Bivariate Threshold Methods for Extremes

H. Joe, R.L. Smith & I. Weissman

Technical Report #137

SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

HEALTH SCIENCES PROGRAM
RAND CORPORATION

DEPARTMENT OF APPLIED EARTH SCIENCES
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
Bivariate Threshold Methods for Extremes

by

H. Joe, R.L. Smith and I. Weissman

Technical Report 137

August 1989

Prepared in part under support to SIMS from the United States Environmental Protection Agency.

Department of Statistics, University of British Columbia, 2021 West Mall, Vancouver, British Columbia, CANADA V6T 1W5.
This is one of a series of technical reports produced under a Cooperative Agreement, "Study on Statistics and Environmental Factors in Health (Acid Deposition, Human Exposure, Quality Assurance, and Global Climate Change)" sponsored by the U.S. Environmental Protection Agency. The study is supported by a grant to SIMS (Societal Institute of the Mathematical Sciences) from the EPA. Cooperating members of the study are Rand Corporation, Stanford University (Departments of Statistics and Applied Earth Sciences), University of British Columbia (Department of Statistics), University of Washington (Department of Statistics), and SIMS.

SIMS is a nonprofit corporation founded in 1973 to "foster the application of mathematics to societal problems". To achieve this objective, SIMS has organized appropriate research and education studies and conferences, with mathematics and statistics playing a principal role.

Support for SIMS' programs has been forthcoming from private foundations, industry, and government agencies. Governance of SIMS resides with a Board of Directors from academe, government/industry, and the private sector. SIMS' research results have been recorded in a series of Technical Reports, which have been distributed to those requesting them. Most of the Technical Reports have been published in scientific journals.

SIMS Technical Reports are available from: SIMS Publications
97 Parish Road South
New Canaan, Connecticut
06840

PRINCIPAL INVESTIGATORS

Donald L. Thomsen, Jr.
Societal Institute of the Mathematical Sciences

Naihua Duan
The Rand Corporation

A. John Petkau,
U. of British Columbia, Statistics

Peter Guttrop
U. of Washington, Statistics

Paul Switzer
Stanford University, Statistics

SIMS BOARD OF DIRECTORS

Joel E. Cohen
Rockefeller University

William T. Golden
New York, New York

Ettore F. Infante
University of Minnesota

William H. Kruskal
University of Chicago

Brockway McMillan
Bell Telephone Laboratories

Emanuel R. Piore
Rockefeller University

Howard Raiffa
Harvard University

Mina S. Rees
City University of New York

Fred S. Roberts
Rutgers University

I. Richard Savage
Yale University

Donald L. Thomsen, Jr.
Societal Institute of the Mathematical Sciences

Arthur C. Upton
New York University Medical Center
Bivariate Threshold Methods for Extremes
H. Joe, R.L. Smith, I. Weissman
University of British Columbia, University of Surrey, Technion University

SUMMARY

Extreme quantiles of the concentration of a pollutant are needed for comparison with air quality standards. Threshold methods have become widely used for estimating extreme quantiles or tails of distributions. In place of the traditional methods based on the extreme value limiting distributions (Gumbel 1958, Galambos 1987), they are based on fitting a probability model to the observations exceeding a high threshold or a small number of extreme order statistics. The Generalized Pareto Distribution is used as a model for the distribution of excesses over a high threshold, and may be given a justification analogous to classical extreme value theory.

Air quality standards exist for several pollutants, with “danger” levels being determined separately for each pollutant. However it may be possible that, in some regions, concentrations of some pollutants are highly dependent, even in the extremes. If this is so, the health and other effects of several pollutants all having high levels should be considered and air quality standards modified to take this into account.

Bivariate (and multivariate) extremes have not been widely studied, though there is extensive probability theory summarized in the books of Galambos (1987) and Resnick (1987). Statistical applications are in Pickands (1981), Tiago de Oliveira (1984) and Tawn (1988). These are all based on the traditional model for bivariate extremes, which concerns limiting joint distributions for the extreme values in each margin of a bivariate or multivariate sample.

This paper is concerned with bivariate threshold methods, in which a model is fitted to the joint distribution of a bivariate observation subject to some function of the components exceeding a high threshold. There are at least two reasons for pursuing such an approach. First, the increasingly extensive applications of threshold methods in univariate extreme value theory (for example, van Montfort and Witter 1986, Davison and Smith 1989) suggest that this approach is gradually taking over from the classical approach to univariate extremes; therefore, it makes sense to develop similar methods in the multivariate case also. Secondly, the traditional approach to multivariate extremes is based on componentwise maxima which is not always an appropriate approach, whereas the approach developed here is potentially applicable to answer any question about the joint distribution of two variables in the tail.

Applications of the bivariate threshold method include (i) estimating the probability that thresholds are simultaneously exceeded by 2 variables (say, concentrations of two different pollutants), (ii) inferences on dependence of two variables in the extremes, and (iii) estimation of a bivariate upper quantile curve (a curve where the probability of exceedance for both variables is a constant \( p \) for all points on the curve). Similarly to the univariate threshold method, for inference with extremes, the sample size must be at least in the order of hundreds.
The basic theory behind our approach is a point process model for extreme observations. Point process methods were introduced into univariate extreme value theory by Pickands (1971) and have been widely used to establish probabilistic properties; the monographs of Leadbetter, Lindgren and Rootzén (1983) and Resnick (1987) both make heavy use of these techniques. Exploratory data analysis techniques are given as an aid for deciding on a model. If a parametric model for the point process has been decided on, estimation of parameters and other inferences can proceed using the maximum likelihood method.

An example with sulphate/nitrate data from a major study (MAPS/PCN) of acid rain is used to illustrate the theory and techniques in this paper. A plot of the bivariate upper .01-quantile curve with pointwise radial standard error bands is included, as well as several plots and tables useful for checking of assumptions and models.
Bivariate Threshold Methods for Extremes
H. Joe, R.L. Smith, I. Weissman
University of British Columbia, University of Surrey, Technion University

ABSTRACT

An extension in the threshold method for extreme values is developed, to consider the joint distribution of extremes of two variables. The methodology is based on the point process representation of bivariate extremes developed by Pickands, de Haan and Resnick. Both parametric and nonparametric models are considered. The simplest case to handle is that in which both marginal distributions are known. For the more realistic case in which the marginal distributions are unknown, a mixed parametric-nonparametric method is proposed. The techniques are illustrated with data on sulphate and nitrate levels taken from a major study of acid rain.

