A Bayesian Analysis of Extreme Rainfall Data

Author(s): Stuart G. Coles and Jonathan A. Tawn

Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 45, No. 4 (1996), pp. 463-478

Published by: Wiley for the Royal Statistical Society

Stable URL: http://www.jstor.org/stable/2986068

Accessed: 11-11-2015 05:54 UTC

Appl. Statist. (1996)
45, *No.* 4, *pp.* 463–478

A Bayesian Analysis of Extreme Rainfall Data

By STUART G. COLES† and JONATHAN A. TAWN

Lancaster University, UK

[Received June 1994. Final revision March 1996]

SUMMARY

Understanding and quantifying the behaviour of a rainfall process at extreme levels has important applications for design in civil engineering. As in the extremal analysis of any environmental process, estimates often are required of the probability of events that are rarer than those already recorded. As data on extremes are scarce, all available sources of information should be used in inference. Consequently, research has focused on the development of techniques that make optimal use of available data. In this paper a daily rainfall series is analysed within a Bayesian framework, illustrating how the careful elicitation of prior expert information can supplement data and lead to improved estimates of extremal behaviour. For example, using the prior knowledge of an expert hydrologist, a Bayesian 95% interval estimate of the 100-year return level for daily rainfall is found to be approximately half of the width of the corresponding likelihood-based confidence interval.

Keywords: Bayesian models; Extreme value theory; Markov chain Monte Carlo method; Point processes; Prior information; Rainfall

1. Introduction

Extreme value modelling of environmental processes is standard practice for the design of many large scale constructions. Using historical data, estimates are required of design parameters which would lead to the construction having a specified low level failure probability. Throughout this paper we highlight the particular difficulties encountered in modelling the extremes of a rainfall process, though the techniques that we propose have wider applicability. Our data correspond to a 54-year series of daily rainfall aggregates measured at a location in the south-west of England. A time series plot of the data given in Fig. 1 highlights the problem: at the most extreme levels data are scarce, whereas for design purposes it is the behaviour at these levels which is of greatest interest.

The use of extreme value models in such a context has a long history, but classical methods are not efficient in terms of data usage—typically, only the most extreme observation per year is modelled. A dominant research theme over the last decade has been the development of inferential procedures that maximize the use of available data. In particular, two modelling approaches have been developed: one based on point process characterizations of extremes, to facilitate the maximal use of data on a single process (Smith (1989), for example); the other based on multivariate

†*Address for correspondence*: Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.
E-mail: s.coles@lancaster.ac.uk

 0035–9254/96/45463

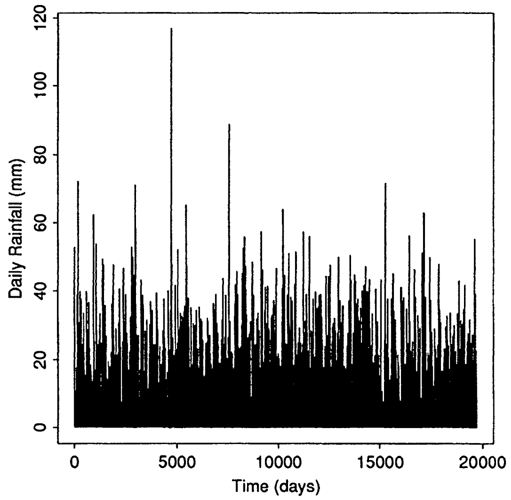

Fig. 1. Time series of daily rainfall aggregates over a 54-year period at a location in south-west England

techniques, in order that information from several processes can be pooled (Coles and Tawn, 1990, 1991).

Given the value of information in the context of extreme value modelling, it is natural to consider including other sources of knowledge in the analysis. This may take the form of known physical constraints, on the maximum value possible for example, or may be derived from an understanding of related processes, perhaps the same variable at a different location. Furthermore, the period over which data have been observed may be known not to be completely representative, and there may be historical evidence, though not in the form of data, of behaviour which is substantially more extreme than that which has been measured. Consequently there are several reasons why it is to be expected that an expert with knowledge of the physical processes involved may have information that is relevant to extremal behaviour, which is independent of the available data. This leads naturally to the Bayesian inferential framework as a basis for undertaking an extreme value analysis.

Very few attempts have been made to incorporate Bayesian methodology into extreme value analysis. Perhaps the most comprehensive analysis is by Smith and Naylor (1987), who examined the effect of different (but arbitrary) prior assumptions on the posterior distributions of the parameters of a Weibull distribution. Pickands (1994) considered Bayesian estimation of extreme quantiles from a theoretical viewpoint, but with limited regard for the issue or importance of determining prior structure. Other investigations have focused on rather specialized subfamilies of models to circumvent the analytical and computational difficulties which arise; see the review by Coles and Powell (1996).

As in many other contexts, the advent of Markov chain Monte Carlo methodology (Gelfand and Smith (1990), for example) has made redundant any major concerns about the computational aspects of a Bayesian analysis of extremes. Thus, we are free to explore more fundamental issues.

(a) Can 'experts' be expected to have meaningful information about extremal behaviour?
(b) How is prior information for extremes best elicited?
(c) How sensitive are extrapolations to changes in prior specification?
(d) How do the results compare with a classical likelihood-based analysis?

