Session 5. Bayesian inference for extremes

5.1 General theory

5.2 Markov chain monte Carlo (MCMC)

- 5.2.1 The Gibbs sampler
- 5.2.2 Metropolis–Hastings sampling
- 5.2.3 Hybrid methods

5.3 Bayesian inference for extremes

- 5.3.1 Example: Port Pirie, South Australia
- 5.3.2 More complex structures: A random effects model

5. Bayesian inference for extremes

Throughout this short course, the method of maximum likelihood has provided a general and flexible technique for parameter estimation.

In fact, there are numerous other inferential procedures available, such as:

- Method of moments;
- Probability weighted moments;
- L–moments;
- Ranked set estimation; and, of course,
- **Bayesian inference**

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 $\mathbf{x} = (x_1, \ldots, x_n)$ to be realisations of a random variable whose density falls within a parametric family $\mathcal{F} = \{f(\mathbf{x}; \psi) : \psi \in \mathbf{\Psi}\}$.

However, parameters of a distribution are now treated as random variables, for which we specify **prior distributions**.

5.1 General theory

This often provides the main argument for – and against (!) – the use of Bayesian methods:

- 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.
- **But** since different analysts might specify different priors, conclusions become subjective.

5.1 General theory

Suppose we model our observed data $\mathbf{x}$ using the probability density function $f(\mathbf{x}; \boldsymbol{\psi})$.

The likelihood function for $\boldsymbol{\psi}$ is therefore $L(\boldsymbol{\psi}|\mathbf{x}) = f(\mathbf{x}; \boldsymbol{\psi})$.

Also, suppose our prior beliefs about likely values of $\boldsymbol{\psi}$ are expressed by the probability density function $\pi(\boldsymbol{\psi})$.

5.1 General theory

We can combine both pieces of information using Bayes Theorem, which states that

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

where

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

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

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

5.1 General theory

In equation (16), $\pi(\boldsymbol{\psi}|\mathbf{x})$ 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.

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 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.

5.2.1 The Gibbs sampler

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

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

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

5.2.1 The Gibbs sampler

Each simulated value depends only on the previous simulated value, and not any other previous values.

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 $p(\boldsymbol{\psi}_{i+1}, \boldsymbol{\psi}_i)$ 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 $p(\boldsymbol{\psi}^{(k-1)}, \boldsymbol{\psi}')$;

5.2.2 Metropolis–Hastings sampling

3. Evaluate the *acceptance probability* $A(\boldsymbol{\psi}^{(k)}, \boldsymbol{\psi}')$ of the proposed move, where

$$A(\boldsymbol{\psi}, \boldsymbol{\psi}') \quad = \quad \min\left\{1, \frac{\pi(\boldsymbol{\psi}')L(\boldsymbol{\psi}'|\mathbf{x})p(\boldsymbol{\psi}', \boldsymbol{\psi})}{\pi(\boldsymbol{\psi})L(\boldsymbol{\psi}|\mathbf{x})p(\boldsymbol{\psi}, \boldsymbol{\psi}')}\right\};$$

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.

5.2.2 Metropolis–Hastings sampling

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.

5.2.2 Metropolis–Hastings sampling

Common choices for the proposal distribution include:

- *symmetric chains*, where

$$p(\boldsymbol{\psi}, \boldsymbol{\psi}') \quad = \quad p(\boldsymbol{\psi}', \boldsymbol{\psi})$$

- *random walk chains*, where the proposal $\boldsymbol{\psi}'$ at iteration k is

$$\boldsymbol{\psi}' \quad = \quad \boldsymbol{\psi} + \varepsilon_k,$$

where the ε_k are IID random variables.

5.2.2 Metropolis–Hastings sampling

According to Coles (2001),

> "*...it's almost like magic... regardless of the choice of p, the rejection steps involved ensure that the simulated values have, in a limiting sense, the desired marginal distribution*"

In reality, there is alot more to the story, since choosing p so as to ensure a short "settling–in" period and low dependence between successive values can be difficult to arrange.

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
- Apply them to each component in turn for each iteration

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 (compelling?) reasons for preferring a Bayesian analysis of extremes over the more traditional likelihood approach.

- Extreme data are (by their very nature) scarce, so the ability to incorporate other sources of information through a prior distribution has obvious appeal
- Bayes' Theorem leads to an inference that comprises a complete distribution
- It is not dependent on the regularity assumptions required by the theory of maximum likelihood
- Implicit in the Bayesian framework is the concept of the **predictive distribution**

5.3 Bayesian inference for extremes

This *predictive distribution* describes how likely are different outcomes of a future experiment.

The predictive probability density function is given by

$$f(y|\mathbf{x}) = \int_{\boldsymbol{\psi}} f(y|\boldsymbol{\psi})\pi(\boldsymbol{\psi}|\mathbf{x})d\boldsymbol{\psi}$$

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.

5.3 Bayesian inference for extremes

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 $\mathbf{x} = (x_1, \ldots, x_n)$.

Thus, in the Bayesian framework, we would have

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

This equation gives the distribution of a future threshold excess, allowing for both parameter uncertainty and randomness in future observations.