Key words: Extreme values, threshold method, point process, Generalized Pareto distribution.

1. Introduction.

Threshold methods have become widely used for estimating extreme quantiles or tails of distributions. In place of the traditional methods based on the extreme value limiting distributions (Gumbel 1958, Galambos 1987), they are based on fitting a probability model to the observations exceeding a high threshold or a small number of extreme order statistics (Weissman 1978). The Generalized Pareto Distribution is used as a model for the distribution of excesses over a high threshold, and may be given a justification analogous to classical extreme value theory (Pickands 1975, Smith 1984, Joe 1987).

Bivariate extremes have been less widely studied, though there is extensive probability theory summarized in the books of Galambos (1987) and Resnick (1987). For statistical applications see Pickands (1981), Tiago de Oliveira (1984) and Tawn (1988). These are all based on the traditional model for bivariate extremes, which concerns limiting joint distributions for the extreme values in each margin of a bivariate or multivariate sample. A review of current developments is by Smith, Tawn and Yuen (1989).

The present paper is concerned with bivariate threshold methods, in which a model is fitted to the joint distribution of a bivariate observation subject to some function of the components exceeding a high threshold. There are at least two reasons for pursuing such an approach. First, the increasingly extensive applications of threshold methods in univariate extreme value theory (for example, van Montfort and Witter 1986, Davison and Smith 1989) suggest that this approach is gradually taking over from the classical approach to univariate extremes; therefore, it makes sense to develop similar methods in the multivariate case also. Secondly, the traditional approach to multivariate extremes is based on componentwise maxima which is not always an appropriate approach, whereas the approach developed here is potentially applicable to answer any question about the joint distribution of two variables in the tail. Applications of the bivariate threshold method include (i) estimating the probability that thresholds are simultaneously exceeded by 2 variables (say, concentrations of two different pollutants), (ii) inferences on dependence of two variables in the extremes,
and (iii) estimation of a bivariate upper quantile curve (a curve where the probability of exceedance for both variables is a constant $p$ for all points on the curve). Similarly to the univariate threshold method, for inference with extremes, the sample size must be at least in the order of hundreds.

The basic idea behind our approach is a point process model for extreme observations. Point process methods were introduced into univariate extreme value theory by Pickands (1971) and have been widely used to establish probabilistic properties; the monographs of Leadbetter, Lindgren and Rootzén (1983) and Resnick (1987) both make heavy use of these techniques. Statistical applications have not been so widely developed so far, though Smith (1989a) used these ideas for a complex example involving ozone and it seems likely that other applications will follow. In the case of multivariate extremes, a point process approach seems essential for a full understanding of the theory, and indeed forms the foundation for Chapter 5 of Resnick (1987), the most complete account published so far. It is therefore natural for our statistical development.

The point process theory we need is presented in Section 2. Section 3 consists of methods for assessing the amount of dependence in bivariate extremes and for deciding on the dependence structure in the bivariate extremes. Some models for the dependence structure are given, and some methods are described for nonparametric estimation and for checking the fit of a parametric model. Details of inference based on the likelihood function of a model are given in Section 4 and an example with sulphate/nitrate data in Section 5 illustrates the theory.

In some parallel work Coles and Tawn (1989) have also developed threshold models for bivariate extremes, and have proposed extensions to the general multivariate case.

2. Point process theory.

By way of introduction, we start with the univariate case. Let $X_1, X_2, \ldots$, be a sequence of independent and identically distributed random variables with distribution function $F$ and let $M_n = \max\{X_1, \ldots, X_n\}$. In accordance with classical extreme value theory, we assume there are constants $a_n, b_n$ such that $(M_n - b_n)/a_n$ converges in distribution to a nondegenerate distribution $G$. It follows that $G$ is a Generalized Extreme Value distribution function

$$G(x) = \exp[-(1 + \gamma(x - \mu)/\phi)^{-1/\gamma}], \quad \text{for } x : 1 + \gamma(x - \mu)/\phi > 0, \quad (2.1)$$

where $\phi > 0$ and $\gamma, \mu$ are reals. Now let $Y_i = (X_i - b_n)/a_n, i = 1, \ldots, n$. Let $P_n$ denote the point process on the real line with points at $Y_i$. Then $P_n$ converges weakly to a nonhomogeneous Poisson process under a topology which essentially excludes sets bordering on the lower endpoint, where points of $P_n$ will tend to cluster. The intensity measure of the limiting process is

$$\Lambda([x, \infty)) = [1 + \gamma(x - \mu)/\phi]^{-1/\gamma}, \quad \text{for } x : 1 + \gamma(x - \mu)/\phi > 0. \quad (2.2)$$

This is easily related to (2.1): the limiting probability that $(M_n - b_n)/a_n \leq x$ is just the probability that the limiting point process has no points in $(x, \infty)$, and this is given by (2.1).
The Generalized Pareto Distribution may also be derived within this framework. Restrict the process to an interval \((u, \infty)\) for fixed \(u\): then for an arbitrary point of the process the probability that it lies in the interval \((u, u + x)\) is

\[
\Lambda((u, x + u))/\Lambda((u, \infty)) = 1 - (1 + \gamma x/\sigma)^{-1/\gamma}
\]  

(2.3)

with \(\sigma = \phi + \gamma(u - \mu)\). The distribution function (2.3) is the Generalized Pareto Distribution, used in univariate extreme value theory to represent excesses over thresholds.

