International Statistical Review (1996), **64**, 1, 119–136, Printed in Mexico

Bayesian Methods in Extreme Value Modelling: A Review and New Developments

Stuart G. Coles[1] and Elwyn A. Powell[2]

[1]*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK*
[2]*Department of Mathematics, University of Nottingham, Nottingham NG7 2RD, UK*

Summary

Extreme value problems are characterized by a scarcity of data and the requirement of modelling where the data are most sparse. This presents a dilemma when considering a Bayesian approach to inference: the value of additional prior information is likely to be substantial, but the plausibility of formulating such prior knowledge for extremal behaviour is questionable. In this paper we review the literature linking the themes of Bayesian and extreme value analysis, and use recent advances in Bayesian computational tools to assess the utility of a Bayesian extreme value analysis in three different situations: one where an expert is available to supply prior information; the second where maximum likelihood fails; and the third where spatial information on related variables is used to formulate an empirical prior.

Key words: Bayesian inference; Empirical Bayes; Extreme value models; Gibbs sampler; Markov Chain Monte Carlo; Prior elicitation; Spatial modelling.

1 Introduction

Extreme value modelling and Bayesian inference have one thing in common: both are anathema to large numbers of statisticians. Objections are made to extreme value modelling on the grounds that extrapolation of statistical models to domains on the fringe of, or beyond, observed data is scientifically unreasonable. Antagonists of the Bayesian philosophy argue that it is similarly unreasonable to use statistical procedures which are intrinsically subjective. Without dwelling on issues of philosophy or semantics, but taking pragmatism as motivation, this paper seeks to examine the utility of a Bayesian approach to the inference of extreme value models.

The literature on this subject, reviewed in Section 3, is sparse, though there are a large number of related papers in the field of reliability. Most of what has been written explicitly for extreme value modelling has focused on the use of Bayesian techniques as a technical inferential device by specifying uninformative priors, rather than attempting to exploit the Bayesian mechanism for incorporating genuine prior information into the analysis. Why should this be so? Computation is undoubtedly part of the problem—only relatively recent developments in numerical quadrature and, more latterly, Markov Chain Monte Carlo have enabled some of the difficulties in Bayesian computations to be overcome. But more fundamentally it is natural to be sceptical about the reasonableness of expecting even the most knowledgeable scientist to be able to supply meaningful prior information about the extremal nature of a particular process. To put it bluntly, if even the data on extremes are so scarce, what hope is there of an expert being able independently to formulate his or her prior beliefs?

The alternative view, however, is equally compelling: the very scarcity of extreme data implies that the inferential value of even the slightest piece of additional information can be enormous. Thus if an expert can provide meaningful prior information it is likely to substantially improve the quality of the inference.

One theme of this paper then is to assess the viability of using an expert to supply prior information as an aid to inference in extreme value problems. This requires consideration of how to elicit prior information in this context; the impact of prior choice on posterior inference; and the consequences of prior mis-specification. These issues are addressed in Section 4 where the case study on extreme rainfall carried out by Coles & Tawn (1996a) is reviewed and summarized.

Section 5 considers the use of Bayesian inference for extreme value models in a situation where maximum likelihood fails. This occurs when the parent population from which data are drawn has a very short tail, leading to models whose likelihood is unbounded. This adds to the computational burden of a Bayesian analysis, but does not in principle prevent Bayes estimators from being obtained. Again, we will use a particular data set, in this case a series of survival times, to illustrate the issues.

In Section 6 we consider a novel approach to the estimation of extreme wind speeds in the presence of spatial information. Formal spatial modelling of such processes over large geographical scales is often impossible due to the disturbance effects of localized topography and the incoherence of meteorological systems at large distances. On the other hand, without any physical evidence to the contrary, it should be anticipated that wind speed behaviour at any given location is not likely to be atypical relative to that of other sites in the region. Our approach therefore is to use available spatial information to construct prior distributions which can then be applied at any arbitrary site. We perceive that the benefit of such an analysis is likely to be greatest when trying to model extremal behaviour at a site where very few data are available. Maximum likelihood estimators are unacceptably variable in this case, whereas the inclusion of the spatial information in the form of a prior distribution is shown to stabilize the estimates without incurring bias. In this way the spatial information acts as an empirical surrogate for the information which could be provided by an expert, or by having a more extensive data record. We illustrate this approach using wind speed records from 106 sites in the United States as the basis for formulating prior information to assist the analysis at a further 9 sites.

In each of Sections 4 to 6 the basic computational tool for the Bayesian analyses is the Gibbs sampler. Details are not given here as the application is completely standard and the general methodology of Markov Chain Monte Carlo is well referenced in the literature (Smith & Roberts, 1993, or Gilks *et al.*, 1995, and references therein, for example). We begin in Section 2 by giving a brief overview of the models and asymptotic motivation which form the core of extreme value methodology.

2 Extreme Value Models

Extreme value modelling is widely applied, particularly in engineering design situations, where accurate assessment of the behaviour of a process at high levels is required. The simplest model construction for extremal behaviour is where we have a series $X_1, X_2, \ldots, X_n$ of independent random variables with common distribution function F and seek an estimate of the upper tail behaviour of F. Lack of data in the tail region prevents the use of standard density estimation techniques. For this reason, a widely accepted rationale for developing tail models for F is the use of asymptotic argument; the examples given in subsequent sections are based on the two most commonly used models. See Smith (1991) for a review of the general area of extreme value modelling.

The first model is formulated by considering the asymptotic behaviour of the series $M_n = \max\{X_1, X_2, \ldots, X_n\}$. With normalizing sequences $\{a_n\}$ and $\{b_n\}$, the complete class of non-degenerate limiting distributions of $(M_n - a_n)/b_n$ as $n \to \infty$ is found to be the Generalized Extreme Value (GEV) family. The distribution function for this family takes the form

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} \tag{2.1}$$

The parameters μ and σ are location and scale parameters, while ξ is a shape parameter which determines the weight of the tail of G, and consequently F. The special case of $\xi = 0$, interpreted in (2.1) as $\xi \to 0$, gives the so-called Gumbel model with distribution function

$$G(z) = \exp\left[-\exp\left\{-\frac{z-\mu}{\sigma}\right\}\right]. \tag{2.2}$$

In some areas of the literature (2.2) is referred to as *the* extreme value distribution, with the implication that this is the only plausible limit distribution for maxima. In this spirit much of the literature on the Bayesian modelling of extremes has focused on this family (c.f. Section 3). More generally, it is usual to adopt (2.1) as an approximation to the distribution of $(M_N - a_N)/b_N$ for some large value of N—typically the number of observations in a year, so that M_N is the annual maximum—which implies that M_N itself has a GEV distribution.

A limitation with (2.1) as an inferential model is its restriction to annual maximum data when in fact the series $X_1, X_2, \ldots, X_n$ may contain many other data which are informative about the tail of F. Consequently, several other models of a similar genesis have been proposed to incorporate a greater amount of the data into the inference. Conceptually easiest is the threshold approach in which all observations in the series $X_1, X_2, \ldots, X_n$ which exceed a high threshold u are parametrically modelled. This serves as an approximation to the tail of F itself. Pickands (1971) showed that, in an asymptotic sense, consistency with the basis of model (2.1) is achieved by specifying the conditional distribution of threshold exceedances to follow the Generalized Pareto Distribution (GPD). Thus the distribution function of $Y_i = X_i - u | X_i > u$ is taken as

$$H(y) = 1 - (1 + \xi y/\tilde{\sigma})_+^{-1/\xi}, \tag{2.3}$$

where $\tilde{\sigma} = \sigma + \xi(\tilde{u} - \mu)$, and μ, σ and ξ are the GEV parameters of (2.1). Taking into account also the crossing rate of the threshold u gives a three-parameter model equivalent to the three-parameter GEV model (2.1). Inference based on (2.3) is generally superior since it applies to more data. For both (2.1) and (2.3) inference typically proceeds by numerical maximization of the corresponding likelihood function. Smith (1985) has shown that maximum likelihood is regular in the case $\xi > -0.5$, and that when $-1 < \xi < -0.5$, maximum likelihood estimators exist but are non-regular. If $\xi < -1$, corresponding to very short tailed distributions, the likelihood is unbounded and maximum likelihood fails. Fortunately, for environmental data, this situation rarely occurs.

