5 Bayesian inference for extremes

Throughout this short course, the method of maximum likelihood has provided a general and flexible technique for parameter estimation. Given a (generic) parameter vector $\boldsymbol{\psi}$ within a family $\boldsymbol{\Psi}$, the likelihood function is the probability (density) of the observed data as a function of $\boldsymbol{\psi}$. Values of $\boldsymbol{\psi}$ that have high likelihood correspond to models which give high probability to the observed data. The principle of maximum likelihood estimation is to adopt the model with greatest likelihood; of all the models under consideration, this is the one that assigns the highest probability to the observed data. Other inferential procedures, such as "method of moments", provide viable alternatives to maximum likelihood estimation; moments–based techniques choose $\boldsymbol{\psi}$ optimally by equating model–based and empirical moments, and solving for $\boldsymbol{\psi}$ to obtain parameter estimates. These, and other procedures (such as probability weighted moments, L–moments and ranked set estimation), are discussed in detail in, amongst other places, Kotz and Nadarajah (2000).

5.1 General theory

Bayesian techniques offer an alternative way to draw inferences from the likelihood function, which many practitioners often prefer. As in the non–Bayesian setting, we assume data $\boldsymbol{x} = (x_1, \ldots, x_n)$ to be realisations of a random variable whose density falls within a parametric family $\mathcal{F} = \{f(\boldsymbol{x}; \boldsymbol{\psi}) : \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$. However, parameters of a distribution are now treated as random variables, for which we specify *prior distributions* – distributions of the parameters *prior* to the inclusion of data. The specification of these prior distributions enables us to supplement the information provided by the data – which, in extreme value analyses, is often very limited – with other sources of information. At the same time, it can be contended that, since different analysts might specify different priors, conclusions become subjective.

Leaving aside the arguments for and against the Bayesian methodology, suppose we model our observed data $\boldsymbol{x}$ using the probability density function $f(\boldsymbol{x}; \boldsymbol{\psi})$. The likelihood function for $\boldsymbol{\psi}$ is therefore $L(\boldsymbol{\psi}|\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\psi})$. Also, suppose our prior beliefs about likely values of $\boldsymbol{\psi}$ are expressed by the probability density function $\pi(\boldsymbol{\psi})$. We can combine both pieces of information using Bayes Theorem, which states that

$$\pi(\boldsymbol{\psi}|\boldsymbol{x}) \quad = \quad \frac{\pi(\boldsymbol{\psi}) L(\boldsymbol{\psi}|\boldsymbol{x})}{f(\boldsymbol{x})}, \tag{16}$$

where

$$f(\boldsymbol{x}) \quad = \quad \begin{cases} \displaystyle\int_{\boldsymbol{\Psi}} \pi(\boldsymbol{\psi}) L(\boldsymbol{\psi}|\boldsymbol{x}) d\boldsymbol{\psi} & \text{if } \boldsymbol{\psi} \text{ is continuous,} \\ \displaystyle\sum_{\boldsymbol{\Psi}} \pi(\boldsymbol{\psi}) L(\boldsymbol{\psi}|\boldsymbol{x}) & \text{if } \boldsymbol{\psi} \text{ is discrete.} \end{cases}$$

Since $f(\boldsymbol{x})$ is not a function of $\boldsymbol{\psi}$, Bayes Theorem can be written as

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

In equation (16), $\pi(\boldsymbol{\psi}|\boldsymbol{x})$ is the *posterior* distribution of the parameter vector $\boldsymbol{\psi}$, $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, i.e. the distribution of $\boldsymbol{\psi}$ *after* the inclusion of the data. This prior distribution is often of great

interest, since the prior–posterior changes represent the changes in our beliefs after the data has been included in the analysis. However, computation of the denominator in (16) can be problematic, and usually analytically intractable. There is nothing particularly special about the fact that equation (16) represents a Bayesian posterior; given any complex non–standard probability distribution, we need ways to understand it, to calculate its moments, to compute its conditional and marginal distributions and their moments, all of which could require troublesome integration as in the denominator of equation (16). We need a way of understanding posterior densities which does not rely on being able to analytically integrate the kernel of the posterior; stochastic simulation is one possible solution.

5.2 Markov chain Monte Carlo

The recent explosion in Markov chain Monte Carlo (MCMC) techniques owes largely to their application in Bayesian inference. The idea here is to produce simulated values from the posterior distribution – not exactly, as this is usually unachievable, but through an appropriate MCMC technique.

5.2.1 The Gibbs sampler

The Gibbs sampler is a way of simulating from multivariate distributions based only on the ability to simulate from conditional distributions. Suppose the density of interest (usually the posterior density) is $\pi(\boldsymbol{\psi})$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d)'$, and that the full conditionals

$$\pi(\psi_i|\psi_1, \ldots, \psi_{i-1}, \psi_{i+1}, \ldots, \psi_d) = \pi(\psi_i|\psi_{-i}) = \pi_i(\psi_i), \qquad i = 1, \ldots, d$$

are available for simulating from (ψ_{-i} denotes the parameter vector $\boldsymbol{\psi}$ excluding ψ_i). The Gibbs sampler uses the following algorithm:

1. Initialise the iteration counter to $k = 1$. Initialise the state of the chain to $\boldsymbol{\psi}^{(0)} = (\psi_1^{(0)}, \ldots, \psi_d^{(0)})'$;

2. Obtain a new value $\boldsymbol{\psi}^{(k)}$ from $\boldsymbol{\psi}^{(k-1)}$ by successive generation of values

$$\begin{aligned}
\psi_1^{(k)} &\sim \pi(\psi_1|\psi_2^{(k-1)}, \ldots, \psi_d^{(k-1)}) \\
\psi_2^{(k)} &\sim \pi(\psi_2|\psi_1^{(k)}, \psi_3^{(k-1)}, \ldots, \psi_d^{(k-1)}) \\
&\vdots \qquad\quad \vdots \\
\psi_d^{(k)} &\sim \pi(\psi_d|\psi_1^{(k)}, \ldots, \psi_{d-1}^{(k)});
\end{aligned}$$

3. Change counter k to $k + 1$, and return to step 2.

Each simulated value depends only on the previous simulated value, and not any other previous values or the iteration counter k. The Gibbs sampler can be used in isolation if we can readily simulate from the full conditional distributions; however, this is not always the case. Fortunately, the Gibbs sampler can be combined with Metropolis–Hastings schemes when the full conditionals are difficult to simulate from.

