J. R. Statist. Soc. B (1991)
53, *No.* 2, *pp.* 377–392

Modelling Extreme Multivariate Events

By STUART G. COLES and JONATHAN A. TAWN†

University of Sheffield, UK

[Received June 1989. Final revision January 1990]

SUMMARY

The classical treatment of multivariate extreme values is through componentwise ordering, though in practice most interest is in actual extreme events. Here the point process of observations which are extreme in at least one component is considered. Parametric models for the dependence between components must satisfy certain constraints. Two new techniques for generating such models are presented. Aspects of the statistical estimation of the resulting models are discussed and are illustrated with an application to oceanographic data.

Keywords: EXTREME VALUE THEORY; GENERALIZED PARETO DISTRIBUTION; MAXIMUM LIKELIHOOD; MULTIVARIATE ORDERING; NON-HOMOGENEOUS POISSON PROCESS; SIMPLEX MEASURES

1. INTRODUCTION

Recently there have been significant advances in the modelling procedures available for univariate extreme values. Two approaches, in particular, have been advocated in preference to the classical procedure of fitting a generalized extreme value distribution to the annual maxima of a series. One approach is based on the asymptotic joint distribution of a fixed number of extreme order statistics (Weissman, 1978; Smith, 1986; Tawn, 1988a). An alternative method is to model independent exceedances of a suitably high threshold using a generalized Pareto distribution (GPD) (Pickands, 1975; Davison and Smith, 1990). Through the inclusion of additional relevant data both procedures lead to greater estimation precision than the classical approach. Here our aim is to develop procedures which offer similar improvements to the analysis of multivariate extreme value data.

Problems concerning environmental extremes are often multivariate in character. An example is wind speed data where maximum hourly gusts, maximum hourly mean speeds and the dependence between them are relevant to building safety. Problems involving spatial dependence are also of much interest to engineers: for example, floods may occur at several sites along a coastline, or at various rain-gauges in a national network. In such situations knowledge of the spatial dependence of extremes is essential for regional risk assessment. Correct modelling of spatial dependence also leads to greater estimation precision of marginal parameters.

To date, the analysis of multivariate extreme values has been based on a componentwise ordering: for an independent and identically distributed (IID) vector series $(Y_{i,1}, \ldots, Y_{i,p})$, $i = 1, \ldots, n$, the vector of componentwise maxima, $\mathbf{M}_n$, is given by $\mathbf{M}_n =$

†*Address for correspondence*: Department of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, UK.

 0035–9246/91/53377 $2.00

$(M_{n,1}, \ldots, M_{n,p})$ where $M_{n,j} = \max\{Y_{1,j}, \ldots, Y_{n,j}\}$. Under general conditions $\mathbf{M}_n$, suitably normalized, converges in distribution to a member of the class of multivariate extreme value distributions. These distributions have received theoretical consideration (de Haan and Resnick, 1977; Resnick, 1987), and recent papers have explored their statistical application (Tawn, 1988b, 1990; Smith *et al.*, 1990). In applications $\mathbf{M}_n$ is generally taken as the vector of annual maxima for the relevant variables, although $\mathbf{M}_n$ may not correspond to an observed event. As with the classical univariate approach this procedure is wasteful of data. A further weakness with existing applications is that the margins are estimated as a preliminary step to estimating the dependence structure so that any potential for improving marginal estimation precision by incorporating dependence is lost.

To develop a procedure which utilizes more of the available data, and permits simultaneous estimation of the marginal and dependence structures, we exploit a representation in de Haan (1985). This is based on the limiting point process of actual observations which are extreme in at least one margin, giving a multivariate analogue of the GPD, and has the advantage that it does not require multivariate ordering (Barnett, 1976). In Section 2 the point process representation is reviewed and its connections with componentwise ordering and convex hulls are examined. With standardized margins the limiting point process is a non-homogeneous Poisson process whose intensity is characterized by a constrained measure on the unit simplex. This measure embodies the dependence structure of large observations. Two new techniques for generating suitably constrained parametric models for the measure are presented and demonstrated in Sections 3 and 4 respectively. Aspects of the statistical estimation of these models are discussed in Section 5 and are illustrated, in Section 6, with a trivariate application to oceanographic data.

2. POINT PROCESS THEORY OF MULTIVARIATE EXTREMES

In essence our approach follows de Haan and Resnick (1977), de Haan (1985) and Resnick (1987). Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of IID random vectors on $\mathbf{R}^p_+$ whose distribution function F is in the domain of attraction of a multivariate extreme value distribution G. We suppose that the marginal components are identically distributed with a unit Fréchet distribution: $\exp(-1/x)$, $x > 0$. This assumption is not restrictive because a suitable transformation can be applied otherwise.

Consider the process of points P_n, on $\mathbf{R}^p_+$, where $P_n = \{n^{-1}\mathbf{X}_i: i = 1, \ldots, n\}$. The choice of divisor n to normalize the $\mathbf{X}$ variables results from the max-stability property of unit Fréchet variables. Then P_n converges in distribution to a non-homogeneous Poisson process P on $\mathbf{R}^p_+ \setminus \{\mathbf{0}\}$. Defining pseudoradial and angular coordinates

$$r_i = \sum_{j=1}^{p} X_{i,j}/n, \qquad w_{i,j} = X_{i,j}/nr, \qquad i=1, \ldots, n, \quad j=1, \ldots, p, \tag{2.1}$$

where $X_{i,j}$ is the jth component of $\mathbf{X}_i$, the intensity measure μ of the limiting process P then satisfies

$$\mu(\mathrm{d}r \times \mathrm{d}\mathbf{w}) = \frac{\mathrm{d}r}{r^2}\, \mathrm{d}H(\mathbf{w}). \tag{2.2}$$

Here H is a positive finite measure on the $(p-1)$-dimensional unit simplex,

$$S_p = \{(w_1, \ldots, w_p): \sum_{j=1}^{p} w_j = 1, \quad w_j \geqslant 0, \quad j=1, \ldots, p\}.$$

By equation (2.2) μ factorizes into a known function of the radial component and a measure H of the angular components, which describes the dependence structure of all extreme observations. Although an arbitrary finite positive measure, H must satisfy

$$\int_{S_p} w_j \, \mathrm{d}H(\mathbf{w}) = 1, \qquad j=1, \ldots, p, \tag{2.3}$$

in order that the margins have the correct form. The only constraints on H are given by equation (2.3) so no finite parameterization exists for this measure.