There is, however, a parallel of all the preceding models in which the case of $\xi < -1$ arises often. If interest is in the lower tail rather than the upper tail then we may apply similar arguments as above to the series $m_n = \min\{X_1, X_2, \ldots, X_n\}$. This leads to the asymptotic distribution of minima having the form

$$G(z) = 1 - \exp\left\{-\left[1 - \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{2.4}$$

This is a direct analogue of (2.1) for minima, and there are corresponding analogues of (2.2) and (2.3). This model, which includes the traditional Weibull distribution (in the case $\xi < 0$) is often applied to model failure time data, for which it often happens that there are many failures close to the start of the experiment. This leads to an extremely short lower tail with model (2.4) having $\xi < -1$, in which case maximum likelihood fails. The example in Section 5 is of precisely this nature.

3 Literature Review

There are only very few papers linking the themes of extreme value modelling and Bayesian inference, most of which are restrictive either in the extreme value family adopted, or in the informativeness of the prior specification. Within the reliability literature there are considerably more cross-references, presumably because there are fewer conceptual problems in formulating prior beliefs in this context compared with extreme value modelling.

In this section we give a brief review of the literature, focussing in particular on aspects which are equally relevant to the principal theme of this paper: the assessment of Bayesian procedures as a means to incorporate genuine scientific belief into data analytic extreme value problems.

3.1 Reliability Models

The application of Bayesian inference within the field of reliability has a relatively long history; early references include Basu (1964), Holla (1966) and Bhattacharya (1967). We review here a number of the most relevant works, all of which take the Weibull distribution as their basis. The usual parameterization of the distribution function of the Weibull distribution is

$$F(x) = 1 - \exp\left\{-\left(\frac{x-c}{a}\right)_+^b\right\} \tag{3.1}$$

This is a re-parameterized sub-family of the GEV model for minima (equation 2.4) in the case $\xi < 0$.

Sinha & Sloan (1988) work with the full 3-parameter Weibull distribution. Even with non-informative priors, such a specification leads to intractable posterior calculations. Sinha and Sloan propose the use of Bayes Linear Estimates to approximate the posterior expectations in this case, and formulate the corresponding calculations, both for the Weibull parameters and for the survivor (or reliability) function. In the context of this paper, the limitation to uninformative priors is overly restrictive, though it may be that more general prior specifications for the Weibull parameters are also amenable to treatment using Bayes Linear Estimation. Even so, the method yields only the first and second posterior moments rather than the full posterior distribution.

In greater generality, Smith & Naylor (1987) work with the 3-parameter Weibull distribution, comparing maximum likelihood with Bayesian estimators in the case of informative priors, using specially adapted versions of numerical quadrature to perform the posterior calculations. Though the priors they work with are arbitrary, they are chosen to reflect a range of potential scientific hypotheses. Thus their paper is apparently the first to illustrate the computational feasibility within extreme value models of working with prior distributions which are scientifically motivated but analytically intractable. Furthermore, they report that the Bayesian inferential framework as a whole proved more satisfactory for their data analysis than the corresponding likelihood-based analysis.

The issue of prior elicitation is pursued by Singpurwalla & Song (1988), who restrict attention to the 2-parameter Weibull model ($c = 0$ in equation 3.1), but formulate priors to reflect genuine scientific belief. Their prior is hierarchical: moments are specified by an expert, which are then incorporated into an informative prior specification by an analyst, who seeks to correct for anticipated bias and over-confidence in the expert's assessment. Laplace approximations then lead to Bayes' estimators of the Weibull parameters.

In similar spirit, a generalization of the 2-parameter Weibull model is considered by Berger & Sun (1993). The generalization is essentially a mixture of 2-parameter Weibull distributions, and again the focus of the paper is on prior elicitation and specification. The extra complexity of this model leads to an increase in computational burden, which the authors address by the use of the Gibbs sampler, which is also the algorithm adopted in subsequent sections of this paper. An earlier paper, Day & Lee (1992), also suggests the use of the Gibbs sampler for carrying out the calculations required in the Bayesian analysis of reliability data, though the basic model they work with is the 2-parameter exponential distribution instead of the Weibull distribution.

Naturally, the mathematical equivalence between the extreme value models used in a reliability context, and those used in a more explicit extreme value context, means that theoretical issues are identical in the two fields. However, the difference in application means that a number of practical issues, especially concerning the prior formulation and inference interpretation, are somewhat different. In the remainder of this section we look at papers which are explicit in their emphasis on extreme value modelling.

3.2 *Prior Choice*

Clearly the GEV model (2.1) admits no conjugate prior distribution. Two approaches based on obtaining conjugate families for sub-classes of models have been proposed. Engelund & Rackwitz (1991) obtain conjugate priors in the case of single parameter subsets of the GEV distribution: a location parameter in the case of $\xi = 0$ and a scale parameter otherwise. For practical application such models are obviously quite restrictive, especially as it is the shape parameter ξ about which there is usually greatest uncertainty. With greater generality, Ashour & El-Adl (1980) consider the two parameter Gumbel family (for minima) specified by the limit in equation (2.4) as $\xi \to 0$, in the presence of left-censored data. In this case the joint conjugate prior takes the form

$$f(\mu, \sigma) \propto \sigma^{-H} \exp\left[-\frac{G}{\sigma} + \frac{H\mu}{\sigma} - \frac{D\exp(\mu - \delta)}{\sigma}\right], \tag{3.2}$$

for parameters D, G, H and δ. The cost of this extra generality is that the normalizing factor is intractable. In many respects this defeats the object of a conjugate analysis whose role usually is to avoid such numerical calculations. However, by using straightforward numerical techniques, Ashour & El-Adl (1980) compare Bayes estimators with the corresponding maximum likelihood estimators for a set of simulated data. They report that the Bayesian estimators are more efficient than the corresponding likelihood-based estimators, though this appears from Table 1 of their paper to be offset by the substantial bias of the Bayes estimators. This signals a warning against the use of Bayesian inference blindly for extreme value problems.

A non-conjugate specification of prior distributions within the GPD class (2.3) is proposed by Pickands (1994). He suggests the prior

$$f(\sigma, \xi) \propto \frac{1}{\tilde{\sigma}} 1_{\{\tilde{\sigma} > 0\}} \tag{3.3}$$

which is equivalent to specifying priors for $\log(\tilde{\sigma})$ and ξ which are independent and uniform on $(-\infty, \infty)$, though a limitation of this proposal is that the parameter $\tilde{\sigma}$ is dependent on the threshold u chosen in (2.2).

3.3 *Parameter Transformation*

The issue of model parameterization is considered by Rodrigues & Louzada Neto (1990). They again restrict attention to the Gumbel class, and consider estimation of the scale parameter σ with μ regarded as a nuisance parameter. Using Laplace integral approximations to the posterior with a non-informative Jeffrey's prior, they compare the marginal distribution for σ both with and without an orthogonalising transformation for the pair (μ, σ). Inevitably, they find very little difference in the marginal posterior inference for σ.

