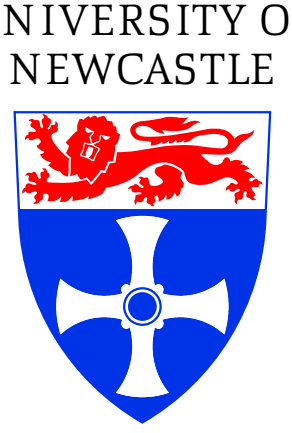
Bayesian Modelling of Extreme Rainfall Data

Elizabeth Smith

A thesis submitted for the degree of Doctor of Philosophy at the

University of Newcastle upon Tyne

September 2005



Bayesian Modelling of Extreme Rainfall Data

Elizabeth Smith

Summary

Inference on the extremes of environmental processes is essential for design-specification in civil engineering. Rainfall is such a process and structures need to be built to withstand the extremal behaviour of this process; for example, a reservoir should be capable of storing the amount of rainfall expected to fall in the region of interest. If the reservoir is not large enough, there is a risk that the water may overtop the dam. The rainfall to be stored in the reservoir may have come from a large region and may also build up over a number of days. For this reason it is important to consider multivariate models to model data from more than one location simultaneously, and also to consider modelling daily rainfall aggregates rather than just the annual maxima of these daily aggregates. Modelling the extremes of rainfall data can also be used to assess the risk of urban flooding and this information is often used for insurance purposes.

Multivariate techniques that can be applied to daily data have been developed but Bayesian techniques for these models have not received any serious consideration. Only univariate extreme value inference has been considered from a Bayesian standpoint. Bayesian inference in the context of extremes has some obvious advantages: extreme observations are naturally scarce, so incorporating any information supplementary to the data, in the form of a prior distribution, could be extremely advantageous. The application of Bayesian techniques has become practical through the recent development of simulation-based techniques such as Markov chain Monte Carlo.

Since the Bayesian framework has much to offer an extreme value analysis, and since multivariate modelling of extreme rainfall data is important, a Bayesian multivariate model for daily rainfall aggregates is developed, and the effect of using expert prior information and a multivariate model is discussed.

Acknowledgements

I am very grateful to my supervisor, Dave Walshaw, for all his help and encouragement over the last four years. I would also like to thank Darren Wilkinson and Richard Boys, who have always been willing to help when asked.

I have really appreciated the friendship and support of so many people within the School of Mathematics and Statistics; I would particularly like to thank Lee and Jonathan for all of their help. I would also like to thank Jill for her help, especially for helping to keep me calm before the viva. Finally, I would like to thank my friends and family for being there when I needed them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Key concepts of extreme value theory

Extreme value theory is defined as the study of the extremal properties of random processes. The objective of an extreme value analysis is to quantify the stochastic behaviour of a process at unusually high or low levels. In particular, an estimate of the probability of observing events more extreme than those already observed is often desired. This area has become a very important statistical discipline in applied sciences and has been used predominantly in modelling environmental phenomena. Other areas of application include finance, insurance, survival analysis, food science and telecommunications.

For environmental processes extreme value theory can be used to estimate the probabilities of extreme levels of the processes. For some processes, such as sea-level and wind speed, this information can help in the design of structures such as sea walls, bridges and buildings. For other processes, such as rainfall and pollution, the information can be used to assess danger due to extreme levels of the process. Extreme value theory can be used in finance to, for example, assess the risks of large insurance claims or predict the probability of rare events. In the area of survival analysis the Weibull distribution, which is one of the 'extreme value distributions', is often used to model remaining lifetimes.

Processes for which an extreme value analysis might be used could be either univariate

or multivariate. An example of a univariate process is sea-level measurement at one port on a coastline. This process could be studied to find the probability of coastal flooding at that particular port. If the sea-level data is available at a number of points along the coastline then the data from all of the ports can be considered as a multivariate process. The joint probability of coastal flooding can then be estimated using multivariate extreme value theory. Both univariate and multivariate processes have been studied extensively from a classical frequentist viewpoint. Recently there has been an increasing interest in Bayesian methods applied to extreme value problems, and there have been a number of studies based on univariate extreme value problems (see for example Coles and Powell (1996) and Coles and Tawn (1996)). There has, however, been little, if any, consideration of multivariate extreme value problems from a Bayesian viewpoint. The idea of this thesis is to extend the Bayesian approach to model the extremes of multivariate processes.

1.2 Thesis structure

In the remainder of Chapter 1, the key results of extreme value theory, which are used throughout the thesis, are described.

In Chapter 2 the rainfall data from eleven locations in south-west England, which are considered in this thesis, are introduced and basic univariate extreme value analyses are carried out at each location. The block maxima method and the threshold method, introduced in Chapter 1, are applied to the annual maxima data and the daily data respectively. This forms a preliminary analysis of the data, the results of which are compared to those obtained in later chapters. The thresholds obtained in this chapter for the analysis of the daily data are also used in Chapter 5 for the multivariate analysis.

Chapter 3 provides an overview of Bayesian modelling and introduces the Markov chain Monte Carlo techniques used throughout the thesis. A Bayesian implementation of the annual maxima model is then carried out for each location, with both non-informative and informative priors. This provides a demonstration of how the Bayesian techniques

are applied and also gives results which are compared with those obtained from bivariate analyses in Chapter 4.

In Chapter 4 Bayesian techniques are used to model bivariate annual maxima data from each pair of sites by using the multivariate theory of Chapter 1. This is a first step towards a Bayesian multivariate model for extremes. Two different models for dependence are considered and both non-informative and informative priors are used. The effect of using the informative prior, and also the bivariate model, as opposed to univariate analyses, are each considered.

Chapter 5 extends the work of Chapter 4 to model all extreme data from all eleven locations simultaneously. The multivariate point process method, described in Chapter 1, is used to achieve this.

The dependence model used in Chapter 5 is not entirely appropriate. In Chapter 6 an alternative model is considered. This model cannot easily be applied to all eleven locations simultaneously, so it is applied to smaller groups of sites. The models for each group are then linked together to provide a model for all eleven locations.

1.3 Classical extreme value theory

1.3.1 The 'extremal types' characterisation

Classical extreme value theory focuses on the behaviour of

$$M_n = \max\{X_1, \ldots, X_n\}$$

where $X_1, \ldots, X_n$ is a sequence of independent and identically distributed (IID) random variables with distribution function F. The properties of M_n as $n \to \infty$ are of particular importance.

Theoretically, the distribution of M_n can be derived as

$$\begin{aligned} Pr\{M_n \leq z\} &= Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\ &= Pr\{X_1 \leq z\} \times \ldots \times Pr\{X_n \leq z\} \\ &= \{F(z)\}^n, \end{aligned}$$

for all n. In practice, however, F is unknown. Standard statistical techniques could be applied to observed data to estimate F leading to an estimate for the distribution of M_n. In this approach, however, small discrepancies in the estimate of F could result in large discrepancies for F^n.

An alternative approach is to estimate approximate families of models for F^n, using only the extreme data. The limiting distribution of M_n, however, is degenerate, since M_n converges to the upper endpoint of F with probability 1. This means that to consider the behaviour of F^n as $n \to \infty$ would not be enough. Considering a linear renormalisation of M_n:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, avoids this problem. The key result, known as the extremal types theorem, gives the entire range of possible limiting distributions, where the limit is taken as $n \to \infty$, for M_n. Different aspects of this result were proved by Fisher and Tippett (1928) and Gnedenko (1943).

Theorem 1.1 (Extremal Types Theorem) *If there exist sequences of constants* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that*

$$Pr\{(M_n - b_n)/a_n \leq z\} \to G(z) \qquad \text{as } n \to \infty, \tag{1.1}$$

where G *is a non-degenerate distribution function, then* G *belongs to one of the following*

families:

$$I: \quad G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \qquad -\infty < z < \infty;$$

$$II: \quad G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases}$$

$$III: \quad G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b; \\ 1, & z \geq b, \end{cases}$$

for parameters $a > 0$, b, and $\alpha > 0$.

Theorem 1.1 says that if there are suitable sequences $\{a_n\}$ and $\{b_n\}$ to stabilise M_n, then M_n^* has a limiting distribution which is one of the types I, II and III. These three families of distributions are known as the *Gumbel*, *Fréchet* and *Weibull* families respectively and are collectively known as the *extreme value distributions*.

The family of extreme value distributions can be equivalently characterised via the class of *max-stable* distributions. A distribution G is said to be max-stable if for $n = 2, 3, \ldots$ there are constants $\alpha_n > 0$ and β_n such that $G^n(\alpha_n x + \beta_n) = G(x)$. The connection between extreme value theory and the max-stability is established by the following theorem (Theorem 1.4.1 of Leadbetter *el al* (1983)). Before stating the theorem it is necessary to define an equivalence class of distributions: the distributions F and F^* are of the *same type* if there are constants a and b such that $F^*(ax + b) = F(x)$ for all x.

Theorem 1.2 *A distribution is max-stable if, and only if, it is of the same type as an extreme value distribution.*

It follows from Theorem 1.2 that the distribution of the maxima of independent samples will be of the same type as that of the underlying population if, and only if, the underlying population itself has a distribution of extreme value type.

1.3.2 The generalised extreme value distribution

The behaviour of the three types of limit arising in Theorem 1.1 are of distinct forms. These distinctions correspond to different forms of tail behaviour for the distribution

function F. The three models can be combined into one family of models known as the *generalised extreme value* (GEV) family with distribution functions of the form

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{1.2}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$ and with parameters satisfying $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. This distribution was independently derived by von Mises (1954) and Jenkinson (1955). The parameters μ, σ and ξ are location, scale and shape parameters respectively. It is the value of the shape parameter, ξ, that distinguishes the three classes of extreme value distribution. Types II and III in Theorem 1.1 correspond to values of $\xi > 0$ and $\xi < 0$ respectively. The case of $\xi = 0$ is interpreted as the limit as $\xi \to 0$ of (1.2). This leads to the Gumbel family of distributions, where the distribution function is

$$G(z) = \exp\left[-\exp\left\{-\left(\frac{z-\mu}{\sigma}\right)\right\}\right], \qquad -\infty < z < \infty. \tag{1.3}$$

In practical applications of Theorem 1.1, the limit (1.1) is interpreted as an approximation for large values of n; i.e.

$$Pr\left\{\left(M_n - b_n\right)/a_n \leq z\right\} \approx G(z).$$

Equivalently,

$$\begin{aligned} Pr\left\{M_n \leq z\right\} &\approx G\left\{(z-b_n)/a_n\right\} \\ &= G^*(z), \end{aligned}$$

where G^* is also a member of the GEV family. So, if the distribution of M_n^* can be approximated by a member of the GEV family for large n, then the distribution of M_n can be approximated by another member of the family. Hence the GEV family can be fitted directly to a series of observations of M_n.

To model the extremes of a series of independent observations, $X_1, X_2, \ldots$, the data can be blocked into m sequences of length n, where n is some large number. Taking the maxima of each block generates a series, $M_{n,1}, \ldots, M_{n,m}$, to which the GEV distribution

can be fitted. The choice of block size, n is important. A value of n which is too small is likely to lead to a poor approximation by the limit model in Theorem 1.1. This would lead to bias in the estimation of parameters and, consequently, in extrapolation. If n is too large, however, too few block maxima will be obtained, resulting in large estimation variances. A balance needs to be found between the bias and the size of variances. Typically, the blocks correspond to a time period of one year (i.e. n is the number of observations in a year) resulting in a series of annual maxima data.

Several techniques for parameter estimation in extreme value models have been considered, but maximum likelihood estimation has generally been considered to be the best method. Hosking (1990), however, suggests that in small samples L-moments may provide a more robust procedure than maximum likelihood. One problem with using maximum likelihood is that the regularity conditions underlying the classical principals of maximum likelihood estimation do not hold for the GEV. This means that standard asymptotic likelihood results are not automatically applicable. Smith (1985) found that for $\xi > -0.5$ maximum likelihood estimators have the usual asymptotic properties; for $-1 < \xi < -0.5$ maximum likelihood estimators are obtainable in general but do not have the standard asymptotic properties; for $\xi < -1$, maximum likelihood estimators are unlikely to be obtainable, and often don't exist.

Simulation based techniques such as Markov chain Monte Carlo (MCMC) have provided a way for Bayesian techniques to be applied. A Bayesian analysis of extreme value data has many potential benefits over the maximum likelihood approach: sources of information other than the data, which are likely to be scarce, can be included; the predictive distribution provides a natural way to estimate the probability of future events being extreme; a Bayesian analysis is not dependent on the regularity assumptions required by the asymptotic theory of maximum likelihood, so provides a reasonable alternative when $\xi < -0.5$.

Estimates of extreme quantiles of the annual maxima are of particular interest in environmental extremes, as they give an estimate of the level the process is expected to

exceed once, on average, in a given number of years. These quantiles are obtained by inverting (1.2) and (1.3):

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \xi \neq 0, \\ \mu - \sigma \log\left[-\log(1-p)\right], & \xi = 0, \end{cases} \tag{1.4}$$

where $G(q_p) = 1 - p$. The quantity q_p is known as the return level associated with the $1/p$-year return period: it is, approximately, the level which is expected to be exceeded on average once every $1/p$ years, or more precisely, it is the level exceeded by the annual maxima in any year with probability p. To obtain estimates of return levels substantially beyond the range of the available data, the relationships given in (1.4) can be extrapolated to the required level. One useful tool for looking at the extrapolations of return levels is the return level plot. A return level plot is a plot of q_p against return period, usually shown on a logarithmic scale. This has the effect of compressing the tail region, so that return levels are displayed for long return periods. Linear plots are obtained for the case $\xi = 0$, providing a baseline against which the effect of the shape parameter can be judged.

When maximum likelihood is used to estimate the parameters of the GEV distribution, the maximum likelihood estimates can be substituted into (1.4) to give a maximum likelihood estimate of q_p, for $0 < p < 1$. To obtain a standard error for q_p, the delta method can be used (Rao, 1973). In a Bayesian analysis, vectors of simulated values from the marginal posterior distributions of the GEV parameters are obtained. Applying (1.4) to each component of the vectors for μ, σ and ξ, results in a sample from the posterior distribution for the $1/p$-year return level from which summary statistics can then be obtained. Predictive distributions for return levels can also be obtained.

1.4 Threshold models

Using the GEV distribution to model block maxima is very wasteful of data if more data on the extremes are available. One method, which was developed to overcome this problem, is the r largest order statistics method. This method considers the limiting joint

distribution of the r largest order statistics, where these statistics are suitably normalised. Initial developments of this approach were carried out by Weissman (1978), with further studies by Smith (1986) and Tawn (1988a). This method, however, can also be wasteful of data. If, for example, one block of data has a lot of extreme observations relative to other blocks, many of these may not be used as only the r largest are modelled, and r is small enough so that the r largest observations in all blocks can be regarded as extreme. An alternative approach is to model all observations exceeding a specified high threshold. This model was originally proposed by Pickands (1975) and a comprehensive treatment of the model is given by Davison and Smith (1990).

Suppose $X_1, X_2, \ldots$ is a sequence of IID random variables with common distribution function F. The X_i exceeding a high threshold u can be regarded as extreme events. Let X denote an arbitrary term of the sequence, then the conditional probability

$$Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \qquad y > 0,$$

describes the stochastic behaviour of extreme events. The distribution of threshold exceedances would therefore be known if F was known. In practice F is unknown, so the distribution must be approximated.

1.4.1 The generalised Pareto distribution

Theorem 1.3 *Let $X_1, X_2, \ldots$ be a sequence of IID random variables with common distribution function F, and let*

$$M_n = \max\{X_1, \ldots, X_n\}.$$

Let X denote an arbitrary term of the X_i sequence. Suppose also that F satisfies Theorem 1.1, so that for large n,

$$Pr\{M_n \leq z\} \approx G(z),$$

where

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

for some μ, $\sigma > 0$ *and* ξ. *Then, for suitably large* u, *the distribution function of* $(X - u)$, *conditional on* $X > u$, *is approximately*

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{1.5}$$

defined on $\{y : y > 0$ *and* $(1 + \xi y/\tilde{\sigma}) > 0\}$, *where*

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

The case $\xi = 0$ is interpreted by taking the limit $\xi \to 0$ in (1.5). This leads to

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0, \tag{1.6}$$

which corresponds to an exponential distribution with parameter $1/\tilde{\sigma}$.

The *generalised Pareto family* is the name given to the family of distributions defined by (1.5). Theorem 1.3 is the key result for modelling threshold exceedances and it implies that, if G is the approximating distribution of block maxima, then there is a corresponding approximate distribution for threshold exceedances from within the generalised Pareto family. Additionally, the parameters of the associated GEV distribution uniquely determine the parameters of the distribution of threshold exceedances for any given threshold u. The shape parameter ξ is, in fact, equal to the corresponding shape parameter of the GEV distribution, while the scale parameter $\tilde{\sigma}$ is equal to $\sigma + \xi(u - \mu)$.

Expressions for return levels in terms of the GPD parameters can be obtained. Suppose a GPD with parameters σ and ξ is an appropriate model for exceedances of a threshold u by the variable X. So, assuming $\xi \neq 0$, for $x > u$

$$Pr\{X > x | X > u\} = \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}.$$

It follows that

$$Pr(X > x) = \zeta_u \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi},$$

where $\zeta_u = Pr(X > u)$. The level x_m that is exceeded on average once every m observations, then, is the solution of

$$\zeta_u \left[1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}.$$

Rearranging gives

$$x_m = u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right], \tag{1.7}$$

provided m is large enough to ensure that $x_m > u$. For $\xi = 0$, x_m can be found similarly from (1.6), giving

$$x_m = u + \sigma \log(m\zeta_u), \tag{1.8}$$

for sufficiently large m. The quantity x_m is the *m-observation return level*. If there are n_y observations per year then the N-year return level is given by

$$z_N = \begin{cases} u + \frac{\sigma}{\xi}\left[(Nn_y\zeta_u)^{\xi} - 1\right], & \xi \neq 0, \\ u + \sigma \log(Nn_y\zeta_u), & \xi = 0, \end{cases} \tag{1.9}$$

The quantity ζ_u has a natural estimator, $\hat{\zeta}_u = k/n$, which is the sample proportion of points exceeding u. To obtain estimates of the return levels, then, the estimate of ζ_u and the maximum likelihood estimates of σ and ξ can be substituted into the expression for the return level. Standard errors can be obtained using the delta method (Rao, 1973). If a Bayesian analysis is carried out, the vectors of realisations from the marginal posteriors of σ and ξ can be substituted into the expression for the return level to give a series of realisations from the posterior distribution of the return level. Return level plots could be obtained by plotting x_m against m on a logarithmic scale.

1.5 The point process characterisation

All of the models introduced so far can be derived as special cases of a point process characterisation given by Pickands (1981). The fundamental result of this characterisation is given in the following theorem.

Theorem 1.4 *Let $X_1, X_2, \ldots$ be a series of independent and identically distributed random variables, with $M_n = \max\{X_1, \ldots, X_n\}$, for which there are sequences of constants $a_n > 0$ and b_n such that*

$$Pr\left\{(M_n - b_n)/a_n \leq z\right\} \to G(z),$$

where

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

and let z_- *and* z_+ *be the lower and upper endpoints of* G *respectively. Then the sequence of point processes*

$$P_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \ldots, n\}$$

converges on regions of the form $(0,1) \times [u, \infty)$*, for any* $u > z_-$*, to a Poisson process, with intensity measure on* $A = [t_1, t_2] \times [z, z_+)$ *given by*

$$\Lambda(A) = (t_2 - t_1)\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}. \tag{1.10}$$

This theory was extended by Coles and Tawn (1991) to a representation for multivariate extremes, which is given in section 1.6.2 and used in Chapter 5.

1.6 Multivariate theory

1.6.1 Componentwise Maxima Method

Let $(X_{i,1}, \ldots, X_{i,p})$, $i = 1, \ldots$ be a sequence of independent and identically distributed realisations of a random vector and define

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,p})$$

to be the vector of componentwise maxima where $M_{n,j} = \max\{X_{1,j}, \ldots, X_{n,j}\}$, $j = 1, \ldots, p$. The index i for which the maximum of each sequence $\{X_{i,j}\}, j = 1, \ldots, p$ occurs will not necessarily be the same for each j. This means that $\boldsymbol{M}_n$ does not necessarily correspond to an observed vector of the series.

Univariate extreme value results can be applied to each component of $\boldsymbol{M}_n$ as for each j $\{X_{i,j}\}$ is a sequence drawn from an independent univariate random variable. For simplicity, assume that for each j $X_{i,j}$ has a standard Fréchet distribution with distribution function $F(z) = \exp(-1/z)$. The Fréchet distribution is a special case of the GEV

distribution with parameters $\mu = 0$, $\sigma = 1$ and $\xi = 1$. Due to the max-stability property of unit Fréchet variables the vector

$$\boldsymbol{M}_n^* = (M_{n,1}/n, \ldots, M_{n,p}/n) \tag{1.11}$$

should be considered in order to obtain standard univariate results for each margin. The limiting joint distribution of M_n^* as $n \to \infty$ is characterised by the following theorem (de Haan and Resnick (1977), Pickands (1981)).

Theorem 1.5 *A distribution function G is a limiting distribution of the random vector (1.11) with unit Fréchet margins if, and only if, it has the form*

$$G(\boldsymbol{x}) = \exp\{-V(\boldsymbol{x})\} \tag{1.12}$$

for $\boldsymbol{x} = (x_1, \ldots, x_p) \in \boldsymbol{R}_+^p$, where

$$V(\boldsymbol{x}) = \int_{S_p} \max_{j=1,\ldots,p} \left(\frac{w_j}{x_j}\right) dH(\boldsymbol{w}) \tag{1.13}$$

for some positive measure H defined on the $(p-1)$-dimensional unit simplex

$$S_p = \left\{ \boldsymbol{w} = (w_1, \ldots, w_p) \in \boldsymbol{R}_+^p : \sum_{j=1}^{p} w_j = 1 \right\}, \tag{1.14}$$

subject to

$$\int_{S_p} w_j dH(\boldsymbol{w}) = 1; \qquad j = 1, \ldots, p. \tag{1.15}$$

The functions V and H are called the exponent measure and the dependence measure respectively. Specification of either V or H would characterise the dependence structure of the limiting distribution G.

Generalising the marginal distributions by letting

$$\tilde{x}_j = \left[1 + \xi_{x_j}\left(\frac{x_j - \mu_{x_j}}{\sigma_{x_j}}\right)\right]^{\xi_{x_j}^{-1}} \tag{1.16}$$

means that the complete family of multivariate extreme value distributions, with arbitrary GEV margins, has a distribution function of the form:

$$G(\boldsymbol{x}) = \exp\{-V(\tilde{\boldsymbol{x}})\}, \tag{1.17}$$

if $\left[1+\xi_{x_j}\left(x_j-\mu_{x_j}\right)/\sigma_{x_j}\right] > 0$ and V satisfies (1.13) for some H. The family of distributions in (1.17), with arbitrary GEV margins, is known as the family of multivariate extreme value distributions.

There is no finite parametrisation for the limit family of (1.17), so parametric sub-families of distributions for H are used. These lead to sub-families of distributions for G. The two families considered in this thesis are the logistic family and the mixed family. The logistic family generates a sub-family of bivariate extreme value distributions which allows all levels of dependence, whereas in the mixed family very strong dependence and complete dependence is not possible.

In practice, the block maxima from each series can be considered as sequences of independent block maxima. The GEV distribution can be used to model these sequences and maximum likelihood estimators can be obtained for the parameters. Using the transformation (1.16), the margins can be transformed to have approximate standard Fréchet distributions. The sequence of vectors obtained after transformation are realizations of a vector with a multivariate extreme value distribution. Maximum likelihood techniques could then be used to get an estimate for the dependence parameter. Alternatively, the transformation and maximum likelihood could be done together. This would require the replacement of x and y in the likelihood with $\tilde{x}$ and $\tilde{y}$.

1.6.2 Multivariate point process method

As with the univariate approach, the block maxima method is wasteful of data. The multivariate point process method, described here, avoids this problem (de Haan and Resnick (1977), deHaan (1985), Resnick (1987)).

Theorem 1.6 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be a sequence of IID random vectors on $\boldsymbol{R}_+^p$, where the margins are identically distributed with unit Fréchet distributions. Suppose the random vectors satisfy the convergence for componentwise maxima*

$$Pr\left\{M_{n,1}^* \leq x_1, \ldots, M_{n,p}^* \leq x_p\right\} \rightarrow G(\boldsymbol{x}),$$

for $\boldsymbol{x} = (x_1, \ldots, x_p)$. *Let* P_n *be a process of points on* $\boldsymbol{R}_+^p$, *where* $P_n = \{n^{-1}\boldsymbol{X}_i : i = 1, \ldots, n\}$. *Then,*

$$P_n \xrightarrow{d} P,$$

where P *is a non-homogeneous Poisson process on* $\boldsymbol{R}_+^p \backslash \{\boldsymbol{0}\}$. *Define*

$$r_i = \sum_{j=1}^{p} X_{i,j}/n, \qquad w_{i,j} = X_{i,j}/nr_i, \qquad i = 1, \ldots, n,\ j = 1, \ldots, p,$$

where the j*th element of* $\boldsymbol{X}_i$ *is given by* $X_{i,j}$. *Then the intensity measure* μ *of* P *satisfies*

$$\mu(dr \times d\boldsymbol{w}) = \frac{dr}{r^2} dH(\boldsymbol{w}), \tag{1.18}$$

where H *is a positive finite measure on the* $(p-1)$*-dimensional unit simplex (1.14) satisfying (1.15), and* H *and* G *are related through (1.12) and (1.13).*

This is the only constraint on H and it must be satisfied so that the margins have the correct form. Since there are no other constraints on H, there is no finite parametrisation for it. Parametric sub-families of distributions for H are used, which result in parametric sub-families of distributions for G.

By taking $A = \boldsymbol{R}_+^p \backslash \{(0, x_1) \times \ldots \times (0, x_p)\}$ it can be seen that the componentwise maxima representation, given in Theorem 1.5, is a consequence of this limiting Poisson process. In this case $Pr\left(n^{-1}\boldsymbol{X}_i \notin A, i = 1, \ldots, n\right) \to \exp\{-\mu(A)\}$ as $n \to \infty$, where

$$\begin{aligned} \mu(A) &= \int_A \frac{dr}{r^2} dH(\boldsymbol{w}) \\ &= \int_{S_p} \int_{r=\min_{1\le j\le p}(x_j/w_j)}^{\infty} \frac{dr}{r^2} dH(\boldsymbol{w}) \\ &= \int_{S_p} \max_{1\le j\le p} \left(\frac{w_j}{x_j}\right) dH(\boldsymbol{w}). \end{aligned} \tag{1.19}$$

Since $Pr(n^{-1}\boldsymbol{X}_i \notin A, i = 1, \ldots, n) = Pr(n^{-1}M_{n,j} \le x_j, j = 1, \ldots, p)$, any limit distribution of normalised componentwise maxima, with unit Fréchet margins, has the form

$$G(\boldsymbol{x}) = \exp\left\{-\int_{S_p} \max_{1\le j\le p} \left(\frac{w_j}{x_j}\right) dH(\boldsymbol{w})\right\}, \tag{1.20}$$

for H satisfying (1.15). This is the same as the distribution given in Theorem 1.5.

1.7 Modelling dependent and non-stationary sequences

In the asymptotic results given for modelling extreme data the underlying processes are assumed to be IID. The data considered in extreme value analyses, however, usually do not have these properties. Short term temporal dependence is often present, since extreme levels of a process often continue over several observations. This is very common in environmental data, such as the rainfall data which are considered in this thesis. Also, processes often have characteristics that change over time, such as seasonal effects or trends. These are examples of non-stationarity and are, again, often present in environmental data. Seasonal effects are particularly common, since different weather patterns are observed at different times of the year. Trends could also be present due to long-term climate change, for example.

Theoretical models have been developed to deal with dependence and non-stationarity but few of these have been used in statistical modelling. It is more common to use statistical procedures which allow the application of the standard theory. Some of the theoretical results and statistical procedures which can be applied to non-IID series are given below.

1.7.1 Dependent sequences

In order to obtain a general characterisation of extremal behaviour certain constraints need to be imposed. Assuming a stationary sequence, it is usual to impose 'Leadbetter's $D(u_n)$ condition' (Leadbetter *et al.* (1983)) on the sequence, which limits the extent of long-range dependence at extreme levels. This condition is given in Definition 1.1 below.

Definition 1.1 *A stationary series $X_1, X_2, \ldots$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,*

$$\begin{aligned}\Big| Pr\left\{X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n\right\} \\ - Pr\left\{X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n\right\} Pr\left\{X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n\right\}\Big| \leq \alpha(n, l), \end{aligned} \quad (1.21)$$

where $\alpha(n, l_n) \to 0$ for some sequence l_n such that $l_n/n \to 0$ as $n \to \infty$.

The difference in probabilities expressed in (1.21) is zero for sequences of independent variables for any sequence u_n. Generally, it is required that the $D(u_n)$ condition holds only for a particular sequence of thresholds u_n that increase with n. Then, for sets of variables far enough apart, the $D(u_n)$ condition ensures that the difference between the probabilities is sufficiently close to zero to have no effect on the limit laws for extremes. The following theorem states this result more formally.

Theorem 1.7 *Let $\tilde{Y}_1, \tilde{Y}_2, \ldots$ be a stationary series satisfying Leadbetter's $D(u_n)$ condition and let $\tilde{M}_n = \max\left\{\tilde{Y}_1, \ldots, \tilde{Y}_n\right\}$. Now let $Y_1, Y_2, \ldots$ be an* independent *series with Y having the same distribution as $\tilde{Y}$, and let $M_n = \max\{Y_1, \ldots, Y_n\}$. Then if M_n has a non-degenerate limit law given by $Pr\{(M_n - b_n)/a_n) \leq y\} \rightarrow G(y)$, it follows that*

$$Pr\left\{(\tilde{M}_n - b_n)/a_n) \leq y\right\} \rightarrow G^{\theta}(y) \tag{1.22}$$

for some $0 \leq \theta \leq 1$.

The parameter θ in equation (1.22) is called the *extremal index*. The max-stability property of the extreme value family, described by Theorem 1.2, means that since $G(.)$ is an extreme value distribution, $G^{\theta}(.)$ must also be an extreme value distribution but with different values for the scale and location parameters. This implies that the GEV distribution is still the appropriate limit distribution for maxima of dependent sequences, satisfying $D(u_n)$.

The extremal index measures the propensity of the process to cluster at extreme levels and one interpretation, given by Leadbetter (1983), is that $1/\theta$ is the limiting mean cluster size of exceedances of increasingly high thresholds. Estimating θ, then, may be of interest and in practice θ is sometimes estimated using empirical realisations of the mean cluster size. This technique is discussed further by Leadbetter *et al.* (1989) and Tawn (1992).

The usual approach to dealing with dependence in block maxima data is just to apply the GEV model to the data, provided that long-range dependence is weak. This is because, by Theorem 1.7, the GEV distribution is still appropriate for modelling block maxima data with short term dependence. If more data are available and a threshold

model is preferred, a different approach is needed to deal with clusters of exceedances. The most commonly used method is *declustering*, where the dependent observations are filtered leaving a set of approximately independent threshold exceedances. An empirical rule is used to define the clusters, then the maximum observation within each cluster is extracted. This filtered series can be assumed to be independent, with the GPD being the distribution of conditional exceedances. The GPD can then be fitted to these cluster maxima. Results can be sensitive to the choice of clusters, so care is needed when determining the empirical rule. Davison and Smith (1990) and Tawn (1988a) discuss this issue for univariate processes, and Coles and Tawn (1991, 1994) consider the generalisation of this procedure to multivariate processes. An "automatic" declustering scheme, which avoids this arbitrary choice of clusters, is proposed by Ferro and Segers (2003). Although declustering is a simple way of dealing with dependence it is also wasteful of data, since all observations apart from the cluster maxima are discarded. Fawcett and Walshaw (2005) considered this issue and found that declustering is usually misleading and unnecessary.

1.7.2 Non-stationary sequences

A process may exhibit non-stationarity in many ways but trends and seasonality are particularly common. Both of these forms of non-stationarity can be accommodated by fitting the usual models but with time-dependent covariates. Examples of this type of modelling are given in Smith (1986), Tawn (1988a) and Davison and Smith (1990). Another technique for dealing with seasonality is to split the data up into seasons, within which the data can be assumed homogeneous, and model the data for each season separately. Examples of this approach are found in Smith (1989) and Walshaw (1994).

Chapter 2

Rainfall data

2.1 Introduction

The data considered throughout consist of daily aggregates of rainfall (measured in millimetres) at a network of 11 sites within a $40km \times 40km$ region of south-west England. A spatial plot of the network of sites is given in Figure 2.1.

This data has been used several times in the past. Coles (1993) uses the data to obtain a model for the whole region using the theory of max-stable processes (de Haan 1984) and Coles and Tawn (1996a) extend this to model the areal rainfall of the region. A temporal study of the data is given by Coles (1994), and a Bayesian analysis of the data at one of the sites is given by Coles and Tawn (1996b). Schlather and Tawn (2003) analyse the data using non-parametric estimators of a dependence measure for multivariate and spatial extremes.

The models discussed in Chapter 1 assume that the data are temporally independent. The temporal aspects of the data were studied by Coles (1994), revealing only weak temporal dependence. Coles (1993) found that there were strong seasonal effects present in the data and restricted his analysis to the data from November to February. Reasonable stationarity was observed in the restricted dataset. He also took the maxima over two-day periods in order to reduce the temporal dependence.

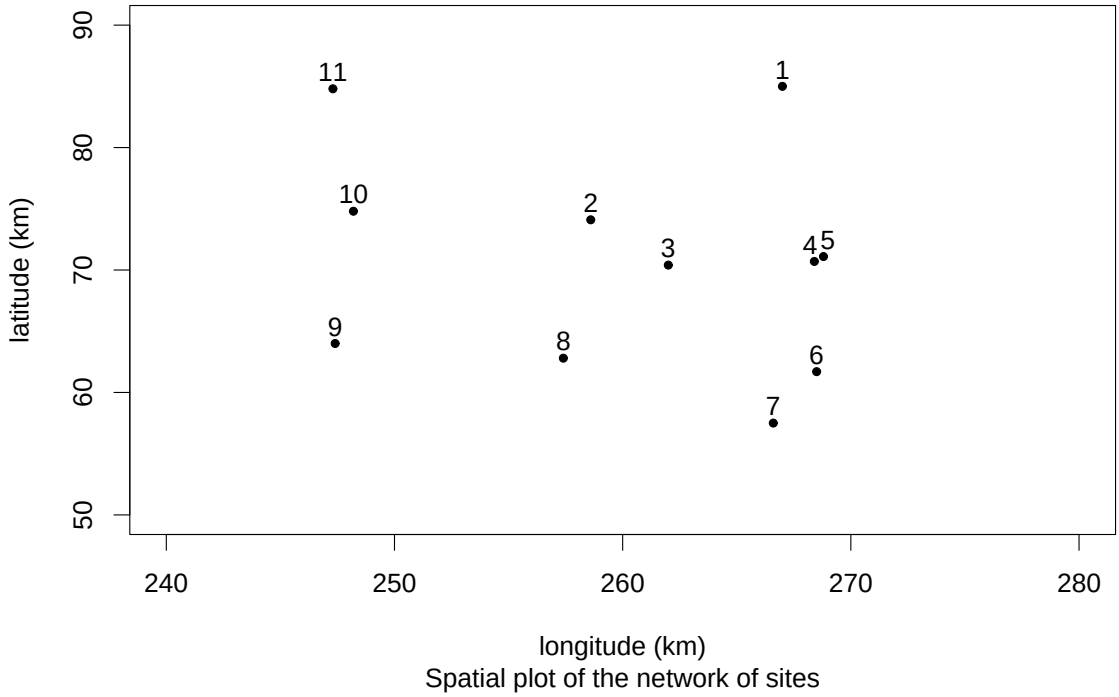


Figure 2.1: Spatial plot of the network of sites

Throughout this thesis models will be applied to both the annual maxima data and the daily data. In this chapter, maximum likelihood techniques are used to apply the univariate models of 1.3 to the annual maxima data (in Section 2.2) and the daily data (in Section 2.3). A discussion of the results obtained for the two models is given in Section 2.4.

2.2 Annual maxima

In Chapter 4 multivariate extreme value models are applied to bivariate annual maxima data from pairs of sites from Figure 2.1. This section provides a preliminary analysis of the annual maxima data by considering each site separately.

Table 2.1 gives the number of annual maxima observations used and the span of the observations for each site. The number of years of data used range from 35 years of data at site 3 to 65 years at site 2.

Site	No. of years	Span
1	56	1916–1973
2	65	1912–1973
3	35	1929–1964
4	60	1927–1988
5	61	1927–1988
6	42	1927–1968
7	63	1921–1988
8	41	1935–1976
9	57	1932–1988
10	58	1912–1978
11	58	1914–1974

Table 2.1: Table showing the number of years of data available and the span of the data for each site

Maximum likelihood techniques are applied to fit the GEV distribution (1.2) to the annual maxima from each site. Let $Z_{1j}, \ldots, Z_{m_j j}$ be the annual maxima for site j, with m_j denoting the number of annual maxima at site j. These are assumed to be independent GEV random variables with parameters μ_j, σ_j and ξ_j for site j. The log-likelihood when $\xi_j \neq 0$ is

$$l(\mu_j, \sigma_j, \xi_j) = -m_j \log \sigma_j - (1 + 1/\xi_j) \times \sum_{i=1}^{m_j} \log \left[1 + \xi_j \left(\frac{z_{ij} - \mu_j}{\sigma_j} \right) \right] - \sum_{i=1}^{m_j} \left[1 + \xi_j \left(\frac{z_{ij} - \mu_j}{\sigma_j} \right) \right]^{-1/\xi_j}, \quad (2.1)$$

provided that

$$1 + \xi_j \left(\frac{z_{ij} - \mu_j}{\sigma_j} \right) > 0, \text{ for } i = 1, \ldots, m_j. \quad (2.2)$$

If (2.2) does not hold, the likelihood is zero. This corresponds to combinations of parameters for which one or more of the observed data fall beyond one of the distribution's

endpoints. For the case where $\xi_j = 0$, the log-likelihood is

$$l(\mu_j, \sigma_j) = -m_j \log \sigma_j - \sum_{i=1}^{m_j} \left(\frac{z_{ij} - \mu_j}{\sigma_j} \right) - \sum_{i=1}^{m_j} \exp \left\{ - \left(\frac{z_{ij} - \mu_j}{\sigma_j} \right) \right\}, \qquad (2.3)$$

which comes from the Gumbel limit of the GEV distribution. Maximum likelihood estimates of the parameters μ_j, σ_j and ξ_j are obtained by maximising (2.1) and (2.3) with respect to the parameter vector (μ_j, σ_j, ξ_j) for $j = 1, \ldots, 11$. The maximum likelihood estimates and standard errors are given in Table 2.2. These maximum likelihood estimates

j	$\hat{\mu}_j$ (s.e.)	$\hat{\sigma}_j$ (s.e.)	$\hat{\xi}_j$ (s.e.)
1	54.558 (2.242)	14.883 (1.680)	0.0860 (0.103)
2	66.709 (1.641)	11.930 (1.263)	0.169 (0.0852)
3	59.365 (1.672)	8.889 (1.302)	0.170 (0.124)
4	66.630 (2.037)	13.449 (1.533)	0.0121 (0.124)
5	61.948 (1.912)	12.692 (1.461)	0.0641 (0.127)
6	59.116 (1.527)	8.893 (1.356)	0.385 (0.123)
7	44.167 (1.060)	7.604 (0.776)	0.0661 (0.0797)
8	46.673 (1.438)	8.321 (1.078)	0.128 (0.0991)
9	43.183 (1.284)	8.536 (1.057)	0.265 (0.112)
10	39.744 (1.516)	10.330 (1.440)	0.480 (0.116)
11	39.899 (1.348)	9.163 (1.042)	0.155 (0.0995)

Table 2.2: Maximum likelihood estimates (standard errors) for the GEV parameters for each site

are then substituted into (1.4) for $p = 0.1, 0.01, 0.001$ to give the maximum likelihood estimates of the 10, 100 and 1000-year return levels for each site. These estimates are given in Table 2.3 with corresponding standard errors, obtained using the delta method.

In Chapter 3 Bayesian models, with both non-informative and informative priors, are applied to the same datasets. Posterior means and standard deviations of the GEV parameters and of the return levels are obtained and are compared with the maximum

Site	Return period (years)		
	10	100	1000
1	91.512 (6.117)	138.547 (21.446)	194.968 (58.873)
2	99.381 (5.400)	149.755 (20.535)	223.077 (57.329)
3	83.746 (5.603)	121.451 (21.952)	176.499 (62.039)
4	97.310 (4.610)	130.252 (16.230)	163.520 (38.199)
5	92.671 (4.912)	129.847 (18.951)	172.217 (48.108)
6	90.965 (8.234)	171.870 (47.550)	366.462 (197.611)
7	62.618 (2.732)	85.052 (8.471)	110.744 (19.960)
8	68.370 (4.250)	98.780 (14.592)	138.984 (37.722)
9	69.549 (5.249)	120.032 (25.456)	212.075 (86.505)
10	81.622 (10.220)	214.164 (72.377)	611.562 (365.188)
11	64.562 (4.327)	101.328 (16.740)	153.061 (46.452)

Table 2.3: Maximum likelihood estimates (standard errors) for the 10, 100 and 1000-year return levels (mm) at each site based on the annual maxima method

likelihood estimates and standard errors obtained here.

2.3 Daily data

Modelling only the annual maxima data when more data are available is extremely wasteful. In Chapter 5 a Bayesian multivariate model is used to model the daily rainfall data at all 11 sites simultaneously. Here, the threshold model, described in Section 1.4, is used to model the daily rainfall data at each site. This provides a preliminary analysis of the daily rainfall data with which later results can be compared.

To obtain series of data which are stationary and independent, the approach taken by Coles (1993) is used. That is, only the data from November to February are used

to obtain a stationary series, and the maxima over two-day periods are used to obtain temporal independence. It would also have been possible to apply other techniques, such as those described in section 1.7.2, to cope with the seasonal variability. This would have enabled the use of more data, which would have been beneficial.

2.3.1 Threshold choice

After obtaining stationary and independent sets of data, suitable thresholds for each dataset need to be chosen. Observations above the chosen threshold are classed as extreme events, so the threshold must be high enough for the asymptotic result of Theorem 1.3 to hold. A threshold which is too low will, therefore, lead to bias. A threshold which is too high, however, will result in few excesses, leading to high variance. It is common practice to choose as low a threshold as possible such that the limit model still provides a reasonable approximation, but the maximum number of extremes are used.

One method for choosing a threshold is based on examination of the *mean residual life plot*. This is a plot of the mean number of excesses of a threshold u, against u, for a range of values of u. At levels of u for which the model is appropriate, the relationship should be linear. The threshold, then, is chosen as the value of u above which the plot is approximately linear in u. This method was used to choose the threshold for each of the datasets and the details for site 1 follow. Figure 2.2 gives the mean residual life plot for site 1 with approximate 95% confidence intervals. Taking the confidence intervals into account, it appears that the graph is approximately linear between $u \approx 50$ and $u \approx 90$. After this point there is a very sharp decline. For large values of u the plot is unreliable since the data on which it is based are limited. So, although it may seem that linearity is not achieved until $u \approx 90$, it is better to assume a value of $u = 50$ as an appropriate threshold, since there is some evidence for linearity above this point. The same procedure was used to obtain the thresholds for the other sites. The thresholds chosen for each site, the number of exceedances and the proportion of observations exceeding the thresholds are given in Table 2.4.

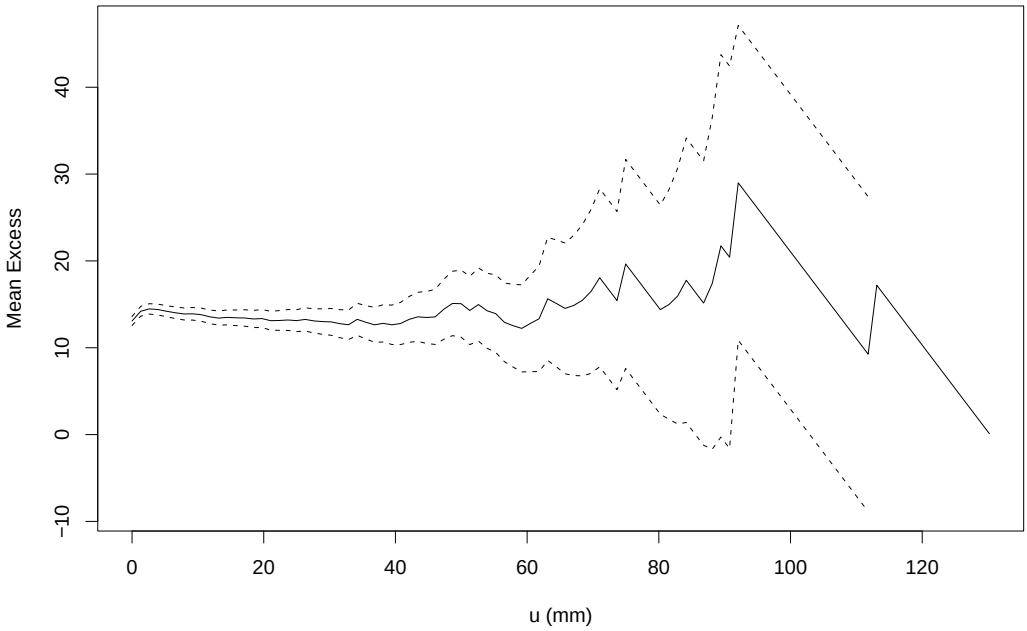


Figure 2.2: Mean residual life plot for the daily rainfall data from site 1

It can be difficult to choose a threshold based on the mean residual life plot. Another procedure is to fit the GPD at a range of thresholds and to look for stability of the parameter estimates. If the GPD is a suitable model with some threshold u_0 then it should also be suitable with any threshold $u > u_0$. For all thresholds $u > u_0$ the distributions would have the same shape parameters, but the scale parameters change with u for $\xi \neq 0$. Reparameterising the scale parameter as $\sigma^* = \sigma - \xi u$ results in a modified scale parameter σ^* which should be constant above u_0. So, for a suitable threshold u_0, estimates of σ^* and ξ should be constant above u_0, after allowing for sampling error. By plotting the estimates of σ^* and ξ against u, with corresponding confidence intervals, an appropriate threshold for the GPD can be chosen as the lowest value of u above which the estimates are approximately constant. Such plots for site 1 are given in Figure 2.3. The threshold of 50, chosen on the basis of the mean residual life plot, does seem appropriate, since above this point the parameter estimates can be assumed constant when taking into account the sampling error. Plots for the other sites were also examined to verify the threshold

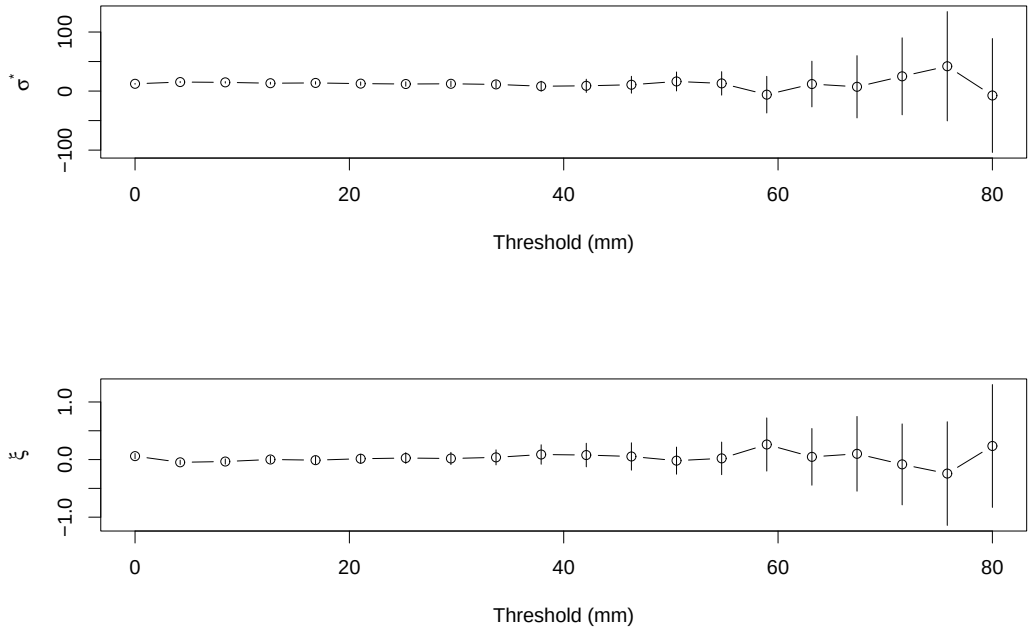


Figure 2.3: Parameter estimates (with confidence intervals) against threshold for the daily rainfall data from site 1

choices for each site.

