Extremes
DOI 10.1007/s10687-007-0054-y

# Bayesian inference for clustered extremes

**Lee Fawcett · David Walshaw**

Received: 19 December 2006 / Revised: 17 December 2007 /
Accepted: 17 December 2007


**Abstract** We consider Bayesian inference for the extremes of dependent stationary series. We discuss the virtues of the Bayesian approach to inference for the extremal index, and for related characteristics of clustering behaviour. We develop an inference procedure based on an automatic declustering scheme, and using simulated data we implement and assess this procedure, making inferences for the extremal index, and for two cluster functionals. We then apply our procedure to a set of real data, specifically a time series of wind-speed measurements, where the clusters correspond to storms. Here the two cluster functionals selected previously correspond to the mean storm length and the mean inter-storm interval. We also consider inference for long-period return levels, advocating the posterior predictive distribution as being most representative of the information required by engineers interested in design level specifications.

**Keywords** Bayesian inference · Extremal index · MCMC · Clusters · Extreme values

**AMS 2000 Subject Classification** 62G32

L. Fawcett · D. Walshaw (✉)
School of Mathematics and Statistics, Newcastle University, Newcastle NE1 7RU, UK
e-mail: david.walshaw@ncl.ac.uk

L. Fawcett
e-mail: lee.fawcett@ncl.ac.uk

## 1 Introduction

In this paper we consider Bayesian inference for characteristics of the clustering behaviour of extremes of dependent stationary series. As a starting point, we consider the extremal index, which characterises the strength of extremal dependence, and for which much work has been published concerning classical (non-Bayesian) inference. Here we adopt Bayesian inference for the extremal index itself, and for certain characteristics of the *clusters* of extremes (defined as exceedances of an appropriate threshold) which will arise as part of such a series. A simple example would be the mean cluster size. Such quantities are often referred to as *cluster functionals* (Smith et al. 1997).

There are many reasons why a Bayesian analysis of extremes might be preferable to a standard classical (frequentist) analysis, not least the ability to supplement information provided by the data with other sources of information, through the *prior distribution*. The output of a Bayesian analysis, which provides posterior information on the status of extremal dependence, can also be exploited to provide posterior inferences on any cluster functional related to this dependence, for example by use of some appropriate cluster identification method within a Bayesian sampling scheme. In this way it becomes possible to assess the variability of estimates of the extremal index, as well as estimates of any other cluster functional, by direct reference to their posterior distributions. Further, estimating *return levels*, i.e. the values exceeded with specified small probability, is a design requirement in many applications, and the Bayesian framework provides a very effective tool for achieving this, via the *posterior predictive distribution*.

We begin with the usual definition of the extremal index. Let $X_1, X_2, \ldots, X_n$ be the first $n$ observations of a stationary series satisfying Leadbetter's $D(u_n)$ condition (Leadbetter et al. 1983), and let $M_n = \max\{X_1, X_2, \ldots X_n\}$. Now let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ be an *independent* series, with $\tilde{X}$ having the same distribution as $X$, and let $\tilde{M}_n = \max\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\}$. Then if $\tilde{M}_n$ has a non-degenerate limit law given by $\Pr\{(\tilde{M}_n - b_n)/a_n \leq x\} \to G(x)$, it follows that

$$\Pr\{(M_n - b_n)/a_n \leq x\} \to G^{\theta}(x), \tag{1}$$

for some $0 \leq \theta \leq 1$ (Leadbetter et al. 1983). The parameter $\theta$ is known as the *extremal index*.

The concept of the extremal index was developed in a series of papers including Newell (1964), Loynes (1965), O'Brien (1974) and Leadbetter et al. (1983), with a review by Leadbetter and Rootzén (1988). From a statistical point of view, since there are already many publications on estimating extreme value distributions from i.i.d. data, it can be seen that estimating the extremal index is a key problem. To this end, the extremal index has been the focus of much recent theoretical and practical research—see Ferro and Segers (2003), Ancona-Navarrete and Tawn (2000), Smith and Weissman (1994) and Smith (1992).

The conventional approach to estimation of the extremal index, as well as the estimation of any other cluster functional, has been to identify

(empirically) clusters of extreme events, and then to use *ad hoc* methods to study within-cluster behaviour (see, for example, Coles 2001). However, estimates are often highly sensitive to the choice of cluster identification procedure, and it is difficult to assess the variability of our estimators using such an empirical approach.

Ancona-Navarrete and Tawn (2000) suggest a maximum likelihood estimator for the extremal index which makes use of the characterisation in expression (1). However, the estimation of cluster functionals still requires the identification of clusters of extremes, and so this problem is still subject to issues of sensitivity.

Smith et al. (1997) use parametric models for the temporal evolution of extremes to *simulate* clusters of extremes for a process, allowing the study of estimates of the extremal index, and other cluster functionals by analysing their sampling distributions obtained via repeated simulation-estimation. However, issues surrounding the choice of model for the dependence structure then arise, which are often non-trivial to address.

In Section 2 we review some methods used for estimating the extremal index, and we recap some techniques which have been proposed to assess the variability of such estimates. We discuss how one of these methods, due to Ferro and Segers (2003), can be extended to cover the cluster identification process and hence can be used to make inferences on various cluster functionals. We also discuss the practicalities of Bayesian inference and how the methods suggested by Ferro and Segers (2003) can be adapted to work within the Bayesian framework. In Section 3 we illustrate the implementation (and assess the performance) of such methods in the Bayesian framework using simulated data, obtaining a posterior distribution for the extremal index and two selected cluster functionals. In Section 4, these techniques are implemented for a time series of genuine wind-speed data obtained at High Bradfield in northern central England. The same two cluster functionals here correspond to the mean storm duration, and the mean duration of calm periods (both of interest to practitioners), and inference on return levels is based on the posterior predictive distribution.