Box–Cox transformations within the Gumbel family are considered by Achcar *et al.* (1987). They model a sequence of survival times $T_1, T_2, \ldots, T_n$ using the model

$$Y \sim G_m(\mu, \sigma) \tag{3.4}$$

where

$$Y = \begin{cases} \frac{T^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(T), & \lambda = 0 \end{cases} \tag{3.5}$$

and $G_m(\mu, \sigma)$ is the Gumbel model for minima described in Section 2.

They use a Jeffreys prior for μ and σ, an improper uniform prior for the Box–Cox parameter λ, and show how Laplace integral approximations lead to the posterior marginal distribution for λ. On the basis of this posterior inference, taking the mode of the posterior distribution, they assess the quality of the transformation by simple probability plot techniques. We will consider a slightly different model for their data in Section 5.

3.4 Prediction

In most practical applications, the role of an extreme value analysis is to characterize the extremal behaviour of the past history of a process with a view to designing against extreme excursions of future values of the process. That is, the principal inferential objective is predictive. Several authors have highlighted the importance of viewing extreme value problems from a predictive perspective, and that from such a perspective Bayesian procedures offer the most elegant solution. The predictive density function (Aitchison & Dunsmore, 1975) is defined as

$$f(y|x) = \int f(y|\theta)f(\theta|x)d\theta \tag{3.6}$$

where x represents historical data, y a future observation, $f(y|\theta)$ the likelihood and $f(\theta|x)$ the posterior distribution of θ given x. Thus, the predictive distribution averages the distribution across the uncertainty in θ as measured by the posterior distribution. For the GEV family, Davison (1987) gives a prior-free approximation to (3.6) based on the approximate predictive likelihood; asymptotically the approximation is exact. Engelund & Rackwitz (1992) calculate the exact forms of the predictive distribution for the one-parameter family of models specified in Section 3.1. In a slightly different context, Lingappaiah (1984) develops bounds for the predictive probabilities of extreme order statistics under a sequential sampling scheme, when sampling is carried out from either an exponential or Pareto population.

4 Prior Elicitation and Impact

From a practical viewpoint, the most important issues arising from the literature identified in Section 3 are the elicitation and formulation of genuine prior information in extreme value problems, and the consequent impact such a specification (or mis-specification) has on subsequent inferences. In a recent case study, Coles & Tawn (1996a) consider a case study in which expert knowledge is sought and formulated into prior information as the basis for a Bayesian analysis of extreme rainfall. Their analysis is based on 54 years of daily rainfall data measured at a single location in south-west England, using a likelihood which is equivalent to the GPD distribution with the GEV parametrization specified by equation (2.3), as a model for daily rainfall excesses of 40mm. Much of their paper is devoted to the issue of prior elicitation in the context of an extremal analysis, and they argue that prior formulation in terms of the GEV parameters is not a sensible approach. Instead, they elicit prior information in terms of extreme quantiles, arguing that this is a scale on which an expert is most likely to be able to accurately quantify their prior beliefs about extremal behaviour. They argue, moreover, that it is not unrealistic to hope that an expert can provide meaningful prior information, since the expert has extensive meteorological knowledge, and detailed topographical knowledge which can be brought to bear when anticipating the behaviour of extreme rainfall at a specific site.

The particular formulation adopted by Coles and Tawn is to elicit prior information in terms of the quantiles of the annual maximum rainfall distribution. With the GEV model (2.1), the $(1-p)$ quantile takes the form

$$q_p = \mu + \sigma\{[-\log(1-p)]^{-\xi} - 1\}/\xi. \tag{4.1}$$

Table 1

Elicited prior medians and 90% quantiles, each in mm, for distributions of $\tilde{q}_i$

p_i	median	90% quantile
0.1	59	72
0.01	43	70
0.001	100	120

Thus, prior specification of q_p for three values of p is equivalent to prior specification of (μ, σ, ξ), but corresponds to a scale which hydrologists are familiar with (q_p is the $1/p$-year return level.) Since there is a natural ordering of the q_p ($q_{p_i} > q_{p_j}$ whenever $p_i < p_j$) specification of independent priors on three different q_{p_i} would not be valid. Hence, independent Gamma priors are adopted for

$$\tilde{q}_1 = q_{p_1} \quad \tilde{q}_2 = q_{p_2} - q_{p_1} \quad \tilde{q}_3 = q_{p_3} - q_{p_2}, \tag{4.2}$$

and choose in particular $p_1 = 0.1,\ p_2 = 0.01,\ p_3 = 0.001$. The parameters of the Gamma priors are chosen to calibrate with the expert hydrologist's prior beliefs which were elicited in terms of quantiles of beliefs about the $\tilde{q}_i$. The actual calibration was made in terms of the prior median and 90% quantile for each of the q_{p_i}, and these are given in Table 1. Further discussion of the elicitation procedure and the justification for this particular prior specification by the expert hydrologist is given by Coles & Tawn (1996a).

The results of the analysis are best illustrated by the three figures included here as Figures 1–3. Figure 1 plots estimates of the annual maximum quantile q_p as a function of return period $(1/p)$. The estimates are based on prior means, posterior means and maximum likelihood respectively. Also shown, within the range of the observed data, are simple empirical estimates. Perhaps surprisingly, the Bayesian estimates appear the most consistent with the most extreme of the observed data, suggesting that the expert hydrologist's prior information has supplemented the data-based analysis with the information that observations substantially more extreme than those already observed are quite plausible. Within the data, all but the most extreme data point suggest this is less likely. However, it is equally clear that once estimation variability is accounted for, both the Bayesian and maximum likelihood estimates provide reasonable descriptions of the observed annual maxima, so the real gain in the Bayesian analysis in this respect is that extrapolations are based on the expert's perceived knowledge as well as the data. In a sense then, the Bayesian analysis admits extrapolation to be based on a compromise of asymptotic argument, and scientific belief formulated from an understanding of general meteorological and topographical issues. Objectors to the Bayesian analysis would be forced to resort to the extrapolation based on data only.

Figure 2 illustrates the value of the information provided by the expert. Again, annual maximum quantiles are plotted against return period, but in this case the Bayesian estimators are obtained using both the expert's prior and an uninformative prior respectively. As in Figure 1, the calibration effect of the expert information can be observed, but there is also a substantial improvement in precision (indicated by narrower credibility intervals) when the expert's prior is used.

Finally, Figure 3 compares the predictive annual maximum distribution (c.f. equation (3.6)) with the corresponding estimative distribution, which is simply the GEV distribution (2.1) with the unknown parameters (μ, σ, ξ) replaced by their maximum likelihood estimates. The importance of accounting for parameter uncertainty when designing on the basis of an extreme value model is clearly illustrated by Figure 3, especially when designing against events with extremely low probability. Designing simply to the maximum likelihood estimate is likely to give a quite false sense of security, compared with designing to the predictive distribution which takes account of uncertainty in the estimation process.