2.3.2 Estimation

Having chosen suitable thresholds, maximum likelihood is used to fit the GPD to the daily data from each site. Suppose $y_{1,j}, \ldots, y_{k,j}$ are the k excesses of the threshold u_j for site j. The log-likelihood, derived from (1.5), is

$$l(\sigma_j, \xi_j) = -k_j \log \sigma_j - (1 + 1/\xi_j) \sum_{i=1}^{k_j} \log(1 + \xi_j y_{i,j}/\sigma_j),$$

if $(1+\sigma_j^{-1}\xi_j y_{i,j}) > 0$ for $i = 1, \ldots, k_j$; otherwise, $l(\sigma_j, \xi_j) = -\infty$. For $\xi_j = 0$ the likelihood is derived from (1.6) as

$$l(\sigma_j) = -k_j \log \sigma_j - \sigma_j^{-1} \sum_{i=1}^{k_j} y_{i,j}.$$

The log-likelihood is maximised to obtain maximum likelihood estimates of the parameters σ_j and ξ_j for each site $j = 1, \ldots, 11$. These estimates are given, with standard errors, in Table 2.4.

j	u_j	n_j	p_j	$\hat{\sigma_j}$	$\hat{\xi}_j$
1	50	56	0.0166	15.793 (2.795)	-0.0314 (0.166)
2	50	113	0.0287	15.615 (1.959)	0.0416 (0.0831)
3	35	135	0.0624	14.522 (1.868)	-0.128 (0.0738)
4	30	407	0.113	15.940 (1.122)	-0.0340 (0.0500)
5	40	190	0.0522	14.822 (1.568)	-0.00422 (0.0770)
6	35	156	0.0629	12.385 (1.373)	0.166 (0.0777)
7	20	495	0.129	10.884 (0.645)	-0.142 (0.0392)
8	30	121	0.0500	8.629 (1.091)	-0.0302 (0.0878)
9	35	88	0.0261	6.772 (1.068)	0.186 (0.118)
10	30	104	0.0288	6.529 (0.982)	0.447 (0.122)
11	20	325	0.0928	7.134 (0.489)	0.107 (0.0616)

Table 2.4: Chosen thresholds u_j for site j, with n_j exceedances. The proportion of observations for site j exceeding the threshold u_j is denoted by p_j. Maximum likelihood estimates of the GPD parameters are denoted by $\hat{\sigma_j}$ and $\hat{\xi}_j$ for site j.

Estimates of the N-year return level for $N = 10, 100, 1000$ are obtained by substituting these estimates for σ_j and ξ_j into (1.7) for $p = 0.1, 0.01, 0.001$, where $m = Nn_y$ and n_y is the number of observations per year. These estimates are given, with standard errors, in Table 2.5. These return levels differ quite significantly from the equivalent return levels obtained using the annual maxima data (Table 2.3). This emphasises the importance of using as much of the available data as possible. All return level estimates based on the threshold method, apart from the 10-year return level estimate for site 6, are lower than the equivalent estimate based on the annual maxima method. The standard deviations

Site	Return period (years)		
	10	100	1000
1	84.983 (4.941)	117.627 (13.005)	147.997 (28.981)
2	97.149 (5.280)	139.622 (15.631)	186.365 (36.276)
3	77.074 (3.594)	95.258 (7.823)	108.789 (13.427)
4	92.593 (4.436)	123.184 (10.797)	151.471 (20.050)
5	90.682 (4.716)	124.156 (13.525)	157.306 (28.387)
6	96.668 (8.265)	160.050 (28.319)	252.887 (74.508)
7	55.340 (1.754)	66.885 (3.525)	75.216 (5.596)
8	57.888 (2.995)	75.208 (7.994)	91.364 (16.091)
9	59.325 (3.784)	91.752 (15.811)	141.472 (48.010)
10	67.541 (7.537)	161.231 (50.028)	423.199 (241.910)
11	55.825 (3.567)	84.457 (11.107)	121.086 (26.040)

Table 2.5: Maximum likelihood estimates (standard errors) for the 10, 100 and 1000-year return levels (mm) at each site based on the threshold method

are also lower when the threshold method is used, with the exception of the standard error for the 10-year return level for site 6. This reflects the benefit of using the threshold method. The difference between the estimates and standard errors based on the two models is greater for the longer return periods. For site 10 the 1000-year return level estimates obtained using both methods are extremely high, as are the standard errors. It is likely that this is due to the very limited amount of data available at high levels.

2.4 Discussion

This chapter has demonstrated the use of the block maxima method and the threshold method for modelling univariate rainfall data. Maximum likelihood techniques have been used to obtain maximum likelihood estimates of the model parameters, which have then

been used to obtain maximum likelihood estimates of the 10, 100 and 1000-year return levels. The same techniques have been applied to the data from each of the eleven locations. This analysis has provided a preliminary analysis of the data considered throughout the thesis.

To model the daily data the issues of temporal dependence and seasonal variation are addressed in a very simple way. Following Coles (1993), the analyses were restricted to the data from November to February and the maxima over two-day periods. This resulted in approximately stationary, independent series. This technique was used for consistency with the technique used in later chapters, where it was also used, for simplicity. Other techniques, such as those described in Section 1.7.2 would, however, have been better for addressing the issue of non-stationarity. Restricting the analyses to the data from November to February resulted in much smaller datasets and therefore in fewer threshold exceedances to be modelled. For most of the sites, though, the threshold method used considerably more data than the annual maxima method.

Although taking the maxima over two-day periods was found to give series which are independent to a good approximation, it is still possible that the resulting seriescould contain consecutive values which have come from the same storm. It may, then, be worth investigating the sensitivity of the results to the assumption of independence for the two-day maxima.

Appropriate thresholds for the the threshold method were chosen through examination of mean residual life plots and plots of a modified scale parameter and the shape parameter obtained for a range of thresholds.

The return level estimates using the threshold method were found to be lower than the equivalent return level estimates obtained with the annual maxima method. The standard errors were also found to be lower with the threshold method, reflecting the importance of using as much of the available data as possible. One exception was with the 10-year return level for site 6, where the return level estimate and standard error were both slightly higher with the threshold method than with the annual maxima method.

The parameter and return level estimates obtained in this chapter will be compared with those obtained later in the thesis.

Chapter 3

Bayesian modelling

3.1 Introduction

Inference on the extremes of environmental processes is essential for design specification in civil engineering. Naturally, data at extreme levels are scarce. Using Bayesian inference in these problems would allow any additional information about the processes to be incorporated as prior information. The benefits of using any information available are likely to be great, due to the lack of data; however, there is some concern that it may not be possible to formulate such prior information. Coles and Powell (1996) comment that if the data on the extremes are so scarce, then it may not be possible for an expert to independently formulate prior beliefs about the process. Some work has been carried out using the Bayesian approach in univariate extreme value problems but so far, multivariate problems have not received any significant consideration.

This chapter introduces the idea of Bayesian modelling with the use of Markov chain Monte Carlo (MCMC) techniques. Section 3.1.1 outlines the basic theory behind Bayesian modelling and Section 3.2 gives details of two MCMC techniques which are used in the remainder of this thesis. The MCMC techniqes are then applied to give Bayesian analyses of the annual maxima rainfall data at each site, using non-informative priors (in Section 3.3) and informative priors (in Section 3.4). The informative priors are based on infor-

mation provided by a hydrologist. In Section 3.5, the effect the informative prior has on posterior inference is considered, by comparing the results obtained from the analysis with the non-informative prior and those obtained from the analysis with the informative prior. The posterior inference is taken further, in Section 3.6, by considering the predictive distribution of return levels based on both the non-informative prior and the informative prior. The results of this chapter are then discussed in Section 3.7.

3.1.1 Basic theory

Suppose the data $\boldsymbol{x} = (x_1, \ldots, x_n)$ are realizations of a random variable with a density from the parametric family $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta\}$. Also, suppose that prior beliefs about θ can be formulated and expressed by a probability density function $\pi(\theta)$ with no reference to the data. The likelihood for θ is

$$\begin{aligned} L(\theta|\boldsymbol{x}) &= f(\boldsymbol{x}|\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta). \end{aligned}$$

The prior information and the likelihood can be combined using Bayes' theorem to give a *posterior* distribution for θ as follows:

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &= \frac{\pi(\theta)L(\theta|\boldsymbol{x})}{f(\boldsymbol{x})} \\ &= \frac{\pi(\theta)L(\theta|\boldsymbol{x})}{\int_\Theta \pi(\theta)L(\theta|\boldsymbol{x})d\theta} \end{aligned} \tag{3.1}$$

This distribution expresses the beliefs about θ after observing the data and can be rewritten as

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) \times L(\theta|\boldsymbol{x}) \\ \text{i.e. posterior} &\propto \text{prior} \times \text{likelihood} \end{aligned}$$

Often, the function of an extreme value analysis is to describe the extremal behaviour of an observed process in order to find the probability of extreme events occurring in the future. Within the Bayesian framework, prediction is possible through the *predictive*

distribution, which describes how likely different outcomes of a future experiment are. Let y denote a future observation with probability density function $f(y|\theta)$, and $\pi(\theta, \boldsymbol{x})$ is the posterior distribution of θ given the data $\boldsymbol{x}$. Then,

$$f(y|\boldsymbol{x}) = \int_{\Theta} f(y|\theta)\pi(\theta|\boldsymbol{x})d\theta \tag{3.2}$$

is the predictive distribution of y given $\boldsymbol{x}$. So, if a suitable prior distribution can be specified, there are good reasons to choose Bayesian procedures. One reason for rejecting these Bayesian procedures is the difficulty in computing the integral in (3.1). This problem can be overcome by using simulation based techniques such as Markov chain Monte Carlo (MCMC) to simulate realisations of the posterior distribution. Estimates of the posterior distribution could then be obtained from the simulated sample.

3.2 MCMC techniques

MCMC techniques provide a way of simulating from complex distributions by simulating from Markov chains which have the target distributions as their stationary distributions. There are many MCMC techniques of which two are described below. The specific details of how these techniques are used will be given later. There is much literature available on the theory behind MCMC techniques and on applications of the techniques. Introductions to the area are provided by Besag *et al.* (1995), Gamerman (1997) and Besag (2001).

3.2.1 The Gibbs sampler

The Gibbs sampler was used by Geman and Geman (1984) for models with the Gibbs distribution and was extended to the general form given here by Gelfand and Smith (1990). The Gibbs sampler enables simulation from multivariate distributions by simulating only from the conditional distributions. So, suppose the density of interest is $\pi(\theta)$, where $\theta = (\theta_1, \ldots, \theta_d)'$ and the full conditionals are given by

$$\pi(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d) = \pi(\theta_i|\theta_{-i}), \qquad i = 1, \ldots, d.$$

If it is possible to simulate from the full conditionals then the Gibbs sampler can be used by using the following algorithm:

1. Initialise the counter to $j = 1$ and the state of the chain to $\theta^{(0)} = \left(\theta_1^{(0)}, \ldots, \theta_d^{(0)}\right)'$.

2. Obtain a new value $\theta^{(j)}$ from $\theta^{(j-1)}$ by successive simulation from the full conditionals

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi\left(\theta_1 | \theta_2^{(j-1)}, \ldots, \theta_d^{(j-1)}\right) \\
\theta_2^{(j)} &\sim \pi\left(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_d^{(j-1)}\right) \\
&\vdots \\
\theta_d^{(j)} &\sim \pi\left(\theta_d | \theta_1^{(j)}, \ldots, \theta_{d-1}^{(j)}\right).
\end{aligned}$$

3. Increase counter from j to $j + 1$ and return to step 2.

If it is possible to simulate from the full conditionals of the posterior distribution (3.1), then it is also possible to simulate from the posterior itself. The Gibbs sampler should be run after initialising the sampler somewhere in the support of θ. The resulting chain will converge, after an initial "burn-in" period, to the posterior distribution.

3.2.2 Metropolis-Hastings sampling

The Gibbs sampler provides a way of simulating from multivariate distributions provided that the full conditional distributions can be simulated from. It may not be straightforward to simulate from these full conditionals but Metropolis-Hastings schemes provide a way. These schemes come from work by Metropolis *et al.* (1953) and Hastings (1970). Given a distribution of interest, π, a reversible Markov chain, which has this distribution as its stationary distribution, can be constructed. Simulating from such a Markov chain will result in values from the distribution of interest.

The procedure is to construct a transition kernel $p(\theta, \phi)$ such that the equilibrium distribution of the chain is π. This transition kernel is made up of two elements; an arbitrary transition kernel $q(\theta, \phi)$ also known as the *proposal distribution*, and an acceptance

probability $a(\theta, \phi)$. The acceptance probability

$$a(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi) q(\phi, \theta)}{\pi(\theta) q(\theta, \phi)} \right\} \tag{3.3}$$

was suggested by Hastings (1970). The algorithm below can then be followed to obtain a chain with limiting distribution π.

1. Initialise the counter to $j = 1$ and the chain to $\theta^{(0)}$.
2. Simulate a proposed value ϕ using the kernel $q\left(\theta^{(j-1)}, \phi\right)$.
3. Find the acceptance probability of the proposed value $a\left(\theta^{(j-1)}, \phi\right)$.
4. Accept $\theta^{(j)} = \phi$ with probability $a\left(\theta^{(j-1)}, \phi\right)$ and take $\theta^{(j)} = \theta^{(j-1)}$ otherwise.
5. Increase the counter from j to $j + 1$ and return to step 2.

The particular type of MCMC method used in this thesis is based on simulation of a *random walk* chain. Here, the proposed value ϕ at point j is $\phi = \theta^{(j-1)} + \omega_j$. The ω_j are IID random variables and have density $f(\cdot)$. Supposing $f(\cdot)$ is easy to simulate from, an *innovation*, ω_j, can be simulated. The *candidate* point is then set to $\phi = \theta^{(j-1)} + \omega_j$ and the transition kernel is given by $q(\theta, \phi) = f(\phi - \theta)$. This is then used to calculate the acceptance probability. The variance of the innovation affects the acceptance probability: if the variance is too low most proposals will be accepted, resulting in very slow convergence, and if it is too high very few will be accepted and the moves in the chain will often be large.

3.3 Bayesian analysis of the annual maxima rainfall data using non-informative priors

In this section the annual maxima data from each of the 11 sites are modelled using a Bayesian model with non-informative prior distributions. The priors are constructed by assuming there is no information available about the process apart from the data.

3.3.1 Prior specification

Although the description below is in terms of an analysis at one location, the same model is applied to all 11 locations. The annual maxima data have a GEV distribution i.e. $Z_i \sim GEV(\mu, \sigma, \xi)$, so the likelihood is

$$\begin{aligned} L(\mu, \sigma, \xi | \boldsymbol{z}) &= \prod_{i=1}^{n} \frac{1}{\sigma} \left[1 + \xi \frac{(z_i - \mu)}{\sigma}\right]^{-1/\xi - 1} \exp\left\{-\left[1 + \xi \frac{(z_i - \mu)}{\sigma}\right]^{-1/\xi}\right\} \\ &= \frac{1}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \left[1 + \xi \frac{(z_i - \mu)}{\sigma}\right]^{-1/\xi}\right\} \prod_{i=1}^{n} \left[1 + \xi \frac{(z_i - \mu)}{\sigma}\right]^{-1/\xi - 1} \end{aligned} \tag{3.4}$$

For specification of the prior, the parametrisation $\phi = \log \sigma$ is easier to work with since σ is constrained to be positive. The prior density was chosen to be

$$\pi(\mu, \phi, \xi) = \pi_\mu(\mu)\, \pi_\phi(\phi)\, \pi_\xi(\xi), \tag{3.5}$$

where the marginal priors, $\pi_\mu(\cdot), \pi_\phi(\cdot)$ and $\pi_\xi(\cdot)$, are

$$\begin{aligned} \mu &\sim N(0, 1000); \\ \phi &\sim N(0, 100); \\ \xi &\sim N(0, 10). \end{aligned}$$

These are independent normal priors with large variances. The variances are chosen large enough to make the distributions almost flat, corresponding to prior ignorance.

3.3.2 MCMC algorithm

The density of interest is the posterior

$$\pi(\mu, \sigma, \xi | \boldsymbol{z}) \propto \pi(\mu, \phi, \xi)\, L(\mu, \phi, \xi | \boldsymbol{z}),$$

with the prior given in (3.5) and the likelihood in (3.4) with σ replaced by e^ϕ. Since this is a multivariate density a Gibbs sampler is used, and a Metropolis step with random walk updates is used to simulate from each of the full conditionals. The posterior is of the form

$$\pi(\mu, \phi, \xi)\, L(\mu, \phi, \xi | \boldsymbol{z}) = \pi_\mu(\mu)\, \pi_\phi(\phi)\, \pi_\xi(\xi)\, L(\mu, \phi, \xi | \boldsymbol{z}),$$

so the full conditionals are of the form

$$\begin{aligned}\pi(\mu|\phi,\xi) &= \pi_\mu(\mu)\, L(\mu,\phi,\xi|\boldsymbol{z}),\\ \pi(\phi|\mu,\xi) &= \pi_\phi(\phi)\, L(\mu,\phi,\xi|\boldsymbol{z}),\\ \pi(\xi|\mu,\phi) &= \pi_\xi(\xi)\, L(\mu,\phi,\xi|\boldsymbol{z}).\end{aligned}$$

The three transition densities are denoted by q_μ, q_ϕ and q_ξ, and the proposed values for each variable are given by

$$\begin{aligned}\mu^* &= \mu + \omega_\mu,\\ \phi^* &= \phi + \omega_\phi,\\ \xi^* &= \xi + \omega_\xi,\end{aligned}$$

where ω_μ, ω_ϕ and ω_ξ are normally distributed with zero means, and variances $\epsilon_\mu = 4$, $\epsilon_\phi = 0.3$ and $\epsilon_\xi = 0.1$ respectively. The variances were chosen to give acceptance rates of approximately 30%

The full details of the algorithm are as follows:

1. Initialise the chain at $\theta^{(0)} = \left(\mu^{(0)}, \phi^{(0)}, \xi^{(0)}\right)$ and the counter at $j = 1$.

2. Simulate $\omega_\mu \sim N(0, 4)$.

3. Put $\mu^* = \mu^{(j-1)} + \omega_\mu$.

4. Accept $\mu^{(j)} = \mu^*$ with probability $a\left(\mu^{(j-1)}, \mu^*\right) = \min\{1, A\}$ where

$$A = \frac{\pi\left(\mu^*|\phi^{(j-1)}, \xi^{(j-1)}\right) q_\mu\left(\mu^{(j-1)}, \mu^*\right)}{\pi\left(\mu^{(j-1)}|\phi^{(j-1)}, \xi^{(j-1)}\right) q_\mu\left(\mu^*, \mu^{(j-1)}\right)}$$

and $\mu^{(j)} = \mu^{(j-1)}$ otherwise. Now $q_\mu\left(\mu^{(j-1)}, \mu^*\right) = f_\mu\left(\mu^{(j-1)} - \mu*\right)$, where $f_\mu(\cdot)$ is the density function of ω_μ. Since the distribution of ω_μ is symmetric about zero $q_\mu\left(\mu^{(j-1)}, \mu^*\right) = q_\mu\left(\mu^*, \mu^{(j-1)}\right)$, so

$$A = \frac{\pi\left(\mu^*|\phi^{(j-1)}, \xi^{(j-1)}\right)}{\pi\left(\mu^{(j-1)}|\phi^{(j-1)}, \xi^{(j-1)}\right)}.$$

5. Simulate $\omega_\phi \sim N(0, 0.3)$.

6. Put $\phi^* = \phi^{(j-1)} + \omega_\phi$.

7. Accept $\phi^{(j)} = \phi^*$ with probability $a\left(\phi^{(j-1)}, \phi^*\right) = \min\{1, A\}$ where

$$A = \frac{\pi\left(\phi^* | \mu^{(j)}, \xi^{(j-1)}\right)}{\pi\left(\phi^{(j-1)} | \mu^{(j)}, \xi^{(j-1)}\right)}$$

   and $\phi^{(j)} = \phi^{(j-1)}$ otherwise.

8. Simulate $\omega_\xi \sim N(0, 0.1)$.

9. Put $\xi^* = \xi^{(j-1)} + \omega_\xi$.

10. Accept $\xi^{(j)} = \xi^*$ with probability $a\left(\xi^{(j-1)}, \xi^*\right) = \min\{1, A\}$ where

$$A = \frac{\pi\left(\xi^* | \mu^{(j)}, \phi^{(j)}\right)}{\pi\left(\xi^{(j-1)} | \mu^{(j)}, \phi^{(j)}\right)}$$

    and $\xi^{(j)} = \xi^{(j-1)}$ otherwise.

11. Increase counter from j to $j + 1$ and return to step 2.

3.3.3 Application

The algorithm in 3.3.2 was applied to all 11 annual maxima datasets. In each case 30000 iterations of the algorithm were carried out, of which the first 10000 were discarded. The chains had all converged well before this point (See Figure 3.1). The remaining 20000 simulations can then be regarded as realizations of the marginal distributions of the posterior. The MCMC trace plots and estimated posterior densities for the GEV parameters for site 1 are given in Figure 3.1. To check that the chains had converged to the correct place, the same algorithm was carried out using different starting points. The chains all converged to the same places. The chains for the other 10 sites also converged very well within the first 10000 iterations.

The posterior means and standard deviations for the GEV parameters for each site are given in Table 3.1. The posterior means are very close to the maximum likelihood

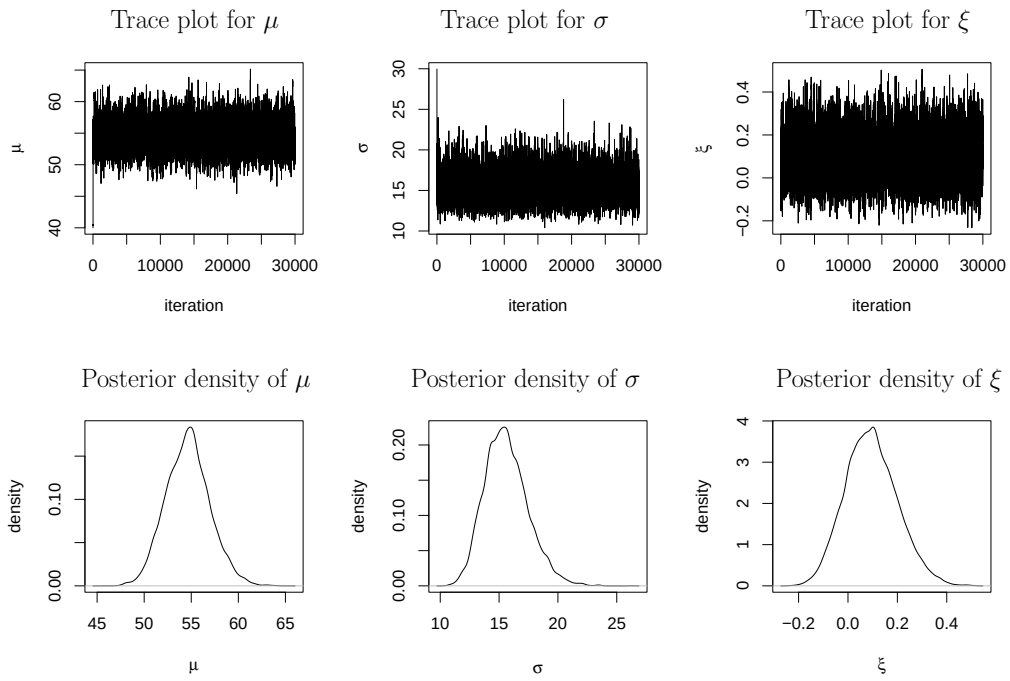


Figure 3.1: Trace plots and posterior densities of the GEV parameters using non-informative priors for site 1

estimates of the GEV parameters given in Table 2.2 for all sites apart from site 10. It is to be expected that posterior means would be close to the maximum likelihood estimates, since the priors were almost flat and added very little information to the likelihood. For site 10, however, there is a noticeable difference between the posterior means obtained here and the maximum likelihood estimates; the difference is particularly large for the shape parameter ξ.

By substituting the vectors of observations from the marginal posterior distributions of μ, σ and ξ into (1.4), for $0 < p < 1$, samples from the posterior distribution of return levels can be obtained. This procedure was carried out for $p = 0.1, 0.01, 0.001$ to obtain the posterior distributions of the 10, 100 and 1000-year return levels. Plots of the posterior densities for site 1 are given in Figure 3.2.

Due to the positive skew of the posterior distributions, seen in Figure 3.2, the posterior medians are considered to be more suitable measures of location than the posterior means.

Site	μ	σ	ξ
1	54.570 (2.357)	15.605 (1.839)	0.0982 (0.105)
2	66.766 (1.729)	12.480 (1.389)	0.184 (0.0897)
3	59.468 (1.828)	9.707 (1.576)	0.177 (0.136)
4	66.646 (2.104)	14.057 (1.659)	0.0253 (0.123)
5	62.023 (1.955)	13.292 (1.581)	0.0670 (0.129)
6	59.272 (1.651)	9.586 (1.589)	0.406 (0.128)
7	44.171 (1.084)	7.922 (0.843)	0.0830 (0.0810)
8	46.625 (1.538)	8.846 (12.490)	0.159 (0.106)
9	43.223 (1.351)	8.971 (1.167)	0.284 (0.114)
10	37.633 (1.278)	8.930 (0.948)	0.0578 (0.0828)
11	39.962 (1.422)	9.642 (1.146)	0.161 (0.102)

Table 3.1: Posterior means (standard deviations) for the GEV parameters for each site using non-informative priors

The same was true for all of the sites. Posterior medians and 95% credibility intervals for the three return levels for each site are given in Table 3.2.

With the exception of site 10, the posterior medians of the return levels are all close to, but also higher than, the maximum likelihood estimates given in Table 2.3. The posterior medians for site 10 are all considerably lower than the maximum likelihood estimates. This is due to the large difference between the posterior mean of ξ_{10} and the maximum likelihood estimates of ξ_{10}. In fact, the maximum likelihood estimate of the return levels for site 10 are not even contained within the 95% credibility intervals. The posterior medians for site 10, particularly for the 100 and 1000-year return levels, seem to be more sensible estimates for the return levels than the maximum likelihood estimates.

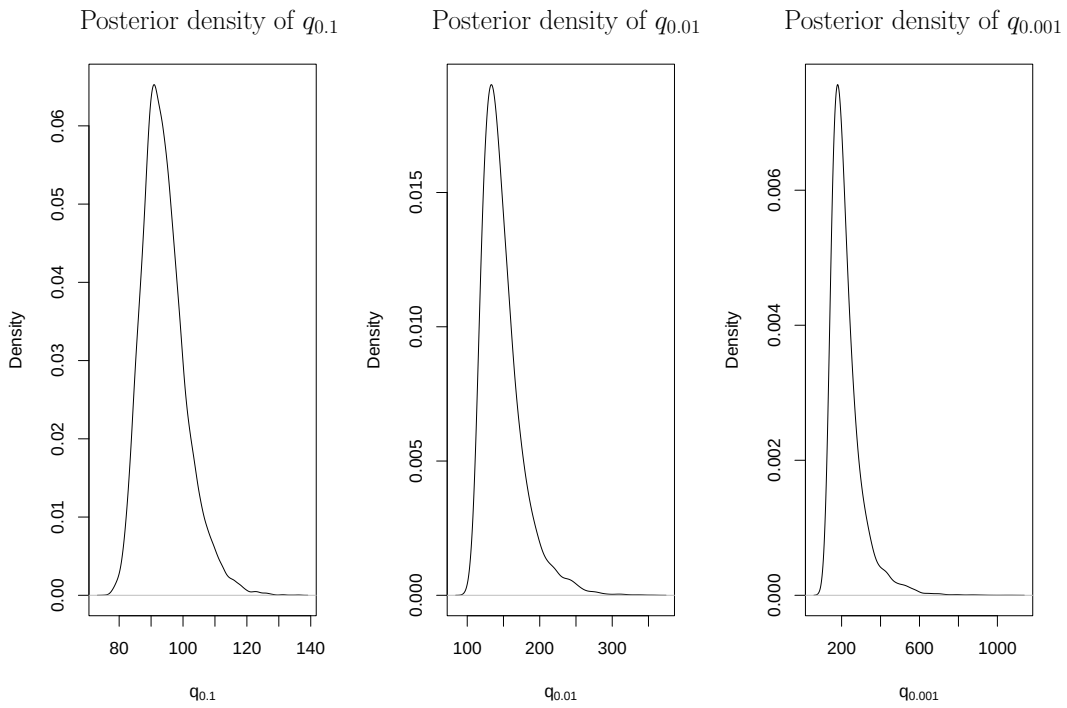


Figure 3.2: Posterior densities of the 10, 100 and 1000-year return levels for site 1, using a non-informative prior

3.4 Bayesian analysis of the annual maxima rainfall data using informative priors

Coles and Tawn (1996a) carried out a sophisticated Bayesian analysis of extremes from one of the 11 locations in Figure 2.1. It seems that the location considered is location 9, since the data spans from 1932 to 1988 in both cases. The expert prior was based on information provided by a hydrologist, Duncan Reed, who was asked to express his beliefs about extreme quantiles in the distribution of annual maxima at that location. The procedure used by Coles and Tawn is described in 3.4.1, together with a correction for one of the elicited priors. The prior is then generalised in 3.4.2 to be used at any one of the 11 locations.

Site	Return period (years)		
	10	100	1000
1	93.137 (83.084,110.94)	141.950 (113.789,223.249)	201.034 (137.688,454.780)
2	100.882 (92.030,116.128)	153.932 (125.707,226.231)	232.147 (160.901,483.716)
3	85.445 (76.638,103.997)	124.216 (97.710,225.217)	180.070 (114.536,564.886)
4	98.522 (90.964,112.062)	132.402 (112.902,193.319)	167.353 (125.64,346.834)
5	93.720 (85.823,108.302)	130.990 (108.076,205.000)	173.614 (121.294,410.958)
6	93.087 (80.643,123.917)	179.724 (122.574,421.166)	393.193 (186.494,1800.500)
7	63.365 (58.660,71.108)	87.025 (75.228,117.338)	114.516 (89.519,199.700)
8	69.916 (62.612,84.116)	103.459 (83.426,169.216)	149.299 (103.687,376.302)
9	70.743 (62.529,87.034)	123.789 (91.928,231.043)	222.111 (126.595,706.058)
10	58.848 (53.492,67.480)	83.586 (70.761,117.036)	111.075 (84.552,200.969)
11	65.600 (58.554,77.715)	103.547 (81.592,163.158)	157.408 (102.775,361.587)

Table 3.2: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels (mm) at each site using non-informative priors

3.4.1 Prior elicitation

The $1-p$ quantile of the annual maximum distribution is given by

$$q_p = \begin{cases} \mu + \frac{\sigma}{\xi}\left[\{-\log(1-p)\}^{-\xi} - 1\right] & \text{if } \xi \neq 0 \\ \mu - \sigma \log y_p & \text{if } \xi = 0 \end{cases} \tag{3.6}$$

where $y_p = -\log(1-p)$. This is also known as the return level associated with the $\{-\log(1-p)\}^{-1}$-year return period. For the location in question, the information requested concerned the three quantities

$$\tilde{q}_1 = q_{p_1} \qquad \tilde{q}_2 = q_{p_2} - q_{p_1} \qquad \tilde{q}_3 = q_{p_3} - q_{p_2}, \tag{3.7}$$

where the quantiles chosen were $(p_1, p_2, p_3) = (0.1, 0.01, 0.001)$. The use of differences ensures the correct ordering of quantiles. It is assumed that the priors on these quantities

Quantile	Median (mm)	90% quantile (mm)	λ_i	ν_i
$\tilde{q}_1$	59	72	38.9	0.67
$\tilde{q}_2$	43	70	7.1	0.16
$\tilde{q}_3$	100	120	47.0	0.39

Table 3.3: Elicited prior medians and 90% quantiles for distributions of $\tilde{q}_i$ with associated gamma parameters for the prior distribution

Quantile	Median	90% quantile
$\tilde{q}_1$	58	70
$\tilde{q}_2$	42	67
$\tilde{q}_3$	120	144

Table 3.4: Medians and 90% quantiles of the distributions for $\tilde{q}_1$, $\tilde{q}_2$ and $\tilde{q}_3$

are independent, and take the form

$$\tilde{q}_i \sim \Gamma(\lambda_i, \nu_i); \quad i = 1, 2, 3. \tag{3.8}$$

The parameters λ_i and ν_i are determined by measures of location and variability in prior belief. Here the expert was asked to estimate the median and 90% quantile of each of the $\tilde{q}_i$. The elicited values, together with the solutions in terms of the gamma parameters found by Coles and Tawn are shown in Table 3.3.

After simulating 10^7 observations from the distributions of $\tilde{q}_1$, $\tilde{q}_2$ and $\tilde{q}_3$, the medians and 90%-quantiles were calculated and are given in Table 3.4. For the simulated data, the medians and 90% quantiles for $\tilde{q}_1$ and $\tilde{q}_2$ are approximately the same as those given by the expert. For $\tilde{q}_3$, however, there is a discrepancy. The median and 90% quantile for the simulated data are 120 and 144 respectively, whereas the values given by the expert are 100 and 120. Assuming that the expert'ss values for the median and 90% quantile are correct, the gamma parameters were recalculated using a normal approximation for the gamma distribution. Given the median and 90% quantile of a gamma distribution, the

gamma parameters can not be calculated directly. For the normal distribution the median is equal to the mean and the quantiles can be obtained through the $N(0,1)$ distribution. This enables the parameters of a normal distribution to be obtained from the given median and 90% quantile. The parameters of the gamma distribution can then be approximated. Suppose

$$\tilde{q}_3 \sim \Gamma(\lambda_3, \nu_3) \simeq N\left(\frac{\lambda_3}{\nu_3}, \frac{\lambda_3}{\nu_3^2}\right).$$

The median and 90% quantile for $\tilde{q}_3$ are 100 and 120 respectively, so

$$\frac{\lambda_3}{\nu_3} = 100 \tag{3.9}$$

and

$$\frac{\lambda_3}{\nu_3} + 1.2816\frac{\sqrt{\lambda_3}}{\nu_3} = 120. \tag{3.10}$$

Substituting (3.9) in (3.10) gives

$$\frac{12.816}{\sqrt{\nu_3}} = 20,$$

so $\nu_3 = 0.4106$ and $\lambda_3 = 100\nu_3 = 41.06$. Simulating from a $\Gamma(41, 0.41)$ distribution gives a median of 99 and a 90% quantile of 121. These values are very close to those given by the expert, suggesting that the gamma parameters for $\tilde{q}_3$ obtained by Coles and Tawn were incorrect. The new parameters are therefore assumed to be the correct parameters.

From (3.7) and (3.8), the joint prior for the q_{p_i} is found to be

$$\pi(q_{p_1}, q_{p_2}, q_{p_3}) \propto q_{p_1}^{\lambda_1 - 1} \exp(-\nu_1 q_{p_1}) \prod_{i=2}^{3} (q_{p_i} - q_{p_{i-1}})^{\lambda_i - 1} \exp\left\{-\nu_i (q_{p_i} - q_{p_{i-1}})\right\}, \tag{3.11}$$

on $0 \leq q_{p_1} \leq q_{p_2} \leq q_{p_3}$. Substituting the quantile expression (3.6) in (3.11), and multiplying by the Jacobian of the transformation $(q_{p_1}, q_{p_2}, q_{p_3}) \rightarrow (\mu, \sigma, \xi)$, leads directly to an expression for the prior in terms of the GEV parameters. The Jacobian is given by

$$\begin{aligned} J(\mu, \sigma, \xi) &= -\frac{\sigma}{\xi^2}\{[\log(1-p_2)\log(1-p_3)]^{-\xi}[\log(-\log(1-p_2)) - \log(-\log(1-p_3))] \\ &+ [\log(1-p_1)\log(1-p_3)]^{-\xi}[\log(-\log(1-p_3)) - \log(-\log(1-p_1))] \\ &+ [\log(1-p_1)\log(1-p_2)]^{-\xi}[\log(-\log(1-p_1)) - \log(-\log(1-p_2))]\}. \end{aligned} \tag{3.12}$$

To assess the coherence of the expert's prior specification, Coles and Tawn asked for prior medians and 90% quantiles of the 30, 300 and 3000 year return levels. They compared these values with the corresponding statistics found on the basis of the specified prior for $(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3)$. There was excellent agreement for the 30 and 300 year return levels but not for the 3000 year return level. This was attributed to a disbelief in the likelihood model at very extreme levels; however, the inconsistency could largely be due to the aforementioned error in the gamma parameters for $\tilde{q}_3$ in Coles and Tawn (1996a). Simulating from the prior distribution for (μ, σ, ξ) and then solving (3.6) for $p = 1/30, 1/300, 1/3000$, gives realizations from the distributions of the 30, 300 and 3000 year return levels. The medians and 90% quantiles of these distributions, together with those obtained by Coles and Tawn and the expert's values, are given in Table 3.5. For the 30 and 300 year return

	New		Coles & Tawn's		Expert's	
Return level	median	90% qu.	median	90% qu.	median	90% qu.
30 year	74.5	90.8	74.0	89.3	75	95
300 year	137.9	167.5	144.0	173.3	140	170
3000 year	290.7	349.9	350.7	419.3	260	300

Table 3.5: Medians and 90 % quantiles for the 30, 300 and 3000-year return levels using the new gamma parameters, Coles & Tawn's gamma parameters and the values specified by the expert

levels, both the new parameters and those used by Coles and Tawn lead to medians and 90% quantiles which are close to those given by the expert. For the 3000 year return level, however using Coles and Tawn's gamma parameters leads to a large discrepancy with the expert's values. With the new gamma parameters, this discrepancy is greatly reduced and the values are much closer to those given by the expert.

3.4.2 Priors for each site

The prior used by Coles and Tawn was for modelling the data of site 9 from the network in Figure 2.1. This prior was based on expert information which took into account features of the local terrain for that site, together with information about the mean rainfall at that location. The rainfall climate is relatively homogeneous over the small scale of the region from which the data have been obtained, so Coles and Tawn's prior forms the key component of the prior knowledge about rainfall extremes at all locations in Figure 2.1.

Here, the prior beliefs about annual maxima at each site are expressed by inflating the uncertainty about the $\tilde{q}_i$. The gamma priors are modified by inflating the variance by a factor c, while keeping the mean the same. So, $\tilde{q}_i \sim \Gamma\left(\frac{\lambda_i}{c}, \frac{\nu_i}{c}\right)$. To make an appropriate choice for c, the medians and 90% quantiles of the distributions of the $\tilde{q}_i$ for a range of values of c were considered. These values are given in Table 3.6. Plots of the densities of the $\tilde{q}_i$ for $c = 1, \ldots, 4$ were also considered and are given in Figure 3.3. For $\tilde{q}_1$ and $\tilde{q}_3$ the medians decrease as c increases but only by a small amount. For $\tilde{q}_2$ however, the median decreases more significantly. This is also true of the mode, as can be seen from the plots in Figure 3.3. For all of the $\tilde{q}_i$, the 90% quantiles increase quite significantly as c increases.

To see the effect of different values of c on the prior distributions of the GEV parameters, simulations of 60000 observations from the marginal priors for $c = 1, \ldots, 4$ were obtained. Plots of the estimated densities are given in Figure 3.4. Increasing c to 2 has quite a big effect on the distribution of the GEV parameters. The distributions are much flatter and wider. A value of $c = 3$ was selected here as a subjective choice for an appropriate reflection of the uncertainty associated with the $\tilde{q}_i$.

	$\tilde{q}_1$		$\tilde{q}_2$		$\tilde{q}_3$	
c	median	90% qu.	median	90% qu.	median	90% qu.
1	57.563	70.263	42.310	66.605	99.432	120.732
2	57.068	75.441	40.285	75.954	98.623	129.421
3	56.574	79.443	38.306	83.000	97.816	136.140
4	56.082	82.830	36.377	88.782	97.011	141.826
5	55.592	85.819	34.502	93.715	96.209	146.846
6	55.103	88.522	32.686	98.017	95.410	151.387
7	54.617	91.007	30.931	101.821	94.614	155.561
8	54.132	93.316	29.238	105.215	93.820	159.442
9	53.649	95.480	27.609	108.265	93.029	163.081
10	53.168	97.521	26.046	111.018	92.242	166.515

Table 3.6: Medians and 90 % quantiles of the distributions for the $\tilde{q}_i$s for a range of values of c

3.4.3 Application to annual maxima data

The joint prior for the GEV parameters is

$$\begin{aligned}\pi(\mu,\sigma,\xi) \quad \propto \quad & q_{p_1}^{\lambda_1/3-1}\exp\left(-\frac{\nu_1}{3}q_{p_1}\right) \\ & \times\prod_{i=2}^{3}(q_{p_i}-q_{p_{i-1}})^{\lambda_i/3-1}\exp\left\{-\frac{\nu_i}{3}\left(q_{p_i}-q_{p_{i-1}}\right)\right\}\times J\left(\mu,\sigma,\xi\right), \quad (3.13)\end{aligned}$$

where the q_{p_i} are replaced by the expression given in (3.6) and $J\left(\mu,\sigma,\xi\right)$ is as in (3.12). The posterior then is

$$\pi\left(\mu,\sigma,\xi|\boldsymbol{z}\right)\propto\pi\left(\mu,\sigma,\xi\right)L\left(\mu,\sigma,\xi|\boldsymbol{z}\right),$$

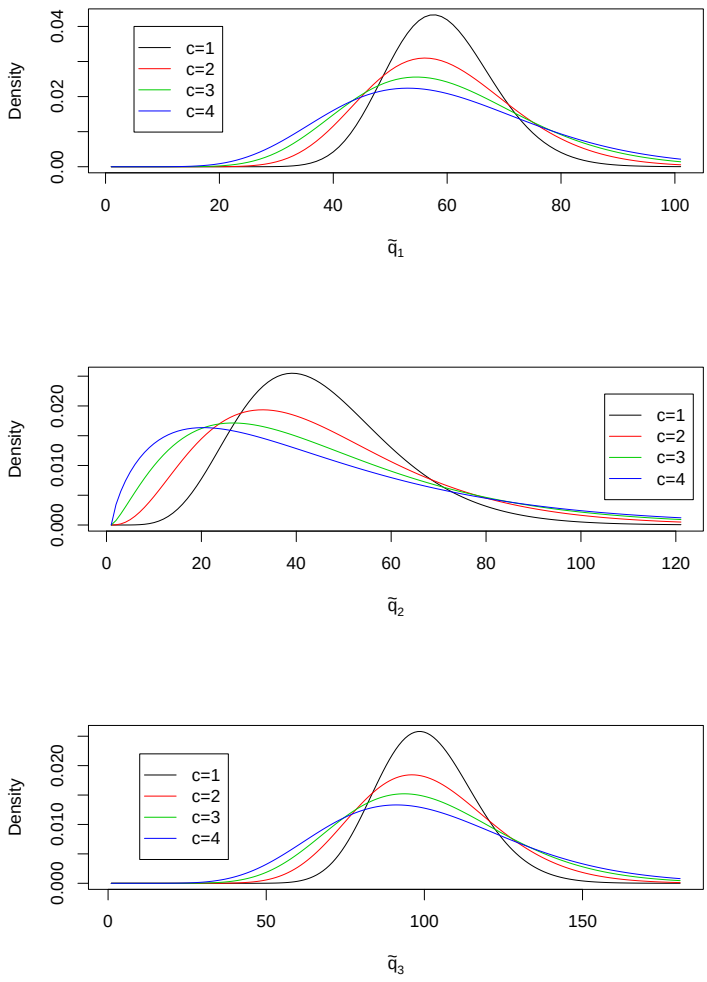


Figure 3.3: Plots of prior distributions of the $\tilde{q}_i$ for different values of c

where $L(\mu, \sigma, \xi|\boldsymbol{z})$ is given in (3.4). The full conditionals are:

$$\pi(\mu|\sigma,\xi) = q_{p_1}^{\lambda_1/3-1} \exp\left(-\frac{\nu_1}{3}q_{p_1}\right) \prod_{i=2}^{3} (q_{p_i} - q_{p_{i-1}})^{\lambda_i/3-1} \exp\left\{-\frac{\nu_i}{3}\left(q_{p_i} - q_{p_{i-1}}\right)\right\} \times L(\mu, \sigma, \xi|\boldsymbol{z}); \quad (3.14)$$

$$\pi(\sigma|\mu,\xi) = q_{p_1}^{\lambda_1/3-1} \exp\left(-\frac{\nu_1}{3}q_{p_1}\right) \prod_{i=2}^{3} (q_{p_i} - q_{p_{i-1}})^{\lambda_i/3-1} \exp\left\{-\frac{\nu_i}{3}\left(q_{p_i} - q_{p_{i-1}}\right)\right\} \times \sigma \times L(\mu, \sigma, \xi|\boldsymbol{z}); \quad (3.15)$$

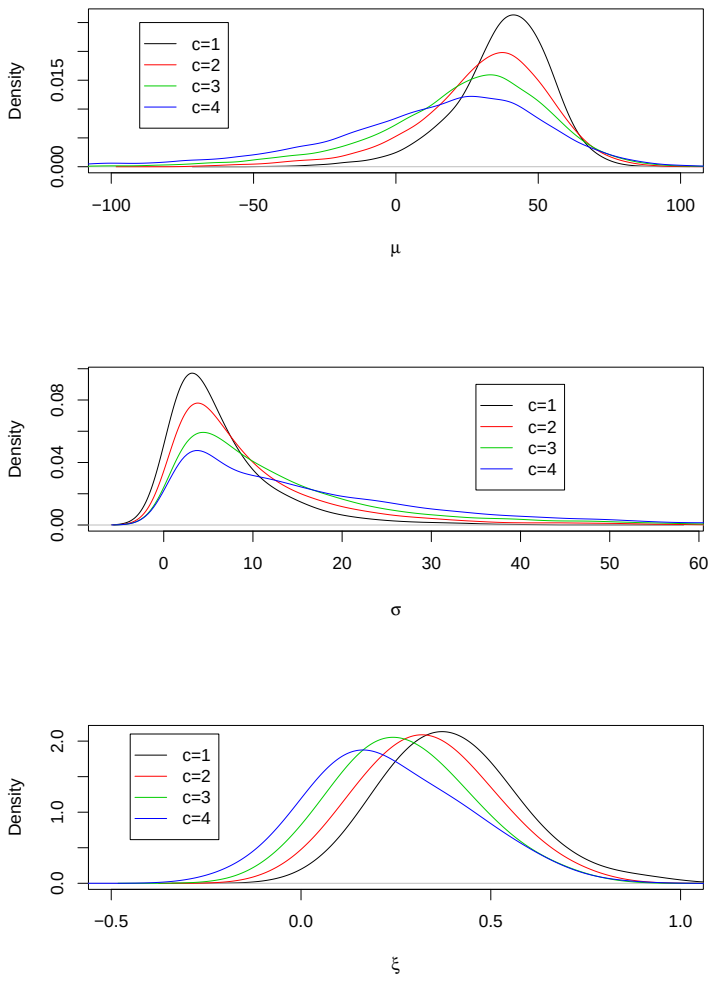


Figure 3.4: Plots of prior distributions of the GEV parameters for different values of c

and

$$\begin{aligned}\pi\left(\xi|\mu,\sigma\right) &= q_{p_1}^{\lambda_1/3-1}\exp\left(-\frac{\nu_1}{3}q_{p_1}\right)\prod_{i=2}^{3}(q_{p_i}-q_{p_{i-1}})^{\lambda_i/3-1}\exp\left\{-\frac{\nu_i}{3}\left(q_{p_i}-q_{p_{i-1}}\right)\right\}\\ &\quad\times\frac{1}{\xi^2}\{(\log(1-p_2)\log(1-p_3))^{-\xi}(\log(-\log(1-p_2))-\log(-\log(1-p_3)))\\ &\quad(\log(1-p_1)\log(1-p_3))^{-\xi}(\log(-\log(1-p_3))-\log(-\log(1-p_1)))\\ &\quad(\log(1-p_1)\log(1-p_2))^{-\xi}(\log(-\log(1-p_1))-\log(-\log(1-p_2)))\}\\ &\quad\times L\left(\mu,\sigma,\xi|\boldsymbol{z}\right). \qquad (3.16)\end{aligned}$$

The MCMC algorithm is the same as for the model using non-informative priors given in 3.3.2 but using the σ parametrisation and with $\epsilon_\mu = 5.0$, $\epsilon_\sigma = 2.0$ and $\epsilon_\xi = 0.1$. Again, 30000 iterations of the algorithm were carried out for each site. The first 10000 simulations were discarded in each case, after which the chains had converged well. To ensure convergence had occurred, the algorithm was carried out with different starting points: all chains converged to the same places. Trace plots and estimated posterior densities of the parameters for site 1 are given in Figure 3.5. The posterior means and standard deviations of the parameters for each site are given in Table 3.7.

