NISS

Multivariate Threshold Methods

Richard L. Smith

Technical Report Number 7
December, 1993

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

To appear in: *Extreme Value Theory and Applications*, J. Galambos (ed.), Kluwer Academic Publishers, 1994.

MULTIVARIATE THRESHOLD METHODS

Richard L. Smith (University of North Carolina)

1. Threshold methods

In this paper, I review a variety of approaches to the estimation of extremal properties of a probability distribution or stochastic process. Statistical methods based on exceedances over a high threshold have gained in popularity in recent years, as compared with the much older methods based directly on the extreme value distributions. Nevertheless, there remain some critical questions about their application.

The simplest approach is based on a sequence of i.i.d. observations $X_1, ..., X_n$, from an unknown distribution function F, in which, to confine attention to the tail, we model only the exceedances over a fixed high threshold u. That is, our information consists of N_n, the number of exceedances in n trials over the threshold u, and the excesses $Y_1, ..., Y_{N_n}$, where if the i'th exceedance occurred on the j'th trial of the original sample, the i'th excess is defined by $Y_i = X_j - u$. It is natural to formulate the problem in terms of $\lambda = 1 - F(u)$, the probability of an exceedance over the threshold, and the conditional distribution function of the excesses,

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}. \tag{1.1}$$

It is obvious how to estimate λ (by the point estimate $\hat{\lambda} = N_n/n$ with associated confidence intervals etc.), but not at all obvious how we should deal with (1.1). Davison and Smith (1990), following earlier papers by Davison (1984), Smith (1984) and others leading back to Pickands

(1975), suggested that an appropriate model for this is the *Generalized Pareto distribution* (GPD) defined by

$$G(y;\sigma,\xi) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)_+^{-1/\xi}, \quad y > 0, \tag{1.2}$$

where $\sigma > 0$, ξ is any real number, and $x_+ = \max(x,0)$. Thus the range of y is $0 < y < \infty$ for $\xi \geq 0$ and $0 < y < -\sigma/\xi$ if $\xi < 0$. The exponential distribution, $1 - e^{-y/\sigma}$, arises naturally as a limiting case when $\xi \to 0$.

The motivation for the Generalized Pareto distribution is based on Pickands' theorem that

$$\inf_{\xi} \lim_{u \uparrow \omega_F} \inf_{\sigma} \sup_{y>0} |F_u(y) - G(y;\sigma,\xi)| = 0$$

if and only if F is in the domain of attraction of one of the standard extreme value distributions. Here $\omega_F = \sup\{x : F(x) < 1\}$ is the right-hand endpoint of F, which may be finite or infinite. In words, there exists a ξ (which does not depend on u) and a σ (which does) such that F_u is closely approximated by $G(\cdot;\sigma,\xi)$ whenever u is sufficiently close to the right-hand endpoint.

If we ignore the approximation and assume F_u is exactly GPD, then estimation by numerical maximum likelihood (MLE) is straightforward in nearly all practical cases. The asymptotic properties are regular whenever $\xi > -\frac{1}{2}$, and alternative remedies are available for $\xi \leq -\frac{1}{2}$. These results follow from general properties of nonregular estimation established by Smith (1985). One serious competitor to maximum likelihood, apart from Bayesian methods, is the probability weighted moments (PWM) method of Hosking and Wallis (1987). The PWM method was earlier developed for the three-parameter generalized extreme value distribution (Hosking, Wallis and Wood, 1985), where it was shown by simulation to be in general more efficient than MLE for the central range of ξ ($-0.2 \leq \xi \leq 0.2$) and moderate values of n (up to 100). However, PWMs are much less flexible than MLEs as a general estimation method — the reader who doubts this is invited to re-cast all the results of this paper in terms of PWMs! Pickands himself proposed a direct method of estimation based on quantiles, and de Haan and co-workers developed methods based on moments (Dekkers and de Haan 1989, Dekkers *et al.* 1989, de Haan and Rootzén 1993), but these also lack the flexibility of MLE to handle more complicated situations.

In practice, of course, we want to apply this method in situations much more complicated than i.i.d. observations, for example, dependent data and situations depending on covariates. Major themes are the following.

(i) Modeling of covariates: The idea here is that threshold exceedances may depend on other measured variables that can be incorporated into the analysis. For example, tropospheric (ground-level) ozone, a common cause on environmental health concern in large cities, is hugely affected by meteorological factors: the worst ozone days occur in hot weather when there is low wind. It is natural to construct a regression model in which the parameters λ, σ and (possibly) ξ are functions of the covariates, through additional parameters which are estimated by MLE. A direct method of doing this was described by Davison and Smith (1990). An alternative approach, based on nonhomogeneous Poisson processes, was developed by Smith (1989) and is outlined below.

(ii) Dependence: Most environmental time series are not independent but exhibit serial correlation. An informal and somewhat historical approach to this is to group the data into clusters of dependent exceedances and to apply the GPD to the maximum within each cluster. A more rigorous justification for such a procedure is based on the theory of extremes in dependent stationary processes (Leadbetter *et al.* 1983, Hsing 1987, Hsing *et al.* 1988, Leadbetter 1991) which shows how clustering naturally arises in the limiting distribution of the point process of threshold exceedance times under a simultaneous rescaling of time and space.

Nevertheless, although the idea of clustering is intuitive, there are a number of approaches to identifying clusters in practice. This problem is closely tied up with *estimating the extremal index*, for which a growing literature now exists (Leadbetter *et al.* 1989, Nandagopalan 1990, Hsing 1991, Smith and Weissman 1994). The extremal index, an important parameter in its own right (Leadbetter *et al.* 1983, Leadbetter 1983) may be defined in the present context as the reciprocal of the mean cluster size in the limiting point process. It is estimated by first identifying clusters of neighboring exceedances, and then calculating the mean cluster size. (This assumes that we do not have in mind any specific model for dependence, since if we did, we might prefer a model-based approach. See section 4 for specific proposals in this case.) From this point of view, the problems of identifying clusters and estimating the extremal index are the same.