In summary, the case study of Coles & Tawn (1996a) demonstrates that elicitation of meaningful prior information for extreme value analyses is at least feasible, and that if such a prior specification

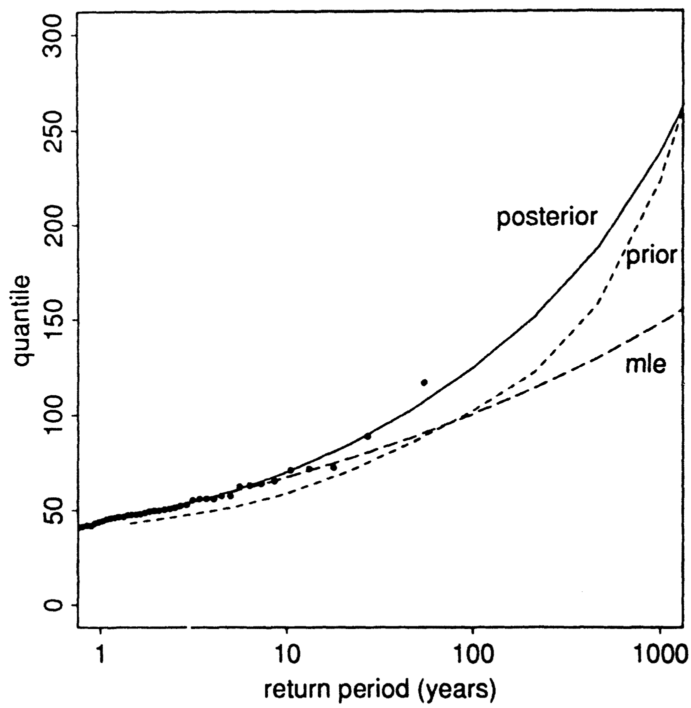

Figure 1. *Quantiles of annual maximum rainfall distribution as function of return period using prior and posterior means and maximum likelihood estimates. Also shown are empirical estimates.*

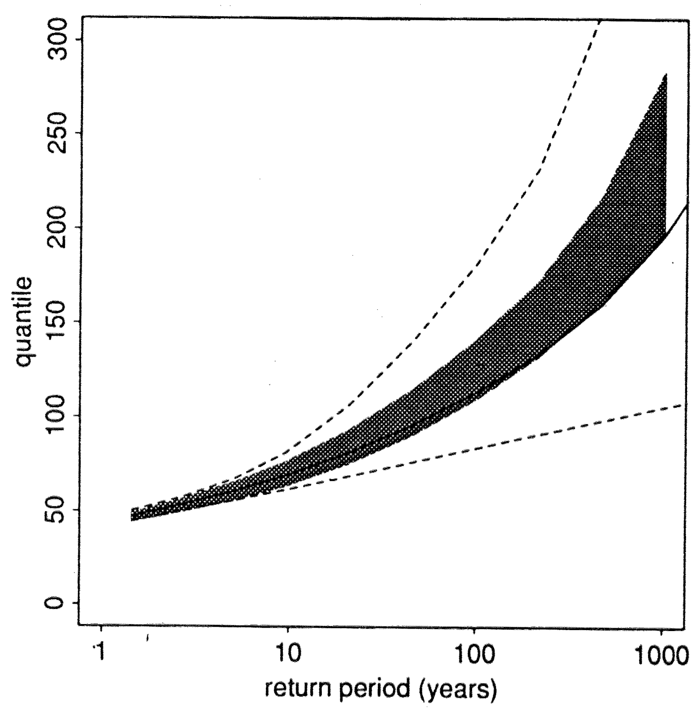

Figure 2. *Comparison of posterior distributions based on informative and non-informative priors respectively. Shaded region gives 95% credibility interval based on informative prior; outer dotted curves give corresponding probability limits based on non-informative prior. Solid line corresponds to posterior means using non-informative prior.*

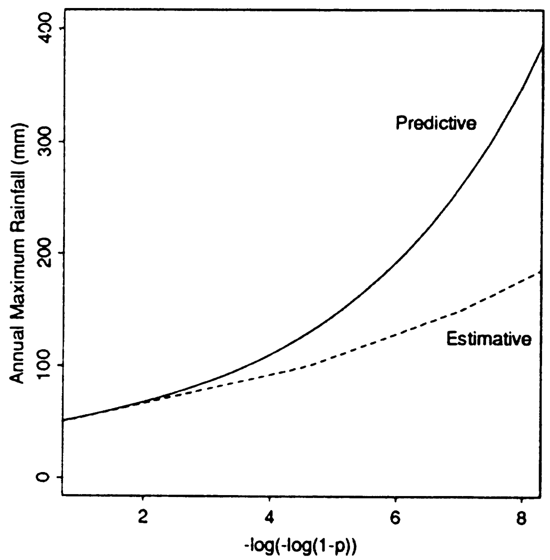

Figure 3. *Predictive and estimative distributions of annual maximum rainfal.*

can be trusted, then there are considerable benefits to be obtained by using a Bayesian analysis in preference to a likelihood-based analysis.

5 Bayesian Estimation for Short-Tailed Distributions

In this section we examine the use of Bayesian techniques in the situation of distributions with very short tails leading to the non-existence or non-regularity of maximum likelihood estimators. For illustration, we analyse the survival time data given by Achcar *et al.* (1987). These represent the survival times in weeks of a set of 21 patients following a particular treatment, and are recorded as:

$$1, 1, 4, 4, 9, 16, 25, 25, 64, 64, 64, 64, 121, 121, 144, 144, 225, 289, 484, 529.$$

Unfortunately, there is no discussion in Achcar *et al.* (1987) of the background to these data, which display an unusually large number of ties. This makes the example somewhat unsatisfactory from a broad statistical viewpoint. Nonetheless, the example represents a genuine data set to which extreme value type models have previously been fitted, so the analysis retains some comparative merit as well as illustrating the utility of the Bayesian approach in this situation. Consequently, we take the data at face value, but also account for the discretization of the data by assuming that an observation recorded as a death at x weeks is known only to have occurred within the interval $(x-1, x]$. Ideally, we should also take account of whatever mechanism has produced the unusual ties in the data, but this information is unavailable.

We consider a slightly different model to the Box–Cox model of Achcar *et al.* (1987) by assuming that the (unobserved) true survival times follow the GEV distribution for minima specified by equation (2.4). Thus, the likelihood for each observed failure time x_i is $F(x_i) - F(x_i - 1)$, with F as in (2.4). Attempts to fit this model by maximum likelihood fail due to the clustering of data close to the start of the experiment. Thus, in this situation, model (2.4) is likely to be best fitted with a shape parameter $\xi < -1$, but this is precisely when maximum likelihood fails due to the unboundedness of the likelihood function.

A Bayesian analysis is not, in principle, affected by the unboundedness so long as the likelihood is integrable with respect to the prior, though the computational burden is substantially increased. Technically, the Gibbs Sampler is highly persistent when sampling in regions close to the end-point of the distribution, leading to a requirement for extremely long samples in order to adequately cover the posterior distribution. Though this is a point of technicality rather than of analytical substance, it may mean that the technique would not be judged sufficiently time-efficient for general application to such problems.