If the effects of a storm can be thought of as additive representation (2.1) has a simple interpretation: r_i is a standardized measure of the strength of the ith spatial storm over the sites and $w_{i,j}$ is the relative strength on this storm at site j. This formulation differs from that of de Haan and Resnick who transform to standard polar coordinates.

The representation for limiting distributions of componentwise maxima is a consequence of the limiting Poisson process with intensity (2.2). This is seen by taking $A = \mathbf{R}_+^p \setminus \{(0,x_1) \times \ldots \times (0,x_p)\}$. Then, as $n \to \infty$, $\Pr(n^{-1}\mathbf{X}_i \notin A, i=1, \ldots, n) \to \exp\{-\mu(A)\}$, where

$$\begin{aligned}\mu(A) &= \int_A \frac{\mathrm{d}r}{r^2} \, \mathrm{d}H(\mathbf{w}) \\ &= \int_{S_p} \int_{r=\min\limits_{1\leqslant j\leqslant p}(x_j/w_j)}^{\infty} \frac{\mathrm{d}r}{r^2} \, \mathrm{d}H(\mathbf{w}) \\ &= \int_{S_p} \max_{1\leqslant j\leqslant p}(w_j/x_j) \, \mathrm{d}H(\mathbf{w}).\end{aligned}$$

As $\Pr(n^{-1}\mathbf{X}_i \notin A, i=1, \ldots, n) = \Pr(n^{-1}M_{j,n} \leqslant x_j, j=1, \ldots, p)$, any limit distribution of normalized componentwise maxima, with unit Fréchet margins, is of the form $G(\mathbf{x}) = \exp\{-V(\mathbf{x})\}$ where

$$V(\mathbf{x}) = \int_{S_p} \max_{1\leqslant j\leqslant p}(w_j/x_j) \, \mathrm{d}H(\mathbf{w}) \tag{2.4}$$

for some measure H satisfying constraints (2.3). This is Pickands's (1981) representation theorem for multivariate extreme value distributions with unit Fréchet margins.

Following de Haan and Resnick (1977) we call V the exponent measure. Pickands's statement of equation (2.4) was given for unit exponential margins and in terms of a dependence function related to V. For technical reasons we work in terms of the exponent measure instead of Pickands's dependence function; see Section 3. Furthermore, except for the bivariate case, the dependence function has no concise expression for unit Fréchet margins.

Various other probabilities of interest can also be calculated from equation (2.2). For example, the region A could be taken to be a convex hull in $\mathbf{R}_+^p$. In some cases closed form expressions can be found: an example is the region $A = \{\mathbf{u} \in \mathbf{R}_+^p: \Sigma_{j=1}^p u_j/x_j \geqslant 1\}$, as

$$\begin{aligned}\mu(A) &= \int_{S_p} \int_{r=(\Sigma_{j=1}^p w_j/x_j)^{-1}}^{\infty} \frac{\mathrm{d}r}{r^2} \,\mathrm{d}H(\mathbf{w}) \\ &= \int_{S_p} \left(\sum_{j=1}^p w_j/x_j \right) \mathrm{d}H(\mathbf{w}) \\ &= \sum_{j=1}^p x_j^{-1}.\end{aligned}$$

For more complicated regions numerical integration of the measure over the region will generally be necessary.

3. TECHNIQUES FOR MODELLING THE MEASURE

A simple nonparametric procedure for modelling the measure function H is described by de Haan (1985). For simultaneous estimation of marginal and dependence structures parametric procedures are preferable. Thus we concentrate on developing suitably flexible parametric models for the measure H. Parametric measures on the simplex have been developed for the study of compositional data (Aitchison, 1986), but these models are not suitable for the present application because of constraints (2.3). Therefore in this section we develop techniques which considerably simplify the non-trivial problem of generating parametric measures which satisfy constraints (2.3).

3.1. *Relating Exponent Measure V to Measure H*

We restrict ourselves to the class of models for which the exponent measure V is differentiable; thus H has densities on the interior, and on each of the lower dimensional boundaries, of the unit simplex S_p. For each $j = 1, \ldots, p$, let $c = \{i_1, \ldots, i_j\}$ be an index variable over the subsets, of size j, of the set $c_p = \{1, \ldots, p\}$ and let $S_{j,c} = \{\mathbf{w} \in S_p: w_k = 0, k \notin c\}$. For each c with $|c| = j$, $S_{j,c}$ is isomorphic to the $(j-1)$-dimensional unit simplex S_j. The measure has a density $h_{j,c}$ on each of the subspaces $S_{j,c}$. Where convenient, the domain of $h_{j,c}$ will be taken as either $S_{j,c}$ or S_j. The density $h_{j,c}$ describes the dependence structure for events which are extreme only in the components $c = \{i_1, \ldots, i_j\}$. Such a hierarchy of densities is required to describe dependence when the marginal components need not all be extreme simultaneously.

Theorem 1. Let V and H be the exponent measure and measure function defined by equation (2.4), and let $h_{j,c}$ be the class of densities of H defined previously. Then for $c = \{i_1, \ldots, i_m\}$

$$\frac{\partial V}{\partial x_{i_1} \ldots \partial x_{i_m}} = -\left(\sum_{j=1}^m x_{i_j} \right)^{-(m+1)} h_{m,c}\left(\frac{x_{i_1}}{\Sigma x_{i_j}}, \ldots, \frac{x_{i_m}}{\Sigma x_{i_j}} \right) \tag{3.1}$$

on $\{\mathbf{x} \in \mathbf{R}_+^p : x_r = 0 \text{ if } r \notin c\}$.

A proof of this result is given in Appendix A.