A number of procedures for identifying clusters exist. The simplest is the *blocks approach* (Leadbetter *et al.* 1989, Hsing 1991) in which the data set is divided up into blocks of consecutive observations, all the exceedances within a single block being defined to form a single cluster. An alternative is the *runs approach* in which two consecutive exceedances are defined to be in the same cluster if and only if they are less than r_n observations apart. Here r_n is a parameter to be determined. Smith and Weissman (1994) give a theoretical argument for preferring the runs approach to the blocks approach and suggest a procedure for determining r_n. In the majority of practical cases, the results are not sensitive to the precise choice of r_n, which may be based on intuitive reasoning of how long one would expect the clusters to last, and consequently such formal procedures are not necessary.

(iii) Seasonality: A common feature of environmental data is that the series is not stationary, but exhibits a strong annual variability and possibly other periodic effects. (For instance, the tidal record is often decomposed into over 100 harmonic components.) Davison and Smith (1990) discussed two broad approaches to dealing with seasonality, the "prewhitening" approach in which the raw data are deseasonalized before applying the threshold, and the "separate seasons" (or blocking) approach in which the year is divided up into blocks with a homogeneous model being assumed for each block. The main example in Smith (1989) also contains an extensive discussion of the blocking feature. The disadvantages of prewhitening are, (a) the method is unsound unless there are strong grounds, which would almost certainly have to be physical rather than statistical, for believing that the prewhitened data really are strictly stationary even in the tails, (b) it can be hard to interpret a threshold analysis of the prewhitened data as an extreme value analysis of the original data. The main disadvantage of the blocks method is that the data are unlikely to be exactly homogeneous across blocks, unless one uses very short blocks, in which case the problem becomes overparametrized. An alternative approach is to treat time of year as another covariate and use a regression approach. This is what I actually do in Section 5 below.

(iv) Selecting the threshold: Another practical problem in these methods is how to decide what threshold u to use. Although theoretical results on this question are available (Smith 1987, Dekkers and de Haan 1989, Dekkers *et al.* 1989), they are rather hard to use in practice be-

cause the theoretical results themselves depend on additional parameters which are unknown. A practical technique is to use the *mean residual life* plot (Davison and Smith 1990, Smith 1990). The motivation for this is the easily checked formula that for $Y \sim G(\cdot;\sigma,\xi)$ and $v > 0$ (assuming $v < -\sigma/\xi$ in the case $\xi < 0$),

$$\mathrm{E}\{Y - v | Y > v\} = \frac{\sigma + \xi v}{1 - \xi}. \tag{1.3}$$

Therefore, an empirical plot of $\mathrm{E}\{Y - v|Y > v\}$ against v, should be approximately a straight line. In practice, with data $X_1, ..., X_n$, we plot

$$\frac{\sum (X_i - u) I(X_i > u)}{\sum I(X_i > u)} \tag{1.4}$$

against u, and look for the smallest u over the region in which this is a straight line. The left hand side of (1.3), or its empirical counterpart in (1.4), is known as the mean residual life in the theory of survival data, hence the name of the plot. This at least gives some empirical support to the practical choice of u, though it requires some care because the plot becomes very irregular as u approaches the upper boundary of the data. Again, from a practical point of view, the problem is often not so serious because the choice of u is either suggested on physical grounds (for example, with tropospheric ozone it is known that different chemical processes affect the ozone levels above about 80 ppb — the current ozone standard is 120 ppb) or else the analysis is not too sensitive to the precise choice of u. Nevertheless, it must be admitted that the choice of threshold remains a major practical issue in this kind of analysis, and may still be the area where further theoretical development is most badly needed.

The Poisson process approach

An alternative approach, introduced by Smith (1989), is based on viewing the two-dimensional process of exceedance times and excesses as a point process in $\mathcal{R}^2$. Limit theorems for such processes were developed by Pickands (1971) and Resnick (1975) in the independent case, and formed a major impetus for the probabilistic work on extremes in

dependent processes, reviewed in Leadbetter *et al.* (1983). However, the *statistical* application of these processes is more recent. Although at first sight more complicated than the GPD approach, it ultimately helps to simplify and unify the whole theory, especially when covariates are present.

To develop this approach, let $X_1, ..., X_n$ be i.i.d. ($\sim F$). Suppose F is in the domain of attraction of an extreme value distribution. One way to express that property is to say that there exist constants $a_n > 0$ and b_n such that

$$n\{1 - F(a_n y + b_n)\} \to V(y) \tag{1.5}$$

for each y, and hence

$$F^n(a_n y + b_n) \to e^{-V(y)}. \tag{1.6}$$

Here we may, without loss of generality, take $V(y) = (1 + \xi y)_+^{-1/\xi}$ for fixed ξ, in which case the limiting distribution function in (1.6) is of generalized extreme value (GEV) form.

Now define a point process P_n on $[0,1] \times \mathcal{R}$ by putting a point at each $\{i/n, (X_i - b_n)/a_n)\}$, $1 \leq i \leq n$. The total number of points in this process is of course n, but the expected number in $[0,1] \times (y, \infty)$ for any fixed y is $n\{1 - F(a_n y + b_n)\}$ which, by (1.5), converges to $V(y)$ (finite). From this it is possible to establish that P_n converges weakly, on all sets for which the second coordinate is bounded away from its lower boundary, to a nonhomogenous Poisson process P whose intensity measure Λ ($\Lambda(A)$ is the expected number of points in A, for $A \subset [0,1] \times \mathcal{R}$) satisfies

$$\Lambda\{(t_1, t_2) \times (y, \infty)\} = (t_2 - t_1)V(y)$$

whenever $0 \leq t_1 < t_2 \leq 1$ and $y \in \mathcal{R}$ ($y > -1/\xi$ if $\xi > 0$).