We have adopted an almost flat prior distribution (reflecting prior ignorance), with zero prior weight on domains for which the lower end-point of the distribution (2.4) would be negative; this eliminates the possibility of posteriors giving positive probability to impossible events, though in some situations this possibility might be regarded as acceptable in order to obtain a better fit to the observed data. Applying the Gibbs sampler, the corresponding marginal posterior distributions for μ, σ and ξ are shown in Figure 4. From this figure it is almost certain that $\xi < -1$, confirming the reasons for the difficulty with maximum likelihood. Specifying point estimates on the basis of this analysis is more troublesome. Ideally, we would return the joint posterior mode as the inferred estimate, but this is not easily identified from the Gibbs sample output. On the other hand, the skewness in the posterior densities, and constraints on the parameter support, mean that the posterior mean $(\overline{\mu}, \overline{\sigma}, \overline{\xi}) = (105.1, 173.6, -1.67)$ is outside of the parameter space. That is, $t < \overline{\mu} + \overline{\sigma}/\overline{\xi}$ for at least one of the observed failure times, t. In order to obtain working estimates in this case, we have adopted the marginal posterior means for σ and ξ, and then adopted μ so that the lower end-point of the resulting model corresponds to the marginal posterior mean of that parameter. This gave the estimator $(\tilde{\mu}, \tilde{\sigma}, \tilde{\xi}) = (105.1, 173.6, 1.67)$. On the basis of this estimator, the transformation

$$Y = -\frac{1}{\xi} \log \left\{ 1 - \xi \left(\frac{T - \mu}{\sigma} \right) \right\} \tag{5.1}$$

produces variables having the standard Gumbel distribution (of minima). Applying equation (5.1) to produce standardized order statistics $y_{(i)}$ from the original order statistics $t_{(i)}$ permits standard graphical checks of model adequacy to be applied, though this doesn't allow for the discretized nature of the data in this example. Accordingly, the probability plot of Figure 5 suggests that the GEV model (2.4) provides as good a description of these data as the Box–Cox model suggested by Achcar *et al.* (1987).

Though it is a somewhat contrived question for these particular data, in many reliability modelling situations it is the lower end-point of the distribution which is of greatest scientific concern. To illustrate the ease with which the sample-based inference extends to functionals of the original parameters, the posterior distribution of the lower end-point, obtained by direct transformation of the Gibbs sample output, is included in Figure 4. As anticipated by the difficulties with maximum likelihood, most of the mass of the posterior distribution of the end-point is close to the lowest recorded data value of 1 week; an alternative analysis which does not account for the discreteness of the data emphasises this point to an even greater extent.

Thus, in situations of short tailed distributions, where $\xi < -1$, we observe that the Bayesian estimator is not adversely affected (except in terms of computation time) by the unboundedness of the likelihood function. It is perhaps also worth noting that since posterior inferences are not dependent on likelihood asymptotics, there is no added complication to the Bayesian analysis in the intermediate case of $-1 < \xi < -0.5$, where maximum likelihood estimators exist but are non-regular. In this situation all information about the regularity of the likelihood surface is automatically summarized within the posterior distribution. Smith (1989) and A.F.M. Smith (1995) have each suggested the use of Bayesian estimators for inference in non-regular likelihood problems.

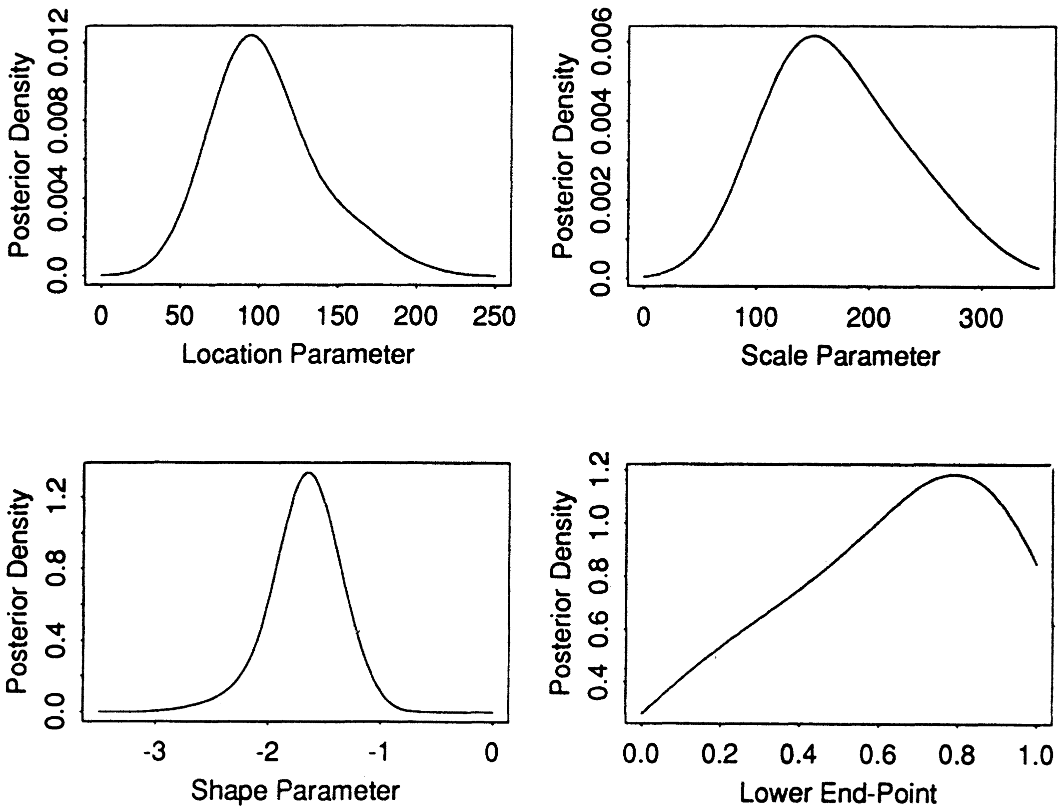

Figure 4. *Posterior distributions of GEV parameters and distributional lower end-point in survival data analysis.*

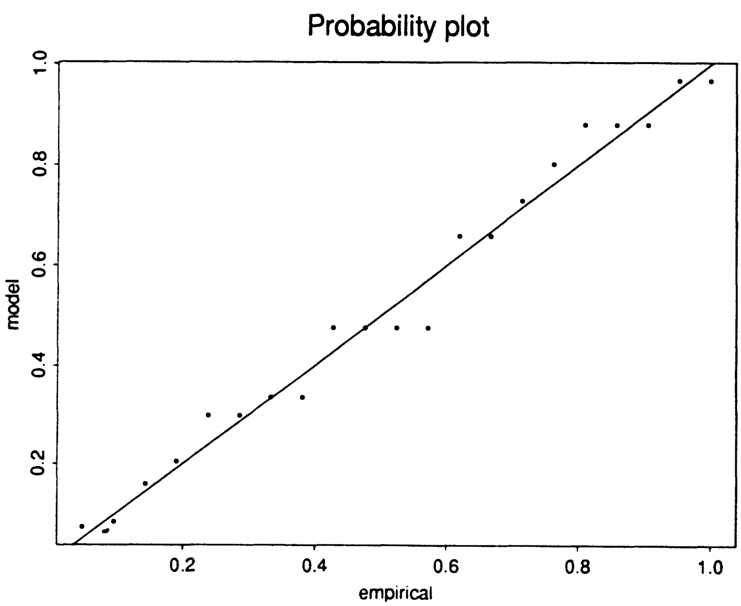

Figure 5. *Gumbel probability plot of transformed survival data.*

6 Spatial Information for Extreme Wind Speed Modelling

In this section we consider the utility of a simple Bayesian approach to the spatial modelling of extreme wind speeds. The specific issue we choose to examine is the extent to which a Bayesian analysis can be used to pool information spatially in situations where very few data for a specific location are available. In such cases conventional estimators can be imprecise and highly variable.

Our data are in the form of historical series of annual maximum wind speeds from 115 sites throughout the United States. The entire data set consists of 129 sites, but it is common to exclude those sites which are prone to hurricane wind speeds, since this phenomenon requires specialized modelling (Simiu & Scanlan, 1987, ch. 3). Assuming a simple GEV model is valid at each site, maximum likelihood estimates of each parameter set (μ_j, σ_j, ξ_j) can be obtained. Histograms of each parameter over all sites are given in Figure 6.

