Accounting for threshold uncertainty in extreme value estimation

A. Tancredi, C. Anderson, A. O'Hagan

2002.12

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Settembre 2002

Accounting for Threshold Uncertainty in Extreme Value Estimation

Andrea Tancredi * Clive Anderson

Dipartimento di Scienze Statistiche Department of Probability and Statistcs

Università di Padova University of Sheffield

Anthony O'Hagan

Department of Probability and Statistcs

University of Sheffield

Summary

Tail data are often modelled by fitting a generalized Pareto distribution (GPD) to the exceedances over high thresholds. In practice, a threshold u is fixed and a GPD is fitted to the data exceeding u. A difficulty in this approach is the selection of the threshold above which the GPD assumption is appropriate. Moreover the estimates of the parameters of the GPD may depend significantly on the choice of the threshold selected. Sensitivity with respect to the threshold choice is normally studied but this does not fully take account of threshold uncertainty.

We propose to model extreme and non extreme data with a distribution composed by a piecewise constant density up to an unknown end point α and by a GPD with threshold α for the remaining tail part. Since we estimate the threshold together with the other parameters of the GPD we take naturally into account the threshold uncertainty. We will discuss this model from a Bayesian point of view and the method will be illustrated using simulated data and two real data sets.

Some key words: Extreme value theory, Generalized Pareto Distribution, Reversible jump algorithm, Threshold estimation, Uniform mixtures

*Corresponding author; Dipartimento di Scienze Statistiche, Università di Padova, Via Cesare Battisti 241-249, 35121, Padova, Italy; E-mail tancredi@stat.unipd.it

1 Introduction

Extreme value theory is concerned with the statistical study of the extremal behaviour of random variables. Practical applications of extreme value modelling can be found in many fields. For example, in environmental sciences, hydrologists may be interested in studying the extreme levels of a river or of the sea in order to build structures able to withstand severe floods. Other important applications can be found in financial analyses. For example, an insurance company may be interested to predict the largest claims which it will have to face in the course of a given year.

From a statistical point of view the problem of extreme value theory is an extrapolation problem. The basic idea to conduct such extrapolation is to find a good parametric model for the tail of the data generating process and then to fit this model with the extreme observations. The extrapolation is thus obtained analysing the behaviour of the model beyond the range of the observed data.

Extreme value models are becoming more popular but also more sophisticated. In fact in recent years great efforts have been made both to take into account covariates to handle stationarity (Smith,1989; Davison and Ramesh, 2000; Pauli and Coles, 2001) and to model multivariate extremes (Coles and Tawn, (1991, 1994)). In this paper we will try to address a methodological problem which arises even in the simplest extreme value models.

Given a time series $(x_1, \ldots, x_n)$ collected in a period of T years, Davison and Smith (1990) propose to model the values greater than a fixed high threshold u as independent observations from a generalized Pareto distribution (GPD) with density

$$\frac{1}{\sigma}\left(1+\xi\frac{x-u}{\sigma}\right)_+^{-1/\xi-1} \tag{1}$$

for $x > u$ where $y_+ = \max(y, 0)$, ξ is a real-valued shape parameter, and $\sigma\,(\sigma > 0)$ is a scale parameter. Moreover they suppose that the number N of exceedances of the threshold u in one year is distributed as a Poisson variable with parameter λ. With these assumptions, justified by probabilistic theory (Pickands (1975); Embrechtes *et al.*, 1997), the mean crossing rate of a value x greater than u in one year is $\lambda(1+\xi(x-u)/\sigma)_+^{-1/\xi}$. Setting this rate equal to $1/n$ and solving for x we obtain the return level associated to a return period of n years

$$q_n = u - \frac{\sigma}{\xi}\{1-(\lambda n)^{\xi}\} \tag{2}$$

which, basically, is the value exceeded in mean once every n years. The extrapolation process is then conducted estimating the return levels (2) corresponding to high return periods.

One crucial assumption of this procedure is the imposition of the threshold value u before the fitting of the model. Such imposition is not troublesome if the scope of the analysis is simply exploratory: identifying a model able to describe the observed high values of a given phenomenon. In this case, it is enough to report the estimated model for the data that have to be described together with a measure of the accuracy of the description, i.e. a measure of goodness of fit. In this situation the choice of a threshold is, in fact, an aspect of the design which must be tackled before the fitting of the model.

On the other hand, the scope of the analysis may be that to give a single estimate of the return level corresponding to a given return period and to attach a measure of the uncertainty to the estimate. In this situation, which we assume to be the case throughout this article, the imposition of a fixed threshold is quite unnatural. In fact, it is reasonable to expect a different estimation of the return levels and a different precision of estimation for each threshold. Thus it may be more sensible to treat the threshold as a parameter of a model. In this way we can mix over all threshold values that are determined as suitable by the the model and the data to get an overall estimate of return levels together with their variability.

Notwithstanding this, it is still standard practice to follow a threshold fixed approach, estimating the lower threshold that is consistent with the GPD assumption and conducting the extrapolation with this threshold. A recent method of threshold estimation for the GPD is given by Dupius (1998), while Guillou and Hall (2001) consider the analogue problem of fixing the number of the largest order statistics. Sensitivity with respect to the threshold is normally studied by comparing the results obtained with different thresholds, but there can be different sensible thresholds which lead to different estimates. Moreover, even if we are able to select an optimal threshold, the problem remains of how incorporate threshold uncertainty when we want to estimate the return levels.

Frigessi *et al* (2001) give a first solution to the problem just described. In fact, for a random sample of positive observations, they propose the following dynamically weighted mixture model

$$f(x) = \frac{[1 - p(x|\mu, \tau)]h(x|\beta, \nu) + p(x|\mu, \tau)g(x|\sigma, \xi)}{\int_0^\infty [1 - p(x|\mu, \tau)]h(x|\beta, \nu) + p(x|\mu, \tau)g(x|\sigma, \xi)dx} \tag{3}$$

where $g(x|\sigma, \xi)$ is the density (1) with $u = 0$, $h(x|\beta, \nu) = \beta\,\lambda^\beta\,x^{\beta-1}\exp-(\lambda\,x)^\beta$ is the density of a Weibull distribution and $p(x|\mu, \tau) = 1/2 + (1/\pi)\arctan((x - \mu)/\tau)$. The mixing function $p(x|\mu, \tau)$ models the way in which the GPD becomes predominant for the fitting of extreme observations. Thus it is expected that $p(x|\mu, \tau)$ goes rapidly from 0 to 1 around the best threshold levels for the fitting of the GPD. The threshold uncertainty is naturally taken into account, but we see some limitations in doing extreme value inference with this model. In fact, the applicability of the model is limited to heavy-tail distributions. With short-tail distributions the tail of the Weibull can, in fact, interfere with the extrapolation process. Moreover the Weibull assumption and the parametric formulation of the mixing function $p(x|\mu, \tau)$ can be inadequate for modelling respectively the bulk of data and the switching from the non-extreme to the extreme observations. In these cases we may have unexpected repercussions on the fitting of the complete model and consequently on the extrapolation.