## 2 Inference for the Extremal Index and Cluster Characteristics

### 2.1 Cluster Size Methods

Section 1 gives a formal definition of the extremal index; an alternative characterisation, provided by Hsing et al. (1988), is that $\theta^{-1}$ is the limiting mean cluster size in the point process of exceedance times over a high threshold. This suggests that a suitable way to estimate the extremal index can be found through methods which identify clusters of extremes, the estimate itself being found as the reciprocal of the mean cluster size. One such approach is 'runs' declustering, where a declustering interval $\kappa$ is chosen such that a cluster of extremes above a threshold is deemed to have terminated when at least $\kappa$

consecutive observations fall below that threshold. Another method is 'blocks' declustering. This method of cluster identification partitions the data into approximately $k$ blocks of length $\tau$ and the threshold exceedances within each block are treated as a single cluster of extremes. As with most other cluster identification procedures, both of these require an arbitrarily chosen auxiliary parameter (e.g. the declustering interval $\kappa$ in 'runs' declustering and the block size $\tau$ in 'blocks' declustering), the choice of which can seriously influence the extremal index estimate. Such estimation procedures are also completely empirical, giving a point estimate for $\theta$ without any natural way to quantify the uncertainty associated with it. For a more detailed discussion of both 'runs' and 'blocks' declustering see, for example, Smith and Weissman (1994).

From the estimate for $\theta$ it is possible to obtain point estimates for certain cluster functionals. However, estimation of such quantities will also be vulnerable to the choice of declustering parameter used in the original cluster identification process, and once again, the variability of the estimates is difficult to quantify.

## 2.2 Maxima Methods

Another class of estimators, based on block maxima and using the characterisation in expression (1), has been developed by, amongst others, Gomes (1993) and Ancona-Navarrete and Tawn (2000). From expression (1) it can be seen that, relative to the independent case, the effect that introducing extremal dependence to the time series has on the limiting distribution of block maxima is to change this distribution from $G(x)$ (for i.i.d. variables) to $G^{\theta}(x)$ (for a stationary dependent sequence). The distributions $G(x)$ and $G^{\theta}(x)$ can be estimated using data on maxima obtained from (respectively) the observed stationary sequence, and a constructed independent sequence with the same marginal distribution as the stationary sequence. Then, by a suitable comparison of these two estimated distributions, $\theta$ itself can be estimated (see Gomes 1993).

First consider the estimation of the distribution of the maximum of any stationary sequence divided into $k$ blocks of length $\tau$. If we denote the maximum of the $i^{th}$ block by $M_{\tau,i}$, for $i = 1, \ldots, k$, then it follows that, for large enough $\tau$, the $M_{\tau,i}$ are approximately independent observations from a distribution $G(x; \mu, \sigma, \xi)$, being a generalised extreme value (GEV) distribution, where

$$G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\}, \tag{2}$$

and $a_+ = \max(0, a)$. The parameters $\mu$ $(\infty < \mu < \infty)$, $\sigma$ $(\sigma > 0)$ and $\xi$ $(-\infty < \xi < \infty)$ are location, scale and shape parameters respectively. These parameters can be estimated by assuming the $M_{\tau,i}$ values are a random sample from the GEV distribution and using, for example, maximum likelihood estimation (see Coles 2001, Chapter 3, for details of both the limiting theory and the estimation ideas).

For any stationary sequence, an exchangeable sequence with the same marginal distribution as the dependent sequence can be generated by randomising the index of the observations. This exchangeable series acts as a good approximation to an associated i.i.d. series, such as the independent series prior to expression (1). For this associated series, denote the block maxima by $\tilde{M}_{\tau,i}$ for $i = 1, \ldots, k$. Now if the $\tilde{M}_{\tau,i}$ were i.i.d. GEV random variables having distribution $G(x; \mu, \sigma, \xi)$ , and if the extremal index of the dependent series is equal to $\theta$, then from Eq. 1, the block maxima of the dependent series, $M_{\tau,i}$, have a distribution $G^{\theta}(x; \mu, \sigma, \xi)$, which is easily seen to be a GEV distribution $G(x; \mu_\theta, \sigma_\theta, \xi_\theta)$, where

$$\begin{aligned} \mu_\theta &= \mu - \sigma(1 - \theta^{\xi})/\xi, \\ \sigma_\theta &= \sigma\theta^{\xi} \qquad \text{and} \\ \xi_\theta &= \xi. \end{aligned}$$

Gomes (1993) proposed estimating $\theta$ from estimates $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and $(\hat{\mu}_\theta, \hat{\sigma}_\theta, \hat{\xi}_\theta)$ obtained from the separate fits to the two sets of block maxima. A pooled estimate of $\xi$ is calculated as

$$\tilde{\xi} = \frac{\hat{\sigma} - \hat{\sigma}_\theta}{\hat{\mu} - \hat{\mu}_\theta}.$$

Then an estimate of $\theta$ is given by

$$\hat{\theta} = \left(\frac{\hat{\sigma}}{\hat{\sigma}_\theta}\right)^{-1/\tilde{\xi}}.$$

Ancona-Navarrete and Tawn (2000) advanced this idea further. They suggested that, instead of separately estimating the two sets of GEV parameters and then manipulating these to estimate $\theta$, it would be better to simultaneously estimate all the components of the augmented parameter vector $(\mu, \sigma, \xi, \theta)$, which is a sufficient statistic for the joint distribution $(M_\tau, \tilde{M}_\tau)$.

To undertake this simultaneous estimation, the components of the vector

$$(\boldsymbol{M}_\tau, \tilde{\boldsymbol{M}}_\tau) = (M_{\tau,1}, \ldots, M_{\tau,k}, \tilde{M}_{\tau,1}, \ldots, \tilde{M}_{\tau,k}) \tag{3}$$

are treated as independent (but not identically distributed) GEV random variables. Thus, the joint log-likelihood is given by

$$\begin{aligned}\ell(\mu,\sigma,\xi,\theta|\boldsymbol{M}_\tau,\tilde{\boldsymbol{M}}_\tau) = &-k(\log\sigma+\xi\log\theta)-\sum_{i=1}^{k}\left[\frac{\sigma+\xi(M_{\tau,i}-\mu)}{\sigma\theta^{\xi}}\right]^{-1/\xi}\\ &-(1/\xi+1)\sum_{i=1}^{k}\log\left[\frac{\sigma+\xi(M_{\tau,i}-\mu)}{\sigma\theta^{\xi}}\right]\\ &-k\log\sigma-\sum_{i=1}^{k}\left[1+\xi\left(\frac{\tilde{M}_{\tau,i}-\mu}{\sigma}\right)\right]^{-1/\xi}\\ &-(1/\xi+1)\sum_{i=1}^{k}\log\left[1+\xi\left(\frac{\tilde{M}_{\tau,i}-\mu}{\sigma}\right)\right]. \qquad (4)\end{aligned}$$

Although the joint vector in Eq. 3 is not independent, Ancona-Navarrete and Tawn (2000) demonstrate that the impact of this approximation is asymptotically zero, and argue that it is always likely to be negligible in practice. The extremal index $\theta$ can then be estimated (simultaneously with the GEV parameters $\mu$, $\sigma$ and $\xi$) via maximisation of $\ell$ in Eq. 4.