In the bivariate case this result reduces to that obtained by Pickands (1981). The importance of theorem 1 is that densities of all orders for the measure H may be obtained for any closed form multivariate extreme value distribution. For example, consider Gumbel's (1960) multivariate logistic model with exponent measure

$$V(\mathbf{x}) = \left(\sum_{j=1}^{p} x_j^{-r} \right)^{1/r}, \qquad r > 1. \tag{3.2}$$

Applying equation (3.1) gives $h_{j,c} \equiv 0, j < p$, and

$$h_{p,c_p}(\mathbf{w}) = \left\{ \prod_{j=1}^{p-1} (jr-1) \right\} \left(\prod_{j=1}^{p} w_j \right)^{-(r+1)} \left(\sum_{j=1}^{p} w_j^{-r} \right)^{1/r-p}, \qquad w \in S_p.$$

In contrast, if $r=1$ in equation (3.2), so that the variables are independent, then $h_{1,(i)} \equiv 1$ for all i and $h_{j,c} \equiv 0$ for all $j > 1$, so the measure places all mass on the vertices of S_p. The drawback to the use of theorem 1 is that it can be applied only to multivariate extreme value distributions, of which very few have been explicitly obtained; see Tawn (1990) and Joe (1989).

3.2. *Parametric Models for the Measure by Transformation*

An alternative technique for generating parametric measures is motivated by Pickands's representation which, in its full generality, characterizes multivariate extreme value distributions with exponential margins but arbitrary means. Equivalently, if $\mathbf{X}'$ has unit Fréchet margins, with ith-component scale parameter equal to m_i, then $\mathbf{X}'$ has a multivariate extreme value distribution if and only if its joint distribution function can be expressed as

$$G_{\mathbf{X}'}(\mathbf{x}') = \exp\left\{ - \int_{S_p} \max_{1 \leqslant j \leqslant p} \left(\frac{u_j}{x_j'} \right) \, \mathrm{d}H^*(\mathbf{u}) \right\},$$

where $\int_{S_p} u_j \, \mathrm{d}H^*(\mathbf{u}) = m_j$ and H^* is a positive finite measure on S_p. Transformation of $\mathbf{X}'$ to a random vector $\mathbf{X}$ with unit Fréchet margins, $X_j = X_j'/m_j, j = 1, \ldots, p$, gives a representation for the distribution function of $\mathbf{X}$ in the form

$$G_{\mathbf{X}}(\mathbf{x}) = \exp\left\{ - \int_{S_p} \max_{1 \leqslant j \leqslant p} \left(\frac{u_j}{m_j x_j} \right) \, \mathrm{d}H^*(\mathbf{u}) \right\}. \tag{3.3}$$

Thus equation (3.3) is a representation for multivariate extreme value distributions, with unit Fréchet margins, in terms of a finite measure H^* whose only constraint is to be positive over S_p. The equivalence of equations (3.3) and (2.4) is exploited in theorem 2 to show that any positive function h^* on S_p, corresponding to the density of H^*, may be transformed into a measure density h_{p,c_p} which satisfies constraints (2.3). We adopt the notation $h \equiv h_{p,c_p}$ if all the mass is in the interior of the simplex. Also, let $\mathbf{m} \cdot \mathbf{w}$ denote $\Sigma m_i w_i$, the scalar product of $\mathbf{m}$ and $\mathbf{w}$.

Theorem 2. If h^* is any positive function on S_p, with finite first moments, then

$$h(\mathbf{w}) = (\mathbf{m}\cdot\mathbf{w})^{-(p+1)} \prod_{j=1}^{p} m_j h^*\left(\frac{m_1 w_1}{\mathbf{m}\cdot\mathbf{w}}, \ldots, \frac{m_p w_p}{\mathbf{m}\cdot\mathbf{w}}\right), \tag{3.4}$$

where

$$m_j = \int_{S_p} u_j h^*(\mathbf{u})\, \mathrm{d}\mathbf{u}, \qquad j = 1, \ldots, p, \tag{3.5}$$

satisfies constraints (2.3) and is therefore the density of a valid measure function H.

This result is also proved in Appendix A.

Theorem 2 can be used to generate a rich class of parametric models for the measure on the interior of S_p. Though not stated here we have generalized this result to generate models with mass on the boundaries of S_p. The only practical restriction in each case is the necessity to calculate the integrals (3.5), either by numerical integration or, preferably, analytically. Take for example

$$h^*(\mathbf{w}) = \left\{\prod_{j=1}^{p} \Gamma(\alpha_j)\right\}^{-1} \Gamma(\boldsymbol{\alpha}\cdot\mathbf{1}) \prod_{j=1}^{p} w_j^{\alpha_j - 1}, \qquad \alpha_j > 0,\ j = 1, \ldots, p, \quad \mathbf{w} \in S_p.$$

By equation (3.5) $m_j = \alpha_j/(\boldsymbol{\alpha}\cdot\mathbf{1})$, and from equation (3.4) it follows that

$$h(\mathbf{w}) = \prod_{j=1}^{p} \frac{\alpha_j}{\Gamma(\alpha_j)} \frac{\Gamma(\boldsymbol{\alpha}\cdot\mathbf{1}+1)}{(\boldsymbol{\alpha}\cdot\mathbf{w})^{p+1}} \prod_{j=1}^{p} \left(\frac{\alpha_j w_j}{\boldsymbol{\alpha}\cdot\mathbf{w}}\right)^{\alpha_j - 1}, \qquad \mathbf{w} \in S_p, \tag{3.6}$$

is a valid measure density. This model will be discussed in Section 4.

4. MODELS FOR THE MEASURE

As no finite parametric family exists we apply the results of Section 3 to develop a few flexible families of models which cover a broad range of possible structures.

4.1. *Asymmetric Logistic Model*

Tawn (1990) developed the asymmetric logistic model as a multivariate extreme value distribution with much flexibility and capacity for physical interpretation. The joint distribution function of this model has the form

$$G(\mathbf{x}) = \exp\left[-\sum_{c\in C}\left\{\sum_{i\in c} (\theta_{i,c}/x_i)^{r_c}\right\}^{1/r_c}\right]$$

where C is the set of all non-empty subsets of $\{1, \ldots, p\}$ and the parameters are constrained by $r_c \geqslant 1$ for all $c \in C$, $\theta_{i,c} = 0$ if $i \notin c$, $\theta_{i,c} \geqslant 0$, $i = 1, \ldots, p$ and $\Sigma_{c\in C}\, \theta_{i,c} = 1$. Applying theorem 1 gives the class of measure densities, for $\mathbf{w} \in S_{j,c}$,

$$h_{j,c}(\mathbf{w}) = \left\{\prod_{k=1}^{j-1} (kr_c - 1)\right\} \left(\prod_{i\in c} \theta_{i,c}\right)^{r_c} \left(\prod_{i\in c} w_i\right)^{-(r_c+1)} \left\{\sum_{i\in c} \left(\frac{\theta_{i,c}}{w_i}\right)^{r_c}\right\}^{1/r_c - j}. \tag{4.1}$$

This model has a density on S_p, and on each lower dimensional boundary of S_p. Each density is a form of weighted logistic density (3.2). A special case of equation (4.1), for $\theta_{i,c_p}=1$, $i=1, \ldots, p$ and $r_{c_p}=r$, is the logistic model. Further properties of the model are discussed in Tawn (1990).