For the bivariate case, let \((X_{1i}, X_{2i})\), \(i = 1, 2, \ldots\), be independent and identically distributed pairs. In order to concentrate on the dependence structure, we assume that the \(X_{ji}\) have been transformed to \(Z_{ji}\) so that \(P(Z_{ji} > z) \sim 1/z, z \to \infty, j = 1, 2\) (details are given in Sections 3 and 4). In this case we may take \(b_n = 0, a_n = n\), whereupon the limiting nonhomogeneous Poisson process intensity measure of both sets of points \(\{Z_{11}/n, \ldots, Z_{1n}/n\}\) and \(\{Z_{21}/n, \ldots, Z_{2n}/n\}\) is \(\Lambda((x, \infty)) = 1/x, x > 0\). Now let \(P_n\) denote the set of pairs \(n^{-1}(Z_{11}, Z_{21}), \ldots, n^{-1}(Z_{1n}, Z_{2n})\), viewed as a point process in \(\mathbb{R}^2\). Then \(P_n\) converges weakly under a suitable topology to a nonhomogeneous Poisson process on \([0, \infty) \times [0, \infty)\) \(\setminus\{(0, 0)\}\) and the limiting intensity measure \(\Lambda\) satisfies

\[
\Lambda(B/m) = m\Lambda(B)
\]

(2.4)

for all measurable sets \(B\) which are bounded away from \((0, 0)\) and reals \(m \geq 1\); see Resnick (1987) for the details of this result. For a \(\Lambda\) satisfying (2.4), let \(\mu(z_1, z_2) = \Lambda\{[(0, z_1) \times (0, z_2)]^c\}\). Then \(\mu\) is homogeneous of order -1, that is, \(\mu(az_1, az_2) = a^{-1}\mu(z_1, z_2)\) for \(a > 0\).

Let \(\Lambda(B)\) be an intensity measure defined for all \(B\) which are bounded from 0. The relationship (2.4) implies that the measure has the polar representation

\[
\Lambda\{(dr, d\omega)\} = \frac{dr}{r^2} H(d\omega)
\]

(2.5)

with a "radial" variable \(r\) and an "angular" variable \(\omega\). For instance de Haan and Resnick (1977), de Haan (1985) and Resnick (1987) took \(r = (z_1^2 + z_2^2)^{1/2}\) and \(\omega = \arccos(z_1/r)\). We also use the alternative representation \(r = z_1 + z_2\) and \(\omega = z_1/r\), related to Pickands (1981). The advantage of this representation is that a number of algebraic expressions take on a particularly simple form (see also Smith, Tawn and Yuen 1989).

The relationship between \(\mu\) and \(H\) is

\[
\mu(z_1, z_2) = \int_0^1 \max[\omega z_1^{-1}, (1 - \omega)z_2^{-1}] H(d\omega) = (z_1^{-1} + z_2^{-1}) A(z_1/(z_1 + z_2)),
\]

where \(A(v) = \int_0^1 \max[\omega(1 - v), (1 - \omega)v] H(d\omega)\). The function \(A(v)\) is convex on \([0, 1]\), bounded above by 1 and below by \(\max\{v, 1 - v\}\). By setting first \(z_1\) and then \(z_2\) equal to \(\infty\), it is easy to see that \(H(d\omega)/2\) is a probability measure with mean 1/2. Let \(h(\omega)/2\) be the probability density if it exists, so that \(h = H'\) is given by \(-\partial^2 \mu/\partial z_1 \partial z_2\) evaluated at \(z_1 = \omega, z_2 = 1 - \omega\).
Some parametric families (models) for (2.4) are given in Section 3.4. A boundary of the class represented by (2.4) is where the extremes are independent; in this case \( \Lambda \{ (x, \infty) \times (y, \infty) \} = 1/xy \), \( x > 0, y > 0 \), all the points occur on the \( x \) and \( y \) axes, and the measure \( H \) is concentrated on \( \{ 0, 1 \} \).

3. Models and exploratory analysis.

3.1. Independent versus dependent bivariate extremes. With some bivariate extreme data, the first thing to do is to assess the amount of dependence in the tails. If the extremes of the two variables are independent, no modelling for (2.4) is needed. The extremes could be asymptotically independent even if \( X_{1i} \) and \( X_{2i} \) are strongly dependent, for example, if \( X_{1i} \) and \( X_{2i} \) have a bivariate normal distribution with correlation less than 1, as the following discussion will make clear.

A simple way to assess the dependence in extremes of two variables is to classify each variable into being above or below a threshold. A \( 2 \times 2 \) table results and measures of dependence such as Kendall’s \( \tau_b \) can be computed.

For a bivariate distribution \( F(x_1, x_2) \) with marginal distributions \( F_1(x_1) \) and \( F_2(x_2) \), let \( p_{11} = F(u_1, u_2) \), \( p_{12} = F(u_1) - p_{11} \), \( p_{21} = F_2(u_2) - p_{11} \), \( p_{22} = 1 - p_{11} - p_{12} - p_{21} \), where \( u_1, u_2 \) are thresholds. Hence the subscript 1 of \( p \) is for “low” and the subscript 2 is for “high” for the resulting ordinal categorical variables with cell probabilities \( p_{ij} \). For the \( 2 \times 2 \) table, Kendall’s \( \tau_b \) can be simplified to

\[
\tau_b = \frac{p_{22} - p_{21} + p_{12} + p_{11}}{\sqrt{p_{22}(1 - p_{22})}}
\]

where the + in the subscript indicates that a subscript has been summed over to get a marginal probability. If \( u_1 \) and \( u_2 \) each correspond to upper \( \pi \)-quantiles of the marginal distributions, the resulting \( \tau_b \) is \( \tau(\pi) = (p(\pi) - \pi^2)/(\pi - \pi^2) \), where \( p(\pi) \) is the probability of exceeding both thresholds. By a condition in Sibuya (1981), if \( F^n(a_{1n}x_1 + b_{1n}, a_{2n}x_2 + b_{2n}) \), for some \( a_{1n}, b_{1n}, a_{2n}, b_{2n} \), converges to a limiting bivariate extreme value distribution, then the limiting distribution is independent if \( \tau(\pi) \rightarrow 0 \) as \( \pi \rightarrow 0 \), and the limiting distribution has dependence if \( \tau(\pi) \rightarrow \tau > 0 \), as \( \pi \rightarrow 0 \).