5.2.2 Metropolis–Hastings sampling

Suppose again that $\pi(\boldsymbol{\psi})$ is the density of interest. Further, suppose that we have some arbitrary transition kernel $p(\boldsymbol{\psi}_{i+1}, \boldsymbol{\psi}_i)$ (which is easy to simulate from) for iterative simulation of successive values. Then consider the following algorithm:

1. Initialise the iteration counter to $k = 1$, and initialise the chain to $\boldsymbol{\psi}^{(0)}$;
2. Generate a proposed value $\boldsymbol{\psi}'$ using the kernel $p(\boldsymbol{\psi}^{(k-1)}, \boldsymbol{\psi}')$;
3. Evaluate the *acceptance probability* $A(\boldsymbol{\psi}^{(k)}, \boldsymbol{\psi}')$ of the proposed move, where
$$A(\boldsymbol{\psi}, \boldsymbol{\psi}') \quad = \quad \min\left\{1, \frac{\pi(\boldsymbol{\psi}')L(\boldsymbol{\psi}'|\boldsymbol{x})p(\boldsymbol{\psi}', \boldsymbol{\psi})}{\pi(\boldsymbol{\psi})L(\boldsymbol{\psi}|\boldsymbol{x})p(\boldsymbol{\psi}, \boldsymbol{\psi}')}\right\};$$
4. Put $\boldsymbol{\psi}^{(k)} = \boldsymbol{\psi}'$ with probability $A(\boldsymbol{\psi}^{(k-1)}, \boldsymbol{\psi}')$, and put $\boldsymbol{\psi}^{(k)} = \boldsymbol{\psi}^{(k-1)}$ otherwise;
5. Change the counter from k to $k + 1$ and return to step 2.

So at each stage, a new value is generated from the proposal distribution. This is either accepted, in which case the chain moves, or rejected, in which case the chain stays where it is. Whether or not the move is accepted or rejected depends on the acceptance probability which itself depends on the relationship between the density of interest and the proposal distribution. Common choices for the proposal distribution include symmetric chains, where $p(\boldsymbol{\psi}, \boldsymbol{\psi}') = p(\boldsymbol{\psi}', \boldsymbol{\psi})$, and random walk chains, where the proposal $\boldsymbol{\psi}'$ at iteration k is $\boldsymbol{\psi}' = \boldsymbol{\psi} + \varepsilon_k$, where the ε_k are IID random variables.

5.2.3 Hybrid methods

Here, we combine Gibbs sampling and Metropolis–Hastings schemes to form hybrid Markov chains whose stationary distribution is the distribution of interest. For example, given a multivariate distribution whose full conditionals are awkward to simulate from directly, we can define a Metropolis–Hastings scheme for each full conditional, and apply them to each component in turn for each iteration. This is similar to Gibbs sampling, but each component update is a Metropolis–Hastings update, instead of a direct simulation from the full conditional. Another scheme, known as "Metropolis within Gibbs", goes through each full conditional in turn, simulating directly from the full conditionals wherever possible, and carrying out a Metropolis–Hastings update elsewhere.

5.3 Bayesian inference for extremes

There are various (and some may say compelling) reasons for preferring a Bayesian analysis of extremes over the more traditional likelihood approach. As already discussed, since extreme data are (by their very nature) quite scarce, the ability to incorporate other sources of information through a prior distribution has obvious appeal. Bayes' Theorem also leads to an inference that comprises a complete distribution, meaning that the variance of the posterior distribution, for example, can be used to summarise the precision of the inference, without having to rely upon asymptotic theory. Also, implicit in the Bayesian framework is the concept of the *predictive distribution*. This distribution describes how likely are different outcomes of a future

experiment. The predictive probability density function is given by

$$f(y|\boldsymbol{x}) \quad = \quad \int_{\boldsymbol{\Psi}} f(y|\boldsymbol{\psi})\pi(\boldsymbol{\psi}|\boldsymbol{x})d\boldsymbol{\psi} \tag{17}$$

when $\boldsymbol{\psi}$ is continuous. From equation (17), we can see that the predictive distribution is formed by weighting the possible values for $\boldsymbol{\psi}$ in the future experiment $f(y|\boldsymbol{\psi})$ by how likely we believe they are to occur after seeing the data. For example, a suitable model for the threshold excess Y of a process is $Y \sim \text{GPD}(\sigma, \xi)$. Estimation of $\boldsymbol{\psi} = (\sigma, \xi)$ could be made on the basis of previous observations $\boldsymbol{x} = (x_1, \ldots, x_n)$. Thus, in the Bayesian framework, we would have

$$\Pr\{Y \leq y|x_1, \ldots, x_n\} \quad = \quad \int_{\boldsymbol{\Psi}} \Pr\{Y \leq y|\boldsymbol{\psi}\}\, \pi(\boldsymbol{\psi}|\boldsymbol{x})d\boldsymbol{\psi}. \tag{18}$$

Equation (18) gives the distribution of a future threshold excess, allowing for both parameter uncertainty and randomness in future observations. Solving

$$\Pr\{Y \leq q_{r,\text{pred}}|x_1, \ldots, x_n\} \quad = \quad 1 - \frac{1}{r}$$

for $q_{r,\text{pred}}$ therefore gives an estimate of the r–year return level that incorporates uncertainty due to model estimation. Though (17) may seem analytically intractable, it can be approximated if the posterior distribution has been estimated using, for example, MCMC. After removal of the "burn–in" period, the MCMC procedure gives a sample $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_B$ that can be regarded as realisations from the stationary distribution $\pi(\boldsymbol{\psi}|\boldsymbol{x})$. Thus

$$\Pr\{Y \leq q_{r,\text{pred}}|x_1, \ldots, x_n\} \quad \approx \quad \frac{1}{B}\sum_{i=1}^{B} \Pr\{Y \leq q_{r,\text{pred}}|\boldsymbol{\psi}_i\},$$

which we can solve for $q_{r,\text{pred}}$ using a numerical solver. Another reason lending appeal to Bayesian inference for extremes is that it is not dependent on the regularity assumptions required by the theory of maximum likelihood. For example, when $\xi < -0.5$, maximum likelihood estimation breaks down – in this situation, a Bayesian approach provides a feasible alternative.