With these broad questions in mind, we focus on the analysis of the extremes of the rainfall series presented in Fig. 1. These data were recorded at one of a network of rain-gauges in the south-west of England and span the period 1932–88, excluding three years of data within which there was a substantial proportion of missing observations. This gives rise to a complete series of 19725 daily observations. The site has an altitude of 30 m and a recorded mean annual rainfall of 1.33 m; these features are fundamental to a prior specification of extremal behaviour. The data have previously been analysed within a classical framework by Coles (1994) and Coles and Tawn (1996). In particular, Coles (1994) identified some seasonal characteristics of the series and found the temporal dependence in the series at extreme levels to be extremely weak (see Fig. 1). Our approach here, disregarding the seasonality, is to take a simple point process model as a description of the extremal behaviour, and to consider model estimation using prior information provided by Dr Duncan Reed of the Institute of Hydrology, an acknowledged expert in the field of hydrological science, with specialized knowledge of rainfall behaviour within this particular region (Reed and Stewart, 1989, 1994; Dales and Reed, 1989).

The paper is organized as follows. In Section 2 we outline the asymptotic model used to describe the extremal behaviour of stationary series in general. In Section 3 we describe the elicitation scheme for the prior distribution of the extremal characteristics of our rainfall data, and we discuss the details of computation in Section 4. The results of our analysis are presented in Section 5, highlighting the potential advantages that the Bayesian method has in this extreme value context over classical methods of inference. Section 6 gives a general discussion with ideas for subsequent modelling and research.

2. Likelihood Model

We require a model for the behaviour of the series observed in Fig. 1 at extreme levels of daily rainfall. Since we have no specified model for the underlying distribution of daily rainfall, and it is the tail behaviour only which is of interest here, it is standard to appeal to an asymptotic argument which characterizes such behaviour. This then forms the basis of an inferential model. In this section we outline this argument and illustrate how earlier characterizations of extremal behaviour are derived as special cases.

Let $X_1, X_2, \ldots, X_n$ be a series of independent random variables with common distribution function F and suppose that interest lies in estimating the upper tail behaviour of F. The most general result based on asymptotic argument, but for simplicity stated here as an approximation, is that for large thresholds u the sequence $\{X_1, X_2, \ldots, X_n\}$, viewed on the interval $[u, \infty)$, is approximately a non-homogeneous Poisson process with intensity function

$$\lambda(x) = \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_{+}^{-(\xi+1)/\xi} \tag{2.1}$$

where $z_+ = \max(z, 0)$. In equation (2.1), μ, σ (> 0) and ξ are parameters determined by the tail behaviour of F. Precise details are given by Pickands (1971) and Smith (1989).

Two immediate consequences of this representation are that a likelihood for $\boldsymbol{\theta} = (\mu, \sigma, \xi)$, based on an observed set of exceedances $x_1, \ldots, x_{n_u}$ of a high threshold u, with N taken for reference as the number of observations in a year, is given by

$$L(\boldsymbol{\theta}|\mathbf{x}) = \exp\left(-\frac{n}{N}\,\Lambda[u, \infty)\right) \prod_{i=1}^{n_u} \lambda(x_i), \tag{2.2}$$

where

$$\Lambda[u, \infty) = \int_u^\infty \lambda(x)\,\mathrm{d}x; \tag{2.3}$$

and that

$$\begin{aligned} \Pr\{\max(X_1, \ldots, X_N) \leqslant x\} &= \exp(-\Lambda[x, \infty)) \\ &= \exp\left[-\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}_{+}^{-1/\xi}\right]. \end{aligned} \tag{2.4}$$

The right-hand side of equation (2.4) is usually referred to as the generalized extreme value (GEV) distribution; μ and σ are location and scale parameters respectively, whereas ξ is a shape parameter determining the heaviness of the tail. In particular, $\xi < 0$ corresponds to distributions with a finite upper end point, whereas if $\xi \geqslant 0$ the upper end point to the distribution is infinite. Equation (2.4) is the classical result of extreme value theory and is used to justify the use of the GEV family as a model for annual maximum data. The advantage of working directly with the point process likelihood (2.2) is that all extreme observations (in the sense of exceeding u) are included in the analysis.

Another immediate consequence of the Poisson process model is that, for all thresholds $\tilde{u} > u$,

$$\Pr(X_i - \tilde{u} \leqslant y | X_i > \tilde{u}) = 1 - (1 + \xi y/\tilde{\sigma})_{+}^{-1/\xi} \qquad y \geqslant 0 \tag{2.5}$$

where $\tilde{\sigma} = \sigma + \xi(\tilde{u} - \mu)$. This is the generalized Pareto distribution (GPD) and many extreme value inferences have been based on this model (Davison and Smith, 1990; Pickands, 1975). Furthermore, Pickands's (1994) study of Bayesian estimation of quantiles focuses in particular on the GPD model, with uninformative independent priors for $\tilde{\sigma}$ and ξ. This approach seems restrictive though, since the scale parameter $\tilde{\sigma}$ is dependent on the choice of threshold $\tilde{u}$, so that an uninformative prior for $\tilde{\sigma}$ becomes informative at other thresholds. Consequently, we prefer working explicitly with the point process model (2.1), in which the parameters are independent of the threshold level.

Maximum likelihood estimation based on equation (2.2), or derived from equations (2.4) or (2.5), is non-regular for $\xi < -0.5$ (Smith, 1985), and although this is uncommon for environmental processes an additional argument in support of a Bayesian analysis is that the issue of non-regularity may be avoided (Coles and Powell, 1996).