5.3 Bayesian inference for extremes

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.

After removal of the "burn–in" period, the MCMC procedure gives a sample $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_B$. 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.

5.3.1 Example: Port Pirie, South Australia

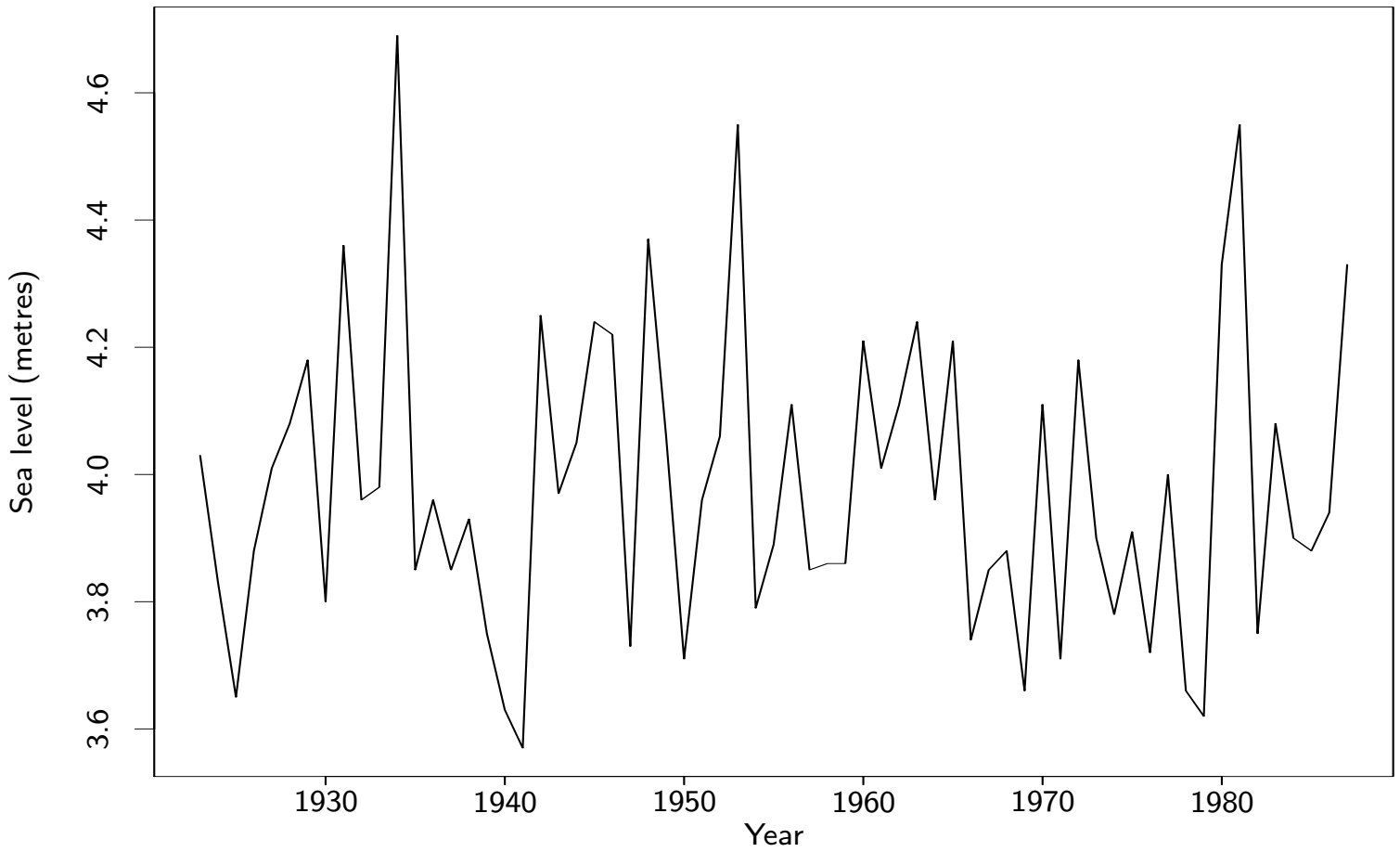
Figure 18 shows a time series plot of annual maximum sea levels at another Australian location – Port Pirie, in South Australia.

Notice that, unlike the corresponding data from Fremantle in Western Australia, there doesn't appear to be any trend in this series.