Our model for extreme value inference is explained in the next section. Basically we try to merge a generalized Pareto with unknown threshold with a piecewise constant density function. Our approach to inference is Bayesian and in the third section we give the details of the prior modelling. Markov chain Monte Carlo methods are needed for estimating the parameters of the model. A suitable algorithm is described in section 4. In section 5 we assess the performance of the model both with simulated data and with real data, while in section 6 we give a brief discussion for subsequent modelling and research.

2 Models for univariate extremes

2.1 Standard models

The generalized extreme value distribution (GEV)

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

plays a central role in modelling maxima and exceedances. In fact, in the simplest case, it represents the complete class of limit distributions for the renormalised maximum of independent observations $X_1, \ldots, X_n$ with a common distribution F. In the classical approach to extreme value inference for univariate processes the GEV is fitted to the maxima of block observations.

To maximise data utility most extreme value models work on the principle of modelling threshold exceedances. A basic formulation for modelling threshold exceedances is given by the Poisson-GPD model. Let (Y_n) be a sequence of independent and identically distributed random variables with common distribution GPD with threshold u scale σ and shape ξ and let N be a random variable having a Poisson distribution with mean λ. Several theoretical properties suggest that the distribution of the vector $(Y_1, \ldots, Y_N)$ can be seen as a limiting approximation for the values and the number of the sample exceedances of an high threshold u. An important link between the GEV and the Poisson-GPD model is given by the fact that

$$Pr\{\max(Y_1, \ldots, Y_N) < x\} = \exp\left\{-\lambda \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}\right\}$$

which turns out to be exactly the GEV distribution with parameters μ, ψ, ξ satisfying

$$\sigma = \psi + \xi(u - \mu). \tag{4}$$

and

$$\lambda = \left(1 + \xi \frac{u-\mu}{\psi}\right)_+^{-1/\xi}. \tag{5}$$

2.2 A threshold mixing model

We suppose the data comprise a series of m independent observations from a random variable X with unknown distribution F. Moreover we suppose that these m observations are collected daily over a period of T years but the methodology that we present can be easily generalized to handle data collected on different scales. We want to estimate the upper tail behaviour of F taking account of the uncertainty of the level at which asymptotic tail models provide reasonable approximations. Let u_0 be a basic low threshold which is supposed to be below such a level. We will assume that the data beyond u_0 are independent and identically distributed observations from a random variable with density

$$f(x) = \begin{cases} (1-w)\, h(x|\omega^{(k)}, a^{(k)}, \alpha) & u_0 < x < \alpha \\ w\, g(x|\alpha, \sigma, \xi) & \alpha \leq x < \infty \end{cases} \tag{6}$$

where $g(x|\alpha, \sigma, \xi)$ is the density (1) with the threshold value u now supposed unknown and called α, w can be interpreted as the probability that an observation from F is greater than α conditional on being greater than u_0 and $h(x|\omega^{(k)}, a^{(k)}, \alpha)$ is a piecewise constant density on $[u_0, \alpha)$ with un unknown number of steps.

The piecewise density $h(x|\omega^{(k)}, a^{(k)}, \alpha)$ has been analyzed in a Bayesian context by Robert (1998) and Robert and Casella (1999). Specifically we have that

$$h(x|\omega^{(k)}, a^{(k)}, \alpha) = \sum_{i=1}^{k} \omega_i I_{[a_i, a_{i+1})}(x) \tag{7}$$

where $a_1 = u_0$, $a_{k+1} = \alpha$ and

$$\sum_{i=1}^{k} \omega_i (a_{i+1} - a_i) = 1. \tag{8}$$

The density (7) is then a step function with an unknown number k of steps at positions $a_1 = u_0 < a_2 < \cdots < a_k < a_{k+1} = \alpha$. The vector $a^{(k)}$ denotes the unknown parameters $(a_2, \ldots, a_k)$. The step function takes the unknown value ω_i, which we call its height, on the subinterval $[a_i, a_{i+1})$ for $i = 1, \cdots, k$. We indicate by $\omega^{(k)}$ the vector $(\omega_1, \ldots, \omega_k)$. By constraint (8) the integral of this step function on $[u_0, \alpha)$ is 1. Moreover, model (7) can be seen as a mixture of k uniform distributions $\mathcal{U}_{[a_i, a_{i+1}]}$ where k is also unknown. In fact if we let

$$\omega_i = \frac{p_i}{a_{i+1} - a_i} \tag{9}$$

we can write density (7) as

$$\sum_{i=1}^{k} p_i \mathcal{U}_{[a_i, a_{i+1}]} \tag{10}$$

with the constraint $\sum_{i=1}^{k} p_i = 1$. In the following we indicate by $p^{(k)}$ the vector of weights $(p_1, \ldots, p_k)$.

To make model (6) consistent with the GEV distribution for the annual maximum it is enough to observe that the annual mean number of exceedences of α can be approximated by $365\, p_{u_0}\, w$ where p_{u_0} is the probability of an observation greater than the basic low threshold u_0. Thus, following parameterization (5), which is usually adopted in the Poisson-GPD model approximation to express the mean number of exceedances in terms of the GEV parameters, we set

$$w = \frac{1}{365\, p_{u_0}} \left(1 + \xi \frac{\alpha - \mu}{\psi}\right)_+^{-1/\xi}. \tag{11}$$

Moreover we adopt for the scale parameter σ exactly expression (4) with the threshold u now replaced by the unknown parameter α.