It is also possible to estimate the limiting mean cluster size by inversion of the extremal index estimator, and assess the variability of this and the estimator for the extremal index itself, by reference to the usual asymptotic properties of maximum likelihood estimators. However, other cluster characteristics require the systematic identification of clusters of extremes using a declustering technique. For such cluster functionals, estimates remain vulnerable to the choice of technique used to identify clusters, and hence the methods based on maxima do not improve over the cluster size methods when such quantities are of interest.

### 2.3 Simulation Methods

Smith et al. (1997) propose a simulation framework for estimating the extremal index and other cluster functionals. They fit a first-order Markov chain model to their data, modelling the distribution of consecutive pairs of extremes using bivariate extreme value distributions. They assess the suitability of the first-order Markov assumption for their data, and then use *ad hoc* procedures to compare different models for the choice of dependence structure. The bivariate logistic model, with dependence parameter $\alpha \in (0, 1]$ (see, for example, Coles 2001), is found to be the most suitable model to represent the temporal evolution of the extremes of their process, and the estimated value of $\alpha$ is then used to repeatedly simulate clusters of extremes from their series. For each series of simulated clusters the extremal index is estimated using the runs estimator, as well as other cluster functionals, and 95% confidence intervals are obtained from this collection of estimates.

Although Smith et al. (1997) find that a first-order Markov assumption seems adequate for the extremes of their process, in common with standard time series modelling, the selection of model-order can be difficult. Testing between orders $d-1$ and $d$ is equivalent to testing for conditional independence of variables within a unit simplex domain (Smith et al. 1997). Standard likelihood procedures can be used, though models will typically be non-nested, and so it may be necessary to use more informal goodness-of-fit assessments based on comparisons of predicted and observed functionals of interest. The same difficulties can also arise when selecting between different dependence models of a given order; work by Fawcett (2005) also shows that estimates of the extremal index and other cluster functionals are highly sensitive both to the choice of model, and the choice of model-order.

### 2.4 Methods Based on Inter-arrival Times of Extremes

Ferro and Segers (2003) provide an alternative approach to inference for clusters of extremes. Consider a high threshold $u$, with marginal exceedance probability $\bar{F}(u)$. Now consider a sequence of exceedances of $u$, and denote by $T(u)$ the random variable corresponding to the inter-arrival times of consecutive exceedances (known as 'inter-exceedance times'). Ferro and Segers (2003) show that under a reasonable mixing condition, and to a first-order approximation, the distribution of inter-exceedance times is given by

$$\Pr\left(T(u) > n\right) = \theta\left(1 - \bar{F}(u)\right)^{n\theta}. \tag{5}$$

As with the maxima method of Ancona-Navarrete and Tawn (2000), a maximum likelihood estimate of $\theta$ can be found by numerical optimisation of the likelihood associated with the model in Eq. 5. Unfortunately, as Ferro and Segers (2003) explain, a combination of two features can cause it to perform poorly. Firstly, the distribution (5) from which the likelihood is constructed is not a good model for the smallest inter-exceedance times. Secondly, the associated log-likelihood is extremely sensitive to the number of inter-exceedance times equal to one; the observed value may be far from the value that is expected under the model, in which case poor estimates result.

Due to the shortcomings of the model in Eq. 5 we choose only to exploit an extension to the inter-exceedance times methodology of Ferro and Segers (2003). This extension provides an "automatic" scheme for the identification of clusters which avoids the need to make an arbitrary choice of declustering parameter, allowing the inter-arrival times between threshold exceedances to identify clusters in a more natural way than those schemes which require an arbitrary auxiliary parameter (such as standard 'runs' or 'blocks' declustering as described in Section 2.1). They outline their declustering method as follows. Inter-arrival times of threshold excesses are classified into two types: independent inter-cluster times and independent sets of intra-cluster times. Given the

set of inter-arrival times $T_i; i = 1, \ldots, N-1$, between threshold exceedances $S_1, \ldots, S_N$, we can assume that the largest $C - 1 = \lfloor \theta N \rfloor$ inter-arrival times are approximately independent inter-cluster times that divide the remainder into approximately independent sets of intra-cluster times. More precisely, if $T_{(C)}$ is the $C$-th largest inter-arrival time and $T_{i_j}$ is the $j$-th inter-arrival time to exceed $T_{(C)}$, then $\{T_{i_j}\}_{j=1}^{C-1}$ is a set of approximately independent inter-cluster times. Let $\mathcal{T}_j = \{T_{i_{j-1}+1}, \ldots, T_{i_j-1}\}$, where $i_0 = 0$, $i_C = N$ and $\mathcal{T}_j = \varnothing$ if $i_j = i_{j-1} + 1$. Then $\{\mathcal{T}_j\}_{j=1}^{C}$ is a collection of approximately independent sets of intra-cluster times. Essentially, this approach is equivalent to runs declustering with $\kappa = T_{(C)}$, where $C = \lfloor \theta N \rfloor + 1$; however, the declustering parameter is no longer chosen arbitrarily—the specific value of $\kappa$ is now governed by the level of extremal dependence in the process, which is quantified by $\theta$.

## 2.5 Bayesian Methods

Before we consider methods for estimating the extremal index (and other characteristics of clustering behaviour) within the Bayesian framework, we give a general introduction to Bayesian methods in extremes.

Although the Bayesian paradigm was quite late to be adopted by statisticians working on extreme value theory and methods, there is now a reasonable body of literature focusing on Bayesian statistics for extremes. For some general background, Coles (2001) devotes a section to this topic in Chapter 9, while Stephenson and Tawn (2004) review the literature in a paper which focuses on accounting for the three extremal types. Coles and Powell (1996) carry out a comprehensive review of the literature up to that date, and analyse wind data from a number of locations in the USA by constructing a prior for the GEV parameters based on estimates obtained at other locations.