4.2. *Negative Asymmetric Logistic Model*

A model obtained by Joe (1989), with similar structure to the asymmetric logistic model, has distribution function

$$G(\mathbf{x}) = \exp\left[-\sum_{j=1}^{p} \frac{1}{x_i} + \sum_{c\in C:|c|\geqslant 2} (-1)^{|c|} \left\{ \sum_{i\in c} \left(\frac{\theta_{i,c}}{x_i}\right)^{r_c} \right\}^{1/r_c} \right]$$

with parameter constraints given by $r_c \leqslant 0$ for all $c\in C$, $\theta_{i,c}=0$ if $i\notin c$, $\theta_{i,c}\geqslant 0$ for all $c\in C$, $\Sigma_{c\in C}(-1)^{|c|}\theta_{i,c}\leqslant 1$. Again by theorem 1, for $\mathbf{w}\in S_{j,c}$,

$$h_{j,c}(\mathbf{w}) = \sum_{\substack{d\in C:\\ c\subset d}} (-1)^{|d|} \left\{ \prod_{k=1}^{j-1} (1-kr_d) \right\} \left(\prod_{i\in c} \theta_{i,d} \right)^{r_d} \left(\prod_{i\in c} w_i \right)^{-(r_d+1)} \left\{ \sum_{i\in c} \left(\frac{\theta_{i,d}}{w_i} \right)^{r_d} \right\}^{1/r_d - j}. \tag{4.2}$$

Because of the similarity of equations (4.1) and (4.2) we term this model the negative asymmetric logistic model.

4.3. *Dirichlet Model*

In Section 3, by application of theorem 2, measure density (3.6) was obtained. This model has the important feature of being asymmetric, allowing for non-exchangeability of the variables, yet having its mass confined to the interior of S_p. For the symmetric model with $\alpha = \alpha_1 = \ldots = \alpha_p$ both total independence and complete dependence are attained as limiting cases by taking $\alpha\to 0$ and $\alpha\to\infty$ respectively.

For $p=2$ density (3.6) gives a new bivariate extreme value distribution. The exponent measure with unit Fréchet margins is

$$V(x_1, x_2) = \frac{1}{x_1}\left\{1-\mathrm{Be}\left(\alpha_1+1, \alpha_2; \frac{\alpha_1 x_1}{\alpha\cdot\mathbf{x}}\right)\right\} + \frac{1}{x_2}\,\mathrm{Be}\left(\alpha_1, \alpha_2+1; \frac{\alpha_1 x_1}{\alpha\cdot\mathbf{x}}\right)$$

where $\alpha_1, \alpha_2 > 0$, $\alpha = (\alpha_1, \alpha_2)'$ and

$$\mathrm{Be}(\alpha_1, \alpha_2; u) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)} \int_0^u w^{\alpha_1-1}(1-w)^{\alpha_2-1}\,\mathrm{d}w,$$

a normalized incomplete beta function. The extension to general multivariate extreme value distributions is complicated although numerically feasible.

4.4. *Bilogistic Model*

Proposed by Smith (1990), the bilogistic model is an asymmetric generalization of the logistic model. The model is motivated by the representation for max-stable processes (de Haan, 1984) and has exponent measure of the form

$$V(x_1, x_2) = \int_0^1 \max\left\{\frac{(1-\alpha)s^{-\alpha}}{x_1}, \frac{(1-\beta)(1-s)^{-\beta}}{x_2}\right\} \mathrm{d}s \qquad (0 \leqslant \alpha \leqslant 1, \quad 0 \leqslant \beta \leqslant 1).$$

Even in this bivariate case no closed form expression can be obtained for the measure density which must be calculated numerically from the mixed derivative of V; see equation (3.1).

4.5. *Nested Logistic Model*

Generalizing the logistic model, McFadden (1978) developed a class of models which involve hierarchical dependence. Tawn (1990) described a physical motivation for these. A special trivariate case of this model has exponent measure $V(\mathbf{x}) = \{(x_1^{-r} + x_2^{-r})^{s/r} + x_3^{-s}\}^{1/s}$, $r \geqslant s \geqslant 1$. The pair $(1, 2)$ has logistic dependence, parameter r, whereas both other pairs have logistic dependence, parameter s; hence $(1, 2)$ has greatest dependence. The trivariate logistic model is obtained when $s = r$. By equation (3.1) the corresponding measure is easily obtained. It places all mass in the interior of S_3 but is slightly unwieldy. Obvious counterparts of this model in higher dimensions, and with higher level hierarchical dependence, are not discussed here.

4.6. *Time Series Logistic Model*

A quite different approach to generating parametric measures uses the theory of domains of attraction for multivariate extreme densities. The technique is illustrated by the following example. Let $X_1, \ldots, X_p$ be a first-order Markov process representing observations of a propagating sea storm at sites ordered along a coast. Suppose that the pairs (X_j, X_{j+1}) follow a bivariate extreme value distribution with unit Fréchet margins and logistic dependence structure, parameter r_j, whose joint and marginal densities we denote by f_j and f respectively. Then the joint density of $X_1, \ldots, X_p$ is

$$g(x_1, \ldots, x_p) = f(x_1) \prod_{j=1}^{p-1} \frac{f_j(x_j, x_{j+1})}{f(x_j)}.$$

Adapting the arguments of Resnick (1987), Smith (1989) and equation (3.1) gives

$$\lim_{t\to\infty} \{t^{p+1} g(tx_1, \ldots, tx_p)\} = (\mathbf{x}\cdot\mathbf{1})^{-(p+1)} h\{x_1/(\mathbf{x}\cdot\mathbf{1}), \ldots, x_p/(\mathbf{x}\cdot\mathbf{1})\}.$$

Hence

$$h(\mathbf{w}) = \frac{1}{w_1^2} \prod_{j=1}^{p-1} \frac{(r_j - 1) w_j^2}{(w_j w_{j+1})^{r_j+1}} (w_j^{-r_j} + w_{j+1}^{-r_j})^{1/r_j - 2}, \qquad \mathbf{w} \in S_p, \tag{4.3}$$

for $r_j \geqslant 1$. Here dependence decays with lag. An extension of this model, not examined here, is based on a higher order Markov sequence with the associated joint density of consecutive values taken as multivariate extreme value with unit Fréchet margins.