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

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

5.3.1 Example: Port Pirie, South Australia



5.3.1 Example: Port Pirie, South Australia

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 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

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

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

5.3.1 Example: Port Pirie, South Australia

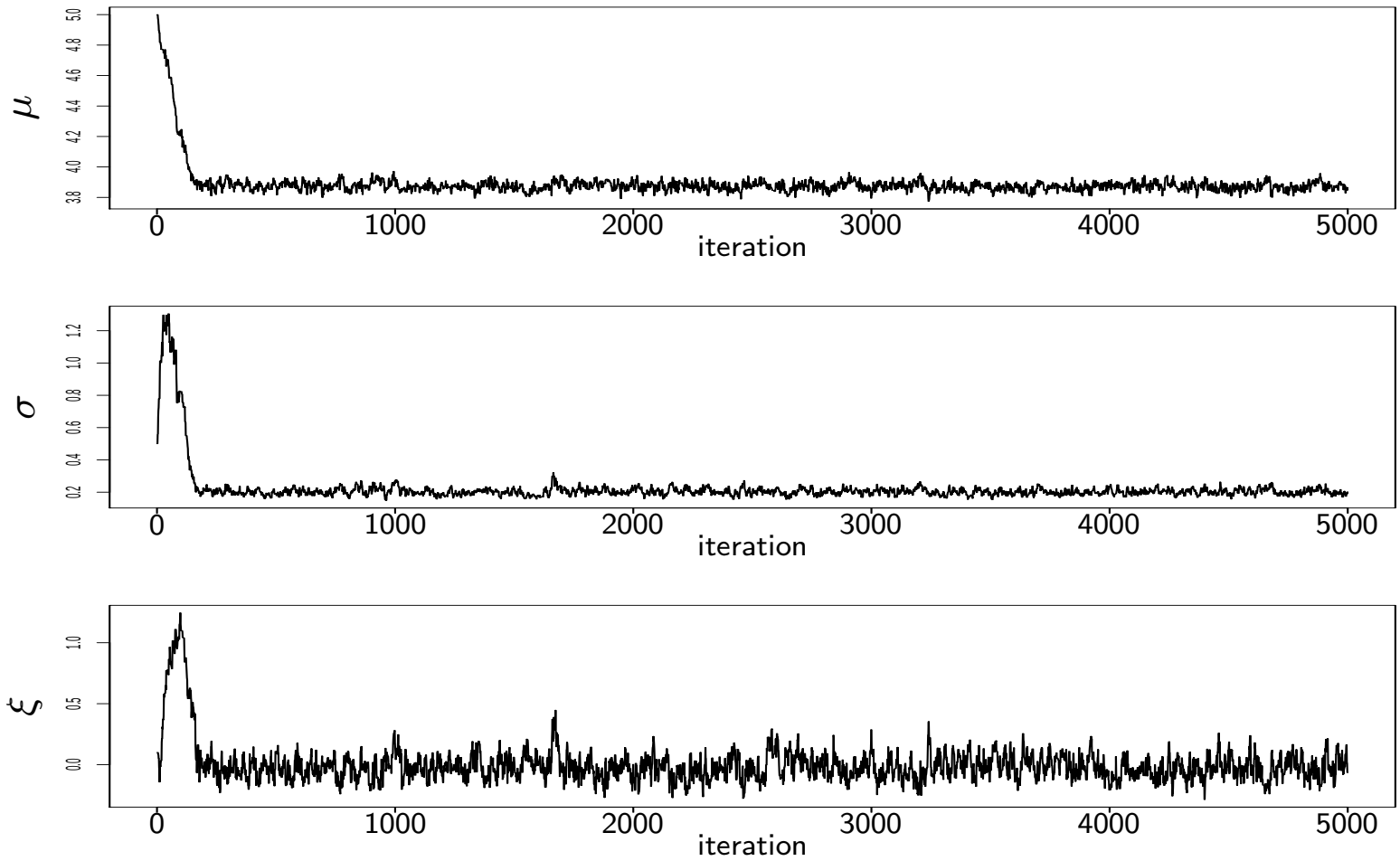
We use a Metropolis–Hastings MCMC sampling scheme since the full conditionals (for Gibbs sampling) are unobtainable.

After setting initial starting values for $\boldsymbol{\psi} = (\mu, \eta, \xi)$, we use a random walk update procedure to generate future values in the chain, i.e.

