\boldsymbol{Z} \in \mathcal{A}), \quad v \geq 0. \tag{23}$$

Hence, if an asymptotic dependence model is wrongly assumed to be valid, probabilities of extremely rare events will be overestimated, and conversely (see Animation 4), potentially leading to serious misestimation of risk (Davison et al. (2013). Further refinements of Equation 22 have been proposed by Ledford & Tawn (1997, 2003), Ramos & Ledford (2009, 2011), and Wadsworth & Tawn (2013), among others. All of these models are generally justified under mild assumptions usually referred to as hidden regular variation (Resnick 2002, Das & Resnick 2011). (Wadsworth & Tawn (2012b) also tackle this problem in the spatial framework and provide hybrid models that can handle both asymptotic dependence and asymptotic independence at different distances. In practice, however, the coefficient of tail dependence is difficult to estimate, as it relates to the joint behavior of the data at infinity. Thus, careful assessment of the plausibility of asymptotic independence is required.

Generalizations of the approaches described above to dimensions greater than three have proven elusive. Heffernan & Tawn (2004) proposed a formulation of the problem that covers both asymptotic dependence and independence cases and may be applied in high dimensions, but their formulation requires conditioning on the value of a particular variable. Let $t_d(\cdot)$ denote a transformation of the original observations $\boldsymbol{Y}_j$ to the Gumbel scale, i.e., $t_d(Y_{j,d}) = -\log[-\log F_d(Y_{j,d})]$, where F_d is the distribution of $Y_{j,d}$. Suppose that there exist $(D-1)$-dimensional functions $\boldsymbol{a}(y)$ and $\boldsymbol{b}(y) > 0$, with respective components $(a_2(y), \ldots, a_D(y))$ and $(b_2(y), \ldots, b_D(y))$, for which

$$\Pr\left\{ \frac{t_d(Y_d) - a_d[t_1(Y_1)]}{b_d[t_1(Y_1)]} \leq z_d, d = 2, \ldots, D \middle| Y_1 > u \right\} \to J(\boldsymbol{z}), \quad u \to y_*,$$

where $y_* = \sup\{y : F_1(y) < 1\}$ is the upper terminal for Y_1, all of the marginal distributions of J are nondegenerate, and $\boldsymbol{z} = (z_2, \ldots, z_D)$. The a_d and b_d here are unrelated to those used in Sections 2 and 3.1. It turns out that the standardized values of the Y_d, i.e., the vector $\boldsymbol{Z}(Y_1)$ with components

$$Z_d(Y_1) = \frac{t_d(Y_d) - a_d[t_1(Y_1)]}{b_d[t_1(Y_1)]}, \quad d = 2, \ldots, D,$$

are asymptotically independent of Y_1, conditional on $Y_1 > u$:

$$\Pr[\boldsymbol{Z}(Y_1) \leq \boldsymbol{z}, (Y_1 - u)/\tau_u > y | Y_1 > u] \to J(\mathbf{z})(1 + \xi y)_+^{-1/\xi}, \quad u \to y_*,$$

where τ_u appears in Equation 5. Thus, the limiting conditional distribution of Y_1 is generalized Pareto, independent of the standardized versions of the other variables.

Many data sets show positive association(s) between the underlying variables, in which case the component functions of $\boldsymbol{a}(y)$ and $\boldsymbol{b}(y)$ generally have the simple forms

$$a_d(y) = \alpha_d y, \quad b_d(y) = y^{\beta_d}, \quad d = 2, \ldots, D,$$

where $0 \leq \alpha_d \leq 1$ and $-\infty < \beta_d \leq 1$. Further constraints on these parameters are given by Keef et al. (2013) and Liu & Tawn (2014). The values of α_d and β_d determine the structure of the model: Y_1 and Y_d are asymptotically dependent only if $\alpha_d = 1$ and $\beta_d = 0$, whereas they are asymptotically independent if $0 \leq \alpha_d < 1$ [see Heffernan & Tawn (2004) for further discussion]. Using these forms for the functions, and for a high conditioning threshold u, this conditional model can be formulated as follows:

$$Y_d = \alpha_d y + y^{\beta_d} Z_d(y), \quad d = 2, \ldots, D, \quad Y_1 = y > u,$$

where the joint distribution j of $Z_2(y), \ldots, Z_D(y)$ is independent of Y_1. This distribution can in fact be any nondegenerate distribution. Inference for this semiparametric model may be performed using a variety of techniques, including quasi-likelihood.

The privileged role of Y_1 makes this model asymmetric and thus not suitable for every situation, but, unlike the models mentioned previously, it may be applied to very high-dimensional data sets. Butler et al. (2007) and Keef et al. (2009), among others, provide examples of its application. Further developments on theory and methods related to asymptotic independence and hidden regular variation include those by Heffernan & Resnick (2005, 2007) and Weller & Cooley (2014).

4. SPATIAL SETTINGS

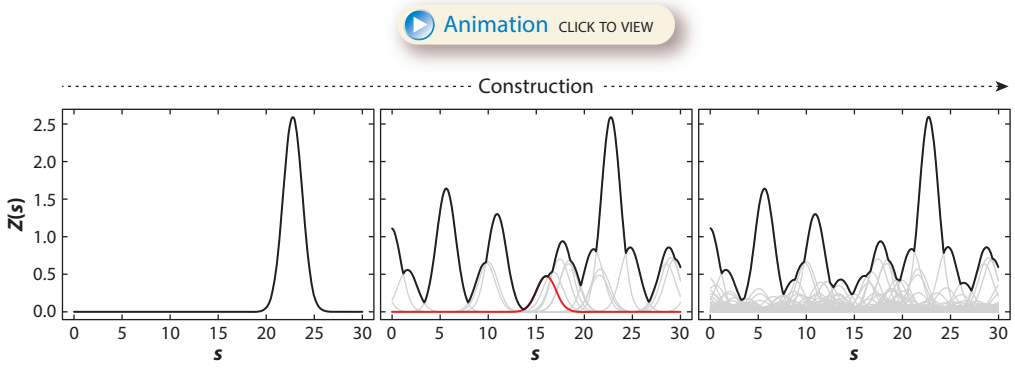
4.1. Max-Stable Processes

Many extreme-value problems are spatial or spatiotemporal in nature. For instance, Westra & Sisson (2011) use spatial extreme-value processes to investigate patterns of extreme precipitation in Australia and to simulate spatial fields comprising observations from multiple point locations. Blanchet & Davison (2011) fit spatial models to extreme snowfall data, with a view toward risk management in Alpine regions. Davison & Gholamrezaee (2012) model annual temperature maxima, a large-scale phenomenon compared with rainfall, and Huser & Davison (2014) extend their ideas to spatiotemporal modeling of extreme rainfall.

The natural way to tackle such problems is to generalize the previous models (Equations 1 and 8) to the spatial framework. Thus random processes valid for spatial extremes should possess a max-stability property analogous to that given by Equation 2. A random process $Z(s)$ defined for $s \in S \subset \mathbb{R}^p$ is said to be max-stable if, for each positive integer n, there exist continuous functions $a_n(s) > 0$ and $b_n(s)$ such that for any function $z(s)$,

$$\{\Pr[Z(s) \leq a_n(s)z(s) + b_n(s), s \in S]\}^n = \Pr[Z(s) \leq z(s), s \in S], \tag{24}$$

or equivalently, $Z(s)$ and the pointwise maximum of n independent copies of $[Z(s) - b_n(s)]/a_n(s)$ have the same finite-dimensional distributions. Clearly, any D-variate sample $\{Z(s_1), \ldots, Z(s_D)\}$ from such a process has a multivariate extreme-value distribution, and its marginal distributions are GEVs. As for the finite-dimensional case, if $Y_1(s), \ldots, Y_n(s)$ denote independent replicates of a continuous random process $Y(s)$ $(s \in S)$, then the only limits for the process of renormalized pointwise maxima $Z_n^*(s) = a_n^{-1}(s)\{\max[Y_1(s), \ldots, Y_n(s)] - b_n(s)\}$ that arise as $n \to \infty$ satisfy Equation 24. Thus, max-stable processes form natural models for spatial maxima. We now discuss some special cases.