Hence for bivariate data, an estimate of \( \tau(\pi) \) for small \( \pi \) will provide a diagnostic for independence versus dependence in the extremes. If the data can be assumed independent in the extremes, then inference for bivariate tail probabilities can be made via inference for the univariate extremes.

The application of this to some data is provided in Section 5. For comparison, we list here \( \tau(\pi) \) for a bivariate normal distribution with correlation 0.5. For \( \pi = .25, .20, .15, .10, .05 \), the corresponding \( \tau(\pi) \) are \( .308, .295, .276, .249, .204 \). As \( \pi \rightarrow 0 \) so \( \tau(\pi) \rightarrow 0 \) and this forms the basis for our earlier assertion that bivariate normal extremes are asymptotically independent. The same is true for any correlation between the components, provided it is not 1.

3.2. Transformations for margins. The model in Section 2 assumes that we can transform each margin to satisfy the condition \( \Pr(Z_j > z) \approx 1/z \) for \( j=1,2 \). There are some advantages in having a “quick and dirty” way to achieve this. A simple approach is to rank each variable and then apply the probability
integral transform to the ranks so that the marginal distributions of the resulting variables are Fréchet, $G(z) = \exp(-1/z), z > 0$. That is, $X_{ji}$ is transformed to $Z_{ji} = 1/\log[n/(R_{ji} - .5)]$, where $R_{ji}$ is the rank of $X_{ji}$ among $X_{j1}, \ldots, X_{jn}, j = 1, 2, \text{rank 1 being smallest}$. Asymptotically, $(Z_{i1}, Z_{2i}), i = 1, \ldots, n$, is like a random sample with Fréchet margins and also the $n^{-1}(Z_{i1}, Z_{2i})$ which are bounded away from 0 are a realization of a nonhomogeneous Poisson process with intensity measure $\Lambda$ satisfying (2.4).

An alternative transformation is based on assuming that the marginal distributions $F_1, F_2$ are known or have been estimated separately. This is pursued in Section 4, where the tail is assumed to be Generalized Pareto but the distribution otherwise known. Then, transformations to Fréchet margins are obtained by $Z_{ji} = -1/\log F_j(X_{ji})$.

3.3. Nonparametric estimates of the intensity measure. By (2.4), for a fixed $\beta > 0,$

$$\mu(y, \beta y) = y^{-1}\mu(1, \beta). \quad (3.3.1)$$

Therefore, (3.3.1) is the intensity measure associated with the limiting nonhomogeneous Poisson process for

$$\{n^{-1}\max(Z_{i1}, \beta^{-1}Z_{2i}) : i = 1, \ldots, n\}. \quad (3.3.2)$$

Since (3.3.2) is a finite sample, the asymptotics apply as an approximation only for the upper portion of (3.3.2). Asymptotically, if $T$ is large, then the points in (3.3.2) which exceed $T$ are conditionally like order statistics from the density $(y^{2T})^{-1}, y > T$. Hence Q-Q plots of the tail of (3.3.2) may be obtained to check whether the sample size and $T$ are large enough for the asymptotics to apply. Furthermore, the expected number of points from (3.3.2) which exceed $T$ is $T^{-1}\mu(1, \beta)$ and hence $A(v) = \mu(1, \beta)/(1 + \beta^{-1}), v = (1 + \beta)^{-1},$ can be estimated pointwise for $0 < v < 1$. This estimate is similar to that in Pickands (1981). The shape of $A(v)$ can help to decide on a parametric model for $\Lambda$. If $T = T[k]$ is the $k^{th}$ largest among (3.3.2), then an estimate of $\mu(1, \beta)$ is $kT$ and an estimate of $A(1/(1 + \beta))$ is $kT/(1 + \beta^{-1})$. This estimate is also a “maximum likelihood” estimate for fixed $\beta$. The argument is as follows. Let $Y_1, \ldots, Y_n$ be the reciprocals of the points in (3.3.2). The smaller $Y$'s are asymptotically a realisation of a homogeneous Poisson point process with intensity rate $\mu(1, \beta)$. From the smallest $k$ $Y$'s, the maximum likelihood estimate is $k/Y[k] = kT[k]$, where $Y[k]$ is the $k^{th}$ smallest among $Y_1, \ldots, Y_n$.

Another nonparametric estimate can be obtained through the polar representation of the intensity measure. For determining a parametric model for $\Lambda$ and for checking whether the sample size is large enough for the asymptotics to be a good approximation, a plot of $r_i = Z_{i1} + Z_{2i}$ versus $\omega_i = Z_{i1}/r_i$ for $r_i > r^*$ for a cutoff point $r^*$ can be used to check whether $r_i$ and $\omega_i$ appear independent as expected by (2.5). This plot will be referred to as the $r - \omega$ plot. If the independence assumption appears to be reasonable, the set

$$\{\omega_i : r_i > r^*\} \quad (3.3.3)$$

can be used to estimate the probability measure $H(dw)/2$, say by density estimation. This is similar in spirit to a suggestion of A.A. Balkema reported by de Haan (1985).
Furthermore, provided the cutoff point \( r^* \) is suitable, (3.3.3) can also be used for maximum likelihood with a parametric model for the probability density \( h(\omega)dw/2 = H(d\omega)/2 \).

3.4. Parametric models. In a detailed study of the traditional approach to bivariate extremes, Tawn (1988) identified some useful parametric models. He found that the logistic model, originally due to Gumbel, is the most useful of the previously existing models, and he also proposed an asymmetric form. The logistic model is easily translated into the present context, but Tawn's asymmetric model is not so easily handled: it turns out that it corresponds to a point process model with degeneracies on each axis, a feature not easily made compatible with real data. Therefore an alternative asymmetric model will be studied, which we call the bilogistic model. This has also been developed for the traditional approach (Smith 1989b), without being clearly either better or worse than Tawn's asymmetric logistic model.