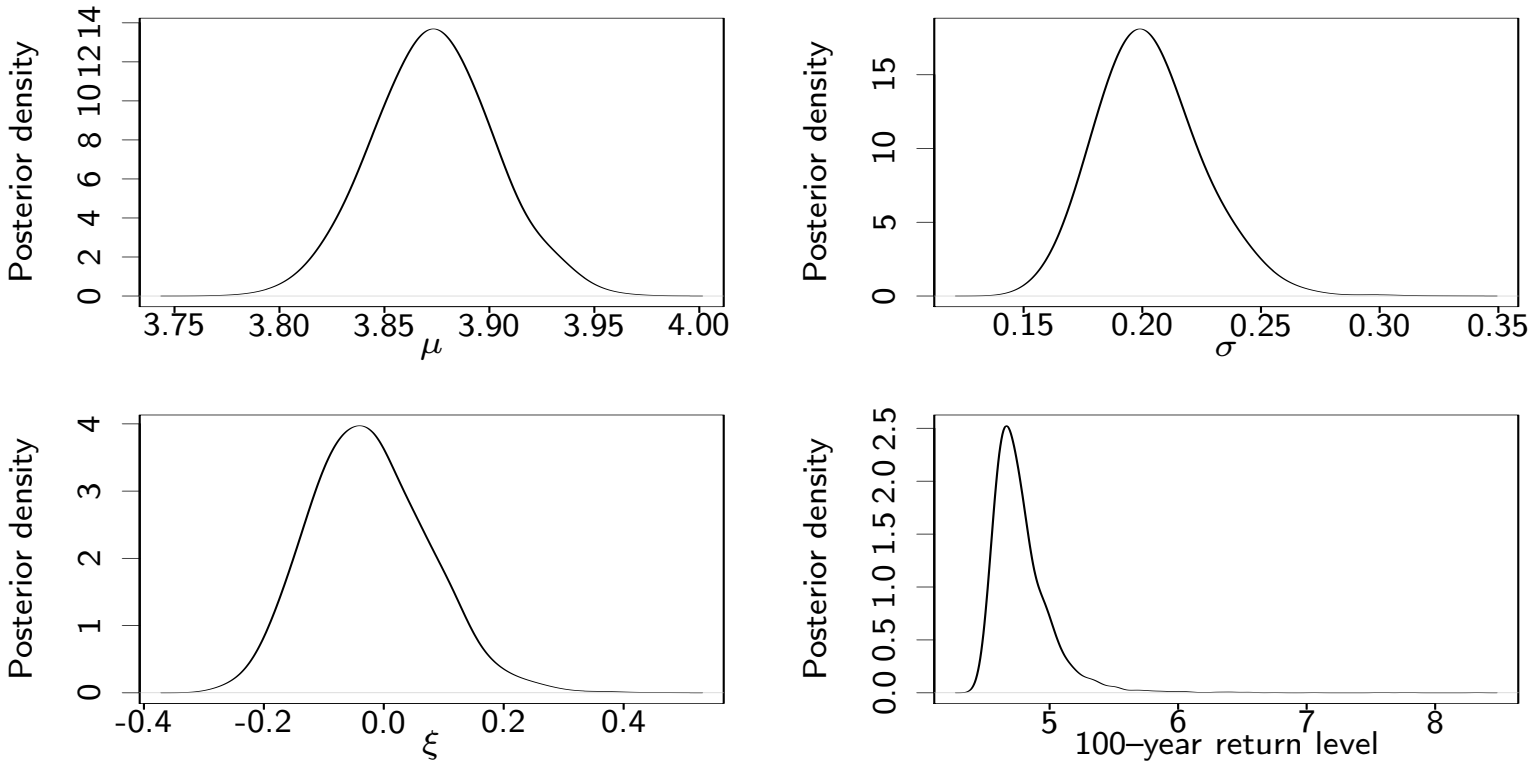
$$\begin{aligned} \mu' &= \mu_i + \epsilon_\mu \\ \eta' &= \eta_i + \epsilon_\eta \qquad \text{and} \\ \xi' &= \xi_i + \epsilon_\xi, \end{aligned}$$

with the ϵ being normally distributed with zero mean and variances v_μ, v_η and v_ξ respectively.

5.3.1 Example: Port Pirie, South Australia



5.3.1 Example: Port Pirie, South Australia



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

5.3.2 More complex structures

In this section we briefly discuss the work which will be presented 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.
- construct a model which is based on a standard limiting extreme value distribution, but
  - incorporates random effects for site variation
  - incorporates random effects for seasonal variation
  - explicitly models the temporal dependence inherent in the data