Among the other significant contributions, Coles and Tawn (1996) use expert knowledge to construct a multivariate prior for the GEV parameters, and Smith and Walshaw (2003) extend this idea to bivariate distributions for extreme rainfall at pairs of locations within a region. Smith (1999) considers predictive inference under the Bayesian and frequentist paradigms, and Smith and Goodman (2000) and Bottolo et al. (2003) construct Bayesian hierarchical models for extreme values in insurance problems. Fawcett and Walshaw (2006a) model extreme wind speeds in a region of the UK using a Bayesian hierarchical model. Fawcett and Walshaw (2006b) consider Bayesian inference for Markov chain models (also for extreme wind speeds) using a simulation framework similar to that used by Smith et al. (1997), outlined in Section 2.3, to obtain estimates of the extremal index.

For Bayesian inference in our general context, we assume that the data $\boldsymbol{x} = (x_1, \ldots, x_n)$ provide our information on clustering behaviour by coming from a density within a parametric family $\mathcal{F} = \{f(\boldsymbol{x}; \boldsymbol{\psi}) : \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$. So, for example, $\boldsymbol{x}$ may be a vector of block maxima as given in Eq. 3, or it may be a vector of inter-exceedance times for observations above a suitable high threshold. We wish to make inference on the parameter vector $\boldsymbol{\psi}$ (in the parameter space $\boldsymbol{\Psi}$). Denote the likelihood function for $\boldsymbol{\psi}$ by $L(\boldsymbol{\psi}|\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\psi})$, and suppose that

our prior beliefs about $\boldsymbol{\psi}$ are expressed by the probability density function $\pi(\boldsymbol{\psi})$. Then Bayes' Theorem gives the posterior density as

$$\pi(\boldsymbol{\psi}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{\psi})L(\boldsymbol{\psi}|\boldsymbol{x})}{f(\boldsymbol{x})}. \tag{6}$$

Computation of the denominator in Eq. 6 can be problematic, as the solution is usually analytically intractable. Stochastic simulation using a Markov chain Monte Carlo (MCMC) method provides a now widely used solution.

The idea behind one such method, the Metropolis–Hastings sampler (see, for example, Smith and Roberts 1993), is to produce simulated values from the posterior distribution in the following way: set an initial value $\boldsymbol{\psi}^{(1)}$, and specify an arbitrary probability rule $q(\boldsymbol{\psi}^{(i+1)}|\boldsymbol{\psi}^{(i)})$ to simulate future values. At each step in the simulation, the probability rule $q(.|\boldsymbol{\psi}^{(i)})$ is used to generate a proposal value $\boldsymbol{\psi}^*$ for $\boldsymbol{\psi}^{(i+1)}$. We then set $\boldsymbol{\psi}^{(i+1)}$ equal to $\boldsymbol{\psi}^*$ with probability

$$A_i = \min\left\{1, \frac{\pi(\boldsymbol{\psi}^*)L(\boldsymbol{\psi}^*|\boldsymbol{x})q(\boldsymbol{\psi}^{(i)}|\boldsymbol{\psi}^*)}{\pi(\boldsymbol{\psi}^{(i)})L(\boldsymbol{\psi}^{(i)}|\boldsymbol{x})q(\boldsymbol{\psi}^*|\boldsymbol{\psi}^{(i)})}\right\},$$

or put $\boldsymbol{\psi}^{(i+1)} = \boldsymbol{\psi}^{(i)}$ otherwise. Whether or not the move is accepted or rejected depends on the acceptance probability, which itself depends on the relationship between the density of interest and the probability rule. A common choice for this probability rule is a random walk chain, where the proposal $\boldsymbol{\psi}^*$ at iteration $r$ is $\boldsymbol{\psi}^* = \boldsymbol{\psi}^{(r)} + \epsilon_r$, where the $\epsilon_r$ are i.i.d. random variables. It can be shown that the generated sequence has a stationary distribution which, under simple regularity assumptions, is the target posterior distribution in Eq. 6. No matter what the choice of $q$, the rejection steps outlined above ensure that the simulated values have, in the limit, the desired marginal distribution.

To implement an estimation procedure for the extremal index $\theta$ (and indeed any other cluster functional) in the Bayesian setting, we require a likelihood function involving $\theta$. In this paper, in order to obtain draws from the posterior distribution for $\theta$, we exploit the log-likelihood $\ell$ given in Eq. 4. However, we then combine this with the inter-exceedance times methodology of Ferro and Segers (2003), outlined in Section 2.4, to obtain draws from the posterior distribution for the declustering interval $\kappa$. We do this by setting $\kappa$ equal to the $C$-th largest inter-exceedance time, where $C = \lfloor \theta N \rfloor + 1$ (see Section 2.4). For each posterior draw for $\kappa$ we can then implement a runs declustering scheme to identify unique clusters of extremes, and so obtain an associated posterior draw for any other cluster functional. Specifically, we implement the following steps:

1. Obtain a posterior sample $\boldsymbol{\psi}^{(1)}, \ldots, \boldsymbol{\psi}^{(R)}$, where $\boldsymbol{\psi} = (\mu, \sigma, \xi, \theta)$, and $R$ is the number of iterations in our MCMC scheme, using log-likelihood $\ell$ (Eq. 4);
2. Calculate $C^{(r)} = \lfloor \theta^{(r)} N \rfloor + 1, r = 1, \ldots, R$;
3. Find $\kappa^{(r)}$, the $C^{(r)}$-th largest inter-exceedance time, that is the $C^{(r)}$-th order statistic in the sequence $T_i, i = 1, \ldots, N-1, r = 1, \ldots, R$;

4. Use each $\kappa^{(r)}, r = 1, \ldots, R$, as the declustering interval to implement a full cluster identification procedure based on runs declustering;
5. Use each set of identified clusters found using $\kappa^{(r)}, r = 1, \ldots, R$, to estimate any other cluster characteristic, say $H^{(r)}$, and so obtain draws from the (approximate) posterior distribution for that functional also.

## 3 Application to Simulated Data

In this section we use simulated data to investigate the performance of the procedure for implementing Bayesian inference, described in Section 2.5. In Section 3.1 we describe the simulation of a dependent series. Section 3.2 contains some preliminary analyses based on using the non-Bayesian methods described in Sections 2.1 and 2.2, and in Section 3.3 we implement our full MCMC scheme.