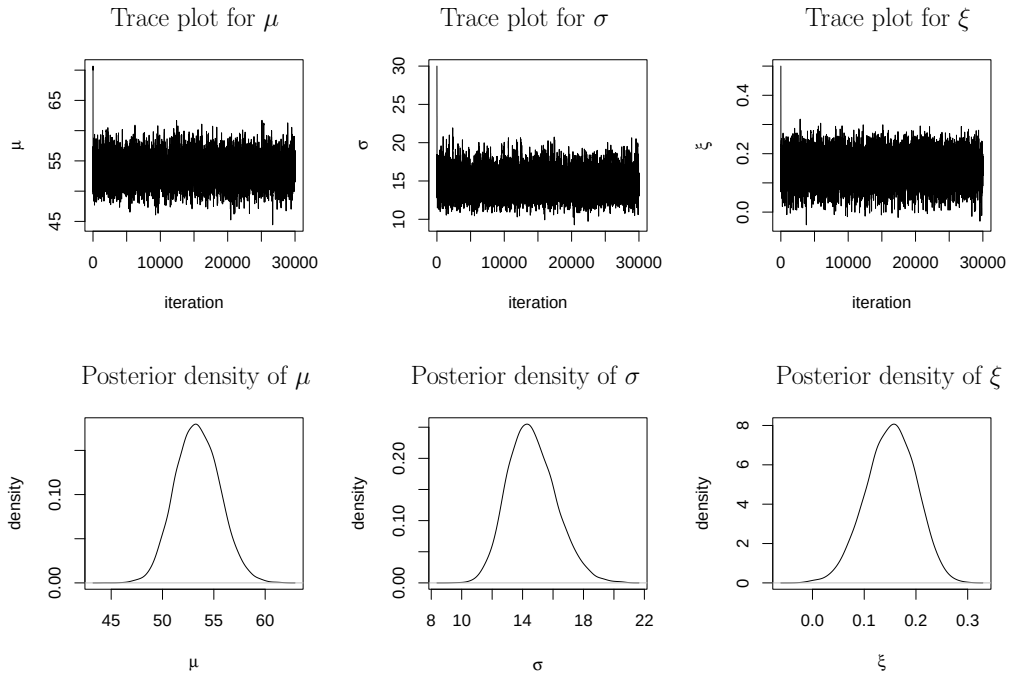


Figure 3.5: Trace plots and posterior densities of the GEV parameters using expert priors at site 1

In order to check that the expert prior had been implemented correctly, posterior densities of the GEV parameters for site 1 were obtained using the expert prior but with $c = 100$. This large value of c results in a large variance and therefore should correspond to prior ignorance. Plots of the posterior densities for site 1 with this prior and with the non-informative prior are given in Figure 3.6. For each parameter, the two densities are

Site	μ	σ	ξ
1	53.468 (2.182)	14.708 (1.610)	0.151 (0.0476)
2	66.198 (1.647)	11.914 (1.222)	0.195 (0.0446)
3	58.837 (1.715)	9.165 (1.278)	0.242 (0.0496)
4	65.177 (1.880)	13.038 (1.439)	0.165 (0.0510)
5	60.910 (1.798)	12.454 (1.389)	0.179 (0.0508)
6	58.892 (1.457)	8.517 (1.206)	0.283 (0.0451)
7	43.789 (1.106)	8.194 (0.811)	0.235 (0.0470)
8	46.251 (1.481)	8.794 (1.118)	0.243 (0.0487)
9	43.034 (1.231)	8.623 (1.036)	0.265 (0.0469)
10	37.238 (1.329)	9.344 (0.920)	0.212 (0.0461)
11	39.596 (1.356)	9.494 (1.043)	0.232 (0.0462)

Table 3.7: Posterior means (standard deviations) for the GEV parameters for each site using expert priors

very close, suggesting that the expert prior has been correctly implemented. The closeness of these two densitites also shows that posterior inference is not especially sensitive to the precise form chosen for the prior.

In the same way as for the model with non-informative priors, the posterior densities of the 10, 100 and 1000-year return levels were obtained. Posterior density plots of the return levels for site 1 are given in Figure 3.7. In contrast to the results for the model with non-informative priors, here there is very little difference between the posterior means and medians. The densities appear to be much more symmetric. The means are still higher than the medians and the difference between the two increases with return period. Even though the posterior densities do appear to be quite symmetric, the medians are considered here as the measures of location. These results can then be compared with those of the model with the non-informative prior. Posterior medians and 95% credibility intervals for all sites are given in Table 3.8.

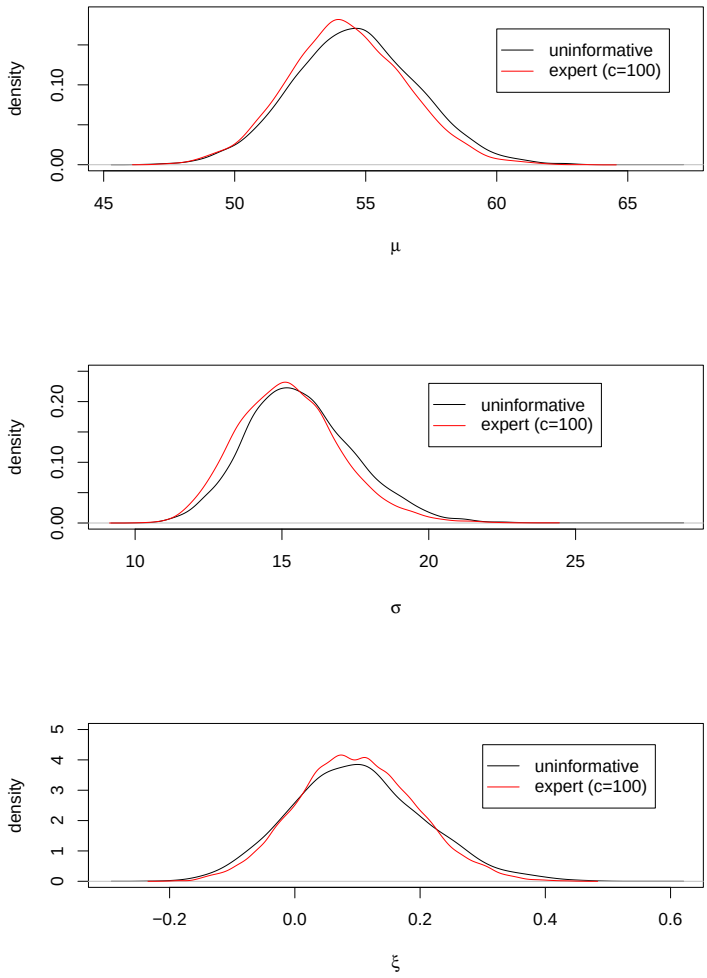


Figure 3.6: Posterior densities of the GEV parameters for site 1 using both non-informative priors and expert priors with c=100

3.5 The effect of the expert prior information

3.5.1 The effect on the GEV parameters

Density plots of simulated realisations from the non-informative and informative marginal priors are given in Figure 3.8. The informative prior for μ is more negatively skewed than the non-informative prior, but there is not much more certainty in the informative prior than in the non-informative prior. For both σ and ξ, the informative priors are much less

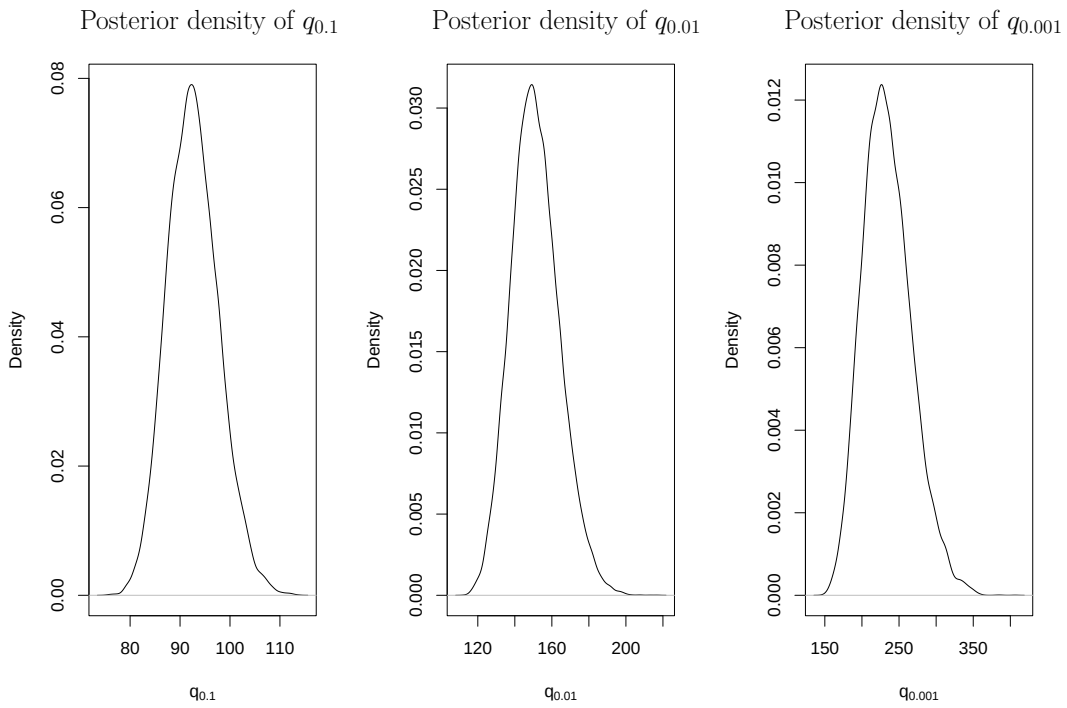


Figure 3.7: Posterior densities of the 10, 100 and 1000-year return levels for site 1 using the informative prior

flat than the non-informative priors. It seems, then, that the informative priors should have very little impact on μ and more impact on σ and ξ.

Figure 3.9 shows plots of the expert prior and posterior distributions for site 1. The priors are very flat compared to the posteriors, so the data has a much larger impact on the posterior than the prior does.

Figure 3.10 shows, for site 1, the estimated posterior densities for the GEV parameters using both the non-informative prior and the expert prior with $c = 3$. When using the expert prior, the posterior modes of μ and σ are lower and the distributions are slightly more concentrated than with the uninformative prior. The posterior mode of ξ is higher and the distribution is much more concentrated. So the expert prior clearly has an effect on the posterior distributions of the GEV parameters, with a greater effect on the shape parameter ξ. A similar effect was found for the other 10 sites. Comparing the posterior means and standard deviations obtained with the two priors also shows this effect. The

Site	Return period (years) 10	100	1000
1	92.344 (83.332,103.159)	150.322 (127.835,178.160)	231.767 (178.099,305.402)
2	99.432 (92.088,108.297)	154.384 (134.356,179.298)	239.698 (188.034,311.755)
3	85.751 (78.001,95.119)	135.436 (114.709,160.494)	221.757 (167.292,295.976)
4	100.282 (92.097,110.153)	154.372 (133.233,180.600)	233.451 (180.515,306.344)
5	94.794 (86.892,104.194)	149.038 (128.247,174.765)	230.753 (177.378,303.240)
6	85.296 (77.981,94.590)	138.720 (119.310,163.798)	240.651 (185.330,318.201)
7	67.822 (62.415,74.642)	111.391 (93.754,133.854)	185.299 (136.803,255.117)
8	72.289 (65.231,81.091)	120.232 (100.506,145.255)	204.001 (149.764,279.172)
9	69.392 (62.860,77.439)	120.459 (101.728,143.999)	213.599 (159.677,289.156)
10	64.024 (57.781,71.781)	109.848 (91.138,134.040)	183.852 (133.836,254.876)
11	67.446 (60.865,75.600)	117.287 (98.179,141.072)	201.426 (149.110,274.975)

Table 3.8: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels (mm) at each site using informative priors

posterior means of μ and σ are slightly lower when using the expert prior, with the exception of σ for sites 7 and 10. All of the posterior standard deviations, apart from those of μ for sites 7 and 10, are slightly reduced when the informative prior is used, suggesting reduced uncertainty. The posterior densities of μ for sites 7 and 10, with non-informative and informative priors, are very similar to those for site 1, given in Figure 3.10, but at lower values of μ. The informative prior has made the posterior density shift slightly to the left, but has not decreased the uncertainty. It seems that, due to the high level of uncertainty in the informative marginal prior for μ (seen in Figure 3.8), the posterior standard deviation based on the informative prior has remained very close to that based on the non-informative prior. The effect of the informative prior on ξ is much greater than for the other two parameters. In most cases the posterior mean of ξ is higher when using the expert prior, but for sites 6 and 9 it is lower. The posterior standard

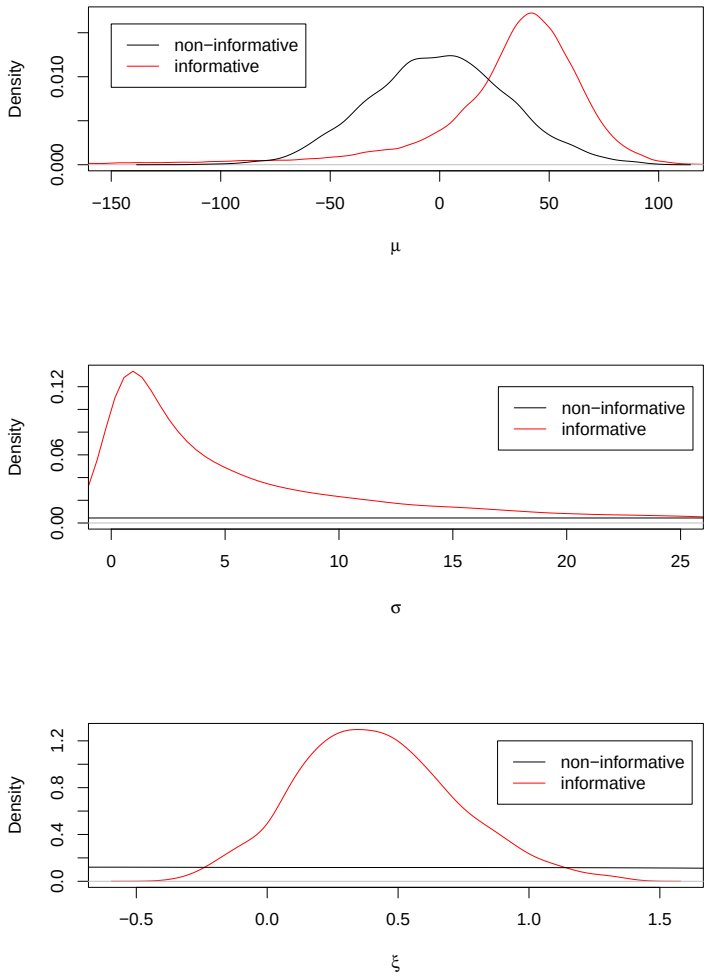


Figure 3.8: Plots of marginal non-informative and informative prior distributions of the GEV parameters

deviation of ξ is always significantly reduced.

3.5.2 The effect on the return levels

To look at how the informative prior information has affected the return levels, the posterior densities of return levels obtained in both cases can be compared. Figure 3.11 gives plots of the posterior densities of the 10, 100 and 1000-year return levels for site 1, for both non-informative and informative priors. From these plots, it is clear that the

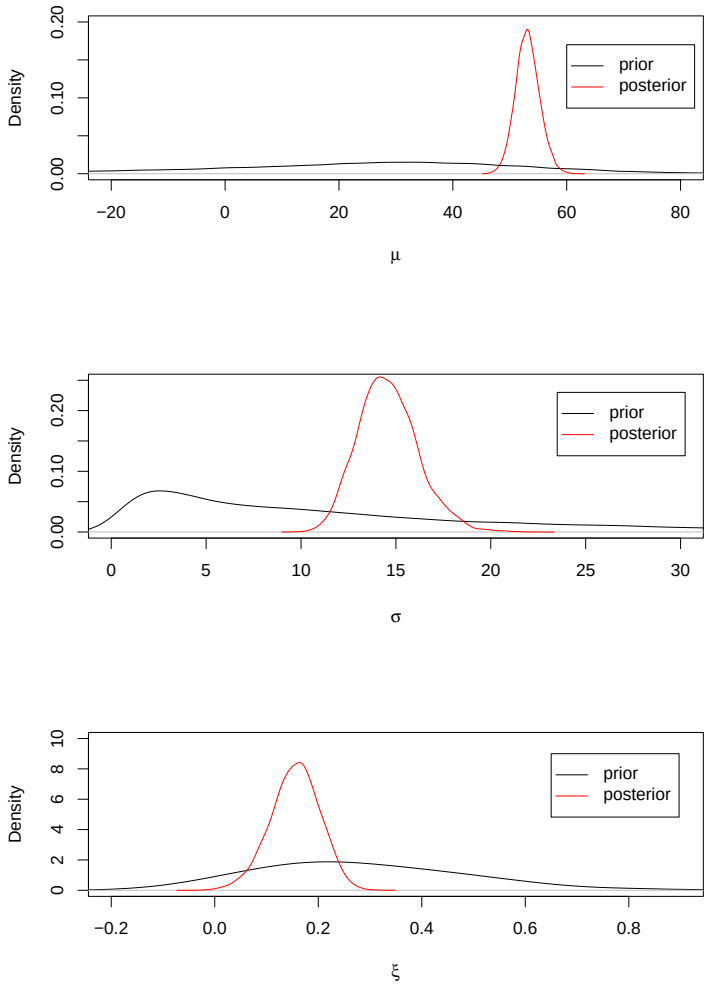


Figure 3.9: Plots of informative prior and posterior distributions of the GEV parameters for site 1

informative prior has had an effect on the distribution of return levels. For the 10-year return level the effect is not great: there is a slight reduction in uncertainty. For the 100 and 1000-year return levels the effect is much greater: in both cases there is a significant reduction in uncertainty. The posterior densities all have a much more symmetric shape when the informative prior is used.

To give an overall view of the effect on the posterior medians for site 1, plots of posterior medians of return levels against return period, for a range of values of p between 0 and 1,

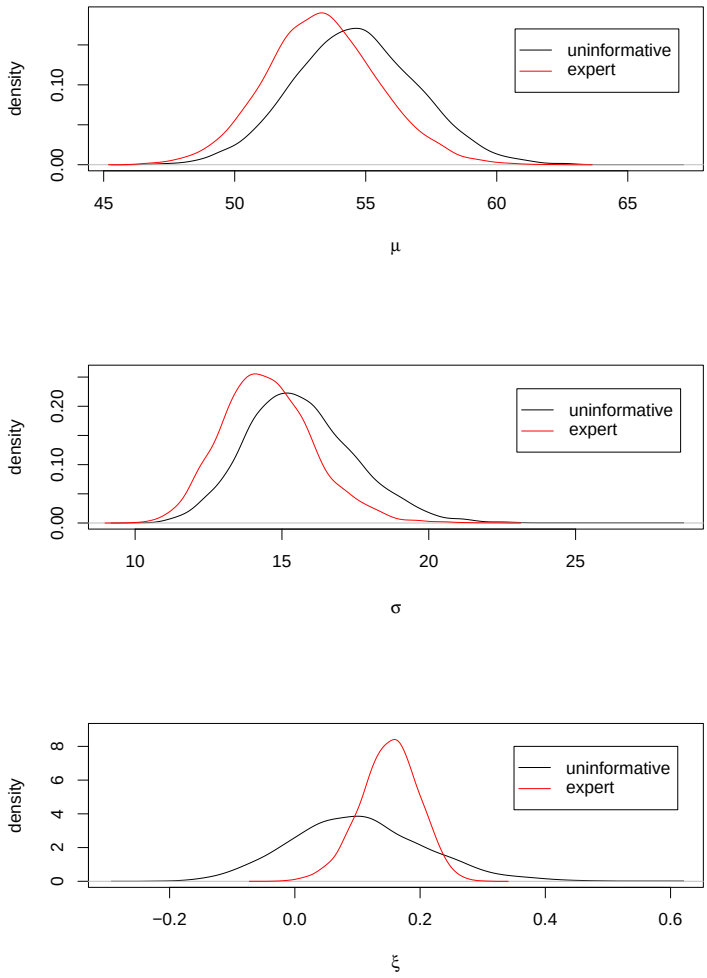


Figure 3.10: Posterior densities of the GEV parameters for site 1 using both non-informative priors and informative priors with c=3

are given for both priors in Figure 3.12. Maximum likelihood estimates of the return levels and medians of a simulated series from the informative prior are also given, together with the empirical estimates. This shows that the posterior medians of the return levels are fairly close together for the two priors at low return periods, but at greater return periods the posterior medians tend to be higher when the informative prior is used. The posterior medians found with non-informative priors are fairly close to the MLEs but do tend to be slightly higher. The two lines based on posterior medians and the MLE line seem to

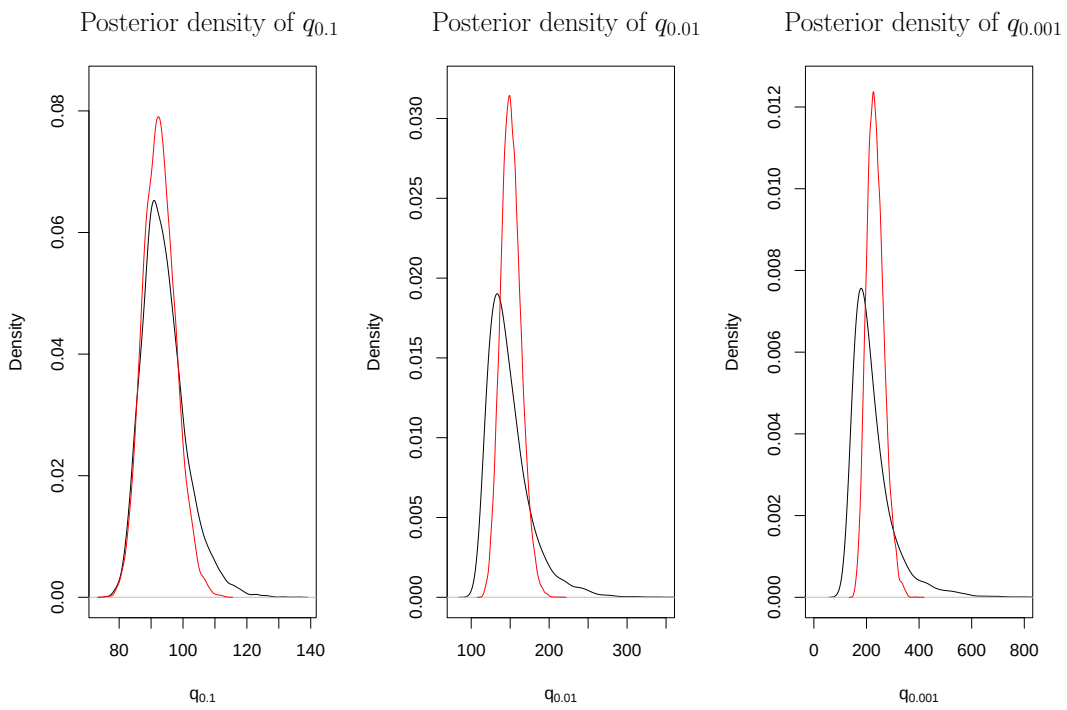


Figure 3.11: Posterior densities of the 10, 100 and 1000-year return levels for site 1, using a non-informative prior (black) and an informative prior (red)

be in close agreement to the empirical estimates, but the posterior means based on the informative prior do seem to be in closest agreement with the most extreme empirical estimates.

Contrary to expectation, the expert posterior curve is not nested between the non-informative posterior curve (or MLE curve) and the prior curve. This effect was also observed in the study of site 9 by Coles and Tawn (1996). They explain that this is due to the substantial skew in the profile likelihood surfaces, especially in the most extreme return levels.

Figure 3.13 shows the return level curve based on the posterior medians obtained using the informative prior, with 95 % credibility intervals based on both non-informative and informative priors. The interval based on the informative prior information is much narrower than that based on the non-informative prior. The informative prior information has clearly reduced the variability in the posterior greatly.

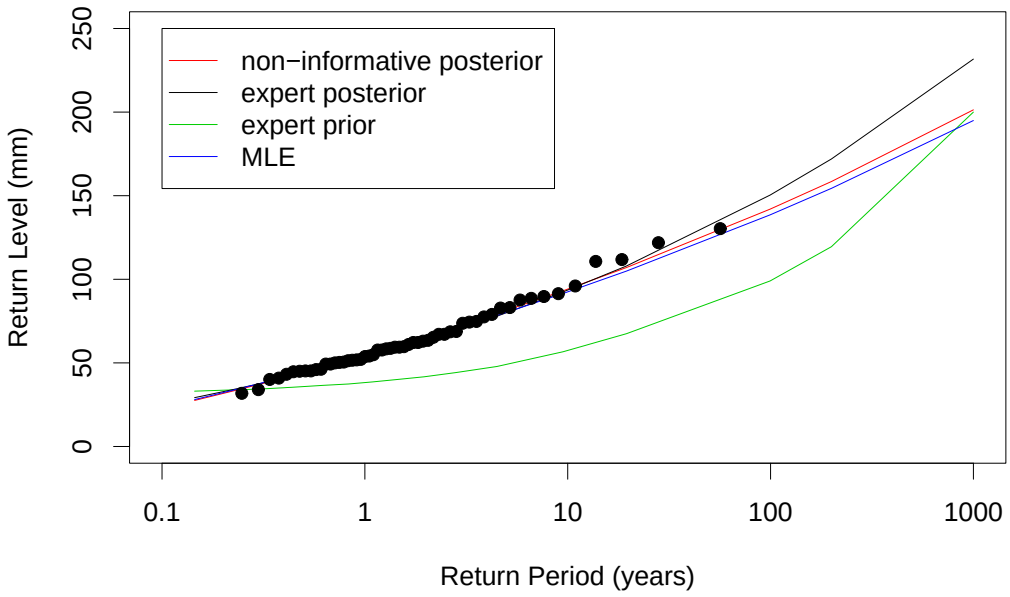


Figure 3.12: Plots of return levels against return period for site 1 based on the non-informative posterior distribution, the informative posterior distribution, the informative prior distribution, the maximum likelihood estimate and empirical estimates (points)

To look more closely at the effect of the informative prior, the posterior medians and 95% credibility intervals obtained for the two priors (given in Table 3.2 and Table 3.8) can be compared: The posterior medians of the 10-year return levels are very close for the two different priors; for both the 100 and 1000-year return levels, most of the posterior medians are higher when the informative prior is used, and the difference is much greater than for the 10-year return level; most of the 95% credibility intervals are narrower when the informative priors are used. For sites 7 and 10 the 95% credibility intervals tend to shift upwards, rather than get narrower, when the informative prior is used. This suggests that the prior and the data are in conflict for these two sites.

The overall effect of the informative prior on the return levels is to give higher posterior medians and much narrower credibility intervals for the high return periods than the non-

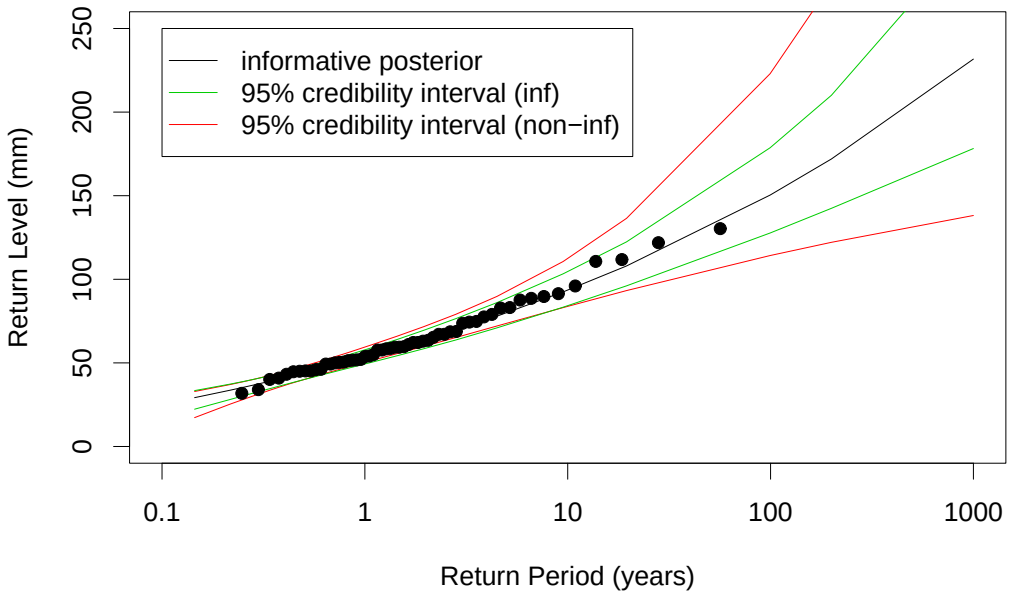


Figure 3.13: Plot of posterior medians of return levels using the informative prior and empirical estimates of return levels (points) with 95 % credibility intervals based on both non-informative and informative priors

informative prior does. Again, this is in agreement with the results for site 9 obtained by Coles and Tawn (1996). They comment that this effect is consistent with the qualitative reasoning supplied by the expert to support his choice of prior.

3.6 Prediction

The predictive distribution, which can be used to find the probability of extreme events occurring in the future, was introduced in section 3.1. The predictive distribution of a future observation y given the data $\boldsymbol{x}$ is given in (3.2). If Z_L is the maximum daily rainfall over a future period of L years, then the predictive distribution of Z_L (Aitchison and Dunsmore, 1975), after allowing for uncertainty in parameter estimation, is defined

as

$$Pr(Z_L < z|\boldsymbol{x}) = \int_{\Theta} Pr(Z < z|\boldsymbol{\theta})^L \pi(\boldsymbol{\theta}|\boldsymbol{x}) d\boldsymbol{\theta}, \tag{3.17}$$

where Z is the annual maximum. A design level which, after allowing for parameter uncertainty, will be exceeded with probability p in an L-year period, is the level z_p such that $Pr(Z_L < z_p|\boldsymbol{x}) = 1 - p$. Davison (1986) gave a likelihood based approximation scheme to find the predictive distribution of Z, where Z has a GEV distribution. Direct estimation of (3.17) can be obtained from the Gibbs sampler output by using

$$\hat{Pr}(Z_L < z|\boldsymbol{x}) = M^{-1} \sum_{m=1}^{M} Pr(Z < z|\theta_m)^L, \tag{3.18}$$

where θ_m is the output from the mth iteration of a sample of size M from the Gibbs sampler of the posterior distribution of θ. To obtain predictive return levels the equation

$$M^{-1} \sum_{m=1}^{M} Pr(Z < z_p|\theta_m)^L = 1 - p \tag{3.19}$$

can be solved for z_p using a numerical solver. Solving with a value of $L = 1$ would give a level which will be exceeded with probability p in a 1-year period. This is equivalent to the definition of a return level, so these predictive return levels can be compared with return levels obtained through the other methods. For the block maxima method, and using L=1, equation (3.19) becomes

$$M^{-1} \sum_{m=1}^{M} \exp\left\{ -\left[1 + \xi\left(\frac{z_p - \mu}{\sigma}\right)\right]^{-1/\xi} \right\} = 1 - p. \tag{3.20}$$

Solving (3.20) for z_p with various values of p and plotting z_p against return level on a logarithmic scale gives a return level plot which allows for parameter uncertainty. Figure 3.14 gives such plots for site 1 using both the non-informative and informative priors. The return level plots based on the posterior medians of the return levels, with both non-informative and informative priors, are also given. In both prior cases, the predictive return level curve is above the corresponding posterior median return level curve at high values of $-1/\log(1-p)$. At other values of $-1/\log(1-p)$ the curves are close together. So, if the posterior medians of return levels were used for design purposes there may be

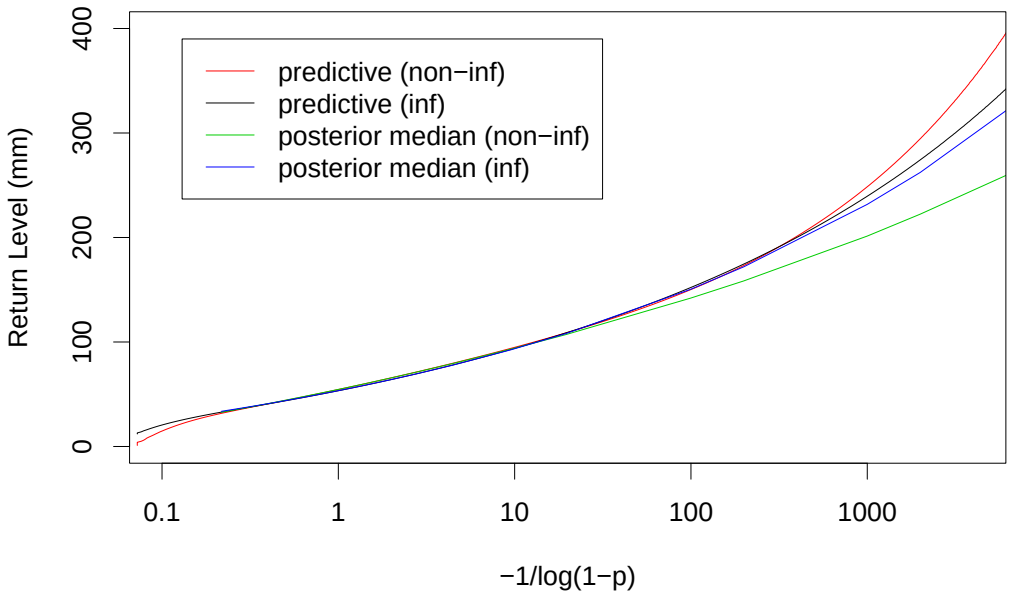


Figure 3.14: Predictive return level plots and return level plots based on posterior median of return level for site 1, using both non-informative and informative priors

a considerable amount of under-protection. This is due to uncertainty in the parameter estimates, which is accounted for by using the predictive estimates. There is a much greater difference between the posterior median and predictive return level curves for the non-informative prior case than for the informative prior case, showing the extent to which the informative prior has reduced uncertainty. The predictive return level curve based on the non-informative prior is noticeably higher than that based on the informative prior for high values of $-1/\log(1-p)$, but the curves are very close at other values. This suggests that due to a lack of certainty in the model with non-informative priors, the predictive return levels are over estimated at high values of $-1/\log(1-p)$.

The 10, 100 and 1000-year predictive return levels for all sites, obtained using both non-informative and informative priors, are given in Table 3.9. The predictive return levels obtained using the non-informative prior can be compared with the posterior medians of

	Non-informative prior			**Informative prior**		
	Return period (years)			Return period (years)		
Site	10	100	1000	10	100	1000
1	93.885	150.096	249.076	92.798	151.980	239.314
2	101.427	160.369	271.297	99.807	155.741	246.344
3	86.198	134.483	246.379	86.134	136.828	228.160
4	99.056	138.893	206.561	100.641	155.823	240.241
5	94.286	138.659	221.423	95.122	150.301	237.515
6	94.359	200.025	468.649	85.740	140.182	246.558
7	63.715	89.978	230.249	67.977	112.271	191.471
8	70.567	110.576	192.060	72.642	121.675	210.500
9	71.329	133.291	292.977	69.670	121.510	219.435
10	59.161	86.812	128.071	64.282	111.155	190.710
11	66.091	109.129	191.487	67.769	118.543	207.853

Table 3.9: Predictive return levels for the 10, 100 and 1000-year return periods

the return levels given in Table 3.2. For all three return periods, the predictive return levels are higher than the posterior medians of the return levels, but there is very little difference between the posterior medians of the return levels and the predictive return levels for the 10-year return period. So, under the non-informative prior, taking into account parameter uncertainty, by using the predictive return levels, results in an increase in the estimate of the return levels. This increase is greater for the longer return periods. A similar effect is seen when using the informative prior, although the difference between the predictive return levels and the posterior medians, given in Table 3.8, is much less. The predictive return levels are higher than the posterior medians for all three return periods and the difference increases with return period, but the differences are much smaller than those observed with the non-informative prior. This can be attributed to the fact that using the informative prior has reduced uncertainty in the model.

3.7 Discussion

This chapter has introduced Bayesian modelling and has demonstrated how MCMC techniques can be used to simulate from posterior distributions. This chapter has also introduced expert prior information which was used to formulate informative priors. These informative priors were used to model the annual maxima rainfall data from each of the eleven sites in Figure 2.1 and will also be used later in the thesis. The results obtained with the informative priors were compared with those obtained with non-informative priors.

Posterior means and standard deviations of the GEV parameters for each site were obtained using both the non-informative priors and the informative priors. The informative prior affected both the posterior means and standard deviations, with the standard deviations being lower for the model with informative priors. This reduction in the standard deviations reflects the decrease in uncertainty due to the expert information. The effect was most noticeable for the shape parameter ξ.

The samples from the marginal posterior distributions of the GEV parameters were used to obtain samples from the posterior distributions of return levels. Since the informative prior had an effect on the distributions of the GEV parameters, there was also an effect on the distributions of the return levels. As with the GEV parameters, both the location and spread of the return levels were affected by the informative prior, with the spread being reduced by the use of expert information.

Predictive return levels for each site were also obtained. These values were found to be higher than the posterior medians of the return levels when both non-informative and informative priors were used. There was also some difference between the predictive return levels obtained with the non-informative prior and those obtained with the informative prior, with greater differences observed at longer return periods.

For sites 7 and 10 the informative prior did not always result in smaller posterior standard deviations (for the GEV parameters) and 95% credibility intervals (for the return levels). Combining prior information and data which suggest different parameter values could result in a more spread out posterior distribution, giving larger measures of spread.

It is possible, then, that for these sites the informative prior and the data are in conflict with each other. This could mean that the data is not very representative of what is typically likely to be observed, or that the prior is not applicable to data from these two sites. Using daily data, rather than just the annual maxima, may be beneficial in this case, as the data may be more representative of the locations. Multivariate models for the daily data are considered in chapters 5 and 6.

This analysis has provided a demonstration of the Bayesian approach to modelling, whilst introducing the informative prior which will be used later for multivariate analyses. The results have demonstrated the importance of using expert information if it is available.

Chapter 4

Modelling bivariate annual maxima data

4.1 Introduction

The work in this chapter is based on the work in Smith and Walshaw (2003) and is a first step towards building a comprehensive Bayesian strategy for inference on multivariate extremes of environmental variables. Coles and Tawn (1996a) carried out a sophisticated Bayesian analysis of extremes from one of the locations of Figure 2.1. However, in a (frequentist) multivariate analysis of extremes from all 11 sites, Coles and Tawn (1996b) demonstrated the importance of dependence across sites in terms of modelling the extremal behaviour of rainfall in this region.

In this chapter, models for the bivariate distribution of extremes at pairs of sites are considered. Using models based on limiting joint distributions, posterior inferences for joint extremal behaviour are made using an MCMC scheme. In Section 4.2, two models for dependence (the logistic model and the mixed model) are described and the procedure for combining marginal and dependence parameters into the model is outlined. The procedure for inference is described in Section 4.3; details of the joint densities for the logistic and mixed models are also given. In Section 4.4, the bivariate logistic and mixed

models are fitted to the data from sites 1 and 2 of Figure 2.1, with non-informative priors, to compare the performance of the two models. One of the models is then chosen for analysis of the data from each pair of sites. In Section 4.5, the model is then applied to the data from each pair of sites, using an informative prior. Informative priors for the marginal behaviour of extremes at individual sites are based on modifications of carefully elicited expert beliefs, while a prior specification for the dependence between sites utilises a model which relates the strength of dependence inversely to the distance separation between sites. In section 4.6, the results from the models with non-informative and informative priors are compared to determine the effect the informative prior has on posterior inference. The results are also compared (in Section 4.7) with the results from the univariate analyses given in Chapter 3. The predictive distribution of return levels for both priors is considered in Section 4.8, and a discussion of the results of the chapter is given in Section 4.9.

4.2 Model building

4.2.1 Models for dependence

Parametric modelling involves identifying parametric sub-families of G in (1.12), which equates to specifying parametric families for H (see (1.15)) on $[0, 1]$ with mean equal to 0.5 for every value of the parameter. In practice it is not trivial to create models with a useful range of dependence; for example, an early attempt by Gumbel (1960) allowed only for negative dependence between variables. Two practical models identified by Tawn (1988a) are considered here.

The *logistic model* has distribution function given by

$$G(x, y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right\}, \qquad x > 0, y > 0, \tag{4.1}$$

for a parameter $\alpha \in (0, 1]$. This is obtained non-trivially by appropriate choice of H in (1.13); see Tawn (1988a) for details. This family is popular because of its flexibility. In

particular, the limits $\alpha = 1$ and $\alpha \to 0$ correspond to complete independence and perfect dependence respectively, so the full range of positive dependence is available with this model.

The *mixed model* has distribution function given by

$$G(x, y) = \exp\left\{-\left(\frac{1}{x} + \frac{1}{y}\right) + \frac{\beta}{x+y}\right\}, \qquad x > 0, y > 0, \tag{4.2}$$

for a parameter $\beta \in [0, 1]$. Here $\beta = 0$ corresponds to independence, but perfect dependence is not possible. In fact the maximum Pearson correlation possible under this model is obtained from an expression given by Tawn (1988) as 0.473. Note that both the logistic model and the mixed model are symmetric, that is to say the variables X and Y are exchangeable.

4.2.2 Incorporating the marginal distributions

A full bivariate model for extremal behaviour will be based on one of the parametric models, and will be completely specified by a set of GEV parameters for each marginal component, and the appropriate dependence parameter. For example, the logistic model would be completely specified by the vector

$$\boldsymbol{\theta} = (\mu_X, \sigma_X, \xi_X, \mu_Y, \sigma_Y, \xi_Y, \alpha) \tag{4.3}$$

where the vectors (μ_X, σ_X, ξ_X) and (μ_Y, σ_Y, ξ_Y) determine the marginal GEV distributions, and α is the dependence parameter in (4.1) after transformation to unit Fréchet margins. The marginal parameters could be estimated first, by using the GEV distribution to model each margin, and then the data transformed to have unit Fréchet margins. Here however, estimation of the marginal parameters and the dependence parameter is carried out simultaneously. This approach is slightly more complicated to implement but is more efficient since it is possible that marginal estimation can be improved by incorporating the dependence into the model.

4.3 Inference

Consider the situation where the data available for modelling are componentwise maxima from two locations, $\boldsymbol{M}_r = (M_n^X, M_n^Y)_r, \quad r = 1, \ldots, t$, taken from t successive time periods of equal length n. This corresponds with the situation where annual maxima are available from the sites of interest. Adopting one of the models in Section 4.1 (thereby assuming exchangeability of maxima between locations), the likelihood can be calculated by substituting the transformed marginal maxima into the appropriate distribution function. The details for the two models are given in 4.3.1 and 4.3.2.

The prior is most naturally specified as a product of joint priors for each of (μ_X, σ_X, ξ_X) and (μ_Y, σ_Y, ξ_Y), and a univariate prior for α. The prior then has the form

$$\pi\left(\mu_x, \sigma_x, \xi_x, \mu_y, \sigma_y, \xi_y, \alpha\right) = \pi_x\left(\mu_x, \sigma_x, \xi_x\right) \pi_y\left(\mu_y, \sigma_y, \xi_y\right) \pi_\alpha\left(\alpha\right). \tag{4.4}$$

It is easy to obtain a sample from the approximate posterior distribution for the complete parameter vector $\boldsymbol{\theta}$ using MCMC methods. Specific details about the priors used are given later.

4.3.1 Logistic Model

From (4.1), for the logistic model

$$G(\tilde{x}, \tilde{y}) = \exp\left\{-(\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^\alpha\right\}, \qquad \tilde{x} > 0,\ \tilde{y} > 0, \tag{4.5}$$

where

$$\tilde{x} = k_x^{(1/\xi_x)} \qquad \text{for} \qquad k_x = \left[1 + \xi_x\left(\frac{x - \mu_x}{\sigma_x}\right)\right], \tag{4.6}$$

and

$$\tilde{y} = k_y^{(1/\xi_y)} \qquad \text{for} \qquad k_y = \left[1 + \xi_y\left(\frac{y - \mu_y}{\sigma_y}\right)\right] \tag{4.7}$$

are the transformations of x and y to unit Fréchet distribution. In this case, from (1.17) and (4.5),

$$V(\tilde{x}, \tilde{y}) = (\tilde{x}^{-1/\alpha} + \tilde{y}^{-1/\alpha})^\alpha.$$

The joint density for a bivariate component wise maximum (x, y) is then given by

$$g(x,y) = \{V_x(\tilde{x},\tilde{y})V_y(\tilde{x},\tilde{y}) - V_{xy}(\tilde{x},\tilde{y})\} \exp\{-V(\tilde{x},\tilde{y})\}, \tag{4.8}$$

where V_x, V_y and V_{xy} represent the partial and mixed derivatives of V.

$$\begin{aligned} V_x(\tilde{x},\tilde{y}) &= \alpha\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-1} \times \left(-\frac{1}{\alpha\xi_x} k_x^{-1/\alpha\xi_x - 1}\right)\frac{\xi_x}{\sigma_x} \\ &= -\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-1} \frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x} \end{aligned}$$

and similarly,

$$V_y(\tilde{x},\tilde{y}) = -\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-1} \frac{k_y^{-1/\alpha\xi_y-1}}{\sigma_y}.$$

$$\begin{aligned} V_{xy}(\tilde{x},\tilde{y}) &= (\alpha-1)\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-2} \frac{1}{\alpha\xi_y} k_y^{-1/\alpha\xi_y-1}\frac{\xi_y}{\sigma_y}\frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x} \\ &= \frac{\alpha-1}{\alpha}\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-2} \frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x}\frac{k_y^{-1/\alpha\xi_y-1}}{\sigma_y}, \end{aligned}$$

so from (4.8)

$$\begin{aligned} g(x,y) &= \left\{\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{2(\alpha-1)} \frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x}\frac{k_y^{-1/\alpha\xi_y-1}}{\sigma_y}\right. \\ &\quad \left. -\frac{\alpha-1}{\alpha}\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-2} \frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x}\frac{k_y^{-1/\alpha\xi_y-1}}{\sigma_y}\right\} \\ &\quad \times \exp\left\{-\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha}\right\} \\ &= \frac{k_x^{-1/\alpha\xi_x-1}}{\sigma_x}\frac{k_y^{-1/\alpha\xi_y-1}}{\sigma_y}\left\{\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{2\alpha-1}\right. \\ &\quad \left. -\frac{\alpha-1}{\alpha}\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha-1}\right\}\exp\left\{-\left(k_x^{-1/\alpha\xi_x} + k_y^{-1/\alpha\xi_y}\right)^{\alpha}\right\}. \end{aligned} \tag{4.9}$$

4.3.2 Mixed Model

For the mixed model the distribution function is given by

$$G(\tilde{x},\tilde{y}) = \exp\left\{-\left(\frac{1}{\tilde{x}} + \frac{1}{\tilde{y}}\right) + \frac{\beta}{\tilde{x}+\tilde{y}}\right\}, \qquad 0 \le \beta \le 1,$$

where $\tilde{x}$ and $\tilde{y}$ are the transformations of x and y to unit Fréchet margins given by (4.6) and (4.7) respectively. The exponent measure then is

$$V(\tilde{x}, \tilde{y}) = \frac{1}{\tilde{x}} + \frac{1}{\tilde{y}} - \frac{\beta}{\tilde{x} + \tilde{y}}.$$

The joint density is given by (4.8), where

$$\begin{aligned} V_x(\tilde{x}, \tilde{y}) &= \frac{\partial \tilde{x}}{\partial x}\left(\frac{\beta}{(\tilde{x}+\tilde{y})^2} - \frac{1}{\tilde{x}^2}\right), \\ V_y(\tilde{x}, \tilde{y}) &= \frac{\partial \tilde{y}}{\partial y}\left(\frac{\beta}{(\tilde{x}+\tilde{y})^2} - \frac{1}{\tilde{y}^2}\right), \\ V_{xy}(\tilde{x}, \tilde{y}) &= -\frac{\partial \tilde{x}}{\partial x}\frac{\partial \tilde{y}}{\partial y}\frac{\beta}{(\tilde{x}+\tilde{y})^3}, \end{aligned}$$

for

$$\frac{\partial \tilde{x}}{\partial x} = \frac{k_x^{1/\xi_x - 1}}{\sigma_x} \quad \text{and} \quad \frac{\partial \tilde{y}}{\partial y} = \frac{k_y^{1/\xi_y - 1}}{\sigma_y}.$$

4.4 Application to spatial data with non-informative priors

The aim here is to fit a bivariate model to the annual maxima data from each pair of sites from the 11 locations in Figure 2.1 first using non-informative priors and then using informative priors in section 4.5. Knowledge about the locations in Figure 2.1 suggests that it is reasonable to assume that annual maxima from the 11 locations are exchangeable. The number of years of data available for the 55 pairs of sites ranges from 28 to 60. The annual maxima data being considered are taken as the maxima of daily rainfall totals over periods of approximately 365 days. Although the assumption of an identical distribution for these daily totals is unlikely to be realistic because of seasonal variations, many analyses both of data from these locations and of rainfall data generally demonstrate that the GEV is nevertheless a very good model for the marginal behaviour of annual maxima, and that there is no evidence for a long-term trend in extremal behaviour of rainfall over the relevant period of the 20th Century.

In this section the procedure is demonstrated by using vague priors. This is enlightening from the point of view of assessing convergence behaviour, and comparing the performances of the logistic and mixed models. The model is fitted to annual maximum data from locations 1 and 2 in Figure 2.1, using the 45 bivariate componentwise maxima available from 1916 to 1973 (some years are missing at each site).