3. Eliciting Prior Information

For many applications a time series of about 20000 values would be considered sufficiently informative; for extreme value analyses, the data are sparse. In this section we articulate the knowledge of an expert hydrologist to obtain information with which to supplement the data. Through a general understanding of the physical generation of rainfall, and a specific knowledge of the rainfall characteristics within the vicinity of the particular data site, it is reasonable to hope that an expert hydrologist should have valuable prior information about extremal behaviour. The remainder of the paper examines the extent to which this hope can be realized.

An immediate problem concerns parameterization. It is unlikely that prior beliefs on extremal behaviour could be adequately elicited directly in terms of the GEV parameters. Indeed, even if marginal prior forms for each parameter were available, it is by no means clear how to build these into appropriate joint priors. In particular, long-range extrapolation is highly sensitive to the weight of the tail, which in turn is determined by the combination of scale and shape parameters, σ and ξ in equation (2.1). Specifically, increasing σ or ξ leads to a longer-tailed distribution, so *a priori* negative dependence between these parameters is expected. Consequently, to avoid the approach of adopting independent priors on the GEV parameters, as suggested by Smith and Naylor (1987), and similarly within the GPD model by Pickands (1994), we have elicited information within a parameterization that corresponds to a scale on which the expert has familiarity, and within which a natural dependence between the prior specifications is constructed.

Inverting equation (2.4), we obtain that the $1-p$ quantile of the annual maximum distribution is given by

$$q_p = \mu + \sigma[\{-\log(1-p)\}^{-\xi} - 1]/\xi. \tag{3.1}$$

In an engineering context, q_p is referred to as the return level associated with a $\{-\log(1-p)\}^{-1}$-year return period. Values of q_p for small p comprise design parameters in various applications. Thus, our approach is to elicit prior information in terms of $(q_{p_1}, q_{p_2}, q_{p_3})$, for specified (small) values of $p_1 > p_2 > p_3$. This is similar to the approach taken by Crowder (1992), who suggested the construction of priors on the space of probabilities, for fixed quantiles. There are also similarities with the general approach of prior construction via quantiles as summarized by Kadane *et al.* (1980) and Garthwaite and Dickey (1985).

One slight complication with our approach is that the parameters must be ordered $q_{p_1} < q_{p_2} < q_{p_3}$. Hence, we work instead with the differences

$$\left.\begin{aligned} \tilde{q}_1 &= q_{p_1} - e_1, \\ \tilde{q}_2 &= q_{p_2} - q_{p_1}, \\ \tilde{q}_3 &= q_{p_3} - q_{p_2}, \end{aligned}\right\} \tag{3.2}$$

where e_1 is a physical lower end point for the process variable. In the rainfall example, naturally, $e_1 = 0$. We also assume that the priors on these quantities are independent, each with support corresponding to the positive real axis. In particular we take marginal priors of the form

$$\tilde{q}_i \sim \text{gamma}(\alpha_i, \beta_i), \qquad i = 1, 2, 3. \tag{3.3}$$

There is a sense of arbitrariness about this construction, since it means that q_{p_1}, q_{p_2} and q_{p_3} are distributed respectively as sums of one, two and three gamma random variables. Moreover, the assumption of prior independence of the $\tilde{q}_i$ imposes some constraints on the joint prior when viewed in terms of the q_{p_i} or even in the $\boldsymbol{\theta}$-space, but *a priori* such constraints seem reasonable.

Returning then to the issue of prior elicitation, the problem remains of how to obtain prior estimates for the gamma parameters α_i, β_i, $i = 1, 2, 3$, in model (3.3). Again there is some arbitrariness, but the parameters are determined by measures of location and variability in prior belief. Thus we asked the expert for estimates of the median and 90% quantiles of each of the $\tilde{q}_i$ and solved to obtain the corresponding gamma parameter estimates. A summary of this information is given in Table 1. Note that $p_1 = 0.1$, $p_2 = 0.01$ and $p_3 = 0.001$ were the levels on which the expert preferred to state his prior information and correspond to design quantiles which applied engineers most commonly work with.

In determining his prior estimates, Duncan Reed took account of the general rainfall climate over the south–west England peninsula, and also more localized knowledge about the particular characteristics of the site in question, such as the altitude and average annual rainfall. No reference was made to the data available from that site. Previous analyses have also sought to exploit this information by formulating spatial regression models (Coles and Tawn, 1996); our Bayesian analysis facilitates the input of spatial information without the need for complex regression models. However, if covariate information were available, then provided that a greater quantity of prior information could be elicited, in principle there would be no difficulty in adapting the current proposal to include these data through a regression model. Similar issues were discussed by Crowder (1992).