Finally, considering the parameterization of the step function in terms of uniform mixture, the model (6) becomes

$$f(x) = \begin{cases} \left(1 - \frac{1}{365\, p_{u_0}} \left(1 + \xi \frac{\alpha-\mu}{\psi}\right)_+^{-1/\xi}\right) \sum_{i=1}^{k} \frac{p_i}{a_{i+1}-a_i} I_{[a_i, a_{i+1})}(x) & u_0 < x < \alpha \\ \frac{1}{365\, p_{u_0}} \left(1 + \xi \frac{\alpha-\mu}{\psi}\right)_+^{-1/\xi} \frac{1}{\psi+\xi(\alpha-\mu)} \left(1 + \xi \frac{x-\alpha}{\psi+\xi(\alpha-\mu)}\right)_+^{-1/\xi-1} & \alpha \le x < \infty \end{cases} \tag{12}$$

where μ, ψ, ξ are the GEV parameters of the corresponding annual maximum distribution.

3 Prior modelling

We consider Bayesian estimation of the model (12) assuming that we do not have prior information on the behaviour of the non extreme and the extreme data. This means that we will try to be uninformative both with the prior distribution of the parameters of the piecewise constant density (7) and with the parameters of the GPD. The joint distribution of all the parameters of the model (12) can be expressed by the factorization

$$\pi(k, p^{(k)}, a^{(k)}\alpha, \mu, \psi, \xi) = \pi(k)\pi(p^{(k)}, a^{(k)}, \alpha, \mu, \psi, \xi|k) \tag{13}$$

In equation (13) it is natural to impose some independencies. For example we can assume that a priori the GEV parameters are independent of all the other parameters of the model. Thus we have that

$$\pi(p^{(k)}, a^{(k)}, \alpha, \mu, \psi, \xi|k) = \pi(k)\pi(p^{(k)}, a^{(k)}, \alpha|k)\pi(\mu, \psi, \xi).$$

3.1 Prior distribution $\pi(k, a^{(k)}, p^{(k)}, \alpha)$

A natural prior distribution for the step positions $a^{(k)} = (a_2, \dots, a_k)$ given α and k is

$$\pi(a^{(k)}|k\,,\alpha) = \frac{\Gamma(kf)}{\Gamma(f)^k}\frac{(a_2-a_1)^{f-1}(a_3-a_2)^{f-1}\cdots(\alpha-a_k)^{f-1}}{(\alpha-a_1)^{kf-1}} \tag{14}$$

where $u_0 = a_1 < a_2 < \cdots < a_k < a_{k+1} = \alpha$. With this prior the mean and the variance for the difference $a_{i+1} - a_i$ conditional on α are

$$E(a_{i+1} - a_i|\alpha, k) = \frac{\alpha - u_0}{k}$$

$$Var(a_{i+1} - a_i|\alpha, k) = (\alpha - u_0)^2\left(\frac{f+1}{(kf+1)k} - \frac{1}{k^2}\right)$$

thus increasing f we favour steps with equal length.

For the prior distribution $\pi(p^{(k)}|a^{(k)}, k)$ we have taken a Dirichlet $(\delta(a_2 - a_1), \dots, \delta(\alpha - a_k))$. Specifically we have that

$$\pi(p^{(k)}|k, a^{(k)}, \alpha) = \frac{\Gamma(\sum_{i=1}^k \delta(a_{i+1}-a_i))}{\prod_{i=1}^k \Gamma(\delta(a_{i+1}-a_i))}\prod_{i=1}^k p_i^{\delta(a_{i+1}-a_i)-1} \tag{15}$$

Taking for α a uniform distribution $\mathcal{U}_{[u_0,l]}$ the prior mean and the prior variance for a single weight p_i given k are

$$E(p_i|k) = \frac{1}{k}$$

$$Var(p_i|k) = \frac{\log(\delta(l-u_0)+1)}{k\delta(l-u_0)}\left(1-\frac{f+1}{kf+1}\right)+\frac{f+1}{k(kf+1)}-\frac{1}{k^2}. \quad (16)$$

Analyzing the prior distribution for the step heights $\omega_i = p_i/(a_{i+1}-a_i)$ obtained by the distributions (14) and (15) we can see how our prior model penalizes the fitting of the tail by the step function. In fact we have that

$$E\left(\frac{p_i}{a_{i+1}-a_i}|k,\alpha\right) = \frac{1}{\alpha-u_0}$$

which means that for a given α we expect the same mean for all the steps of the density function, and that

$$Var\left(\frac{p_i}{a_{i+1}-a_i}|k,\alpha\right) = \left[k\left(\frac{f-1/k}{f-1}\right)-1\right]\frac{1}{(\alpha-u_0)^2(\delta(\alpha-u_0)+1)} \qquad \forall f>1. \quad (17)$$

The coefficient of variation of the heights of the steps conditional on α decreases with α. Thus for high values of α we have a prior distribution for the heights which is more concentrated around the prior means $1/(\alpha-u_0)$ than for low values of α. When the step function fits the extreme data we will have the first step much higher than the tail steps, but the prior distribution will give low probability to such step combinations, penalizing in a natural way the overfitting of the tail by the step function.

The prior model for the uniform mixture is completed by supposing that k is drawn from a binomial distribution

$$\binom{k_{max}}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{k_{max}-k}$$

conditioned on $1 \le k \le k_{max}$.

3.2 Prior distribution for the Gev parameters

Choices for the prior distribution of the parameters of extreme value models are discussed in Coles and Powell (1996), Coles and Tawn (1996) and Coles (2001). Coles and Tawn (1996) give an attractive way of eliciting a subjective prior for the parameters of the GEV distribution. When prior information is available their approach can be easily applied to our model. The advantage of working directly with the GEV parameterization is that the parameters μ, ψ, ξ are independent of the threshold level α. Thus it makes sense to assume μ, ψ, ξ independent on α. Moreover a convenient assumption is to consider μ, ψ, ξ independent a priori. We assume $\pi(\mu)$, $\pi(\psi)$, $\pi(\xi)$ almost non informative. Specifically, normal priors with large variance are assumed for the parameters μ, ξ while a gamma distribution with shape and inverse scale parameters both close to zero is assumed for ψ.

4 Computation

Bayesian inference for the model (12) is possible with Markov Chain Monte Carlo (MCMC) methods, see, for example, Robert and Casella (1999). The principle of MCMC methods is the simulation of a Markov Chain whose equilibrium distribution is exactly the distribution in which we are interested.