In practice, rather than try to construct the limiting process with the renormalization required, a more practical approach is to work directly with a nonhomogeneous Poisson process applied to all observations over a threshold u. Thus, for any observation X_t, taken at time t, for which $X_t > u$, one puts a point at (t, X_t). This is treated as part of a non-homogeneous Poisson process on $\mathcal{R} \times (u, \infty)$ whose intensity measure satisfies

$$\Lambda\{(t_1, t_2), (x, \infty)\} = (t_2 - t_1)\left(1 + \xi\frac{x - \mu}{\psi}\right)_+^{-1/\xi}, \quad t_1 \leq t_2,\ x \geq u. \tag{1.7}$$

It can be seen, from a number of different points of view, that this is equivalent to either of the usual approaches to extreme value theory. If M_T denotes the maximum of the process over time interval $(0,T)$, assumed bigger than u, then the event $M_T \leq x$ is equivalent to saying that the set $(0,T) \times (x,\infty)$ is empty, and this has probability

$$\exp\left\{-T\left(1+\xi\frac{x-\mu}{\psi}\right)_+^{-1/\xi}\right\}, \quad x \geq u,$$

which is of the usual generalized extreme value form. Another fact that follows directly from (1.7) is that, given a point (T_i, X_i) for which $X_i > u$, the conditional probability that $X_i - u > y$ is

$$\frac{\left(1+\xi\frac{u+y-\mu}{\psi}\right)_+^{-1/\xi}}{\left(1+\xi\frac{u-\mu}{\psi}\right)_+^{-1/\xi}} = \left\{1+\frac{\xi y}{\psi+\xi(u-\mu)}\right\}_+^{-1/\xi}$$

which is of GPD form (1.2) with $\sigma = \psi + \xi(u-\mu)$.

Estimation from (1.7) proceeds by constructing the appropriate likelihood for an nonhomogeneous Poisson process (Cox and Lewis 1966, Section 3.3): if we re-define $V(x)$ to be

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

and let $v(x;\mu,\psi,\xi) = -\partial V(x;\mu,\psi,\xi)/\partial x$, then if we observe a random number N of exceedances (T_j, X_j) with $X_j > u$ over a time period $(0,T)$, the approximate likelihood function is

$$L(\mu,\psi,\xi) = \exp\{-TV(u;\mu,\psi,\xi)\} \cdot \prod_{j=1}^{N} v(X_j;\mu,\psi,\xi). \qquad (1.8)$$

Estimation of μ, ψ and ξ proceeds by numerical maximization of (1.8).

Comparison of the GPD and Poisson process approaches

The two approaches are equivalent in the sense that any model which is expressed in terms of one of the two approaches may be rewritten

as an equivalent model in terms of the other. The main advantage of the Poisson approach is that the parametrization is more convenient for complex models involving covariates. Suppose there is a vector of covariates z_t for each time point t, and that the quantities $\mu = \mu_t$, $\psi = \psi_t$, $\xi = \xi_t$ depend on a parameter vector η through functions $\mu_t(\eta) = \mu(z_t;\eta)$, $\psi_t(\eta) = \psi(z_t;\eta)$, $\xi_t(\eta) = \xi(z_t;\eta)$. Then the likelihood, as a function of η, may be expressed in the form

$$L(\eta) = \exp\left\{-\int_0^T V(u;\mu_t(\eta),\psi_t(\eta),\xi_t(\eta))dt\right\} \cdot \\ \cdot \prod_{j=1}^N v(X_j;\mu_{T_j}(\eta),\psi_{T_j}(\eta),\xi_{T_j}(\eta)),$$

extending (1.8). For example, in the not uncommon situation that the location parameter of a distribution is taken to depend on the covariates, with the scale and shape parameters held fixed, we can simplify this by setting $\psi_t(\eta)$ and $\xi_t(\eta)$ equal to constants ψ and ξ. Such a simplification was used by Smith (1989) to model the case of an additive linear trend. It is not so easy to see how to incorporate such a feature into the GPD model. In principle, then, I believe this method to be superior to the GPD approach, though for practical application it is beneficial to have both methods in hand.

2. Multivariate extremes.

Multivariate extreme value theory is concerned with the joint distribution of extremes of two or more dependent random variables. The traditional approach to their definition has been via componentwise maxima (or minima): if $\{(X_{i1},...,X_{ip}),\ i = 1,...,n\}$ is a sample of independent p-vectors, then the vector of componentwise maxima is $(M_{n1},...,M_{np})$, where $M_{nj} = \max(X_{1j},...,X_{nj})$, $j = 1,...,p$. The limiting joint distributions of these maxima $(M_{n1},...,M_{np})$, subject to location-scale renormalization of each component, are the *multivariate extreme value distributions*. The bivariate case $p = 2$ is of particular significance, both because it was the first case to be studied in detail and because much of the theory is simpler for this than the general multivariate case.

The main contrast with the univariate case is that there is no finite-parameter family which exhausts the class of multivariate extreme value distributions, for any $p \geq 2$. Hence there are two basic approaches, namely those based on parametric subfamilies and those which are essentially nonparametric. The discussion in this paper is concerned entirely with parametric families. How severe a restriction this is, from a practical point of view, is still not clear, but there are now several different approaches to the construction of multivariate extreme value families, so this is still a very broad approach.

There are several different but equivalent characterizations of multivariate extreme value distributions, reviewed in detail in the books of Resnick (1987) and Galambos (1987). One feature of multivariate extreme value distributions is that the dependence structure is preserved under transformations of the marginal distributions, so there is no loss of generality in restricting attention to a particular univariate extreme value family. For example, de Haan and Resnick (1977) assumed unit Fréchet ($F(y) = e^{-1/y}$) margins and developed the characterization

$$G(y_1, ..., y_p) = \exp\left\{ - \int_{T_p} \max_j \left(\frac{u_j}{y_j} \right) \mu(du_1, ..., du_p) \right\} \qquad (2.1)$$

where μ is a positive measure on the set

$$T_p = \{(u_1, ..., u_p) : \ u_j \geq 0, \ \sum_j u_j^2 = 1\}$$

subject to the condition

$$\int_{T_p} u_j \mu(du_1, ..., du_p) = 1 \text{ for all } j.$$

An alternative approach, due to Pickands, leads to a representation formula (with GEV margins) of the form

$$G(y_1, ..., y_p) = \exp\left\{ - \left(\sum_{j=1}^{p} t_j \right) A\left(\frac{t_1}{\sum_{j=1}^{p} t_j}, ..., \frac{t_p}{\sum_{j=1}^{p} t_j} \right) \right\}, \qquad (2.2)$$

where $A(x_1, ..., x_p)$ is a convex function over the unit simplex $S_p = \{(x_1, ..., x_p : \ x_1 \geq 0, ..., x_p \geq 0, \sum x_j = 1\}$ which takes the value 1 at the corner points (when one x_j is 1 and the rest 0), and $t_j =$

$\{1+\xi_j(y_j-\mu_j)/\psi_j\}_+^{-1/\xi_j}$. In the bivariate case this reduces to a simpler form

$$G(y_1,y_2) = \exp\left\{-(t_1+t_2)\,A\left(\frac{t_2}{t_1+t_2}\right)\right\} \tag{2.3}$$

with A a convex function on $[0,1]$ satisfying $A(0)=1$, $A(1)=1$, $A(x)\geq \max(x,1-x)$ for $0<x<1$ (the case where equality is attained in the last expression being the case where both components are equal with probability 1, the extreme case of complete dependence).