The expert's perceived knowledge of the localized characteristics led him to express a greater relative certainty in his estimate of $\tilde{q}_3$ than in either of the other two parameters, although this parameter required greater extrapolation. To assess the coherence of his prior specification, we asked additionally for prior medians and percentage points of three other quantiles: the 30-, 300- and 3000-year return levels. A comparison of the expert's specified values and the corresponding statistics obtained on the basis of his specified prior for $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$ is given in Table 2. At both the 30- and the 300-year return levels the agreement is excellent (relative errors in the

TABLE 1
Elicited prior medians and 90% quantiles for distributions of $\tilde{q}_i$ with associated gamma parameters for the prior distribution

p_i	*Median (mm)*	*90% quantile (mm)*	α_i	β_i
0.1	59	72	38.9	0.67
0.01	43	70	7.1	0.16
0.001	100	120	47.0	0.39

TABLE 2
Comparison of elicited and derived prior estimates for statistics of return values associated with a range of return periods

Return period (years)	*Specified prior values (mm)*		*Derived prior values (mm)*	
	Median	*90% quantile*	*Median*	*90% quantile*
30	75	95	74.0	89.3
300	140	170	144.0	173.3
3000	260	300	350.7	419.3

prior mean of 1.3% and 2.8% respectively), suggesting that the expert's prior specification, and our imbedding of that information within the joint gamma structure, is consistent with the expert's true beliefs. However, at the 3000-year level, there is a considerable discrepancy between the values derived from the expert's original prior and his explicit stated beliefs for these quantities (a relative error in the prior mean of 34.9%). The expert's response to this apparent discrepancy was a reaffirment of his original prior with an explanation that, in his view, extrapolations based on the GEV model (2.4) are unlikely to be valid at very extreme levels. Thus the inconsistency is attributable to a disbelief in the likelihood model rather than to any incoherency in the prior. Thus we proceed on the basis of the original prior but maintain caution about the validity of the analysis for extrapolation to very extreme levels.

4. Computation

From the prior specification given by equations (3.2) and (3.3), the joint prior for the q_{p_i} is obtained as

$$f(q_{p_1}, q_{p_2}, q_{p_3}) \propto q_{p_1}^{\alpha_1 - 1} \exp(-\beta_1 q_{p_1}) \prod_{i=2}^{3} (q_{p_i} - q_{p_{i-1}})^{\alpha_i - 1} \exp\{-\beta_i(q_{p_i} - q_{p_{i-1}})\}, \qquad (4.1)$$

on $0 \leqslant q_{p_1} \leqslant q_{p_2} \leqslant q_{p_3}$. Substituting the quantile expression (3.1) in equation (4.1), and multiplying by the Jacobian of the transformation $(q_{p_1}, q_{p_2}, q_{p_3}) \rightarrow \boldsymbol{\theta} = (\mu, \sigma, \xi)$, leads directly to an expression for the prior in terms of the GEV parameters. Multiplication by the likelihood (2.2) then gives the posterior, $\pi(\boldsymbol{\theta}|\mathbf{x})$ say, up to a constant of proportionality.

Explicit analytical calculation of the marginal distributions of π, for example, is completely intractable. However, recent recognition of the powerfulness and simplicity of Markov chain Monte Carlo techniques suggest that direct simulation from a Markov chain whose equilibrium distribution is π is straightforward. Thus, our analysis is based on a Gibbs sampler, successively updating the individual parameters μ, σ and ξ, from π, conditionally on the current values of the other parameters. Because these conditional distributions are not of standard form, a univariate Metropolis step has been included within each iteration of the Gibbs sampler. This amounts to simulating from some arbitrary density, and applying a rejection criterion to the simulated value. With an appropriate rejection criterion, this guarantees a chain with the correct marginal equilibrium distribution. Details of such techniques are now extensively published (Smith and Roberts (1993), for example) and so are omitted here. The procedure is optimized when the probability of rejection is low; we have adopted a pragmatic procedure of basing the simulation around a multivariate normal approximation to the joint posterior, with parameters determined initially by the maximum likelihood estimate, and then updated periodically with empirical estimates based on the history of the chain. This specific approach is also well reported (Tierney (1994), for example). To improve the efficiency of the procedure further, we ran the sampler in the (μ, σ, ξ) GEV parameterization rather than in the quantile parameterization, in the anticipation that the GEV parameterization is likely to have better orthogonality properties than that based on quantiles (Smith, 1987a).

5. Results

5.1. *Estimation*

Determination of an appropriate threshold u usually proceeds by simple exploratory analysis. In particular, the asymptotic assumptions underpinning model (2.1) are such that

$$E(X - x|X > x) = \frac{\sigma + \xi(x - \mu)}{1 - \xi}, \tag{5.1}$$

which is linear in x, for $x > u$ (Davison and Smith, 1990). Thus, inspection of a plot of empirical mean residual life, to determine a threshold above which linearity is observed, is commonly employed. The mean residual life plot for these data is given in Fig. 2. The interpretation is complicated by the increased variability at large values of x, but with this in mind there appears to be evidence of linearity above a threshold of around 40 mm, which gives rise to 86 exceedances. Thus, initially we take $u =$ 40 mm, but we consider the effect of lowering this value later.

The Gibbs sampler was initialized at an arbitrary starting point and run as described in Section 3 for 10000 iterations. After an initial burn-in, the sampler appeared to have reached equilibrium by around 2000 iterations, so subsequent

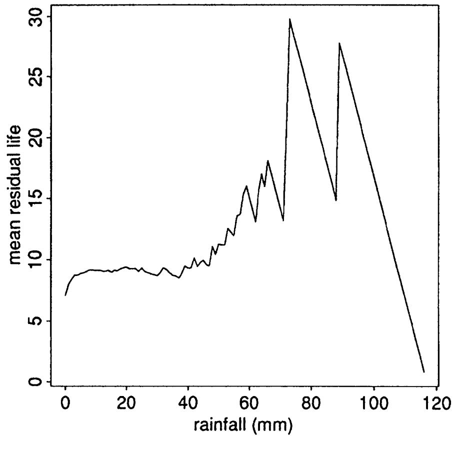

Fig. 2. Mean residual life plot of the rainfall data

analysis is based on iterations 2001–10000. Convergence of the sampler was verified by repeating the sampling for a range of starting values. A summary of the posterior marginal estimates is given in Table 3, which also gives the corresponding maximum likelihood estimates, and prior estimates also obtained from a Gibbs sample. A comparison of the prior and posterior univariate marginal distributions for each of the GEV parameters is also shown in Fig. 3.