5. ESTIMATION

A natural approach to fitting the models of Section 4 is maximum likelihood applied to the likelihood function of the limiting Poisson process of Section 2. Initially we derive the likelihood for $\mathbf{X}_1, \ldots, \mathbf{X}_n$, a sequence of IID random vectors, with each marginal component having a unit Fréchet distribution. By the results of Section 2, for large n, the points of $P_n = \{n^{-1}\mathbf{X}_i: i=1, \ldots, n\}$ in a region A, bounded away from $\mathbf{0}$ by a distance dependent on the rate of convergence, are approximately a non-homogeneous Poisson process with intensity satisfying equation (2.2). Take $\{n^{-1}\mathbf{X}_i, i=1, \ldots, n_A\}$ to be the points of P_n in A. The likelihood over A, L_A, is

$$L_A(\theta; \{n^{-1}\mathbf{X}_i\}) = \exp\{-\mu(A)\} \prod_{i=1}^{n_A} \mu(\mathrm{d}r_i \times \mathrm{d}\mathbf{w}_i) \tag{5.1}$$

where θ are the measure parameters and r_i and $\mathbf{w}_i$ are defined by equations (2.1).

More generally let $\{\mathbf{Y}_i, i=1, \ldots, n\}$ be IID random vectors on $\mathbf{R}^p$. Simultaneous estimation of the marginal and dependence parameters requires an appropriate choice for the region A and a transformation to unit Fréchet margins to be included in equation (5.1). The region $A = \mathbf{R}_+^p \setminus \{(0, v_1) \times \ldots \times (0, v_p)\}$, with thresholds $v_1, \ldots, v_p$, contains all observations which are large in at least one margin. This ensures that the set of points in A is invariant to the marginal transformations (5.3). The marginal transformations above a high threshold, u_j, are determined by the conditional distribution of threshold exceedances. Pickands (1975) showed that these have a GPD form: $\Pr(Y_j > y \mid Y_j > u_j) = \{1 - k_j(y-u_j)/\sigma_j\}^{1/k_j}$, $\sigma_j > 0$, $k_j(y-u_j)/\sigma_j < 1$. Thus for $Y_j > u_j$

$$\Pr(Y_j > y) = p_j\{1 - k_j(y-u_j)/\sigma_j\}^{1/k_j}. \tag{5.2}$$

Here $p_j = \Pr(Y_j > u_j)$ is estimated as the proportion of points exceeding u_j.

Marginally points below the threshold are relatively dense, so we transform these components using the empirical distribution function. Hence,

$$X_j(Y_j) = \begin{cases} -(\log[1 - p_j\{1 - k_j(Y_j-u_j)/\sigma_j\}^{1/k_j}])^{-1} & \text{if } Y_j > u_j, \\ -[\log\{R(Y_j)/(n+1)\}]^{-1} & \text{if } Y_j \leqslant u_j \end{cases} \tag{5.3}$$

has a unit Fréchet distribution where $R(Y_j)$ denotes the rank of Y_j. Thus the thresholds for the limiting process are given by $v_j = n^{-1}X_j(u_j)$, though checks are required to ensure that these are sufficiently high for equation (2.2) to be valid in A.

Incorporating equations (2.2) and (5.3) into likelihood function (5.1) gives, for Cartesian components, the likelihood function $L_A(\theta, \sigma, \mathbf{k}; \{\mathbf{Y}_i\})$ as

$$\exp\{-V(\mathbf{v})\} \prod_{i=1}^{n_A} \Bigg(h(\mathbf{w}_i)(nr_i)^{-(p+1)} \times \prod_{\substack{j=1,\ldots,p: \\ X_{i,j} > nv_j}} [\sigma_j^{-1} p_j^{k_j} X_{i,j}^2 \exp(1/X_{i,j})\{1 - \exp(-1/X_{i,j})\}^{1-k_j}] \Bigg) \tag{5.4}$$

where $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p})$ is the transformation of $\mathbf{Y}_i$ given by equations (5.3).

Generally maximum likelihood estimators behave regularly provided that marginally $k_j < \frac{1}{2}$, $j=1, \ldots, p$ (Smith, 1985). However, in some cases dependence parameters are superefficient. For example in equation (3.2) as $r \downarrow 1$ there is a discontinuity in the associated measure densities: thus within the logistic model any vertex mass implies $r=1$, corresponding to independence of the variables. Tawn (1988b, 1990) discusses this problem further.

As a result of simultaneous estimation covariate modelling and optimal tests of homogeneity of marginal parameters are possible; thus the common characteristics underlying many environmental processes can be exploited. For example, along coastlines it has been found that the generalized extreme value shape parameter for sea-levels is approximately constant (Graff, 1981). Even in the absence of such structure simultaneous estimation gives more precise marginal estimates than univariate GPD analyses as dependence between variables allows transfer of information from one margin to another.

6. OCEANOGRAPHIC APPLICATION

Our data, for the years 1970–76 and 1980–88, consist of simultaneous hourly surge level observations at three sites on the English east coast: Immingham, Lowestoft and Sheerness. The surge levels at a site are defined as the residuals after removal of the astronomically induced tidal component from the sea-level observations. As extreme surges which occur near high tide can cause serious flooding it is important to have both marginal and spatial understanding of this process.

For each site, owing to temporal dependence, extreme surges occur in clusters, forming storms. Consequently, some form of data filtering is necessary to ensure independence of the marginal observations. Simple univariate declustering techniques have been used by Tawn (1988a) and Davison and Smith (1990) for this. An additional feature which must be accounted for in the multivariate case is the time for storm propagation. Here, multivariate declustering is achieved by assuming that independent spatial storms last a standard length of τ hours. The components of each vector in the declustered series are taken as the maximum surge level at each site over a τ-hour event. This approach identifies the total number of independent observations, or equivalently events, n = (number of hourly observations)/τ required for transformations (5.3) and likelihood (5.4). To ensure marginal independence and to allow for propagation time we follow Tawn (1988a) and take $\tau = 40$ h.