We denote by θ the state variable of the chain, i.e. the complete vector of the unknown quantities $\theta = (k, p^{(k)}, a^{(k)}, \alpha, \mu, \psi, \xi)$ and by $\pi(d\theta|x)$ the posterior distribution. Our analysis is based on the reversible jump algorithm (Green, 1995) which can be seen as a generalization of the Metropolis-Hasting algorithm to simulate distributions with variable dimension like $\pi(d\theta|x)$. The basic idea of this algorithm is that the transitions from a state of the chain to another can occur with different kinds of moves where every move leads to the modification of one or more components of the vector θ. For our model a complete updating of the vector θ is made scanning the following moves

(a) updating the parameters $p^{(k)}$

(b) updating the parameters $a^{(k)}$

(c) updating the parameter α

(d) updating the parameters μ, ψ, ξ

(e) splitting one step in two or combining two into one.

Moves (a), (b), (c) and (d) do not modify the dimension of the vector θ and their implementation is quite standard. For example, in the updating of the parameters $p^{(k)}$ we use a Gibbs kernel since it is possible to simulate directly from the posterior distribution of $p^{(k)}$ conditional on all the other parameters. A uniform distribution between a_{i-1} and a_{i+1} or a normal random walk can be used as proposal distribution for the updating of a_i. To update the threshold parameter we have found that a griddy Gibbs sampler step, Tanner (1993) §6.4. can be useful to improve the mixing of the chain when the posterior distribution of α presents two well separated modes. In standard situations a normal random walk performs quite well. For the updating of the GEV parameters μ, ψ, ξ Coles (2001) observes that to move with proposal transition orthogonal to the μ ψ, ξ axes produces slow mixing. In fact, the posterior probability of μ, ψ, ξ lies along the curve that produces the values of w with greater posterior density. Fixing the threshold α, the weight w can be estimated with high precision. Thus the curve in the μ, ψ, ξ space that produces the values of w supported by the data will be quite narrow and proposal values in the μ, ψ, ξ directions will encounter a lot of rejections. The difficulty is easiliy avoided with moves orthogonal to the w, σ, ξ axes.

Move (e) propose change to the parameter k and consequently the dimension of the vector θ, and in this case we use the reversible jump methodology. The details for this move can be found in Robert and Casella (1999) §6.5.

5 Examples

In this section we consider, first, a simulated data set. Then we apply our model to the river Nidd data set analyzed in Davison and Smith (1990) and to a data set consisting of observations from the Dow Jones index..

5.1 Simulated data set

We consider a data set simulated from the model (12) with fixed values of the parameters. The goal here is to show the behaviour of the MCMC algorithm in situations where we know the values of the parameters.

We have considered 365 observations simulated from the following mixture of three uniform components and a generalized Pareto with threshold 4, scale 1 and shape 0:

$$f(x) = 0.85(0.3\mathcal{U}_{[0,1.5]} + 0.35\mathcal{U}_{[1.5,2.6]} + 0.35\mathcal{U}_{[2.6,4]}) + 0.15\,\mathcal{GPD}(4,1,0)$$

To specify the prior distribution $\pi(k, p^{(k)}, a^{(k)}, \alpha, \mu, \psi, \xi)$ we need to fix the values of the hyper-parameters δ, f, k_{max}. We set $f = 2$ and $\delta = 5$. The value of δ has chosen in order to have the prior variance of the weights p_i, which is given by (16), similar to the variance of the components

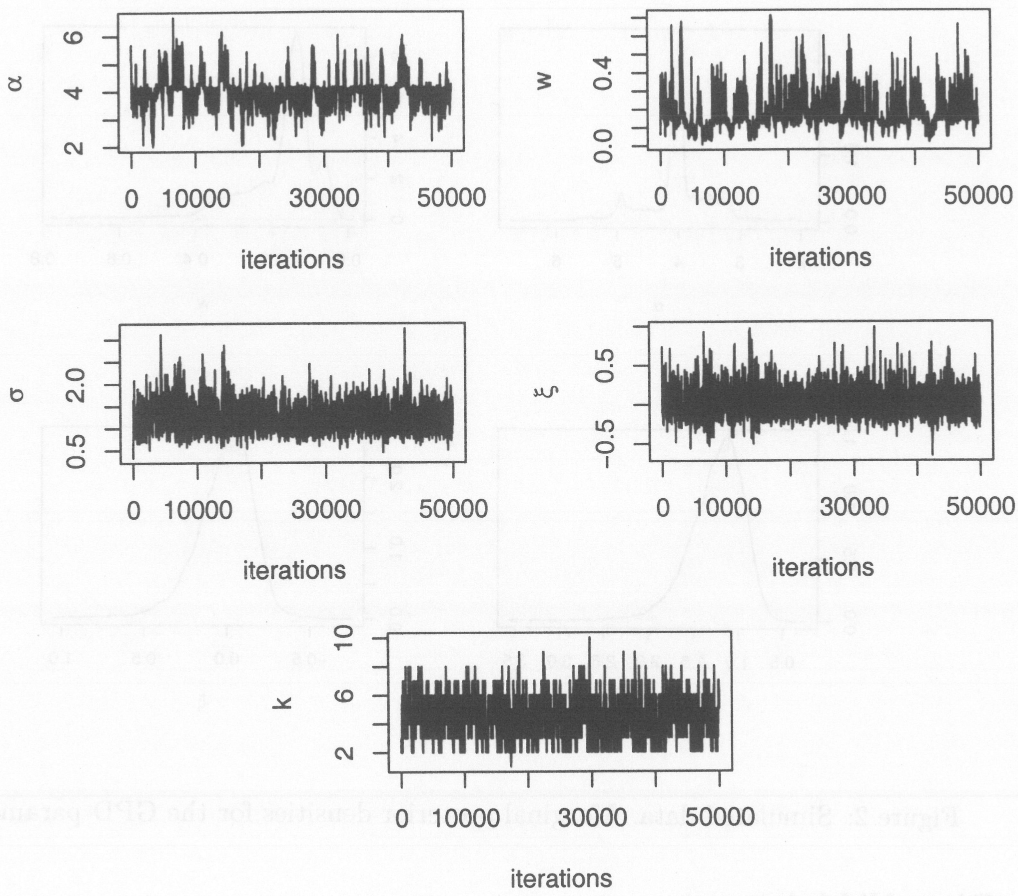

Figure 1: Simulated data. Traces of the MCMC simulations for α, w, σ, ξ and k

of a Dirichlet distribution $D_k(1,\ldots,1)$. For the maximum number of uniform components we set $k_{max} = 10$. As prior distribution for α we took a uniform distribution on the range of data.