Marginally, the location and scale parameter priors are almost non-informative: the prior for μ is extremely flat, whereas that for σ resembles $1/\sigma$. The marginal prior for ξ carries most information since $\xi > 0$ ensures a distribution with an infinite upper end point. There is a physical reasoning why the prior information for the location parameter should be highly diffuse relative to that of the other parameters: the location parameter tends to be highly dependent on localized site-specific characteristics (Coles and Tawn, 1996) which are difficult to calibrate without reference to data. In contrast the scale, and especially the shape, parameters are governed by regional characteristics of the rainfall process, about which prior information is more easily assimilated. This reverses the situation from a likelihood-based analysis in which the relative precision of estimation of the location parameter is much greater than that of the scale or shape parameters (Coles and Tawn, 1990, 1996; Tawn, 1993). The standard errors in Table 3 confirm this.

Comparisons between prior and posterior estimates and between the Bayesian and classical inferences are each of interest. The posterior distributions are entirely consistent with the prior estimates, though substantially more precise. An examination of density estimates of the prior and posterior bivariate marginals, also

TABLE 3
Comparison of prior, posterior and maximum likelihood estimates of GEV parameters at each of two thresholds u†

Parameter	*Prior*	*Results for u = 40*		*Results for u = 30*	
		Maximum likelihood estimate	*Posterior*	*Maximum likelihood estimate*	*Posterior*
μ	41.2	43.8 (1.1)	43.2 (41.6, 45.2)	44.1 (1.1)	44.2 (42.3, 46.4)
σ	4.9	8.5 (1.1)	7.9 (6.1, 10.0)	9.1 (0.73)	10.5 (9.3, 11.7)
ξ	0.47	0.15 (0.12)	0.31 (0.24, 0.38)	0.08 (0.06)	0.25 (0.20, 0.29)

†Figures given for the prior and posterior are simulated means, together with a 95% central credibility interval for the posterior; standard errors of maximum likelihood estimates are given in parentheses.

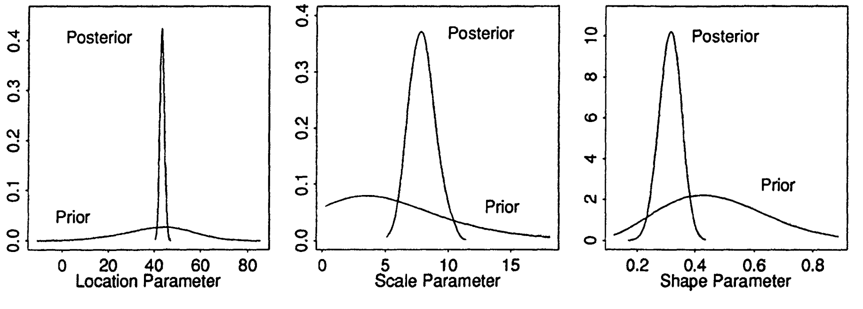

Fig. 3. Univariate marginals of prior and posterior distributions of each GEV parameter

obtained from the Gibbs sampler but not shown here, is equally informative. In addition to the reduction in marginal dispersion, the prior dependence between each pair of parameters is diminished through the posterior analysis, except for the scale–shape parameter pair, for which the prior dependence is largely maintained in the posterior. A comparison between the Bayesian and classical inferences shows that the posterior distributions are in accord with a likelihood-based analysis, in that the posterior means are consistent with the maximum likelihood estimates relative to their asymptotic standard errors. Thus, in terms of the GEV parameters, the posterior analysis appears to have attained a plausible compromise between the prior information and the data. In particular, the marginal posterior means are nested by the prior means and maximum likelihood estimates.

A more informative comparison is based on quantiles, since these are the parameters of primary concern. In this context, it is usual to plot quantiles of the distribution of the annual maximum, as given by equation (3.1), against return period on a logarithmic scale. This choice of scale has the effect of emphasizing the tail region, and giving linear plots for the $\xi = 0$ case. Maximum likelihood estimates for each q_p are obtained by substituting the maximum likelihood estimates for $\boldsymbol{\theta}$ into equation (3.1). In the Bayesian context, a realization from the posterior distribution of any specified q_p is obtained directly by substituting the simulated sample of $\boldsymbol{\theta}$ into equation (3.1). Thus, in Fig. 4, we compare the maximum likelihood estimates of q_p with the prior and posterior means for q_p against return period, on a logarithmic scale. Also shown are estimates based on the empirical distribution function of the annual maximum. Both the likelihood-based analysis and the posterior analysis are reasonably similar and consistent with the data which have exceeded $u = 40$ mm. However, the likelihood analysis graph has less curvature, leading to lower long-range extrapolations, i.e. the estimated tail is lighter. Perhaps surprisingly, it is the curvature of the posterior analysis graph which seems to be in closest agreement with the most extreme of the empirical information. This suggests that the shape of the maximum likelihood estimate curve may be overinfluenced by less extreme data, leading to underestimation in the extrapolation. By contrast, the posterior analysis incorporates perceived knowledge about extreme conditions which, for our data at