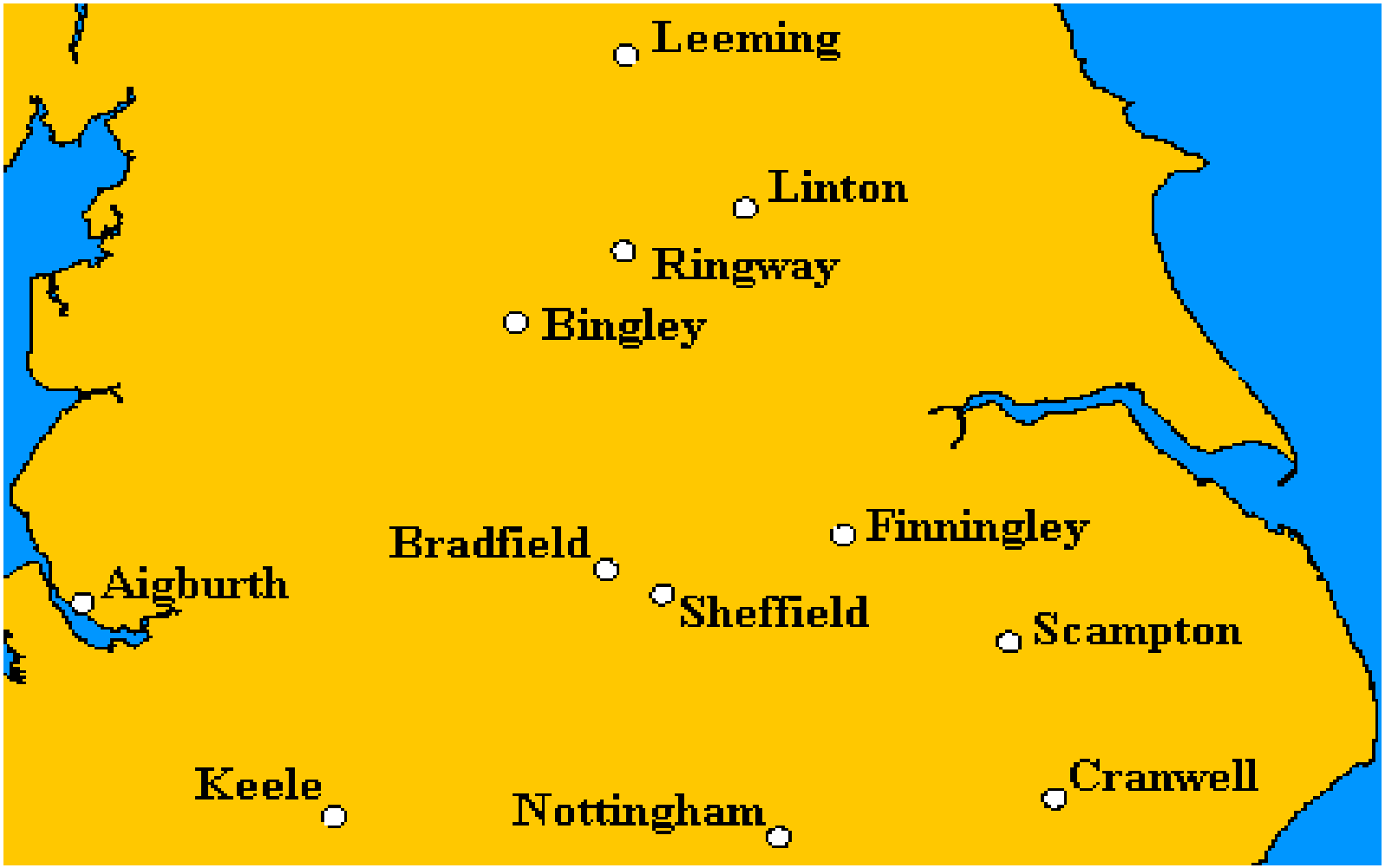
5.3.2 More complex structures

Figure 22 illustrates an exploratory analysis of data from two contrasting sites, Nottingham and Bradfield.

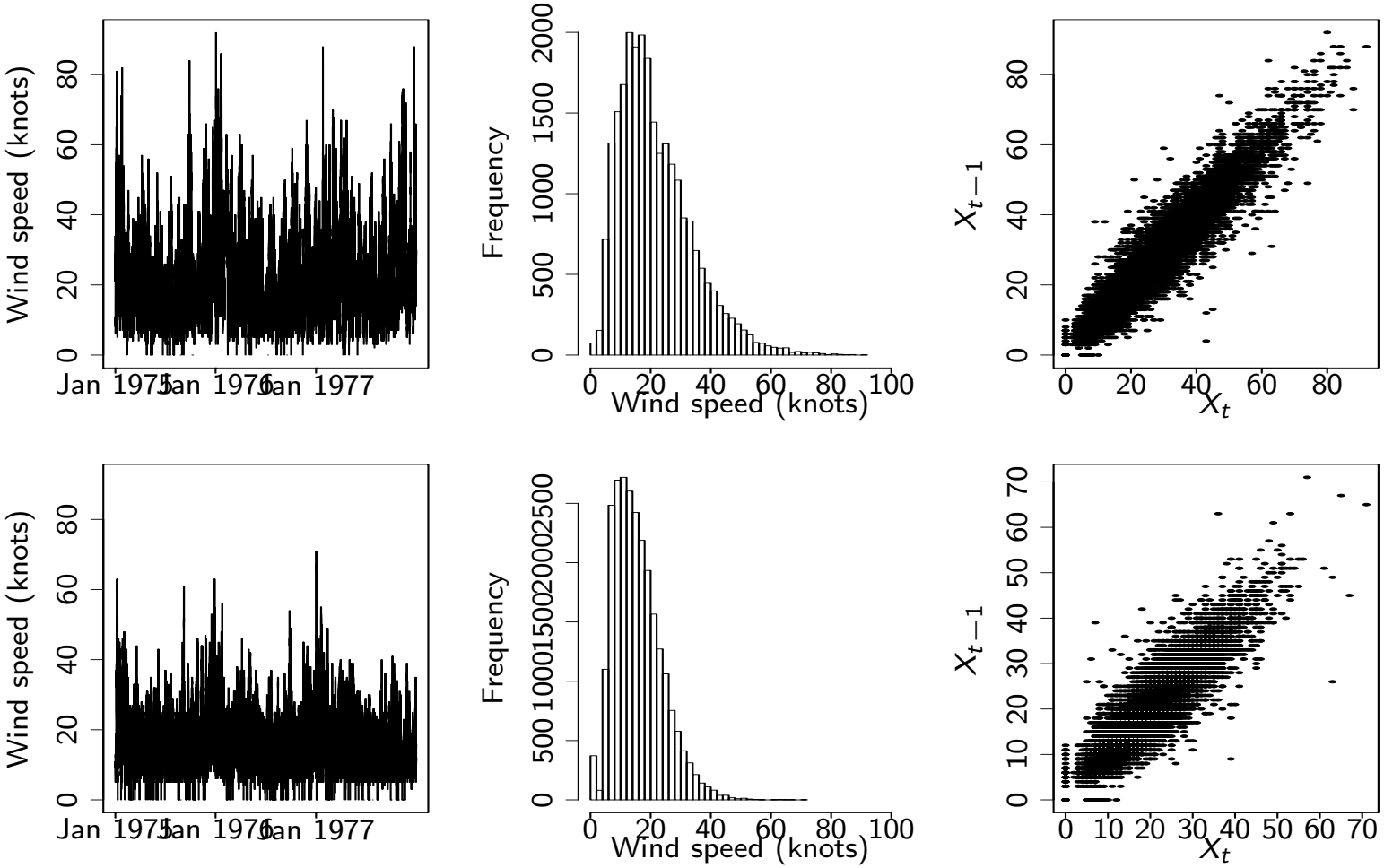
Shown are time series plots of the hourly maxima, histograms, and a plot of the time series against the version at lag 1.