In principle, one can generate multivariate distributions by taking general formulae such as (2.1)-(2.3) and inserting specific functions for μ or A. Simple-minded attempts to do this fail to yield useful or tractable families, but there are by now a number of approaches which do yield usable parametric families. First I consider the bivariate case and then generalize to $p>2$.

Extensive work on bivariate extremes was carried out by Tiago de Oliveira in a long series of papers (for a review see Tiago de Oliveira, 1984), but recent years have seen a number of new models introduced. Among the models currently considered are

(a) *Mixed Model* (Gumbel and Mustafi 1967):

$$A(w) = \theta w^2 - \theta w + 1, \quad \theta \in [0,1],$$

(b) *Logistic Model:* (Gumbel 1960):

$$A(w) = \{(1-w)^{1/\alpha} + w^{1/\alpha}\}^{\alpha}, \quad 0 \leq \alpha \leq 1,$$

the limits $\alpha = 1$ and $\alpha \downarrow 0$ corresponding to independence and complete dependence respectively,

(c) *Asymmetric Mixed Model* (Tawn 1988):

$$A(w) = \phi w^3+\theta w^2-(\theta+\phi)w+1, \quad (\theta \geq 0, \theta+\phi \leq 1, \theta+2\phi \leq 1, \theta+3\phi \geq 0),$$

(d) *Asymmetric Logistic Model* (Tawn 1988):

$$A(w) = \{\theta^{1/\alpha}(1-w)^{1/\alpha} + \phi^{1/\alpha}w^{1/\alpha}\}^{\alpha} + (\theta-\phi)w + 1 - \theta,$$

in which $0 \leq \theta \leq 1, 0 \leq \phi \leq 1, 0 \leq \alpha \leq 1$. The asymmetric versions of the mixed and logistic models allow for non-exchangability between the two components.

(e) *Inverted Logistic Model* (Joe 1990):

$$A(w) = 1-[(\phi_1 w)^{-\tau}+\{\phi_2(1-w)\}^{-\tau}]^{-1/\tau},\ \tau \geq 0, 0 \leq \phi_1 \leq 1, 0 \leq \phi_2 \leq 1.$$

The simplest (symmetric) form is when $\phi_1 = \phi_2 = 1$; in this form the mixed model with $\theta = 1$ arises when $\tau = 1$ while the limits $\tau \to 0, \tau \to \infty$ correspond respectively to independence and complete dependence.

(f) *Gaussian Model*

$$A(w) = (1-w)\Phi\Big\{\frac{a}{2} + \frac{1}{a}\log\Big(\frac{1-w}{w}\Big)\Big\} + w\Phi\Big\{\frac{a}{2} + \frac{1}{a}\log\Big(\frac{w}{1-w}\Big)\Big\}$$

for $a \in [0, \infty]$, where Φ is the normal distribution function. This was introduced independently by Hüsler and Reiss (1989), who derived it as a penultimate approximation for dependent normal extremes, and Smith (1991) as an example of a max-stable process (see Stuart Coles' paper at this meeting for a discussion of max-stable processes).

(g) *Bilogistic Model* (Joe, Smith and Weissman 1992).

$$A(w) = \int_0^1 \max\{(1-\alpha)(1-w)u^{-\alpha}, (1-\beta)w(1-u)^{-\beta}\}du$$

in which $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. This is an alternative to (d) as an asymmetric form of the logistic model, to which it reduces when $\alpha = \beta$.

For multivariate ($p > 2$) cases, the possibilities are even more diverse. In this case an equivalent characterization of Pickands' A function (from (2.2)) is

$$A(w_1, ..., w_p) = \int_{S_p} \max_j (w_j u_j) dH(u_1, ..., u_p), \tag{2.4}$$

H being a positive measure on S_p satisfying

$$\int_{S_p} u_j dH(u_1, ..., u_p) = 1, \quad j = 1, ..., p. \tag{2.5}$$

Suppose we restrict H to be an absolutely continuous measure on the interior of S_p. This is a nontrivial restriction, since many examples require a degenerate H measure — for instance, in the independent case H gives mass 1 to each of the corner points of the simplex. We can construct families of multivariate extreme value distributions by finding

measures H satisfying (2.5) and then applying (2.4). However, it is still not so easy to find suitable H, especially with the restriction (2.5).

One construction suggested by Coles and Tawn (1991) was to take an arbitrary density h^* on S_p with positive first moments, and to define

$$m_j = \int_{S_p} u_j h^*(u_1, ..., u_p) du_1, ..., du_p, \quad j = 1, ..., p.$$

followed by

$$h(w_1, ..., w_p) = \prod_{j=1}^{p} m_j \left(\sum_j m_j w_j\right)^{-p-1} h^*\left(\frac{m_1 w_1}{\sum_j m_j w_j}, ..., \frac{m_p w_p}{\sum_j m_j w_j}\right).$$

Then h is the density of a valid measure H satisfying (2.5), and (2.4) may be used to construct the A function and hence the distribution function.