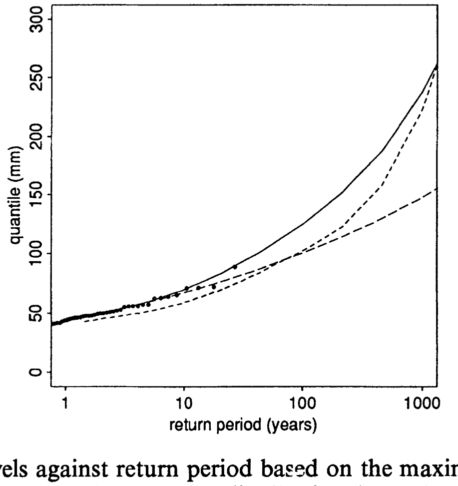

Fig. 4. Plot of return levels against return period based on the maximum likelihood estimate (– – –), the prior distribution (- - - -) and the posterior distribution (———): •, empirical estimates

least, has led to an estimated curve which better reflects the empirical behaviour at the most extreme levels.

It may have been expected that the posterior curve would be nested by the maximum likelihood and prior curves, by analogy with the GEV parameters themselves. That this is not so is a consequence of the substantial skew in the profile likelihood surfaces, particularly for the most extreme return levels (see Davison and Smith (1990) for examples). Furthermore, the flatness of these profile likelihood surfaces explains the tendency for the posterior mean to be closer to the prior mean than the maximum likelihood estimate at long extrapolations.

To assess the effect that the prior assessment has had on the posterior distribution, we repeated the analysis on the basis of an almost non-informative prior. Specifically, independent Gaussian (or log-Gaussian) priors with large variance were used for each of the GEV parameters. Central credibility intervals based on the uninformative prior were found to be almost identical with confidence intervals based on asymptotic theory for the profile likelihood. Fig. 5 compares, as a function of p, the posterior distributions of q_p using the two different prior forms. Within the range of the data the two posteriors for q_p are very similar. However, for longer return periods the effect of the informative prior is to produce posteriors with higher mean and substantially lower variance than in the non-informative case. Both these effects are consistent with the qualitative reasoning that the expert supplied us with in support of his prior choice. Thus, a major benefit of the Bayesian analysis is the improvement in precision of estimates at levels beyond the range of the data that is achieved by the inclusion of informative prior knowledge.

5.2. *Prediction*

The primary objective of an extreme value analysis is often prediction, i.e. given the history of a process $\mathbf{x}$ (and other external information) the main interest is in

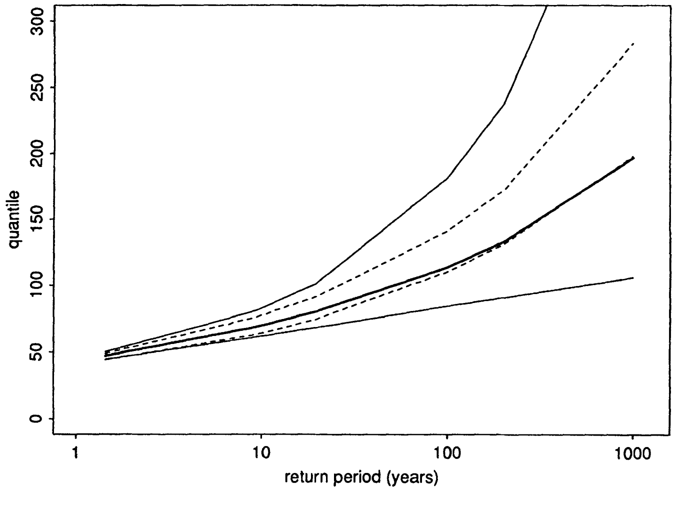

Fig. 5. Comparison of posterior distributions of quantiles as a function of return period based on informative and non-informative priors: - - - -, 95% credibility interval based on an informative prior; ——, corresponding credibility limits based on a non-informative prior; ——, posterior means using the non-informative prior

estimating the extremal characteristics over some future epoch of the process. In the context of the rainfall example, define Z_L to be the maximum daily rainfall over a future period of L years. Then, allowing for the uncertainty in the estimation of the parameter components, the *predictive distribution* of Z_L (Aitchison and Dunsmore, 1975) is defined as

$$\Pr(Z_L < z|\mathbf{x}) = \int_\Theta \Pr(Z < z|\boldsymbol{\theta})^L \, \pi(\boldsymbol{\theta}|\mathbf{x}) \, \mathrm{d}\boldsymbol{\theta}, \tag{5.2}$$

where Z denotes the annual maximum. Consequently, designing to a level z_p such that $\Pr(Z_L < z_p|\mathbf{x}) = 1 - p$ will give a design level which, after allowing for parameter uncertainty, will be exceeded with probability p in an L-year period.