For the logistic model, \( \mu(z_1, z_2; \alpha) = (z_1^{-1/\alpha} + z_2^{-1/\alpha})^\alpha \), where \( 0 < \alpha < 1 \), \( A(v; \alpha) = (v^{1/\alpha} + (1 - v)^{1/\alpha})^\alpha \), and \( 2h(\omega; \alpha) = (1 - \alpha)\omega^{-1}[(1 - \omega)^{-1 - 1/\alpha}(\omega^{-1/\alpha} + (1 - \omega)^{-1/\alpha})^{-2} \). The limiting cases \( \alpha = 1 \) and \( \alpha = 0 \) correspond respectively to independence and perfect dependence.

To develop the bilogistic model, first note that any function of the form

\[
\mu(z_1, z_2) = \int_0^1 \max\{z_1^{-1}g_1(\omega), z_2^{-1}g_2(\omega)\}d\omega,
\]

where \( g_1 \) and \( g_2 \) are probability density functions, has the required homogeneity property and also satisfies \( \mu(z, \infty) = \mu(\infty, z) = 1/z \). This is a special case of de Haan's (1984) representation. The bilogistic model with parameters \( 0 < \alpha, \beta < 1 \) has \( g_1(\omega) = (1 - \alpha)\omega^{-\alpha}, g_2(\omega) = (1 - \beta)(1 - \omega)^{-\beta}, 0 < \omega < 1 \). In this case

\[
2h(\omega; \alpha, \beta) = (1 - \alpha)(1 - \omega)^{-1}\omega^{-2}(1 - u)u^{-1}[\alpha(1 - u) + \beta u]^{-1}
\]

where \( u = u(\omega; \alpha, \beta) \) is the root of

\[
(1 - \alpha)(1 - \omega)(1 - u)^\beta - (1 - \beta)\omega u^\alpha = 0.
\]

From this \( A(\omega; \alpha, \beta) = (1 - \omega)u^{1-\alpha} + \omega(1 - u)^{1-\beta} \). When \( \alpha = \beta \), the bilogistic model reduces to the logistic model. In the bilogistic model it is possible to think of \( (\alpha + \beta)/2 \) as a dependence parameter, measuring the strength of dependence between the extremes of the two variables, and \( (\alpha - \beta) \) as an asymmetry parameter, the case \( \alpha - \beta = 0 \) (logistic) being one in which the two components are exchangeable.

4. Likelihood inference. In this section, the method of maximum likelihood is used to obtain estimates of parameters of a model, including parameters of the univariate margins to transform to the assumed univariate tail for the bivariate intensity measure. The transformations of the two variables come from assuming that the upper tails are Generalized Pareto with unknown parameters and that the remainder of the distributions are arbitrary but known. The latter assumption is reasonable if the data are quite dense except for the tails so that an empirical or smoothed empirical cumulative distribution function can be taken for the "known" part of the distribution. From the estimates of the parameters and their estimated covariance matrix, estimates of functions of the parameters can be obtained together with their standard errors. Of particular interest is the bivariate upper quantile curve; standard errors for this will be obtained pointwise in the radial direction.
As in Section 2, let \( x_i = (x_{1i}, x_{2i}) \), \( i = 1, \ldots, n \), denote the original data and let \( z_i = (z_{1i}, z_{2i}) \), \( i = 1, \ldots, n \), denote the transformed data which have the required tail frequency distribution. The \( x_i \) are independent and identically distributed (iid) realisations of a random vector \( X = (X_1, X_2) \) and the \( z_i \) are treated as iid realisations of a random vector \( Z \). If the \( n^{-1} z_i \) lie in a set \( B \) in which they are a realisation of a nonhomogeneous Poisson point process with measure \( \Lambda(\cdot; \theta) \) and intensity function \( \lambda(\cdot; \theta) \), where \( \theta \) is a parameter vector, then the likelihood for \( \theta \) is

\[
L(\theta) = \prod_{z_i/i/n \in B} \lambda(n^{-1} z_{1i}, n^{-1} z_{2i}; \theta) \exp\{-\Lambda(B; \theta)\}.
\]

Let \( \eta_1, \eta_2 \) be vectors of marginal parameters. Then we may write \( z_i = (t_1(x_{1i}; \eta_1), t_2(x_{2i}; \eta_2)), i = 1, \ldots, n \), where \( t_1, t_2 \) are strictly increasing transformations. The full likelihood for \( \theta, \eta_1, \eta_2 \) is then

\[
L(\theta, \eta_1, \eta_2) = \exp\{-\Lambda(B; \theta)\} \prod_{x_i/i/n \in B} \lambda(n^{-1} z_{1i}, n^{-1} z_{2i}; \theta) n^{-1} \frac{\partial t_1}{\partial x_1}(x_{1i}; \eta_1) n^{-1} \frac{\partial t_2}{\partial x_2}(x_{2i}; \eta_2).
\]  

(4.1)

The transformations \( t_1 \) and \( t_2 \) are derived as follows. Assume that the sample size \( n \) is large, and that \( X_1, X_2 \) have cumulative distribution functions (cdf's) known below thresholds \( u_1, u_2 \) respectively and conditionally Generalized Pareto with unknown parameters \( \eta_1 = (\gamma_1, \phi_1) \) and \( \eta_2 = (\gamma_2, \phi_2) \) above the thresholds. That is, for \( j = 1, 2 \), the cdf's are

\[
F_j(x; \eta_j) = \begin{cases} 
F_j(x; \eta_j), & x_j \leq u_j, \\
1 - (1 - F_j(u_j))(1 + \gamma_j[x_j - u_j]/\phi_j)^{-1/\gamma_j}, & x_j \geq u_j,
\end{cases}
\]