An example for which this construction works is the Dirichlet density,

$$h^*(w_1, ..., w_p) = \left\{\frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)}\right\} \prod_{j=1}^{p} w_j^{\alpha_j - 1}$$

for $\alpha_j > 0$, $j = 1, ..., p$, and $(w_1, ..., w_p) \in S_p$. Then $m_j = \alpha_j / (\sum_k \alpha_k)$ and

$$h(w_1, ..., w_p) = \prod_{j=1}^{p} \left\{\frac{\alpha_j}{\Gamma(\alpha_j)}\right\} \frac{\Gamma(\sum_j \alpha_j + 1)}{(\sum_j \alpha_j w_j)^{p+1}} \prod_{j=1}^{p} \left(\frac{\alpha_j w_j}{\sum_k \alpha_k w_k}\right)^{\alpha_j - 1}.$$

This is therefore known as the Dirichlet model for multivariate extremes.

Other approaches to the construction of multivariate extreme value distributions include a hierarchical approach introduced by Tawn (1990) and in a related way by Joe (1990), time series contructions (cf. Section 4 below) and methods based on specializations of max-stable processes.

Coles and Tawn (1991) reviewed these approaches in more detail. Two models that have been known for a very long time are the p-dimensional generalization of Gumbel's logistic model (Gumbel 1960), $A(w_1, ..., w_p) = \left(\sum w_j^{1/\alpha}\right)^{\alpha}$, and the family introduced by Marshall and Olkin (1967). However, Gumbel's model has the severe restriction that it forces all p components to be exchangeable, while the Marshall-Olkin model can never be absolutely continuous, a fact which restricts its range of applicability in the areas in which we are interested. With the above classes of models, there are now many more possibilities.

Estimation.

The estimation of multivariate extreme value distributions has long been known to be problematic because all of them become nonregular as the independence case is approached. That is, standard maximum likelihood asymptotic theory breaks down, as the Fisher information about the dependence parameter tends to ∞. This problem was, however, essentially solved by Tawn (1988, 1990), who developed stable asymptotic distributions for the score statistic at independence, and used this property to characterize the limiting properties of maximum likelihood in this case. For interior points of the parameter space (i.e. bounded away from independence or any of the other boundary conditions) these difficulties do not arise and maximum likelihood seems to behave perfectly well. In any case, our own preference is now to use standard maximum likelihood theory for all calculations except that of testing independence, for which we know that the standard χ^2 distribution for the likelihood ratio statistic breaks down. Even then, maximum likelihood may not be easy to implement in practice, as it requires being able to evaluate the density, and in several of the cases described above (for instance, the Dirichlet model) this is not available in closed form. Computer algebra is one of the techniques that has been used to get round this problem.

3. Threshold methods for multivariate extremes

The first attempts to construct threshold-based methods of statistical inference in the multivariate cases were those of Coles and Tawn (1991) and Joe *et al.* (1992). Both of these are based on a generalization of the point process approach described in Section 1. If we have multivariate data $\{X_{i1}, ..., X_{ip}\}$ and renormalize each component so that $(M_{nj} - b_{nj})/a_{nj}$ tends to a limiting univariate extreme value distribution, then we can form a point process P_n on $[0,1] \times \mathcal{R}^p$ by putting a point at each $\{i/n, (X_{i1} - b_{i1})/a_{i1}, ..., (X_{ip} - b_{ip})/a_{ip}\}$, $1 \leq i \leq n$. In this case the number of points in any rectangle of the form $(t_1, t_2) \times (y_1, \infty) \times ... \times (y_p, \infty)$ converges to a limiting form, which we may write in the form $(t_2 - t_1)V(y_1, ..., y_p)$, as $n \to \infty$. In fact the limiting extreme value distribution, in the sense defined in Section 2, is

just $\exp\{-V(y_1,...,y_n)\}$ and this result combined with representations such as (2.2) may be used to obtain point-process methods for inference about multivariate extremes. For applications of this technique, I refer to Coles and Tawn (1994) and Jonathan Tawn's paper at this meeting.

For the remainder of this section I discuss an alternative approach, currently being developed by Jonathan Tawn, Stuart Coles and myself (Smith *et al.* 1993), in which we work directly with the exceedances over a threshold, analogous to the GPD approach in the univariate case.

Suppose $(X_1,\ldots,X_p)$ is a typical p-variate data point but that, in the spirit of threshold methods, for each j between 1 and p, we observe not X_j but $Z_j = \max(X_j, u_j)$, together with an indicator δ_j ($\delta_j = 1$ if $X_j > u_j$, 0 otherwise). Here u_j is a fixed threshold, which may be different for each component j. We want to derive an approximation for the joint distribution of $\{(\delta_j, Z_j),\ j = 1,\ldots,p\}$. The underlying assumption is that $(X_1,\ldots,X_p)$ lies in the domain of attraction of a multivariate extreme value distribution. One characterization of that, described in detail by Resnick (1987), is first to transform each margin to unit Fréchet form by means of a probability integral transformation, and then to define the distribution function $F_*(v_1,...,v_p)$ for the transformed distribution. Resnick's Proposition 5.15 shows that the distribution is in the domain of attraction of a multivariate extreme value family if and only if, for all $(v_1,...,v_p)$ in which each $v_j > 0$,

$$\lim_{t\to\infty} \frac{1 - F_*(tv_1,\ldots,tv_p)}{1 - F_*(t,\ldots,t)} = \frac{-\log G_*(v_1,\ldots,v_p)}{-\log G_*(1,\ldots,1)} = \frac{V(v_1,...,v_p)}{V(1,...,1)}. \quad (3.1)$$

where G_* is a multivariate extreme value distribution function with unit Fréchet margins.

By analogy with threshold methods for univariate extremes, in which the approximate generalized Pareto distribution is in effect treated as exact for sufficiently high thresholds, we may consider a scheme in which (3.1) is treated as exact, for sufficiently large t and $v_1,...,v_p$.

The assumption we make about each marginal component is that for sufficiently large threshold u_j, the marginal distribution of $X_j - u_j$ given $X_j > u_j$ is GPD. This allows us to write the j'th marginal distribution function in the form

$$F_j(x) = 1 - \lambda_j\{1 + \xi_j(x - u_j)/\sigma_j\}_+^{-1/\xi_j}, \quad x \geq u_j,$$

where $\lambda_j = 1 - F_j(u_j)$. By analogy, we also take the limiting result (3.1) as an identity for sufficiently large t. In fact, it is more convenient to treat

(3.1) as an identity for some fixed $t = t_c$, provided the v_j are sufficiently large; this is clearly an equivalent interpretation. Specifically, taking $t_c = 1$ and $v_j \geq \lambda_j^{-1}$, $j = 1, \ldots, p$ ensures that the GPD approximation is applicable for each marginal component. Our additional assumption is that these levels are also sufficiently high for the asymptotic dependence structure to be a valid approximation through (3.1).