Davison (1986) gave a likelihood-based approximation scheme for the predictive distribution of Z where the distribution of Z is the GEV, but from the Gibbs sampler output distribution (5.2) is estimated directly by

$$\widehat{\Pr}(Z_L < z|\mathbf{x}) = m^{-1} \sum_{j=1}^{m} \Pr(Z < z|\boldsymbol{\theta}_j)^L, \tag{5.3}$$

where $\boldsymbol{\theta}_j$ is the output from the jth iteration of a sample of size m taken from the Gibbs sampler of the posterior distribution of $\boldsymbol{\theta}$. Again, since it is extreme levels of Z_L that are of most interest, it is natural to plot the predictive distribution of Z_L on a $-\log\{-\log(1-p)\}$-scale. For $L = 1$, this is given in Fig. 6, which may be interpreted as a graph of return levels in which uncertainty due to both process variability and sampling variability has been accounted for. The corresponding estimative return level plot, based on the maximum likelihood estimates, is also included in Fig. 6. Clearly, designing to a level determined by the maximum likelihood estimates may lead to substantial underprotection due to uncertainty in the parameter estimates, which is accounted for in the predictive analysis. In practice, even when using

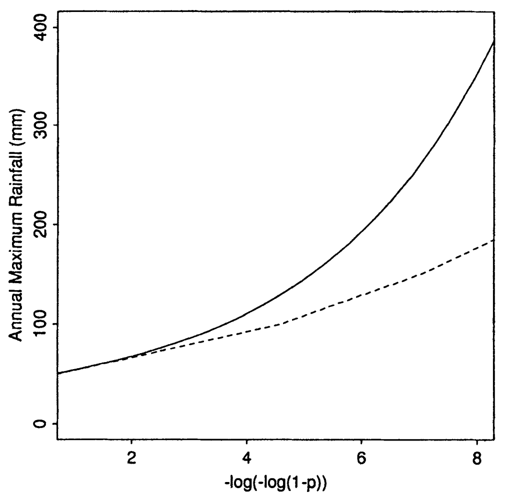

Fig. 6. Plot of z_p against $-\log\{-\log(1-p)\}$ corresponding to annual maximum rainfall based on predictive (———) and estimative (- - - -) distributions

maximum likelihood estimates, some account is taken of parameter uncertainty by perhaps designing to upper end points of confidence intervals. This issue is examined more closely in Table 4. For comparison, we have considered a variety of criteria by which a design may be determined. Under each criterion, a design level is sought that will be exceeded in a given year with probability 0.001. Our comparison includes design levels based on the predictive distribution, the maximum likelihood estimate, a 95% profile likelihood upper end point, the mid-point of a 95% profile likelihood interval and the posterior mean. The levels obtained, together with the associated exceedance probabilities obtained from the predictive distribution, are given in Table 4. As observed from Fig. 6, designing to the maximum likelihood estimate would result in considerable underprotection. It is clear from Table 4, however, that attempting to allow for uncertainty by designing to the upper end point of a profile likelihood interval results in substantial overprotection. These findings have serious implications, since most designs are made on the basis of either of these criteria. In contrast, designing to either the posterior mean or the profile likelihood mid-point gives results which are in close agreement with the predictive distribution in this case. Our view is that the interpretability of the predictive analysis, together with the systematic way that it accounts for all sources of uncertainty while exploiting all sources of information, makes it preferable as the basis for design against extreme events.

5.3. *Choice of Threshold*

One final issue is the question of the choice of threshold. The choice of $u = 40$ mm was made on the basis of Fig. 2, but this gives rise to only 86 exceedances over the whole period. Accordingly, for comparison, we repeat the analysis with $u = 30$ mm. Under model (2.1) the slope of the mean residual life plot at sufficiently high thresholds is $\xi/(1-\xi)$ (see equation (5.1)). Thus, from Fig. 2, a threshold of $u = 30$ mm is likely to result in a likelihood model which is inconsistent with the most extreme data and in particular has a value of ξ which is much closer to 0 than was found with a threshold $u = 40$ mm, and which is in conflict with the prior specification (see Fig. 3). A summary of the likelihood and posterior analysis with this threshold is included in Table 3. The lower threshold gives rise to 284 exceedances, so it might be expected that, as more data have been included, the analysis would result in a posterior distribution which is in closer agreement with a

TABLE 4
Design heights z chosen so that $Pr(Z>z)=0.001$, where Z is the annual maximum rainfall†

Criterion	*Design height (mm)*	*Encounter risk*
Predictive distribution	248	0.001
Maximum likelihood estimate	145	0.0069
Profile likelihood end point	430	0.00018
Profile likelihood mid-point	263	0.00087
Posterior mean	236	0.0013

†Estimates are based on different estimation criteria, and associated with each is the exceedance probability of z based on the predictive distribution.

likelihood analysis than previously. Table 3 suggests that this is not so, as the information that these data supply in terms of model (2.1) is in conflict with the prior information. This prior–data conflict has resulted in a posterior which is a compromise of the two disparate sources of information but is in disaccord with both. This confirms that for these data a threshold has been chosen which is too low for the asymptotic model (2.1) to be valid. However, we subsequently found that the difference in the predictive distribution compared with the previous analysis at $u = 40$ mm is negligible except at the most extreme levels, suggesting that this aspect of a Bayesian analysis may be more robust to the choice of threshold than is the equivalent likelihood-based analysis.