5.3.1 Example: Annual maximum sea levels: Port Pirie, South Australia

Figure 18 shows a time series plot of annual maximum sea levels at another Australian location – Port Pirie, in South Australia. Notice that, unlike the corresponding data from Fremantle in Wester Australia, there doesn't appear to be any trend in this series; in fact, the series appear stationary.

We use the GEV as a model for the annual maximum sea levels at Port Pirie Z_i in year i, i.e.

$$Z_i \quad \sim \quad \text{GEV}(\mu, \sigma, \xi), \qquad i = 1, \ldots, 65.$$

When employing MCMC methods it is common to re–parameterise the GEV scale parameter and work with $\eta = \log(\sigma)$ to retain the positivity of this parameter. In the absence of any expert

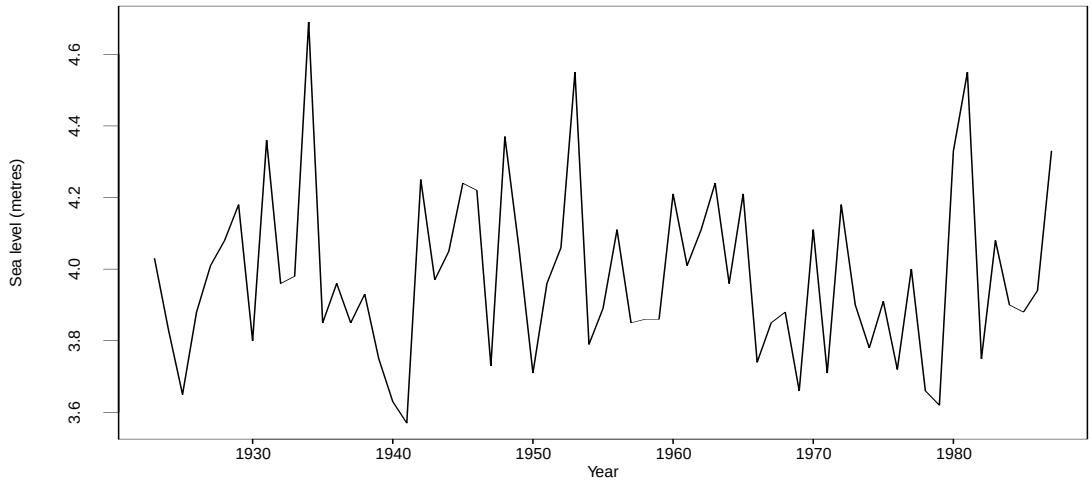

Figure 18: Time series plot of annual maximum sea levels observed at Port Pirie.

prior information regarding the three parameters of the GEV distribution, we adopt a 'naive' approach and use largely non–informative, independent priors for these, namely

$$\begin{aligned} \pi(\mu) &\sim N(0, 10000), \\ \pi(\eta) &\sim N(0, 10000) \qquad \text{and} \\ \pi(\xi) &\sim N(0, 100), \end{aligned}$$

the large variances of these distributions imposing near–flat priors.

We use a Metropolis–Hastings MCMC sampling scheme; after setting initial starting values for $\boldsymbol{\psi} = (\mu, \eta, \xi)$, we use an arbitrary probability rule $p(\boldsymbol{\psi}_{i+1}|\boldsymbol{\psi})$ for iterative simulation of successive values in the chain. Once this rule has been used to generate a candidate value $\boldsymbol{\psi}'$ for $\boldsymbol{\psi}_{i+1}$, we accept this with probability A (see 5.2.2); otherwise, $\boldsymbol{\psi}_{i+1} = \boldsymbol{\psi}_i$. Here, we use a *random walk* procedure to generate candidate values, i.e.

$$\begin{aligned} \mu' &= \mu_i + \epsilon_\mu \\ \eta' &= \eta_i + \epsilon_\eta \qquad \text{and} \\ \xi' &= \xi_i + \epsilon_\xi, \end{aligned}$$

with the ϵ being normally distributed with zero mean and variances v_μ, v_η and v_ξ respectively. In fact, the choice of algorithm and its 'tuning parameters' (v_μ, v_η and v_ξ) does not affect the model. It does, however, affect the efficiency of the algorithm. Some believe there is a 'fine art' to tuning the algorithm used, but it is common to aim for an overall acceptance rate of around 30%.

Initialising with $\boldsymbol{\psi}^{(0)} = (5, 0.5, 0.1)$, we get the following values generated by 5000 iterations of the MCMC scheme (see Figure 19). The settling–in period seems to take around 300 iterations, after which the chain seems to have converged. This settling–in period is often known as the *burn–in*. Thus, after deleting the first 300 simulations, the remaining 4700 simulated values can be treated as dependent realisations whose marginal distribution is the target posterior. Over-leaf, in Figure 20, is a panel of plots corresponding to the sampling distributions

of the three GEV parameters (after the removal of burn–in), as well as the 100–year return level. The sampling distribution for the posterior of the return level has been obtained by inversion of the distribution function for the GEV (Equation 1) and then by repeated substitution of $\mu^{(301)}, \sigma^{(301)}, \xi^{(301)}, \ldots, \mu^{(5000)}, \sigma^{(5000)}, \xi^{(5000)}$.

The posterior means, standard deviations and 95% credible intervals are shown in Table 4, along with the corresponding maximum likelihood estimates for comparison.