As a preliminary step, GPD analyses were performed on each margin. This serves two purposes. Firstly, appropriate thresholds u_i can be determined; in this case 0.8 m, 0.9 m and 1.0 m at Immingham, Lowestoft and Sheerness respectively. Corresponding numbers of independent exceedances were 98, 82 and 89 from 131 independent events which exceeded at least one threshold. A second use of this preliminary investigation is for verification of model validity. On the basis of the marginal GPD fits the vector observations in the bivariate and trivariate cases were transformed to $(r, \mathbf{w})$ using equations (5.3) and (2.1). For bivariate cases Joe *et al.* (1989) have suggested plotting w_1 against $\log r$ as a diagnostic aid. Fig. 1 shows this for Lowestoft with Sheerness. The apparent independence of r and w_1 suggests that the asymptotic property (2.2) applies above the chosen thresholds. Similar conclusions were drawn from plots in the other bivariate cases although these do not exhibit the asymmetry of w_1 in Fig. 1. The w_1 *versus* $\log r$ plot can be further used to identify whether mass is on the boundaries of $\mathbf{w}$. For example in Fig. 1 very few points are near the boundaries, suggesting that models with all mass in the interior of S_2 are appropriate, namely the Dirichlet and symmetric versions of the two logistic models.

Concentrating on the Lowestoft with Sheerness case, standard likelihood ratio tests gave no significant evidence against taking the shape parameter $k = 0$ in each margin,

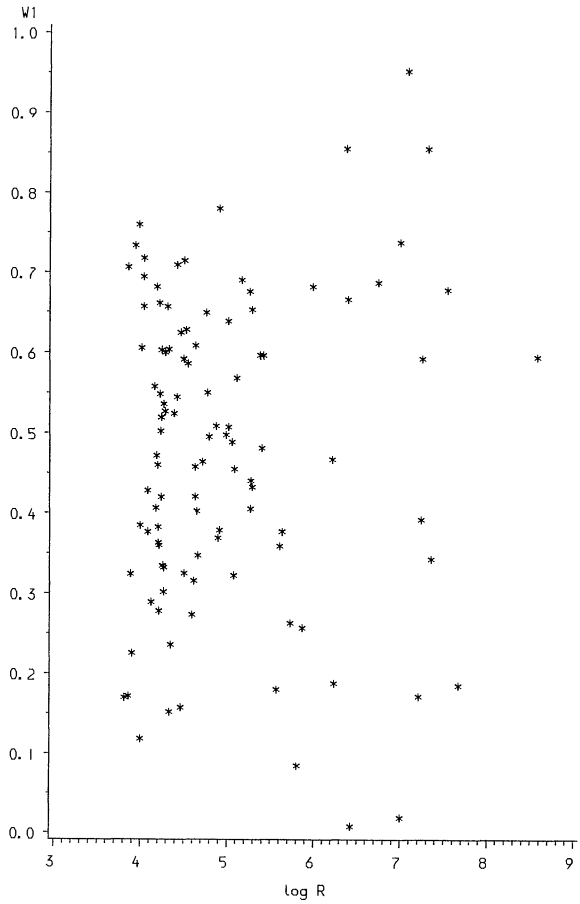

Fig. 1. w_1 *versus* log r plot for Lowestoft with Sheerness

i.e. the tails may be taken as exponential. In this case the following dependence parameter estimates and maximized log-likelihoods L were obtained: $\alpha_1=2.3$, $\alpha_2=13.1$ ($L=-1775.0$) for the Dirichlet model and $\alpha_1=\alpha_2=3.5$ ($L=-1778.0$) for the symmetric Dirichlet model; $r=-2.3$ ($L=-1774.7$) for the negative logistic model; and $r=3.0$ ($L=-1774.7$) for the logistic model. Testing within families is possible using standard likelihood ratio tests; for example there is evidence of asymmetry by testing within the Dirichlet family. Although tests of independence are non-standard a test of 'near independence', $r=1.1$ *versus* $r>1.1$ for the logistic model, gives strong evidence of dependence. For the other pairs symmetry and weaker dependence were detected (Immingham with Lowestoft and Immingham with Sheerness gave logistic parameters $r=2.7$ and $r=2.3$ respectively).

In addition to likelihood methods we have adopted two simple diagnostic procedures to assess model fit and to discriminate between different families of models. One procedure is a graphical comparison of parametric and empirical estimates of the measure, while the other compares model and empirical estimates for probabilities of joint extreme events within the range of the data. In this case both procedures confirm the Dirichlet model as acceptable.