(4.2)

where \( y_+ = \max\{0, y\} \). The assumption on the tail comes from univariate extreme value theory. (The details below can be easily modified if the assumption of known cdf's below the thresholds is not made and there are additional parameters that are to be estimated. The assumption is made partly for simplicity of numerical computations and is reasonable if the data are "dense" below the thresholds. Note that the inferences are much more sensitive to the tails.) The transformation \( Z_j \) of \( X_j \) is such that \( Z_j \) has the Fréchet distribution \( G(z) = \exp(-1/z) \), \( z > 0 \). Note that \( G(z) \sim 1 - 1/z \) for \( z \) large. Therefore,

\[
Z_j = t_j(X_j) = [-\log F_j(X_j; \eta_j)]^{-1}, \quad X_j = F_j^{-1}(\exp(-1/Z_j); \eta_j)
\]

and for \( X_j \geq u_j, Z_j = [-\log\{1 - (1 - F_j(u_j))(1 + \gamma_j[(X_j - u_j)/\phi_j])^{-1/\gamma_j}\}]^{-1} \). For \( x_j > u_j, \frac{\partial t_j}{\partial x_j} \) evaluated at \( z = t_j(x_j) \) becomes \( \phi_j^{-1} x^2 (1 - e^{-1/x})^{1+\gamma_j} (1 - F_j(u_j))^{-1-\gamma_j} \). For (4.1), there is a simplification if \( B \) is chosen to be \([0, n^{-1} t_1(u_1)] \times [0, n^{-1} t_2(u_2)]\)\(^c\), but in general because the \( t_j \) depends on \( \eta_j \) for \( x_j > u_j \), (4.1) can be modified to

\[
L = \exp\{-\Lambda(B; \theta)\} \left[ \prod_{x_i/i/n \in B} \lambda(n^{-1} z_{1i}, n^{-1} z_{2i}; \theta) \right] \\
\left[ \prod_{x_i/i/n \in B, x_1 > u_1} n^{-1} \frac{\partial t_1}{\partial x_1}(x_{1i}; \eta_1) \right] \left[ \prod_{x_i/i/n \in B, x_2 > u_2} n^{-1} \frac{\partial t_2}{\partial x_2}(x_{2i}; \eta_2) \right].
\]  

(4.3)
The logarithm of (4.3) can be maximised using a quasi–Newton routine to get the maximum likelihood estimate of \((\theta, \eta_1, \eta_2)\) and the estimated covariance matrix (negative inverse Hessian). An initial guess for the maximum likelihood estimate can be obtained from separate univariate (Generalized Pareto) likelihoods for \(\eta_1, \eta_2\), and the likelihood

\[
\exp\{-\Lambda(B; \theta)\} \prod_{x_i \in B} \lambda(n^{-1} \tilde{z}_{1i}, n^{-1} \tilde{z}_{2i}; \theta)
\]

for \(\theta\), where \(\tilde{z}_{1i}, \tilde{z}_{2i}\) are transforms using the maximum likelihood estimates \(\hat{\eta}_1, \hat{\eta}_2\) from the univariate likelihoods. Alternatively, an initial guess for \(\theta\) can come from the maximum likelihood for (3.3.3) with the parametric family \(h(\omega; \theta)/2\) for the angles.

For a model such as the asymmetric logistic (referred to in Section 3.4) in which there is mass on the two axes, the likelihood (4.1) is sensitive to the transformations (at points near zero). The transformations above do not allow points on the axes, but asymptotically there would be some for such a model. Therefore in this case, (4.1) should be replaced by

\[
L(\theta, \eta_1, \eta_2) = \exp\{-\Lambda(B; \theta)\} \Lambda(B_1; \theta)^{m_1} \Lambda(B_2; \theta)^{m_2}
\]

\[
\prod_{x_i \in B_0} \lambda(n^{-1} z_{1i}, n^{-1} z_{2i}; \theta)n^{-1} \frac{\partial t_1}{\partial x_1}(x_{1i}; \eta_1)n^{-1} \frac{\partial t_2}{\partial x_2}(x_{2i}; \eta_2)
\]

where \(B_0, B_1, B_2\) is a partition of \(B\) with \(B_0\) bounded away from both axes, \(B_1\) includes a strip along the horizontal axis and \(B_2\) includes a strip along the vertical axis, and \(m_1, m_2\) are respectively the number of \(n^{-1} z_i\) in \(B_1, B_2\). In view of the extra complications this approach is not pursued further in the present paper.

We conclude this section with the upper \(p\)–quantile curve, where \(p\) is small. Let \(\tilde{F}(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2)\). The upper \(p\)–quantile curve is defined to be the set of \((x_1, x_2)\) for which \(\tilde{F}(x_1, x_2) = p\). With the nonhomogeneous Poisson point process models, \(\tilde{z}_1^{-1} + \tilde{z}_2^{-1} - \mu(\tilde{z}_1, \tilde{z}_2; \theta)\) is an estimate of \(\tilde{F}(x_1, x_2)\), where \(\tilde{z}_j = t_j(x_j; \hat{\eta}_j)\) and \(\hat{\eta}_j\) is the maximum likelihood estimate of \(\eta_j\).

If \(X_1 > 0, X_2 > 0\), we can use regular polar coordinates. For the equation,

\[
\tilde{z}_1^{-1} + \tilde{z}_2^{-1} - \mu(\tilde{z}_1, \tilde{z}_2; \theta) = p,
\]

fix \(w = \arctan(x_2/x_1)\) and solve for \(r = (x_1^2 + x_2^2)^{1/2} = r(w)\). That is, (4.4) becomes an equation in one variable. The vectors \(\partial r/\partial \theta, \partial r/\partial \eta_1, \partial r/\partial \eta_2\) evaluated at \((w, r(w), \hat{\theta}, \hat{\eta}_1, \hat{\eta}_2)\) can be computed numerically. From the covariance matrix for \((\hat{\theta}, \hat{\eta}_1, \hat{\eta}_2)\) and \(\partial r/\partial \theta, \partial r/\partial \eta_1, \partial r/\partial \eta_2\), a standard error for \(r(w)\) can be obtained (using the delta method). As \(w\) varies from 0 to \(\pi/2\) radians, the result will be a curve with pointwise standard errors in the radial direction.