Animation 6

Illustration of the construction and simulation of a max-stable process, here a unidimensional Smith model. A large (but in theory, infinite) number of random storms with random decreasing sizes are generated (*light gray curves*), and the resulting max-stable process corresponds to the pointwise supremum.

To construct max-stable processes, we consider independent replicates $\{X_i(s)\}$ of a positive unit mean random process $X(s)$ $(s \in S)$ and a unit rate Poisson process $\{P_i\}$ on the positive halfline, with $0 < P_1 < P_2 < \cdots$. Then under mild conditions, the process defined by

$$Z(s) = \sup_{i=1,2,\cdots} X_i(s)/P_i \tag{25}$$

is max-stable with unit Fréchet margins, and each max-stable process with unit Fréchet margins can be represented using Equation 25 (see de Haan 1984; Schlather 2002; de Haan & Ferreira 2006, chapter 9). Animation 6 illustrates the construction above (Equation 25) for a rather special process $X(s)$.

The representation given by Equation 25 is not only theoretically powerful but also extremely useful in model construction, as different max-stable models stem from different choices of $X(s)$. Expression 25 also suggests a natural simulation algorithm for max-stable processes and gives an elegant expression for the associated finite-dimensional distributions. The joint distribution of $Z(s)$ observed at locations $s_1, \ldots, s_D \in \mathcal{S}$ satisfies Equation 26 with exponent measure

$$V(z_1, \ldots, z_D) = E\left\{\max\left[\frac{X(s_1)}{z_1}, \cdots, \frac{X(s_D)}{z_D}\right]\right\}. \tag{26}$$

Modeling requires flexible processes $X(s)$ for which Equation 26 can be computed explicitly, although in some cases the expectation is tractable only for $D = 2$.

Smith (1990) proposed a max-stable model with $X(s) = \phi_p(s - U; \boldsymbol{\Omega})$, where $\phi_p(0; \boldsymbol{\Omega})$ is the p-dimensional Gaussian density with mean zero and variance matrix $\boldsymbol{\Omega}$, and U is a point of a unit rate Poisson process on $\mathcal{S}$; its exponent measure in dimension $D \leq p + 1$ is provided by Genton et al. (2011). As Animation 6 shows, this process is too smooth for realism in most applications. The extremal t model (Nikoulopoulos et al. 2009, Opitz 2013) takes $X(s) = \sqrt{\pi}2^{1-\nu/2}\Gamma[(\nu + 1)/2]^{-1} \max[0, \varepsilon(s)]^\nu$, where $\nu > 0$, $\Gamma(\cdot)$ is the gamma function and $\varepsilon(s)$ is a stationary standard Gaussian process with correlation function $\rho(h)$. When $\nu = 1$, this model reduces to the Schlather (2002) model, with $X(s) = \sqrt{2\pi}\max[0, \varepsilon(s)]$. The Brown–Resnick model (Brown & Resnick 1977, Kabluchko et al. 2009), which arises as a limiting case when $\nu \to \infty$, sets $X(s) = \exp[\tilde{\varepsilon}(s) - \gamma(s)]$,

where $\tilde{\varepsilon}(s)$ is an intrinsically stationary process with semivariogram $\gamma(h)$, and $\tilde{\varepsilon}(0) = 0\lambda$ almost surely. All of these models are preferable to the Smith model because of their greater flexibility, their stochastic profiles, and their basis in classical geostatistics. Expressions for their exponent measures are given in articles by Opitz (2013) and Huser & Davison (2013). However, these models also have disadvantages.

First, the Schlather and extremal t models cannot capture complete independence in the isotropic case (Davison et al. 2012). In the extremal t model, independence can be reached as $\nu \to \infty$, but practical use of the Schlather model is limited unless a random set element is included (Davison & Gholamrezaee 2012, Huser & Davison 2014). Second, simulation of the Brown–Resnick and extremal t models may be burdensome (Oesting et al. 2011), although fast and exact simulation methods have recently been developed (Oesting et al. 2013, Dieker & Mikosch 2014). Third, computation of V involves the multivariate Gaussian and multivariate Student's t distribution functions, making it demanding in high dimensions. Finally, these models represent just a small parametric corner of the class of max-stable processes and may be too restrictive for some applications. Nonparametric spatial models have yet to be developed for general use (but see Buishand et al. 2008).

As mentioned in Section 3, asymptotic dependence of extremes may be unrealistic in applications, especially for small-scale phenomena such as rainfall. If the data at hand are asymptotically independent, models that explicitly specify the rate of convergence to independence may be preferred. Gaussian processes are examples of such models; their correlation is directly linked to the coefficient of tail dependence in Equation 22 through the expression $\eta(h) = [1 + \rho(h)]/2$. Wadsworth & Tawn (2012b) note that asymptotically independent processes may be obtained through the transformation $Z'(s) = -1/\log\{1 - \exp[-1/Z(s)]\}$, where $Z(s)$ is a max-stable process with unit Fréchet margins; the coefficient of tail dependence for $Z'(s)$ depends on the exponent measure of $Z(s)$. Other asymptotic independence models are discussed in an article by Wadsworth & Tawn (2012a). Wadsworth & Tawn (2012b) also propose max-mixtures of max-stable and inverted max-stable processes to model different types of asymptotic dependence at varying spatial distances, but these mixtures are highly parameterized.

4.2. Pareto Processes

Two ways of dealing with threshold exceedances in the spatial framework have been advocated. The first is to approximate the joint upper tail of the original observations using a max-stable model (recall Equation 20). This approximation is justified when all variables are large; when not all variables are large, a censored likelihood approach may be adopted (see, for example, Huser & Davison 2014). The second is to use Pareto processes (Ferreira & de Haan 2014, Thibaud & Opitz 2014), recently introduced as the process analog of the GPD. In the same vein as Equation 25, a standard Pareto process can be defined as follows:

$$Y_\infty(s) = R\tilde{X}(s), \tag{27}$$

where R has a standard Pareto distribution, i.e., $\Pr(R > r) = 1/r$ $(r \geq 1)$, and $\tilde{X}(s)$ is an independent positive process with $\sup_{s\in\mathcal{S}} \tilde{X}(s) = u_0$ almost surely for some scalar u_0. A generalized Pareto process is defined as $\mu(s) + \sigma(s)[Y_\infty(s)^{\xi(s)} - 1]/\xi(s)$ for the spatially varying location, scale, and shape parameters $\mu(s)$, $\sigma(s) > 0$, and $\xi(s)$, respectively; the case $\xi(s) = 0$ is treated as $\xi(s) \to 0$. Ferreira & de Haan (2014) have shown that these processes arise as limits as $n \to \infty$ for the rescaled process $[Y(s) - b_n(s)]/a_n(s)$, conditional on there being at least one exceedance of some threshold u. Therefore, unlike max-stable processes, they provide a valid approximation

to the upper tail when at least one variable is large. Furthermore, the Pareto process given by Equation 27 that corresponds to a max-stable process defined via Equation 25 satisfies the distributional equality

$$Y_\infty(s) \stackrel{D}{=} \left\{ \lim_{u_0 \to \infty} \frac{RX(s)}{u_0} \middle| \sup_{s \in S} RX(s) \geq u_0 \right\}, \tag{28}$$

which links the processes $X(s)$ and $\tilde{X}(s)$. Variants of Equation 27 may be used to answer different questions: For example, the condition $\sup_{s \in S} \tilde{X}(s) = u_0$ may be replaced by other measures of extremeness in terms of a risk functional ℓ, giving rise to ℓ-Pareto processes (Dombry & Ribatet 2014, Thibaud & Opitz 2014). Similar to the univariate case, the dependence structure of spatial ℓ-exceedances over high thresholds may be approximated by that of a generalized ℓ-Pareto process. One advantage of such processes appears when one compares Equations 25 and 27: The simulation of max-stable processes requires the generation of many independent replicates of $X(s)$, whereas the simulation of generalized Pareto processes requires only a single realization of $\tilde{X}(s)$. There is a close connection between the modeling of spatial maxima with max-stable processes, the modeling of exceedances with generalized Pareto processes, and the unifying approach based on point processes. Thibaud & Opitz (2014) discuss simulation and conditional simulation algorithms for generalized Pareto processes corresponding to max-stable processes based on point process considerations.