### 3.1 Simulation of a Dependent Series

We generate a sequence of artificial data $X_i$; $i = 1, \ldots, n$, where the joint distribution of each pair of consecutive observations is given by the c.d.f.

$$G(x_i, x_{i+1}; \alpha) = \exp\left[-\left\{\exp\left(-x_i/\alpha\right) + \exp\left(-x_{i+1}/\alpha\right)\right\}^{\alpha}\right]; \quad i = 1, \ldots, n-1,$$

for $x_i, x_{i+1} > 0$ and $\alpha \in (0, 1]$. Independence and complete dependence correspond to $\alpha = 1$ and $\alpha \to 0$ respectively. This is the symmetric logistic model for dependence of bivariate extremes, as described by Tawn (1988). In this form, it ensures that the sequence is stationary with identically distributed Gumbel margins (i.e. having $F(x) = \exp\left\{-\exp(-x)\right\}$), and that there is genuine extremal dependence in the time series, the strength of which is controlled by the value of $\alpha$. With this marginal choice the lag 1 autocorrelation is $1 - \alpha^2$. The precise method of simulation utilises an envelope rejection method to simulate each value of $x_{i+1}$ conditional on $x_i$, so that $(x_i, x_{i+1})$ has a bivariate unit exponential distribution. The margins are then transformed to the required Gumbel form (Fawcett 2005, Chapter 5). In the present study, we simulate two series of length $n = 10{,}000$ with a fixed level of dependence—$\alpha = 0.2$ and $\alpha = 0.5$. These values of $\alpha$ give lag 1 autocorrelations of 0.96 and 0.75, and Smith (1992) shows that the corresponding values of $\theta$ are 0.0616 and 0.328 (respectively).

### 3.2 Preliminary Analyses

A 95% marginal quantile was used to identify extreme events in each simulated time series; work by Ancona-Navarrete and Tawn (2000) shows that the choice of threshold used to identify extremes has a very limited effect on

estimates of the extremal index, provided reasonable values are chosen and the threshold level is high. Using each of the 'runs' and 'blocks' declustering methods described in Section 2.1, we have estimated the extremal index as the reciprocal of the observed mean cluster size, using various values of the declustering parameter in each scheme to identify clusters. Figure 1 shows a plot of the declustering parameter against the extremal index estimate for both schemes, and for both simulated series. From these plots, the shortcomings of such cluster size estimation procedures are clear: estimation of $\theta$ is both sensitive to the choice of declustering scheme (here, 'runs' and 'blocks') and the choice of auxiliary parameter within the chosen scheme ($\kappa$ and $\tau$ for 'runs' and 'blocks' respectively). Sensitivity of estimates of the extremal index to the choice of declustering scheme and/or the choice of parameter within that scheme will also lead to the sensitivity of any other cluster characteristic which is a function of $\theta$. Indeed, some cluster functionals will be defined by the way in which clusters are identified, leading to a direct dependence between these functionals and, for example, the value for $\kappa$ used in 'runs' declustering; two examples are the mean cluster length and the mean interval duration between clusters (which we denote by $\rho$ and $\omega$, respectively). Thus it is important to develop methods for estimating the extremal index, and indeed any functional of cluster behaviour, without the need to make arbitrary choices of cluster identification scheme or within-scheme auxiliary parameter.

We can use the approach of Section 2.2 to obtain a maximum likelihood estimator for $\theta$, by numerical optimisation of $\ell$ in Eq. 4. Doing so gives $\hat{\theta} = 0.063$ (0.052) and $\hat{\theta} = 0.297$ (0.047) for the series with $\theta = 0.0616$ and 0.328 (respectively, standard errors in parentheses). Inverting $\hat{\theta}$ gives an estimate of the limiting mean cluster length: $\hat{\rho} = 15.873$ (0.212) and $\hat{\rho} = 3.367$ (0.193) for

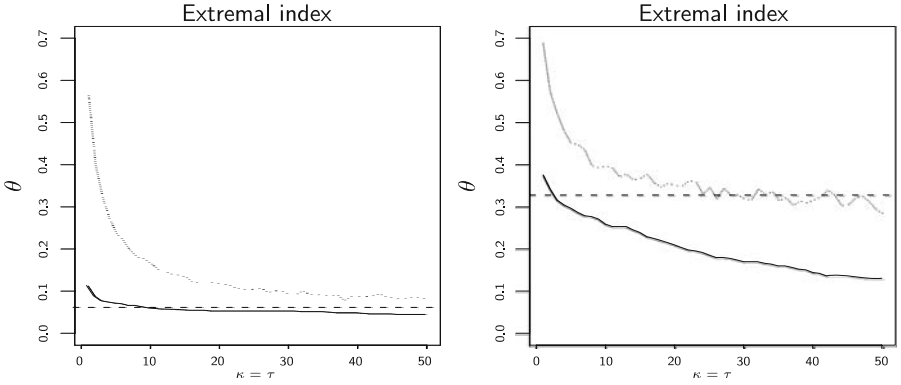


**Fig. 1** Plot of estimates of the extremal index, $\theta$, against the declustering parameter used to identify clusters of extremes ($\kappa$ for 'runs' declustering and $\tau$ for 'blocks' declustering). The *solid lines* indicates estimates obtained using 'runs' declustering, the *dotted lines* 'blocks' declustering. The *horizontal broken lines* represent the true parameter values. *Estimates in the left panel* have been obtained from a series with $\alpha = 0.2$; *in the right panel*, $\alpha = 0.5$.

the same two series (estimates of the standard error have been found via the delta method—see, for example, Coles 2001, page 33). This approach does not extend to the estimation of other cluster functionals, however, which require the identification of clusters of extremes (see above).