5. Example with sulphate/nitrate data. We apply the methodology of the previous sections to some acid rain data from the MAP3S/PCN data base (Gentleman, Zidek and Olsen 1985). These data
consist of measurements taken at a wet deposition monitoring network at rainfall events over a period of six years. The network consists of 9 stations. Many variables are measured at each rainfall event and we will use the sulphate and nitrate concentrations here. From smoothing the time series, there does not appear to be a trend so we consider observations as an iid sample. For some stations, the sulphate and nitrate concentrations are almost independent in the extremes and for other stations there appears to be dependence according the technique with Kendall’s τ in Section 3.1. For the remainder of this section, we use station 65 (Penn State, PA) which has the strongest dependence in the bivariate extremes. There are n = 504 pairs after deleting the pairs for which at least one of the two concentrations is missing.

Table 1 shows approximate upper π-quantile for \( x_1 \) = sulphate concentration and \( x_2 \) = nitrate concentration, the proportion \( \hat{p}(\pi) \) exceeding both upper quantiles, \( \tau(\pi) \) and its standard error.

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( \hat{F}_1^{-1}(\pi) )</th>
<th>( \hat{F}_2^{-1}(\pi) )</th>
<th>( \hat{p}(\pi) )</th>
<th>( \tau(\pi) )</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>62.0</td>
<td>67.0</td>
<td>.131</td>
<td>.365</td>
<td>.084</td>
</tr>
<tr>
<td>.20</td>
<td>69.0</td>
<td>74.0</td>
<td>.0976</td>
<td>.358</td>
<td>.070</td>
</tr>
<tr>
<td>.15</td>
<td>81.0</td>
<td>84.0</td>
<td>.0657</td>
<td>.337</td>
<td>.058</td>
</tr>
<tr>
<td>.10</td>
<td>93.1</td>
<td>100.0</td>
<td>.0388</td>
<td>.330</td>
<td>.047</td>
</tr>
<tr>
<td>.05</td>
<td>110.0</td>
<td>150.0</td>
<td>.0179</td>
<td>.323</td>
<td>.095</td>
</tr>
</tbody>
</table>

The table suggests that \( \tau(\pi) \) is not converging towards 0 and we proceed with further analysis assuming the extremes are dependent.

Assuming \( \pi \) to be fixed and the marginal quantiles to be random, and assuming densities to exist, the asymptotic standard error for \( \tau(\pi) \) from using the delta method and the Bahadur representation (Serfling, 1980, p.91) is

\[
n^{-1} \left[ \left( \frac{\partial \hat{F}}{\partial x_1} \right)^2 \frac{\pi(1-\pi)}{f_1(q_1)} + \left( \frac{\partial \hat{F}}{\partial x_2} \right)^2 \frac{\pi(1-\pi)}{f_2(q_2)} + 2 \frac{\partial \hat{F}}{\partial x_1} \frac{\partial \hat{F}}{\partial x_2} \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \frac{p(\pi) - \pi^2}{f_1(q_1)f_2(q_2)} \right].
\]

(5.1)

where \( q_1 = F_1^{-1}(\pi) \), \( q_2 = F_2^{-1}(\pi) \), \( f_1, f_2 \) are the derivatives of \( F_1 \) and \( F_2 \). Finite difference approximations to the derivatives using points a distance \( \Delta \) on both sides of \( \hat{F}_1^{-1}(\pi) \) and \( \hat{F}_2^{-1}(\pi) \) were used, and standard errors in the above table come from the largest estimate of (5.1) from choosing \( \Delta = 2.5, 5, 7.5 \), although there was not much difference among these three \( \Delta \) values.

Now we go on to some exploratory analysis of the dependence structure for the bivariate extremes. Let \((Z_{1i}, Z_{2i})\) be as defined through the rank transformations in Section 3.2. (The transformations into \( Z_{1i} \) based on (4.2) leads to similar results.) Let \( r_i = Z_{1i} + Z_{2i} \) and let \( \omega_i = Z_{1i}/r_i \). Using a cutoff point of \( r^* = 12 \), the \( r - \omega \) plot with a log axis for \( r \) is in Figure 1. For the 84 pairs \((r_i, \omega_i)\) such that \( r_i > 12 \), Spearman’s rho is .0046, suggesting that the assumption of independence of \( r \) and \( \omega \) required for the asymptotic theory is satisfactory here.

A nonparametric density estimate for the \( \omega_i \) such that \( r_i > 12 \) is given in Figure 2. Superimposed are the parametric densities \( h(\omega)/2 \) from the logistic model and the bilogistic model (see Section 3.4) with the
maximum likelihood estimates for the parameters. From maximising the likelihood with the data satisfying (3.3.3), the estimates are \( \hat{\alpha} = .581 \) with a standard error of .026 for the logistic model, and the estimates are \( \hat{\alpha} = .365 \) with a standard error of .069 and \( \hat{\beta} = .716 \) with a standard error of .040 for the bilogistic model. The difference in log–likelihoods is 4.6. Both Figure 2 and the difference in log–likelihoods suggest that the bilogistic model is an improvement to the logistic model, and Figure 2 suggests that further improvements might be made with a model with more parameters.

Next we compare parametric and nonparametric estimates of \( A(v) \). The Q–Q plot suggested in Section 3.3 is equivalent to checking if \( T[i] \) is approximately constant for a fixed \( \beta \), where \( T[i] \) is the \( i^{th} \) largest among \( n^{-1} \max(Z_{1i}, Z_{2i}/\beta) \). For \( \beta = v^{-1} - 1 \) with \( v \) ranging from .1 to .9 in increments of .1, this seemed to be the case, that is, there were fluctuations but not a trend. In Table 2, the nonparametric estimates of \( A(v) \) is \( \sum_{k=30}^{50} kT_{[k]}/[21(1+\beta^{-1})] \).