5. INFERENCE

5.1. Generalities

Misspecification is an overarching issue in the analysis of data on extremes. The extremal models described above arise as limits that are never attained in practice, so their appropriateness as approximations is always open to question. Of course this is a common issue in statistical modeling, but it is particularly important in this case because the fitted models are often intended to be used for extrapolation beyond the data. There are essentially two approaches to dealing with this issue. The first is to fit broader classes of distributions that may arise en route to the limits. The most widespread such approach presupposes that the underlying random sample $Y_1, \ldots, Y_n$ arises from a regularly varying distribution of index $-1/\xi$, i.e.,

$$1 - F(y) \sim y^{-1/\xi} L(y), \quad y \to \infty, \quad \xi > 0,$$

where the function L is slowly varying, so the limit for maxima falls into the Fréchet class. This approach is equivalent to attributing an approximate exponential distribution with mean ξ to large $\log Y_j$, and it yields the celebrated Hill (1975) estimator

$$\hat{\xi}_{\mathrm{H},k} = k^{-1} \sum_{j=1}^{k} \log[Y_{(n+1-j)}/Y_{(n-k)}], \tag{29}$$

where $Y_{(1)} \leq \cdots \leq Y_{(n)}$ denote the ordered Y_j. This estimator uses the top k sample order statistics, and the choice of k may strongly influence the estimate. A vast literature focuses on the Hill estimator and variants that are insensitive to the choice of k, reduce bias in the estimator, or estimate high quantiles of F rather than the tail index ξ [see, for example, Weissman (1978), Hall & Welsh (1985), and Gomes & Pestana (2007)]. Apart from the plethora of competing estimators,

the main difficulty with this semiparametric approach is the lack of a general paradigm for the treatment of censoring, seasonality, and other issues met in applications.

The second approach to estimation, which we discuss in the next section, is to fit limiting models directly to the data. Of course, doing so requires careful assessment of the fit of such models, as well as sensitivity analysis with a particular eye to high quantiles.

5.2. Parametric Estimation

Inference for extremes based on the parametric distributions given by Equations 1 and 5 is often based on likelihood methods, which provide a unified approach to modeling, estimation, and uncertainty assessment. The log-likelihood is easily obtained for independent data, and maximum likelihood estimates and their standard errors may be obtained numerically by, for example, using routines provided in the packages `evd`, `ismev`, or `Pot` in the statistical computing environment `R` (R Development Core Team 2014). However, the regularity conditions that underpin classical large-sample likelihood approximations hold only when $\xi > -1/2$ (Smith (1985). [Martins & Stedinger (2000) ensure that $\hat{\xi} > -1/2$ by placing a penalty on ξ.] Uncertainty concerning high quantiles of the underlying distribution, termed return levels, is typically highly asymmetric; confidence intervals for these quantiles should be based on the profile log-likelihood.

Maximum likelihood estimators are asymptotically unbiased and have the smallest possible variances in large samples, but they may be unstable in small samples. Hosking et al. (1985) and Hosking & Wallis (1987) proposed the use of probability-weighted moments estimation, which skirts the nonexistence of the ordinary moments of Equations 1 and 5 for large ξ. These estimates, and the corresponding quantile estimators, may perform better in very small samples but lack the flexibility of likelihood methods. Maximum likelihood estimators are also sensitive to model misspecification and to outliers. D.J. Dupuis has investigated the use of weighting to make these estimators more robust in a series of articles (Dupuis & Field 1998, Dupuis & Morgenthaler 2002, Dupuis 2005), and deHaan & Ferreira (2006, section 3.4) discuss the behavior of maximum likelihood estimators for the GPD under a penultimate tail model.

The incorporation of external information and borrowing strength across related time series may require a Bayesian approach; a Markov chain Monte Carlo approximation to the posterior density of the GEV parameters based on independent maxima is implemented in the `R` package `evdbayes`. Such methods are widely used in more complex problems (e.g., Cooley et al. 2007, Sang & Gelfand 2009, 2010, Reich & Shaby 2012, Shaby & Reich 2012, Reich et al. 2013).

The choice of block for analysis of sample maxima using the GEV is generally determined by the context: For example, annual blocks of environmental data avoid the need to model seasonality but result in a loss of information, so it may be useful to take monthly blocks and to model any resulting seasonality, hoping that any bias due to the use of shorter blocks is limited. The choice of threshold for use with the GPD is akin to the choice of block length under the block maximum approach and involves a similar bias–variance trade-off; threshold choice is equivalent to the choice of k for the Hill estimator. Graphical approaches are commonly used and are supplemented by goodness-of-fit considerations such as the use of quantile–quantile plots or tests of fit (e.g., Choulakian & Stephens 2001). Then u is generally chosen as the lowest threshold for which the estimates are stable and the fit seems adequate. However, graphical procedures are difficult to apply when many series must be analyzed, and more formal approaches are often preferred. A variety of ideas have been proposed. Dupuis (1998) suggested downweighting badly fitting observations using bias-robust estimators, whereas Danielsson et al. (2001) and Ferreira et al. (2003) choose the number of threshold exceedances to minimize the mean square error of parameter estimators. Other authors (e.g., Beirlant et al. 1999, Guillou & Hall 2001) have based threshold choice on second-order

approximations to the decay of the tail probability or have suggested that the relation between cluster size and threshold can be used to diagnose poor fit (Sveges & Davison 2010). Wadsworth & Tawn (2012a) propose the use of a penultimate approximation based on a nonhomogeneous Poisson process that extends Equation 4, allowing the use of likelihood ratio testing to assess at what threshold the process appears to become homogeneous and the use of Bayesian model averaging to provide robust quantile estimates. Seasonality is a nuisance for the point process approach; using a constant threshold can lead to a severe loss of information, but applying a time-varying threshold may require a two-stage approach to analysis (see Eastoe & Tawn 2009, Northrop & Jonathan 2011).

5.3. More Complex Settings

Likelihood inference for parametric models for multivariate maxima based on Equation 8 raises no new issues in principle, although in the case of numerical difficulties in the simultaneous estimation of the marginal and the dependence parameters, adopting a two-stage approach in which the former are first estimated separately may help. Doing so, however, leads to some loss of estimation efficiency.

Nonparametric estimation of the joint distribution or the Pickands function has been discussed by various authors; for example, the reader is referred to Pickands (1981) and Capéraà et al. (1997). Einmahl & Segers (2009) and de Carvalho & Davison (2014) use empirical likelihood to impose the mean constraint (Equation 14) when estimating bivariate extreme-value distributions, and this seems a promising approach that can be broadly applied. Semiparametric models that satisfy the mean constraints have been proposed and fitted using Markov chain Monte Carlo algorithms by Boldi & Davison (2007), Guillotte et al. (2011), and Sabourin & Naveau (2014).

The fitting of a Poisson process model with the measure given by Equation 17 entails the choice of a set within which the model is believed to be adequate. Because this limiting approximation will often be questionable near the coordinate axes, fitting a censored likelihood is a common practice (Coles & Tawn 1991): Taking the bivariate case for simplicity, with thresholds u_1 and u_2 for the two variables, and denoting derivatives of the exponent measure V by subscripts, the likelihood contribution from an observation (y_1, y_2) is taken to be $V_{1,2}(y_1, y_2)\exp[-V(y_1, y_2)]$ $(y_1 > u_1, y_2 > u_2)$, $V_1(y_1, u_2)\exp[-V(y_1, u_2)]$ $(y_1 > u_1, y_2 \le u_2)$, $V_2(u_1, y_2)\exp[-V(u_1, y_2)]$ $(y_1 \le u_1, y_2 > u_2)$, and $\exp[-V(u_1, u_2)]$ $(y_1 \le u_1, y_2 \le u_2)$. Thus, a variable that exceeds its threshold is regarded as observed exactly; otherwise it is treated as being censored below the threshold. Asymptotic independence models require regions such as $\{(y_1, y_2) : y_1 > u_1, y_2 > u_2\}$ that do not approach the axes. The censored likelihood approach may be extended to higher dimensions, and even to spatial and space–time models (Davison et al. 2013, Thibaud et al. 2013, Huser & Davison 2014). It may also be extended to time series exceedances, for example, using Markov chain models for the dependence of consecutive values (y_1, y_2), which may or may not exceed their common threshold (Smith et al. 1997, Bortot & Tawn 1998, Bortot & Coles 2003, Bortot & Gaetan 2014), thereby providing parametric models for clustering of exceedances. Similar ideas may be used in combination with the Heffernan–Tawn (2004) model to estimate the ratio of subasymptotic extremal indexes $\theta(y, r)/\theta(u, r)$ and thereby to provide improved estimation of high quantiles based on clustered exceedances [see Sections 2.3 and 3.3, as well as Eastoe & Tawn (2012)].