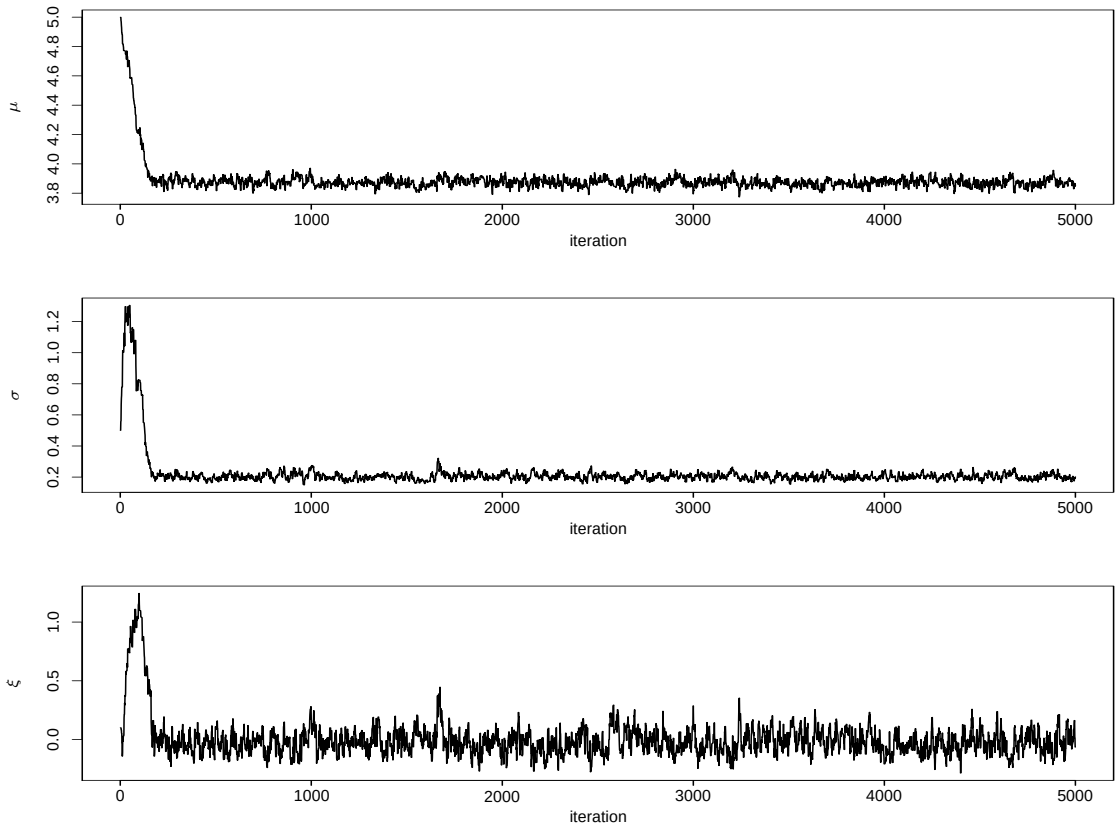

Figure 19: MCMC realisations of the GEV parameters in a Bayesian analysis of the Port Pirie sea level maxima.

		μ	σ	ξ	q_{100}
Posterior	mean (st. dev.)	3.874 (0.028)	0.203 (0.021)	–0.024 (0.098)	4.788 (0.255)
distribution	95% CI	(3.819, 3.932)	(0.166, 0.249)	(–0.196, 0.182)	(4.516, 5.375)
Maximum	m.l.e. (s.e.)	3.872 (0.028)	0.198 (0.020)	-0.040 (0.098)	4.692 (0.158)
likelihood	95% CI	(3.821, 3.930)	(0.158, 0.238)	(–0.242, 0.142)	(4.501, 5.270)

Table 4: Summary statistics for the posterior location, scale and shape, and the 100–year return level. Shown also, for comparison, are the corresponding m.l.e.s, the confidence interval for the return level being found via profile likelihood.

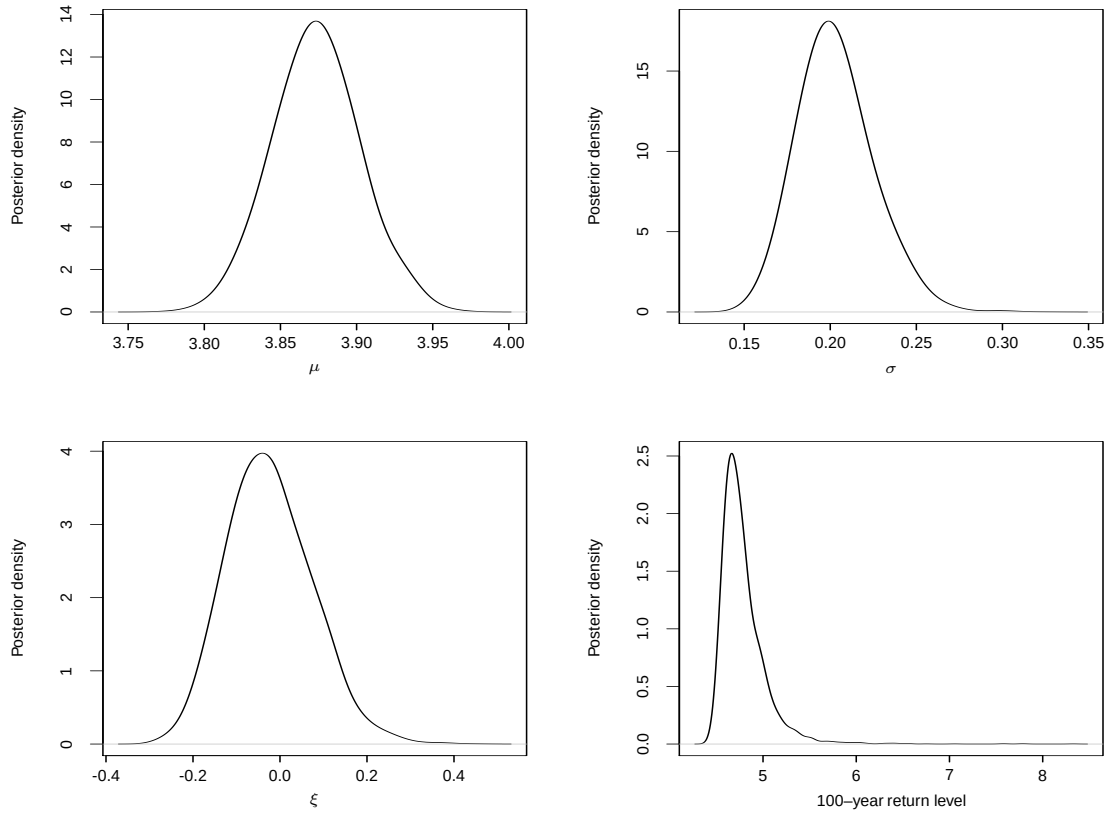

Figure 20: Sampling distributions for the posterior densities of μ, σ, ξ and the 100–year return level.