### 3.3 Implementation of the MCMC Scheme for Bayesian Inference

Here we implement the Bayesian inference procedure developed in Section 2.5. We specify independent, highly uninformative prior distributions for the GEV parameters in Eq. 2. Specifically, we use

$$\begin{aligned} \mu &\sim N\left(0, 10^4\right), \\ \log(\sigma) &\sim N\left(0, 10^4\right) \qquad \text{and} \\ \xi &\sim N\left(0, 10^2\right), \end{aligned}$$

where the Normal distribution here is specified by its mean and variance. We work with $\log(\sigma)$ for computational convenience and to retain the positivity of the scale parameter $\sigma$; simple exponentiation of the posterior draws for $\log(\sigma)$ can then be used to obtain posterior draws for $\sigma$ itself. In the absence of any useful prior information about the extremal index, we again choose an uninformative option, specifically the flat prior:

$$\theta \sim U(0, 1).$$

Though not attempted in this paper, any subsequent Bayesian inference for the extremal index, or for any cluster functional derived from this using the scheme outlined at the end of Section 2.5, has the potential to be improved by using prior distributions for $\mu$, $\log(\sigma)$, $\xi$ and $\theta$ which are derived from carefully elicited prior information about these parameters, instead of the largely non-informative priors used here.

We use a Metropolis–Hastings sampling scheme with random walk updates to simulate from the posterior density for $\boldsymbol{\psi} = (\mu, \sigma, \xi, \theta)$. We implement the Metropolis–Hastings sampler over $R = 50{,}000$ iterations, to obtain a posterior sample $\boldsymbol{\psi}^{(1)}, \ldots, \boldsymbol{\psi}^{(50{,}000)}$, and then use the posterior draws $\theta^{(1)}, \ldots, \theta^{(50{,}000)}$ to obtain posterior samples for the runs declustering parameter $\kappa$, as well as the two cluster functionals $\rho$ and $\omega$ (using steps 3–5 of the sampling scheme discussed in Section 2.5). The sampler was initialised at various starting points to confirm convergence to the stationary distribution; trace plots from the MCMC output (not shown here) confirmed quite rapid convergence, although to be safe we discard the first 2,000 observations as 'burn-in'. Posterior densities after the removal of burn-in are summarised in Table 1. The posterior densities for $\mu, \sigma$ and $\xi$ are not of primary interest here and so summaries for these parameters are omitted from this table. Also shown in Table 1, for comparison, are the corresponding maximum likelihood estimates for $\theta$ for both simulated series (see Section 3.2) as well as the maximum likelihood estimates for the (limiting) mean cluster size $\rho$, with asymptotic standard errors in parentheses.

**Table 1** Posterior means (and standard deviations in parentheses) for the extremal index ($\theta$), the declustering parameter ($\kappa$) and the two cluster functionals ($\rho$ and $\omega$), as well as 95% credible intervals

| | $\theta$ | $\kappa$ | $\rho$ | $\omega$ |
|---|---|---|---|---|
| | (=0.062) | | | |
| Posterior mean (SD) | 0.065 (0.050) | 14.035 (4.002) | 14.648 (0.707) | 53.343 (10.546) |
| 95% credible interval | (0.031, 0.165) | (6, 23) | (12.203, 18.321) | (32.222, 77.372) |
| m.l.e. (asymp. SE) | 0.058 (0.052) | – | 13.317 (1.193) | – |
| | (=0.328) | | | |
| Posterior mean (SD) | 0.319 (0.048) | 5.974 (3.678) | 3.048 (0.237) | 39.662 (5.440) |
| 95% credible interval | (0.225, 0.416) | (2, 16) | (2.618, 3.420) | (29.715, 50.867) |
| m.l.e. (asymp. SE) | 0.297 (0.047) | – | 3.361 (0.233) | – |

Also shown, for comparison and where available, are the corresponding maximum likelihood estimates (m.l.e.s; and associated asymptotic standard errors in parentheses)

Overall, this simulation study has demonstrated the viability of our approach to inference. The Bayesian sampling scheme provides a more complete inferential procedure than non-Bayesian techniques which are based on the empirical identification of clusters of extremes (see Section 2.1), whilst also avoiding the somewhat complex and non-trivial modelling issues surrounding simulation-based inference schemes such as those in Smith et al. (1997). Further, it enables direct inference on the cluster functionals of interest.

## 4 Application to Bradfield Wind Speed Data

### 4.1 Markov Chain Monte Carlo Simulations

We now apply the MCMC scheme used for the simulated data in Section 3 to make inferences for the extremal index and the two cluster functionals for a set of real-life wind speeds. The data we use consist of hourly maximum observations collected over a period of 10 years (January 1st 1975–December 31st 1984) at High Bradfield in the Peak District of Northern England (see Fawcett 2005, Chapter 2). A time series plot of the first three years of this dataset is shown in Fig. 2, and the strong seasonal element in our data is clear. We deal with this by analysing each 'season' separately—here, as closely as possible, we divide the year into twelve 'months' of equal length, and use these as our seasons. The fitting of distinct model components for each month is a tried-and-tested compromise between retaining a sufficient amount of data for modelling, while reducing seasonal variation within each modelling unit to a level which can be ignored (see, for example, Walshaw 1994). Also shown in Fig. 2 is a plot of the time series against the version at lag 1, to illustrate the degree of temporal dependence between successive observations. The lag 1 autocorrelation coefficient for our series is 0.960.

As in the simulation study, we specify independent, non-informative prior distributions for each of the GEV parameters and the extremal index itself. We use the same MCMC scheme to sample 50,000 iterations on $\theta$ and $\kappa$. Samples

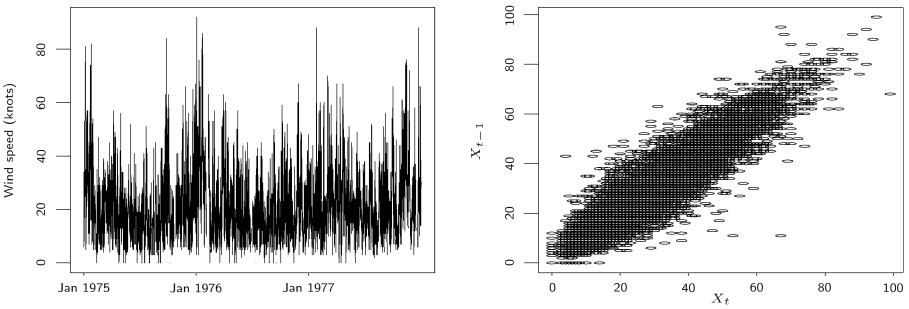


**Fig. 2** Time series plot for the first three years (1975–1977 inclusive) of hourly maximum wind speeds at High Bradfield. Also shown is a plot of the time series against the version at lag 1.

of $\rho$ and $\omega$ (mean cluster length and mean interval duration between clusters respectively) are then obtained as before.