The fitting of max-stable processes to spatial or spatiotemporal block maxima is awkward because the likelihood involves all derivatives of Equation 26 up to order D and has B_D terms, where the Bell numbers B_D grow very rapidly with D: $B_2 = 2$ and $B_5 = 52$, but $B_{10} = 115,975$. As D is often large in spatial and spatiotemporal applications, computation of this likelihood is generally

infeasible, although simplification may be possible if the times of events are available (Stephenson & Tawn 2005). This potential computational explosion led Padoan et al. (2010) to suggest the use of composite likelihoods based on the marginal distributions of pairs of observations (Varin 2008). This approach, which has become widely used (e.g., Blanchet & Davison 2011, Westra & Sisson 2011, Shang et al. 2011, Davison & Gholamrezaee 2012, Davison et al. 2012), can in some cases be extended to higher-order margins (Genton et al. 2011, Huser & Davison 2013), but in addition to failing to exploit the data fully owing to the use of block maxima, the approach entails a loss of estimation efficiency. Engelke et al. (2012) and Wadsworth & Tawn (2014) overcome both difficulties by using the point process representation to construct full likelihoods based on threshold exceedances for individual events. Originally proposed for Gaussian-based models, this approach has been extended to the extremal-t case by Thibaud & Opitz (2014). Unfortunately it applies only to asymptotic dependence models, whereas composite likelihoods may also be used in situations of asymptotic independence. Similar ideas can also be used to fit Pareto processes.

The `R` packages `SpatialExtremes` and `RandomFields` may be used to fit and to simulate from max-stable processes.

6. APPLICATIONS

6.1. Introduction

We now discuss elements of risk assessment of the vulnerability of a major building to lightning and heavy rainfall to illustrate recurrent practical issues in the application of statistics of extremes. The discussion is based on a real context but is bowdlerized for reasons of confidentiality. Extrapolation to events having annual frequencies of 10^{-4} or even lower was required, based on much shorter data series. Quantitative analysis provided by the statistics of extremes is just one element of risk assessment in such contexts, but it is an important one despite the many uncertainties that surround it. These uncertainties include the following:

1. Measurement uncertainty: Measurements of equipment with well-understood behavior under average conditions may become unreliable in extreme circumstances. Such concerns are additional to issues that arise as a result of the homogenization of time series when measuring equipment is renewed or the surrounding environment changes, e.g., owing to urbanization.
2. Stochastic uncertainty: Even if a probability model were known to describe the measured phenomenon perfectly, future values would vary.
3. Estimation uncertainty: A fitted model will never perfectly match the true model, even if the correct family of models is known.
4. Model uncertainty: In practice, the family of models is chosen from the data, but another family might have been chosen if the data had been different.

The last of these sources (model uncertainty) includes misspecification uncertainty, as mentioned in Section 5.1. Moreover, because data are inevitably observed under particular physical regimes, there may be a change of regime beyond the observed data that renders irrelevant any extrapolation from a model fitted to the available data. This concern is of particular relevance here, as just a few years of data gathered in the late 20th century must be used to estimate probabilities corresponding to events that would occur, on average, once every 10,000 or more years in a stationary system.

6.2. Lightning

Data on the locations and sizes of lightning flashes for approximately 14 years were available from the Euclid collaboration (**http://www.euclid.org**) but were subject to substantial measurement

error. The locations are said to be accurate to approximately 600 m, and there are approximately 80,000 flashes within 30 km of the building; those within a radius of 200 m centered at the building are regarded as potentially hazardous. Flashes below 150 kA in absolute magnitude are thought to have errors of ±20%, whereas those above 150 kA are thought to be of unknown magnitude, although presumably their magnitudes are greater than 150 kA. Spatial analysis suggests that flashes may be treated as a spatially homogeneous Poisson process over the region considered, but they show seasonality and extreme temporal clustering. Essentially no events occur during the winter months, and flashes occur at intervals of a few milliseconds during storms, interspersed with gaps of days or even weeks. No temporal dependence can be discerned, even among near-simultaneous flashes.

In the absence of measurement error and under the assumption that observations are independent, the likelihood based on the Poisson process model for flashes $y_1, \ldots, y_{n_u}$ exceeding a threshold of u kA may be written as follows:

$$L(\mu, \sigma, \xi) = \exp\left[-C\left(1+\xi\frac{u-\mu}{\sigma}\right)_+^{-1/\xi}\right] \times \prod_{j=1}^{n_u} \frac{1}{\sigma}\left(1+\xi\frac{y_j-\mu}{\sigma}\right)_+^{-1/\xi-1}, \qquad (30)$$

where C is chosen so that the parameters μ, σ, and ξ correspond to the GEV (Equation 1) for the maximum of interest. Here the data are observed for 14 years over a radius of 30 km, whereas the target of inference is the maximum for a single year over a radius of 200 m, so we take $C = (14\pi 30^2)/(\pi 0.2^2)$.

One interpretation of the error on flash sizes is that a value recorded as y kA is uniformly distributed over the interval $(0.8y, 1.2y)$ $(y < 150)$, whereas a value $y > 150$ kA is right-censored. This implies replacing the terms in the product in Equation 30 by $\int_{0.8y_j}^{1.2y_j} \sigma^{-1}[1+\xi(x-\mu)/\sigma]_+^{-1/\xi-1}\mathrm{d}x$ or $\int_{y_j}^{\infty} \sigma^{-1}[1+\xi(x-\mu)/\sigma]_+^{-1/\xi-1}\mathrm{d}x$, depending on whether y_j is smaller or larger than 150 kA; here we suppose that if $y_j < u$, then the true value of the flash is also less than u. Refinements of this or other interpretations of the measurement error could be handled similarly. **Figure 1** shows that with this change in Equation 30, the maximum likelihood estimates of the parameters change strongly with u, although they are so uncertain that many confidence intervals

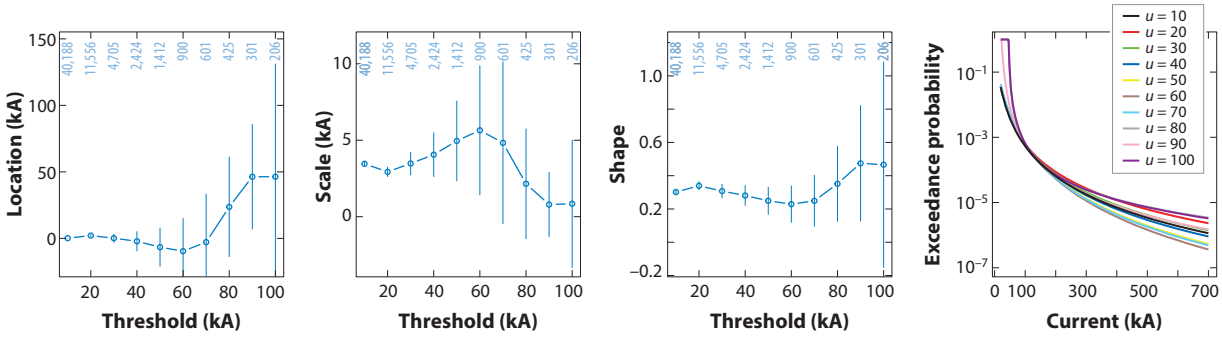


Figure 1

Lightning data analysis. The three leftmost panels show the estimated location, scale, and shape parameters for the Poisson process model with measurement error as functions of the threshold u, with approximate 95% pointwise confidence intervals. The numbers at the tops of these panels represent the numbers of observations exceeding the thresholds. The rightmost panel shows the fitted exceedance probabilities for events within 200 m of the site when $u = 10$ (*black*), 20 (*red*), 30 (*green*), 40 (*blue*), 50 (yellow), 60 (*brown*), 70 (*cyan*), 80 (*gray*), 90 (*pink*), and 100 (*purple*) kA.