5.3.2 More complex structures: A random effects model for extreme wind speeds

In this section we briefly discuss the work which Lee Fawcett will present at next week's TIES conference. In this work, we develop a hierarchical model for hourly maximum wind speeds over a region of central and northern England. The data used consist of hourly gust maximum wind speeds recorded for the British Meteorological Office at twelve locations (see Figure 21). We construct a model which is based on a standard limiting extreme value distribution, but incorporates random effects for the sites, for seasonal variation, and for the serial dependence inherent in the time series of hourly maximum speeds obtained at each site. The Bayesian paradigm provides the most feasible modelling approach to capture the rich meteorological structure present in these data. Figure 22 illustrates an exploratory analysis of data from two contrasting sites, Nottingham and Bradfield. Shown are time series plots of the hourly maxima, histograms, and a plot of the time series against the version at lag 1. The first three years of data only are used in each case, to best illustrate the relevant data characteristics. We now (very briefly) outline the model structure used.

Modelling threshold exceedances

We will start with the Generalised Pareto Distribution as a model for threshold excesses; by doing so, we can incorporate more extreme data in our analysis than if we were to select "block maxima", and so increase the precision of our analysis. Thus, wind speed excesses over a high threshold will be modelled with a GPD(σ, ξ).

Figure 21: Location of wind speed stations.

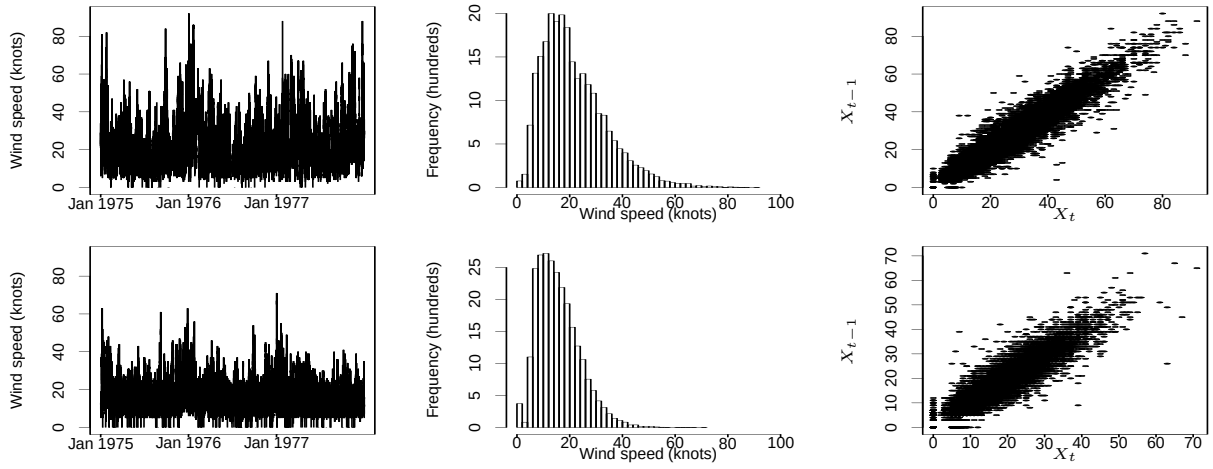

Figure 22: Time series plots and histograms of hourly gusts observed at Bradfield (top row) and Nottingham (bottom row) over a three year period (1975–1977 inclusive). Also shown are plots of the time series against the lagged series.

Site and seasonal variability

For our purposes, we need the GPD parameters to vary across sites, and seasonally. We take a pragmatic approach to seasonality, partitioning the annual cycle into twelve 'months'. Thus our hierarchical model will need to yield parameter pairs $(\sigma_{m,j}, \xi_{m,j})$ for $m = 1, \ldots, 12$ and $j = 1, \ldots, 12$, where m and j are indices of season and site respectively. It is also necessary to allow the threshold u used for excesses modelled by the GPD to vary, since different criteria about what constitutes an extreme value will be in play for each combination of season and site. We will denote by $u_{m,j}$ the value of the exceedance threshold for month m and site j.

Temporal dependence

To account for the presence of temporal dependence within each season and site, we now adopt approach 3 outlined in Section 2.1.3; specifically, we use bivariate extreme value theory discussed in Part 4 of this short course to formulate a simple first–order Markov chain structure for successive extreme wind speeds. As with fitting to all threshold exceedances and then ad-

justing the inferences accordingly (as we recommended with the Newlyn sea–surge data in Section 2.1.4), this approach avoids the need to arbitrarily identify clusters of extremes and filter out a set of independent extreme values (thus discarding many precious extremes!), but also quantifies the extent of extremal dependence at each site. Put simply, at each site, the logistic model with parameter α_j (discussed in Part 4) is used to model each successive pair of threshold exceedances (say (x_i, x_{i+1})) at site j. The parameter $\alpha_j \in (0, 1]$ measures the strength of dependence between consecutive extremes, smaller values indicating stronger dependence. Independence and complete dependence are obtained when $\alpha_j = 1$ and $\alpha_j \searrow 0$ respectively. Following work in Fawcett (2005), which suggests that the serial dependence in extremes is fairly constant across all seasons, we assume that the Markov chain model describes the dependence over all seasons at site j.

Threshold stability property
In order to ensure a threshold stability property in our models, we use $\tilde{\sigma}_{m,j} = \sigma_{m,j} - \xi_{m,j} u_{m,j}$ in place of the usual scale parameter $\sigma_{m,j}$. With this parameterisation, if $(X - u^*_{m,j})$ is distributed GPD($\tilde{\sigma}_{m,j}, \xi_{m,j}$), then for all values $u_{m,j} > u^*_{m,j}$, we have that $(X - u_{m,j})$ is also GPD($\tilde{\sigma}_{m,j}, \xi_{m,j}$) distributed (e.g. see Coles (2001)). This is useful here, because it allows comparisons of the GPD scale and shape parameters across seasons and sites. It also allows us to specify prior information for both parameters without having to worry about the additional complications that would arise for parameters which were threshold dependent.