To confirm convergence, the MCMC was run from multiple starting points. Again, convergence was achieved quite rapidly. Table 2 summarises the results of the MCMC scheme after the removal of 'burn-in' (again, 2,000 iterations) and, as with Table 1, compares these posterior summaries with the corresponding m.l.e.s (where available). In the context of the Bradfield wind speed data, the functionals $\rho$ and $\omega$ now correspond to the mean storm duration, and the mean interval duration between storms, respectively. The information provided in Table 2 regarding these functionals would be of interest to meteorologists and engineers concerned with the physical properties of storms which these functionals represent.

### 4.2 Return Levels and Predictive Inference

So far, we have discussed how Bayesian inference for the extremal index $\theta$ can be extended to make inferences for cluster characteristics of practical interest. In a practical setting, we are often interested in the estimation of extreme quantiles, or *return levels*, as a primary aim, and so we describe how this can be conducted as part of our procedure for inference on clustered extremes.

**Table 2** Posterior means (and standard deviations in parentheses) for the extremal index ($\theta$), the declustering parameter ($\kappa$) and the two cluster functionals ($\rho$ and $\omega$), as well as 95% credible intervals, for the Bradfield wind speed data (January)

| | $\theta$ | $\kappa$ | $\rho$ | $\omega$ |
|---|---|---|---|---|
| Posterior mean (SD) | 0.243 (0.047) | 5.266 (5.840) | 4.924 (0.637) | 82.747 (15.271) |
| 95% credible interval | (0.162, 0.347) | (2, 24) | (4.289, 6.246) | (56.435, 117.259) |
| m.l.e. (asymp. SE) | 0.207 (0.042) | – | 4.833 (0.578) | – |

Also shown, for comparison and where available, are the corresponding maximum likelihood estimates (m.l.e.s; and associated asymptotic standard errors in parentheses)

A typical application of the generalised extreme value (GEV) distribution is to fit Eq. 2 to a series of independent block maxima. Estimates of extreme quantiles can then be obtained by inversion of Eq. 2, to give

$$z_r = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-\log(1 - 1/r)\right\}^{-\xi}\right],$$

for $\xi \neq 0$, where $G(z_r) = 1 - r^{-1}$ and $z_r$ is the return level associated with the $r$-year return period. In the case of $\xi = 0$, the c.d.f. is of Gumbel form, obtained as the limit as $\xi \to 0$ of Eq. 2, and given by $G(x; \mu, \sigma) = \exp\{-\exp[-(x - \mu)/\sigma]\}$. Inverting this gives

$$z_r = \mu - \sigma \log\left[-\log(1 - 1/r)\right].$$

Now for a dependent series with extremal index $\theta$, standard arguments in Leadbetter et al. (1983), Chapter 3, show that, for large $n$ and $x$, typically

$$\Pr\left\{\max\left(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\right) \leq x\right\} \approx G^{n\theta}(x). \tag{7}$$

Setting $x = z_r$ in approximation (7), equating this to $1 - r^{-1}$ and solving for $z_r$ gives, to a good approximation, the $r$-year return level of the process.

With the Bradfield wind speed data, we have the added complexity of seasonal variation. The seasonal estimates of the GEV parameters $\mu$, $\sigma$ and $\xi$, and the extremal index $\theta$, can be recombined to give overall return level estimates for Bradfield. The *annual* exceedance rate $z_r$ is given by

$$\sum_{m=1}^{12} \left\{1 - G_m^{h_m\theta_m}(z_r)\right\}, \qquad m = 1, \ldots, 12, \tag{8}$$

where $1 - G_m^{h_m\theta_m}(z_r)$ is the annual exceedance rate of $z_r$ in month $m$ (obtained from approximation (7)), $G_m$ is the GEV distribution function in month $m$ with parameters $\mu_m$, $\sigma_m$ and $\xi_m$ (from model (2)), $h_m$ is the number of hours in month $m$, and $\theta_m$ is the extremal index in month $m$. Expression (8) can then be set equal to $r^{-1}$ and solved for $z_r$ for each posterior draw of $\mu_m$, $\sigma_m$, $\xi_m$ and $\theta_m$ to obtain posterior draws for the $r$-year return level $z_r$.

If we denote by $\pi(\boldsymbol{\psi}|\boldsymbol{x})$ the posterior distribution for $\boldsymbol{\psi} = (\mu_m, \sigma_m, \xi_m, \theta_m)$ given data $\boldsymbol{x} = (x_1, \ldots, x_n)$, then the *posterior predictive distribution* is given by:

$$\Pr(X \leq x|x_1, \ldots, x_n) = \int_{\Psi} \Pr(X \leq x|\boldsymbol{\psi})\pi(\boldsymbol{\psi}|\boldsymbol{x})d\boldsymbol{\psi}. \tag{9}$$

Solving

$$\Pr(X \leq z_{r,\text{pred}}|x_1, \ldots, x_n) = 1 - r^{-1}$$

for $z_{r,\text{pred}}$ therefore gives the posterior predictive estimate of the $r$-year return level. This estimate incorporates uncertainty due to model estimation, in addition to the uncertainty associated with future observations given any particular model. Although Eq. 9 is analytically intractable, it is easily approximated through the estimated posterior distribution (which is found by simulation).

**Table 3** Posterior means (and standard deviations in parentheses) for the 10, 50, 200 and 1000 year return levels for the Bradfield wind speed data, along with 95% credible intervals

| | $z_{10}$ | $z_{50}$ | $z_{200}$ | $z_{1000}$ |
|---|---|---|---|---|
| Posterior mean (s.d.) | 96.48 (2.38) | 103.86 (4.51) | 109.65 (7.74) | 116.73 (12.16) |
| 95% credible interval | (91.42, 102.77) | (96.82, 115.55) | (99.88, 132.01) | (104.62, 149.33) |
| m.l.e. (asymp. s.e.) | 96.75 (2.86) | 103.24 (5.93) | 108.15 (8.79) | 113.31 (12.22) |
| 95% profile confidence interval | (93.70, 100.35) | (99.77, 107.91) | (103.88, 113.09) | (108.52, 118.79) |
| Posterior predictive | 100.71 | 111.96 | 124.28 | 144.94 |

Also shown, for comparison, are the corresponding m.l.e.s (and asymptotic standard errors in parentheses), and 95% confidence intervals based on the profile likelihood. The bottom row of the table gives the *predictive* estimates of these return levels

After deletion of burn-in, we have a sample $\boldsymbol{\psi}^{(1)}, \ldots \boldsymbol{\psi}^{(B)}$ which may be regarded as observations from the stationary distribution $\pi(\boldsymbol{\psi}|\boldsymbol{x})$. Thus,

$$\Pr(X \leq z_{r,\text{pred}}|x_1, \ldots, x_n) \approx \frac{1}{B}\sum_{r=1}^{B} \Pr(X \leq z_{r,\text{pred}}|\boldsymbol{\psi}^{(r)}), \tag{10}$$

and we can solve approximation (10) by using a standard numerical method (see Coles and Tawn 1996, for details of these ideas).