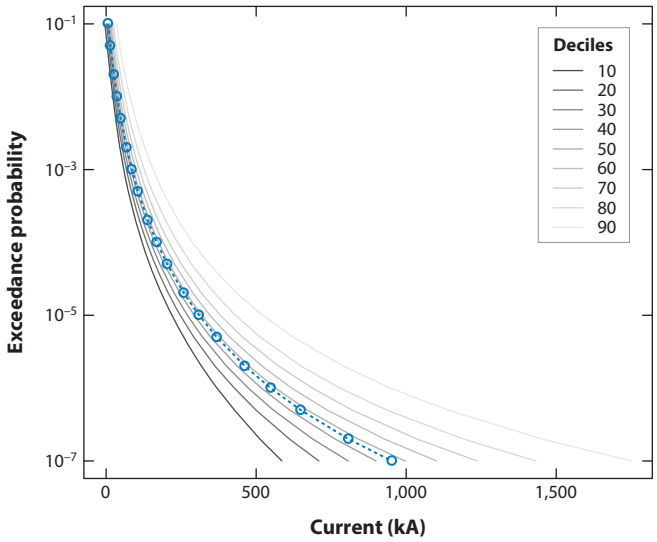


Figure 2

Lightning data analysis. For a threshold at 60 kA, the plot shows the estimated distributions of future lightning strikes within 200 m of the building site, with exceedance probabilities ranging from 10^{-1} (once every 10 years) to 10^{-7} (once every 10 million years). The distribution for 10^{-4} (once every 10 thousand years) was needed for the application. The blue line shows the maximum likelihood estimates of the corresponding quantiles, and the gray lines show deciles of the estimated predictive distributions.

overlap. The estimated shape parameter is quite strongly positive, resulting in large and rather unstable estimates of the upper quantiles.

Although risk estimation is often based on extreme quantiles, these quantiles do not account for stochastic variation. Such variation is fairly easily accommodated within a Bayesian framework by using the posterior predictive distribution of the largest event expected over some future period of m years, conditional on the observed data. Unlike the corresponding point estimate of the m-year return level, or equivalently the $1 - 1/m$ quantile of the annual maximum distribution, such a predictive distribution gives some idea of the likely variation in the future event. A crude frequentist analog may be provided by simulating many values from the asymptotic density of the parameter estimates and basing a prediction on each such parameter value. An example is provided by **Figure 2**, which shows both the rapid increase of the quantiles as the exceedance probability decreases, owing to the positive shape parameter, and very large uncertainty for the sizes of future strikes. A fully Bayesian analysis would in some ways be more satisfactory, but as in many such applications, difficulties in agreeing on a suitable formulation of prior information meant that the analysis described here was more broadly acceptable to the stakeholders concerned.

This cursory account illustrates both the difficulties of extreme tail estimation, even based on an apparently large sample, and the flexibility of the parametric approach to inference, which makes managing the idiosyncrasies of the data relatively straightforward.

6.3. Rainfall

Rainfall data, available from late 1988 to the end of February 2013, comprise totals over 10-minute intervals, and these data are aggregated to obtain totals over overlapping 30-minute, 60-minute, and 120-minute intervals. Because the data for all aggregations show marked seasonality and extremes appear mainly during the summer months, only data from May–September were retained

for the extremal analyses. Furthermore, to avoid having to choose a time-varying threshold above which the GPD (Equation 5) might be fitted, we fit the GEV (Equation 1) with time-varying parameters to the monthly maxima; this approach uses five times more observations than would fitting to annual maxima, and it should reduce uncertainty somewhat. To account for seasonality and possible trend, we suppose that the location parameter satisfies the following equation:

$$\mu(t) = \mu_0 + \mu_1(t - t_0) + \sum_{j=1}^{L} [c_j \cos(2\pi jt/365) + s_j \sin(2\pi jt/365)], \tag{31}$$

where t_0 is a reference date and c_j, s_j $(j = 1, \ldots, L)$ are the amplitudes of the harmonic terms. We make similar suppositions for the scale parameter $\sigma(t) > 0$ and the shape parameter $\xi(t)$. The simplest approach is to fit these models separately to the data for each different aggregation interval, but this approach has the unsatisfactory consequence that extremes forecast for longer intervals may be smaller than those for shorter intervals if the corresponding shape parameters differ. We therefore fit these models simultaneously to the data at all levels of aggregation, imposing a common shape parameter $\xi(t)$ by maximizing an independence likelihood (Chandler & Bate 2007), bootstrapping months to obtain standard errors, and comparing models using the composite likelihood information criterion (CLIC) (Varin 2008). The parameter estimates based on these fits are very similar to those for separate fits, but extrapolations to high quantiles are coherent when using this simultaneous fitting approach. Following Coles & Tawn (1996), Ferreira et al. (2012) give theoretical support for the use of a constant shape parameter for extremes at different aggregations. The values of the CLIC suggest that a good overall model has constant shape and scale parameters over time, no time trend, and $L = 1$ in Equation 31. This model for $\xi(t)$ is fortunate because the shape parameter is difficult to estimate, and including seasonality in it would greatly increase estimation uncertainty. The estimates of μ_0 and σ increase with the length of the aggregation interval, thus ensuring that, as a common shape parameter is used for all intervals, the quantiles are coherent. The shape parameter estimate is $\hat{\xi} = 0.14$ and significantly positive, implying that summer rainfall is heavy tailed. Fits to data from nearby sites show broad spatial coherence. **Figure 3** compares quantiles for this site, S1, and another site for which $\hat{\xi} = 0.26$, and shows that a small increase in $\hat{\xi}$ can have a dramatic effect on high quantiles.

Annual return levels z_p were estimated by numerical solution of the equation

$$\prod_{j=May}^{September} \hat{G}_j(z) = 1 - 1/p, \tag{32}$$

where $\hat{G}_j(z)$ denotes the GEV fit for month j, and their uncertainties were assessed by the posterior predictive approach outlined in Section 6.2, which attempts to pool stochastic and estimation uncertainties. More precisely, sets of parameters were simulated from their estimated asymptotic distributions, then used in Equation 32 to express the global uncertainty that surrounds z_p. **Figure 3** is instructive: The higher shape parameter at site S2 not only affects the return level estimates, but also their variability. In addition, confidence bands for site S2 are dramatically wider than those for site S1: The larger the estimated shape parameter, the more variable is the estimation of return levels and the less reliable is the extrapolation. A full Bayesian analysis would allow the inclusion of other sources of uncertainty, but such an analysis was regarded as unacceptable for legal reasons.