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.

5.3.2 More complex structures



5.3.2 More complex structures



5.3.2 More complex structures

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).

5.3.2 More complex structures: Seasonal variation

Possible solution: Restrict an extreme value analysis to the 'season' which contains the 'most extreme' extremes (e.g. Coles and Tawn, 1991)

We want our model to **take account** of seasonal variability and identify all gusts which are large *given the time of year* as extreme!

Our solution: Fit a seasonally–varying GPD!

- For wind speed data, there is no natural partition into separate seasons (in the UK)
- We partition the annual cycle into 12 'seasons' (we use calendar months)
  - reflects well the continuous nature of seasonal climate changes
  - still enough data within each season for analysis!

5.3.2 More complex structures: Site variability

We take the same approach to allow for site variation.

Thus, our model will yield parameter pairs

$$(\sigma_{m,j}, \xi_{m,j}), \qquad \text{for } m, j = 1, \ldots, 12,$$

where m and j are indices of season and site (respectively).

We also need our threshold u to vary, since different criteria for what constitutes an extreme will be in play for each combination of season and site.

We denote by $u_{m,j}$ the threshold for identifying extremes in month m at site j.

5.3.2 More complex structures: Temporal dependence

The plot of the time series against the version at lag 1, for each site, shows the presence of substantial serial correlation between successive extremes.

What can be done?

1. '**Remove**' it – use "Peaks over threshold" (Davison and Smith, 1990)
2. '**Ignore**' it – initially, but then adjust standard errors post–analysis (Smith, 1991)
3. **Model** it

5.3.2 More complex structures: Temporal dependence

We have already seen – in Part 2 of this short course – the shortcomings of declustering (and using POT).

That leaves us with:

(a) Adjust inference post–analysis to account for temporal dependence

(b) Explicitly model the dependence present

We opt for (b), because:

- Our intention in this work is to investigate complex structures in the data, not *ignore* or *remove* them!
- Exploratory analyses support a simple first–order Markov model for the serial dependence (Fawcett and Walshaw, 2006)

5.3.2 More complex structures: Temporal dependence

The **logistic model** is one of the most flexible and accessible of these models (for example, see Tawn, 1988).

For consecutive threshold exceedances, the appropriate form of this model is given by:

$$F(x_i, x_{i+1}) = 1 - \left(Z(x_i)^{-1/\alpha} + Z(x_{i+1})^{-1/\alpha} \right)^{\alpha}, \qquad x_i, x_{i+1} > u,$$

where the transformation Z is given by

$$Z(x) = \lambda^{-1} \{1 + \xi(x-u)/\sigma\}_+^{1/\xi},$$

and ensures that the margins are of GPD form.

Independence and complete dependence are obtained when $\alpha = 1$ and $\alpha \searrow 0$ respectively.

5.3.2 More complex structures: Threshold stability

Recall that we denote by $(\sigma_{m,j}, \xi_{m,j})$ the parameters of the GPD assumed to be valid for threshold excesses in season m and site j.

To ensure **threshold stability** in our models, we now use

$$\tilde{\sigma}_{m,j} = \sigma_{m,j} - \xi_{m,j} u_{m,j}$$

5.3.2 More complex structures: Threshold stability

With this parameterisation, if $X - u^*_{m,j}$ follows a GPD $(\tilde{\sigma}_{m,j}, \xi_{m,j})$, where $u_{m,j} > u^*_{m,j}$, then

- $X - u_{m,j}$ also follows the same GPD,
- which is useful for comparisons across different sites and seasons.
- It also allows us to specify prior information about both parameters without having to worry about threshold dependency.

5.3.2 More complex structures: The model

With these assumptions in mind, we build the following **random effects model**:

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

where 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 $\tilde{\sigma}_{m,j}$.

5.3.2 More complex structures: The model

All random effects for $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$ are taken to be normally and independently distributed:

$$\begin{aligned} \gamma_{\tilde{\sigma}}^{(m)} &\sim N_0(0, \tau_{\tilde{\sigma}}) \qquad \text{and} \\ \gamma_{\xi}^{(m)} &\sim N_0(0, \tau_{\xi}), \qquad m = 1, \ldots, 12, \end{aligned}$$

for the **seasonal** effects, and

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

for the **site** effects.