4.4.1 Prior choice

As for the univariate example in section 3.4.1, the parametrisation $\phi_X = \log \sigma_X$ and $\phi_Y = \log \sigma_Y$ is used for specification of uninformative priors. The six parameters $\mu_X, \phi_X, \xi_X, \mu_Y, \phi_Y$ and ξ_Y are chosen to have independent univariate normal priors with large variances. In particular,

$$\mu_X \sim N(0, 1000) \qquad \phi_X \sim N(0, 100) \qquad \xi_X \sim N(0, 10)$$

$$\mu_Y \sim N(0, 1000) \qquad \phi_Y \sim N(0, 100) \qquad \xi_Y \sim N(0, 10).$$

These priors are denoted by $\pi_{\mu_x}(\mu_x)$, $\pi_{\phi_x}(\phi_x)$, $\pi_{\xi_x}(\xi_x)$,$\pi_{\mu_y}(\mu_y)$, $\pi_{\phi_y}(\phi_y)$ and $\pi_{\xi_y}(\xi_y)$. The dependence parameters of the two models, α and β, both range between 0 and 1. Assuming nothing is known about the dependence, $U(0,1)$ priors are appropriate as all values in the range are equally likely. The joint prior then is given by

$$\pi(\mu_x, \phi_x, \xi_x, \mu_y, \phi_y, \xi_y, \alpha) = \pi_x(\mu_x, \phi_x, \xi_x)\, \pi_y(\mu_y, \phi_y, \xi_y),$$

where

$$\pi_x(\mu_x, \phi_x, \xi_x) = \pi_{\mu_x}(\mu_x)\, \pi_{\phi_x}(\phi_x)\, \pi_{\xi_x}(\xi_x),$$

and

$$\pi_y(\mu_y, \phi_y, \xi_y) = \pi_{\mu_y}(\mu_y)\, \pi_{\phi_y}(\phi_y)\, \pi_{\xi_y}(\xi_y).$$

4.4.2 Posterior inference

From Bayes' theorem, the posterior has the form

$$\pi(\theta|\boldsymbol{x}, \boldsymbol{y}) \propto \pi_x(\mu_x, \phi_x, \xi_x)\, \pi_y(\mu_y, \phi_y, \xi_y) \prod_{i=1}^{n} g(x_i, y_i),$$

where g is as the joint distribution function for either the logistic model or the mixed model and n is the number of observations.

The full conditionals have the form

$$\pi(\mu_x|\cdot) = \pi_{\mu_x}(\mu_x)L(\theta|\boldsymbol{z}), \qquad \pi(\mu_y|\cdot) = \pi_{\mu_y}(\mu_y)L(\theta|\boldsymbol{z}),$$

$$\pi(\phi_x|\cdot) = \pi_{\phi_x}(\phi_x)L(\theta|\boldsymbol{z}), \qquad \pi(\phi_y|\cdot) = \pi_{\phi_y}(\phi_y)L(\theta|\boldsymbol{z}),$$

$$\pi(\xi_x|\cdot) = \pi_{\xi_x}(\xi_x)L(\theta|\boldsymbol{z}), \qquad \pi(\xi_y|\cdot) = \pi_{\xi_y}(\xi_y)L(\theta|\boldsymbol{z})$$

$$\pi(\alpha|\cdot) = L(\theta|\boldsymbol{z}) \quad \text{for the logistic model,}$$

$$\pi(\beta|\cdot) = L(\theta|\boldsymbol{z}) \quad \text{for the mixed model.}$$

The MCMC algorithm used here follows the same pattern as the univariate algorithm given in section 3.3.2, with additional steps for the extra parameters. The variances of the innovations were chosen, by trial and error, as values which resulted in good convergence of the chains. The variances vary slightly for each pair of sites but for the pair (1,2), for which the details are given here, the innovations are simulated from the following distributions:

$$\omega_{\mu_x} \sim N(0, 4.2), \qquad \omega_{\phi_x} \sim N(0, 0.3), \qquad \omega_{\xi_x} \sim N(0, 0.2),$$

$$\omega_{\mu_y} \sim N(0, 4), \qquad \omega_{\phi_y} \sim N(0, 0.3), \qquad \omega_{\xi_y} \sim N(0, 0.2),$$

$$\omega_\alpha \sim N(0, 0.16) \quad \text{for the logistic model,}$$

$$\omega_\beta \sim N(0, 0.2) \quad \text{for the mixed model.}$$

To compare the performance of the logistic and mixed models, the two models were fitted to the data from sites 1 and 2. Figure 4.1 shows traces and densities of the dependence parameters for each of the logistic and mixed models for the models applied to sites 1 and 2. Autocorrelation plots are given in Figures 4.2 and 4.3, for all parameters of the logistic and mixed models respectively.

The trace plots show all 30000 iterations whilst for the density plots, the first 10000 iterations are discarded as burn-in. Convergence appears to occur rapidly for both models.

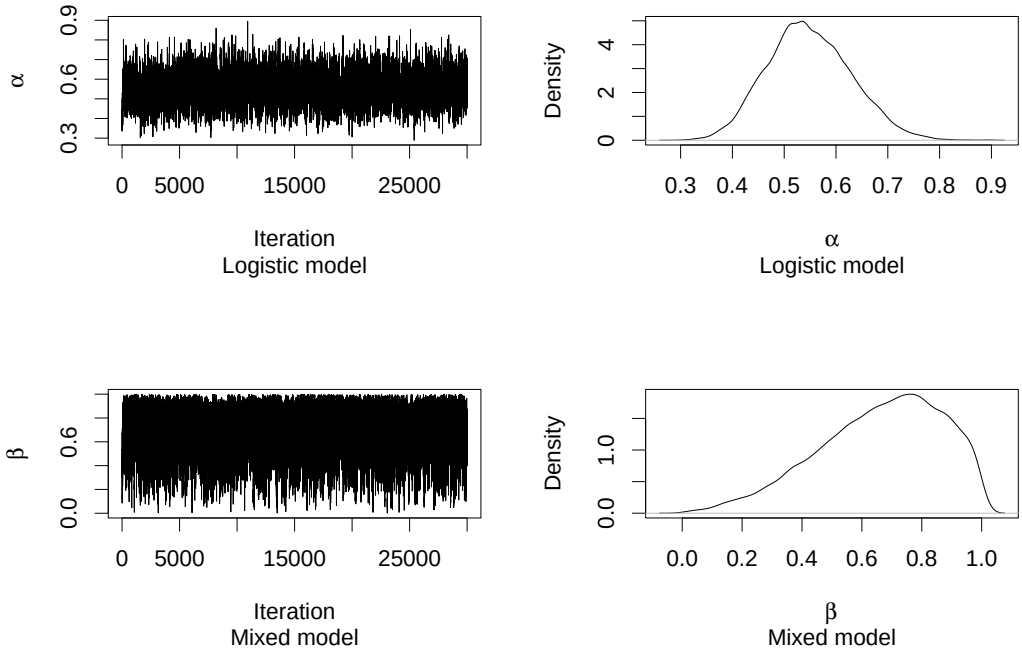


Figure 4.1: Trace and density plots of the dependence parameter for logistic and mixed models for sites 1 and 2

For the mixed model, however, the trace and density for β indicate that the posterior density remains substantial as $\beta \to 1$. Given that this limit corresponds to only moderate dependence between extremes at the two locations, it appears that the mixed model may be too restrictive to represent the levels of extremal dependence that exist between locations in this region. By contrast, the posterior density for α in the logistic model appears to be well contained within the interior of $[0, 1]$, suggesting that the capacity of this model to represent the full range of positive extremal dependence is sufficient to overcome the limitations of the mixed model. This, together with the good convergence and mixing properties exhibited by the time-series plots, suggests using the logistic model in preference to the mixed model for the purpose of serious application to the spatial data collected at the 11 locations in Figure 2.1. The autocorrelation plots do show, however, that the autocorrelations of the parameters reduce more quickly in the mixed model than in the logistic model. This difference does not appear to be substantial though.

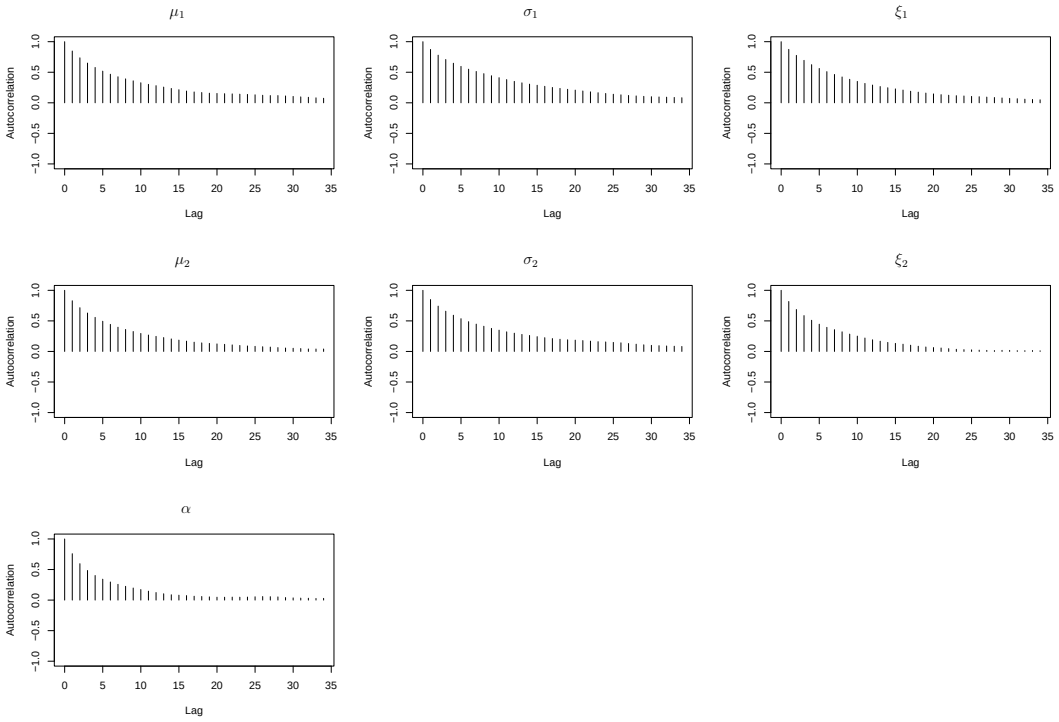


Figure 4.2: Autocorrelation plots for the parameters of the logistic model

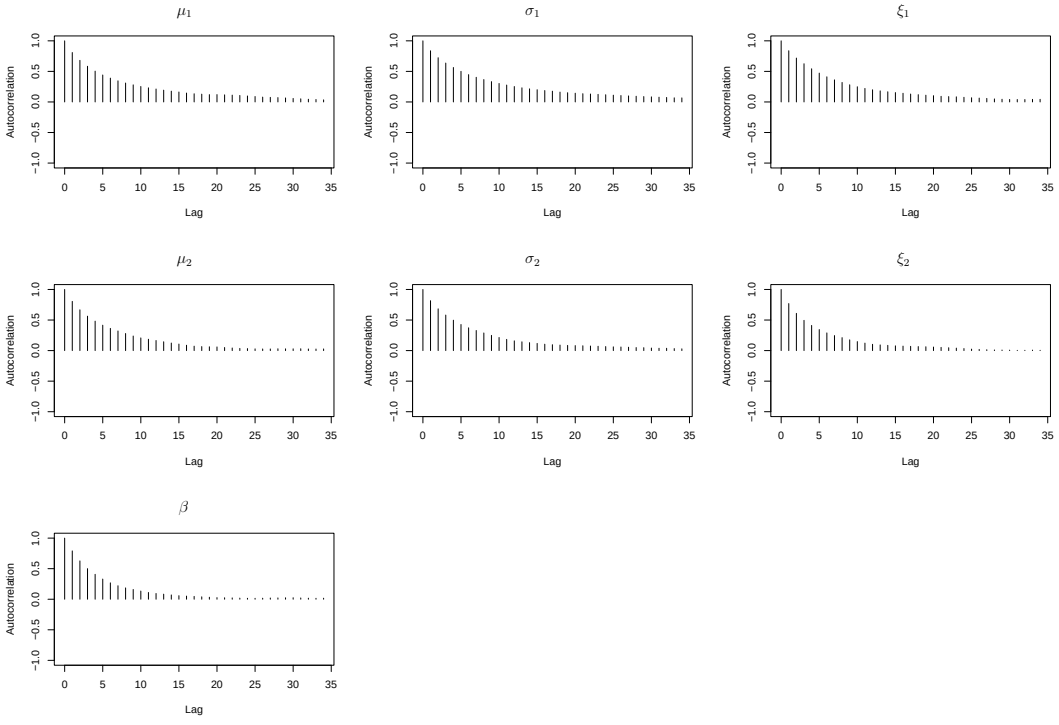


Figure 4.3: Autocorrelation plots for the parameters of the mixed model

The logistic model then is used to model all 55 pairs of sites. The posterior means and standard deviations of each parameter for all pairs are given in table A.1 and table A.2 respectively.

The distances between the two sites of each pair vary between 0.5657 km and 33.4332 km. The posterior means and standard deviations of the parameters for three particular pairs, chosen to have widely different distance separations, are given in Table 4.1. The first pair is (4,5), which are the two closest sites. The second pair is (2,6), which have a distance separation in the middle of the range (15.8673 km). The third pair are the two sites furthest apart, (7,11).

	$i = 4,\ j = 5$		$i = 2,\ j = 6$		$i = 7,\ j = 11$	
Parameter	Mean	S.D.	Mean	S.D.	Mean	S.D.
μ_i	65.3760	1.8879	66.7578	2.2961	44.6228	1.2768
σ_i	13.4410	1.4929	13.1464	2.1067	8.3212	0.9396
ξ_i	0.1206	0.0958	0.3688	0.1408	-0.0237	0.0994
μ_j	62.7338	1.8780	59.2407	1.6699	40.5698	1.4905
σ_j	13.4834	1.4834	9.6765	1.6270	9.1998	1.3352
ξ_j	0.0750	0.0966	0.4443	0.1346	0.3249	0.1578
α	0.3342	0.0504	0.6056	0.0996	0.9076	0.0651

Table 4.1: Posterior means and standard deviations for the bivariate model with non-informative priors applied to the pairs of sites (4,5), (2,6) and (7,11)

The posterior means of the dependence parameter α are very different for the three different pairs. For the pair closest together, (4,5), the posterior mean is very low (0.3342), indicating strong dependence. The pair furthest apart, (7,11), has a very high posterior mean (0.9076), indicating very weak dependence. The pair with intermediate distance separation, (2,6), has a posterior mean of 0.6056, which indicates moderate dependence. These results confirm the belief that dependence increases with decreasing distance separation.

A plot of the posterior mean of α for each pair against the distance between the two sites is given in Figure 4.4. There does seem to be a relationship between the dependence

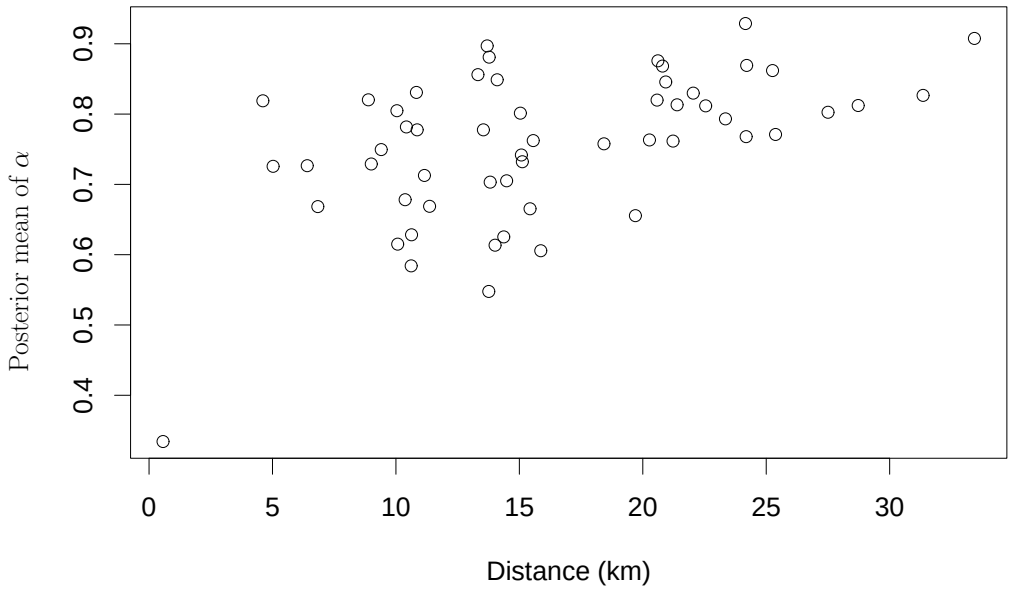


Figure 4.4: Plot of posterior mean of α (using non-informative priors) against distance between sites

parameter and the distance between sites. At the greater distances, the posterior means of α do tend to be high, meaning there is weak dependence. As the distance separation decreases, there are occurrences of lower posterior means but there are still some quite high values. The only posterior mean below 0.5 is for the pair of sites that are the closest together (4,5). For the two sites which have the second smallest distance separation (4.6098km) the posterior mean is 0.8189. This indicates very weak dependence. On the other hand, there are no cases of strong dependence at the higher distance separations. The variability in the dependence parameter for any distance separation suggests that, while the dependence does increase with decreasing distance, there are additional factors affecting the dependence.

As for the univariate model, the vectors of observations from the marginal distributions

of the GEV parameters can be substituted into (1.4) to get realisations from the posterior densities of return levels. This could be carried out for both of the sites concerned in the bivariate analysis. Since the results for site 1 have been given in the univariate analysis, the results obtained for site 1, based on a bivariate analysis of sites 1 and 2, are given here. Table 4.2 gives posterior means, medians, SDs and IQRs for the 10, 100 and 1000-year return levels.

	Return period (years)		
	10	100	1000
Posterior mean	95.54	154.6	252.5
Posterior median	94.20	143.5	205.4
Posterior SD	8.610	45.44	196.3
Posterior IQR	9.37	35.9	99.7

Table 4.2: Posterior means, medians, standard deviations (SDs) and inter-quartile ranges (IQRs) for the 10, 100 and 1000-year return levels for site 1, using the non-informative prior

The posterior distributions of the return levels appear to be positively skewed as was observed in for the univariate case; the posterior means are greater than the posterior medians, especially for the larger return periods. The posterior medians and IQRs, then, are taken as the appropriate measures of location and spread.

4.5 Application using informative priors

The models applied in 4.4 to the bivariate annual maxima data are now extended to incorporate the expert prior information described in 3.4. Additionally, an informative prior is given for the dependence parameter of the logistic model, α. The model is applied to the data from all 55 pairs of sites.

4.5.1 Prior choice

In section 3.4.2, expert information is used to formulate a prior distribution in terms of the GEV parameters which can be used for modelling the data at each of the sites. This prior can be used as the priors for each of the two marginal distributions of the bivariate model. So, $\pi_x(\mu_x, \sigma_x, \xi_x)$ has the same form as (3.13), with μ, σ and ξ replaced by μ_x, σ_x and ξ_x respectively and the q_{p_i} replaced by $q_{p_i,x}$. Similarly, $\pi_y(\mu_y, \sigma_y, \xi_y)$ has the form of (3.13), with μ, σ and ξ replaced by μ_y, σ_y and ξ_y respectively and the q_{p_i} replaced by $q_{p_i,x}$.

Although no expert opinion is available on the extremal dependence between sites, prior knowledge exists in the form of experience in fitting the logistic model to bivariate annual maxima for rainfall data. This is most easily expressed in terms of the relationship between the dependence parameter α and the distance between sites, d. The functional form

$$\alpha = 1 - \exp(-\tau d^t), \tag{4.10}$$

which has the desired property of varying continuously between perfect dependence for $d = 0$ and complete independence as $d \to \infty$, is chosen here. The choice of the power of d, $t = 1/2$, is somewhat arbitrary, but gives a functional form which reflects the likely decay of the strength of dependence with increasing distance. For lower values of t, α appears to decay too slowly and for higher values it decays too rapidly. This is shown by Figure 4.5 which, for a value of $\tau = 0.4$, shows plots of the dependence parameter α against distance for a range of values of t.

Prior information for the dependence is now expressed as a distribution for the parameter τ. A $\Gamma(1.2, 4)$ distribution, having mean 0.3 and variance 0.075, was chosen for τ. This choice of distribution for τ gives distributions for α over the range of distances between sites in Figure 2.1 which reflect prior beliefs about the strength of dependence for these distance separations. Figure 4.6 shows density plots of α, based on 10^5 simulated values of τ, for a range of distances d. These plots show that for the smallest and largest distance separations, there is less uncertainty about the dependence than for other

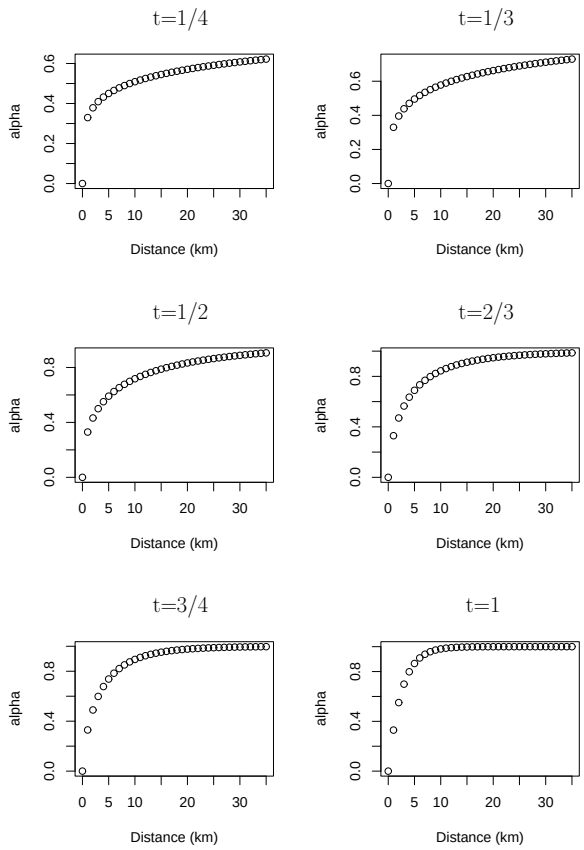


Figure 4.5: Plots of α against distance for $t = 1/4, 1/3, 1/2, 2/3, 3/4, 1$, and for a value of $\tau = 0.4$

distance separations. For small distances the mode is low and for the highest distance separations the mode is high. These features reflect the beliefs that dependence will be strong for locations close together and become weaker as distance separation increases. Less is known about exactly how α changes with distance, which is reflected in the flatter shaped distributions for intermediate distances.

From (4.10), an expression for τ in terms of α is

$$\tau = -\frac{\log(1-\alpha)}{d^{1/2}}. \tag{4.11}$$

The prior of α can be obtained by substituting (4.11) into the prior for τ and multiplying

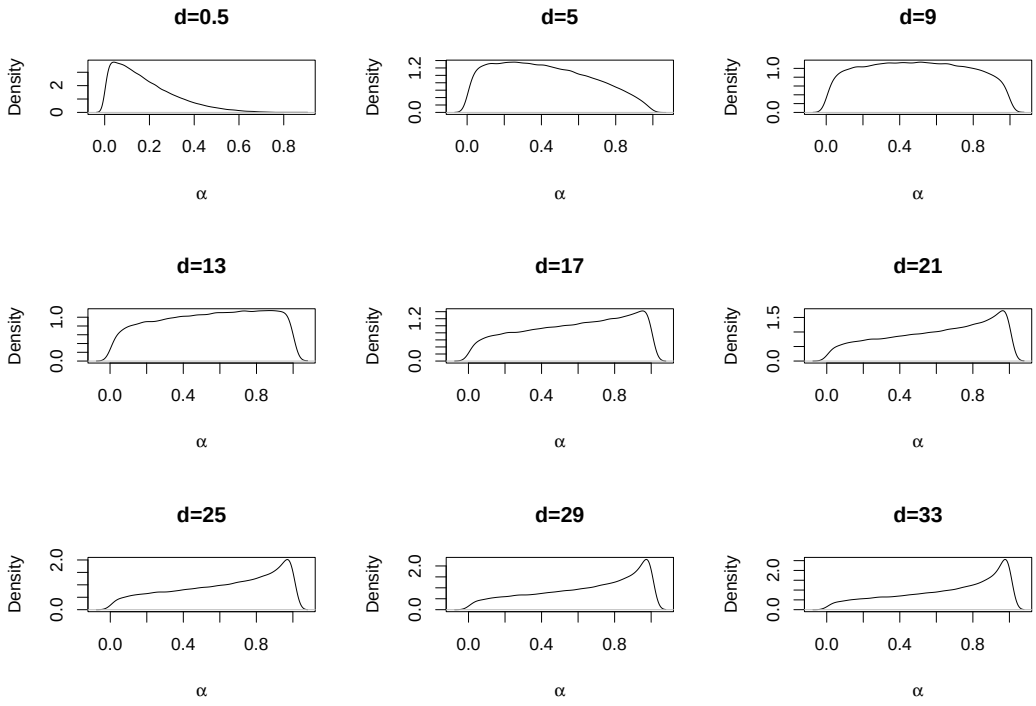


Figure 4.6: Density plots of α based on 100000 simulated values of τ, for $d = 0.5, 5, 9, 13, 17, 21, 25, 29, 33$

the Jacobian of the transformation. So,

$$\begin{aligned}\pi_\alpha(\alpha) &= \pi_\tau \times \left|\frac{d\tau}{d\alpha}\right| \\ &= \pi_\tau \times \frac{1}{d^{1/2}(1-\alpha)} \\ &= \frac{4^{1.2}}{d^{1/2}(1-\alpha)\Gamma(1.2)}\left(-\frac{\log(1-\alpha)}{d^{1/2}}\right)^{0.2} \\ &\quad \times \exp\left\{-4\left(-\frac{\log(1-\alpha)}{d^{1/2}}\right)\right\}. \end{aligned} \tag{4.12}$$

The joint prior, which has the form of (4.4), is a product of the two joint marginal priors, for sites X and Y, and the prior for α.

4.5.2 Posterior inference

The posterior distribution is

$$\pi(\theta|\boldsymbol{x},\boldsymbol{y}) \propto \pi_x(\mu_x,\sigma_x,\xi_x)\pi_y(\mu_y,\sigma_y,\xi_y)\pi_\alpha(\alpha)\prod_{i=1}^{n} g(x_i,y_i),$$

where g is the joint distribution function for the logistic model, given in (4.9).

The full conditionals for the two sets of GEV parameters have the same form as (3.14), (3.15) and (3.16), but with the new bivariate likelihood, $L(\theta|\boldsymbol{x},\boldsymbol{y}) = \prod_{i=1}^{n} g(x_i,y_i)$, replacing the univariate likelihood. The full conditional for α is

$$\pi(\alpha|\cdot) \propto \pi_\alpha(\alpha)\prod_{i=1}^{n} g(x_i,y_i),$$

where $\pi_\alpha(\alpha)$ is given in (4.12).

The MCMC algorithm is carried out in the same way as for the model with non-informative priors in 4.4. The proposal variances vary slightly for different pairs of sites but for the pair (1,2) the distributions are

$$\omega_{\mu_x} \sim N(0,3.2), \qquad \omega_{\sigma_x} \sim N(0,2.5), \qquad \omega_{\xi_x} \sim N(0,0.1),$$

$$\omega_{\mu_y} \sim N(0,3.2), \qquad \omega_{\sigma_y} \sim N(0,2.5), \qquad \omega_{\xi_y} \sim N(0,0.1),$$

$$\omega_\alpha \sim N(0,0.2).$$

The algorithm was carried out with data from each pair of sites for 30000 iterations. The sample trace plots, given in figure (4.7), are for the parameters of the model applied to the data from sites 1 and 2. Three chains of length 30000 are plotted, each starting with

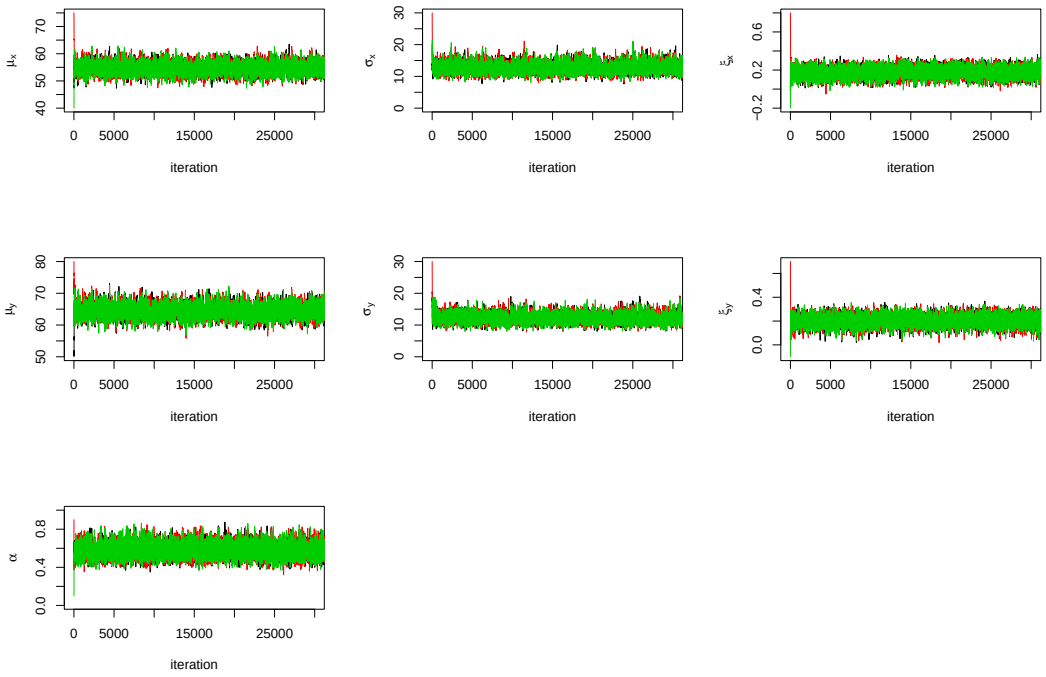


Figure 4.7: Trace plots of the seven parameters for the bivariate model applied to sites 1 and 2 with expert priors

different initial parameter values. For each parameter the three chains converge to the same place, confirming that convergence has been achieved. Convergence is achieved very quickly in this case and for all other pairs of sites. In each case the first 10000 iterations were removed as burn in. The remaining 20000 iterations were regarded as observations from the posterior distribution. Density plots based on the last 20000 observations are given in Figure 4.8 for the model applied to sites 1 and 2.

Posterior means and standard deviations for each parameter of the model applied to every pair of sites are given in tables A.3 and A.4 respectively. The results for the three pairs (4,5), (2,6) and (7,11) are given in Table 4.3.

The posterior means for the dependence parameter, α, follow the same pattern as

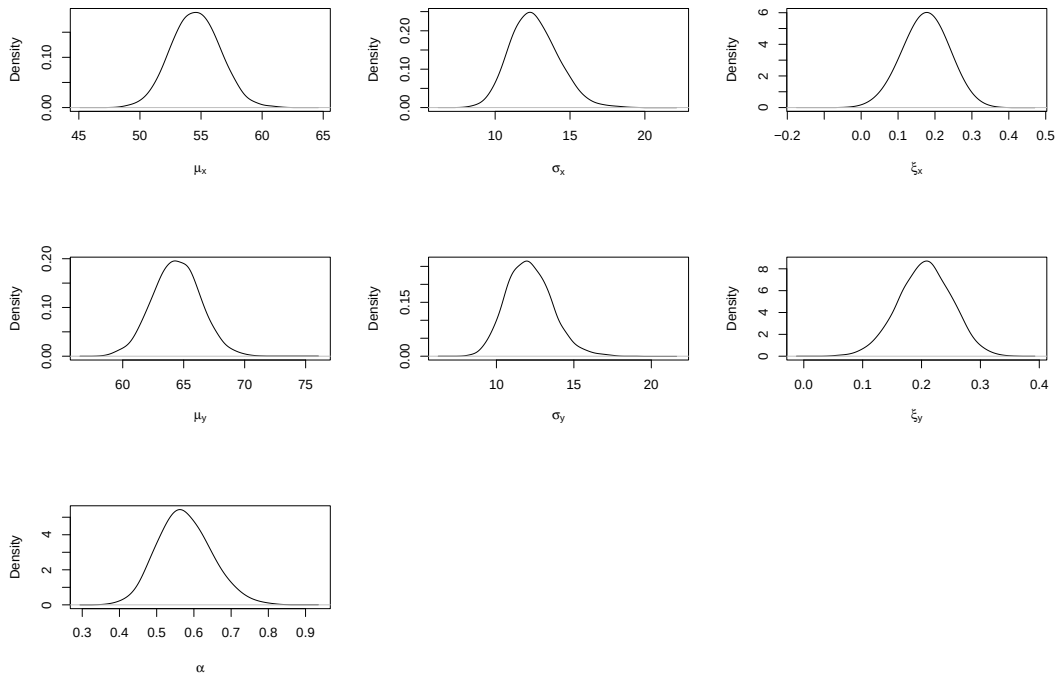


Figure 4.8: Posterior density plots of the seven parameters for the bivariate model applied to sites 1 and 2 with expert priors

	$i = 4,\ j = 5$		$i = 2,\ j = 6$		$i = 7,\ j = 11$	
Parameter	Mean	S.D.	Mean	S.D.	Mean	S.D.
μ_i	63.8581	1.7854	65.5976	1.9440	43.8489	1.2911
σ_i	12.6821	1.3019	11.0743	1.4447	8.5647	0.9389
ξ_i	0.1881	0.0448	0.2300	0.0460	0.2263	0.04939
μ_j	61.1411	1.7352	58.3755	1.4340	40.4026	1.3372
σ_j	12.5171	1.2413	7.9834	1.0706	8.7702	1.1896
ξ_j	0.1721	0.0459	0.2925	0.0449	0.2672	0.0506
α	0.3175	0.0405	0.6850	0.0829	0.9265	0.0636

Table 4.3: Posterior means and standard deviations for the bivariate model with informative priors applied to the pairs of sites (4,5), (2,6) and (7,11)

for the uninformative prior model: the two sites closest together (4,5) have the smallest posterior mean, indicating fairly strong dependence; the two furthest apart (7,11) have a high posterior mean, indicating weak dependence; and the pair with a distance separation in the middle of the range have an intermediate value for the posterior mean, indicating moderate dependence. Also, the plot of the posterior means of α against the distance separations, given in Figure 4.9, reveals approximately the same shape as for the uninformative prior model in Figure 4.4. In the next section, the results from the model with uninformative priors are compared in more detail with these results to examine the effect of the informative prior.

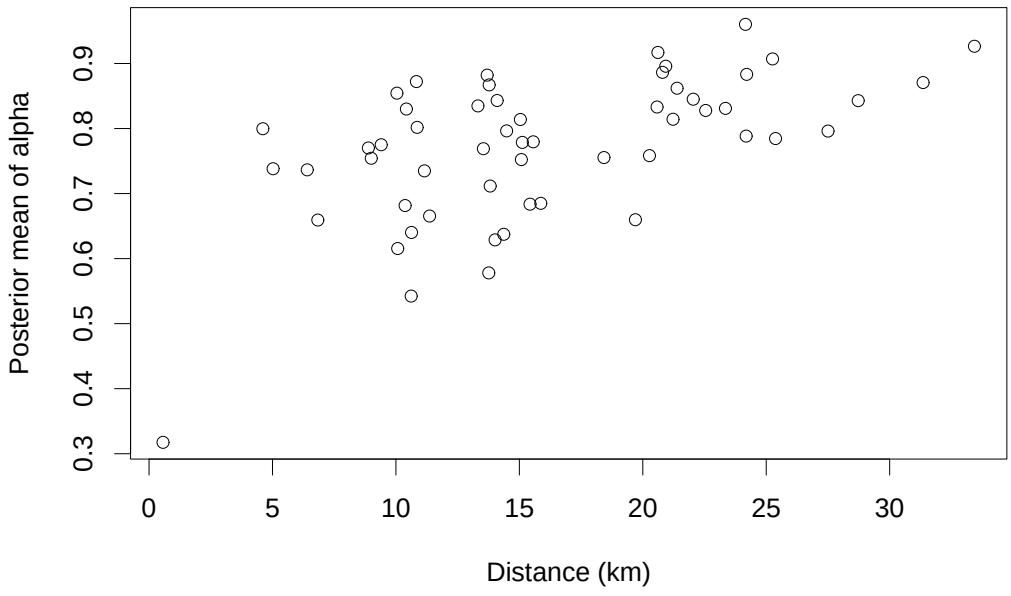


Figure 4.9: Plot of posterior mean of α (using informative priors) against distance between sites

Again, the vectors of observations from the marginal posterior densities of the GEV parameters can be substituted into (1.4) to give realisations from the posterior distribution of return levels. Posterior means, medians, SDs and IQRs of the 10, 100 and 1000-year return levels are given in Table 4.4, for site 1 based on the bivariate analysis of sites

1 and 2. Here, the difference between the posterior means and medians is very small, unlike for the model using the non-informative prior. This is due to there being much less uncertainty in the model with the informative prior, as the non-informative prior tends to allow high return levels which are very unlikely to occur.

	Return period (years)		
	10	100	1000
Posterior mean	89.42	144.6	228.4
Posterior median	89.21	144.0	225.6
Posterior SD	4.902	11.43	31.01
Posterior IQR	6.42	15.6	42.2

Table 4.4: Posterior means, medians, standard deviations (SDs) and inter-quartile ranges (IQRs) for the 10, 100 and 1000-year return levels for site 1, using the informative prior

4.6 The effect of the expert prior information

4.6.1 The pair (1,2)

As a detailed example, the results from the analyses of the pair of sites (1,2) with non-informative and informative priors are compared here. A more general comparison for all sites is given in the next section, followed by a closer look at the effect of the informative prior on the dependence parameter for different values of distance separation.

Since the informative prior is made up of independent priors for the two sets of GEV parameters and for α, plots of the prior densities of the GEV parameters would look the same as those given in Figure 3.8. So, again, it would be expected that the informative prior would have some effect on μ but more so on σ and ξ. For α, the non-informative prior is $U(0,1)$ and so can easily be compared with one of the prior densities given in

Figure 4.6, depending on the distance between the sites. For the pair of sites (1,2) the distance separation is 13.76km, so the informative prior distribution would look very much like the prior for $d = 13$ in Figure 4.6. This prior is rather similar to a $U(0, 1)$ distribution and is therefore a fairly non-informative prior. It would be expected, then, that in this case the informative prior would have very little effect on he posterior distribution of α.

Figure 4.10 shows plots of the expert prior and posterior densities for the pair (1,2). Most of the priors are almost flat in comparison to the posterior densities, so the data has

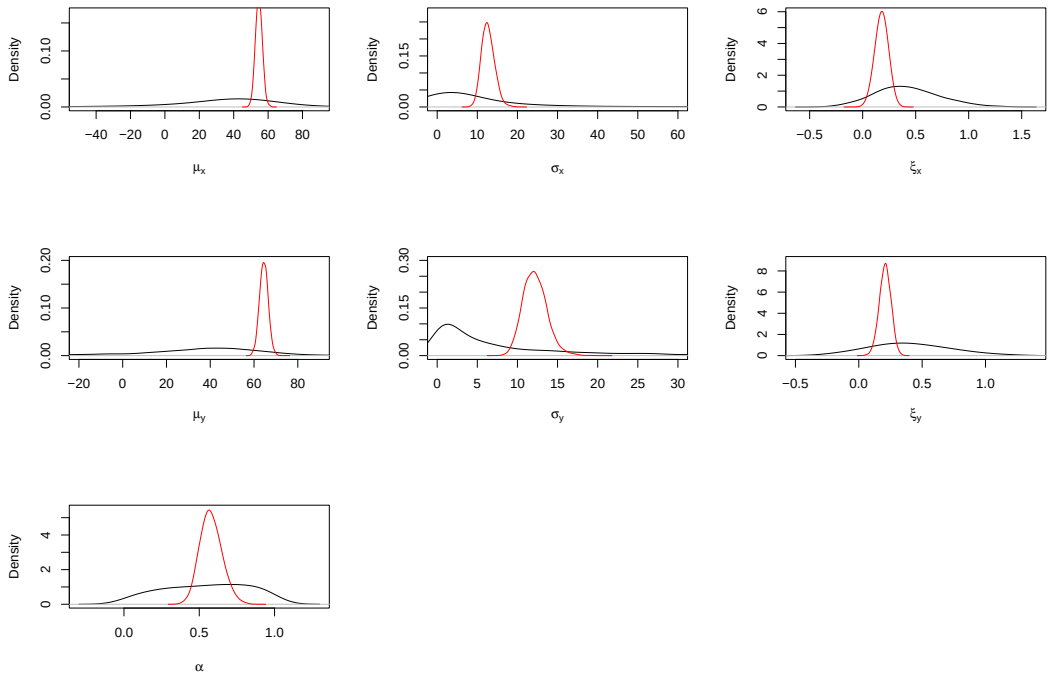


Figure 4.10: Prior (in black) and posterior densities (in red) of the seven parameters for sites 1 and 2

had a more significant impact on these posteriors than the prior has. This is particularly the case for the two location parameters, μ_x and μ_y. For ξ_x and σ_y, however, the priors are not quite as flat. The prior and posterior modes are fairly close for all of the parameters except for the two scale parameters: for both σ_x and σ_y the posterior mode appears to be quite a bit higher than the prior mode.

Figure 4.11 shows, for the pair of sites (1,2), the posterior densities of the parameters

using an uninformative prior (in black) and an informative prior (in red). The expert prior

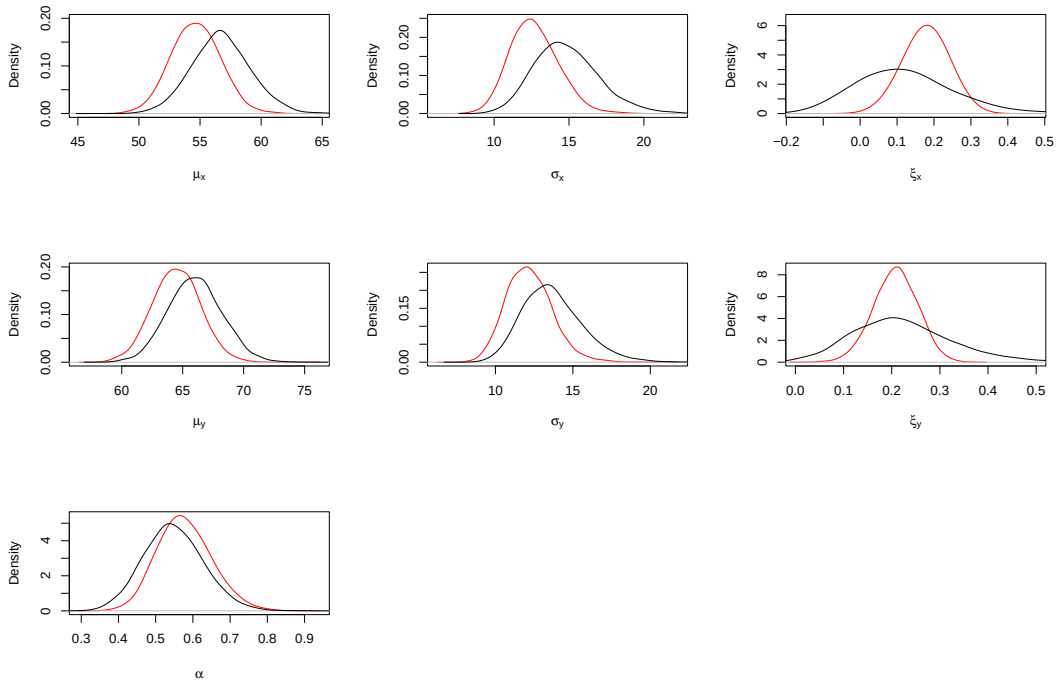


Figure 4.11: Posterior density plots of the seven parameters for the bivariate model applied to sites 1 and 2 with uninformative priors (in black) and expert priors (in red)

information used to formulate the priors on the margins does seem to have had an effect on the marginal posterior distributions. All of these posterior densities change significantly when this information is used. The posterior modes of the location and scale parameters are lower when the informative prior is used than when the uninformative prior is used. The posterior means are also lower for these parameters when the informative prior is used. For the location parameters, the shapes of the posterior densities are similar for both priors, although the densities have slightly narrower spreads and higher peaks when the informative prior is used. This reflects an increase in certainty, when the informative prior is used, which is backed up by the decrease in posterior standard deviations. The posterior standard deviation for μ_x goes down from 2.4788 to 2.0122 and the posterior standard deviation for μ_y goes from 2.2257 to 1.9480. The posterior densities of the shape

parameters also reflect a greater amount of certainty with the informative prior. The densities have higher peaks and are less spread out. Again, there is a reduction in the posterior standard deviations: the posterior standard deviation of σ_x goes from 2.2121 to 1.6180 and the posterior standard deviation for σ_y goes from 1.9141 to 1.4928. The shapes of the posterior densities for the shape parameters show the greatest difference between the two priors. For ξ_x, the posterior mode is higher when the informative prior is used, and the density has a higher peak due to the greater amount of certainty. For ξ_y, the two posterior modes look similar, but again there is more certainty in the model with informative priors, resulting in a higher peak. The decrease in posterior standard deviations reflects this great reduction in uncertainty: the posterior standard deviations go from 0.1351 to 0.0523 for ξ_x and from 0.1052 to 0.0457 for ξ_y. The informative prior has also had an effect on the posterior density of the dependence parameter but this effect is not as significant as the the effect on the other parameters. The posterior densities of the dependence parameter are fairly close for the two priors, although the posterior mode is slightly higher and the spread is slightly narrower for the model with the informative prior. The posterior mean increases slightly from 0.5477 to 0.5780 and the posterior standard deviation decreases from 0.0803 to 0.0720.

Autocorrelation plots of the seven parameters for the model with informative priors are given in Figure 4.12. A comparison of these autocorrelations with those of the uninformative prior model, given in 4.2, reveal that for this pair of sites the autocorrelations become insignificant much quicker when the informative prior is used. This is the case for all of the parameters but is particularly noticeable for the dependence parameter, α.

Plots of the posterior densities of the 10, 100 and 1000-year return levels, using both the non-informative and informative priors, are given in Figure 4.13 for both sites 1 and 2. These plots show that the informative prior information has clearly had an impact on the posterior densities of return levels. For both sites 1 and 2, and for all three return levels, the amount of uncertainty in the posterior has been reduced with the use of the informative prior. This reduction in uncertainty is most significant for the higher return

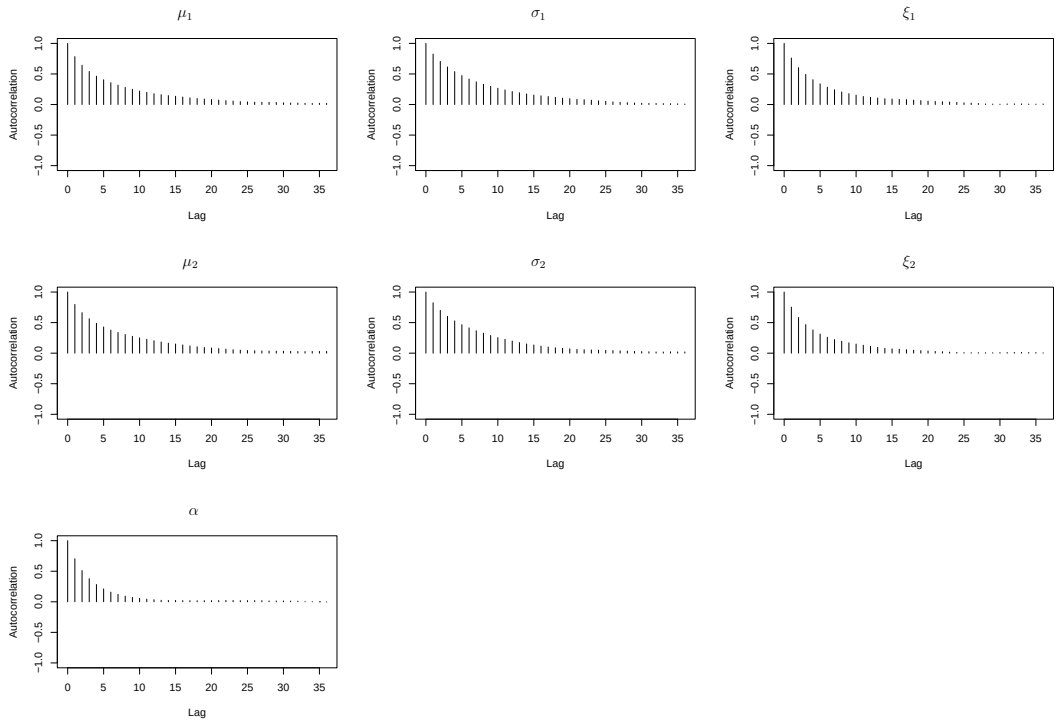


Figure 4.12: Autocorrelation plots for the parameters of the logistic model applied to sites 1 and 2 with informative priors

periods. For site 1, the posterior modes of $q_{0,01}$ and $q_{0.001}$ appear to have increased with the use of the informative prior, while the posterior mode of $q_{0.1}$ has decreased. For site 2, the posterior modes for the two priors appear to be very close for both $q_{0.01}$ and $q_{0.001}$ but again, the posterior mode of $q_{0.1}$ has decreased when the informative prior is used.

A comparison of the posterior medians obtained with the non-informative and informative priors, given in Table 4.2 and Table 4.4 respectively, shows that: the posterior median of $q_{0.1}$ is slightly lower when the informative prior is used; the posterior median of $q_{0.01}$ is slightly higher when the informative prior is used; the posterior median of $q_{0.001}$ is considerably higher when the informative prior is used. The posterior IQRs, also given in Table 4.2 and Table 4.4 for the non-informative and informative priors respectively, confirm that the variability has been reduced by the use of the informative prior: all of the posterior IQRs are lower when the informative prior is used, especially those for the higher return periods.