Spatial modelling of extreme value data has been considered by a number of authors (R.L. Smith, 1995; Coles & Tawn, 1990, 1996b). In most cases models are formulated to relate the GEV parameters (μ_j, σ_j, ξ_j) to site location or other covariate information such as altitude. Simple exploratory plots for these data fail to reveal any clear relationships. For example, in Figure 7 we show plots of the maximum likelihood estimates of the 50 and 1000 year return levels as functions of the latitude and longitude of the data site. On the other hand, in the absence of information to the contrary, we'd expect the extremal wind speeds at some arbitrary location to be not atypical to those at other locations in the country. Thus, it seems reasonable to adopt the information acquired spatially as a prior belief for the behaviour at the arbitrary site. As data become available, if they are consistent with data from elsewhere, our faith in the prior would be reinforced; otherwise, we would modify our beliefs accordingly. This, of course, is precisely what a Bayesian analysis does. In this case the prior beliefs are formulated empirically on the basis of the spatial information.

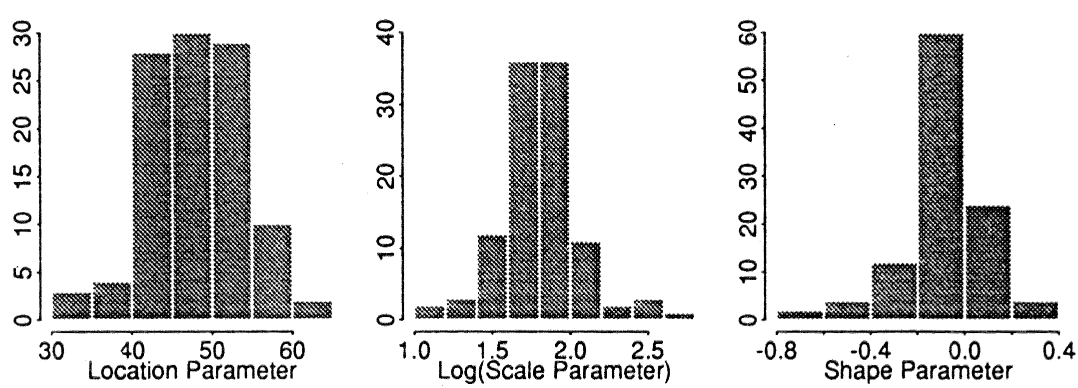

Figure 6. *Histograms of maximum likelihood estimates of GEV parameters for annual maximum wind speed at 106 locations in the USA.*

We have adopted the following scheme to investigate the effects of such an analysis. We selected 9 sites which give a wide geographical coverage and seemingly diverse extremal wind climates. Data from the remaining 106 sites were used to formulate prior beliefs about (μ, σ, ξ) for any of the 9 selected sites. In particular it was assumed that

$$(\mu, \log(\sigma), \xi) \sim MVN(\phi, \Sigma) \tag{6.1}$$

where ϕ and Σ are the empirical means and covariances of the maximum likelihood estimates over the 106 sites. In this way, conceptually at least, the GEV parameters, *a priori*, are presumed to follow the multivariate Normal distribution specified by equation (6.1). The precise values of ϕ and Σ for these data, measured in miles per hour, are found to be

$$\phi = (47.9, 1.79, -0.083), \quad \Sigma = \begin{pmatrix} 35.7 & 0.793 & -0.470 \\ 0.793 & 0.0656 & -0.0266 \\ -0.470 & -0.0266 & 0.0283 \end{pmatrix} \tag{6.2}$$

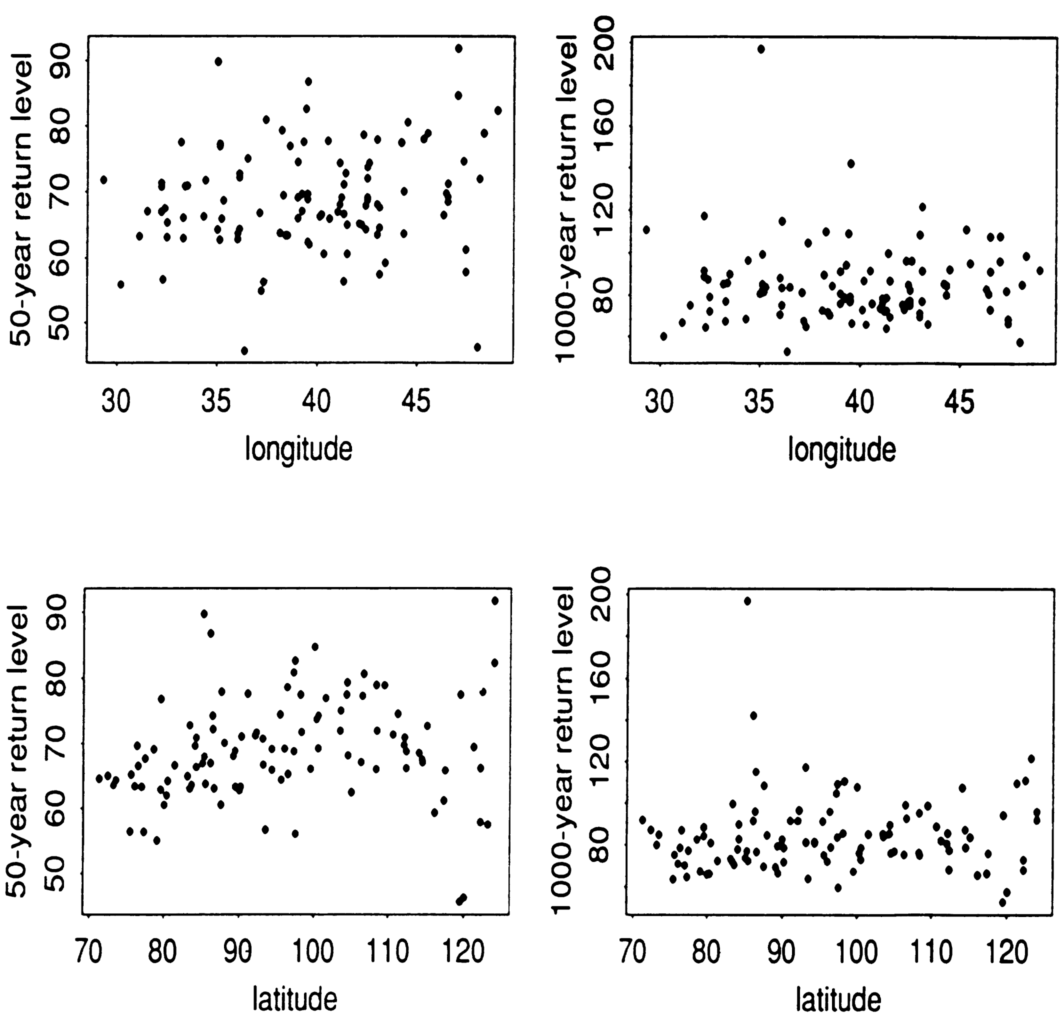

Figure 7. *Maximum likelihood estimates of annual maximum return levels versus longitude and latitude for 106 locations in the USA.*

The choice of multivariate Normality is arbitrary, and Figure 6 suggests that such an assumption is not overly precise given the observed estimates, though it is unlikely that the subsequent analysis is adversely affected by such a mis-specification. This is confirmed by Coles & Powell (1995) which gives an extended analysis of these data. A particular consequence of model (6.1) is that no *a priori* spatial structure is being assumed. The absence of any clear relationships (c.f. Figure 7) also make this assumption reasonable, but there is a potential biasing effect or at least a lack of efficiency if there is spatial structure which we are ignoring. We return to this point below.