The model
We then specify the following random effects model for our extreme wind speeds:

$$\begin{aligned} \log(\tilde{\sigma}_{m,j}) &= \gamma^{(m)}_{\tilde{\sigma}} + \epsilon^{(j)}_{\tilde{\sigma}}, \\ \xi_{m,j} &= \gamma^{(m)}_{\xi} + \epsilon^{(j)}_{\xi} \qquad \text{and} \\ \alpha_j &= \epsilon^{(j)}_{\alpha}, \end{aligned}$$

where, generically, γ and ϵ represent seasonal and site effects respectively. We work with $\log(\tilde{\sigma}_{m,j})$ for computational convenience, and to retain the positivity of the scale parameter $\tilde{\sigma}_{m,j}$. All random effects for $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$ are taken to be normally and independently distributed:

$$\gamma^{(m)}_{\tilde{\sigma}} \sim N_0(0, \tau_{\tilde{\sigma}}) \qquad \text{and} \tag{19}$$

$$\gamma^{(m)}_{\xi} \sim N_0(0, \tau_{\xi}), \qquad m = 1, \ldots, 12, \tag{20}$$

for the seasonal effects, and

$$\begin{aligned} \epsilon^{(j)}_{\tilde{\sigma}} &\sim N_0(a_{\tilde{\sigma}}, \zeta_{\tilde{\sigma}}) \qquad \text{and} \\ \epsilon^{(j)}_{\xi} &\sim N_0(a_{\xi}, \zeta_{\xi}), \qquad j = 1, \ldots, 12, \end{aligned}$$

for the site effects, where $N_0(\eta, \rho)$ is the normal distribution with mean η and *precision* ρ (used for notational convenience). We choose the mean of the normal distribution of the seasonal effects to be fixed at zero in (19) and (20) in order to avoid over–parameterisation and problems of model identifiability, although we could equally well have fixed the mean for the distribution of the *site* effects to achieve this. In the absence of any prior knowledge about α_j, we set the prior by specifying

$$\epsilon^{(j)}_{\alpha} \sim U(0, 1).$$

The final layer of the model is the specification of prior distributions for the random effect distribution parameters. Here we adopt conjugacy wherever possible to simplify computations, specifying:

$$\begin{aligned} a_{\tilde{\sigma}} &\sim N_0(b_{\tilde{\sigma}}, c_{\tilde{\sigma}}), \qquad a_{\xi} \sim N_0(b_{\xi}, c_{\xi}); \\ \tau_{\tilde{\sigma}} &\sim Ga(d_{\tilde{\sigma}}, e_{\tilde{\sigma}}), \qquad \tau_{\xi} \sim Ga(d_{\xi}, e_{\xi}); \\ \zeta_{\tilde{\sigma}} &\sim Ga(f_{\tilde{\sigma}}, g_{\tilde{\sigma}}), \qquad \zeta_{\xi} \sim Ga(f_{\xi}, g_{\xi}); \end{aligned}$$

subject to the choice of arguments for these functions, i.e. the hyper–parameters which determine the precise Normal and Gamma distributions.

MCMC algorithm
We use a hybrid scheme (see Section 5.2.3) – specifically 'Metropolis with Gibbs' – to sample form the posteriors. This means we update each component singly using a Gibbs sampler where the conjugacy allows straightforward sampling from the full conditionals, and a Metropolis step elsewhere.

Some results
Some results are shown in Figures 23–26 and in Table 5. The main points to notice are listed below:

— Advantage of the hierarchical model over a standard likelihood–based analysis: a reduction in sampling variation (posterior standard deviations in the bottom portion of Table 5 are substantially smaller than the corresponding standard errors) due to the pooling of information across sites and seasons

— Figure 25 further highlights this reduction in variability – notice the *shrinkage* in estimates of the GPD shape parameter ξ in the Bayesian analysis relative to the standard likelihood–based analysis

— Separate seasonal parameters are recombined for each site to obtain site–by–site estimates of return levels (see Figure 25, bottom right); notice that estimates of extreme quantiles using maximum likelihood estimation can be very unstable, whereas the hierarchical model achieves a greater degree of stability through the pooling of information across sites

— Figure 26 shows an extension to *predictive return levels*, which cannot be achieved under the classical approach to inference