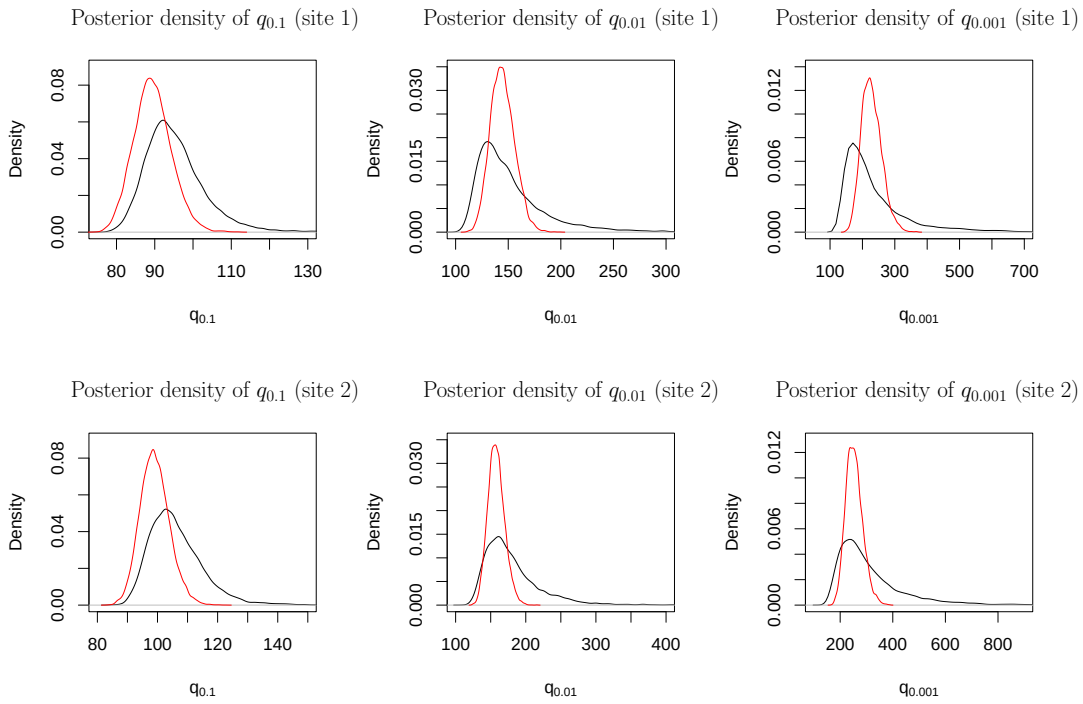


Figure 4.13: Posterior densities of the 10, 100 and 1000-year return levels for sites 1 and 2 based on bivariate analyses of sites 1 and 2 with both non-informative and informative priors

Return level plots based on the posterior medians, for the models with both the non-informative and informative priors, are given in Figure 4.14. Return level curves based on maximum likelihood estimates and on medians of the informative prior are also given together with the empirical estimates. This shows the effect of the informative prior on return levels in general. The return level curves based on the posterior medians with the two priors seem to be very close at low return periods and only start to show a difference at return periods over 100 years. There is a much greater difference between the maximum likelihood curve and the posterior medians curve with the informative prior. It is noticeable that in this bivariate analysis, the informative prior has had less of an effect on the return levels than in the univariate analysis of site 1: in the univariate analysis (see Figure 3.12), the difference between the two curves becomes clear at a much lower return period. For the high return periods, the posterior medians curve based on the informative

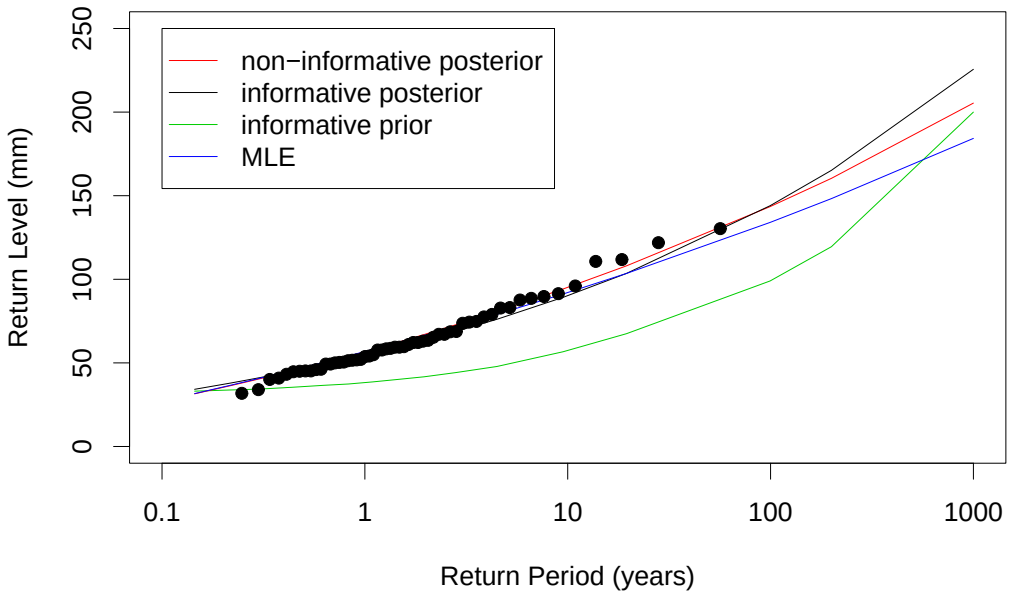


Figure 4.14: Plots of return levels against return period for site 1 based on the uninformative posterior distribution, the informative posterior distribution, the informative prior distribution and the maximum likelihood estimate

prior is not nested between the informative prior curve and the posterior curve with non-informative priors. This is the same effect as was observed in the univariate analysis. At lower return periods, the posterior medians curve with informative priors is lower than the one with non-informative priors, but the two are very close. The two curves based on posterior medians and the maximum likelihood curve are in close agreement to the empirical estimates but the two based on the Bayesian analysis appear to be in better agreement. Up to the point of the most extreme empirical estimate there is very little difference between the two posterior curves, but since the posterior curve based on the informative prior is higher for high return periods, it would be wise to consider this curve for prediction purposes.

Figure 4.15 shows the return level curve based on the posterior medians obtained using the informative prior, with 95 % credibility intervals based on both non-informative and

informative priors. As expected, the credibility interval based on the informative prior is

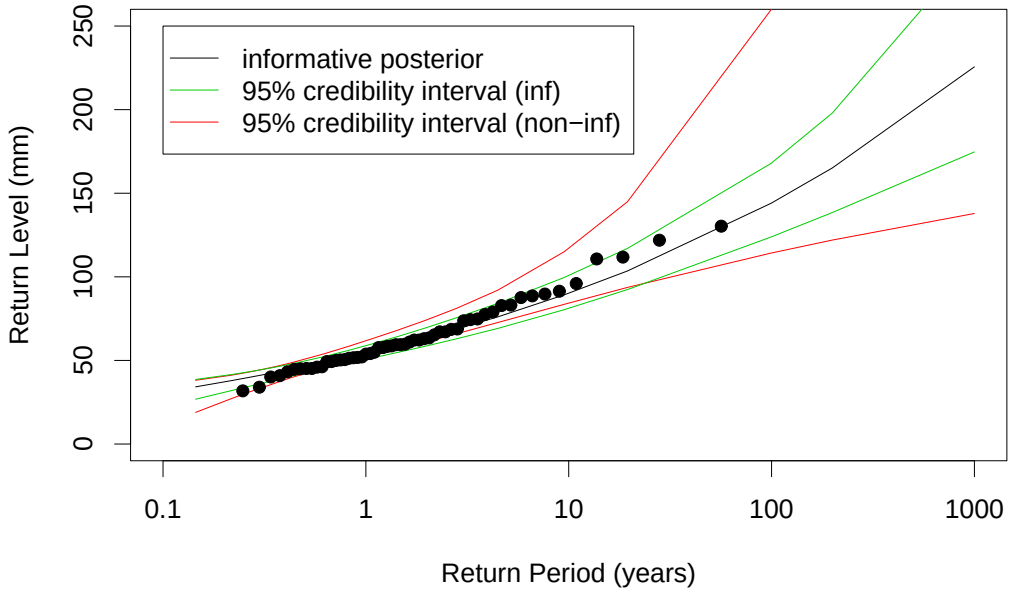


Figure 4.15: Plot of return level against return period for site 1 based on the informative posterior distribution with 95% credibility intervals based on both the informative prior and the non-informative prior

much narrower than the interval based on the non-informative prior. This confirms that the informative prior has greatly reduced the variability in the posterior distribution of return levels. The empirical estimates all lie in the interval based on the informative prior except for one, which appears to be very close to the upper bound.

4.6.2 Overall effect

Figures 4.10 and 4.11 show the effect of the informative prior on the model parameters for the pair of sites (1,2). Plots have not been produced for every pair, but the type of effect is similar for all pairs.

Comparing the posterior means from the model with non-informative priors (Table

A.1) with those from the model with informative priors (Table A.3) reveals the following: all of the posterior means for the location parameters, μ_x and μ_y, are lower when the informative prior is used; most of the posterior means for the scale parameters, σ_x and σ_y, are lower when the informative prior is used; for the shape parameters, ξ_x and ξ_y, some of the posterior means are higher and some are lower when the informative prior is used; for the dependence parameter, α, the majority of the posterior means increased when the informative prior was used. Also, all of the negative posterior means for the shape parameters obtained with the non-informative prior were found to be positive and much greater when the informative prior was used. For example, for the pair (2,5) the posterior mean for ξ_y was found to be -0.0561 when uninformative priors were used and 0.1645 when informative priors were used. In this case, the standard deviation decreased from 0.1239 to 0.0525, suggesting reduced uncertainty. It seems, then, that the negative posterior mean obtained under the non-informative prior can be attributed to a large amount of uncertainty. After taking into account the expert information, then, a positive value for ξ_y seems much more likely.

The posterior standard deviations for the models with non-informative priors, given in Table A.2 and Table A.4 respectively, show that most of the posterior standard deviations are lower when the informative prior is used. The exceptions are for some of the location parameters, μ_x and μ_y, and some of the dependence parameters, α. All of the pairs for which this occurs include either site 7 or site 8. The differences between the posterior standard deviations for the two priors in these cases, though, is very small. It seems, then, that for these two sites the prior and the data may be in conflict, resulting in increased uncertainty for some parameters. Comparing the data from sites 7 and 8 with the data from the other sites, it does appear that in general the rainfall levels are lower at these two sites than at the other sites. Both sets of data also have one observation, from 1969, which is much higher than the others. This could be the cause of the conflict between the data and the prior resulting in the increased standard deviations. It was noted in Section 3.5.1 that the informative marginal prior for μ is not much more informative than the non-

informative prior: although the posterior mode is shifted, there is still a lot of uncertainty. This could contribute to making the posterior standard deviations of μ remain large when the informative prior is used. The posterior standard deviation of μ for site 7 obtained in the univariate analyses was also slightly higher when the informative prior was used than when the non-informative prior was used. Additionally, in the univariate analysis, the posterior standard deviation of μ for site 10 was higher when the informative prior was used than when the non-informative prior was used. In the bivariate analyses, however, there were no occurrences of higher posterior standard deviation for for μ of site 10 under the informative prior.

The return level plots in Figure 4.14 and Figure 4.15 show the effect of the informative prior on the return levels for site 1 based on a bivariate analysis of sites 1 and 2. Although summary statistics of the return levels and return level plots are not given for all sites and all pairs modelled, the fact that the informative prior has impacted upon the model parameters means that it also impacts upon the return levels. The effect displayed in Figure 4.14 and Figure 4.15 is typical of the effect for both sites of all pairs modelled: the curves based on the informative prior tend to be higher with narrower credibility intervals than those based on the non-informative prior.

4.6.3 Effect on the dependence parameter

The dependence parameter is of particular interest as it indicates how much dependence there is between any pair of sites. The informative prior used for the dependence parameter had the distance between sites as a covariate.

To compare the effect of the prior on the dependence parameter at different distance separations, the results for three pairs of sites are given. The posterior means and standard deviations for the uninformative prior and informative prior models were given in Tables 4.1 and 4.3 respectively. A comparison of the posterior means of α for the two priors reveals that the same pattern is evident for both priors: the two sites closest together have a low posterior mean; the two sites furthest apart have a high posterior mean; the

two sites with an intermediate distance separation have a moderate posterior mean. The informative prior does seem to have had some effect on these values though. For the pair (4,5), the posterior mean has gone down from 0.3342 to 0.3175. For the pair (2,6), the posterior mean has increased from 0.6056 to 0.6850. For the pair (7,11), the posterior mean has increased from 0.9076 to 0.9265. These changes are not surprising, since the priors of α, plotted in Figure 4.6 for various values of d, suggest lower values of α are more likely at small distance separations and high values are more likely at greater separations. This is quite a significant change from the completely flat priors used previously.

Figure 4.16 shows posterior density plots for α, using both non-informative and informative priors, for each of the three pairs of sites, together with the informative prior density. In all three case, the expert prior density appears completely flat in comparison to the posterior densities. Despite this, it is clear that the informative prior has affected the posterior distribution. In each case, the posterior density has a higher peak when the informative prior is used. This increase in certainty is also reflected by the decrease in the posterior standard deviations of α. For the first pair, (4,5), the posterior mode shifts very slightly to the left when the informative prior is used. The posterior mode for the second pair, (2,6), is slightly higher when the informative prior is used. For the third pair, (4,11), the posterior mode is slightly higher when the informative prior is used. As with the changes in posterior means, these changes in the posterior densities are not surprising. The informative prior suggests stronger dependence at sites close together than the uninformative prior does, and weaker dependence at sites further apart.

Plots of the posterior mean of the dependence parameter α against the distance between sites d, for both uninformative and informative priors, are given in Figure 4.17. It is clear from this plot that, although the informative prior does have an effect on the posterior means, the relationship between the dependence parameter and the distance separation is very similar for the two priors.

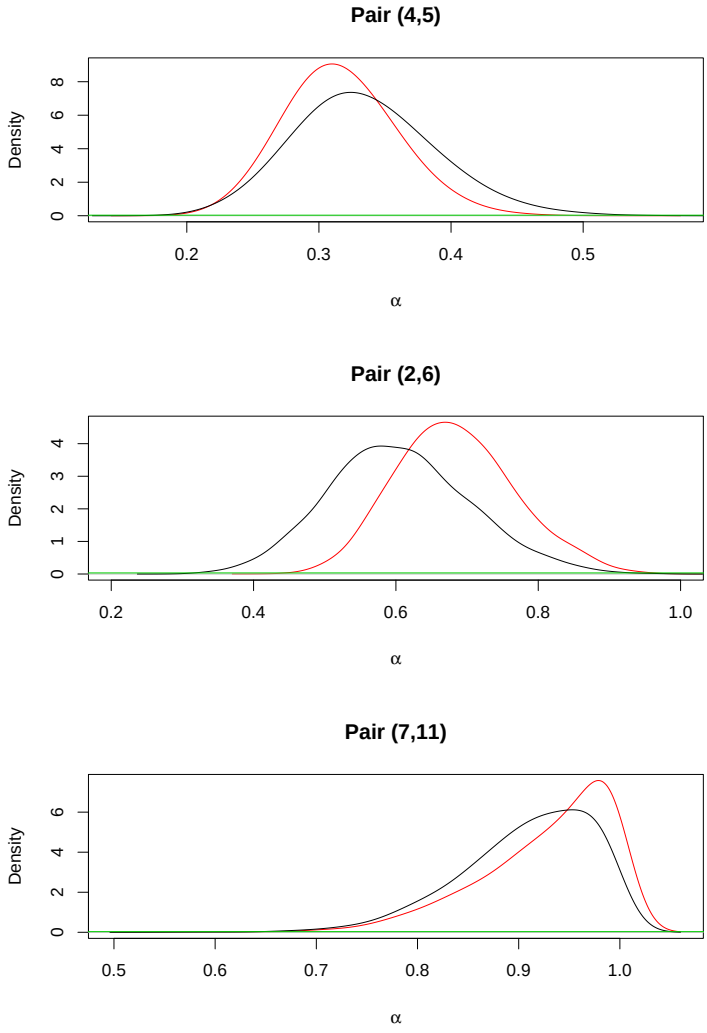


Figure 4.16: Plots of the prior density of α (in green), the posterior density of α with uninformative priors (in black) and the posterior density of α with informative priors (in red) for the three pairs of sites (4,5), (2,6) and (7,11)

4.7 Effect of using a bivariate model

It is clear that the informative prior has had an effect on the posterior distributions of the parameters and hence on the posterior distributions of return levels. Modelling two sites simultaneously, rather than modelling each site individually, is also likely to have an effect on the posterior distributions of return levels. Using a bivariate model to model two

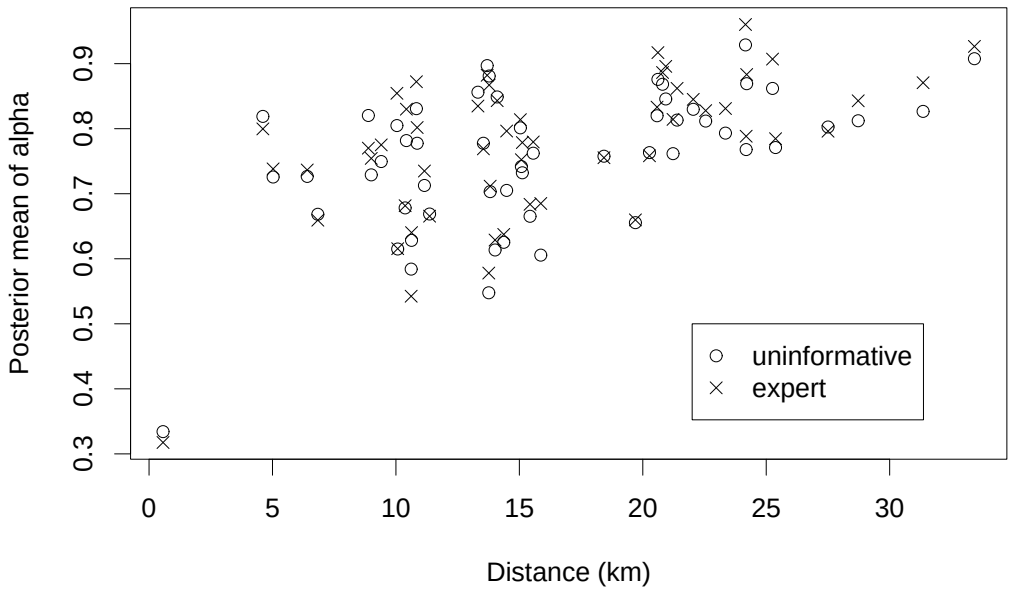


Figure 4.17: Plot of posterior mean of α against distance between sites for models with uninformative and expert priors

sites together allows information to be pooled across sites, resulting in better inference.

To examine the effect of using the bivariate model, return level plots for sites 1 and 2 based on both univariate and bivariate models are given in Figure 4.18 and Figure 4.19 respectively. The plots have been produced using the posterior medians of the return levels and the results for the bivariate model are those of the model applied to sites 1 and 2 simultaneously.

For site 1 there is very little difference between return level curves for low return periods. For higher return periods the curve based on the bivariate model is lower but the difference appears quite small. For site 2 the difference appears to be much greater. At very small and very large return periods the bivariate curve is considerably higher. At return periods in the middle of the range the two lines are close but the bivariate curve is always higher. So, although the difference was not great for site 1, the use of a bivariate model can result in significant differences in the posterior distributions. It

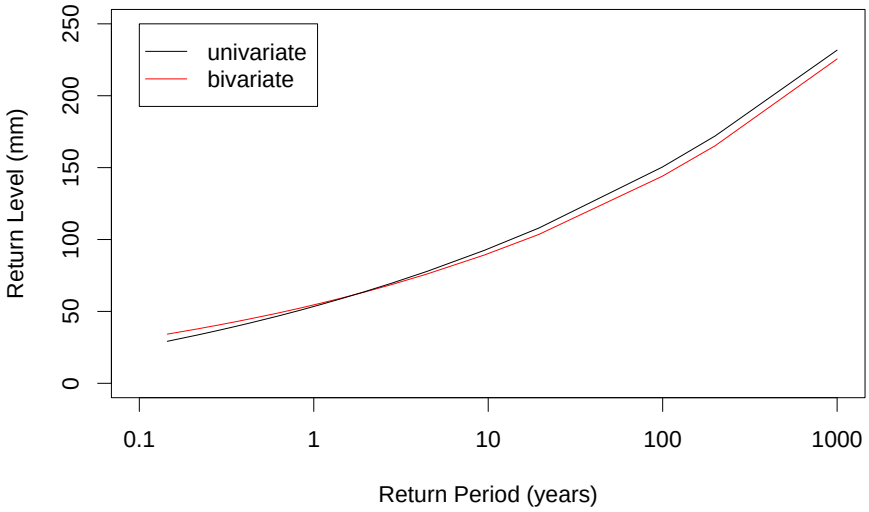


Figure 4.18: Plots of posterior medians of return levels against return period for site 1 based on the informative prior distribution for the univariate and bivariate models

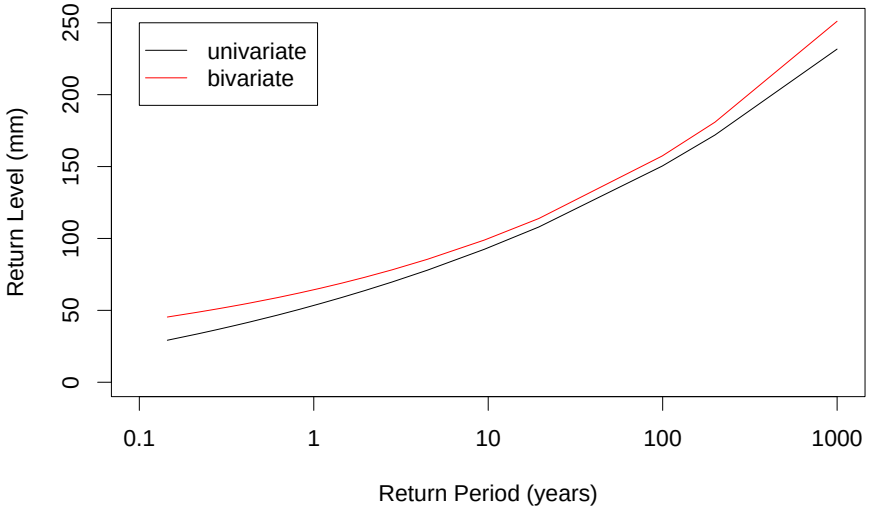


Figure 4.19: Plots of return levels against return period for site 2 based on the informative posterior distribution for the univariate and bivariate models

seems, therefore, to be worthwhile considering models for multivariate extremes of higher dimension.

4.8 Prediction

Predictive return levels can be calculated in the same way as in section 3.6. The predictive return levels for site 1, after applying the bivariate model to sites 1 and 2, are given in Table 4.5. These can be compared with the posterior means and medians obtained from applying the Bayesian model to sites 1 and 2, which are given in Table 4.2 and Table 4.4 for the non-informative and informative priors respectively. The same effect observed in

	Return period (years)		
	10	100	1000
Non-informative prior	95.178	155.167	287.630
Informative prior	89.690	145.375	232.121

Table 4.5: Predictive return levels for site 1 for the 10, 100 and 1000-year return periods, using the bivariate model on sites 1 and 2

the univariate analysis in section 3.6 is observed here. With the non-informative prior the predictive return levels are higher than the posterior medians for all three return periods, and are higher than the posterior means for the 100 and 1000-year return levels. For the return period of 10 years the predictive return level is close to both the posterior mean and median. With the informative prior the predictive return levels are higher than the posterior medians and means for all three return periods. As was observed in the univariate analysis, the distance between predictive return level and posterior median of return level increases with return period, and the distances are greater when the non-informative prior is used than when the informative prior is used.

Predictive return level plots based on both priors are given in Figure 4.20. The return level curves based on the posterior medians of return levels are also given. The pattern

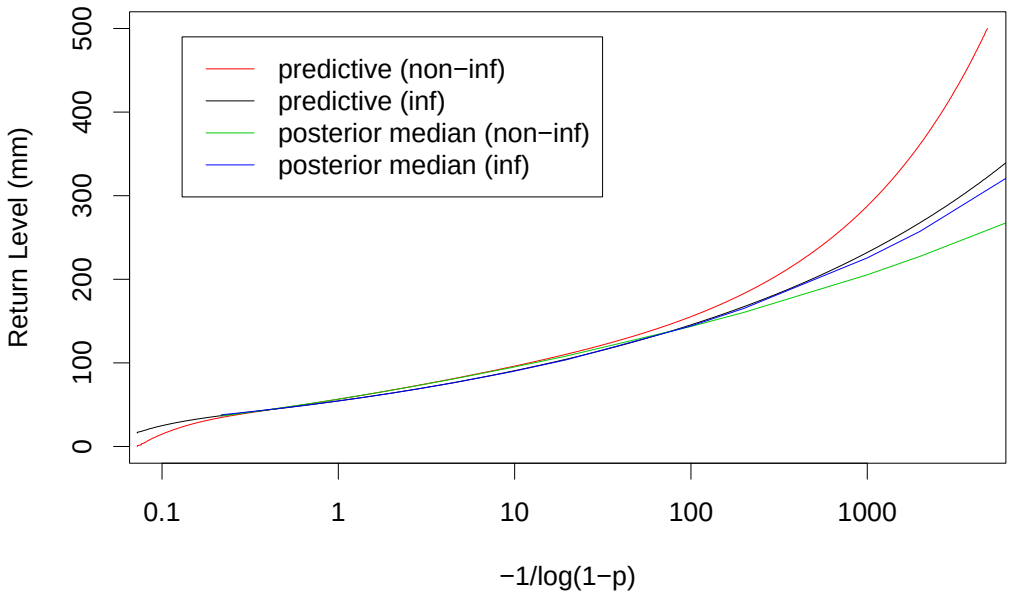


Figure 4.20: Predictive return level plots and return level plots based on the posterior medians of return level for site 1, using both non-informative and informative priors for the bivariate model applied to sites 1 and 2

of curves displayed is very similar to the corresponding pattern of curves based on the univariate model, given in Figure 3.14. There is, however, one noticeable difference: the predictive return level curve based on the non-informative prior extends to much higher return levels in the bivariate model than in the univariate model.

For both priors, the predictive return level curves are above the respective posterior median curves at high values of $-1/\log(1-p)$. For values of $-1/\log(1-p)$ below 100 the differences between the curves are only very slight. The difference between the predictive return level curve and the posterior median curve is much greater for the model with the non-informative prior than for the model with the informative prior. In fact, with the informative prior there is very little difference at all between the two curves. With the non-informative prior it appears that using the posterior medians of return levels for design purposes would lead to substantial under protection. When using the informative

prior there may be some under protection.

The use of an informative prior has had a considerable effect on the predictive return levels. The predictive return level curve based on the non-informative prior is much higher than that based on the informative prior at high values of $-1/log(1-p)$. The two curves for the informative prior are very close, suggesting that using the informative prior has greatly reduced the uncertainty in the model. It seems that, due to uncertainty, the predictive return levels are over estimated when the non-informative prior is used.

4.9 Discussion

In this chapter the Bayesian model for univariate annual maxima data, used in Chapter 3, was extended to a bivariate model which was then used to model the annual maxima data from each pair of sites.

Two models for dependence were considered; the logistic model and the mixed model. These models were applied to the data from sites 1 and 2 with non-informative priors on both margins and on the dependence parameter. The logistic model was found to have the capability of modelling the levels of dependence expected, whereas the mixed model was found to be too restrictive. The logistic model, then, was used to model the data from all pairs of sites, with both non-informative and informative priors.

Informative priors on the two margins were based on expert prior information, and the informative prior on the dependence parameter was based on an assumed relationship between dependence and the distance between sites.

Posterior means and standard deviations for the seven parameters of each site were obtained after using both non-informative and informative priors. For both priors there appeared to be a relationship between the dependence parameter and the distance between the two sites being modelled: strong to moderate dependence was often observed for sites close together and weak dependence for sites further apart. This confirmed the belief that dependence is related to distance separation but it was also apparent that

dependence was not determined by the distance separation alone. A comparison of the results obtained with non-informative and informative priors revealed that the informative priors did significantly affect the posterior distributions of all parameters. In general, the posterior standard deviations for the models with informative priors were lower than those for the models with non-informative priors.

Posterior distributions of the 10, 100, and 1000-year return levels for site 1 based on the analysis of sites 1 and 2 were obtained by transforming the samples of the marginal GEV parameters for site 1. The posterior medians and IQRs were chosen as appropriate measures of location and spread respectively due to the distributions being skewed. Return level curves for posterior medians based on the non-informative and informative priors revealed that estimates of return levels tended to be higher and credibility intervals narrower when based on the informative prior. These results were compared with those obtained in the univariate analysis in Chapter 3 to determine the effect of using a bivariate model, rather than a univariate model, on the posterior distribution of return levels. This comparison showed that using the bivariate analysis did, in some cases, have a significant effect on the return levels.

Predictive return levels for site 1 based on the bivariate model for sites 1 and 2, with both non-informative and informative priors, were obtained. The informative prior had more of an effect on the predictive return levels than on the posterior medians of the return levels: the predictive return levels based on the non-informative prior are much higher than those based on the informative prior for high values of $-1/\log(1-p)$. There is also a big difference between the predictive return level curve based on the bivariate model and that based on the univariate model when the non-informative prior is used: for the bivariate model the curve extends to much higher return levels.

This chapter has demonstrated the importance of using expert prior information and of using a multivariate model. The information from both the informative priors and from pooling information across sites has impacted upon the inference and has reduced uncertainty. In this chapter, estimates of return levels and predictive return levels of the

rainfall process at each site were obtained. In practice, however, the extremal behaviour of some combination of the variables $Z = \phi(M_x, M_y)$ may be of more more interest than the variables themselves. If this is the case then estimates of the return levels of the *structure variable* Z could be found.

Chapter 5

Multivariate modelling

5.1 Introduction

This work extends the Bayesian approach for modelling bivariate annual maximum rainfall data to model daily rainfall data at all 11 sites simultaneously. Using only the annual maxima is inefficient as much extreme data is wasted. To model daily data, the *multivariate point process method*, described in section 1.6.2, is used. This method was used by Coles and Tawn (1991) in a frequentist analysis of oceanographic data from three locations in England. A frequentist multivariate analysis of extremes from all 11 sites of Figure (2.1) was carried out by Coles and Tawn (1996b), demonstrating the importance of dependence across sites when modelling extreme rainfall in this region. Modelling the rainfall at all 11 locations simultaneously means that not only bivariate dependencies are being considered, but potentially all higher order dependencies.

Section 5.2 outlines the procedure for modelling with the multivariate logistic and using the multivariate point process method. In Section 5.3, after accounting for non-stationarity and dependence, the model is applied to the daily rainfall data from the 11 sites, using a Bayesian approach with non-informative priors on all parameters. Posterior inferences for the joint extreme behaviour are made through the use of an MCMC scheme. Informative priors on the marginal GPD parameters, based on the expert prior information

given in Section 3.4, are then used in Section 5.4. Again, posterior inferences are made through the use of an MCMC scheme. The results obtained with the non-informative and informative priors are then compared in Section 5.5, to assess the the effect the informative prior has on posterior inference. In Section 5.6, the posterior inference is taken further by considering the predictive distribution of return levels for each site. The results obtained throughout the chapter are then discussed in Section 5.7.

5.2 Modelling

Given a sequence of IID random vectors with the margins having identical unit Fréchet distributions, the multivariate point process theory of Section 1.6.2 can be applied. Parametric sub-families of the measure H, satisfying the condition 1.15, are needed since there is no finite parameterisation. This section describes how H is modelled and gives the function $h(.)$ for the multivariate logistic model, which is considered in this chapter. The likelihood equation for modelling the marginal and dependence parameters simultaneously, derived by Coles and Tawn (1991), is also given.

5.2.1 Modelling the measure H

Coles and Tawn (1991) prove a theorem which relates the exponent measure V to the measure H for the class of models for which V is differentiable. This theorem is given below, as Theorem 5.1.

For each $j = 1, \ldots, p$, let $c = \{i_1, \ldots, i_j\}$ be an index variable over the subsets, of size j, of the set $c_p = \{1, \ldots, p\}$ and let $S_{j,c} = \{\boldsymbol{w} \in S_p : w_k = 0, k \notin c\}$. For each c with $|c| = j$, $S_{j,c}$ is isomorphic to the unit simplex S_j defined by (1.14). On each subspace $S_{j,c}$ the measure has density $h_{j,c}$. The domain of $h_{j,c}$ is taken as either $S_{j,c}$ or S_j, as is convenient. The dependence structure for events which are extreme only in the components $c = \{i_1, \ldots, i_j\}$ is described by the density $h_{j,c}$. This provides a hierarchy of densities which is needed to describe dependence when all the marginal components are

not necessarily extreme simultaneously.

Theorem 5.1 *Let V and H be the exponent measure and measure function defined by (1.13), and let $h_{j,c}$ be the class of densities defined above. Then for $c = \{i_1, \ldots, i_m\}$*

$$\frac{\partial V}{\partial x_{i_1} \ldots \partial x_{i_m}} = -\left(\sum_{j=1}^{m} x_{i_j}\right)^{-(m+1)} h_{m,c}\left(\frac{x_{i_1}}{\sum_j x_{i_j}}, \ldots, \frac{x_{i_m}}{\sum_j x_{i_j}}\right) \tag{5.1}$$

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

5.2.2 The multivariate logistic model

The bivariate logistic model used in Chapter 4 can be extended to a multivariate model where the exponent measure is

$$V(\boldsymbol{x}) = \left(\sum_{j-1}^{p} x_j^{-1/\alpha}\right)^{\alpha}, \qquad 0 < \alpha < 1.$$

Applying (5.1) gives

$$h_{p,c_p}(\boldsymbol{w}) = \left\{\prod_{j=1}^{p-1}\left(\frac{j}{\alpha} - 1\right)\right\}\left(\prod_{j=1}^{p} w_j\right)^{-(1/\alpha+1)}\left(\sum_{j=1}^{p} w_j^{-1/\alpha}\right)^{\alpha-p}, \qquad \text{for } w \in S_p \tag{5.2}$$

and $h_{j,c} \equiv 0$ for $j < p$.

In this chapter the multivariate logistic model is the only model considered. It does have some disadvantages, however, so other models are considered in Chapter 6.

5.2.3 The likelihood

Using the point process theory described in section 1.6.2, Coles and Tawn (1991) derive the likelihood for a sequence of i.i.d. random vectors, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, with unit Fréchet margins. Let $P_n = \{n^{-1}\boldsymbol{X}_i : i = 1, \ldots, n\}$ and take $\{n^{-1}\boldsymbol{X}_i, i = 1, \ldots, n_A\}$ to be the points of P_n in a region A, bounded away from $\mathbf{0}$ by a distance dependent on the rate of convergence. These points are approximately a non-homogeneous Poisson process with an intensity measure satisfying (1.15). The likelihood over the region A is given by

$$L_A(\theta; \{n^{-1}\boldsymbol{X}_i\}) = exp\{-\mu(A)\}\prod_{i=1}^{n_a} \mu(dr_i \times d\boldsymbol{w}_i) \tag{5.3}$$

where θ are the measure parameters.

In general, the margins do not have unit Fréchet distribution. In chapter 4 estimation of the marginal and dependence parameters was carried out simultaneously. The marginal parameters could have been estimated first and then, after transforming to unit Fréchet margins using these estimates, the bivariate model could have been applied to estimate the dependence parameter. By estimating all of the parameters in one step, information can be transfered across variables, resulting in more efficient estimation. Coles (1991) has shown that most of the loss of efficiency when using the two step approach is in the estimation of the shape parameters.

Coles and Tawn (1991) provide a method for estimating the parameters of the point process model in one step. They do this by including transformations of the margins to the unit Fréchet distribution in the likelihood equation (5.3), and by an appropriate choice of the region A.

Suppose $\{\boldsymbol{Y}_i, i = 1, \ldots, n\}$ are i.i.d. random vectors on $\boldsymbol{R}^p$. Let $A = \boldsymbol{R}^p_+ \setminus \{(0, \nu_1) \times \ldots \times (0, \nu_p)\}$, where $\nu_1, \ldots, \nu_p$ are the thresholds. This region contains all observations which exceed the threshold of at least one margin and the points in the region are invariant to the marginal transformations. To transform points above a high threshold, u_j, the conditional distribution of threshold exceedances is used. It was shown by Pickands (1975) that this conditional distribution is of generalised Pareto form: $Pr\left(Y_j > y | Y_j > u_j\right) = \{1 + \xi_j(y - u_j)/\sigma_j\}^{-1/\xi_j}$, $\sigma_j > 0$, $-\xi_j(y - u_j)/\sigma_j < 1$. So for $Y_j > u_j$

$$Pr(Y_j > y) = p_j \left\{1 + \xi_j(y - u_j)/\sigma_j\right\}^{-1/\xi_j},$$

where $p_j = Pr(Y_j > u_j)$ is estimated as the proportion of points exceeding u_j. The empirical distribution function is used to transform points below the thresholds. Let $R(Y_j)$ denote the rank of Y_j, then

$$X_j(Y_j) = \begin{cases} -(\log[1 - p_j\{1 + \xi_j(Y_j - u_j)/\sigma_j\}^{-1/\xi_j}])^{-1} & \text{if } Y_j > u_j, \\ -[\log\{R(Y_j)/(n+1)\}]^{-1} & \text{if } Y_j \leq u_j \end{cases} \tag{5.4}$$

has a unit Fréchet distribution and the limiting process has thresholds $\nu_j = n^{-1} X_j(u_j)$.

The likelihood is obtained by incorporating the transformations (5.4) and the intensity measure (1.18) into (5.3):

$$L_A(\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\xi}; \{\boldsymbol{Y}_i\}) = \exp\{-V(\boldsymbol{v})\} \prod_{i=1}^{n_A} \Bigg(h(\boldsymbol{w}_i)(nr_i)^{-(p+1)} \times \prod_{\substack{j=1,\ldots,p: \\ X_{i,j} > n\nu_j}} [\sigma_j^{-1} p_j^{-\xi_j} X_{i,j}^2 \exp(1/X_{i,j})\{1 - \exp(-1/X_{i,j})\}^{1+\xi_j}] \Bigg). \tag{5.5}$$

To obtain the likelihood for the multivariate logistic model the function $h(.)$ given in equation (5.2) is substituted into (5.5).

5.3 Application to rainfall data using non-informative priors

Here, the multivariate point process method is applied to the daily rainfall data from all 11 sites simultaneously using a non-informative prior. The models of Chapters 3 and 4 considered only annual maxima data, so non-stationarity and temporal dependence did not need to be dealt with. With the daily rainfall data, however, it is necessary to account for these characteristics of the data. After obtaining an independent, stationary sequence of vectors the multivariate point process is applied with non-informative priors, using an MCMC scheme.

5.3.1 Accounting for non-stationarity and dependence

In Chapter 2 stationary, independent data were obtained by taking only the data from November to February and the maxima over two-day periods. Here, the data from November to February can also be used to remove seasonality but the issue of temporal dependence requires more care for a multivariate process.

When considering a process such as rainfall at more than one location it is necessary to consider the time for storm propagation. Coles and Tawn (1991) apply a univariate

procedure similar to that of Davison and Smith (1990) and then concatenate the maxima of each constituent process within a cluster allowing for a time lag between the variables.

In an analysis of this data by Coles (1994) it was found that the meteorology of the region tends to result in extreme conditions occurring at all sites simultaneously. This suggests that the two-day period used in the univariate analysis in Chapter 2 is probably sufficient to account for storm propagation over the region. Here, then, the maxima over two-day periods are obtained and concatenated to obtain a stationary multivariate process. The resulting data has 1451 observations of which 295 are extreme in at least one margin, where the thresholds are taken as those used for the univariate analysis in Chapter 2.

5.3.2 Prior choice

When working with the annual maxima model, the parameterisation $\phi_j = \log \sigma_j$, $j = 1, \ldots, 11$, was used in order to specify the non-informative priors. For the threshold method, however, a different parameterisation is used. The scale parameters σ_j are dependent on the choice of threshold u_j, unless $\xi_j = 0$. This means that a non-informative prior for σ_j becomes informative at other thresholds. To avoid this problem, the scale parameters are reparameterised as $\sigma_j^* = \sigma_j - \xi_j u_j$, $j = 1, \ldots, 11$.

The parameters σ_j^* and ξ_j, $j = 1, \ldots, 11$, are given independent univariate normal priors with zero means and large variances:

$$\sigma_j^* \sim N(0, 100) \qquad \xi_j \sim N(0, 10).$$

These priors are denoted by $\pi_{\sigma_j^*}(.)$ and $\pi_{\xi_j}(.)$ for σ_j^* and ξ_j respectively. The prior on the dependence parameter α is, as for the bivariate model in Chapter 4, a uniform prior with parameters 0 and 1. The joint prior, then, is

$$\pi(\boldsymbol{\sigma}^*, \boldsymbol{\xi}, \alpha) = \prod_{j=1}^{11} \pi_j(\sigma_j^*, \xi_j), \tag{5.6}$$

where $\pi_j(\sigma_j^*, \xi_j) = \pi_{\sigma_j^*}(\sigma_j^*)\pi_{\xi_j}(\xi_j)$ is the joint marginal prior for site j, and $\boldsymbol{\sigma}^*$ and $\boldsymbol{\xi}$ are the vectors of the parameters σ_j^* and ξ_j respectively, for $j = 1, \ldots, 11$.

5.3.3 Posterior inference

The posterior distribution has the form

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{11}) \propto \pi(\boldsymbol{\sigma}^*,\boldsymbol{\xi},\alpha) \times L_A(\boldsymbol{\sigma}^*,\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}), \tag{5.7}$$

where $L_A(.)$ is the likelihood given in (5.5) in terms of $\boldsymbol{\sigma}* = \boldsymbol{\sigma} - \boldsymbol{\xi}\boldsymbol{u}$ for the function $h(.)$ given in (5.2). The full conditionals have the form

$$\begin{aligned}
\pi(\sigma_j^*|.) &= \pi_{\sigma_j^*}(\sigma_j^*)L_A(\boldsymbol{\sigma}^*,\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}) \\
\pi(\xi_j|.) &= \pi_{\xi_j}(\xi_j)L_A(\boldsymbol{\sigma}^*,\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}) \\
\pi(\alpha|.) &= L_A(\boldsymbol{\sigma}^*,\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\})
\end{aligned}$$

The MCMC algorithm used to simulate from the posterior densities follows the same pattern as the bivariate algorithm given in 4.4.2: a Gibbs sampler with random walk metropolis steps with normal innovations for each parameter. The innovations all have normal distributions with zero mean but the variances differ: for the scale parameters σ_j^* the variances range between 3 and 19; for the shape parameters ξ_j the variances range between 0.1 and 0.4; the variance of the innovation for α is 0.03.

5.3.4 Results

The MCMC algorithm was carried out for 200000 iterations. Trace plots and density plots of the parameters σ_1, ξ_1 and α are given in Figure 5.1 and trace plots and density plots for all parameters are given in Figures B.1 and B.2 in Appendix B. The trace plots are based on all iterations and the density plots are based on the last 100000 iterations. Although the parameters appear to have converged, there does appear to be a large amount of autocorrelation present. This is confirmed by looking at the autocorrelation plots given in Figure B.3: for many parameters there is substantial autocorrelation even after lags of over 200. After removing the first 10^5 iterations as burn in, the posterior means and standard deviations were calculated and are given in Table 5.1 for the marginal

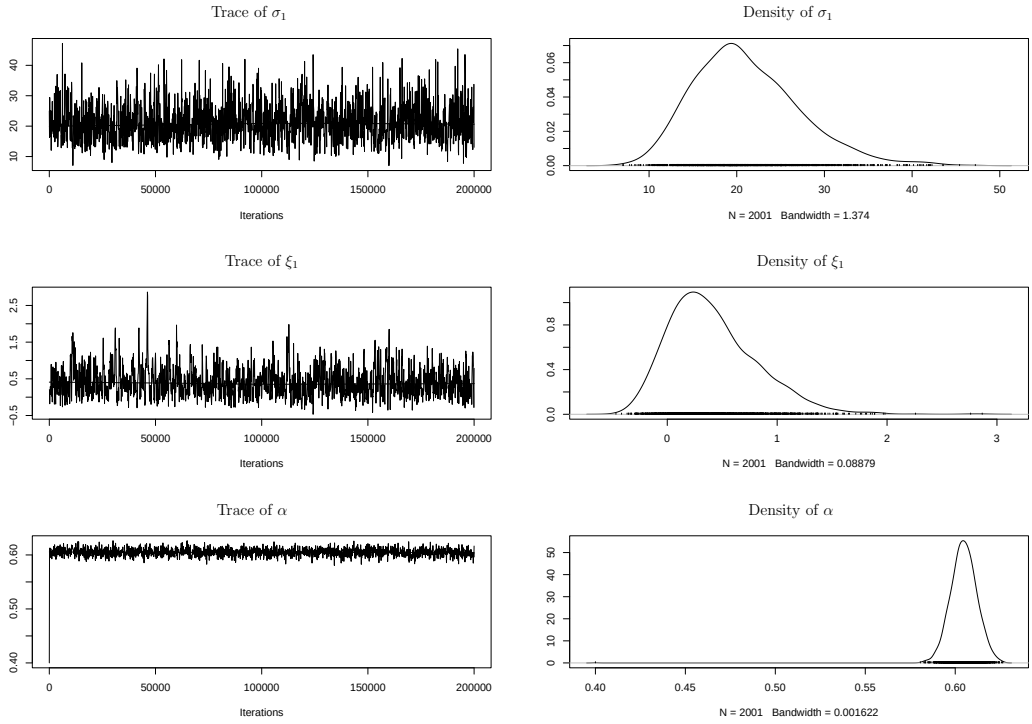


Figure 5.1: Trace plots and posterior density plots of σ_1, ξ_1 and α based on the non-informative prior

parameters. The posterior mean of the dependence parameter is 0.605 and the posterior standard deviation is 0.00706.

Instead of estimating all parameters simultaneously, the maximum likelihood estimates for the scale and shape parameters given in Table 2.4 can be used to transform the data first. The simpler model without the transformations can then be applied to the transformed data. This method results in a posterior mean of 0.632 for α and a posterior standard deviation of 0.00674. These values are fairly close to those obtained by simultaneous estimation of all parameters.

A comparison of the posterior means of the σ parameters with the maximum likelihood estimates given in Table 2.4 suggests that combining estimation of marginal and dependence parameters has had an effect on the the estimation of the σ parameters. With the priors being non-informative, it would be expected for the posterior means to be close to the maximum likelihood estimates. For some σ_j the posterior means are very close to

j	σ_j	ξ_j
1	21.281 (6.080)	0.403 (0.398)
2	19.717 (4.041)	0.369 (0.252)
3	18.300 (2.642)	-0.174 (0.169)
4	12.966 (1.468)	0.175 (0.121)
5	12.747 (2.371)	0.340 (0.216)
6	7.933 (1.759)	0.951 (0.279)
7	9.807 (0.959)	0.137 (0.0989)
8	12.640 (1.978)	0.0961 (0.177)
9	11.387 (2.608)	0.557 (0.327)
10	10.552 (2.978)	1.129 (0.399)
11	6.877 (0.921)	0.472 (0.150)

Table 5.1: Posterior means (standard deviations) for the marginal parameters when using non-informative priors

the corresponding maximum likelihood estimates, but others are quite different. There appears to be a greater difference between the posterior means and maximum likelihood estimates of the ξ parameters. This is to be expected, since Coles (1991) found that the gain in efficiency from estimating the parameters simultaneously is in the estimation of the ξ parameters. All of the posterior means of the ξ parameters are higher than the corresponding maximum likelihood estimates: the posterior means of ξ_6, ξ_9, ξ_{10} and ξ_{11} are particularly high.

The samples of the marginal parameters were transformed using (1.9) to obtain samples from the posterior distributions of the 10, 100 and 1000-year return levels. The posterior medians and 95% credibility intervals of these return levels are given in Table 5.2.

The high values observed for the parameters ξ_6, ξ_9, ξ_{10} and ξ_{11} have resulted in very high return levels for the 1000-year return level and in some cases for the 100-year return

Site	Return period (years)					
	10		100		1000	
1	122.062	(93,311)	274.524	(132,5632)	604.272	(156,118808)
2	142.923	(109,291)	317.106	(162,2184)	697.861	(213,18470)
3	82.931	(73,113)	100.490	(79,197)	112.026	(81,327)
4	109.008	(88,165)	181.892	(118,442)	289.380	(145,1212)
5	118.817	(91,235)	244.192	(126,1437)	503.915	(159,9644)
6	276.341	(130,993)	2205.078	(352,31267)	19119.150	(976,1044429)
7	77.391	(64,106)	124.019	(85,240)	187.245	(104,542)
8	78.973	(66,120)	120.414	(81,338)	170.390	(91,994)
9	102.172	(74,292)	299.337	(118,5550)	942.546	(170,120709)
10	226.812	(104,1346)	2522.036	(327,130480)	30426.83	(1021,13349970)
11	100.227	(69,184)	280.642	(122,1028)	805.131	(207,6103)

Table 5.2: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels at each site, obtained using non-informative priors

level. In general, the posterior medians of the return levels are much higher than the maximum likelihood estimates obtained with univariate analyses of each site (Table 2.5). Only the results for site 3 remain close. It seems that this could be attributed to the change in the estimates of the shape parameters, due to the simultaneous estimation of the parameters.

The 95% credibility intervals for the parameters ξ_6, ξ_9, ξ_{10} and ξ_{11} are all very wide, as are many of the others. Site 3 has much narrower intervals than all of the other sites. The width of these intervals reflects the great amount of uncertainty in the model and demonstrates the importance of incorporating any additional information that is available into the model. In the next section informative priors are used, with the intention of reducing the large amount of uncertainty observed here.