The main issues we wish to address are:

(1) With very short data records, is the spatial information which is accumulated in the form of the prior distribution a useful surrogate for the sparsity of information in the form of data for each of the nine sites?
(2) As more data become available for a particular site, does the prior have a biasing effect, particularly for sites whose data seem atypical according to the prior information?

The nine selected data sites have between 37 and 48 years of annual maximum data, which are typical data lengths for the entire data set of 115 sites. For each site we mimic the effect of having sparse data by sequentially fitting models to the first 5, 10, and 20 years of data only, as well as to the entire data record respectively. With the reduced data sets this allows us to examine the informativeness of the spatial information in the absence of longer data records, as well as observing the effect of the prior on the inference as the data length increases. We report here the results for just three of the sites—San Diego, Portland and Minnesota; the behaviour at the other six sites was similar. The results for each site are given in Tables 1–3. These compare, in each case, for increasing amounts of data, the posterior means and standard deviations of the parameters (μ, σ, ξ) and the annual maximum quantiles $q_{.02}$ and $q_{.001}$ with the corresponding maximum likelihood estimates and their standard errors.

For all three sites it is seen that the prior means differ quite substantially from the maximum likelihood estimates based on all the available data. Thus the comparison is not unfairly biased because of having chosen sites which are unduly typical. We can summarize the main conclusions of the analysis as follows:

(1) Even with just 5 years of data, the estimates of μ in the Bayesian analysis become extremely well calibrated despite the uncertainty in the prior information. This parallels the observation made by Coles & Tawn (1996a) for the case study summarized in Section 4. In that case, they found the expert hydrologist's prior information was relatively weak with respect to μ because of the high spatial variation of this parameter, but that relatively few data were required to calibrate the posterior distribution of μ. By contrast, considerably more data are required to accurately calibrate σ and ξ so that the information provided by the prior for these parameters remains dominant until there is a substantial amount of data;
(2) With short data records (5 or 10 years) the maximum likelihood estimates are both unreliable and low in precision. By contrast, the corresponding Bayesian estimates appear consistent with estimates based on more complete data, and have acceptable levels of precision;
(3) The maximum likelihood estimates oscillate wildly as increasing amounts of data become available. In comparison, the Bayesian estimates are much more stable, especially for the shape parameter ξ, which has greatest influence on long-term extrapolation;
(4) The magnitude of standard errors of maximum likelihood estimates of ξ are influenced heavily by the precise estimate, $\hat{\xi}$, of ξ itself. In particular, when $\hat{\xi}$ is large and negative, corresponding to an estimated distribution with a very short upper tail, the standard errors are very small. Observe, for example, the analysis of the Minnesota data in Table

Table 2

Comparison of prior and posterior mean and standard deviations with maximum likelihood estimates and standard errors for annual maximum wind speeds at San Diego.

	Years of Data	μ	σ	ξ	$q_{.02}$	$q_{.001}$
Prior		47.9 (6.0)	5.99 (1.59)	−0.08 (0.17)	67.9 (7.8)	79.4 (15.8)
Posterior Mean (SD)	5	38.9 (2.1)	4.87 (1.04)	0.04 (0.15)	60.0 (6.1)	83.3 (24.2)
	10	34.8 (2.0)	5.58 (0.92)	−0.05 (0.13)	55.0 (5.2)	70.6 (16.0)
	20	31.7 (1.2)	4.90 (0.72)	0.04 (0.13)	52.7 (5.1)	74.4 (20.5)
	48	32.8 (0.7)	4.68 (0.47)	0.03 (0.07)	52.2 (3.3)	69.6 (10.6)
MLE (SE)	5	36.9 (1.4)	2.39 (1.38)	0.58 (0.73)	71.9 (65.1)	253.9 (821)
	10	34.3 (2.1)	5.93 (1.60)	−0.30 (0.29)	47.9 (4.0)	51.5 (8.3)
	20	31.4 (1.3)	4.91 (0.96)	−0.07 (0.22)	48.1 (5.6)	58.1 (16.2)
	48	32.7 (0.7)	4.49 (0.49)	−0.01 (0.08)	49.8 (2.9)	62.5 (7.7)

Table 3

Comparison of prior and posterior mean and standard deviations with maximum likelihood estimates and standard errors for annual maximum wind speeds at Portland.

	Years of Data	μ	σ	ξ	$q_{.02}$	$q_{.001}$
Prior		47.9 (6.0)	5.99 (1.59)	−0.08 (0.17)	67.9 (7.8)	79.4 (15.8)
Posterior Mean (SD)	5	47.3 (3.6)	6.83 (1.61)	−0.03 (0.15)	72.4 (5.8)	93.2 (18.6)
	10	49.0 (2.9)	7.74 (1.54)	−0.08 (0.14)	75.2 (5.2)	93.3 (16.7)
	20	47.6 (1.8)	7.08 (1.13)	−0.03 (0.12)	73.9 (5.0)	94.8 (17.5)
	45	44.0 (0.9)	5.37 (0.67)	0.11 (0.09)	70.3 (4.9)	102.5 (21.0)
MLE (SE)	5	46.3 (4.4)	7.39 (3.80)	0.29 (0.70)	100.2 (77.5)	212 (490)
	10	48.9 (3.0)	9.00 (2.43)	0.02 (0.38)	85.1 (23.3)	114.7 (81.0)
	20	47.3 (2.0)	7.19 (1.55)	0.11 (0.29)	82.0 (17.9)	120.6 (75.0)
	45	43.7 (0.9)	5.11 (0.76)	0.22 (0.14)	75.3 (10.0)	126.4 (48.5)

Table 4

Comparison of prior and posterior mean and standard deviations with maximum likelihood estimates and standard errors for annual maximum wind speeds at Minnesota.

	Years of Data	μ	σ	ξ	$q_{.02}$	$q_{.001}$
Prior		47.9 (6.0)	5.99 (1.59)	−0.08 (0.17)	67.9 (7.8)	79.4 (15.8)
Posterior Mean (SD)	5	42.8 (2.5)	5.55 (1.19)	−0.01 (0.15)	64.5 (5.8)	84.5 (21.0)
	10	44.6 (1.8)	5.71 (0.99)	−0.08 (0.14)	64.3 (5.1)	78.6 (15.5)
	20	43.8 (1.2)	5.19 (0.87)	0.13 (0.10)	70.4 (5.7)	105.7 (25.6)
	42	44.7 (0.9)	5.25 (0.67)	0.13 (0.09)	71.7 (4.9)	106.7 (21.4)
MLE (SE)	5	41.1 (2.3)	3.79 (2.06)	0.41 (0.71)	77.9 (61.2)	191 (520)
	10	44.8 (2.0)	5.41 (1.49)	−0.26 (0.30)	58.0 (4.4)	62.0 (9.6)
	20	43.4 (1.2)	4.71 (0.99)	0.33 (0.20)	80.4 (19.5)	166 (118)
	42	44.5 (0.9)	4.97 (0.73)	0.25 (0.14)	77.5 (11.0)	137 (56.9)

3. In this case the upper limit of a 95% confidence interval for $q_{.001}$ with 10 years of data is only slightly greater than the maximum likelihood estimate of $q_{.02}$ based on 20 years of data. By comparison, the values of precision of the Bayesian estimators are more stable with respect to the observed data.