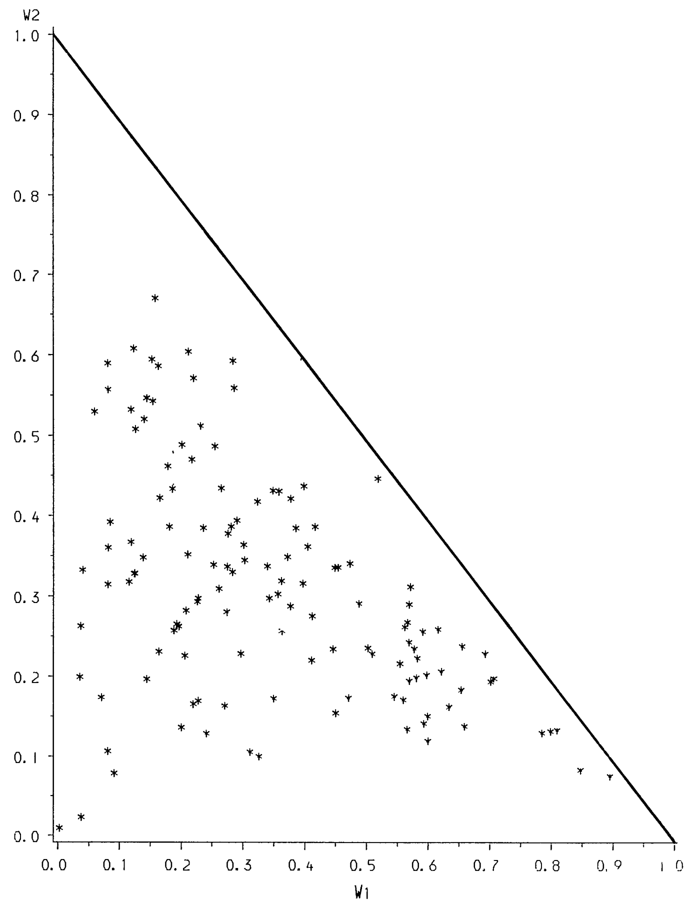

Fig. 2. w_1 *versus* w_2 plot for Immingham, Lowestoft and Sheerness

For the trivariate case Fig. 2 plots w_1 *versus* w_2 based on univariate fits, with Immingham, Lowestoft and Sheerness numbered 1, 2 and 3 respectively. This suggests that the mass of the measure is confined to the interior of S_3. In this higher dimensional case more of the models discussed in Section 4 are suitable for consideration. For each model, likelihood ratio tests gave no evidence against the homogeneity hypothesis $k_i = 0$, $i = 1, 2, 3$; corresponding scale and dependence parameter estimates are given in Table 1. On the basis of both the likelihood values and a comparison of density contours the time series logistic model gives the best fit. This is not surprising as the structure of this model most closely resembles the dynamical process of surge propagation.

The improved efficiency achieved by maximizing likelihood (5.4) is illustrated by comparing the length of profile likelihood intervals with the corresponding intervals from marginal GPD analyses. For example, a GPD fit to the Lowestoft margin gave 95% profile likelihood intervals of (0.20, 0.43) and (−0.16, 0.48) for the scale and shape parameters respectively. The corresponding intervals using the bivariate analysis for Lowestoft with Sheerness were (0.19, 0.37) and (−0.38, 0.21). Further

TABLE 1
Estimated dependence models

Model	*L*	*Parameter estimates* Marginal (σ_1, σ_2, σ_3)	*Dependence*
Symmetric logistic	−2845.1	0.25, 0.30, 0.34	$r=2.5$
Symmetric negative logistic	−2844.4	0.27, 0.29, 0.37	$r=-1.9$
Dirichlet	−2837.3	0.26, 0.30, 0.37	$\alpha_1=1.8$, $\alpha_2=4.4$, $\alpha_3=3.7$
Dirichlet ($\alpha_2=\alpha_3=\alpha$)	−2837.5	0.26, 0.30, 0.37	$\alpha_1=1.8$, $\alpha=4.0$
Nested logistic	−2827.0	0.26, 0.30, 0.35	$r=3.1$, $s=2.4$
Time series logistic	−2822.5	0.27, 0.32, 0.36	$r_1=2.8$, $r_2=3.3$

evidence of the increased precision obtained by using simultaneous estimation comes from simulation studies in which substantial reductions in mean-squared errors of marginal parameter estimates were found.

ACKNOWLEDGEMENTS

S. G. Coles's work was carried out while under the funding of a Science and Engineering Research Council award. We would like to thank Dr C. W. Anderson, the Editor and a referee for helpful discussions and suggestions, and the Proudman Oceanographic Laboratory for providing the data.

APPENDIX A

Proof of Theorem 1

Let C be the set of all non-empty subsets of $\{1, \ldots, p\}$. Consider the partition $(C_1, \ldots, C_p)$ of C where C_j is the set of subsets of $\{1, \ldots, p\}$ of size j. An alternative expression for the exponent measure, equation (2.4), is obtained by the following steps: first split equation (2.4) into a sum of integrals over each boundary space; then transform each integrand into Cartesian components and further separate the integrals by use of the inclusion–exclusion formula. This leads to

$$V(\mathbf{x}) = \sum_{j=1}^{p} \sum_{c \in C_j} \sum_{\substack{d \subset c:\\ d \neq \emptyset}} (-1)^{|d|-1} I(c; d) \tag{A.1}$$

where, if $d \subset c = \{t_1, \ldots, t_j\}$,

$$I(c; d) = \int_0^\infty \ldots \int_0^\infty \int_{x_{t_1}}^\infty \ldots \int_{x_{t_{|d|}}}^\infty h_{j,c}\left(\frac{u_1}{\mathbf{u}\cdot\mathbf{1}_j}, \ldots, \frac{u_j}{\mathbf{u}\cdot\mathbf{1}_j}\right)(\mathbf{u}\cdot\mathbf{1}_j)^{-(j+1)} \,\mathrm{d}\mathbf{u}$$

and $\mathbf{1}_j$ is the j-dimensional unit vector. Now define the operator

$$\mathbf{D}_{m,\tilde{c}} = \frac{\partial^m}{\partial x_{l_1}, \ldots, \partial x_{l_m}}\bigg|_{x_r=0 \text{ if } r \notin \tilde{c}}$$

where $\tilde{c} = \{l_1, \ldots, l_m\}$ and $|\tilde{c}| = m$. Then, if $\tilde{c} \not\subset d$, $D_{m,\tilde{c}}\{I(c; d)\} \equiv 0$ and, if $\tilde{c} \subset d$, $D_{m,\tilde{c}}\{I(c; d)\} \equiv D_{m,\tilde{c}}\{I(c; c)\}$. Hence, by equation (A.1),

$$
\begin{aligned}
D_{m,\bar{c}}\{V(\mathbf{x})\} &= \sum_{j=1}^{p} \sum_{c \in C_j} \sum_{\substack{d \subset c: \\ d \neq \emptyset}} (-1)^{|d|-1} D_{m,\bar{c}}\{I(c;d)\} \\
&= \sum_{j=1}^{p} \sum_{c \in C_j} \sum_{\substack{d \subset c: \\ \bar{c} \subset d}} (-1)^{|d|-1} D_{m,\bar{c}}\{I(c;c)\} \\
&= \sum_{j=1}^{p} \sum_{c \in C_j} \sum_{r=m}^{j} \binom{j-m}{r-m} (-1)^{r-1} D_{m,\bar{c}}\{I(c;c)\}.
\end{aligned}
$$

But as

$$
\sum_{r=m}^{j} \binom{j-m}{r-m} (-1)^{r-1} = \begin{cases} (-1)^{m-1} & \text{if } j = m \\ 0 & \text{if } j > m \end{cases}
$$

then

$$
\begin{aligned}
D_{m,\bar{c}}\{V(\mathbf{x})\} &= (-1)^{m-1} \sum_{c \in C_m} D_{m,\bar{c}}\{I(c;c)\} \\
&= (-1)^{m-1} D_{m,\bar{c}}\{I(\bar{c};\bar{c})\} \\
&= -\left(\sum x_{l_i}\right)^{-(m+1)} h_{m,\bar{c}}\left(\frac{x_{l_1}}{\Sigma x_{l_i}}, \ldots, \frac{x_{l_m}}{\Sigma x_{l_i}}\right).
\end{aligned}
$$