5.4 Application to rainfall data using informative priors

The model used in the previous section is used again here, but now the priors used are based on expert information.

5.4.1 Prior formulation

The informative priors used here make use of the expert information given in Coles and Tawn (1996) that was used in chapters 3 and 4. They were working with the annual maxima model and defined

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

to be the $1-p$ quantile of the annual maximum distribution. This is also known as the $\{-log(1-p)\}^{-1}$-year return period. They looked at the quantities

$$\tilde{q}_1 = q_{p_1} \qquad \tilde{q}_2 = q_{p_2} - q_{p_1} \qquad \tilde{q}_3 = q_{p_3} - q_{p_2}$$

for $(p_1, p_2, p_3) = (0.1, 0.01, 0.001)$ and assumed they had independent priors of the form

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

A hydrologist, Duncan Reed, gave estimates of the medians and 90% quantiles of the $\tilde{q}_i$ for one of the 11 locations. Coles and Tawn then used these estimates to obtain gamma parameter estimates for the corresponding priors. The quantiles q_{p_1}, q_{p_2} and q_{p_3} are approximately the 10, 100 and 1000-year return levels respectively.

In this model there are only two parameters for each margin since the GPD is used to transform the marginal data. The expression for the N-year return level, then, is that given in (1.9). In this case $n_y \approx 60$ since only the data from November to February are included and the maxima over two-day periods are taken. Since each margin has only two parameters, only two of the $\tilde{q}_i$ are needed: $\tilde{q}_1$ and $\tilde{q}_2$ are used here.

In the same way as for the univariate and bivariate analyses, the variance of the gamma parameters for the $\tilde{q}_i$ was increased by a factor of $c = 3$, since the original prior was specific to site 9.

Let $q_{p_{1,j}}$ and $q_{p_{2,j}}$ be the 10 and 100-year return levels respectively for site j; then the joint prior for margin j in terms of $q_{p_{1,j}}$ and $q_{p_{2,j}}$ is

$$\pi\left(q_{p_{1,j}}, q_{p_{2,j}}\right) \propto q_{p_{1,j}}^{\alpha_1 - 1} \exp\left(-\beta_1 q_{p_{1,j}}\right) \left(q_{p_{2,j}} - q_{p_{1,j}}\right)^{\alpha_2 - 1} \exp\left\{-\beta_2 \left(q_{p_{2,j}} - q_{p_{1.j}}\right)\right\},$$

where

$$q_{p_{1,j}} = u_j + \frac{\sigma_j}{\xi_j}\left[\left(600\zeta_{u_j}\right)^{\xi_j} - 1\right]$$

and

$$q_{p_{2,j}} = u_j + \frac{\sigma_j}{\xi_j}\left[\left(6000\zeta_{u_j}\right)^{\xi_j} - 1\right],$$

for $\xi_j \neq 0$ (from (1.9) with $n_y = 60$). In Chapter 4 σ_j was replaced with $\sigma_j^* + \xi_j u_j$ since σ_j^* is not threshold dependent: the same is done here.

To express the prior for margin j in terms of σ_j and ξ_j, the expressions for $q_{p_{1,j}}$ and $q_{p_{2,j}}$ are substituted into $\pi\left(q_{p_{1,j}}, q_{p_{2,j}}\right)$ and then it is multiplied by the Jacobian of the transformation. The Jacobian is

$$\begin{aligned} J(\sigma_j, \xi_j) \;=\; & \frac{1}{\xi_j^2}\left[(10n_y\zeta_{u_j})^{\xi_j}(100n_y\zeta_{u_j})^{\xi_j}\log(100n_y\zeta_{u_j})\sigma_j\right. \\ & +(10n_y\zeta_{u_j})^{\xi_j}(100n_y\zeta_{u_j})^{\xi_j}\log(100n_y\zeta_{u_j})\xi_j u_j \\ & -(10n_y\zeta_{u_j})^{\xi_j}\log(100n_y\zeta_{u_j})\sigma_j - (100n_y\zeta_{u_j})^{\xi_j}\log(100n_y\zeta_{u_j})\xi_j u_j \\ & -(10n_y\zeta_{u_j})^{\xi_j}\log(10n_y\zeta_{u_j})\sigma_j(100n_y\zeta_{u_j})^{\xi_j} + (10n_y\zeta_{u_j})^{\xi_j}\log(10n_y\zeta_{u_j})\sigma_j \\ & \left.-(10n_y\zeta_{u_j})^{\xi_j}\log(10n_y\zeta_{u_j})\xi_j u_j(100n_y\zeta_{u_j})^{\xi_j} + (10n_y\zeta_{u_j})^{\xi_j}\log(10n_y\zeta_{u_j})\xi_j u_j\,\right]. \end{aligned}$$

The prior on the dependence parameter α is $U(0,1)$ since we have no information about α to formulate an informative prior from. The joint prior is given by

$$\pi(\boldsymbol{\sigma}, \boldsymbol{\xi}, \alpha) = \prod_{j=1}^{11} \pi_j(\sigma_j, \xi_j), \tag{5.8}$$

where $\pi_j(\sigma_j, \xi_j)$ is the marginal prior for site j, and $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$ are the vectors of the parameters σ_j and ξ_j respectively, for $j = 1, \ldots, 11$.

5.4.2 Posterior inference

The posterior distribution has the form

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{11}) \propto \pi(\boldsymbol{\sigma},\boldsymbol{\xi},\alpha) \times L_A(\boldsymbol{\sigma},\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}), \tag{5.9}$$

where $L_A(.)$ is the likelihood given in (5.5) for the function $h(.)$ given in (5.2). The full conditionals have the form

$$\begin{aligned}
\pi(\sigma|.) &= \pi_j(\sigma_j,\xi_j)L_A(\boldsymbol{\sigma},\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}) \\
\pi(\xi_j|.) &= \pi_j(\sigma_j,\xi_j)L_A(\boldsymbol{\sigma},\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\}) \\
\pi(\alpha|.) &= L_A(\boldsymbol{\sigma},\boldsymbol{\xi},\alpha;\{\boldsymbol{Y}_i\})
\end{aligned}$$

The MCMC algorithm used to simulate from the posterior densities is, again, a Gibbs sampler with random walk metropolis steps with normal innovations for each parameter. The innovations all have means of zero and have different variances: for the σ_j the variances range from 1.8 to 12; for the ξ_j the variances range from 0.06 to 0.13; the variance of the innovation for α is 0.025.

5.4.3 Results

Samples of size 10000 were obtained from the marginal posterior distributions of the parameters. Trace plots and density plots of the parameters σ_1, ξ_1, and α are given in Figure 5.2, and trace and density plots for all parameters are given in Figures B.4 and B.5 of Appendix B. The trace plots are based on all 10000 iterations and the density plots are based on the last 5000 iterations. All parameters appear to have converged well, and using different starting points for the chains gave very similar results. The first 5000 iterations were removed and the posterior means and standard deviations were obtained: the posterior means and standard deviations of the marginal GPD parameters are given in Table 5.3; the posterior mean and standard deviation of α are 0.615 and 0.00717 respectively.

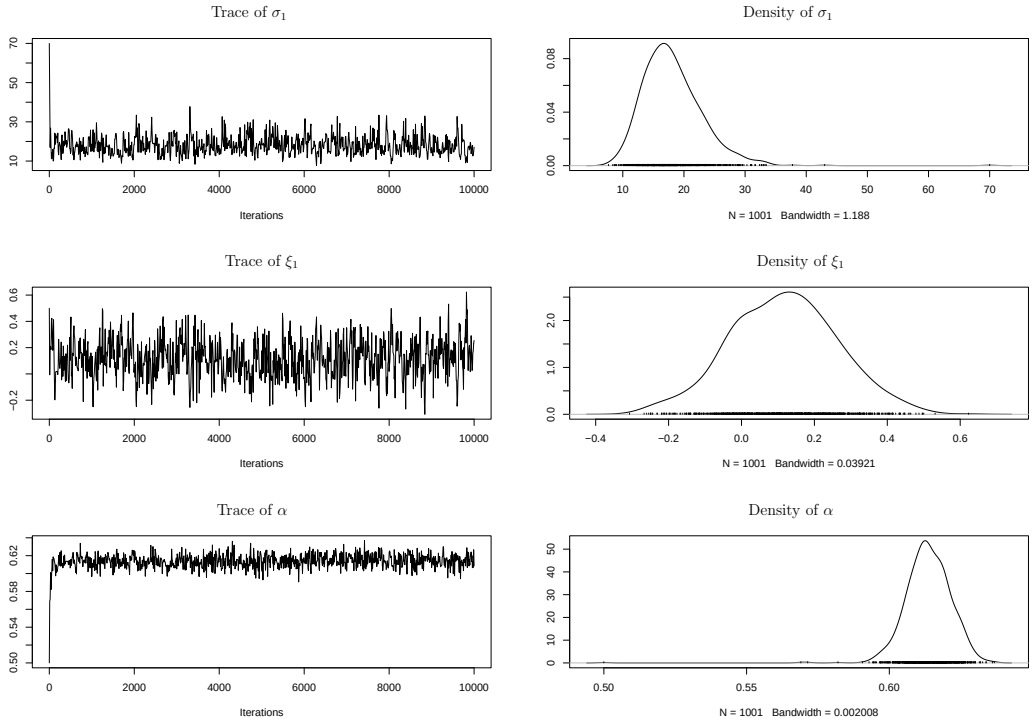


Figure 5.2: Trace plots and posterior density plots of σ_1, ξ_1 and α based on the informative prior

The samples of the marginal parameters were transformed using (1.9) to obtain samples from the posterior distributions of the 10, 100 and 1000-year return levels. The posterior medians and 95% credibility intervals of these return levels are given in Table 5.4.

5.5 The effect of the informative priors

In this section, the results of section 5.4 are compared with those of section 5.3 to assess the effect of using the informative priors.

j	σ_j	ξ_j
1	18.245 (4.597)	0.115 (0.150)
2	17.863 (3.105)	0.121 (0.0969)
3	16.074 (2.077)	-0.0759 (0.119)
4	13.137 (1.223)	0.0850 (0.0703)
5	12.944 (1.999)	0.136 (0.100)
6	10.366 (1.423)	0.302 (0.0703)
7	9.627 (0.823)	0.112 (0.0691)
8	12.049 (1.609)	0.0664 (0.106)
9	11.173 (2.137)	0.251 (0.117)
10	10.720 (2.230)	0.373 (0.0937)
11	7.312 (0.818)	0.283 (0.0737)

Table 5.3: Posterior means (standard deviations) for the marginal parameters when using informative priors

5.5.1 The effect on the GPD parameters and the dependence parameter

Figure 5.3 shows the estimated posterior densities for σ_1, ξ_1 and α, using both non-informative and informative priors. In all three plots there is a clear difference between the density based on the non-informative prior and that based on the informative prior: for σ_1, the posterior mode is slightly lower when the informative prior is used and there is a slight decrease in the spread; for ξ_1, the posterior modes are approximately equal for the two priors, but there is much greater certainty in the distribution when the informative prior was used; for α, the posterior distribution has shifted to the right when the informative prior is used, but the spread of the two distributions appears to be similar.

To assess the affect of the informative prior on all margins, the posterior means and standard deviations obtained with the non-informative and informative priors, given in

Site	Return period (years)					
	10		100		1000	
1	96.628	(85,114)	157.334	(122,231)	237.326	(148,540)
2	109.970	(97,127)	175.001	(139,240)	259.336	(176,478)
3	85.534	(75,102)	111.479	(86,165)	133.313	(91,259)
4	96.331	(84,113)	144.106	(112,200)	202.308	(135,345)
5	96.225	(85,112)	151.339	(114,225)	227.264	(141,468)
6	103.264	(90,119)	204.814	(151,289)	407.796	(242,769)
7	73.717	(64,88)	114.858	(86,165)	167.927	(107,299)
8	75.561	(66,90)	112.787	(84,172)	156.081	(98,329)
9	78.492	(68,93)	146.830	(106,232)	265.357	(148,668)
10	83.191	(70,101)	193.934	(141,292)	457.299	(254,979)
11	74.403	(63,90)	147.935	(102,222)	289.178	(155,576)

Table 5.4: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels at each site, obtained using informative priors

Table 5.1 and Table 5.3 respectively, can be compared. The posterior means of σ_j, $j = 1, \ldots, 11$, have changed slightly with the use of the informative prior, but those for $j = 1, 2, 3, 6$ have changed the most. The informative prior has affected posterior means of ξ_j, $j = 1, \ldots, 11$ more significantly than the σ_j: the effect on ξ_6 and ξ_{10} is particularly significant. The posterior mean of α was 0.605 with the non-informative prior and 0.615 with the informative prior: this change was also noticed in the analysis of Figure 5.3. The posterior standard deviations of all of the marginal GPD parameters are lower when the informative prior is used: this reflects the reduced uncertainty due to incorporating the expert information. No informative prior information was available for the dependence parameter α; consequently the posterior standard deviation increased slightly, from 0.00674 to 0.00717, with the use of the informative prior.

Comparing the autocorrelation, given in Figure B.3 and Figure B.6 for the non-

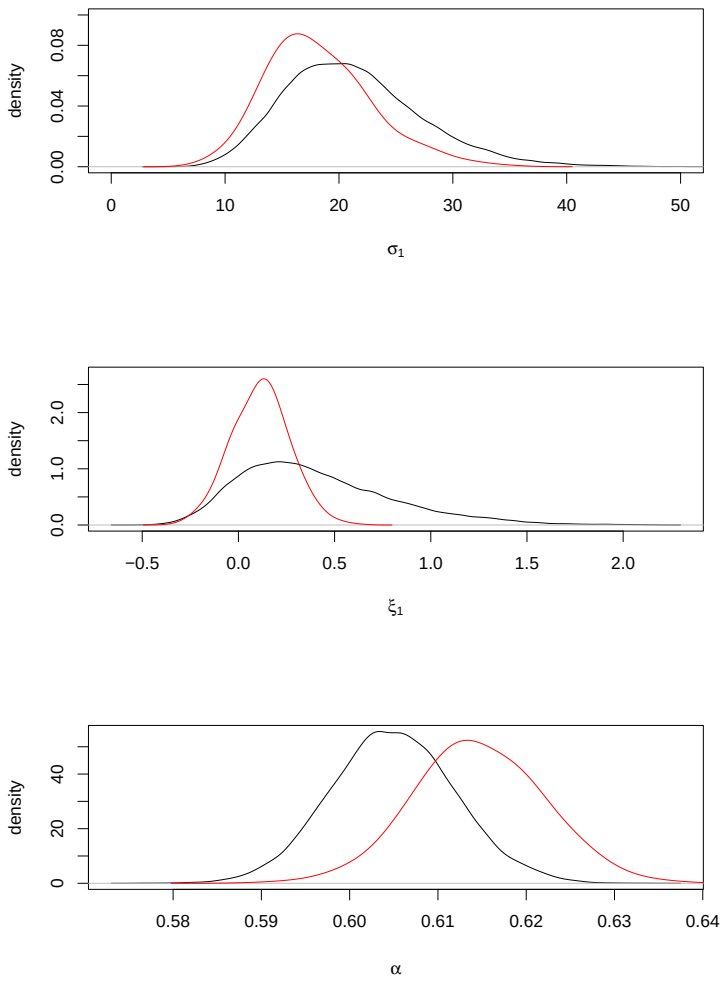


Figure 5.3: Posterior density plots of σ_1, ξ_1 and α based on the non-informative prior (in black) and the informative prior (in red)

informative prior and the informative prior respectively, shows that the informative prior has significantly reduced the autocorrelation.

5.5.2 The effect on the return levels

Posterior densities of the 10, 100 and 1000-year return levels for site 4, using both non-informative and informative priors, are given in Figure 5.4. These plots show that using the informative prior has had an effect on the distribution of return levels for site 4. For

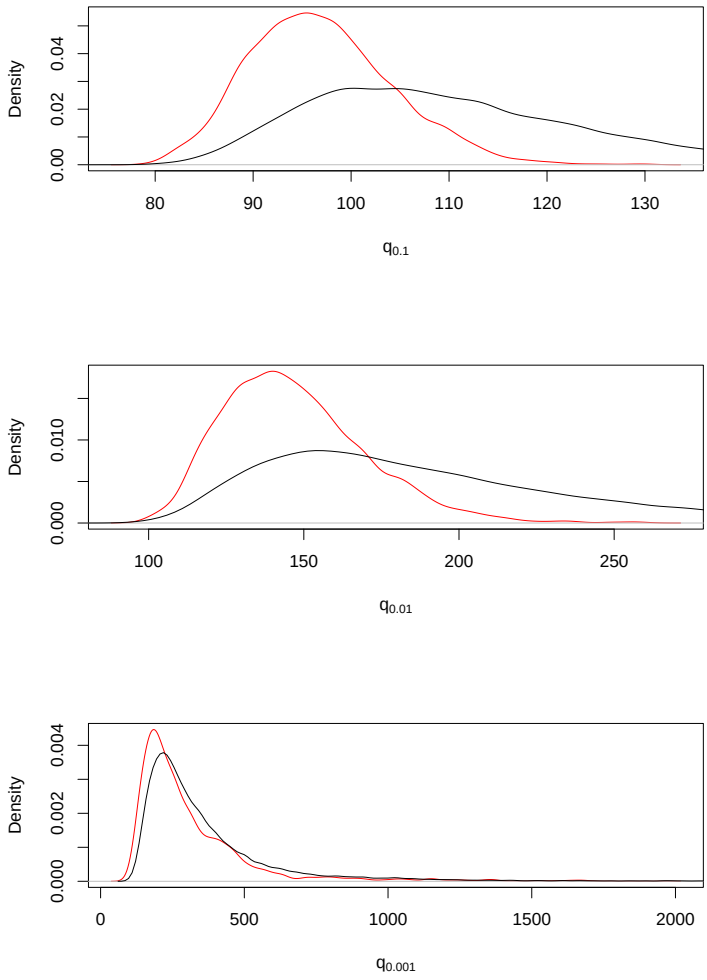


Figure 5.4: Posterior density plots of $q_{0.1}$, $q_{0.01}$ and $q_{0.001}$ for site 4 based on the non-informative prior (in black) and the informative prior (in red)

all three return levels, the informative prior has resulted in the posterior mode being shifted slightly to the left. The informative prior has also resulted in increased certainty, particularly noticeable for $q_{0.1}$ and $q_{0.01}$.

Figure 5.5 gives return level curves for site 4, based on the posterior medians, for both non-informative and informative priors. Estimates based on the empirical distribution function of the annual maximum are also given. This plot shows that using the non-informative prior has the effect of pushing up the return levels, especially for the longer

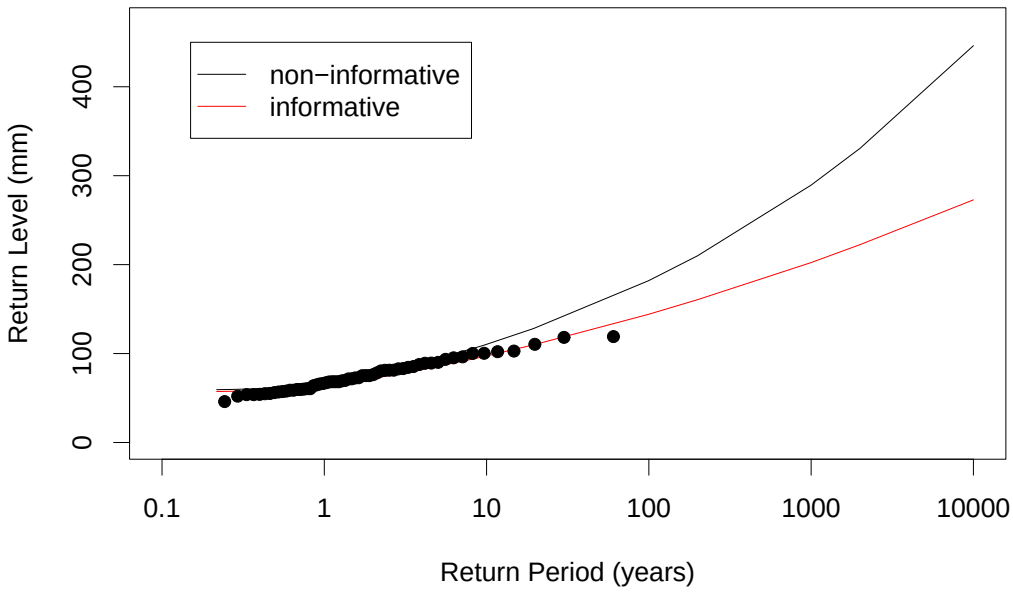


Figure 5.5: Return level plots for site 4, based on the posterior medians of return levels, using both non-informative and informative priors: empirical estimates are given by the points

return periods. This effect was also observed in the univariate and bivariate analyses, although to a lesser extent. The curve based on the informative prior follows more closely the empirical estimates at the longer return periods, but the two curves are very close at short return periods.

The return level curve for site 4 based on the informative prior is given with 95% credibility intervals based on both priors in Figure 5.6. The credibility interval based on the non-informative prior is much wider than that based on the informative prior. The lower bounds are very close, with the one based on the non-informative prior being slightly higher. The upper bound based on the non-informative prior is much higher than that based on the informative prior. This reflects the increase in certainty achieved by using the expert information.

To examine the effect of the informative prior on the return levels at all of the sites,

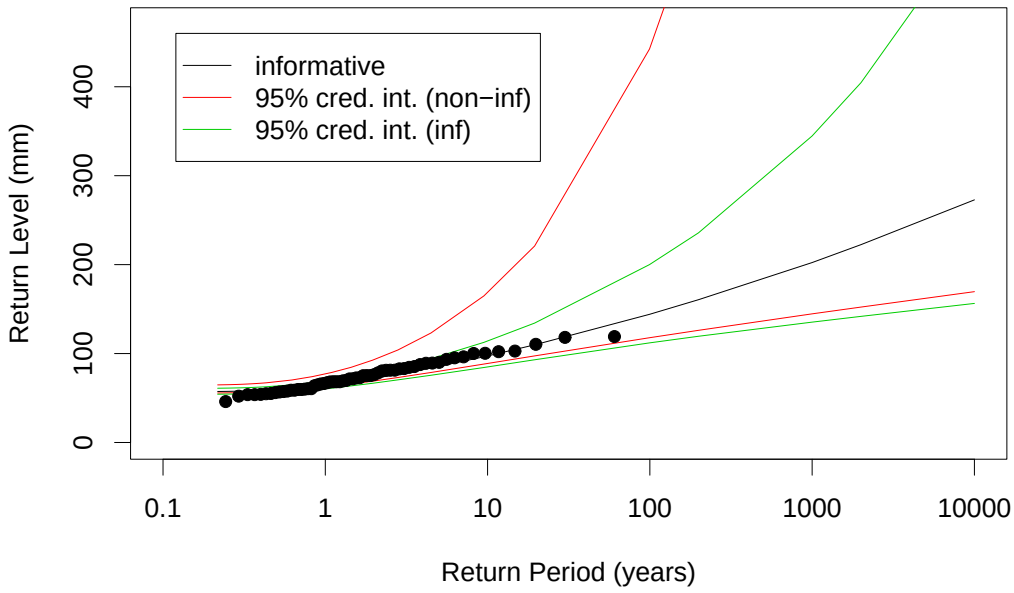


Figure 5.6: Return level plot for site 4 using the informative prior and 95% credibility intervals based on both non-informative and informative priors

the posterior medians and 95% credibility intervals of the 10, 100 and 1000-year return levels for both priors, given in Table 5.2 and Table 5.4, can be compared. For all sites, the posterior medians of the three return levels are lower when the informative prior is used, except for those of site 3. For sites 6 and 10 the difference between the posterior medians for the two priors is particularly great: for these sites, it seems that the return levels are greatly overestimated when the non-informative prior is used. The 95% credibility intervals of the return levels based on the informative prior are all narrower than the respective intervals based on the non-informative prior. Some of the intervals; for example those of sites 6 and 10, are much narrower. The lower bounds are often close for the two priors, but the upper bounds based on the informative prior are usually much higher. One exception is site 3, where the posterior medians and 95% credibility intervals are very close for the two priors.

5.6 Prediction

Predictive return levels for each site can be obtained by solving equation (3.19) for z_p. For this model, equation (3.19) becomes

$$M^{-1}\sum_{m=1}^{M} 1 - n_y\zeta_u\left[1+\xi\left(\frac{z_p-u}{\sigma}\right)\right]^{-1/\xi} = 1-p. \tag{5.10}$$

Solving this equation would give a level z_p which would be exceeded with probability p in a 1-year period.

Predictive return level plots for site 4 based on the non-informative prior and the informative prior are given in Figure 5.7: the return level plots based on the posterior medians are also given. The predictive return level curves, for both priors, are above the the corresponding posterior median return level curves: the difference between them increases with $-\log(1-p)$. This suggests that using the posterior medians of the return levels for design purposes may lead to under-protection, especially for long return periods.

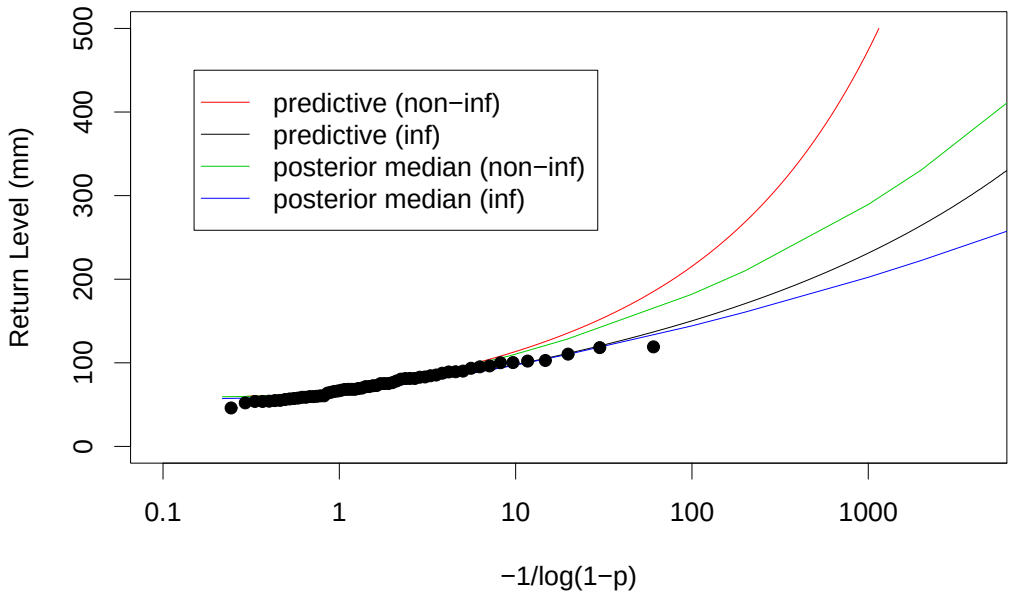


Figure 5.7: Predictive return level plots and return level plots with non-informative and informative priors, and empirical estimates for site 4

Using the predictive return levels, then, has accounted for the parameter uncertainty in the models. There is also quite a large difference between the two predictive return level curves, with the one based on the non-informative prior being higher, particularly for high values of $-1/\log(1-p)$. Using the non-informative prior, then, seems to result in the predictive return levels being over-estimated. This is because there is more uncertainty when no expert information is used.

The predictive return levels for the 10, 100 and 1000-year return periods were obtained using the non-informative and informative priors and are given, for each site, in Table 5.5. All of the predictive return levels are higher than the corresponding posterior medians of the return levels, given in Table 5.2 and Table 5.4 for non-informative and informative priors respectively. Taking parameter uncertainty into account, then, has resulted in increased estimates of return levels. The difference between the posterior

	Non-informative prior			**Informative prior**		
	Return period (years)			Return period (years)		
Site	10	100	1000	10	100	1000
1	127.113	510.617	5493.289	97.581	162.814	284.380
2	148.859	465.497	2418.407	110.607	180.876	297.831
3	84.470	119.012	190.032	86.046	119.891	169.733
4	112.056	215.109	474.036	96.879	150.106	231.224
5	124.064	347.770	1513.531	96.749	158.058	269.366
6	298.323	3597.945	63122.280	103.354	210.318	449.551
7	79.022	139.604	263.845	74.217	120.246	193.836
8	81.027	150.062	342.097	76.015	119.449	192.735
9	108.189	532.187	6229.658	79.234	152.728	318.574
10	253.741	5483.210	231003.984	84.169	199.284	505.797
11	104.214	353.053	1476.615	74.825	153.983	329.377

Table 5.5: Predictive return levels for the 10, 100 and 1000-year return periods

medians of return level and the predictive return levels increases with return period, and the difference is greater when the non-informative prior is used. The informative prior, then, has accounted for some of the uncertainty. The predictive return levels based on the non-informative prior are, in general, much greater than those based on the informative prior, with greater differences observed for higher return periods. The only exceptions are for site 3, where the two sets of predictive return levels are very close; in fact, for the return periods of 10 and 100 years the predictive return levels based on the informative prior are the highest. For many of the sites, the predictive return levels obtained with the non-informative prior were very high due to the high values of the shape parameters. The informative prior has reduced the predictive return levels to more credible values. A difference between these predictive return levels and those obtained with the univariate analysis of the annual maxima (Table 3.9) can also be observed. For the non-informative prior, the predictive return levels based on the multivariate point process method are all much higher, with greater differences observed for longer return periods. For the informative prior the two sets of results are much closer, with most values being higher for the multivariate point process method.

5.7 Discussion

In this chapter, the Bayesian model for bivariate annual maxima, considered in Chapter 4, was extended to a model for the daily rainfall data at all 11 sites. The multivariate point process method was used to achieve this. As in chapters 3 and 4, both a non-informative prior and an informative prior were used, with the informative priors for the margins being based on the expert information described in section 3.4. No information about the dependence parameter was available, so a non-informative prior was used throughout. Non-stationarity and dependence in the data were accounted for by using only the data from November to February and by taking the maxima over two-day periods. After applying an MCMC scheme to simulate from the marginal posterior distributions, posterior

means and standard deviations of all parameters were obtained.

A comparison of the results obtained with the non-informative prior and the maximum likelihood estimates obtained in Chapter 2 showed that combining the estimation of the marginal parameters and the dependence parameter resulted in different estimates for the scale and shape parameters: the dependence parameter was not changed significantly. For sites 6, 9, 10 and 11, the non-informative prior resulted in high posterior means for the shape parameters, which resulted in very high return level estimates. In general, the posterior medians of the return levels were much higher than the maximum likelihood estimates obtained in Chapter1: site 3 was an exception, with the maximum likelihood estimates and the posterior medians remaining close. Using the informative prior significantly affected the posterior means of the GPD parameters, particularly those of the shape parameters for sites 6 and 10. The very large values observed for the shape parameters under the non-informative prior were significantly reduced by using the informative prior. This resulted in significantly reduced return level estimates. The posterior standard deviations of the GPD parameters were all greatly reduced when the informative prior was used: the 95% credibility intervals of the return levels were also greatly reduced. This demonstrates the benefit of incorporating expert prior information into the model. Again, site 3 was an exception, with the posterior medians and 95% credibility intervals remaining very close under the two priors. The predictive return levels were, in general, higher than the posterior medians of the return levels, with greater differences for longer return periods and for the model with the non-informative prior. The predictive return levels for site 3 remained very close to the posterior medians and remained very close under both priors.

Overall, this chapter has demonstrated the importance of considering a multivariate model for all sites and of incorporating expert prior information into the model. Using a multivariate model allows dependence to be modelled and, by combining the estimation of the dependence parameters with the estimation of the marginal parameters, enables a potential gain in efficiency due to the transfer of information across variables. The

multivariate model has resulted in different marginal parameter estimates to the univariate models and therefore in different return level estimates. For sites where the return level estimates were extremely high, perhaps due to insufficient data, the informative prior has reduced the estimates to more credible levels. Also, the informative prior has reduced the uncertainty in the distribution of the parameters. Using a Bayesian analysis has enabled the expert prior information to be used and has provided a way of obtaining predictive return level estimates, which take into account parameter uncertainty.

There are some potential problems with using the multivariate logistic model, and with using the traditional methods for multivariate extremes. These problems are described below and attempt at addressing them is made in Chapter 6.

The logistic model is very restrictive: the model has only one dependence parameter and it results in all variables being exchangeable. Another model for dependence, the Dirichlet model, is considered in Chapter 6.

Unfortunately, the strength of dependence of some multivariate processes weakens at high levels. This results in the most extreme events being near-independent. In such cases, the traditional methods for multivariate extremes may lead to misleading results. This issue has received much consideration by Ledford and Tawn (1996,1997 and 1998). Heffernan and Tawn (2004) say that, in practice, this problem restricts existing methods to applications in two or three dimensions. In Chapter 6, the Dirichlet model is applied to sub-groups, of size 2 or 3, of the 11 sites. This is an attempt at avoiding the the aforementioned problem, whilst still using the traditional methods.

Chapter 6

Further modelling

6.1 Introduction

The applications in chapters 4 and 5 use the logistic model for the dependence structure. The logistic model, however, has only one dependence parameter and this results in the variables being exchangeable. In this chapter the Dirichlet model, derived by Coles and Tawn (1991), is considered. This model incorporates different levels of pairwise dependence and bivariate asymmetry. Another problem with the model used in Chapter 5 is addressed in this chapter: the limiting arguments on which the model is based are not always appropriate and as a result existing methods are often only valid in two or three dimensions. Ledford and Tawn (1996, 1997 and 1998) showed that the multivariate threshold method is not appropriate for extrapolation of a variable with components that are dependent but asymptotically independent; the limiting conditions underlying the method are invalidated in this case.

The Dirichlet model (outlined in Section 6.2) is more difficult to implement than the logistic model, and the time needed to implement it increases greatly with the number of dimensions. To overcome this problem, the Dirichlet model is applied to each of four subgroups of the network of sites (Section 6.3). The same model is then applied again, but assuming that the shape parameter is the same for all of the sites. This links the

four Dirichlet models to give a combined model for all 11 sites. In both models, the non-informative and informative priors used in Chapter 5 are applied.

By using the Dirichlet model on sub-groups of sites of size 2 or 3, taken from the 11 sites, it is possible to avoid the problem of invalid limiting arguments. In order for the model to be suitable, however, the groups of sites need to be chosen appropriately. The choice of groups in Section 6.3 was fairly arbitrary and the suitability of this choice is assessed in Section 6.4.

A discussion of the results of this chapter and of the model choice is given in Section 6.5.

6.2 The Dirichlet model

The Dirichlet model is given by Coles and Tawn (1991). From 3.3 of Coles and Tawn (1991), the distribution function of a random vector with unit Fréchet margins is

$$G_{\boldsymbol{X}}(\boldsymbol{x}) = \exp\left\{-\int_{S_p} \max_{1\le j\le p}\left(\frac{u_j}{m_j x_j}\right) dH^*(\boldsymbol{u})\right\}, \tag{6.1}$$

where $\int_{S_p} u_j dH^*(\boldsymbol{u}) = m_j$ and H^* is a positive finite measure on S_p. For the Dirichlet model,

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

By theorem 2 of Coles and Tawn (1991) it can be shown that for the Dirichlet model, $m_j = \alpha_j/(\boldsymbol{\alpha.1})$ and the measure density is

$$h(\boldsymbol{w}) = \prod_{j=1}^{p}\frac{\alpha_j}{\Gamma(\alpha_j)}\frac{\Gamma(\boldsymbol{\alpha.1}+1)}{(\boldsymbol{\alpha.w})^{p+1}}\prod_{j=1}^{p}\left(\frac{\alpha_j w_j}{\boldsymbol{\alpha.w}}\right)^{\alpha_j-1}, \qquad w\in S_p. \tag{6.2}$$

Since $G(\boldsymbol{x}) = \exp\{-V(\boldsymbol{x})\}$, from (6.1)

$$\begin{aligned} V(\boldsymbol{x}) &= \int_{S_p}\max_{1\le j\le p}\left(\frac{u_j}{m_j x_j}\right) dH^*(\boldsymbol{u}) \\ &= \int_{S_p}\max_{1\le j\le p}\left(\frac{u_j}{m_j x_j}\right) h^*(\boldsymbol{u})\, d\boldsymbol{u}. \end{aligned}$$

So, for two dimensions ($p = 2$) and for the Dirichlet model,

$$\begin{aligned} V(x_1, x_2) &= \int_{S_p} \max\left(\frac{w}{m_1 x_1}, \frac{1-w}{m_2 x_2}\right) \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w^{\alpha_1-1}(1-w)^{\alpha_2-1}\, dw \\ &= \int_{S_p} \max\left(\frac{w(\alpha_1+\alpha_2)}{\alpha_1 x_1}, \frac{(1-w)(\alpha_1+\alpha_2)}{\alpha_2 x_2}\right) \times \\ &\quad \times \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w^{\alpha_1-1}(1-w)^{\alpha_2-1}\, dw \\ &= \int_{S_p} (\alpha_1+\alpha_2) \max\left(\frac{w}{\alpha_1 x_1}, \frac{1-w}{\alpha_2 x_2}\right) \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \times \\ &\quad \times w^{\alpha_1-1}(1-w)^{\alpha_2-1}\, dw. \end{aligned} \tag{6.3}$$

Now,

$$\begin{aligned} \frac{w}{\alpha_1 x_1} > \frac{1-w}{\alpha_2 x_2} &\Leftrightarrow w\alpha_2 x_2 > \alpha_1 x_1 - w\alpha_1 x_1 \\ &\Leftrightarrow w(\alpha_2 x_2 + \alpha_1 x_1) > \alpha_1 x_1 \\ &\Leftrightarrow w > \frac{\alpha_1 x_1}{\boldsymbol{\alpha.x}}, \end{aligned}$$

So (6.3) can be written as

$$\begin{aligned} &\int_0^{\frac{\alpha_1 x_1}{\alpha.\mathbf{x}}} \frac{1-w}{\alpha_2 x_2}(\alpha_1+\alpha_2)\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w^{\alpha_1-1}(1-w)^{\alpha_2-1}\, dw \\ &+ \int_{\frac{\alpha_1 x_1}{\boldsymbol{\alpha.x}}}^1 \frac{w}{\alpha_1 x_1}(\alpha_1+\alpha_2)\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w^{\alpha_1-1}(1-w)^{\alpha_2-1}\, dw \\ = \frac{1}{x_2}&\int_0^{\frac{\alpha_1 x_1}{\boldsymbol{\alpha.x}}} \frac{\Gamma(\alpha_1+\alpha_2+1)}{\Gamma(\alpha_1)\Gamma(\alpha_2+1)} w^{\alpha_1-1}(1-w)^{\alpha_2}\, dw \\ &+ \frac{1}{x_1}\left(1 - \int_0^{\frac{\alpha_1 x_1}{\boldsymbol{\alpha.x}}} \frac{\Gamma(\alpha_1+\alpha_2+1)}{\Gamma(\alpha_1+1)\Gamma(\alpha_2)} w^{\alpha_1}(1-w)^{\alpha_2-1}\, dw\right), \end{aligned}$$

which is equivalent to the expression given by Coles and Tawn (1991) in section 4.3. For more than two dimensions, the extension is complicated but evaluation of the integral is

possible numerically. In general,

$$\begin{aligned}
V(\boldsymbol{x}) &= \int_{S_p} \max_{1\le j\le p}\left(\frac{w_j}{m_j x_j}\right) h^*(\boldsymbol{w})\, d\boldsymbol{w} \\
&= \int_{S_p} \max_{1\le j\le p}\left(\frac{w_j}{m_j x_j}\right) \left\{\prod_{j=1}^{p}\Gamma(\alpha_j)\right\}^{-1} \Gamma(\boldsymbol{\alpha.1}) \prod_{j=1}^{p} w_j^{\alpha_j-1} d\boldsymbol{w} \\
&= \frac{\Gamma(\boldsymbol{\alpha.1})}{\prod_{j=1}^{p}\Gamma(\alpha_j)} \int_{S_p} \max_{1\le j\le p}\left(\frac{w_j \boldsymbol{\alpha.1}}{\alpha_j x_j}\right) \prod_{j=1}^{p} w_j^{\alpha_j-1} d\boldsymbol{w} \\
&= \frac{\Gamma(\boldsymbol{\alpha.1}+1)}{\prod_{j=1}^{p}\Gamma(\alpha_j)} \int_{S_p} \max_{1\le j\le p}\left(\frac{w_j}{\alpha_j x_j}\right) \prod_{j=1}^{p} w_j^{\alpha_j-1} d\boldsymbol{w} \\
&= \frac{\Gamma(\boldsymbol{\alpha.1}+1)}{\prod_{j-1}^{p}\Gamma(\alpha_j)} \int_0^1 \int_0^{1-w_1} \int_0^{1-w_1-w_2} \cdots \int_0^{1-\sum_{j=1}^{p-2} w_j} \max_{1\le j\le p}\left(\frac{w_j}{\alpha_j x_j}\right) \times \\
&\quad \times \prod_{j=1}^{p} w_j^{\alpha_j-1} dw_{p-1} \ldots dw_1. \qquad (6.4)
\end{aligned}$$

To use this model for more than two dimensions, numerical integration techniques or Monte Carlo integration can be used to approximate the integrals.

It can be shown that if $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ is the parameter vector of a p-dimensional Dirichlet model then the bivariate pair with margins i and j follows a bivariate Dirichlet model with parameter vector (α_i, α_j) (Coles and Tawn, 1994). In the bivariate case, measures of asymmetry and dependence strength are given by reparameterising:

$$r_1 = \frac{(\alpha_i - \alpha_j)}{2}, \qquad r_2 = \frac{(\alpha_i + \alpha_j)}{2},$$

where r_1 and r_2 measure the asymmetry and strength of dependence respectively.

6.3 Application to the rainfall data

As mentioned in the introduction, the Dirichlet model is more time consuming to use than the logistic, particularly in high dimensions. This is due to the numerical evaluation of the integral in (6.4). Also, due to the findings of Ledford and Tawn (1996, 1997 and 1998), an application of this model will probably be invalid in more than three dimensions. To

avoid using an invalid model, the Dirichlet model can be applied to sub-groups of size 2 or 3 of the 11 sites. This will also avoid a very time consuming procedure needed for a high dimensional Dirichlet model.

In this section, the Dirichlet model is applied to four sub-groups of the 11 sites. Initially, the model is applied independently to each group. Then, the models for the four groups are linked by assuming a constant shape parameter for all 11 sites. This a way of linking the four models to give one model for all 11 sites, without assuming a high dimensional Dirichlet model. Details of the model are given below in Section 6.3.1.

6.3.1 Model

Suppose that in group h, $h = 1, \ldots, 4$ there are p_h sites and n_{A_h} observations at each site. Then, from (5.5), for cluster h the likelihood component is

$$L_{A_h}(\boldsymbol{\theta_h}, \boldsymbol{\sigma_h}, \boldsymbol{\xi_h}; \{\boldsymbol{Y}_{hi}\}) = exp\{-V(\boldsymbol{v_h})\} \prod_{i=1}^{n_{A_h}} \Bigg(h(\boldsymbol{w}_{hi})(n_h r_{hi})^{-(p_h+1)} \times \prod_{\substack{j=1,\ldots,p_h: \\ X_{hij} > n_h \nu_{hj}}} [\sigma_{hj}^{-1} p_{hj}^{-\xi_{hj}} X_{hij}^2 exp(1/X_{hij}) \{1 - exp(-1/X_{hij})\}^{1+\xi_{hj}}] \Bigg). \quad (6.5)$$

The full likelihood then is

$$L\left(\Theta, \Sigma, \Xi; y\right) = \prod_{h=1}^{4} L_{A_h}\left(\boldsymbol{\theta_h}, \boldsymbol{\sigma_h}, \boldsymbol{\xi_h}; \{\boldsymbol{Y_{hi}}\}\right). \quad (6.6)$$

Two different approaches to modelling the data are used:

Model 1. Each cluster is modelled separately with a multivariate model with no link assumed. The likelihood is as in (6.6).

Model 2. The groups are linked together by assuming a constant shape parameter. In this case the likelihood is as in (6.6) with the shape parameters such that $\xi_{hj} = \xi_h = \xi$.

The choice of the sub-groups was fairly arbitrary, but the physical location of sites was taken into account. The four groups chosen are:

- group 1: sites 1, 4 and 5;

- group 2: sites 2, 3 and 6;
- group 3: sites 7, 8 and 9;
- group 4: sites 10 and 11.

6.3.2 Prior choice

For the marginal GPD parameters the non-informative and informative priors are identical to those used in Chapter 5: details are given in sections 5.3.2 and 5.4.1 for the non-informative and informative priors respectively. The dependence parameters $\alpha_j, j = 1, \ldots, 11$, are restricted to being positive but no other information about them is known. A normal prior with a variance of 1000 is used for the α_j but the prior is truncated below at zero, so that negative values are not allowed: the prior for α_j is denoted by $\pi_j(\alpha_j)$.

6.3.3 Posterior inference

For the model with non-informative priors on the margins the posterior has the form

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_1 1) \propto \prod_{j=1}^{11} \pi_j(\sigma_j^*, \xi_j)\pi_j(\alpha_j) \times L(\Theta, \Sigma^*, \Xi; \boldsymbol{Y}_i), \tag{6.7}$$

where $L(.)$ is the likelihood given in (6.6), in terms of $\sigma^* = \sigma - \xi u$ for the function $h(.)$ given in (6.2). The full conditionals have the form

$$\begin{aligned}
\pi(\sigma_j^*|.) &= \pi_{\sigma_j^*}(\sigma_j^*) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h^*, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right), \\
\pi(\xi_j^*|.) &= \pi_{\xi_j^*}(\xi_j) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h^*, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right), \\
\pi(\alpha_j^*|.) &= \pi_{\alpha_j}(\alpha_j^*) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h^*, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right),
\end{aligned}$$

where site j is from group h.

For the model with informative priors on the margins the posterior has the same form

as (6.7) but in terms of σ_j, not σ_j^*. The full conditionals have the form

$$\begin{aligned}
\pi(\sigma_j^*|.) &= \pi_j(\sigma_j, \xi_j) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right), \\
\pi(\xi_j|.) &= \pi_j(\sigma_j, \xi_j) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right), \\
\pi(\alpha_j|.) &= \pi_{\alpha_j}(\alpha_j) L_{A_h}\left(\boldsymbol{\theta}_h, \boldsymbol{\sigma}_h, \boldsymbol{\xi}_h; \{\boldsymbol{Y}_{hi}\}\right).
\end{aligned}$$

The same type of MCMC algorithm used in Chapter 5 is used here: a Gibbs sampler with random walk metropolis steps with normal innovations for each parameter.

6.3.4 Results

For model 1 the MCMC algorithms were carried out for 3500 and 2500 iterations for the models with non-informative and informative priors respectively. For model 2 the algorithms were carried out for 3000 and 1000 iterations for the models with non-informative and informative priors respectively. Trace plots of the first 1000 iterations of σ_1, for both models with both priors, are given in Figure 6.1. These plots show that, for σ_1, convergence is achieved very quickly in model 2 with both priors, and also in model 1 when the informative prior is used: the same effect was observed for all other parameters. For model 1 with the non-informative prior, σ_1 has not converged well within 1000 iterations. A similar effect was observed for all scale and shape parameters under model 1 with the informative prior. The dependence parameters, for both models and priors, converged very quickly. The number of iterations obtained in each of the four cases appeared to be adequate, but running the algorithms for longer would verify that the chains had converged.

Posterior densities of σ_1 for both models and priors, obtained after removal of observations from the burn-in periods, are given in Figure 6.2. For model 1, the posterior density based on the informative prior is slightly to the right of that based on the non-informative prior, and it is less spread out. This shows that the informative prior has reduced the uncertainty and has resulted in slightly higher values of σ_1 being favoured. For model 2, the difference between the posterior densities is much greater, with the density based on

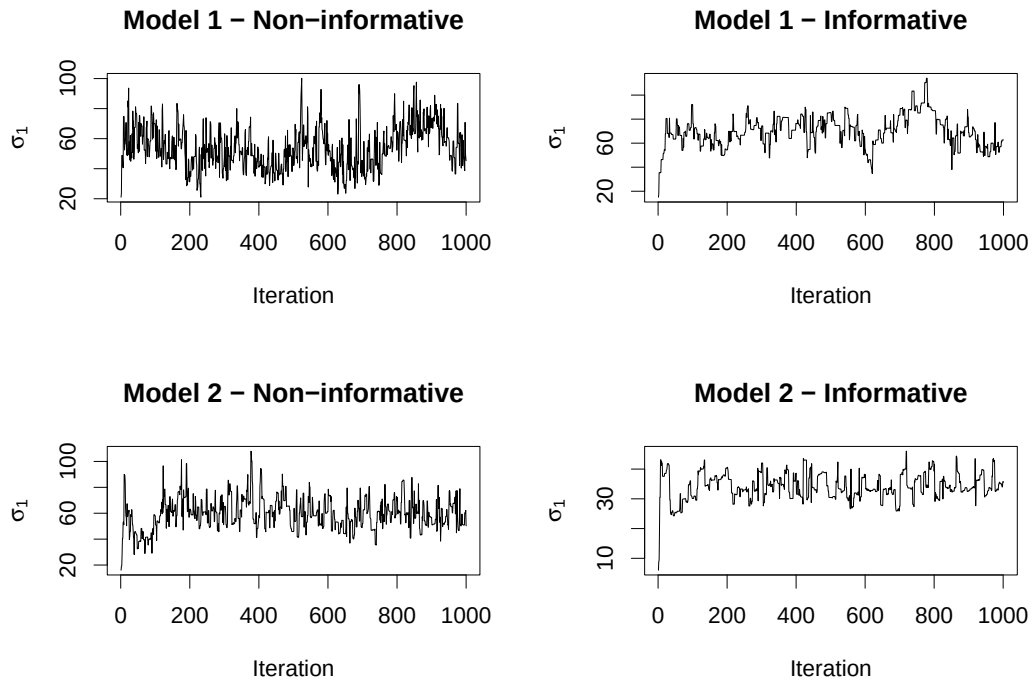


Figure 6.1: Trace plots of σ_1 for models 1 and 2 with non-informative and informative priors

the informative prior being to the left of the density based on the non-informative prior. The density based on the informative prior is much less spread out, suggesting a great reduction in uncertainty. There is a noticeable difference between the posterior densities of model 1 and model 2, with the posterior density of model 2 with the informative prior being significantly to the left of the corresponding posterior density of model 2. This suggests that assuming a constant shape parameter for all sites has had a large effect on the posterior distributions and may not be a sensible assumption to make.