(5) With complete data records, the Bayesian estimators are not inconsistent with the maximum likelihood estimates. This suggests that with enough local data, a Bayesian analysis is not distorted by any mis-specification of the empirical prior;

(6) At several sites (though not those considered in Tables 1 – 3) with either 5 or 10 years of data, the numerical maximization routine we use to find maximum likelihood estimators was unable to converge. Thus Bayesian estimates are attainable in situations where maximum likelihood estimates are not.

In each of the above respects, it appears that the Bayesian analysis performs favourably, especially for short data sets. The most important conclusion is that when only limited data at a given site are available, the spatial information seems genuinely informative when constructed as the prior in a Bayesian analysis, leading to improvements in the precision of inference. Again, refer to Coles & Powell (1995) for a more detailed study of these data. They examine, in particular, the role of the prior in obtaining the inferences made here. As in the case study of Coles and Tawn presented in Section 4, they find that analysis based on an uninformative prior is almost identical to an analysis based on profile likelihoods. This however is quite different to an analysis based on maximum likelihood and asymptotic standard errors due to the substantial skewness in profile likelihood surfaces for return levels based on very few years of data. The point is that a likelihood analysis is recovered in the Bayesian inference by working with flat priors, but that the empirical prior we have adopted is genuinely informative leading to substantially different inferences, particularly when few data are used.

As mentioned earlier, one implication of model (6.1) is that there is no spatial coherence to the GEV parameters. Whilst diagnostic plots such as those of Figure 7 confirm a lack of systematic relationship, it is possible, and indeed likely on meteorological grounds, that at least on a very localized basis, there should be some spatial structure within the true GEV parameters and consequently the maximum likelihood estimates. In an attempt to exploit this we have tried repeating the above analysis, but for each of the nine sites, in place of model (6.1), we have used an empirical prior which gives greater weight to estimates from closer sites. It transpires that such a modification has very little effect on the constructed priors, and consequently only a negligible effect on the posterior analysis. This suggests that with data on this geographical scale, there is indeed no spatial structure within the GEV parameters which can be exploited to improve on the earlier analysis.

7 Discussion

In a technical sense, many of the issues we have addressed in this paper have already been examined in the literature: prior choice, parameterization, computation and prediction. What we have tried to demonstrate here is that for genuine extreme value data modelling, the Bayesian paradigm is a powerful inferential mechanism with the potential to substantially improve upon the informativeness of an analysis which is solely data-based. Naturally, this calls into question the validity of beliefs which are independent of the observed data, but our concern is lessened by analyses we have reported here. In the cases of both expert and empirical priors, the posterior analyses we have obtained appear at least as trustworthy as maximum likelihood estimates, but considerably more flexible and informative. The viability of a Bayesian analysis when maximum likelihood fails has also been demonstrated.

We conclude by noting that in the Bayesian literature, the computational technique of Markov Chain Monte Carlo has been used to make inferences on models which are substantially more complex than those considered in this paper. Having now gained some reassurance that in relatively simple extreme value models there is some gain in adopting a Bayesian approach, we perceive that an active area of future research will be the application of Bayesian techniques to extreme value models of a greater sophistication than have been considered so far.

Acknowledgements

EAP is supported by a grant from the ESPRC. The work was carried out on equipment also provided by the ESPRC under the Complex Stochastic Systems Initiative. We are grateful to Richard Smith for comments on an earlier version of this paper, which was the basis of an invited paper for the 50th meeting of the I.S.I. held in Beijing, 1995.

References

Achcar, J.A., Bolfarine, H. & Pericchi, L.R. (1987). Transformation of survival data to an extreme value distribution. *The Statistician*, **36**, 229–234.

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

Ashour, S.K. & El-Adl, Y.M. (1980). Bayesian estimation of the parameters of the extreme value distribution. *Egyptian Statist. Jnl.*, **24**, 140–152.

Basu, A.P. (1964). Estimates of reliability for some distributions useful in life testing. *Technometrics*. **6**, 215–219.

Bhattacharya, S.K. (1967). Bayesian approach to life testing and reliability estimation. *J.A.S.A*, **62**, 48–62.

Berger, J.O. & Sun, D. (1993). Bayesian analysis for the Poly–Weibull distribution. *J.A.S.A.*, **88**, 1412–1418.

Coles, S.G. & Powell, E.A. (1995). Information utility in the modelling of extreme wind speeds. Submitted to Jnl. Wind Engineering.

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

Coles, S.G. & Tawn, J.A. (1996a). Modelling extremes: a Bayesian approach. To appear in *Appl. Statist.*

Coles, S.G. & Tawn, J.A. (1996b). Modelling extremes of the areal rainfall process. *J. Roy. Statist. Soc., B*, **58**.

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

Dey, D.K. & Lee, T. (1992). Bayes computation for life testing and reliability estimation. *IEEE Trans. Reliab.*, **41**, 621–626.

Engelund, S. & Rackwitz, R. (1992). On predictive distribution functions for the three asymptotic extreme value distributions. *Structural Safety*, **11**, 255–258.

Gilks, W.R., Richardson, S. & Spiegelhalter, D.J. (eds.) (1995). *Markov Chain Monte Carlo in practice*. London: Chapman & Hall.

Holla, M.S. (1966). Bayesian estimates of reliability function. *Australian J. Statist.*, **8**, 32–35.

Lingappaiah, G.S. (1984). Bayesian prediction regions for the extreme order statistics. *Biom. J.*, **26**, 49–56.

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

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

Rodrigues, J. & Louzada Neto, F. (1990). A note on the orthogonal reparameterization in Bayesian inference for the extreme value distribution. *Rev. Mat. Estatist.*, **8**, 95–100.

Simiu, E. & Scanlan, R.H. (1986). *Wind effects on structures*, New York: Wiley.

Singpurwalla, N.Z. & Song, M.S. (1988). Reliability analysis using Weibull lifetime data and expert opinion. *IEEE Trans. Reliab.*, **37**, 340–347.

Sinha, S.K. & Sloan, J.A. (1988). Bayes estimation of the parameters and reliability function of the 3-parameter Weibull distribution. *IEEE Trans. Reliab.*, **37**, 364–368.

Smith, A.F.M. (1995). Contribution to discussion of Non-regular maximum likelihood problems by Cheng, R.C.H. and Traylor, L. *J. Roy Statist. Soc.*, **57**, 30.

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

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

Smith, R.L. (1989). A survey of nonregular problems. *Proc. 47th Session of the I.S.I.*, 353–372.

Smith, R.L. (1992). Extreme value theory. In *Handbook of Applicable Mathematics*, **7**, 437–471, ed. W. Ledermann, Chichester: John Wiley.

Smith, R.L. (1995). Regional estimation from spatially dependent data. Submitted.

Smith, R.L. & Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Appl. Statist.*, **36**, 358–369.

Résumé

Les problèmes des valeurs extrèmes sont caractérisées par la pénurie des données et par l'exigence de modèlisation où les données sont le plus clairsemées. Ce fait devient un vrai dilemme quand on abord l'inférence d'une manière Baysienne: le valeur des renseignements antérieurs est probablement important, mais il nést pas du tout certain qu'on peut bien appliquer ces renseignements aux valeurs extrèmes. Dans cet article nous passons en revue la littérature qui relie les thèmes de l'analyse Baysienne et de l'analyse des valeurs extrèmes, et utilisons des avances récents en outils informatiques Bayesiens pour juger la valeur d'analyse Bayesien des valeurs extrèmes en trois situations différents: premièrement, où un expert est disponible pour fournir des renseignements antérieurs; deuxièmement, où 'maximum likelihood' échoue; et troisièmement, où on utilise des renseignements en espace sur des variables liées pour formuler un 'empirical prior'.

[*Received October 1995, accepted October 1995*]