In the absence of any prior knowledge about α_j, we set

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

5.3.2 More complex structures: The model

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:

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

5.3.2 More complex structures: MCMC

Estimation of the model is made via a **Metropolis within Gibbs** algorithm

- Here, we update each component singly using a Gibbs sampler where conjugacy allows;
- Elsewhere, we adopt a Metropolis step

5.3.2 More complex structures: MCMC

The **full conditionals** for the Gibbs sampling are:

$$
\begin{aligned}
a_\bullet|\ldots &\sim N\left(\frac{b_\bullet c_\bullet + \zeta_\bullet \sum \epsilon_\bullet^{(j)}}{c_\bullet + n_s\zeta_\bullet}, c_\bullet + n_s\zeta_\bullet\right);\\
\zeta_\bullet|\ldots &\sim Ga\left(f_\bullet + \frac{n_s}{2}, g_\bullet + \frac{1}{2}\sum(\epsilon_\bullet^{(j)} - a_\bullet)^2\right);\\
\tau_\bullet|\ldots &\sim Ga\left(d_\bullet + \frac{n_m}{2}, e_\bullet + \frac{1}{2}\sum(\gamma_\bullet^{(m)})^2\right);
\end{aligned}
$$

where $n_m = 12$ and $n_s = 12$.

The complexity of the GPD likelihood means that conjugacy is unattainable for the random effects parameters, and a Metropolis step is used here.

5.3.2 More complex structures: MCMC

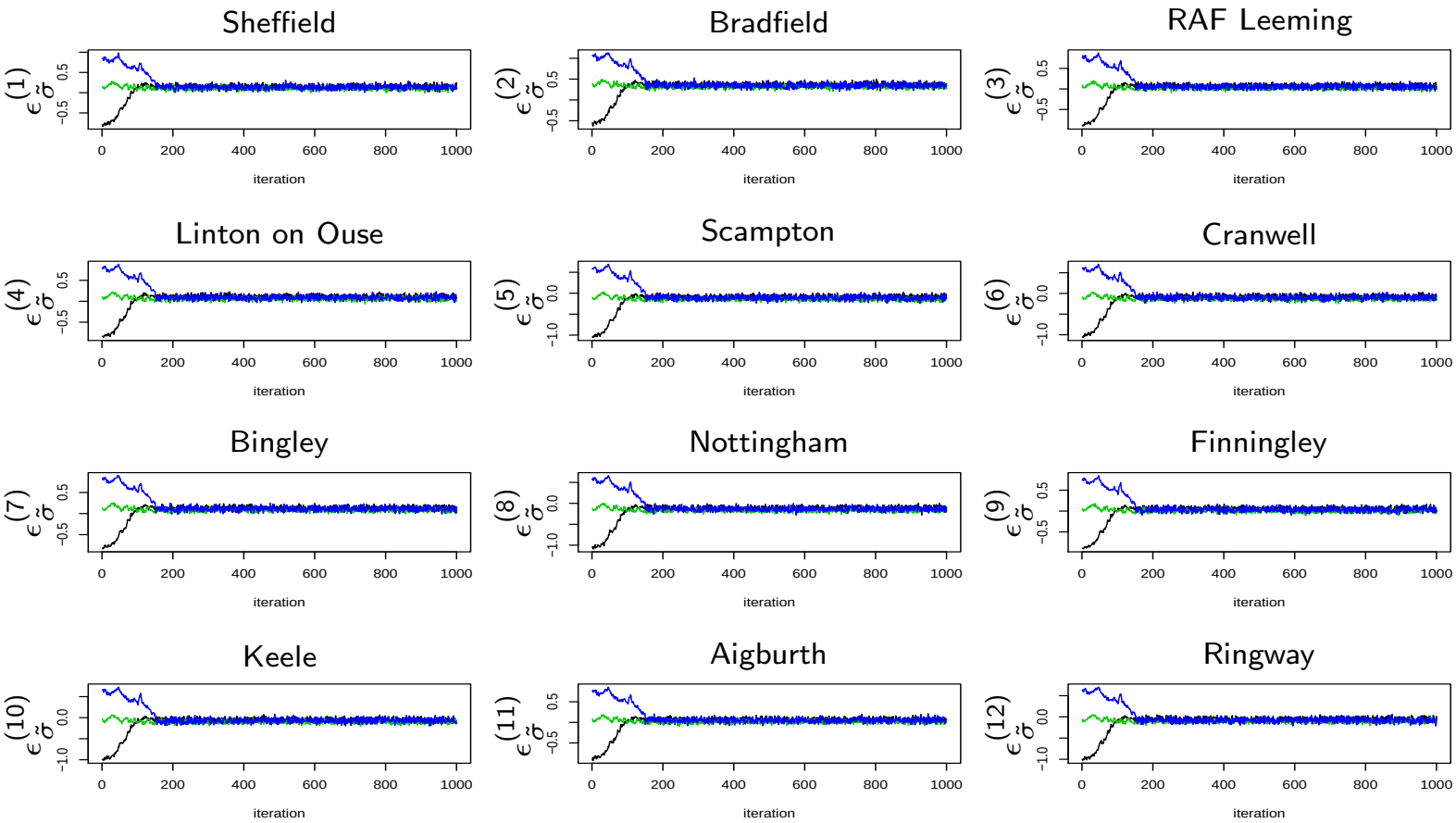
Obviously, we first need to specify appropriate hyper–parameters. In the absence of any expert prior knowledge, we use:

$$b_{\cdot} = 0, \qquad c_{\cdot} = 10^{-6}, \qquad d_{\cdot} = e_{\cdot} = f_{\cdot} = g_{\cdot} = 10^{-2}.$$