By assuming this form of F_j, transforming to unit Fréchet (in the tails) and applying (3.1), the form of multivariate joint distribution that arises is

$$F(x_1, \ldots, x_p) = 1 - V\left\{\lambda_1^{-1}\left(1 + \xi_1 \frac{x_1 - u_1}{\sigma_1}\right)_+^{1/\xi_1}, \ldots, \lambda_p^{-1}\left(1 + \xi_p \frac{y_p - u_p}{\sigma_p}\right)_+^{1/\xi_p}\right\}, \tag{3.2}$$

valid for $x_1 \geq u_1, \ldots, x_p \geq u_p$. For example, in the $p = 2$ case with the logistic model of Section 2, we have $V(y_1, y_2) = (y_1^{-1/\alpha} + y_2^{-1/\alpha})^\alpha$, so equation (3.2) becomes, for $x_1 \geq u_1$, $x_2 \geq u_2$,

$$F(x_1, x_2) = 1 - \left\{\lambda_1^{1/\alpha}\left(1 + \xi_1 \frac{x_1 - u_1}{\sigma_1}\right)_+^{-1/(\alpha\xi_1)} + \lambda_2^{1/\alpha}\left(1 + \xi_2 \frac{y_2 - u_2}{\sigma_2}\right)_+^{-1/(\alpha\xi_2)}\right\}^\alpha. \tag{3.3}$$

This formula still has one disadvantage: when $\alpha = 1$ it does not reduce exactly to the independent GPD case. To get around this problem, Jonathan Tawn and Anthony Ledford have suggested an alternative construction, starting from replacing $1 - F_*$ with $-\log F_*$ in (3.1), which leads to the formula

$$F(x_1, \ldots, x_p) = \exp\left[-V\left\{\left(-\log\left(1 - \lambda_1\left(1 + \xi_1 \frac{x_1 - u_1}{\sigma_1}\right)_+^{-1/\xi_1}\right)\right)^{-1}, \ldots, \left(-\log\left(1 - \lambda_p\left(1 + \xi_p \frac{x_p - u_p)}{\sigma_p}\right)_+^{-1/\xi_p}\right)\right)^{-1}\right\}\right] \tag{3.4}$$

which shares the same asymptotic properties as (3.2) while avoiding this one undesirable feature. In this case, (3.3) becomes

$$F(x_1,x_2) = \exp\left[-\left\{\left(-\log\left(1-\lambda_1\left(1+\xi_1\frac{x_1-u_1}{\sigma_1}\right)_+^{-1/\xi_1}\right)\right)^{1/\alpha} + \left(-\log\left(1-\lambda_2\left(1+\xi_2\frac{x_2-u_2}{\sigma_2}\right)_+^{-1/\xi_2}\right)\right)^{1/\alpha}\right\}^{\alpha}\right]. \tag{3.5}$$

The main complication in the estimation of these distributions is the form of the likelihood. The contribution to the likelihood from a vector $\{(\delta_j, Z_j)\}$ is constructed by differentiating the joint distribution function with respect to the observed components. As an example, suppose $p = 3$ and $\delta_1 = \delta_3 = 1$, $\delta_2 = 0$ (observations 1 and 3 above the threshold, 2 below). The contribution to the likelihood function from this triple is

$$\frac{\partial^2 F(x_1,x_2,x_3)}{\partial x_1 \partial x_3}\Big|_{x_1=X_1, x_2=u_2, x_3=X_3}.$$

Maximum likelihood estimation in these cases is therefore straightforward in principle. In terms of its asymptotic properties, maximum likelihood is regular at interior points of the parameter space, but there are still difficulties in testing independence which require the construction of special tests (work by Ledford and Tawn, in progress).

4. Markov chains

One consequence of the methodology of Section 3 is that it gives us a new way to think about extreme values in univariate time series. The work described here is based on Smith *et al.* (1993).

The threshold approach in Section 2 is based on identifying clusters of neighboring exceedances, and then using the GPD to model the largest value over the threshold within each cluster. This is open to two major objections, first that the method of identifying clusters is arbitrary, and second that it seems inefficient to model only the largest value within each cluster when all of them contain useful information. An alternative approach would be to assume some model for the dependence in the time series. However, the two major classes of models used in time

series analysis are Gaussian models and linear models. Gaussian models are not very satisfactory in extreme value analysis because, subject to a very weak long-range independence condition, they always have extremal index 1, and therefore necessarily fail to capture the features we are trying to model. Linear models are in principle a more general class, but the required extremal computations are not easy, and there remains the question of how to verify the correctness of the linearity assumption, which could be crucial to any calculation of extreme value properties.

The approach suggested here is to model the short-term dependence by assuming the process to be a k-th order Markov chain, i.e. the distribution of X_n given X_m, $m < n$ is determined completely by a transition density depending only on $X_{n-k}, ..., X_{n-1}$. We also assume the process to be stationary. Such a process is, of course, completely specified by its consecutive $(k+1)$'st order joint distributions, and if we assume these are in the domain of attraction of a $(k+1)$-dimensional Markov chain, an attractive theory results.

Suppose the consecutive j-dimensional joint densities are denoted by $f_j(x_1, ..., x_j)$. Then the joint density of a realization $(X_1, ..., X_n)$ for $n > k$ is given by

$$\frac{\prod_{j=1}^{n-k} f_{k+1}(X_j, ..., X_{j+k})}{\prod_{j=2}^{n-k} f_k(X_j, ..., X_{j+k-1})}. \tag{4.1}$$

Thus, given parametric families for f_k and f_{k+1}, the likelihood function is a ratio of two terms, each of which is a constructed in the same way as a likelihood for independent multivariate data.

Now suppose the parametric family is not completely specified, but only determined in the tails, as in Section 3. Then we may approximate each of the numerator and denominator in (4.1) by an appropriate limiting form. Our specific proposal is to use the approximations developed in Section 3 to do this.