For Bradfield, Table 3 shows the posterior predictive estimates of the return levels $z_r$, together with the posterior means (and standard deviations in parentheses), for return periods $r = 10$, 50, 200 and 1000 years. Also shown are 95% credible intervals taken from the simulated posterior distribution for these return levels. For comparison, the corresponding m.l.e.s (and associated asymptotic standard errors) are also shown. Due to the severe asymmetry of the likelihood surface often encountered for return levels, in the likelihood analysis we report 95% confidence intervals based on the *profile likelihood* (see, for example, Venzon and Moolgavkar 1988). The posterior predictive return levels shown in Table 3 are arguably the most useful information for anyone interested in design-level specifications, as they provide a pure representation of the exceedance probability for a given wind-speed. Unlike any point estimates provided by classical inference, these values have taken into account the model uncertainty, in addition to the within-model process variability.

**Acknowledgements** We thank the U.K. Meteorological Office for supplying the wind speed data used in this paper. We also thank a referee for helpful comments. Lee Fawcett's work was part-funded by an EPSRC Doctoral Training Award.

## References

Ancona-Navarrete, M.A., Tawn, J.A.: A comparison of methods for estimating the extremal index. Extremes **3**, 5–38 (2000)

Bottolo, P., Consonni, G., Dellaportos, P., Lijoi, A.: Bayesian analysis of extreme values by mixture modelling. Extremes **6**, 25–48 (2003)

British Standards Institution: Code of Basic Data for the Design of Buildings; CP 3, ch. V, Loading; part 2, Wind Loads. British Standards Institution, London. Waves in a record. Proc. R. Soc. Lond. A **247**, 22–48 (1997)
Coles, S.G.: An introduction to statistical modeling of extreme values. Springer, London (2001)
Coles, S.G., Powell, E.: Bayesian methods in extreme value modelling: a review and new developments. Internat. Statist. Review. **64**, 119–136 (1996)
Coles, S.G., Tawn, J.A.: A bayesian analysis of extreme rainfall data. Appl. Statist. **45**, 463–478 (1996)
Fawcett, L.: Statistical methodology for the estimation of environmental extremes. Ph.D. thesis, University of Newcastle upon Tyne (2005)
Fawcett, L., Walshaw, D.: A hierarchical model for extreme wind speeds. Appl. Stat. **55**(5), 631–646 (2006a)
Fawcett, L., Walshaw, D.: Markov chain models for extreme wind speeds. Environmetrics **17**(8), 795–809 (2006b)
Fawcett, L., Walshaw, D.: Improved estimation for temporally clustered extremes. Environmetrics **18**(2), 173–188 (2007)
Ferro, C.A.T., Segers, J.: Inference for clusters of extreme values. J. R. Statist. Soc. B **65**, 545–556 (2003)
Gomes, M.I.: On the estimation of parameters of rare events in environmental time series. In: Barnett, V., Turkman K.F. (eds.) Statistics for the Environment 2: Water Related Issues, pp. 225–241 (1993)
Hsing, T., Hüsler, J., Leadbetter, M.R.: On the exceedance point process for a stationary sequence. Prob. Theory Rel. Fields **78**, 97–112 (1988)
Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Series. Springer-Verlag, New York (1983)
Leadbetter, M.R., Rootzén, H.: Extremal theory for stochastic processes. Ann. Probab. **16**, 431–476 (1988)
Loynes, R.M.: Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Statist. **36**, 993–999 (1965)
Newell, G.F.: Asymptotic extremes for *m*-dependent random variables. Ann. Math Statist. **35**, 1322–1325 (1964)
O'Brien, G.L.: The maximum term of uniformly mixing stationary processes. Z. Wahrscheinlichkeitsth. **30**, 57–63 (1974)
Smith, R.L.: The extremal index for a Markov chain. J. Appl. Prob. **29**, 37–45 (1992)
Smith, R.L.: Bayesian and frequentist approaches to parametric predictive inference (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.) Bayesian Statistics, vol. 6, pp. 589–612. Oxford University Press (1999)
Smith, R.L., Goodman, D.J.: Bayesian risk analysis. In: Embrechts, P. (ed.) Extremes and Integrated Risk Management, pp. 235–251. Risk Books: London (2000)
Smith, A.F.M., Roberts, G.O.: Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. R. Statist. Soc., B **55**, 3–23 (1993)
Smith, E.L., Walshaw, D.: Modelling bivariate extremes in a region. In: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M. (eds.) Bayesian Statistics, vol. 7, pp. 681–690. Oxford University Press (2003)
Smith , R.L., Weissman, I.: Estimating the extremal index. J. R. Statist. Soc., B **56**, 515–528 (1994)
Smith, R.L., Coles, S.G., Tawn, J.A.: Markov chain models for threshold exceedances. Biometrika **84**, 249–268 (1997)
Stephenson, A., Tawn, J.A.: Bayesian inference for extremes: accounting for the three extremal types. Extremes **7**, 291–307 (2004)
Tawn, J.A.: Bivariate extreme value theory: models and estimation. Biometrika **75**, 397–415 (1988)
Venzon, D.J., Moolgavkar, S.H.: Profile-likelihood-based confidence intervals. Appl. Stat. **37**, 87–94 (1988)
Walshaw, D.: Getting the most from your extreme wind data: a step by step guide. J. Res. Natl. Inst. Stand. Technol. **99**, 399–411 (1994)