<table>
<thead>
<tr>
<th>( v )</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonparametric</td>
<td>.912</td>
<td>.868</td>
<td>.839</td>
<td>.805</td>
<td>.774</td>
<td>.757</td>
<td>.763</td>
<td>.800</td>
<td>.875</td>
</tr>
<tr>
<td>logistic</td>
<td>.912</td>
<td>.842</td>
<td>.790</td>
<td>.759</td>
<td>.748</td>
<td>.759</td>
<td>.790</td>
<td>.842</td>
<td>.912</td>
</tr>
<tr>
<td>bilogistic</td>
<td>.922</td>
<td>.860</td>
<td>.811</td>
<td>.775</td>
<td>.754</td>
<td>.752</td>
<td>.773</td>
<td>.824</td>
<td>.903</td>
</tr>
</tbody>
</table>

The conclusion here is the same as from Figure 2. The bilogistic model is a better fit than the logistic model, but maybe it can be improved.

The two variables are quite dense below the upper 10th quantiles \((u_1 = 93, u_2 = 100)\) with a minimum value of 0. Therefore we may take \( F_j \) to be a smoothed empirical distribution function below \( u_j \); these smoothed empirical distributions are piecewise linear with piecewise constant densities on intervals of lengths 10 to 20. The tail above \( u_j \) is assumed to be Generalized Pareto with unknown parameters. Fitting the univariate tails separately leads to \( \hat{\gamma}_1 = .351, \hat{\phi}_1 = 25.5, \hat{\gamma}_2 = .366, \hat{\phi}_1 = 74.4 \) and Q–Q plots with these estimated parameters have approximately straight lines. The values of the \( \gamma \) parameters mean that the sulphate concentrations have a longer tail than exponential and the nitrate concentrations have a shorter tail than exponential. These are used together with the estimates of dependence parameters obtained from the rank transformation analysis as initial estimates for the likelihood in (4.3). With the choice of \( B = ([0, n^{-1}t_1(u_1)] \times [0, n^{-1}t_2(u_2)])^c \), the maximum likelihood estimates for the logistic model are

\[
\hat{\gamma}_1 = .481, \quad \hat{\phi}_1 = 31.4, \hat{\gamma}_2 = -.376, \quad \hat{\phi}_2 = 84.3, \quad \hat{\alpha} = .538.
\]

For the same \( B \), the maximum likelihood estimates for the bilogistic model are

\[
\hat{\gamma}_1 = .544, \quad \hat{\phi}_1 = 32.0, \hat{\gamma}_2 = -.379, \quad \hat{\phi}_2 = 83.1, \quad \hat{\alpha} = .334, \quad \hat{\beta} = .658.
\] (5.2)

The log–likelihoods are -1032.9 and -1028.0 respectively indicating that the bilogistic model with one extra parameter is a much better fit to the data. Other choices of \( u_1 \) and \( u_2 \) lead to similar results. The
corresponding estimated covariance matrix for the bilogistic model is

\[
\begin{bmatrix}
.039 & -.783 & -.0028 & .543 & -.0020 & .00034 \\
-.783 & 41.3 & -.048 & 9.10 & -.039 & -.059 \\
-.0028 & -.048 & .0112 & -.121 & .00028 & .00030 \\
.543 & 9.10 & -.121 & 17.0 & -.028 & -.091 \\
-.0020 & -.039 & -.00028 & -.028 & .0039 & -.0013 \\
.00034 & -.059 & .00030 & -.091 & -.0013 & .0017
\end{bmatrix}
\] (5.3)

From (5.2) and (5.3), the estimated bivariate upper .01 quantile curve in Cartesian and regular polar coordinates and the standard errors in the radial direction are given in Table 3 for selected angles. The plot with pointwise radial standard error bands is in Figure 3.

Table 3

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$w$</th>
<th>$r$</th>
<th>SE($r$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>239.7</td>
<td>24.1</td>
<td>.1</td>
<td>240.9</td>
<td>32.2</td>
</tr>
<tr>
<td>239.4</td>
<td>74.1</td>
<td>.3</td>
<td>250.6</td>
<td>33.5</td>
</tr>
<tr>
<td>236.6</td>
<td>129.3</td>
<td>.5</td>
<td>269.6</td>
<td>34.5</td>
</tr>
<tr>
<td>231.6</td>
<td>158.4</td>
<td>.6</td>
<td>280.6</td>
<td>32.2</td>
</tr>
<tr>
<td>217.7</td>
<td>183.4</td>
<td>.7</td>
<td>284.7</td>
<td>23.6</td>
</tr>
<tr>
<td>183.6</td>
<td>199.3</td>
<td>.8</td>
<td>277.8</td>
<td>16.2</td>
</tr>
<tr>
<td>165.5</td>
<td>208.6</td>
<td>.9</td>
<td>266.3</td>
<td>14.9</td>
</tr>
<tr>
<td>137.7</td>
<td>214.5</td>
<td>1.0</td>
<td>254.9</td>
<td>13.6</td>
</tr>
<tr>
<td>111.3</td>
<td>218.7</td>
<td>1.1</td>
<td>245.4</td>
<td>12.5</td>
</tr>
<tr>
<td>86.3</td>
<td>221.9</td>
<td>1.2</td>
<td>238.1</td>
<td>11.5</td>
</tr>
<tr>
<td>62.1</td>
<td>223.7</td>
<td>1.3</td>
<td>232.2</td>
<td>12.4</td>
</tr>
</tbody>
</table>

6. Conclusions. This paper has proposed threshold methods for bivariate data which are consistent with bivariate extreme value theory. It is obvious that the same ideas may be applied to the general multivariate case, but in this case there is a need for new parametric models for the resulting point process. Coles and Tawn (1989) have made some proposals here. On the statistical side, both nonparametric and parametric approaches have been considered, the parametric approach assuming the marginal distributions below the threshold to be known. There is scope for further investigation here. Finally, the methods have been applied to data on acid rain and the computation of bivariate quantiles has been discussed.

Acknowledgements. This research has been partially supported by an NSERC Canada grant and by the U.S. Environmental Protection Agency through a cooperative research agreement with SIMS. The second and third authors would like to thank the members of the Department of Statistics, University of British Columbia, for their hospitality. The authors thank Dr. Jonathan Tawn for pointing out an error in an earlier version.

References.


Figure 1. r-w plot
Figure 2: Density estimates for w
Figure 3: upper .01 quantile curve (bilogistic model)