6. Discussion

Our objective was to elicit prior information for extreme rainfall in such a way that when combined with data through a Bayesian analysis the posterior analysis obtained would provide a rational basis for extrapolation. So far as is possible to tell, this has been achieved. The inclusion of the prior information gives a tail estimate which is in closer accord with the most extreme of the observed data and, despite having a longer tail than the maximum likelihood estimate, is more precise in terms of credibility intervals. The ability to calculate the predictive distribution has been shown to summarize the extremal information within a predictive framework, which represents the most appropriate criterion for design purposes, and leads to significantly different design recommendations than would be given under current practices. Also, considerably more complex prior information than we have used, perhaps relating to the physical constraints on a particular process, is easily included within the Gibbs sampler mechanism. With these points in mind, our view is that, if reliable prior information concerning the extremes of a process is available, then the arguments in favour of a Bayesian analysis are compelling. However, the structure of the asymptotic models for extremes is such that considerable care is required in determining appropriate forms for the prior information.

One feature that makes the analysis of extremes particularly unusual from a Bayesian perspective is that of data–prior dominance. In most modelling situations, once more data become available they add weight to the likelihood, which eventually dominates the prior. In an extremal analysis, however, the primary concern is almost invariably about behaviour beyond the range of observed data. So, if the prior is specified for the extreme tail and we focus attention on the posterior distribution of a quantile which remains above the maximum observed datum as the sample size increases to infinity, then the data necessarily fail to dominate the prior. As a simple illustrative example, let $X_1, \ldots, X_n$ be independent and identically distributed random variables, given ϕ, from $U[0, \phi]$, and denote the prior for ϕ by $\pi(\phi)$. The parameter ϕ is an end point, so prior knowledge about the parameter is always informative relative to the data as $\pi(\phi|\mathbf{x}) \propto \phi^{-n}\,\pi(\phi)$ for $\max(X_1, \ldots, X_n) < \phi$. Moreover, for an extremal analysis, once more data become available, because the likelihood is itself an asymptotic approximation, the optimal procedure is not necessarily to include more data in the likelihood, but possibly to raise the threshold u for the point process model (2.1), i.e. there is a formal bias–variance trade-off in the choice of threshold (Smith, 1988), which may be resolved by raising the threshold at a rate that prevents prior dominance by the data. The use of higher order penultimate models (Smith, 1987b) in preference to model (2.1) may give a deeper insight into this issue.

Acknowledgements

We are indebted to Duncan Reed for his work in specifying the prior information for the rainfall data and for more general background advice. We are also grateful to Martin Crowder and Richard Smith for useful discussions, and to the Institute of Hydrology for the data. Both authors are supported by an Engineering and Physical Sciences Research Council complex stochastic systems award.

References

Aitchison, J. and Dunsmore, I. R. (1975) *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.

Coles, S. G. (1994) A temporal study of extreme rainfall. In *Statistics for the Environment 2: Water Related Issues* (eds V. Barnett and K. F. Turkman), pp. 61–78. Chichester: Wiley.

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

Coles, S. G. and Tawn, J. A. (1990) Statistics of coastal flood prevention. *Phil. Trans. R. Soc. Lond.* A, **332**, 457–476.

——(1991) Modelling extreme multivariate events. *J. R. Statist. Soc.* B, **53**, 377–392.

——(1996) Modelling extremes of the areal rainfall process. *J. R. Statist. Soc.* B, **58**, 329–347.

Crowder, M. (1992) Bayesian priors based on a parameter transformation using the distribution function. *Ann. Inst. Statist. Math.*, **44**, 405–416.

Dales, M. Y. and Reed, D. W. (1989) Regional flood and storm hazard assessment. *Report 102*. Institute of Hydrology, Wallingford.

Davison, A. C. (1986) Approximate predictive likelihood. *Biometrika*, **73**, 323–332.

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

Garthwaite, P. H. and Dickey, J. M. (1985) Double- and single-bisection methods for subjective probability assessment in a location-scale family. *J. Econometr.*, **29**, 149–163.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980) Interactive elicitation of opinion for a normal linear model. *J. Am. Statist. Ass.*, **75**, 845–854.

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

——(1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.

——(1994) Bayes quantile estimation and threshold selection for the Generalized Pareto family. In *Extreme Value Theory and Applications* (eds J. Galambos, J. Lechner and E. Simiu), pp. 123–138. Dordrecht: Kluwer.

Reed, D. W. and Stewart, E. J. (1989) Focus on rainfall growth estimation. In *Proc. 2nd Natn. Hydrology Symp.*, pp. 3.57–3.67. Wallingford: British Hydrological Society.

——(1994) Inter-site and inter-duration dependence in rainfall extremes. In *Statistics for the Environment 2: Water Related Issues* (eds V. Barnett and K. F. Turkman), pp. 125–143. Chichester: Wiley.

Smith, A. F. M. and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 3–23.

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

——(1987a) Discussion on Parameter orthogonality and approximate conditional inference (by D. R. Cox and N. Reid). *J. R. Statist. Soc.* B, **49**, 21–22.

——(1987b) Approximations in extreme value theory. *Technical Report 205*. University of North Carolina, Chapel Hill.

——(1988) Bias and variance approximations for estimators of extreme quantiles.
——(1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sci.*, **4**, 367–393.
Smith, R. L. and Naylor, J. C. (1987) A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Appl. Statist.*, **36**, 358–369.
Tawn, J. A. (1993) Extreme sea-levels. In *Statistics for the Environment 2: Water Related Issues* (eds V. Barnett and K. F. Turkman), pp. 243–263. Chichester: Wiley.
Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, **22**, 1701–1762.