It is natural to wonder whether estimation can be improved by using data from nearby locations. To investigate this, we considered two additional locations, sites S3 and S4, and fitted bivariate extreme-value models independently to all distinct pairs of locations (Si,Sj) and all time intervals. We first transformed the monthly maxima to the unit Fréchet scale using the fitted

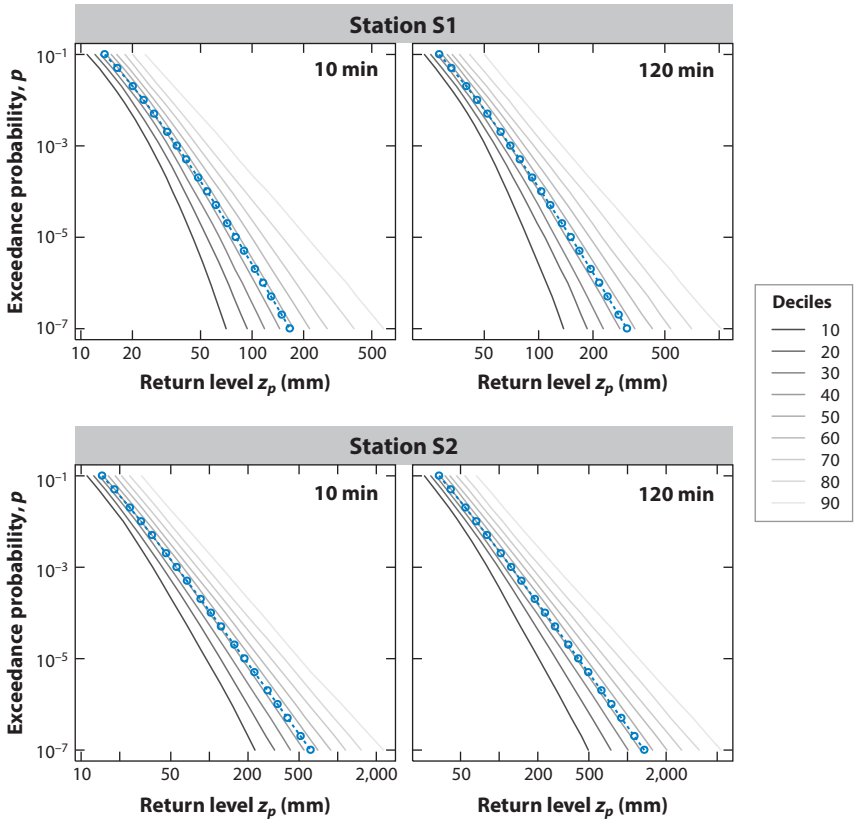


Figure 3

Estimated quantiles (*blue dots*) and predictive distributions (*gray curves*) for annual maxima of (*a*) 10-minute and (*b*) 120-minute rainfall (mm) at stations S1 (*top*) and S2 (*bottom*). Quantiles are displayed on logarithmic scales.

GEV marginals, and we then fitted the distribution given by Equation 8 with the Hüsler & Reiss (1989) exponent measure

$$V(z_1, z_2) = \frac{1}{z_1}\Phi\left[\frac{\sqrt{2\gamma}}{2} - \frac{1}{\sqrt{2\gamma}}\log\left(\frac{z_1}{z_2}\right)\right] + \frac{1}{z_2}\Phi\left[\frac{\sqrt{2\gamma}}{2} - \frac{1}{\sqrt{2\gamma}}\log\left(\frac{z_2}{z_1}\right)\right]. \qquad (33)$$

In Equation 33, $\Phi(\cdot)$ is the normal cumulative distribution function and $\gamma \geq 0$ is a dependence parameter; increasing values of γ indicate weakening dependence. Estimated parameters are plotted against the distance separating the two locations in the first row of **Figure 4**. For each time interval, we also fitted the Brown–Resnick model (recall Section 4.1), which is a spatial extension of the Hüsler–Reiss distribution, and overlaid the estimated semivariogram $\hat{\gamma}(h)$.

Extremal dependence seems to be slightly stronger at short distances for 30-minute intervals, but this variation with distance is less obvious for the other intervals, for which the strength of dependence is more or less constant with distance and rather weak. This lack of a coherent pattern may suggest that marginal fits are poor, making the data appear more independent than they should be, or it may suggest simply that extreme rainfall is a small-scale phenomenon for which an asymptotic independence process might be preferable.

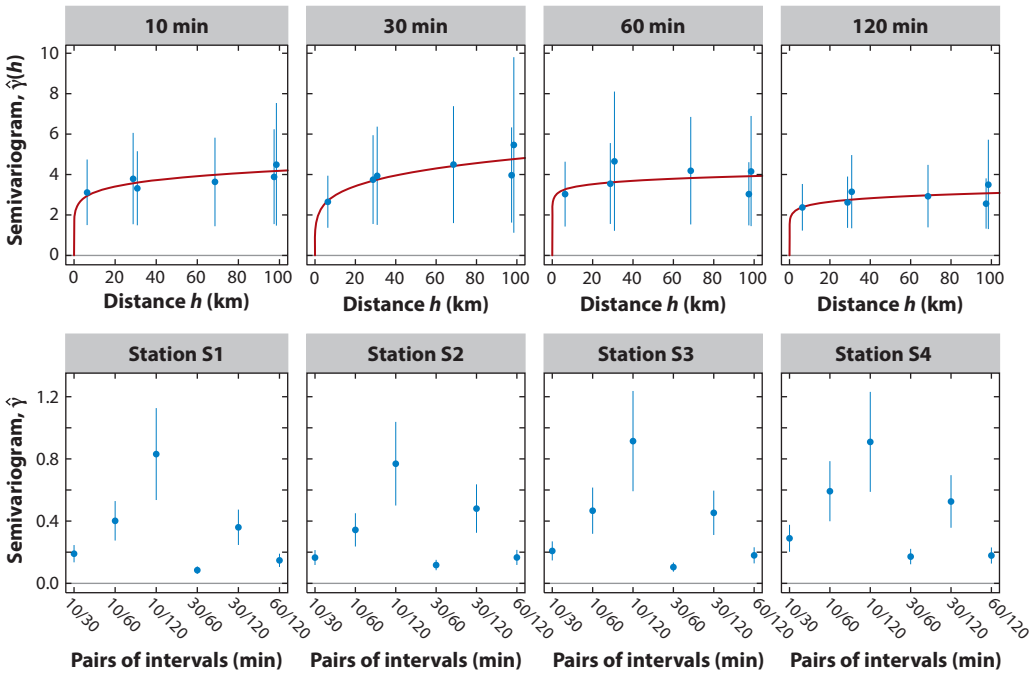


Figure 4

Estimated dependence parameters $\hat{\gamma}$ (*blue dots*), with 95% confidence intervals (*vertical segments*) stemming from the fit of the Hüsler–Reiss model (Equation 33) to (*top row*) all pairs of stations for the different time intervals (*columns*) and (*bottom row*) all pairs of time intervals for the different stations (*columns*). The solid red curves are the estimated semivariograms $\hat{\gamma}(h)$ corresponding to the fit of the Brown–Resnick process.

We also fitted the bivariate Hüsler–Reiss model to all pairs of time intervals for each station. The estimated dependence parameters are reported in the second row of **Figure 4**. Not surprisingly, extremal dependence appears to be quite strong overall, and systematic patterns exist: For example, the pair of 10–30-minute time intervals appears to be more strongly dependent than do the 10–60-minute and 10–120-minute pairs. In general, dependence for the pair of time intervals $I_1 - I_2$ increases as the ratio I_1/I_2 approaches unity and as $\min(I_1, I_2)$ gets smaller. This dependence structure suggests that the joint behavior for all intervals may be described by a more complex multivariate model with fewer parameters overall and that estimation precision may therefore be gained. Building multivariate max-stable processes is, however, not an easy task, and thus far, very few attempts have been made to tackle this challenging issue. Moreover, data from many more nearby sites would be needed for realistic spatial and spatiotemporal modeling.

7. DISCUSSION

The Literature Cited section shows that the statistics of extremes is a highly active research field, and in the space available, mentioning all current developments is impossible. In addition to long-standing topics such as threshold choice for single samples, the demands of applications suggest many avenues for future exploration. Among these are the identification of extremes in settings that have many parallel time series and seasonality, allied to appropriate treatment of possibly unreliable data. A particular application of interest is climate model output. In such cases, robust inference may be useful, although it may be difficult to distinguish outliers, which

should be downweighted, from extreme observations, which may be highly informative. The area of experimental design for computer models is currently an active one, although very little discussion of design is specifically oriented to extremes (but see Heffernan & Tawn 2001, 2003). Other important topics that have seen little development so far are the joint modeling of extremes of several variables in the spatial and space–time settings, taking into account different decay rates for tail dependence, and nonstationarity in the dependence structure. Moreover, there has been little systematic discussion of nonparametric and semiparametric inference for high-dimensional data or for complex applications, although the curse of dimensionality may make progress in these areas difficult. Finally, a perennial issue is the injection of subject-matter information: The models described in this article stem from mathematical and statistical considerations, but the incorporation of well-understood physical models of underlying phenomena may be valuable or even essential.