One consequence of adopting this approach is that it is still necessary to calculate extremal properties of the process for the fitted Markov chain. There has been considerable recent progress in this area. For example, Smith (1992) developed a method of calculating the extremal index for a Markov chain under the main assumption that the consecutive bivariate distributions lie in the domain of attraction of a bivariate extreme value distribution. Yun (1993) has extended this calculation to k'th order Markov chains. Smith *et al.* (1993) discuss a variety of other functions such as the sum of exceedances within a cluster. Anderson

and Dancy (1992) refer to this as the severity of an extreme event, and develop a characterization of its asymptotic form in terms of the limiting point process of crossings of the threshold. With the exception of Smith (1992), which reduced to the numerical solution of a Wiener-Hopf integral equation, all of these solutions rely on simulation to calculate limiting distributions of functionals within a cluster, but such methods seem entirely within the scope of available computing resources.

Smith *et al.* (1993) applied these ideas (with $k = 1$) to model seasonal minima in a long-term temperature series, and a follow-up paper by Coles *et al.* (1993) will develop this analysis further.

5. An application: High-level exceedances of tropospheric ozone

The ideas developed in this paper have proved to be valuable in a major applied study currently in progress, concerning urban ozone. Only a bare outline is presented here; full details are available in the technical report of Smith and Huang (1993).

Ozone is produced by complex chemical processes in the lower atmosphere as a result of the emissions into the atmosphere of hydrocarbons and nitrogen oxides. According to the current national ambient air quality standard (NAAQS), at any individual monitoring site, the number of exceedances by daily ozone maxima of the level of 120 parts per billion (ppb) should not exceed 3 in any 3-year period. Regions which violate this are deemed non-attainment regions. These non-attainment regions are generally in major cities such as New York, Chicago, Houston and Los Angeles.

The principal purpose of the present study is to decide whether there are trends in measured ozone data, after taking account of what are known to be substantial meteorological factors. In earlier work as part of the same project, Bloomfield *et al.* (1993) fitted nonlinear regression models to daily ozone maxima, and for an extensive data set from the Chicago area, concluded that there was a significant downward trend in ozone levels during the 1980s after adjusting for a variety of meteorological variables. Apart from the use of daily maxima, however, there was no specific attempt in their work to focus on the extreme values of the process. Concern about ozone is usually focussed on extreme values, as

is reflected in the ozone standard, and this served to motivate the alternative analysis discussed here, which is based explicitly on exceedances over a high threshold.

The data analyzed were daily maximum ozone levels for the summer months from 1981 to 1991, a total of 2354 daily values. Raw data were available for 45 monitoring stations in greater Chicago; for the purpose of the present analysis, a set of "network maxima" based on 16 of those stations was used (Bloomfield *et al.*, 1993). Meteorological variables include noon values of temperature, windspeed and a number of other variables from the meteorological station at O'Hare Airport. There also appears to be a seasonal factor over and above what is explainable in terms of meteorology, and this is modeled by introducing covariates $\cos(2\pi d/366)$, $\sin(2\pi d/366)$, where d is calendar day within year ($1 \leq d \leq 366$). Finally, the year itself is treated as a covariate in the analysis.

A natural starting point is to consider just the crossing probability of a single threshold (taken as 120 ppb) and to look for trends in that. For this purpose we use the logistic link function:

$$\log\left\{\frac{p(t)}{1-p(t)}\right\} = \sum_j \beta_j z_j(t) \tag{5.1}$$

where $p(t)$ is the probability of exceeding the threshold on day t and $z_j(t)$, $j = 1, 2, \ldots$, are known covariates and $\beta_1, \beta_2, \ldots$ unknown covariates to be estimated. Equation (5.1) is fitted by maximum likelihood to the binary data consisting of 1 if there is an exceedance of the threshold on day t, 0 otherwise.

In the present case, fitting (5.1) by stepwise selection of variables resulted in a regression model including year and the seasonal variables as covariates, as well as a number of meteorological variables: temperature, windspeed factored into directional components, pressure, visibility, relative humidity and a temperature-windspeed interaction term. Significance of the yearly trend may be assessed either from the standard error of the estimated coefficient, or from comparisons of log likelihoods (or deviances) of models with and without the fitted trend. The results do confirm the significance of the yearly trend. At this stage, we are implicitly assuming separate days to be independent.

A full discussion of the problem, however, requires more than the crossing probabilities for a single threshold. For example, one question of importance is to explain the very high ozone levels of 1988, when

the highest reading in Chicago was 223 ppb. This cannot be answered purely in terms of the probability of exceeding 120. Continuing to assume independent days, the obvious analysis is to use the GPD of equation (1.2) to model excesses over the threshold, together with (5.1) for the exceedance probabilities. The analysis uses a logarithmic link function for σ (i.e. we assume that $\log \sigma$ is a linear function of the covariates), with constant ξ. The results imply a significant trend in the excess values, as well as the exceedance probabilities.

We can now ask the question of whether the assumption of independent days is in fact a reasonable one. The current thinking of most researchers working in this area is that raw ozone data are serially correlated from day to day, but this is due primarily to the persistence of meteorological influences, and once these are taken into account, independence is a reasonable assumption. Bloomfield *et al.* (1993) assumed independence based primarily on this reasoning.

We can study this issue by assuming a first-order Markov structure with (3.5) to model the joint distribution of consecutive days. Here λ_1 and λ_2 are the threshold crossing probabilities, modeled by (5.1), and we assume $\xi_1 = \xi_2$ independent of covariates, $\log \sigma_1$ and $\log \sigma_2$ as linear functions of the covariates on the respective days. Fitting this model, with the same covariates as in the previous analyses, resulted in an estimate $\hat{\alpha} = 0.94$ with an estimated standard error of 0.32, and a deviance statistic (twice the difference of log likelihoods) of 6.33 compared with the independence model $\alpha = 1$. As previously pointed out, this cannot be assessed against the usual χ^2_1 distribution, because the asymptotic theory justifying this breaks down, and it is somewhat ambiguous whether this is a significant result or not. In contrast, if we omit the meteorological covariates, there is no ambiguity at all: in that case estimated values of α are in the range 0.75–0.8 and the deviance statistic much larger, indicating clear serial dependence in daily ozone.