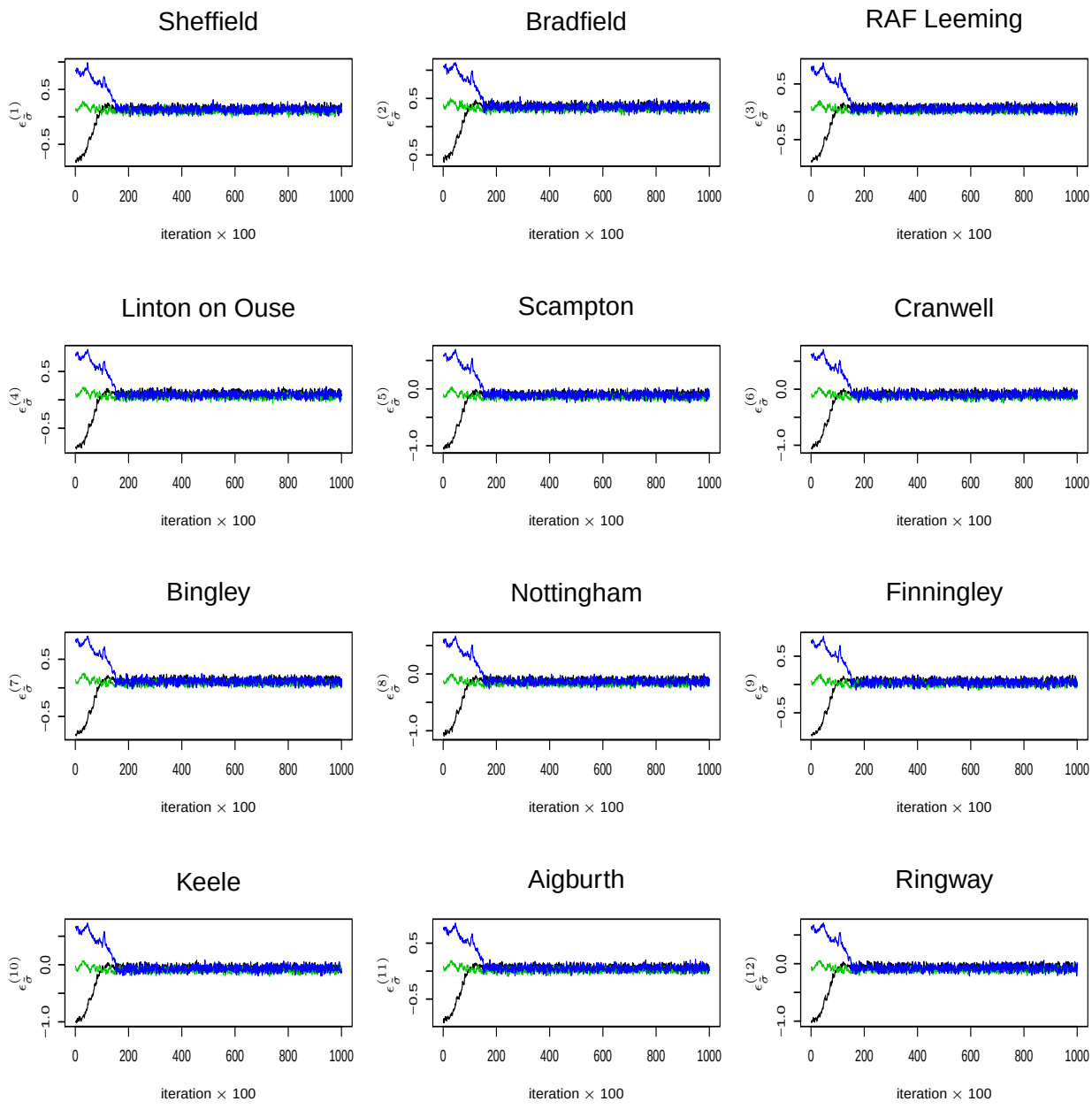

Figure 23: Trace plots of the site effects for $\log(\tilde{\sigma})$ for each site in the study

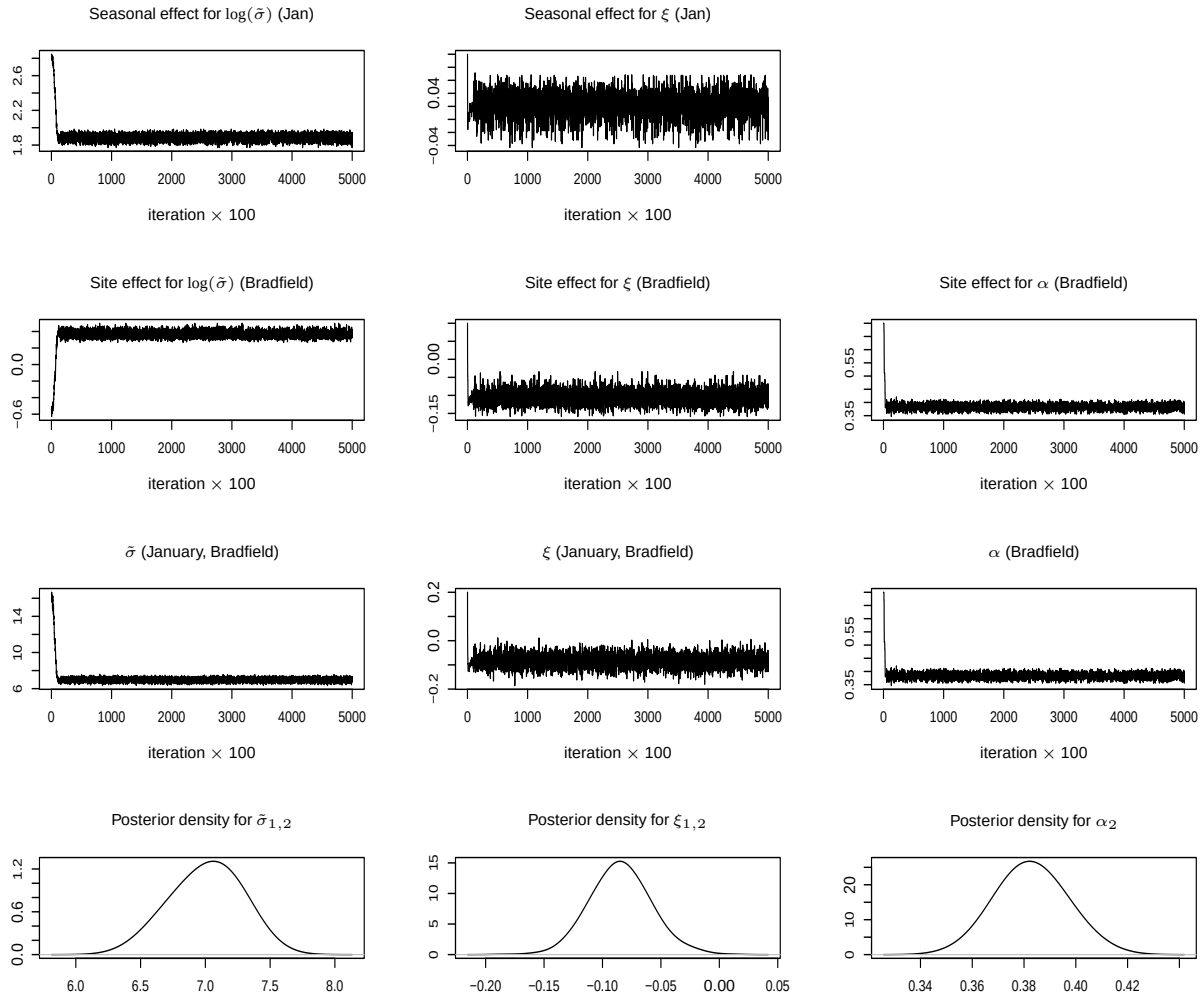

Figure 24: MCMC output for Bradfield in January

	Bradfield, January	**Nottingham, July**
	Mean (st. dev.) *MLE (asymp. s.e.)*	Mean (st. dev.) *MLE (asymp. s.e.)*
$\gamma_{\tilde{\sigma}}^{(m)}$	1.891 (0.042)	1.294 (0.042)
$\gamma_{\xi}^{(m)}$	0.021 (0.018)	0.002 (0.018)
$\epsilon_{\tilde{\sigma}}^{(j)}$	0.367 (0.044)	–0.121 (0.041)
$\epsilon_{\xi}^{(j)}$	–0.105 (0.020)	–0.059 (0.017)
$\epsilon_{\alpha}^{(j)}$	0.385 (0.009)	0.300 (0.011)
$\tilde{\sigma}_{m,j}$	7.267 (0.211) *8.149 (0.633)*	3.234 (0.061) *2.914 (0.163)*
$\xi_{m,j}$	–0.084 (0.015) *–0.102 (0.055)*	–0.057 (0.013) *0.018 (0.044)*
α_j	0.385 (0.009) *0.368 (0.012)*	0.400 (0.011) *0.412 (0.020)*