A single run of 50000 iterations of the MCMC algorithm discussed previously was used. The variances of the proposal distributions for the GPD parameters were tuned to have an acceptance rate approximately equal to 0.3. Fig. 1 illustrates the changes in the values of the parameters $\alpha, w, \sigma, \xi, k$ against iterations. The chain appears to reach equlibrium with few iterations and then to mix quite well. In fig 2 we give the marginal posterior densities of the GPD parameters α, σ, ξ and the weight w. For all these parameters the posterior distribution are very concentrated around the values which have been used to generate the data.

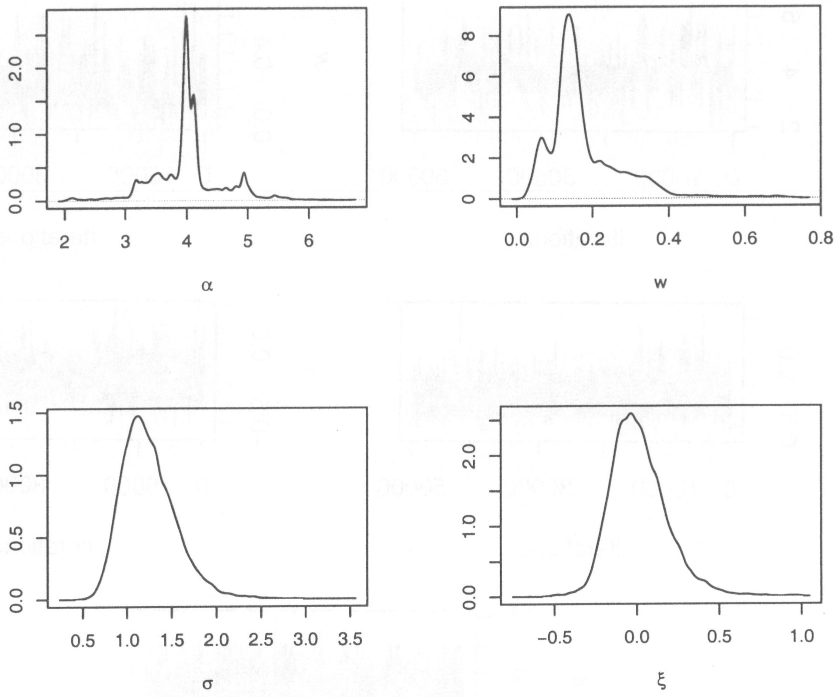

Figure 2: Simulated data. Marginal posterior densities for the GPD parameters

5.2 River Nidd data set

We have applied the model (12) to the river Nidd data set. These data consist of 154 declusterd exceedances of the threshold level 65 m^3s^{-1} by the river Nidd at Hunsingore Weir (U.K.) from 1934 to 1969, see Fig 3. The source is the *Flood Studies Report* (NERC, 1975), volume 4. Previous analyses of these data are in Hosking and Wallis (1987) and Davison and Smith (1990).

We begin by reporting the results, obtained with 100000 iterations of the simulation algorithm, and corresponding to the following settings for the hyperparameters: $\delta = 0.1$, $f = 4\, k_{max} = 20$. In fixing δ our goal is to have a prior variance for the weights simalar to that produced by a Dirichlet $D_k(1,\ldots,1)$. For the threshold α we have taken again a uniform distribution on the range of data.

A density estimation based on the model (12) is displayed in Fig.4. The density estimation is obtained by averaging $f(\cdot|\theta)$ across the MCMC run, hence it takes account of the uncertainty of all the parameters, including k, the number of steps before the threshold. It can be seen like the overall Bayesian predictive density of the distribution of the data. It is evident that the first part of the data, basically the observations before 100, have been modelled semiparametrically by the step function while all the observations above 130 have been modelled by the GPD.

The posterior distribution for k, Fig 5, favours 5 to 9 steps. At least 5 steps are in fact necessary to give the bimodality present in the density estimation before the tail region. While the other few steps are concentrated just before the tail starts.

To assess the performance of the model in terms of threshold estimation we have checked if our estimates are consistent with those obtained with other methods. A graphical approach to threshold estimation is based on the plot of the empirical mean residual life. In fact, if the GPD model starts to be appropriate at the point α, we have that for $\forall x > \alpha$

$$E(X - x|X > x) = \frac{\psi + \xi(x - \mu)}{1 - \xi}. \qquad (18)$$

Thus, plotting the empirical mean $E(X - x|X > x)$ against x we expect to see a linear trend after the threshold α. For the Nidd data, this plot is given by Davison and Smith (1990) and is reproduced here in Fig. 6. Taking account of the increased variability for large threshold values, Davison and Smith (1990) see a rapid increse up to around the threshold 110 and a linear trend after this value. Hence in their analysis they consider threshold values around 110.

The posterior distribution of the threshold obtained with our model (Fig 7.a) is multimodal with the highest mode just around 110. In Fig. 7(b,c,d) we report also the posterior distribution of the parameters μ, ψ, ξ.

Fig. 8 plots the posterior means of the return levels against return periods on a logarithmic scale. We report also the Bayesian estimates with the fixed threshold approach (obtained always averaging the return levels across the MCMC simulations and uninformative priors) and the empirical estimates. With the threshold fixed to 90 the line of the return levels is quite far from the empirical estimate corresponding to the last observation. On the other side, with the threshold equal to 120 the model-based estimation for the last two observations is similar to the empirical estimation. The threshold values 100 and 110 produce a good fit for all the data and the line of the return levels obtained with our model overlaps the line corresponding the threshold 100. Moreover we observe that all the threshold fixed estimates of the return levels are inside the 95% credibility interval obtained taking account of the threshold uncertainty.