SUMMARY POINTS

1. Extreme Value Theory and statistics may be used for extrapolation outside the range of an observed sample, in order to estimate probabilities of rare events, for example, for risk assessment.
2. Use of standard statistical techniques based on the Gaussian distribution can lead to dramatic underestimation of rare event probabilities.
3. Often, data are limited in relation to the rarity of the events of interest, so inferences about the extremes typically have large uncertainty.
4. Max-stable distributions and processes provided approximate probability models for maxima and minima of blocks of random observations, and generalized Pareto distributions and processes provide approximate probability models for rare threshold exceedances.
5. The stochastic properties of extreme events, both block maxima and threshold exceedances, can be characterized and unified using point process theory.
6. Asymptotic independence is a degenerate case in the max-stable paradigm, and it needs to be addressed using alternative models and techniques.
7. Extremal models can provide approximate probabilities for rare events, under broad mathematical conditions, but the adequacy of the models and the effects of possible departures from those conditions will need careful consideration in any given application.
8. Likelihood-based approaches to inference are commonly used because of their flexibility and their efficiency, although moment and nonparametric estimators are also useful.

FUTURE ISSUES

1. Modeling and inference for extremes observed at many locations, such as in the case of climate model output.
2. Modeling complex spatially heterogeneous processes at extreme levels.
3. Modeling of temporal nonstationarity resulting from factors such as seasonality and climate change.

4. Incorporation of substantive knowledge, for example, from physical processes, into probability models.
5. Bayesian modeling of extremal processes.
6. Flexible joint modeling of both nonextreme data, i.e., the bulk of the distribution, and extreme data.
7. Development of flexible models that encompass both max-stable and asymptotic independence cases.
8. Development of methods for detecting rare events in large, highly multivariate, data sets.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation. We thank David Morton de Lachapelle for help with the graphics, and we thank Subhajit Dutta and Nancy Reid for comments on a previous version. Parts of this article are based on Huser (2013).

LITERATURE CITED

Anderson CW. 1971. *Contributions to the Asymptotic Theory of Extreme Values*. PhD Thesis, University of London, London, UK

Anderson CW, Coles SG, Hüsler J. 1997. Maxima of Poisson-like variables and related triangular arrays. *Ann. Appl. Probab.* 7:953–71

Ballani F, Schlather M. 2011. A construction principle for multivariate extreme value distributions. *Biometrika* 98:633–45

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, Teugels J, Segers J. 2004. *Statistics of Extremes: Theory and Applications*. New York: Wiley

Blanchet J, Davison AC. 2011. Spatial modelling of extreme snow depth. *Ann. Appl. Stat.* 5:1699–725

Boldi M-O, Davison AC. 2007. A mixture model for multivariate extremes. *J. R. Stat. Soc. B* 69:217–29

Bortot P, Coles SG. 2003. Extremes of Markov chains with tail-switching potential. *J. R. Stat. Soc. B* 65:851–67

Bortot P, Gaetan C. 2014. A latent process model for temporal extremes. *Scand. J. Stat.* 41:606–21

Bortot P, Tawn JA. 1998. Models for the extremes of Markov chains. *Biometrika* 85:851–67

Brown BM, Resnick SI. 1977. Extreme values of independent stochastic processes. *J. Appl. Probab.* 14:732–39

Buishand TA, de Haan L, Zhou C. 2008. On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.* 2:624–42

Butler A, Heffernan JE, Tawn JA, Flather RA. 2007. Trend estimation in extremes of synthetic North Sea surges. *Appl. Stat.* 56:395–414

Capéraà P, Fougères AL, Genest C. 1997. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84:567–77

Chandler RE, Bate S. 2007. Inference for clustered data using the independence loglikelihood. *Biometrika* 94:167–83

Choulakian V, Stephens MA. 2001. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* 43:478–84

Coles SG. 2001. *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer
Coles SG, Heffernan J, Tawn JA. 1999. Dependence measures for extreme value analyses. *Extremes* 2(4):339–65
Coles SG, Tawn JA. 1991. Modelling extreme multivariate events. *J. R. Stat. Soc. B* 53:377–92
Coles SG, Tawn JA. 1996. Modelling extremes of the areal rainfall process. *J. R. Stat. Soc. B* 58:329–47
Cooley D, Davis RA, Naveau P. 2010. The pairwise beta distribution: a flexible parametric multivariate model for extremes. *J. Multivar. Anal.* 101:2103–17
Cooley D, Nychka D, Naveau P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* 102:824–40
Danielsson J, de Haan L, Peng L, de Vries CG. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivar. Anal.* 76:226–48
Das B, Resnick SI. 2011. Conditioning on an extreme component: model consistency with regular variation on cones. *Bernoulli* 17(1):226–52
Davison AC, Gholamrezaee MM. 2012. Geostatistics of extremes. *Proc. R. Soc. Lond. A* 468:581–608
Davison AC, Huser R, Thibaud E. 2013. Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.* 45:511–29
Davison AC, Padoan SA, Ribatet M. 2012. Statistical modelling of spatial extremes. *Stat. Sci.* 27:161–86; discussion pp. 187–201
Davison AC, Smith RL. 1990. Models for exceedances over high thresholds. *J. R. Stat. Soc. B* 52:393–425; discussion pp. 425–42
de Carvalho MB, Davison AC. 2014. Spectral density ratio models for multivariate extremes. *J. Am. Stat. Assoc.* 109:764–76
de Haan L. 1984. A spectral representation for max-stable processes. *Ann. Probab.* 12:1194–204
de Haan L, Ferreira A. 2006. *Extreme Value Theory: An Introduction*. New York: Springer
Dieker AB, Mikosch T. 2014. Exact simulation of Brown–Resnick random fields. arXiv:1406.5624v1 [math.PR]
Dombry C, Ribatet M. 2014. Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface*. In press
Dupuis DJ. 1998. Exceedances over high thresholds: a guide to threshold selection. *Extremes* 1:251–61
Dupuis DJ. 2005. Ozone concentrations: a robust analysis of multivariate extremes. *Technometrics* 47:191–201
Dupuis DJ, Field CA. 1998. Robust estimation of extremes. *Can. J. Stat.* 26:199–215
Dupuis DJ, Morgenthaler S. 2002. Robust weighted likelihood estimators with an application to bivariate extreme value problems. *Can. J. Stat.* 30:17–36
Eastoe EF, Tawn JA. 2009. Modelling non-stationary extremes with application to surface level ozone. *Appl. Stat.* 58:25–45
Eastoe EF, Tawn JA. 2012. Modelling the distribution for the cluster maxima of exceedances of sub-asymptotic thresholds. *Biometrika* to appear
Einmahl JHJ, Li J, Liu RY. 2009. Thresholding events of extreme in simultaneous monitoring of multiple risks. *J. Am. Stat. Assoc.* 104(487):982–92
Einmahl JHJ, Segers J. 2009. Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Stat.* 37:2953–89
Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer
Engelke S, Malinowski A, Kabluchko Z, Schlather M. 2014. Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. B.* In press
Fawcett L, Walshaw D. 2007. Improved estimation for temporally clustered extremes. *Environmetrics* 18:173–88
Fawcett L, Walshaw D. 2012. Estimating return levels from serially dependent extremes. *Environmetrics* 23:272–83
Ferreira A, de Haan L, Peng L. 2003. On optimising the estimation of high quantiles of a probability distribution. *Statistics* 37:401–34
Ferreira A, de Haan L, Zhou C. 2012. Exceedance probability of the integral of a stochastic process. *J. Multivar. Anal.* 105:241–57
Ferreira A, de Haan L. 2014. The generalized Pareto process; with a view towards application and simulation. *Bernoulli*. In press

Fisher RA, Tippett LHC. 1928. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Math. Proc. Camb. Philos. Soc.* 24:180–90

Fréchet M. 1927. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polonaise Math.* 6:93–116

Galambos J. 1978. *The Asymptotic Theory of Extreme Order Statistics*. New York: Wiley