Table 5: Bayesian random effects analysis of extreme wind speeds – Bradfield (January) and Nottingham (July)

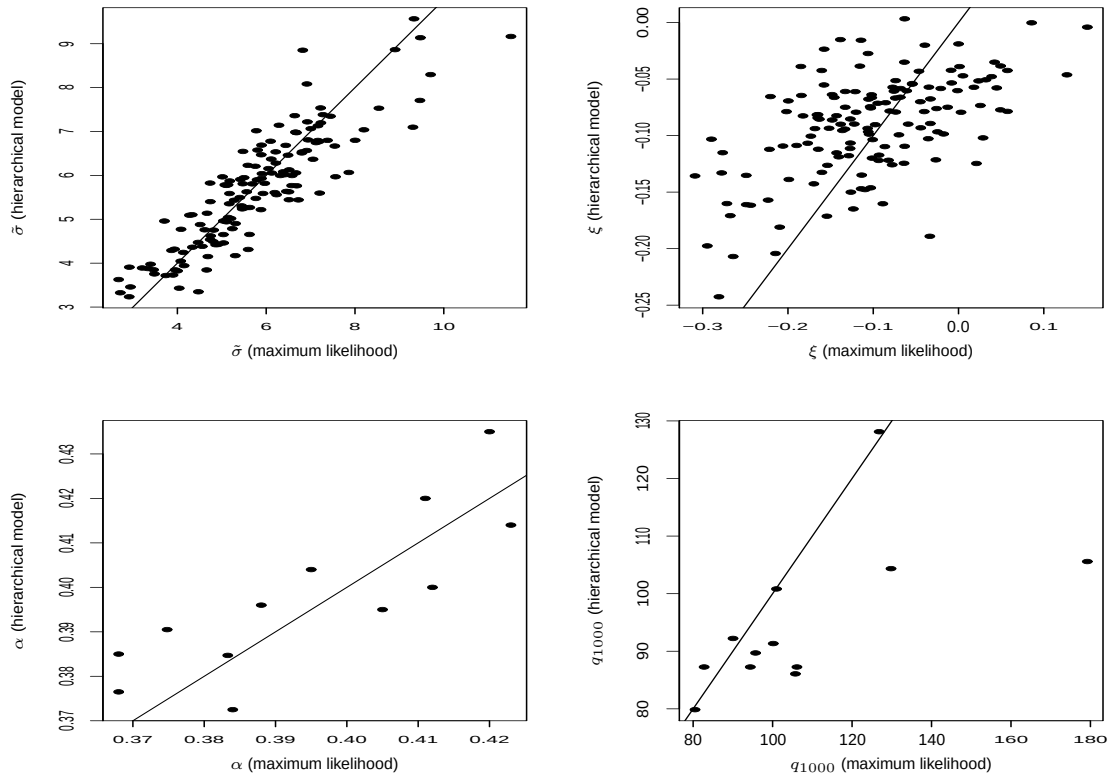

Figure 25: Posterior means against maximum likelihood estimates of GPD parameters, logistic dependence parameter and 1000–year return level

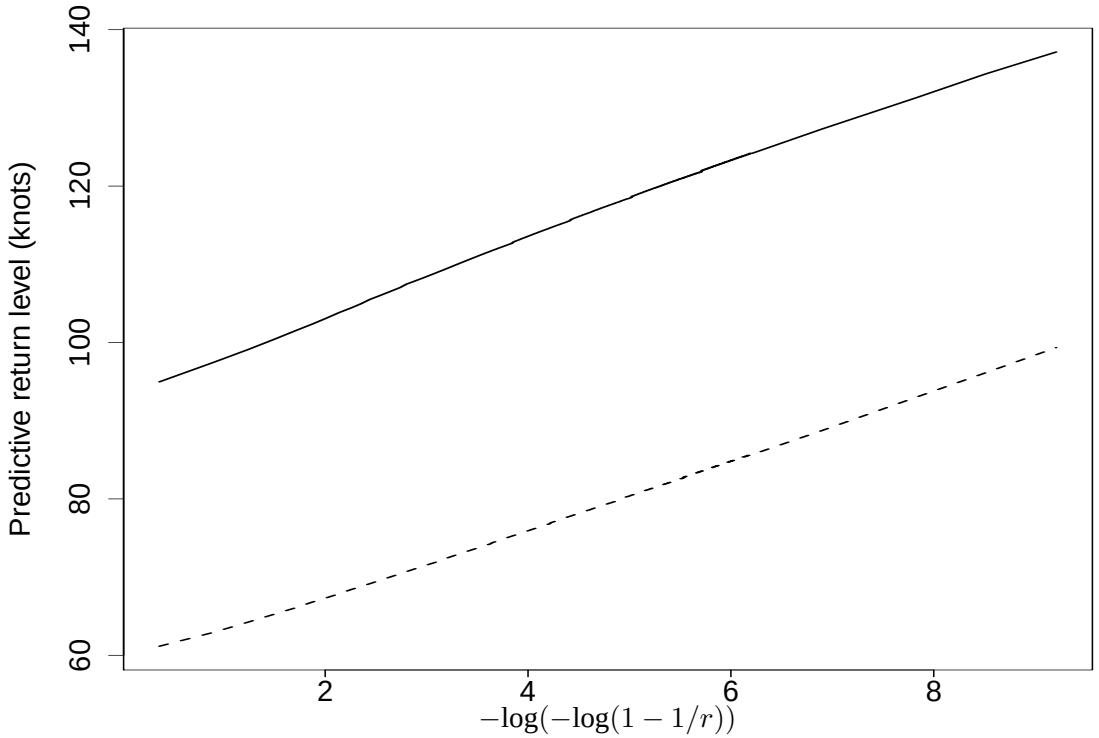

Figure 26: Predictive return level curves for Bradfield (—) and Nottingham (- - - -).