We now consider the sensitivity of results with respect to prior assumptions. In Tables 1 and 2 we give the posterior means respectively for the threshold parameter α and for the shape parameter ξ obtained with different choices of the hyperparameters f, δ, k_{max}. Such estimates are

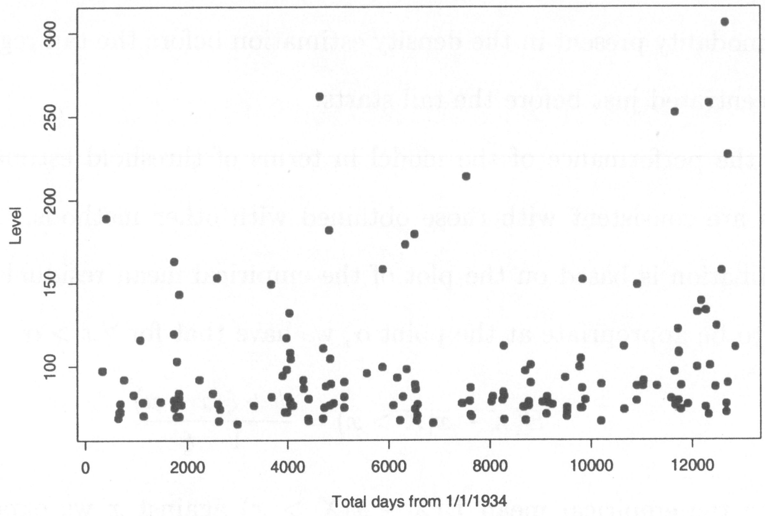

Figure 3: Nidd data. Exceedences of the threshold level 65 $m^3 s^{-1}$.

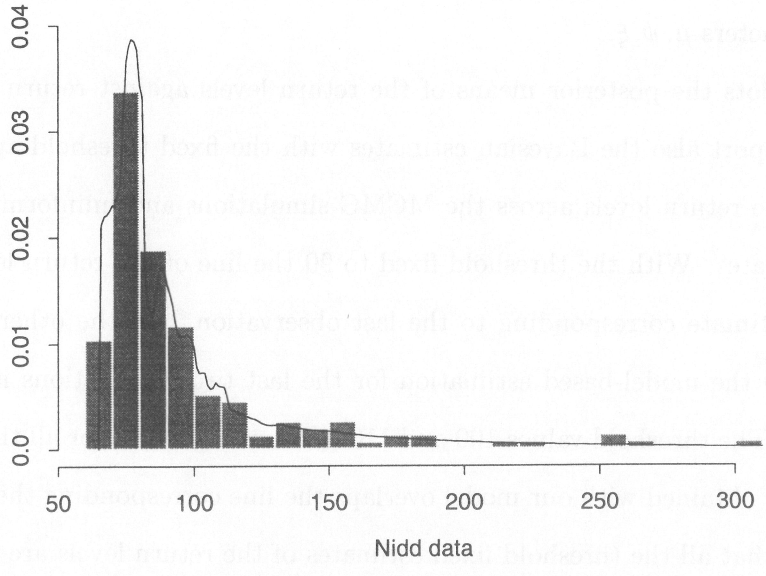

Figure 4: Nidd data. Histogram and posterior density estimation.

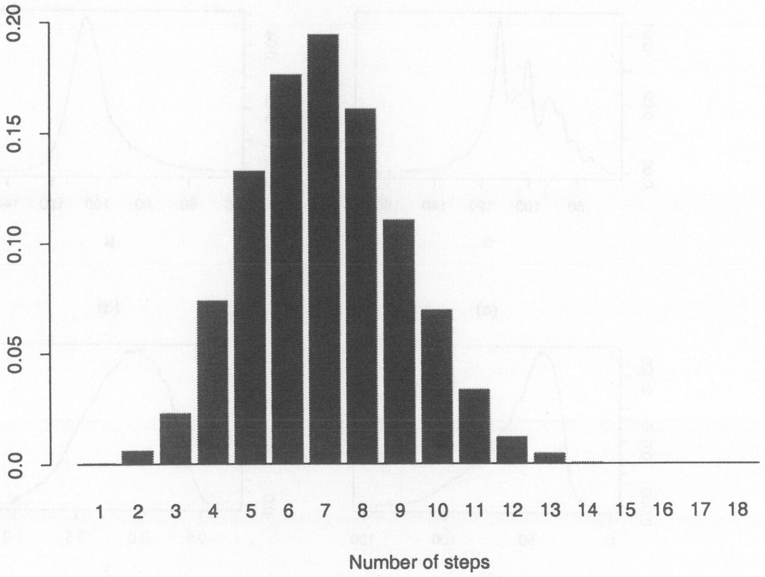

Figure 5: Nidd data. Posterior distribution for k, the number of steps before the threshold .

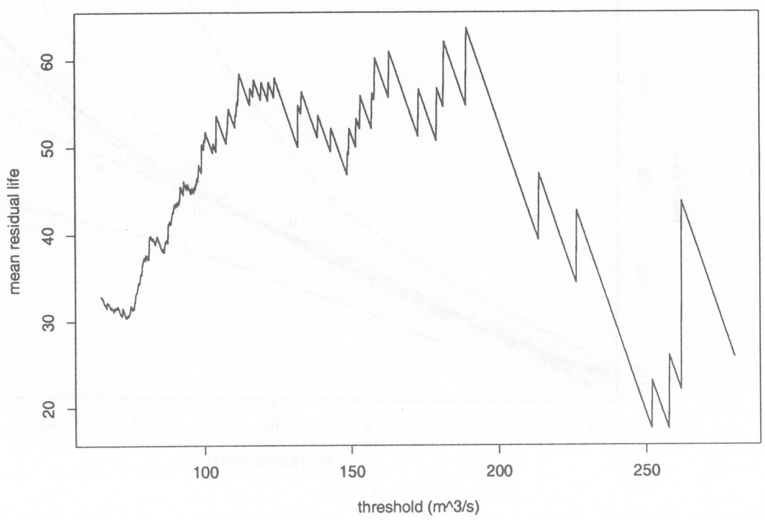

Figure 6: Nidd data. Mean residual life plot.