Proof of Theorem 2

Let h^* be any positive function on S_p with finite first moments m_j, given by equation (3.5). Then by equation (3.3)

$$
G(\mathbf{x}) = \exp\left\{-\int_{S_p} \max\left(\frac{u_j}{m_j x_j}\right) h^*(\mathbf{u})\, d\mathbf{u}\right\} \tag{A.2}
$$

is a multivariate extreme value distribution with unit Fréchet margins. Applying the change of variables $u_j = m_j w_j/(\mathbf{m}\cdot\mathbf{w}), j=1, \ldots, p$, in equation (A.2) gives

$$
G(\mathbf{x}) = \exp\left\{-\int_{S_p} \max\left(\frac{w_j}{x_j}\right) h^*\left(\frac{m_1 w_1}{\mathbf{m}\cdot\mathbf{w}}, \ldots, \frac{m_p w_p}{\mathbf{m}\cdot\mathbf{w}}\right) \frac{J(\mathbf{w})}{\mathbf{m}\cdot\mathbf{w}}\, d\mathbf{w}\right\}
$$

where $J(\mathbf{w})$ is the Jacobian of the transformation $\mathbf{u} \rightarrow \mathbf{w}$. By comparison with equation (2.4) it follows that

$$
h(\mathbf{w}) = \frac{J(\mathbf{w})}{\mathbf{m}\cdot\mathbf{w}} h^*\left(\frac{m_1 w_1}{\mathbf{m}\cdot\mathbf{w}}, \ldots, \frac{m_p w_p}{\mathbf{m}\cdot\mathbf{w}}\right)
$$

is a measure density on S_p satisfying constraints (2.3). It remains to evaluate $J(\mathbf{w})$. As

$$
\begin{aligned}
\partial u_j/\partial w_j &= m_j\{\mathbf{m}\cdot\mathbf{w} + (m_p - m_j) w_j\}/(\mathbf{m}\cdot\mathbf{w})^2, && j=1, \ldots, p, \\
\partial u_j/\partial w_k &= (m_p - m_k) m_j w_j/(\mathbf{m}\cdot\mathbf{w})^2, && j, k=1, \ldots, p, \quad j \neq k,
\end{aligned}
$$

letting $\mathbf{w}_* = (w_1, \ldots, w_{p-1})'$, $\mathbf{m}_* = (m_p - m_1, \ldots, m_p - m_{p-1})'$, and I_j be the $j \times j$ identity matrix it follows that

$$
J(\mathbf{w}) = (\mathbf{m}\cdot\mathbf{w})^{2(1-p)} \left(\prod_{j=1}^{p-1} m_j\right) \det\{(\mathbf{m}\cdot\mathbf{w}) I_{p-1} + \mathbf{w}_* \cdot \mathbf{m}_*'\}
$$

$$= (\mathbf{m}\cdot\mathbf{w})^{1-p}\left(\prod_{j=1}^{p-1} m_j\right)\det\{I_{p-1} + (\mathbf{w}_*\cdot\mathbf{m}_*')/(\mathbf{m}\cdot\mathbf{w})\}$$

$$= (\mathbf{m}\cdot\mathbf{w})^{1-p}\left(\prod_{j=1}^{p-1} m_j\right)\det\{I_1 + (\mathbf{m}_*'\cdot\mathbf{w}_*)/(\mathbf{m}\cdot\mathbf{w})\}$$

by use of a special case of a determinant property in Mardia *et al.* (1979). Hence

$$J(\mathbf{w}) = (\mathbf{m}\cdot\mathbf{w})^{1-p}\left(\prod_{j=1}^{p-1} m_j\right)\{1 + (m_p - \mathbf{m}\cdot\mathbf{w})/(\mathbf{m}\cdot\mathbf{w})\}$$

$$= (\mathbf{m}\cdot\mathbf{w})^{-p}\prod_{j=1}^{p} m_j.$$

REFERENCES

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Barnett, V. (1976) The ordering of multivariate data (with discussion). *J. R. Statist. Soc.* A, **139**, 318-354.

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

Graff, J. (1981) An investigation of the frequency distributions of annual maxima at ports around Great Britain. *Est. Coast. Shelf Sci.*, **12**, 389-449.

Gumbel, E. J. (1960) Distributions de valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Paris*, **9**, 171-173.

de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. Probab.*, **12**, 1194-1204.

—— (1985) Extremes in higher dimensions: the model and some statistics. *Proc. 45th Sess. Int. Statist. Inst.*, paper 26.3.

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

Joe, H. (1989) Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.*, **9**, 75-82.

Joe, H., Smith, R. L. and Weissman, I. (1989) Bivariate threshold methods for extremes. *J. R. Statist. Soc.* B, to be published.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*, A.2.3.n. London: Academic Press.

McFadden, D. (1978) Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models* (eds A. Karlqvist, L. Lundquist, F. Snickers and J. Weibull), pp. 75-96. Amsterdam: North-Holland.

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

—— (1981) Multivariate extreme value distributions. *Proc. 43rd Sess. Int. Statist. Inst.*, 859-878.

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

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

—— (1986) Extreme value theory based on the *r* largest annual events. *J. Hydrol.*, **86**, 27-43.

—— (1989) The extremal index for a Markov chain. Submitted to *J. Appl. Probab.*

—— (1990) Extreme value theory. In *Handbook of Applicable Mathematics* (ed. W. Ledermann), vol. 7. Chichester: Wiley.

Smith, R. L., Tawn, J. A. and Yuen, H. K. (1990) Statistics of multivariate extremes. *Int. Statist. Inst. Rev.*, **58**, 47-58.

Tawn, J. A. (1988a) An extreme value theory model for dependent observations. *J. Hydrol.*, **101**, 227–250.
—— (1988b) Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
—— (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.
Weissman, I. (1978) Estimation of parameters and quantiles based on the k largest observations. *J. Am. Statist. Ass.*, **73**, 812–815.