A closer look at the results from the two models is achieved by considering the posterior means and standard deviations of all parameters. The posterior means and standard deviations for model 1, with the non-informative and informative priors, are given in Table 6.1.

Comparing the results obtained with the non-informative and informative priors shows that the informative prior has affected the posterior means and standard deviations of

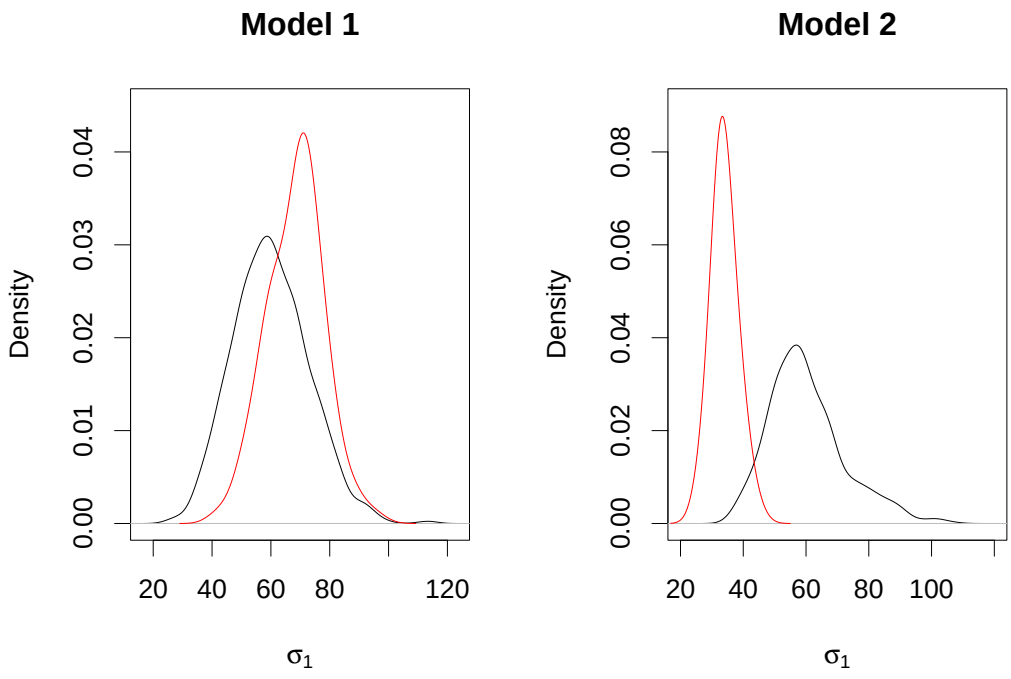


Figure 6.2: Posterior density plots of σ_1 for models 1 and 2 with non-informative (black) and informative (red) priors

the marginal GPD parameters, with the effect being most noticeable on the shape parameters. For sites 6, 9 and 10, the effect of the informative prior on the shape parameter is particularly noticeable, with great reductions in the posterior means and standard deviations being observed. Also, the change in posterior mean of ξ_1 is very significant, changing from 0.567 to -0.609 when the informative prior is used. This is likely to have a great impact on the posterior distribution of the return levels. The posterior standard deviations of almost all of the marginal GPD parameters are reduced by the inclusion of the informative prior, showing the extent to which the informative prior has reduced the uncertainty. The exceptions are the standard deviations of σ_4, σ_5 and ξ_5, which increase slightly when the informative prior is used. Although non-informative priors were assumed for the dependence parameters, the informative priors for the marginal GPD parameters have resulted in changes in the posterior means of the dependence parameters: all of the posterior means, apart from that of α_{10}, are slightly lower when the informative

prior is used.

In general, the posterior means of the scale parameters obtained in model 1 are fairly close to those obtained from using the logistic model (see Table 5.1 and Table 5.3 for the results from the logistic model with non-informative and informative priors respectively). The one noticeable difference is for σ_4, for which the posterior means based on the Dirichlet model, with both priors, are much higher than those based on the logistic model. In general, the posterior means of the shape parameters are higher for the Dirichlet model than for the logistic model. The two exceptions are for ξ_1 and ξ_9 under the informative prior. There is only a very slight difference between the posterior means of ξ_9 for the two models. There is a large difference between the posterior means of ξ_1 for the two models, with the posterior mean based on the Dirichlet model being negative and the posterior mean based on the logistic model being positive.

Posterior means and standard deviations of the scale and dependence parameters for model 2, with the non-informative and informative priors, are given in Table 6.2. The posterior means (standard deviations) of the shape parameter were 0.368 (0.0359) and 0.354 (0.0132) for the non-informative and informative priors respectively. Looking at the posterior means of the shape parameters obtained for model 1, the assumption of a constant shape parameter for all sites does not really seem reasonable. The range of posterior means for the shape parameters of model 1, for both priors, is very wide. Assuming a constant shape parameter, then, is likely to significantly affect the posterior distribution of the return levels. The posterior means of the scale and dependence parameters obtained with model 2 do differ from those obtained with model 1, although the difference is often quite small.

Using the informative prior does seem to have some effect on the posterior distribution of the scale parameters, with the posterior means changing slightly and the posterior standard deviations being reduced. The posterior mean of σ_1 changes greatly when the informative prior is used. The posterior mean of the shape parameter for model 2 is slightly different when the informative prior is used and the posterior standard deviation

is slightly reduced. This change is very small in comparison to the change, due to the informative prior, in the posterior means of the shape parameters of model 1.

To obtain realisations from the posterior distribution of return levels, the vectors of realisations from the marginal posterior densities of $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$ can be substituted into (1.9) for various values of N. Figure 6.3 gives return level plots, based on posterior median of return level, for models 1 and 2 with non-informative and informative priors. Both curves for model 1 are significantly lower than those for model 2. It seems, then, that assuming a constant shape parameter for all sites would result in the return levels being overestimated. For both models, the curve based on the non-informative prior is higher than the curve based on the informative prior, suggesting that return levels are overestimated when the expert prior information is not taken into account. For models 1 and 2, Figure 6.4

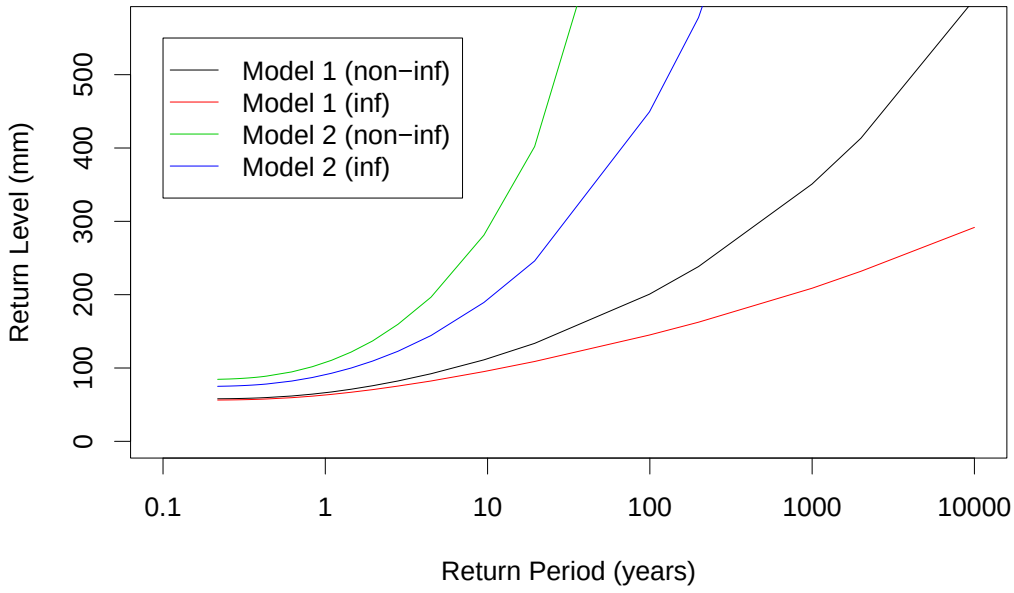


Figure 6.3: Return level plots for site 4 based on models 1 and 2 with non-informative and informative priors

gives the return level curve based on the informative prior with 95% credibility intervals based on both priors. In both cases, the bounds of the credibility intervals based on

the non-informative prior are higher than those based on the informative prior, and the intervals based on the non-informative prior do not contain the posterior medians from the informative prior for long return periods.

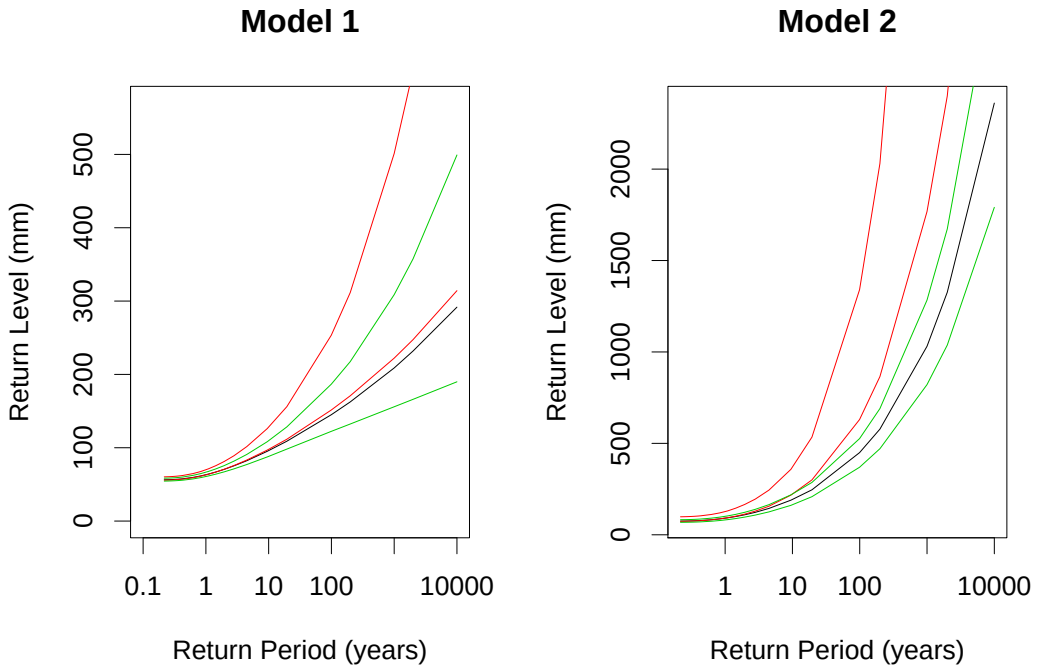


Figure 6.4: Return level plots for site 4 based on models 1 and 2 with the informative prior and 95% credibility intervals based on both priors

Posterior medians and 95% credibility intervals of the 10, 100 and 1000-year return levels for all sites are given for model 1 and for model 2 in Table 6.3 and Table 6.4 respectively. In general, the same pattern observed for site 4 was observed for all sites. Most of the posterior medians of the return levels were lower for model 1 than for model 2. Exceptions include the posterior medians observed with the non-informative prior for sites, such as 6, 9 and 10, which had large posterior mean shape parameters under model 1. Using model 2 reduces the posterior mean shape parameters, resulting in lower return level estimates. In all cases, using the informative posterior results in lower return level estimates. Also, under both models, the 95% credibility intervals based on the non-informative prior usually do not include the posterior medians based on the informative prior. This reflects

the benefit of using the expert information, since under the non-informative prior the 95% credibility intervals for the return levels often include implausibly high values. The informative prior reduces these values to more reasonable levels.

6.4 Validity of the model

The choice of sub-groups for the model used in Section 6.3 was fairly arbitrary. In order for the multivariate Dirichlet model to be valid for each group, however, the variables are required to be *asymptotically dependent*. In this section, graphical methods are used to determine whether the assumption of asymptotic dependence was valid for the groups of sites used in Section 6.3.

6.4.1 Asymptotic dependence

Suppose X and Y are random variables with distribution functions F_X and F_Y respectively. Define

$$\chi = \lim_{u \to 1} Pr\{F_Y(Y) > u | F_X(X) > u\},$$

and also, for $0 < u < 1$,

$$\begin{aligned} \chi(u) &= 2 - \frac{\log Pr\{F_X(X) < u, F_Y(Y) < u\}}{\log Pr\{F_X(X) < u\}}. \\ &= 2 - \frac{\log Pr\{F_X(X) < u, F_Y(Y) < u\}}{\log u}. \end{aligned}$$

It can be shown that

$$\chi = \lim_{u \to 1} \chi(u).$$

The properties of χ are:

1. $0 \leq \chi \leq 1$;

2. for the bivariate extreme value distribution with distribution function $G(x, y) = \exp\{-V(x, y)\}$, $\chi = 2 - V(1, 1)$;

3. $\chi = 0$, for asymptotically independent variables;

4. within the class of asymptotically dependent variables, the value of χ increases with strength of dependence at extreme levels.

A simple measure of extremal dependence within the class of asymptotically dependent variables is provided by χ. No information about the relative strength of dependence for asymptotically independent distributions is provided by χ, however. Another measure, $\bar{\chi}$, is used to overcome this. For $0 < u < 1$, let

$$\begin{aligned} \bar{\chi}(u) &= \frac{2\log Pr\{F_X(X) > u\}}{\log Pr\{F_X(X) > u, F_Y(Y) > u\}} - 1 \\ &= \frac{2\log(1-u)}{\log Pr\{F_X(X) > u, F_Y(Y) > u\}} - 1 \end{aligned}$$

and

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u).$$

The properties of $\bar{\chi}$ are:

1. $-1 \leq \bar{\chi} \leq 1$;

2. $\bar{\chi} = 1$, for asymptotically dependent variables;

3. $\bar{\chi} = 0$, for independent variables;

4. for asymptotically independent variables, $\bar{\chi}$ increases with strength of dependence at extreme levels.

Within the class of asymptotically independent variables, $\bar{\chi}$ provides a measure of strength of dependence. So, the pair $(\chi, \bar{\chi})$ provides a summary of extremal dependence for an arbitrary random vector. If $\bar{\chi} = 1$, the variables are asymptotically dependent and χ summarises the strength of extremal dependence. If $\bar{\chi} < 1$, then $\chi = 0$ and the variables are asymptotically independent. In this case $\bar{\chi}$ can be used as a measure of strength of extremal dependence.

Simple empirical estimates of the functions $\chi(u)$ and $\bar{\chi}(u)$ can be calculated by replacing the probabilities with the observed proportions. These empirical estimates can then be used to assess the model by plotting them as functions of u. The behaviour of the functions as $u \to 1$ can then be examined. It is then possible to determine whether a pair of sites is asymptotically independent or asymptotically dependent. This method is used in Section 6.4.2 to assess the suitability of the subgroups chosen for analysis in Section 6.3.

6.4.2 Assessment of the choice of sub-groups

Plots of $\chi(u)$ and $\bar{\chi}(u)$, with approximate 95% confidence intervals, were obtained for each pair of the sites of each sub-group. Plots for the pairs (4,5), (2,6) and (3,6) are given in figures 6.5, 6.6 and 6.7 respectively.

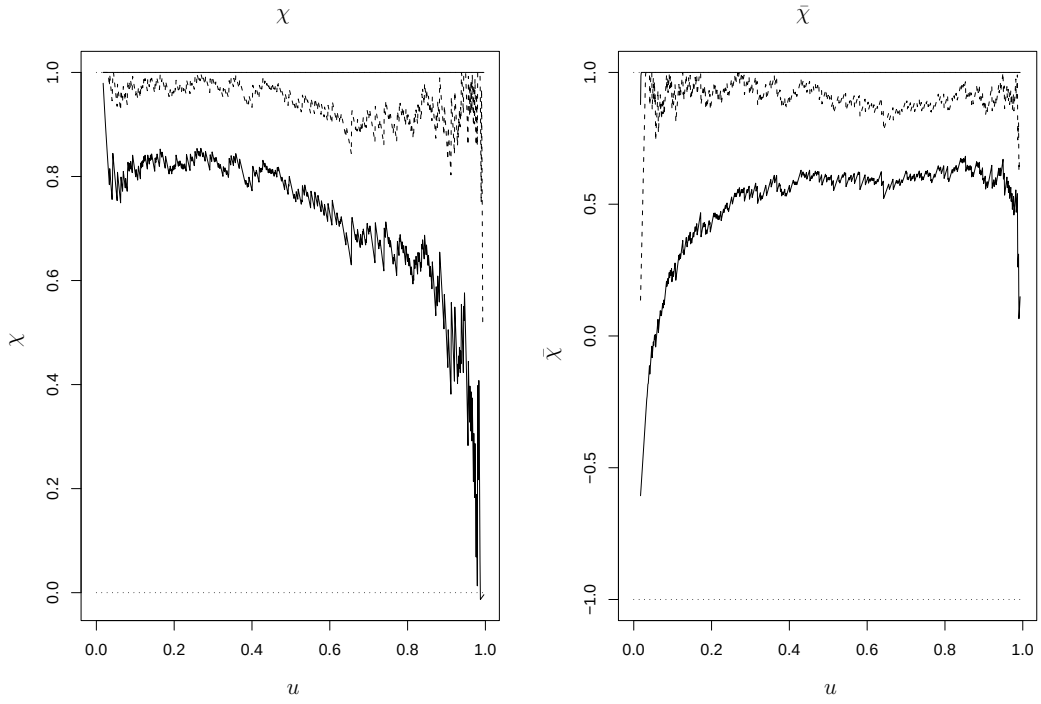


Figure 6.5: Plots of $\chi(u)$ and $\bar{\chi}(u)$ for the pair of sites (4,5)

For the pair (4,5), it is clear that $\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u) = 1$, so the assumption of asymptotic

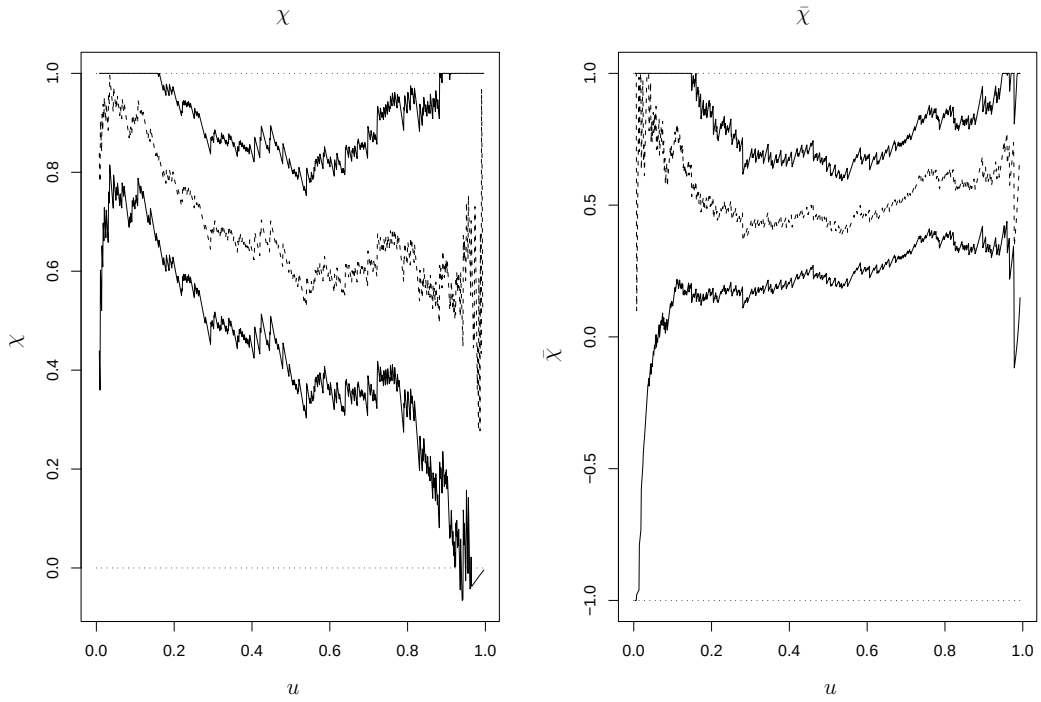


Figure 6.6: Plots of $\chi(u)$ and $\bar{\chi}(u)$ for the pair of sites (2,6)

dependence is valid for this pair. The upper confidence bound for this pair is not shown. For the pair (2,6), $\bar{\chi} = 1$ seems to be a possible limit. Also, $\chi = \lim_{u \to 1} \chi(u) \neq 0$ which suggests that the variables are not asymptotically independent. So, in this case, the assumption of asymptotic dependence may be valid. For other pairs of sites, however, the assumption does not appear to be valid. The plots for the pair (3,6), for example, show that $\bar{\chi} \neq 1$ and $\chi = 0$, suggesting asymptotic independence. Other pairs that appear to be asymptotically independent are (1,4), (1,5), (2,3), and (7,9). It appears then that the combinations of sites used in the four sub-groups is not appropriate.

After plotting and examining the functions $\chi(u)$ and $\bar{\chi}$ for all 55 pairs of sites, a more suitable choice of sub-groups appears to be (1,2,6), (3,4,5), (7,8) and (9,10,11). In this case, all pairs of sites of the sub-groups appear to be asymptotically dependent. Carrying out the analysis of Section 6.3 with these groups, then, would be more appropriate.

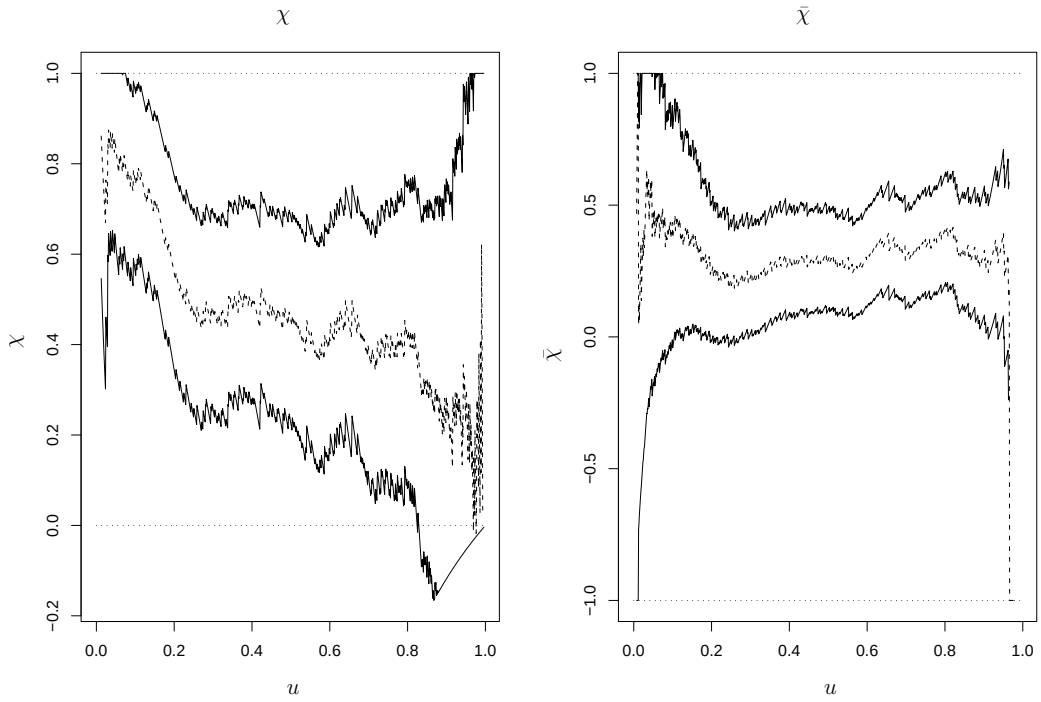


Figure 6.7: Plots of $\chi(u)$ and $\bar{\chi}(u)$ for the pair of sites (3,6)

6.5 Discussion

In this chapter, the multivariate Dirichlet model is used to model the daily rainfall data from the 11 sites in Figure 2.1. Unlike the logistic model, this model allows for different levels of pairwise dependence and bivariate asymmetry. To use a high dimensional Dirichlet model would be very time consuming, and Ledford and Tawn (1996, 1997, 1998) showed that using such a multivariate model in high dimensions is not appropriate. To overcome these problems, the multivariate Dirichlet model was applied to four subgroups of the sites in Figure 2.1. Initially, the models were applied independently to the data from each group of sites. Then a model which linked the groups together, by assuming a constant shape parameter, was used. In both models, non-informative and informative priors were used in the same way as in Chapter 5.

Assuming a constant shape parameter was not found to be a reasonable assumption, since under the first model the range of posterior means of the shape parameters was

very wide. The effect of assuming a constant shape parameter was noticed particularly in the return level estimation, with the return level estimates often being higher than those obtained under model 1.

As in Chapter 5, the informative prior had a significant effect on parameter and return level estimation. The posterior medians of return levels were lower under the informative prior than under the non-informative prior, with the levels obtained under the informative prior generally being more plausible. The informative prior also resulted in reduced uncertainty, with lower posterior standard deviations for the parameters and narrower credibility intervals generally being observed.

In section 6.4, the validity of the model used was checked. To do this, the assumption of asymptotic dependence was checked. It was found that this assumption was valid for some pairs of sites but not all. A more suitable partition of the sites was suggested.

This chapter, along with Chapter 5, has demonstrated the value of expert prior information. It has also shown that for the rainfall data from the region in 2.1 it is unreasonable to assume a constant shape parameter for each site. In addition to the results obtained in this chapter, predictive return level estimates could also have been obtained in the same way as those obtained in Chapter 5. Any further work, however, should be carried out with a partition of sites such as that suggested in Section 6.4.

	Non-informative prior			Informative prior		
j	σ_j	ξ_j	α_j	σ_j	ξ_j	α_j
1	60.083	0.567	0.365	68.363	-0.609	0.361
	(12.821)	(0.667)	(0.0254)	(9.733)	(0.189)	(0.0256)
2	25.828	0.618	3.867	20.798	0.177	3.141
	(5.398)	(0.251)	(0.485)	(2.898)	(0.0737)	(0.381)
3	20.985	0.101	1.755	20.745	-0.0800	1.559
	(2.773)	(0.194)	(0.166)	(2.237)	(0.108)	(0.146)
4	11.774	0.213	83.634	12.386	0.103	77.300
	(0.528)	(0.0519)	(10.192)	(0.534)	(0.0453)	(12.671)
5	10.625	0.488	34.104	11.182	0.270	31.262
	(0.844)	(0.0329)	(4.086)	(0.853)	(0.0537)	(3.835)
6	6.510	1.451	3.904	10.255	0.396	3.692
	(1.142)	(0.182)	(0.471)	(1.298)	(0.0599)	(0.487)
7	9.480	0.214	2.832	9.221	0.185	2.536
	(0.907)	(0.0891)	(0.319)	(0.720)	(0.0554)	(0.272)
8	12.344	0.210	3.598	11.301	0.180	3.271
	(1.398)	(0.104)	(0.403)	(1.223)	(0.0840)	(0.332)
9	13.776	1.542	1.941	17.791	0.249	1.919
	(3.180)	(0.457)	(0.203)	(3.079)	(0.0949)	(0.197)
10	10.097	1.662	11.592	9.913	0.450	15.105
	(2.582)	(0.327)	(7.181)	(1.530)	(0.0586)	(10.775)
11	7.091	0.562	1.858	7.781	0.324	1.540
	(0.928)	(0.123)	(0.210)	(0.675)	(0.0563)	(0.160)

Table 6.1: Posterior means (posterior standard deviations) of parameters in the Dirichlet model with non-constant shape parameter using both non-informative and informative priors

	Non-informative prior		Informative prior	
j	σ_j	α_j	σ_j	α_j
1	60.173	0.513	34.023	0.341
	(12.148)	(0.0364)	(3.752)	(0.0238)
2	23.980	3.553	21.943	3.329
	(3.628)	(0.432)	(1.820)	(0.361)
3	16.691	1.698	16.315	1.646
	(1.794)	(0.174)	(1.286)	(0.147)
4	10.438	88.502	11.820	85.853
	(0.380)	(8.851)	(0.306)	(10.587)
5	14.312	39.553	15.603	37.283
	(0.866)	(4.393)	(0.698)	(3.669)
6	11.994	3.951	13.880	3.914
	(1.187)	(0.483)	(0.787)	(0.500)
7	7.500	2.801	8.871	2.826
	(0.532)	(0.305)	(0.312)	(0.314)
8	10.640	3.677	12.484	3.658
	(1.101)	(0.411)	(0.865)	(0.358)
9	21.312	2.027	19.643	2.000
	(3.520)	(0.202)	(1.943)	(0.194)
10	15.427	22.388	15.719	16.250
	(3.180)	(19.275)	(1.774)	(10.045)
11	7.241	1.636	8.772	1.698
	(0.575)	(0.192)	(0.541)	(0.184)

Table 6.2: Posterior means (posterior standard deviations) of the scale and dependence parameters in the Dirichlet model with constant shape parameter using both non-informative and informative priors

	Non-informative prior			Informative prior		
	Return period (years)			Return period (years)		
Site	10	100	1000	10	100	1000
1	321.867	1355.077	5416.865	133.080	153.395	158.753
	(140,2623)	(164,204545)	(172,20380264)	(122,157)	(134,205)	(135,237)
2	240.143	966.484	3889.928	126.400	221.238	364.968
	(160,512)	(288,6316)	(454,85691)	(113,141)	(183,297)	(261,623)
3	123.558	198.744	294.286	100.615	134.684	163.620
	(95,210)	(115,638)	(127,2003)	(87,119)	(101,190)	(108,276)
4	111.115	200.669	350.932	95.129	144.998	208.773
	(97,126)	(151,253)	(222,501)	(87,108)	(122,187)	(156,309)
5	133.783	360.427	1039.156	102.918	189.862	351.448
	(117,157)	(299,528)	(821,1918)	(93,119)	(153,261)	(248,609)
6	815.424	19577.460	495873.300	117.192	278.679	674.559
	(447,2098)	(6710,116978)	(101032,6535703)	(104,133)	(215,364)	(430,1101)
7	88.035	161.552	283.211	81.508	140.611	230.510
	(72,116)	(104,286)	(140,698)	(71,96)	(109,199)	(155,411)
8	90.618	163.901	278.649	82.571	140.903	225.683
	(75,117)	(109,280)	(147,704)	(73,97)	(108,212)	(145,456)
9	587.515	18067.330	555152.800	104.524	212.882	406.310
	(190,4221)	(1048,1232728)	(5823,359604956)	(90,120)	(157,296)	(234,778)
10	668.848	30377.020	10001386104	86.702	231.810	642.503
	(222,3191)	(2589,602938)	(31263,125316472)	(74,103)	(173,310)	(382,1000)
11	130.225	480.812	1832.313	83.920	181.061	384.779
	(88,189)	(189,1025)	(400,4924)	(74,98)	(135,253)	(233,669)

Table 6.3: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels at each site for model 1, obtained using non-informative and informative priors

	Non-informative prior			Informative prior		
	Return period (years)			Return period (years)		
Site	10	100	1000	10	100	1000
1	305.515	1151.897	3884.934	169.075	441.077	1049.007
	(230,461)	(782,1910)	(2375,7072)	(145,201)	(358,545)	(850,1334)
2	203.171	649.533	2116.922	157.492	368.335	853.938
	(161,260)	(461,930)	(1354,3406)	(139,179)	(312,448)	(686,1080)
3	211.977	684.314	2218.410	155.080	365.231	837.418
	(169,263)	(486,961)	(1350,3647)	(135,177)	(305,425)	(674,1026)
4	187.293	587.878	1902.833	144.825	330.140	748.733
	(162,218)	(444,816)	(1230,3172)	(135,157)	(296,379)	(643,919)
5	178.671	555.285	1778.810	144.541	330.296	750.619
	(152,211)	(413,773)	(1146,2995)	(134,160)	(292,388)	(635,935)
6	161.528	498.715	1597.026	137.136	314.273	711.140
	(137,191)	(377,686)	(1051,2620)	(124,155)	(270,375)	(580,902)
7	140.792	448.908	1454.741	111.244	256.737	588.775
	(121,171)	(338,640)	(968,2466)	(102,125)	(227,307)	(491,750)
8	127.370	390.146	1252.028	111.768	257.320	582.479
	(108,155)	(301,562)	(831,2152)	(100,128)	(223.313)	(495,753)
9	161.077	535.457	1768.606	127.663	313.076	732.217
	(129,211)	(398,798)	(1170,3085)	(107,144)	(250,375)	(563,927)
10	126.538	411.391	1335.048	105.885	256.819	597.887
	(97,185)	(276,665)	(801,2319)	(93,128)	(216,320)	(484,772)
11	114.780	361.176	1174.465	97.573	228.080	519.961
	(101,142)	(283,525)	(798,2026)	(88,110)	(198,270)	(438,649)

Table 6.4: Posterior medians (95% credibility intervals) for the 10, 100 and 1000-year return levels at each site for model 2, obtained using non-informative and informative priors

Appendix A

Appendix to Chapter 4

Table A.1: Posterior means of the bivariate EV parameters for each pair of sites using uninformative priors and logistic dependence

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(1, 2)$	56.7516	14.8932	0.1208	66.0336	13.6993	0.2222	0.5477
$(1, 3)$	55.6363	14.4580	0.2583	59.4606	9.8256	0.2080	0.6653
$(1, 4)$	56.9034	15.2997	0.1843	63.9620	13.0166	0.0655	0.6235
$(1, 5)$	56.5324	15.0617	0.1559	60.6995	12.3605	0.0897	0.6135
$(1, 6)$	56.2702	14.2099	0.2276	59.1780	9.4797	0.4037	0.7932
$(1, 7)$	55.4529	16.5589	0.0916	44.3848	8.6284	0.1068	0.8026
$(1, 8)$	54.2212	13.3145	0.2557	46.2980	9.0538	0.1907	0.7679
$(1, 9)$	55.4751	14.0017	0.1822	43.6768	9.8282	0.3241	0.8122
$(1, 10)$	54.3820	16.370	0.1398	40.1861	11.6256	0.5239	0.8132
$(1, 11)$	53.9564	15.1283	0.0989	39.5705	8.0720	0.2964	0.6555
$(2, 3)$	67.6880	13.4272	0.2989	59.4669	7.6391	0.4273	0.7257
$(2, 4)$	66.5687	12.7051	0.2128	66.2755	13.9956	0.0143	0.6782
$(2, 5)$	66.4967	12.4783	0.1884	62.6523	14.1350	-0.0561	0.6283

Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(2,6)$	66.7578	13.1464	0.3688	59.2407	9.6765	0.4443	0.6056
$(2,7)$	66.7405	13.2355	0.2272	44.3375	8.4819	0.1177	0.7577
$(2,8)$	65.5839	12.4459	0.2233	47.0801	9.4294	0.1913	0.6688
$(2,9)$	66.4015	11.8680	0.1849	43.3463	9.3014	0.3065	0.7418
$(2,10)$	68.3501	14.5551	0.2150	41.3731	12.4578	0.4993	0.7817
$(2,11)$	66.4858	13.7547	0.2310	41.3992	9.0270	0.2393	0.7623
$(3,4)$	59.3359	7.5138	0.4677	62.2624	12.5873	0.0920	0.7266
$(3,5)$	59.4154	9.6728	0.1811	60.0624	11.7370	0.1876	0.6684
$(3,6)$	59.7108	10.1107	0.2152	59.0174	10.2170	0.4778	0.7776
$(3,7)$	59.9318	10.2889	0.1958	44.7091	9.5213	0.0037	0.8970
$(3,8)$	57.7663	8.0424	0.1206	46.8128	7.6771	0.2050	0.8203
$(3,9)$	58.8085	9.0494	0.1593	43.6688	10.4864	0.3271	0.8013
$(3,10)$	59.4291	10.2232	0.2414	42.9482	15.5912	0.6043	0.7051
$(3,11)$	59.6661	10.0035	0.2133	42.5681	10.3822	0.2278	0.8199
$(4,5)$	65.3760	13.4410	0.1206	62.7338	13.4834	0.0750	0.3342
$(4,6)$	64.2939	13.5504	0.0373	59.3822	9.7559	0.4189	0.7290
$(4,7)$	66.9497	14.1899	-0.0022	44.1885	8.1917	0.1016	0.8561
$(4,8)$	62.7578	11.9503	0.0025	46.7360	8.9814	0.1366	0.7776
$(4,9)$	65.1840	13.1160	0.0631	43.0854	9.0612	0.3016	0.8298
$(4,10)$	65.7279	14.4082	0.0466	41.7246	12.9449	0.4991	0.8758
$(4,11)$	63.7365	12.8224	0.0203	42.0808	10.1467	0.2333	0.7710
$(5,6)$	59.8136	12.2679	0.2134	59.1962	9.5727	0.4097	0.7495
$(5,7)$	63.0146	13.3611	0.0756	44.1477	8.1085	0.1050	0.8810
$(5,8)$	58.9642	10.8308	0.1144	46.5497	8.9653	0.1660	0.8488
$(5,9)$	60.7631	12.1978	0.0670	43.0211	8.9634	0.3104	0.8117
$(5,10)$	61.9419	13.5240	0.1437	41.1549	12.6702	0.5243	0.8457

Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(5,11)$	60.1513	12.0636	0.1528	41.6852	9.8239	0.2328	0.7032
$(6,7)$	59.0654	9.6958	0.4583	45.2930	8.9643	-0.0240	0.8189
$(6,8)$	57.5358	8.4227	0.4457	47.7240	7.8845	0.0625	0.7127
$(6,9)$	58.0940	8.9550	0.4724	43.7881	9.8076	0.2883	0.7616
$(6,10)$	59.6279	10.1362	0.4252	42.2278	13.5165	0.5199	0.9288
$(6,11)$	59.3502	9.7634	0.4242	41.8480	9.8663	0.2536	0.8266
$(7,8)$	44.5972	9.1406	0.0777	46.7154	8.9085	0.1434	0.5841
$(7,9)$	43.9184	8.0974	0.1321	43.3773	9.0296	0.3031	0.7632
$(7,10)$	44.4905	8.4150	0.1071	39.9356	11.2519	0.5222	0.8619
$(7,11)$	44.6228	8.3212	-0.0237	40.5698	9.1998	0.3249	0.9076
$(8,9)$	46.6403	9.0441	0.2004	43.2398	9.3556	0.3413	0.6150
$(8,10)$	46.7346	9.2292	0.2194	39.7204	11.0819	0.4771	0.7322
$(8,11)$	47.0795	8.6601	-0.0554	40.9906	8.4064	0.3132	0.8692
$(9,10)$	43.5190	9.5339	0.3690	40.3147	11.9533	0.5347	0.8308
$(9,11)$	43.5718	9.3114	0.2730	40.9708	8.9628	0.2718	0.8683
$(10,11)$	39.6956	11.1102	0.5273	40.5464	8.6344	0.3989	0.8048

Table A.2: Posterior standard deviations of the bivariate EV parameters for each pair of sites using uninformative priors and logistic dependence

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(1, 2)$	2.4788	2.2121	0.1351	2.2257	1.9141	0.1052	0.0803
$(1, 3)$	2.7226	2.4443	0.1713	1.8226	1.5549	0.1302	0.1012
$(1, 4)$	2.5987	2.2567	0.1497	2.1989	1.8150	0.1272	0.0838
$(1, 5)$	2.4958	2.1805	0.1376	2.0890	1.7757	0.1319	0.0809
$(1, 6)$	2.4598	2.1302	0.1512	1.5596	1.4688	0.1213	0.0920
$(1, 7)$	2.5276	2.0084	0.1163	1.3225	1.0218	0.0908	0.0848
$(1, 8)$	2.4284	2.1890	0.1734	1.6120	1.3279	0.1130	0.0962
$(1, 9)$	2.5220	2.1503	0.1549	1.7040	1.5627	0.1309	0.0847
$(1, 10)$	2.5239	2.0274	0.1167	1.7865	1.8123	0.1302	0.0836
$(1, 11)$	2.2529	1.7401	0.1001	1.2560	1.1229	0.1343	0.0814
$(2, 3)$	2.5052	2.2511	0.1356	1.5157	1.5329	0.1922	0.1016
$(2, 4)$	1.8509	1.5009	0.0863	2.0600	1.6656	0.1088	0.0785
$(2, 5)$	1.7922	1.4238	0.0810	2.1668	1.8120	0.1239	0.0742
$(2, 6)$	2.2961	2.1067	0.1408	1.6699	1.6270	0.1346	0.0996
$(2, 7)$	1.9732	1.6336	0.1033	1.2775	0.9925	0.0892	0.0910
$(2, 8)$	2.1664	1.8019	0.1175	1.6547	1.3702	0.1132	0.1020
$(2, 9)$	1.7317	1.3652	0.0875	1.4096	1.2440	0.1157	0.0867
$(2, 10)$	2.2260	1.8003	0.1074	1.9504	1.9199	0.1285	0.0892
$(2, 11)$	2.2237	1.8195	0.1114	1.5059	1.2976	0.1314	0.0878
$(3, 4)$	1.4873	1.5029	0.1901	2.4692	2.0926	0.1651	0.1009
$(3, 5)$	1.7753	1.4889	0.1262	2.1806	1.9611	0.1451	0.0989

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(3, 6)$	1.8463	1.5839	0.1355	1.8996	1.9370	0.1461	0.1040
$(3, 7)$	1.9365	1.6543	0.1438	1.8281	1.4203	0.1413	0.0765
$(3, 8)$	1.6332	1.3229	0.1314	1.6732	1.4705	0.2120	0.1064
$(3, 9)$	1.7373	1.4704	0.1323	2.0687	1.9225	0.1720	0.1018
$(3, 10)$	1.9262	1.6953	0.1460	3.0337	3.3242	0.1801	0.1074
$(3, 11)$	1.8410	1.5610	0.1370	1.9812	1.7289	0.1630	0.0977
$(4, 5)$	1.8879	1.4929	0.0958	1.8780	1.4834	0.0966	0.0504
$(4, 6)$	2.4154	2.0006	0.1664	1.6846	1.6086	0.1271	0.0993
$(4, 7)$	2.1381	1.6970	0.1230	1.2189	0.9248	0.0844	0.0747
$(4, 8)$	2.1106	1.7346	0.1296	1.5671	1.2741	0.0977	0.0920
$(4, 9)$	1.9795	1.5409	0.1176	1.3536	1.1938	0.1147	0.0784
$(4, 10)$	2.4196	1.8970	0.1476	2.1023	2.1233	0.1372	0.0745
$(4, 11)$	2.2137	1.8183	0.1509	1.7362	1.5513	0.1447	0.0879
$(5, 6)$	2.1344	1.8420	0.1476	1.6326	1.5351	0.1241	0.0950
$(5, 7)$	2.0400	1.6432	0.1315	1.1947	0.9222	0.0856	0.0702
$(5, 8)$	1.9020	1.5768	0.1398	1.5435	1.2679	0.1032	0.0798
$(5, 9)$	1.8381	1.4352	0.1191	1.3230	1.1682	0.1143	0.0785
$(5, 10)$	2.2112	1.8505	0.1538	2.0391	2.1005	0.1379	0.0825
$(5, 11)$	2.0060	1.6847	0.1363	1.6840	1.4995	0.1408	0.0869
$(6, 7)$	1.6412	1.6074	0.1334	1.5300	1.1509	0.1118	0.0992
$(6, 8)$	1.5843	1.5607	0.1334	1.5522	1.2783	0.1611	0.1102
$(6, 9)$	1.5731	1.5746	0.1347	1.7751	1.5931	0.1424	0.1072
$(6, 10)$	1.7646	1.7013	0.1358	2.4173	2.4895	0.1515	0.0592

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(6, 11)$	1.6503	1.5960	0.1301	1.7462	1.5552	0.1543	0.0896
$(7, 8)$	1.5529	1.2517	0.1006	1.5054	1.2387	0.0911	0.0928
$(7, 9)$	1.2033	0.9242	0.0891	1.3710	1.2207	0.1184	0.0898
$(7, 10)$	1.2505	0.9553	0.0886	1.6667	1.6910	0.1304	0.0770
$(7, 11)$	1.2768	0.9396	0.0994	1.4905	1.3352	0.1578	0.0651
$(8, 9)$	1.5194	1.2302	0.1003	1.6045	1.4785	0.1306	0.0931
$(8, 10)$	1.5541	1.2937	0.1150	1.9162	1.9087	0.1515	0.1030
$(8, 11)$	1.5777	1.1538	0.1202	1.5906	1.5024	0.1635	0.0804
$(9, 10)$	1.5957	1.5310	0.1331	1.9713	2.0458	0.1451	0.0906
$(9, 11)$	1.6107	1.4139	0.1373	1.5716	1.3814	0.1500	0.0838
$(10, 11)$	1.6799	1.7379	0.1306	1.3761	1.2799	0.1559	0.0903

Table A.3: Posterior means of the bivariate EV parameters for each pair of sites using expert priors and logistic dependence

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(1, 2)$	54.6489	12.6762	0.1784	64.4871	12.1892	0.2078	0.5780
$(1, 3)$	54.2854	12.7551	0.1992	58.1771	8.6687	0.2498	0.6837
$(1, 4)$	55.3736	13.8644	0.1826	61.8648	11.2877	0.1923	0.6374
$(1, 5)$	55.0131	13.6788	0.1803	58.9443	10.7943	0.2027	0.6288
$(1, 6)$	54.9225	12.5117	0.1944	58.5148	8.2286	0.2885	0.8310
$(1, 7)$	54.2162	15.5914	0.1453	43.7293	8.5977	0.2363	0.7961
$(1, 8)$	53.3079	12.0202	0.2083	45.5213	8.5111	0.2506	0.7884
$(1, 9)$	54.0950	12.4114	0.1925	43.1210	8.9705	0.2649	0.8429
$(1, 10)$	52.6218	14.5627	0.1544	39.810	9.9997	0.2845	0.8620
$(1, 11)$	52.6784	14.1274	0.1552	39.2007	7.6681	0.2893	0.6598
$(2, 3)$	66.3618	11.5334	0.2173	58.8440	6.6105	0.3173	0.7382
$(2, 4)$	65.7221	11.9931	0.2065	64.3122	12.5874	0.1642	0.6816
$(2, 5)$	65.5767	11.9778	0.2061	60.3172	12.2597	0.1645	0.6402
$(2, 6)$	65.5976	11.0743	0.2300	58.3755	7.9834	0.2925	0.6850
$(2, 7)$	66.1301	12.4673	0.2007	43.6786	8.4178	0.2368	0.7554
$(2, 8)$	64.5665	11.4203	0.2103	46.1625	8.8823	0.2451	0.6655
$(2, 9)$	65.5979	11.2087	0.2011	42.9975	8.7904	0.2670	0.7523
$(2, 10)$	66.9196	12.8926	0.1825	40.8377	10.6322	0.2705	0.8300
$(2, 11)$	65.5236	12.5749	0.1975	40.7480	8.3667	0.2641	0.7796
$(3, 4)$	59.0123	6.8415	0.3197	60.1636	10.6605	0.2111	0.7365
$(3, 5)$	58.3510	8.9237	0.2452	58.8241	10.5586	0.2242	0.6593

Pair (i, j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(3, 6)$	58.5748	9.0328	0.2439	58.2988	8.5480	0.2919	0.8018
$(3, 7)$	59.1210	9.5696	0.2390	43.5422	9.2054	0.2326	0.8821
$(3, 8)$	57.0180	8.0373	0.2601	46.2037	7.1682	0.2958	0.7701
$(3, 9)$	57.8992	8.5447	0.2496	43.0065	9.4947	0.2544	0.8139
$(3, 10)$	58.0141	8.7073	0.2495	42.0295	12.5591	0.2428	0.7964
$(3, 11)$	58.8016	9.1778	0.2460	41.7854	9.5278	0.2441	0.8331
$(4, 5)$	63.8581	12.6821	0.1881	61.1411	12.5171	0.1721	0.3175
$(4, 6)$	61.9613	11.5681	0.1912	58.8079	8.5519	0.2890	0.7543
$(4, 7)$	65.4445	13.3425	0.1625	43.6649	8.3562	0.2402	0.8348
$(4, 8)$	61.1904	10.9933	0.1986	46.0421	8.9252	0.2438	0.7690
$(4, 9)$	63.8254	12.1340	0.1777	42.8012	8.6406	0.2702	0.8451
$(4, 10)$	63.6195	12.6382	0.1755	41.2936	11.1187	0.2624	0.9170
$(4, 11)$	61.9101	11.3990	0.1939	41.4209	9.4671	0.2480	0.7846
$(5, 6)$	58.4907	10.7268	0.2177	58.5610	8.3131	0.2898	0.7751
$(5, 7)$	61.9130	12.5488	0.1809	43.6913	8.3061	0.2390	0.8671
$(5, 8)$	57.9959	10.1593	0.2216	46.0018	8.8178	0.2457	0.8431
$(5, 9)$	59.5155	11.3861	0.1899	42.7395	8.5188	0.2741	0.8278
$(5, 10)$	60.2705	11.7949	0.1966	0.7200	10.8317	0.2684	0.8958
$(5, 11)$	59.0007	10.9790	0.2109	40.9710	9.1019	0.2538	0.7116
$(6, 7)$	58.7287	8.5402	0.2946	44.0453	8.8781	0.2248	0.7997

Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(6,8)$	57.1794	7.4178	0.3154	46.4171	6.8831	0.2862	0.7349
$(6,9)$	57.5635	7.5660	0.3104	42.9717	8.7595	0.2599	0.8143
$(6,10)$	59.0540	8.7467	0.2801	41.4442	11.1351	0.2571	0.9602
$(6,11)$	58.8595	8.5368	0.2858	41.1045	8.9525	0.2546	0.8707
$(7,8)$	43.8099	9.1088	0.2305	46.1882	9.0689	0.2437	0.5424
$(7,9)$	43.4256	8.1184	0.2420	43.2070	8.7206	0.2744	0.7583
$(7,10)$	44.0005	8.5601	0.2286	39.8184	10.1143	0.2847	0.9070
$(7,11)$	43.8489	8.5647	0.2263	40.4026	8.7702	0.2672	0.9265
$(8,9)$	46.0305	8.6550	0.2431	42.9325	8.8200	0.2752	0.6155
$(8,10)$	45.9781	8.5998	0.2454	39.4176	9.7180	0.2744	0.7787
$(8,11)$	45.9202	8.6391	0.2354	40.7969	7.9847	0.2846	0.8833
$(9,10)$	43.0274	8.5409	0.2728	39.8342	10.0396	0.2795	0.8722
$(9,11)$	43.1934	8.7908	0.2605	40.5285	8.3762	0.2690	0.8864
$(10,11)$	39.3926	9.5412	0.2892	40.2013	7.8900	0.2891	0.8544

Table A.4: Posterior standard deviations of the bivariate EV parameters for each pair of sites using expert priors and logistic dependence

Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(1,2)$	2.0122	1.6180	0.0523	1.9480	1.4928	0.0457	0.0720
$(1,3)$	2.3492	1.9402	0.0537	1.5649	1.1190	0.0490	0.0908
$(1,4)$	2.2333	1.8738	0.0518	1.7809	1.3482	0.0511	0.0770
$(1,5)$	2.1621	1.8104	0.0518	1.6818	1.2918	0.0508	0.0766
$(1,6)$	2.1325	1.7238	0.0511	1.3896	1.0978	0.0447	0.0790
$(1,7)$	2.3245	1.7944	0.0478	1.2927	0.9437	0.0400	0.0879
$(1,8)$	2.1119	1.8386	0.0528	1.4885	1.1014	0.0499	0.0905
$(1,9)$	2.1256	1.7442	0.0520	1.5053	1.2939	0.0497	0.0767
$(1,10)$	2.2482	1.6772	0.0480	1.5394	1.3499	0.0431	0.0666
$(1,11)$	2.0449	1.4637	0.0465	1.1432	0.9873	0.0495	0.0737
$(2,3)$	2.1368	1.6561	0.04800	1.2285	1.1300	0.0551	0.0818
$(2,4)$	1.6840	1.2594	0.0425	1.7653	1.3309	0.0513	0.0716
$(2,5)$	1.6735	1.2608	0.0417	1.6880	1.3006	0.0525	0.0706
$(2,6)$	1.9440	1.4447	0.0460	1.4340	1.0706	0.0449	0.0829
$(2,7)$	1.8586	1.4162	0.0435	1.2740	0.8996	0.0476	0.0866
$(2,8)$	2.0079	1.4591	0.0461	1.5532	1.1434	0.0488	0.0894
$(2,9)$	1.6238	1.1721	0.0439	1.2877	1.0482	0.0457	0.0803
$(2,10)$	2.0047	1.4593	0.0450	1.6780	1.3994	0.0426	0.0692
$(2,11)$	2.0331	1.5473	0.0459	1.3510	1.0876	0.0501	0.0789
$(3,4)$	1.2933	1.2150	0.0530	1.9637	1.5249	0.0529	0.0851
$(3,5)$	1.6163	1.1933	0.0477	1.8861	1.5361	0.0520	0.0879

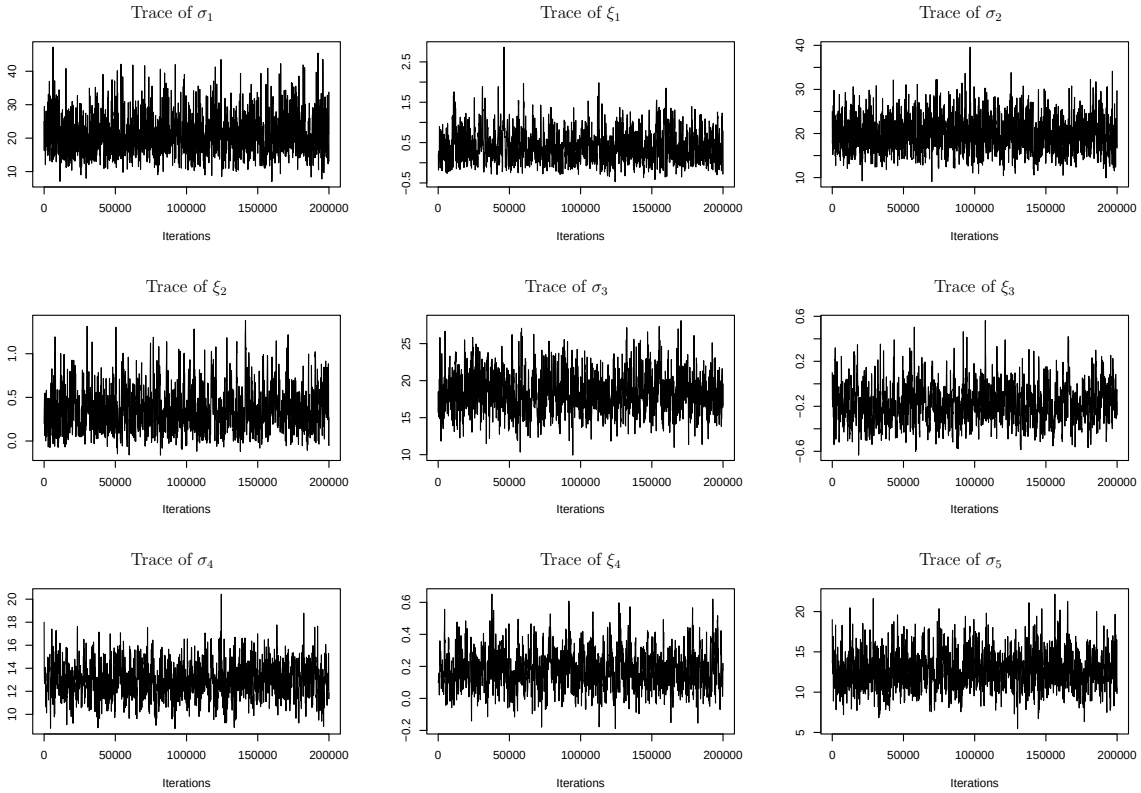
Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(3,6)$	1.6296	1.2171	0.0481	1.5873	1.3210	0.0468	0.0861
$(3,7)$	1.7337	1.3780	0.0481	1.6461	1.2691	0.0524	0.0795
$(3,8)$	1.6349	1.1813	0.0494	1.4310	1.2777	0.0561	0.1052
$(3,9)$	1.6205	1.1897	0.0484	1.8104	1.5851	0.0540	0.0943
$(3,10)$	1.6549	1.2280	0.0481	2.4550	2.2486	0.0502	0.0839
$(3,11)$	1.6505	1.2762	0.0479	1.7564	1.4417	0.05335	0.0916
$(4,5)$	1.7854	1.3019	0.0448	1.7352	1.2413	0.0459	0.0405
$(4,6)$	1.9800	1.5065	0.0515	1.4611	1.1497	0.0441	0.0853
$(4,7)$	1.9259	1.5164	0.0503	1.2213	0.8858	0.0474	0.0813
$(4,8)$	1.8568	1.4590	0.0513	1.5335	1.1704	0.0498	0.0935
$(4,9)$	1.7757	1.3490	0.0487	1.2987	1.0618	0.0463	0.0754
$(4,10)$	2.0354	1.5530	0.0505	1.7916	1.5669	0.0442	0.0573
$(4,11)$	1.8796	1.4454	0.0510	1.5192	1.3111	0.0517	0.0875
$(5,6)$	1.7921	1.4262	0.04987	1.4225	1.1057	0.0442	0.0797
$(5,7)$	1.8236	1.4538	0.0488	1.1955	0.8548	0.0475	0.0753
$(5,8)$	1.7322	1.3633	0.0504	1.4738	1.1217	0.0488	0.0819
$(5,9)$	1.6809	1.2963	0.0492	1.2600	1.0493	0.0465	0.0773
$(5,10)$	1.8908	1.4907	0.0511	1.7567	1.5099	0.0440	0.0643
$(5,11)$	1.7920	1.4160	0.0492	1.4793	1.2568	0.0522	0.0815
$(6,7)$	1.4527	1.2021	0.0437	1.4780	1.0470	0.0515	0.0829

Pair (i,j)	μ_i	σ_i	ξ_i	μ_j	σ_j	ξ_j	α
$(6,8)$	1.3796	1.1010	0.0435	1.3017	1.0527	0.0564	0.0973
$(6,9)$	1.3549	1.0957	0.0440	1.5708	1.2692	0.0513	0.0901
$(6,10)$	1.5337	1.2628	0.0450	1.9477	1.7217	0.0472	0.0391
$(6,11)$	1.4266	1.1747	0.0444	1.5053	1.2558	0.0521	0.0779
$(7,8)$	1.5295	1.1243	0.0485	1.5115	1.1315	0.0470	0.0817
$(7,9)$	1.1982	0.8613	0.0470	1.3118	1.1017	0.0463	0.0873
$(7,10)$	1.2838	0.9395	0.0475	1.5416	1.4035	0.0431	0.0605
$(7,11)$	1.2911	0.9389	0.04939	1.3372	1.1896	0.0506	0.0636
$(8,9)$	1.4413	1.0623	0.0466	1.4834	1.2578	0.0480	0.0822
$(8,10)$	1.4596	1.0972	0.0475	1.6755	1.4236	0.0469	0.0864
$(8,11)$	1.5220	1.1318	0.0509	1.4161	1.2526	0.0518	0.0820
$(9,10)$	1.4148	1.1898	0.0472	1.7017	1.4496	0.0447	0.0666
$(9,11)$	1.4762	1.2209	0.0485	1.4290	1.2046	0.0509	0.0763
$(10,11)$	1.4808	1.2862	0.0422	1.2377	1.0984	0.0505	0.0681

Appendix B

Appendix to Chapter 5

Figure B.1: Trace plots of the parameters for the multivariate logistic model based on non-informative priors



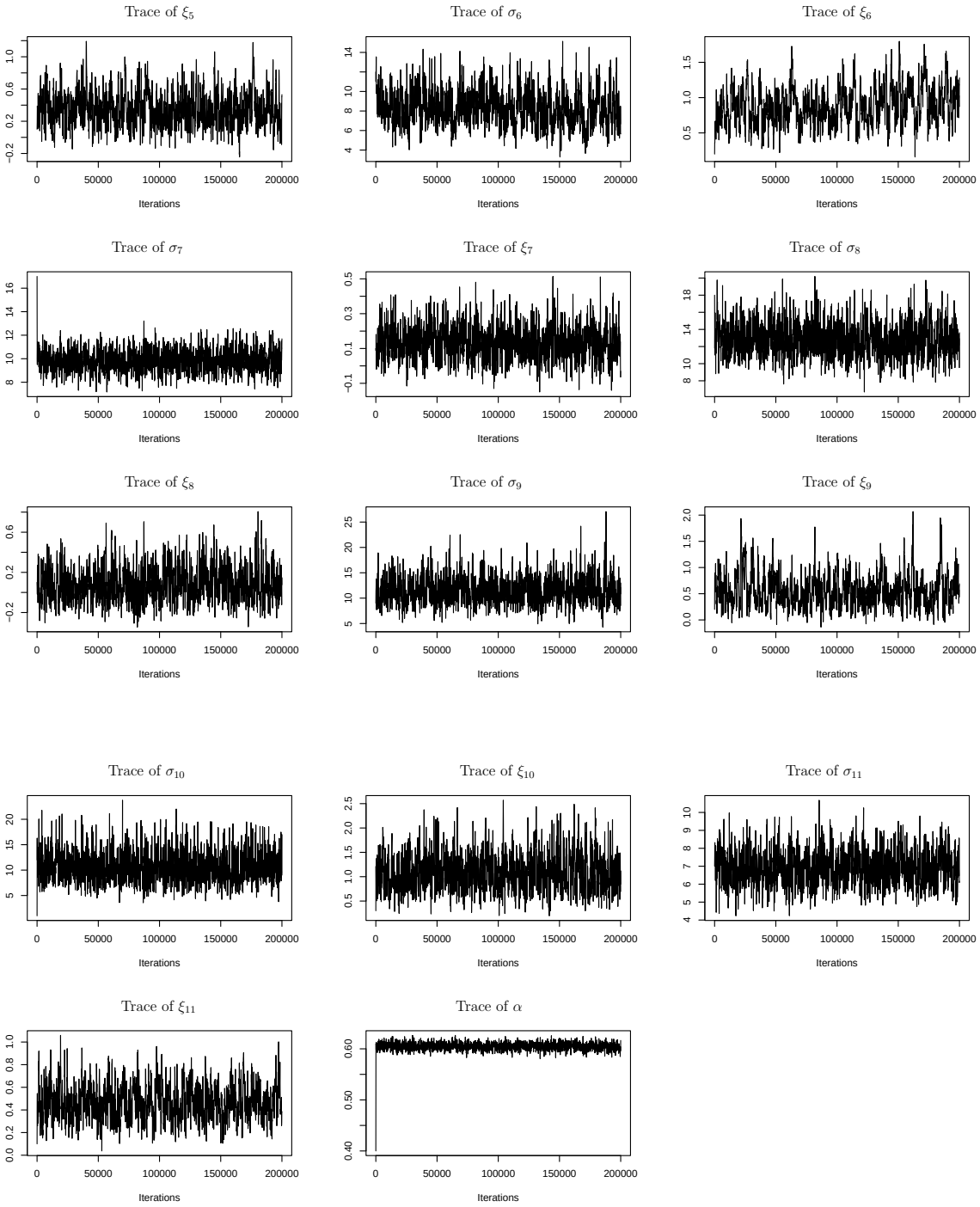
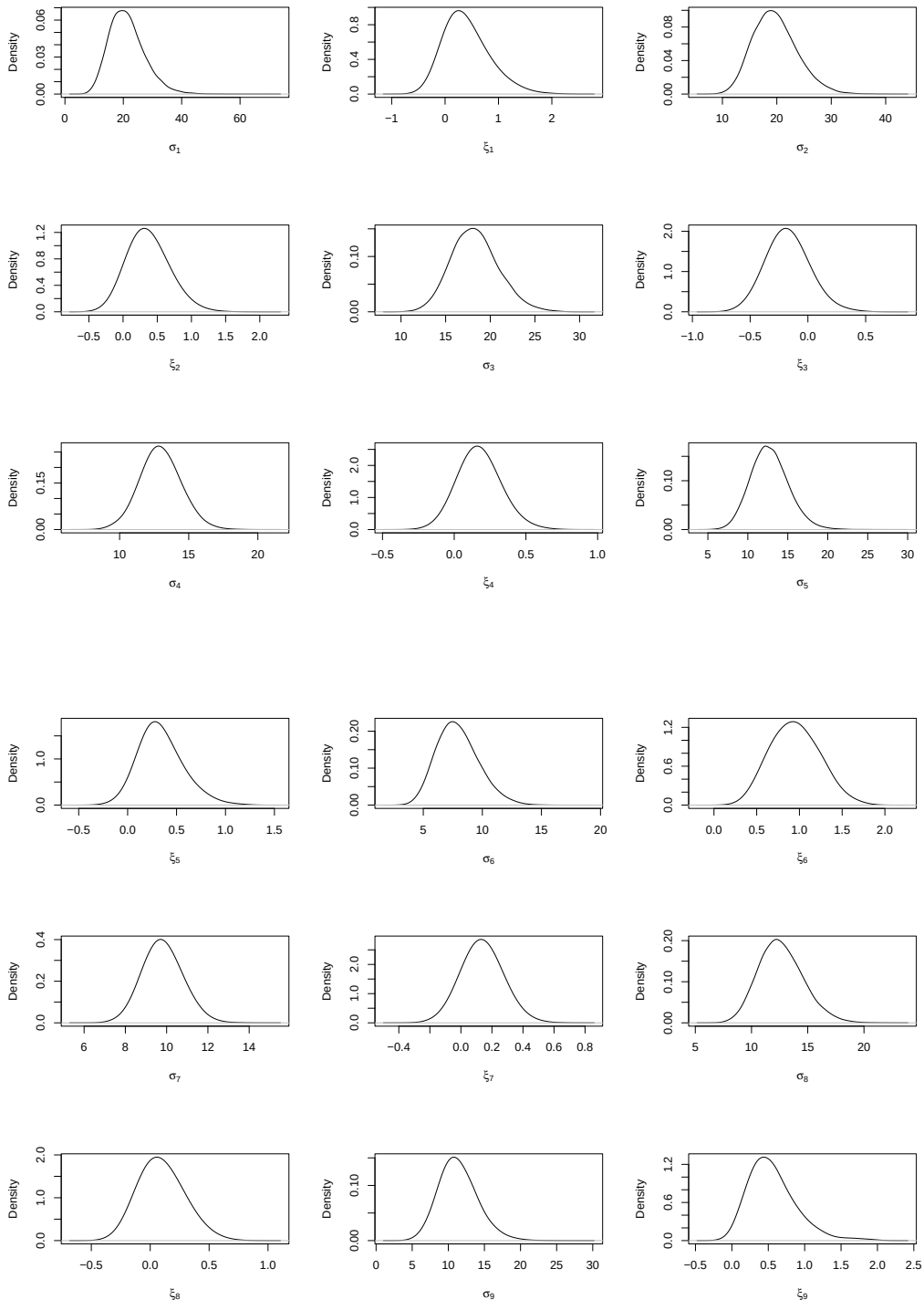

Trace of ξ_5
Trace of σ_6
Trace of ξ_6
Trace of σ_7
Trace of ξ_7
Trace of σ_8
Trace of ξ_8
Trace of σ_9
Trace of ξ_9
Trace of σ_{10}
Trace of ξ_{10}
Trace of σ_{11}
Trace of ξ_{11}
Trace of α
Iterations
0
50000
100000
150000
200000

Figure B.2: Posterior density plots of the parameters of the multivariate logistic model based on non-informative priors



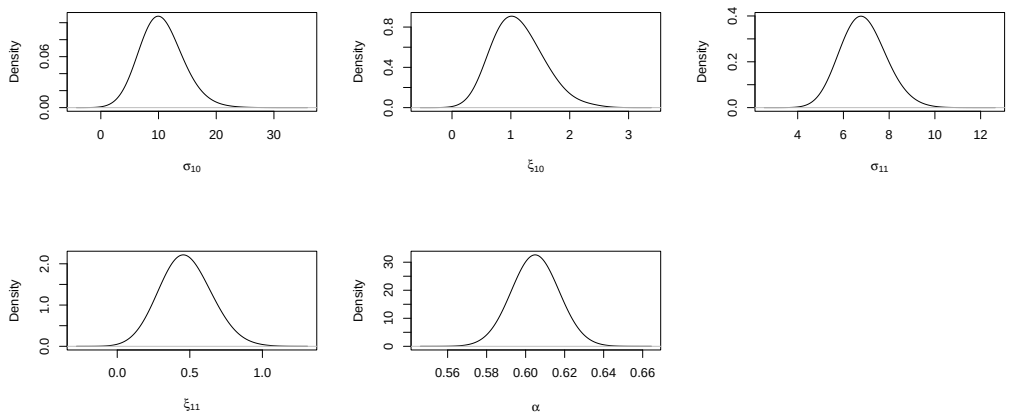
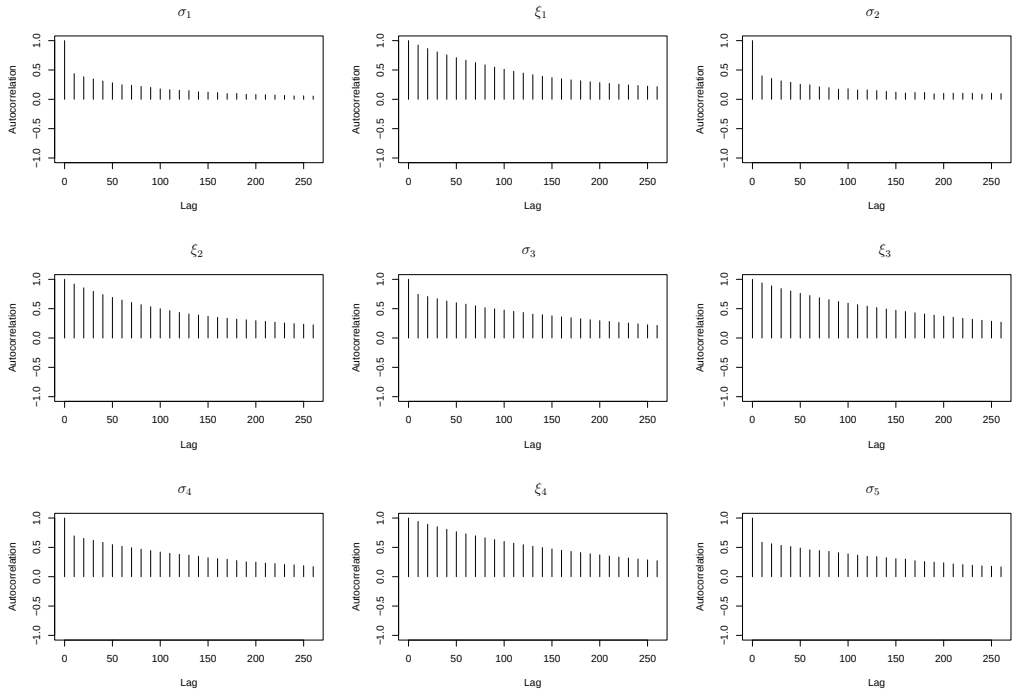


Figure B.3: Autocorrelation plots of the parameters of the multivariate logistic model based on non-informative priors



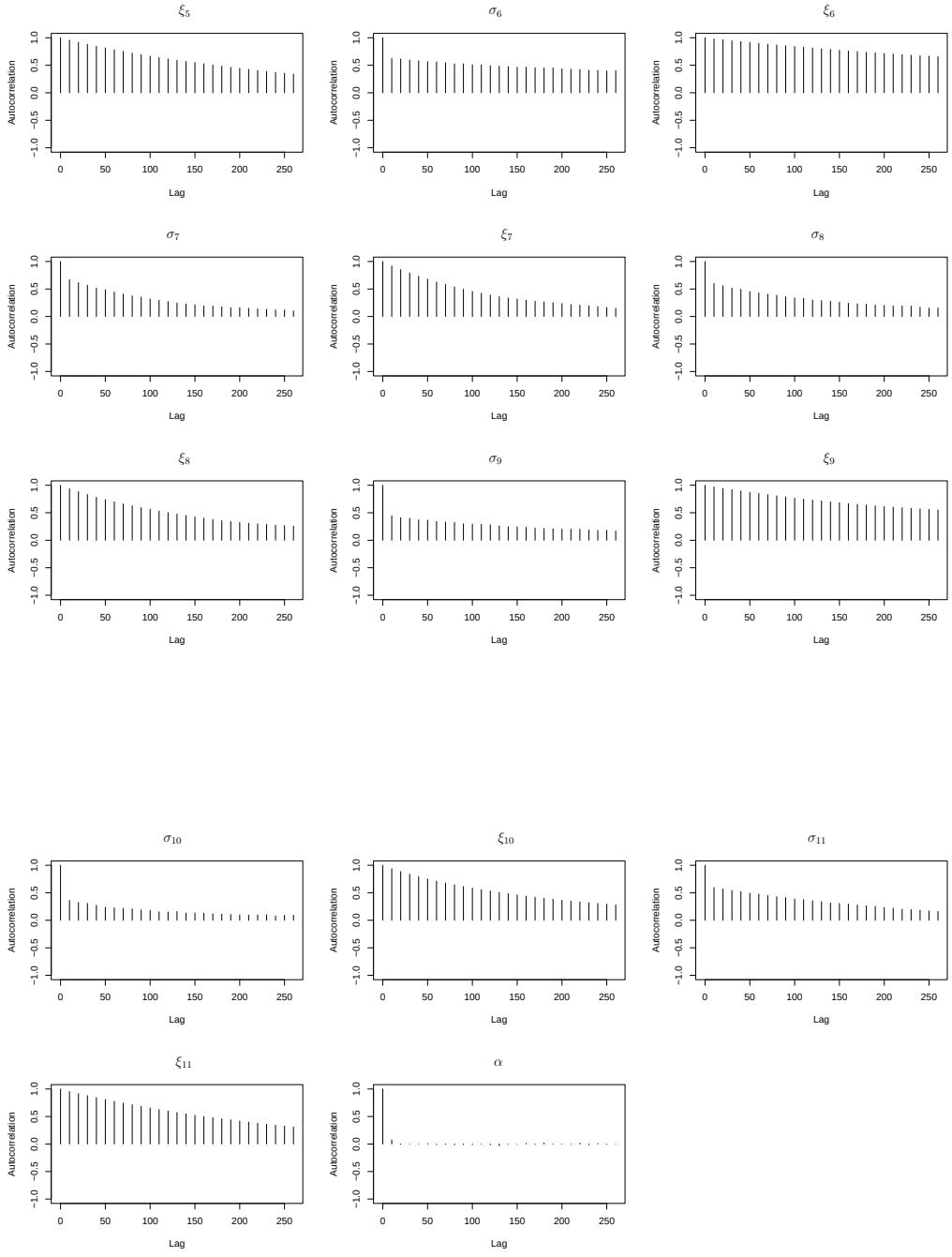
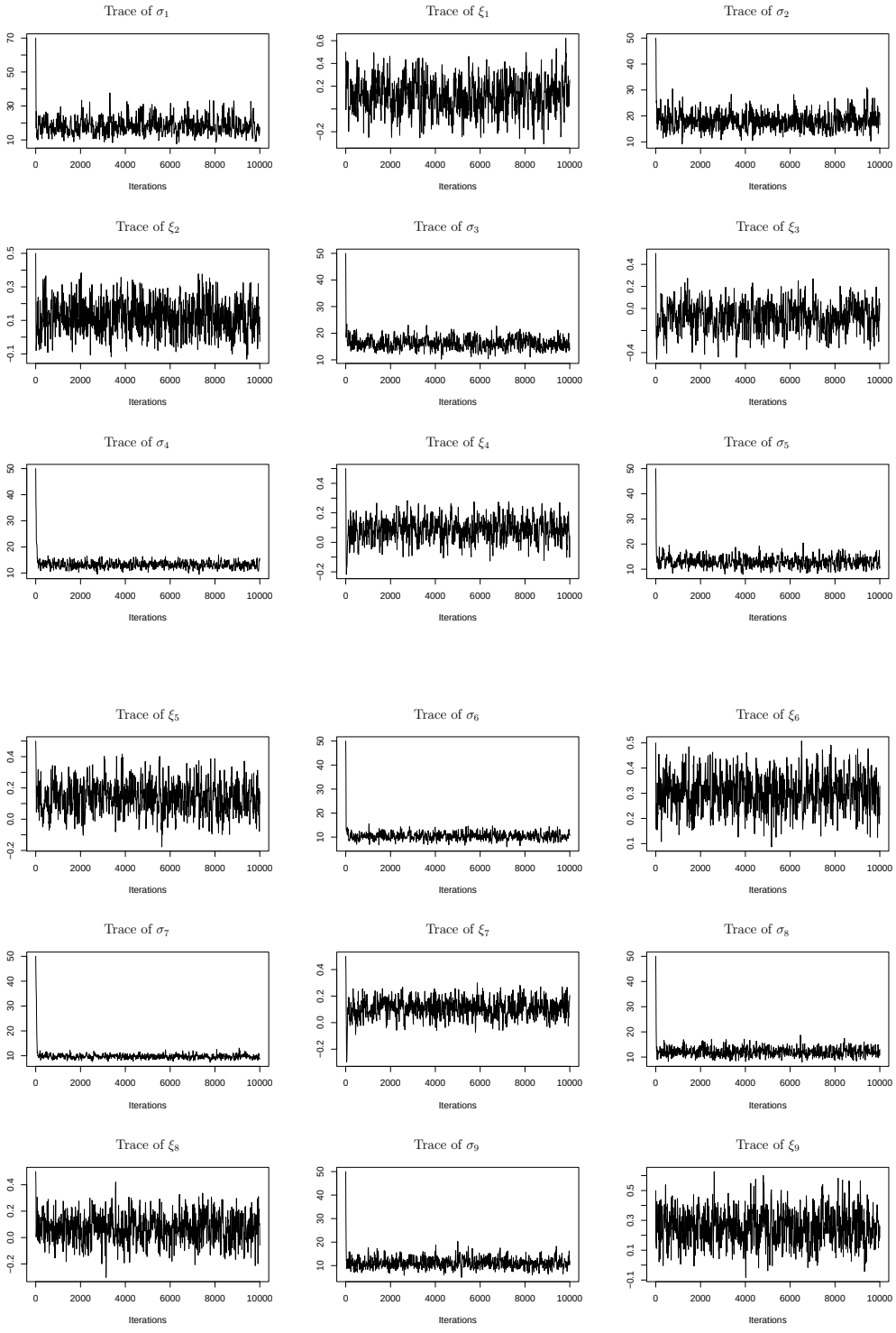

ξ_5
σ_6
ξ_6
σ_7
ξ_7
σ_8
ξ_8
σ_9
ξ_9
σ_{10}
ξ_{10}
σ_{11}
ξ_{11}
α
Autocorrelation
Lag

Figure B.4: Trace plots of the parameters for the multivariate logistic model based on informative priors



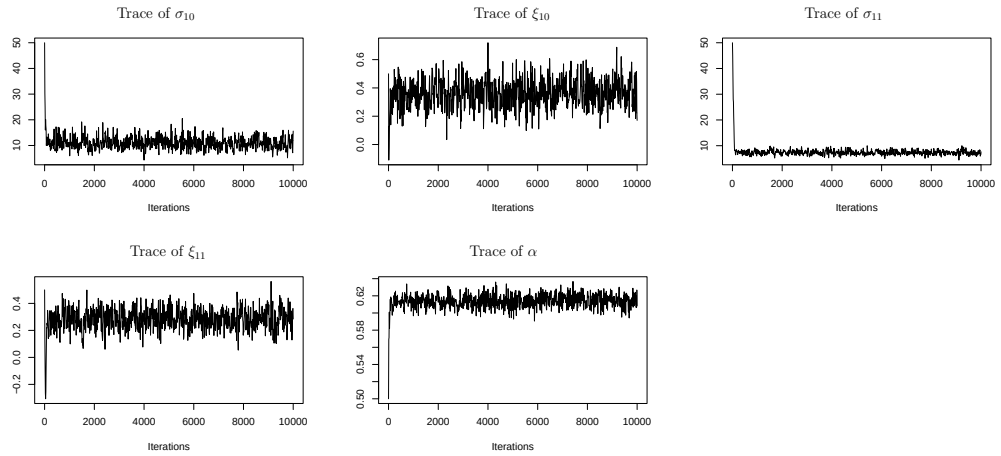
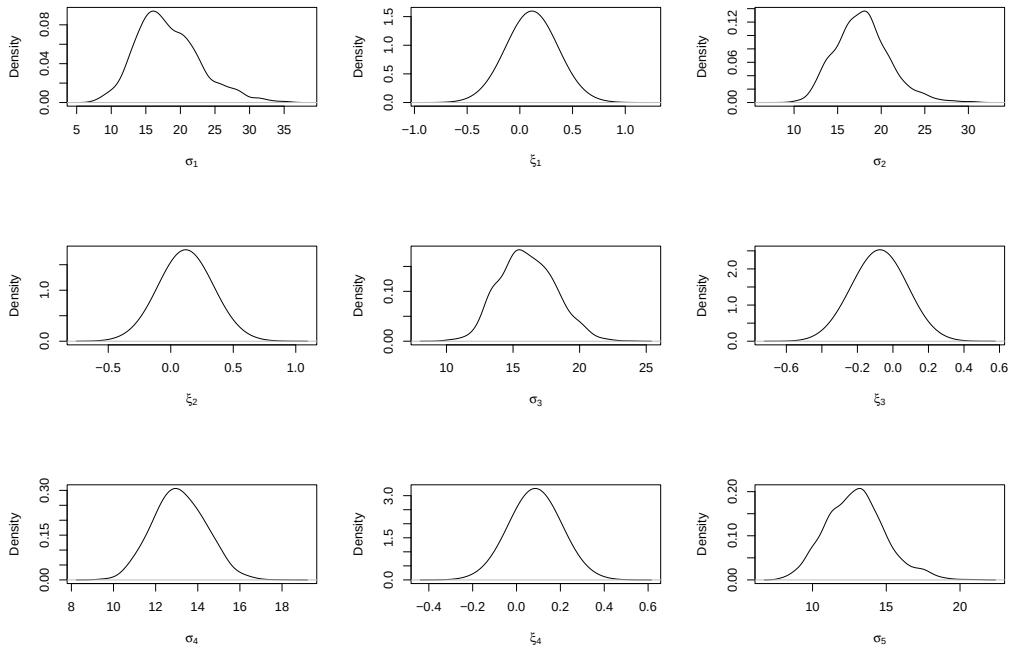


Figure B.5: Posterior density plots of the parameters of the multivariate logistic model based on informative priors



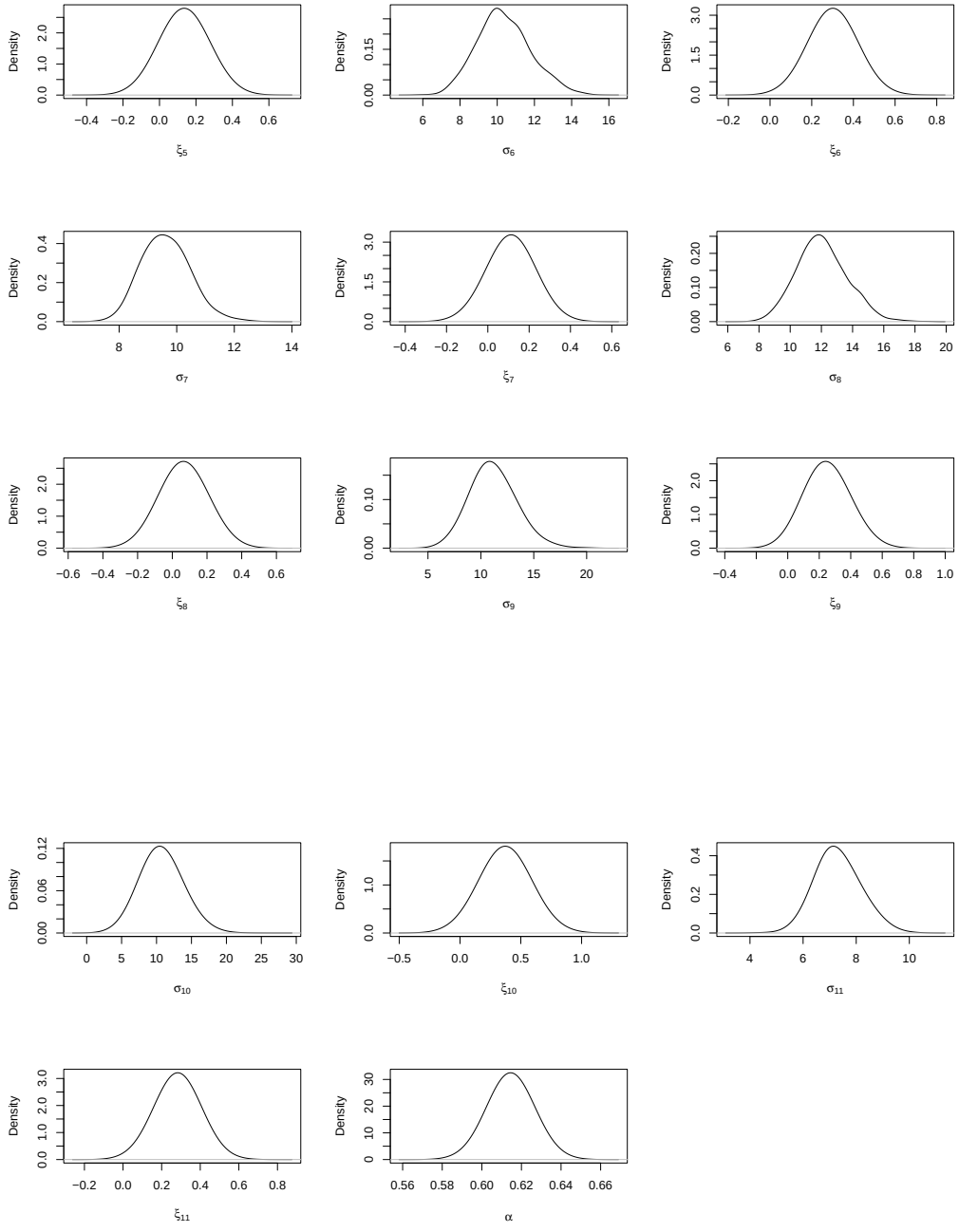
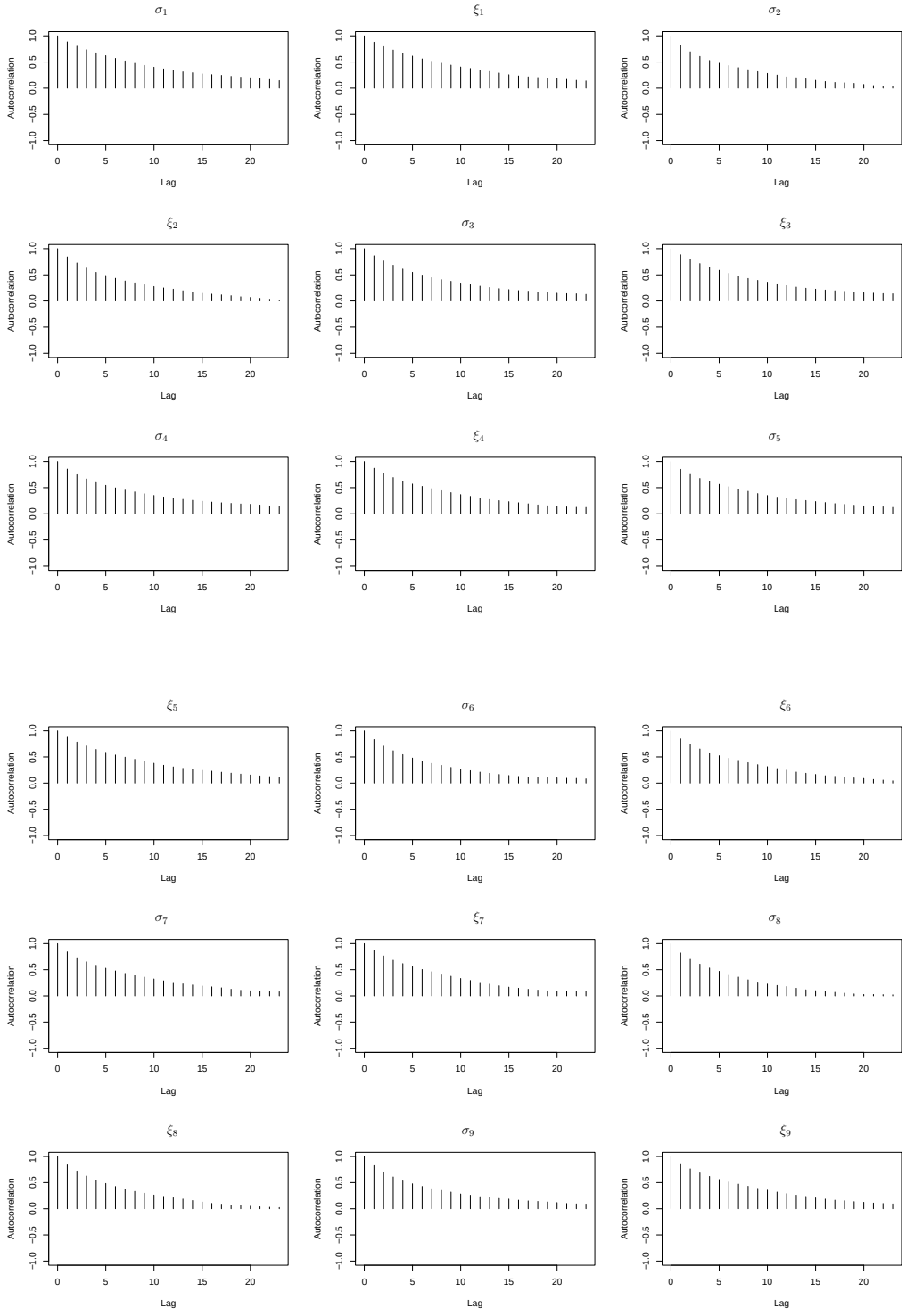
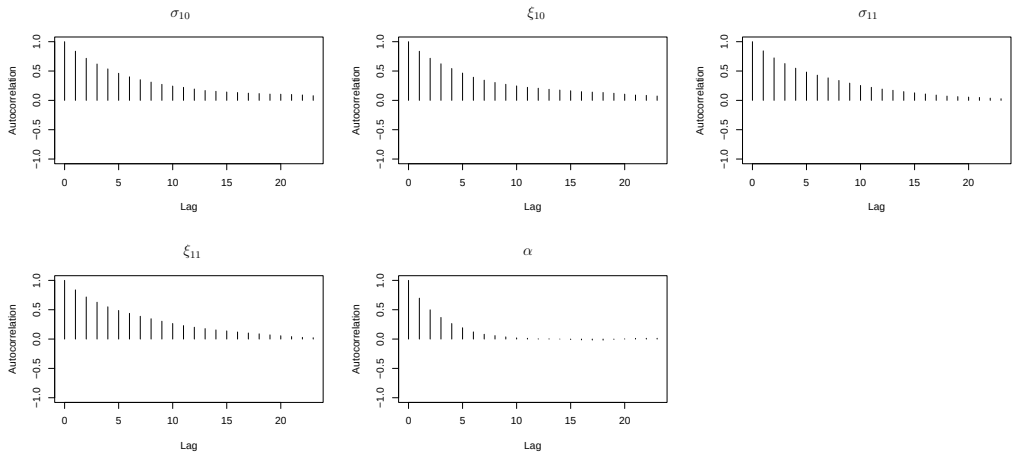

Density
ξ5
σ6
ξ6
σ7
ξ7
σ8
ξ8
σ9
ξ9
σ10
ξ10
σ11
ξ11
α

Figure B.6: Autocorrelation plots of the parameters of the multivariate logistic model based on informative priors




σ_{10}
ξ_{10}
σ_{11}
ξ_{11}
α
Autocorrelation
Lag
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20

Bibliography

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

Barnett, V. and Turkman, K.F. (1994). *Statistics for the environment: water–related issues*. Wiley, Chichester.

Besag, J. (2001). Markov chain Monte Carlo for statistical inference. Center for Statistics and the Social Sciences working paper, September 2001.

Besag, J., Green, P., Higdon, D. and Mengerson, K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science*, **10**, 3–41.

Coles, S.G. (1991). Statistical Methodology for the Multivariate Analysis of Environmental Extremes. *PhD thesis*, University of Sheffield, Sheffield.

Coles, S.G. (1993). Regional Modelling of Extreme Storms via Max-stable Processes. *J. R. Statist. Soc. B*, **55**, 797–816.

Coles, S.G. (1994). A Temporal Study of Extreme Rainfall. In *Statistics for the environment 2: Water Related Issues* (V. Barnett and K.F. Turkman), 61–78.

Coles, S.G. (1999). *Extreme value theory and applications*. `www.maths.lancs.ac.uk/coless/notes/`. (30 Sep. 2000).

Coles, S.G. (2001). *An introduction to statistical modeling of extreme values*. Springer, London.

Coles, S.G., Heffernan, J.E. and Tawn, J.A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes*, **2**, 339—365.

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

Coles, S.G. and Tawn, J.A. (1991). Modelling Extreme Multivariate Events. *J. R. Statist. Soc., B*, **53**, 377—392.

Coles, S.G. and Tawn, J.A. (1994). Statistical Methods for Multivariate Extremes: an application to Structural Design (with discussion). *Appl. Statist.*, **43**, 1—48.

Coles, S.G. and Tawn, J.A. (1996a). A Bayesian Analysis of Extreme Rainfall Data. *Appl. Statist.*, **45**, 463—478.

Coles, S.G. and Tawn, J.A. (1996b). Modelling Extremes of the Areal Rainfall Process. *J. R. Statist. Soc. B*, **58**, 329—347.

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.

Fawcett, L. and Walshaw, D. (2005). Estimation of Extremes Occurring in Clusters. Submitted.

Ferro, C.A.T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc., B*, **65**, 545—556.

Fisher, R.A. and Tippett, L.H.C. (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, **24**, 180—190.

Galambos, J., Lechner, J. and Simiu, E. (1993). *Extreme Value Theory and Applications.* Kluwer, Dordrecht.

Gamerman, D. (1997). *Markov Chain Monte Carlo. Stochastic Bayesian Inference*. Chapman & Hall, London.

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

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721—741.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, **44**, 423—453.

Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

Gumbel, E.J. (1960). Bivariate Exponential Distributions. *J. Am. Statist. Ass.*, **55**, 698—707.

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

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

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

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97—109.

Heffernan, J.E. and Tawn, J.A. (2004). A conditional approach for multivariate extreme values. *J. R. Statist. Soc., B, Part 2*, **66**, 1—34.

Hosking, J.R.M. (1990). *L*–moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc., B*, **52**, 105—124.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Met. Soc.*, **81**, 158—171.

Joe, H., Smith, R.L. and Weissman, I. (1992). Bivariate threshold models for extremes. *J. R. Statist. Soc., B*, **54**, 171—183.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications.* Imperial College Press, London.

Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitsth*, **65**, 291—306.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series.* Springer–Verlag, New York.

Leadbetter, M.R., Weissman, I., de Haan, L. and Rootzen, H. (1989). On clustering of high values in statistically stationary series. *Proc. 4th International conference on meteorology*, Raurora, New Zealand, 217 - 222.

Maindonald, J.H. (2000). *Using R for Data Analysis and Graphics – An Introduction.* Australian National University, Canberra.

Metropolis, N., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem. Phys.*, **21**, 1087—1091.

Pauli, F. and Coles, S.G. (2001). Penalized likelihood inference in extreme value analyses. *J. Appl. Stat.*, **28**, 547—560.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected papers, Vol. II*, pp. 271—294. Amer. Math. Soc., Providence R.I.

Pickands, J. (1975). Statistical Inference using Extreme Order Statistics. *Ann. Statist.*, **3**, 119—131.

Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst.*, XLIX (Book 2), 859—878.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edn. Wiley, New York.

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

Shlather, M. and Tawn, J.A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, **90**, 139—156.

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, E.L. and Walshaw, D. (2003). Modelling Bivariate Extremes in a Region. *Bayesian Statistics*, **7**, 681—690.

Smith, R.L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications* (ed. J. Tiago de Oliveira), pp. 621—638. Reidel, Dordrecht.

Smith, R.L. (1985). Maximum likelihood estimation in a class of non–regular cases. *Biometrika*, **72**, 67—92.

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

Smith, R.L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground–Level Ozone. *Statistical Science*, **4**, 367—393.

Smith, R.L. (1991). Regional Estimation from Spatially Dependent Data. Preprint.

Smith, R.L. and Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three–parameter Weibull distribution.

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

Stephenson, A. (2003). *A User's Guide to the evd Package (Version 2.0).* `www.maths.lancs.ac.uk/~stephena/`. (20 Jul. 2003).

Tawn, J.A. (1988a). Bivariate extreme value theory: Models and estimation. *Biometrika*, **75**, 397—415.

Tawn, J.A. (1988b). An extreme value theory model for dependent observations. *J. Hydrol.*, **101**, 227—250.

Tawn, J.A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika*, **77**, 245—253.

Tawn, J.A. (1992). Estimating Probabilities of Extreme Sea-levels. *Appl. Statist.*, **41**, 77—93.

Tiago de Oliveira, J. (1984). Bivariate models for extremes; statistical decision. In *Statistical Extremes and Applications* (Ed. J. Tiago de Oliveira), pp. 131—153. Reidel, Dordrecht.

Walshaw, D. (1991). Statistical Analysis of Extreme Wind Speeds. *PhD Thesis.* University of Sheffield, Sheffield.

Walshaw, D. (1994). Getting the Most From Your Extreme Wind Data: A Step by Step Guide. *J. Res. Natl. Inst. Stand. Technol.*, **99**, 399—411.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. In *J. Am. Statist. Ass.*, **73**, 812—815.