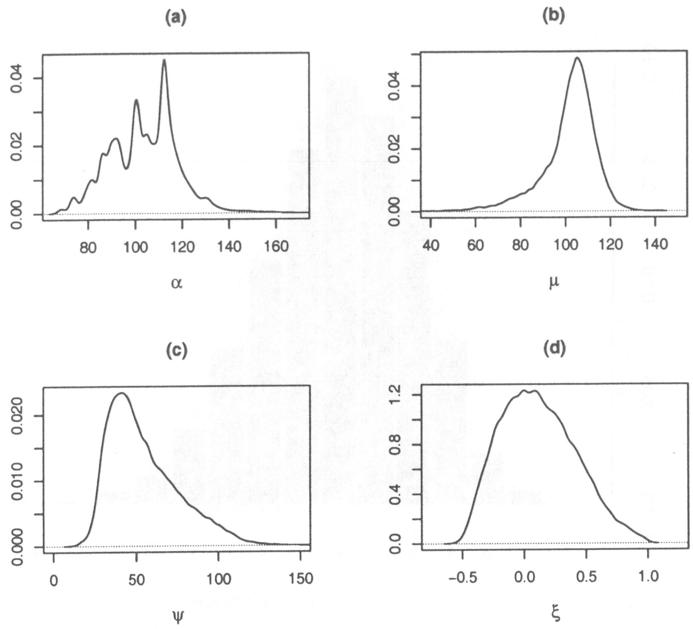

Figure 7: Nidd data. Posterior distribution for the threshold and the GEV parameters

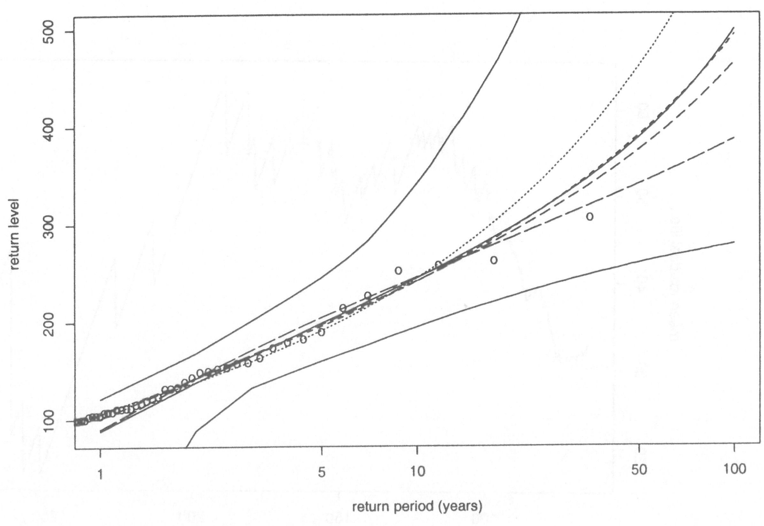

Figure 8: Nidd data. Posterior return levels. Solid line: model (12), dots: u=90, short dashes: u=100, dashes u=110, long dashes: u=120, (○): empirical estimates.

$k_{max} = 10$			
	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.5$
$f = 2$	99	98	90
$f = 3$	100	101	94
$f = 4$	106	100	94
$f = 5$	107	105	92
$f = 6$	110	98	94

$k_{max} = 20$			
	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.5$
$f = 2$	129	97	91
$f = 3$	128	116	94
$f = 4$	129	116	97
$f = 5$	125	113	92
$f = 6$	136	115	96

Table 1: *Nidd data. Posterior means of the threshold parameter* α *for different choices of the hyperparameters* k_{max}, δ, f

	$k_{max} = 10$		
	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.5$
$f = 2$	0.19	0.15	0.27
$f = 3$	0.18	0.15	0.21
$f = 4$	0.09	0.15	0.24
$f = 5$	0.11	0.07	0.24
$f = 6$	0.09	0.18	0.21
	$k_{max} = 20$		
	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.5$
$f = 2$	0.01	0.22	0.28
$f = 3$	-0.06	0.03	0.24
$f = 4$	-0.13	-0.01	0.18
$f = 5$	0.06	0.03	0.26
$f = 6$	-0.06	0.04	0.18

Table 2: *Nidd data. Posterior means of the shape parameter ξ for different choices of the hyperparameters k_{max}, δ, f*

always obtained with 100000 iterations of the simulation algorithm during the equilibrium period.

As expected, when δ increases we have lower values for the posterior mean of α. In fact, δ controls the variability of the step heights around their means with a greater concentration when δ assumes high values, see (17). To have steps with equal lengths we must reduce the support of the step function, thus increasing δ favours low values for α. Increasing the maximum number of steps increases the flexibility of the step function. We expect then a posterior distribution for the threshold translated towards the right. This has been confirmed by our experiments when we have changed the value for k_{max} from 10 to 20. Finally, we observe that there are not substantial differences in the posterior means for α and ξ when we change the hyperparameter f.

5.3 Dow Jones Index data set

We consider a time series given by the logarithm of the ratio of subsequent observations of the Dow Jones index. This data set, which consists of 1304 observations collected over the period 1995-2000 has been previously analyzed by Coles (2001). Fig 9 (a) shows the the time series and Fig 9 (b) shows the mean residual life plot. For these data Coles (2001) suggests to take a threshold equal to 2. For our analysis we took a basic low threshold equal to 1 and we set the following values for the hyperparameters $k_{max} = 20, \delta = 10, f = 3$. Fig 9 (c) shows the data density estimation while Fig. 9 (d) shows the posterior distribution of the threshold α. Again, the posterior distribution confirms the information given by the mean residual life plot. Finally in Fig. 10 we report the estimate of the return levels both with the threshold fixed approach and the mixing threshold approach.

6 Discussion

In this work we have described a Bayesian model for extreme value estimation which takes account of the threshold uncertainty. The key idea has been to merge a very flexible model like a mixture of uniforms with the generalized Pareto distribution. The point where the two models merge is the threshold which is estimated with all the data. The prior distribution has been formulated so that the uniform mixture is appropriate only for the non extreme data where the the GPD model is naturally mispecified. Thus the posterior distribution of the threshold concentrates where the tail starts.