Genton MG, Ma Y, Sang H. 2011. On the likelihood function of Gaussian max-stable processes. *Biometrika* 98:481–88

Gnedenko BV. 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44:423–53

Gomes MI, de Haan L. 1999. Approximation by penultimate extreme value distributions. *Extremes* 2:71–85

Gomes MI, Pestana D. 2007. A sturdy reduced-bias extreme quantile (VaR) estimator. *J. Am. Stat. Assoc.* 102:280–92

Guillotte S, Perron F, Segers J. 2011. Nonparametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. B* 73:377–406

Guillou A, Hall PG. 2001. A diagnostic for selecting the threshold in extreme-value analysis. *J. R. Stat. Soc. B* 63:293–305

Gumbel EJ. 1958. *Statistics of Extremes*. New York: Columbia Univ. Press

Hall PG, Welsh AH. 1985. Adaptive estimates of parameters of regular variation. *Ann. Stat.* 13:331–41

Heffernan JE, Resnick SI. 2005. Hidden regular variation and the rank transform. *Adv. Appl. Probab.* 37:393–412

Heffernan JE, Resnick SI. 2007. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* 17:537–71

Heffernan JE, Tawn JA. 2001. Extreme value analysis of a large designed experiment: a case study in bulk carrier safety. *Extremes* 4:359–78

Heffernan JE, Tawn JA. 2003. An extreme value analysis for the investigation into the sinking of the M. V. Derbyshire. *Appl. Stat.* 52:337–54

Heffernan JE, Tawn JA. 2004. A conditional approach for multivariate extreme values. *J. R. Stat. Soc. B* 66:497–530; discussion pp. 530–46

Hill BM. 1975. A simple general approach to inference about the tail of a distribution. *Ann. Stat.* 3:1163–74

Hosking JRM, Wallis JR. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29:339–49

Hosking JRM, Wallis JR, Wood EF. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27:251–61

Hsing T. 1987. On the characterization of certain point processes. *Stoch. Process. Appl.* 26:297–316

Hsing T, Hüsler J, Leadbetter MR. 1988. On the exceedance point process for a stationary sequence. *Probab. Theory Rel. Fields* 78:97–112

Huser R. 2013. *Statistical Modeling and Inference for Spatio-Temporal Extremes*. PhD Thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Huser R, Davison AC. 2013. Composite likelihood estimation for the Brown–Resnick process. *Biometrika* 100:511–18

Huser R, Davison AC. 2014. Space-time modelling of extreme events. *J. R. Stat. Soc. B* 76:439–61

Hüsler J, Reiss R. 1989. Maxima of normal random vectors: between independence and complete dependence. *Stat. Probab. Lett.* 7:283–86

Kabluchko Z, Schlather M, de Haan L. 2009. Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* 37:2042–65

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

Keef C, Papastatopoulos I, Tawn JA. 2013. Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Multivar. Anal.* 115:396–404

Keef C, Tawn J, Svensson C. 2009. Spatial risk assessment for extreme river flows. *J. R. Stat. Soc. C* 58:601–18

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

Leadbetter MR. 1991. On a basis for 'Peaks over Threshold' modelling. *Stat. Probab. Lett.* 12:357–62

Leadbetter MR, Lindgren G, Rootzén H. 1983. *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer

Ledford AW, Tawn JA. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83:169–87

Ledford AW, Tawn JA. 1997. Modelling dependence within joint tail regions. *J. R. Stat. Soc. B* 59:475–99

Ledford AW, Tawn JA. 2003. Diagnostics for dependence within time series extremes. *J. R. Stat. Soc. B* 65:521–43

Liu Y, Tawn JA. 2014. Self-consistent estimation of conditional multivariate extreme distributions. *J. Multivar. Anal.* 127:19–35

Martins E, Stedinger J. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* 36:737–44

Nadarajah S, Mitov K. 2002. Asymptotics of maxima of discrete random variables. *Extremes* 5:287–94

Nikoloulopoulos AK, Joe H, Li H. 2009. Extreme value properties of multivariate *t*copulas. *Extremes* 12:129–48

Northrop PJ, Jonathan P. 2011. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics* 22:799–809; discussion pp. 810–16

Oesting M, Kabluchlo Z, Schlather M. 2011. Simulation of Brown–Resnick processes. *Extremes* 15:89–107

Oesting M, Schlather M, Zhou C. 2013. On the normalized spectral representation of max-stable processes on a compact set. arXiv:1310.1813 [math.PR]

Opitz T. 2013. Extremal *t* processes: elliptical domain of attraction and a spectral representation. *J. Multivar. Anal.* 122:409–13

Padoan SA, Ribatet M, Sisson SA. 2010. Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.* 105:263–77

Pickands J. 1981. Multivariate extreme value distributions. *Bull. Int. Stat. Inst.* 49:859–78

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna: R Found. Stat. Comput. **http:// www.R-project.org**

Ramos A, Ledford AW. 2009. A new class of models for bivariate joint extremes. *J. R. Stat. Soc. B* 71:219–41

Ramos A, Ledford AW. 2011. An alternative point process framework for modeling multivariate extreme values. *Commun. Stat. Theory Methods* 40:2205–24

Reich BJ, Shaby BA. 2012. A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* 6:1430–51

Reich BJ, Shaby BA, Cooley D. 2013. A hierarchical model for serially dependent extremes: a study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.* 19:119–35

Resnick SI. 1987. *Extreme Values, Regular Variation, and Point Processes*. New York: Springer

Resnick SI. 2002. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* 5:303–36

Resnick SI. 2006. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer

Rootzen H, Tajvidi N. 2006. Multivariate generalized Pareto distributions. *Bernoulli* 12:917–30

Sabourin A, Naveau P. 2014. Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Comput. Stat. Data Anal.* 71:542–67

Sang H, Gelfand AE. 2009. Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* 16:407–26

Sang H, Gelfand AE. 2010. Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* 15:49–65

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

Shaby BA, Reich BJ. 2012. Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* 23:638–48

Shang H, Yan J, Zhang X. 2011. El Nino–Southern Oscillation influence on winter maximum daily precipitation in California in a spatial model. *Water Resour. Res.* 47:W11507

Smith RL. 1985. Maximum likelihood estimation in a class of non-regular cases. *Biometrika* 72:67–92

Smith RL. 1990. *Max-stable processes and spatial extremes*. Work. Pap., Dept. Math., Univ. Surrey, Guildford, UK. **http://www.stat.unc.edu/postscript/rs/spatex.pdf**

Smith RL, Tawn JA, Coles SG. 1997. Markov chain models for threshold exceedances. *Biometrika* 84:249–68

Stephenson AG, Tawn JA. 2005. Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* 92:213–27

Süveges M, Davison AC. 2010. Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.* 4:203–21

Taleb NT. 2007. *The Black Swan*. New York: Random House.

Thibaud E, Mutzner R, Davison AC. 2013. Threshold modeling of extreme spatial rainfall. *Water Resour. Res.* 49:4633–44

Thibaud E, Opitz T. 2014. Efficient inference and simulation for elliptical Pareto processes. *Biometrika* Under review

von Mises R. 1936. La distribution de la plus grande de *n* valeurs. *Rev. Math. l'Union Interbalk.* (*Athens*) 1:141–60

Varin C. 2008. On composite marginal likelihoods. *Adv. Stat. Anal.* 92(1):1–28

Wadsworth JL, Tawn JA. 2012a. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. R. Stat. Soc. B* 74:543–67

Wadsworth JL, Tawn JA. 2012b. Dependence modelling for spatial extremes. *Biometrika* 99:253–72

Wadsworth JL, Tawn JA. 2013. A new representation for multivariate tail probabilities. *Bernoulli* 19:2689–714

Wadsworth JL, Tawn JA. 2014. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* 101:1–15

Weissman I. 1978. Estimation of parameters and large quantiles based on the *k* largest observations. *J. Am. Stat. Assoc.* 73:812–15

Weller GB, Cooley D. 2014. A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika* 101:17–36

Westra S, Sisson SA. 2011. Detection of non-stationarity in precipitation extremes using a max-stable process model. *J. Hydrol.* 406:119–28