Introducing dependence into the model makes little difference to the estimated trends, which are clearly significant under either analysis. However, it does affect the way we judge the model fit: the extreme values of 1988 are clear outliers as judged against the independence model, but do seem to be consistent with the first-order Markov model as judged by some predictive diagnostics developed in Smith and Huang (1993). For this reason, our current thinking is that it is important to take serial dependence into account when trying to model the most extreme levels

of ozone.

Acknowledgement. This research was supported by the National Science Foundation. In addition, the work reported in Section 5 is based on a study carried out at the National Institute of Statistical Sciences under Cooperative Agreement #CR819638-01-0 from the U.S. Environmental Protection Agency. I thank Jonathan Tawn for suggesting the formulation leading to equations (3.4) and (3.5).

References

Anderson, C.W. and Dancy, G. (1992), The severity of extreme events. Preprint, Sheffield University.

Bloomfield, P., Royle, A. and Yang, Q. (1993), Accounting for meteorological effects in measuring urban ozone levels and trends. Technical Report, National Institute of Statistical Sciences, P.O. Box 14162, Research Triangle Park, N.C. 27709-4162.

Coles, S.G. and Tawn, J.A. (1991), Modelling extreme multivariate events. *J. R. Statist. Soc. B* **53**, 377-392.

Coles, S.G. and Tawn, J.A. (1994), Statistical methods for multivariate extremes: An application to structural design (with discussion). *Applied Statistics*, to appear.

Coles, S.G., Tawn, J.A. and Smith, R.L. (1993), A seasonal Markov model for extremely low temperatures. Submitted to *Environmetrics.*

Cox, D.R. and Lewis, P.A.W. (1966), *The Statistical Analysis of Series of Events.* Methuen, London.

Davison, A.C. (1984), Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications*, J. Tiago de Oliveira (ed.), Reidel, Dordrecht, 461-482.

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

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

1795-1832.

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

Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics* (2nd. edn.). Krieger, Melbourne, Fl. (First edn. published 1978 by John Wiley, New York.)

Gumbel, E.J. (1960), Distributions des valeurs extremes en plusiers dimensions. *Publ. Inst. Statist. Univ. Paris* **9**, 171-173.

Gumbel, E.J. and Mustafi, C.K. (1967), Some analytic properties of bivariate extremal distributions. *J. Amer. Statist. Assoc.* **62**, 569-588.

Haan, L. de and Resnick, S.I. (1977), Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie v. Geb.* **40**, 317-337.

Haan, L. de and Rootzén, H. (1993), *Journal of Statistical Planning and Inference.*

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

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

Hsing, T. (1987), On the characterization of certain point processes. *Stoch. Proc. Appl.* **26**, 297-316.

Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988), On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* **78**, 97-112.

Hsing, T. (1991), Estimating the parameters of rare events. *Stoch. Proc. Appl.* **37**, 117-139.

Hüsler, J. and Reiss, R.-D. (1989), Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters* **7**, 283-286.

Joe, H. (1990), Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics and Probability Letters* **9**, 75-81.

Joe, H., Smith, R.L. and Weissman, I. (1992), Bivariate threshold methods for extremes. *J.R. Statist. Soc. B.*, **54**, 171-183.

Leadbetter, M.R. (1983), Extremes and local dependence in stationary sequences. *Z. Wahrsch. v. Geb.* **65**, 291-306.

Leadbetter, M.R. (1991), On a basis for 'peaks over thresholds' modeling. *Statistics and Probability Letters* **12**, 357-362.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Series.* Springer Verlag, New York.

Leadbetter, M.R., Weissman, I., de Haan, L. and Rootzén, H. (1989), On clustering of high levels in statistically stationary series. *Proceedings of the Fourth International Meeting on Statistical Climatology*, ed. John Sansom. New Zealand Meteorological Service, P.O. Box 722, Wellington, New Zealand.

Marshall, A.W. and Olkin, I. (1967), A multivariate extremal distribution. *J. Amer. Statist. Assoc.* **62**, 30-44.

Nandagopalan, S. (1990), Multivariate extremes and the estimation of the extremal index. Ph.D. dissertation available as Technical Report Number 315, Center for Stochastic Processes, Department of Statistics, University of North Carolina.

Pickands, J. (1971), The two-dimensional Poisson process and extremal processes. *J. Appl. Prob.* **8**, 745-756.

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

Resnick, S. (1975), Weak convergence to extremal processes. *Ann. Probab.* **3**, 951-960.

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

Smith, R.L. (1984), Threshold methods for sample extremes. In *Statistical Extremes and Applications*, J. Tiago de Oliveira (ed.), Reidel Dordrecht, 621-638.

Smith, R.L. (1985), Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67-90.

Smith, R.L. (1987), Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.

Smith, R.L. (1989), Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* 4, 367-393.

Smith, R.L. (1990), Extreme value theory. Chapter 14 of *Handbook of Applicable Mathematics*, W. Ledermann (ed.). John Wiley, Chichester, pp. 437-472.

Smith, R.L. (1991), Max-stable processes and spatial extremes. Preprint under revision.

Smith, R.L. (1992), The extremal index for a Markov chain. *J. Applied Probability* **29**, 37-45.

Smith, R.L. and Huang, L.-S. (1993), Modeling high threshold exceedances of urban ozone. Technical Report, National Institute of Statistical Sciences, P.O. Box 14162, Research Triangle Park, N.C. 27709-4162.

Smith, R.L., Tawn, J.A. and Coles, S.G. (1993), Markov chain models for threshold exceedances. Submitted to *Biometrika*.

Smith, R.L. and Weissman, I. (1994), Estimating the extremal index. To appear, *J.R. Statist. Soc. B*.

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

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

Tiago de Oliveira, J. (1984b), Bivariate models for extremes. In *Statistical Extremes and Applications*, J. Tiago de Oliveira (ed.), 131-153, Reidel, Dordrecht.

Yun, S. (1993), The extremal index of a higher-order stationary Markov chain. Manuscript in preparation.

Richard L. Smith
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260, U.S.A.