Clearly, the choice to use a mixture of uniforms to model the non extreme data is quite ar-

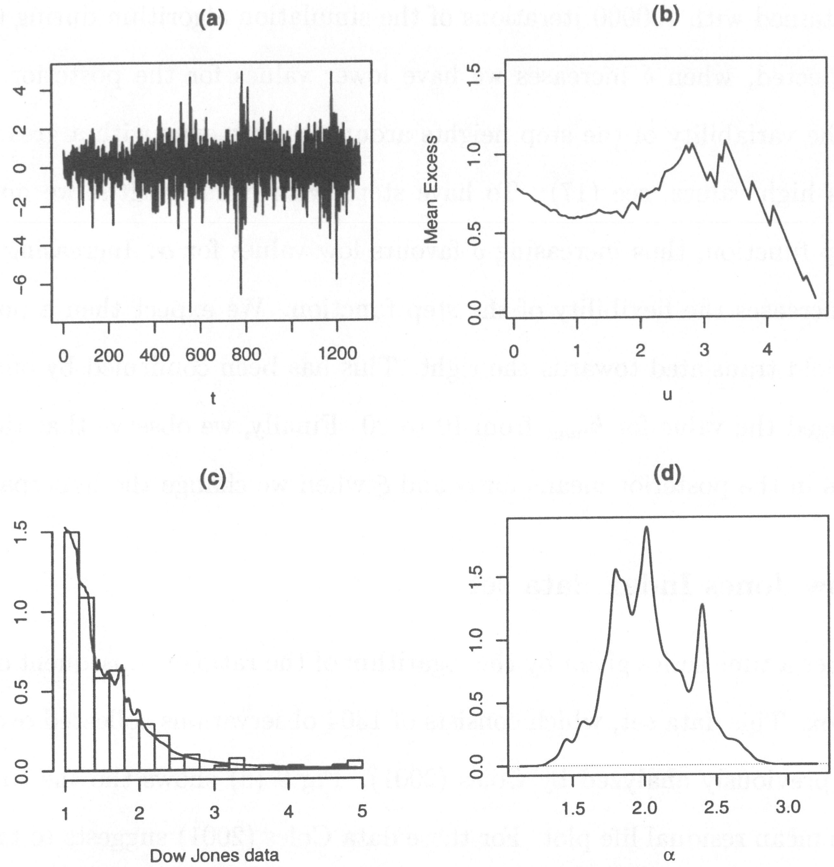

Figure 9: Dow Jones index data: time series (a), mean residual life (b), predictive density (c), posterior distribution for α (d).

bitrary. Other kinds of model like mixture of generalized beta distributions or piecewise linear density functions can be used. However increasing the flexibility of the model for the non extreme observations we increase the risk for an overfitting of the tail, and more involved penalizations may be required to avoid such behaviour.

We are considering two possible extensions of our model. The first regards the situations where temporal covariates must be taken into account. In fact, it is standard practice to assume a fixed threshold also when the other parameters of the GPD are assumed time dependent. In these cases it may be more appropriate to specify a time-varying threshold, and we believe that our model can be easily generalized to handle a temporal dependence of all the GPD parameter including the threshold. The second regards the possibility to extend the threshold mixing model in the case of stationary time series. In particular we are exploring the use of hidden markov models to handle temporal dependence both at extreme and non-extreme levels.

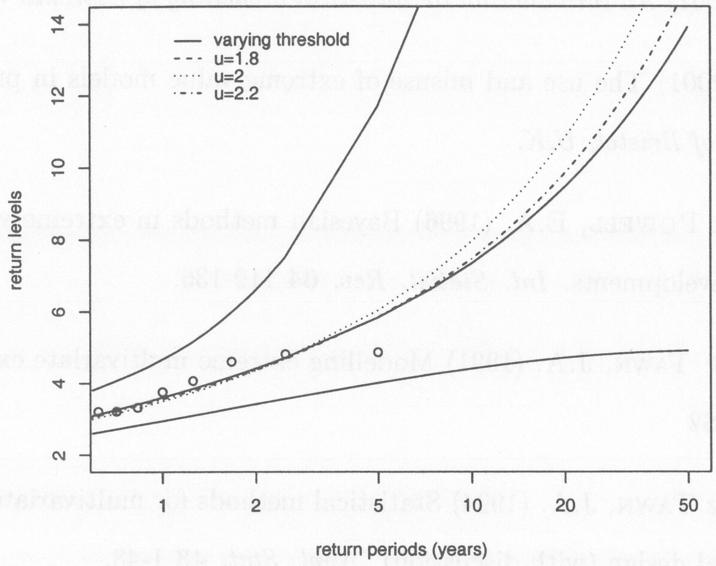

Figure 10: Dow Jones index data: posterior means of the return levels, 95% credibility intervals, posterior means of the returns with fixed thresholds and empirical estimates.

References

COLES, S.G. (2001) *An Introduction to Statistical Modelling of Extreme Values.* London: Springer.

COLES, S.G. (2001) The use and misuse of extreme value models in practice. *Technical report, University of Bristol, U.K.*

COLES, S.G. & POWELL, E.A. (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int. Statist. Rev.* **64** 119-136.

COLES, S. G. & TAWN, J.A. (1991) Modelling extreme multivariate extreme. *J. R. Statis. Soc. B* **53** 377-392

COLES, S. G. & TAWN, J.A. (1994) Statistical methods for multivariate extreme: an application to structural design (with discussion). *Appl. Stat.* **43** 1-48.

COLES, S. G. & TAWN, J.A. (1996) A Bayesian analysis of extreme rainfall data. *Appl. Stat.* **45** 463-478

DAVISON, A. C. & RAMESH, N. I. (2000) Local likelihood smoothing of sample extremes. *J. R. Statis. Soc. B* **62** 191-208

DAVISON, A. C. & SMITH, R. L. (1990). Models for excedances over high threshold (with discussion). *J. R. Statis. Soc. B* **52** 393-442

DUPIUS, D.J (1998). Exceedances over high thresholds: a guide to threshold selection. *Extremes* **1** 251-261

FRIGESSI, A. HAUG, O. RUE, H (2001). Tail estimation with the Generalised Pareto distribution without threshold selection. *submitted to Extremes*

GUILLOU, A. & HALL, P (2001). A diagnostic for selecting the threshold in extreme value theory. *J. R. Statis. Soc. B* **63** 293-305

EMBRECHTS, P. KLUPPELBERG, C. MIKOSH T. (1997) *Modelling Extremal Events.* New York: Springer.

GREEN, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination (1995) *Biometrika* **82** 711-732

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

NATURAL ENVIRONMENT RESEARCH COUNCIL (1975) Flood studies report London: NERC.

PAULI F. & COLES S. G. (2001) Penalized likelihood inference in extreme value theory. *Journal of Applied Statistics* **28** 547-560

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

ROBERT, C.P (1999) *Discretization and MCMC convergence assessment. Lecture Notes in Statistics 135* New York: Springer-Verlag.

ROBERT, C.P. & CASELLA G. (1999) *Monte Carlo Statistical Methods.* New York: Springer.

TANNER, M.A (1993) *Tools For Statistical Inference,* 2nd edn. New York: Springer-Verlag.