The implementation of the MCMC scheme then yields samples from the approximate posterior distributions for

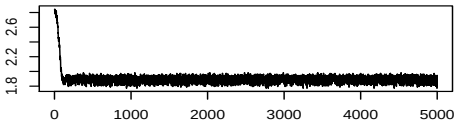
- the 12 site effect parameters for each of $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$;
- the 12 seasonal effect parameters for each of $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$, and
- the 12 site effect parameters for the dependence parameter α_j.

5.3.2 More complex structures: Some results

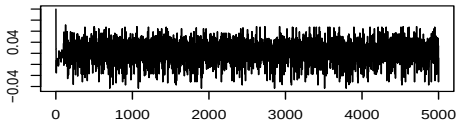


5.3.2 More complex structures: Some results

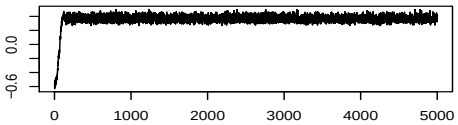
Seasonal effect for $\log(\tilde{\sigma})$ (Jan)



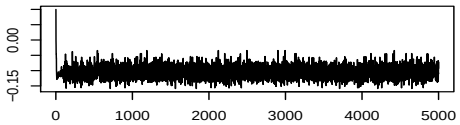
Seasonal effect for ξ (Jan)



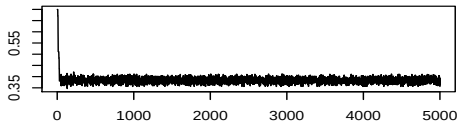
Site effect for $\log(\tilde{\sigma})$ (Bradfield)



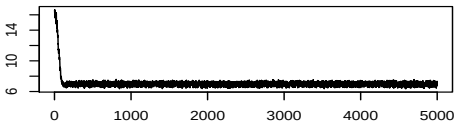
Site effect for ξ (Bradfield)



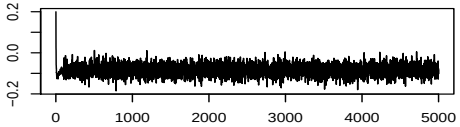
Site effect for α (Bradfield)



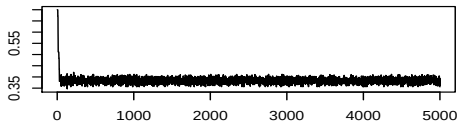
$\tilde{\sigma}$ (January, Bradfield)



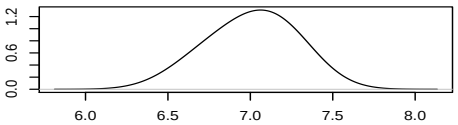
ξ (January, Bradfield)



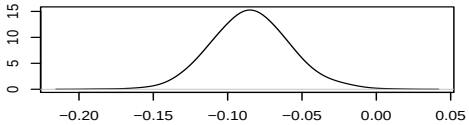
α (Bradfield)



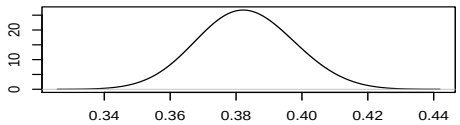
Posterior density for $\tilde{\sigma}_{1,2}$



Posterior density for $\xi_{1,2}$



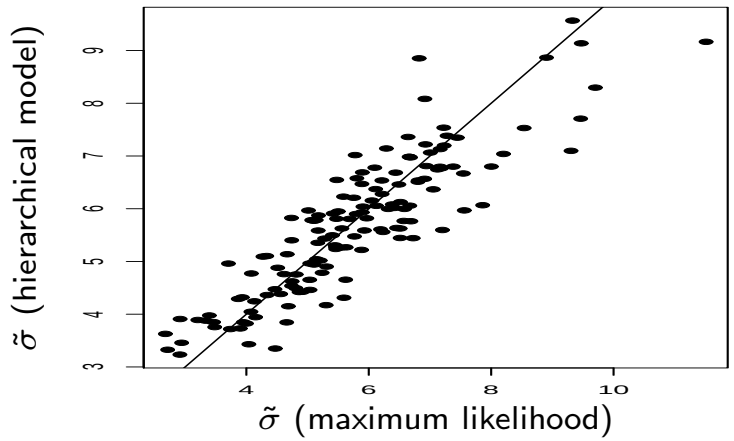
Posterior density for α_2

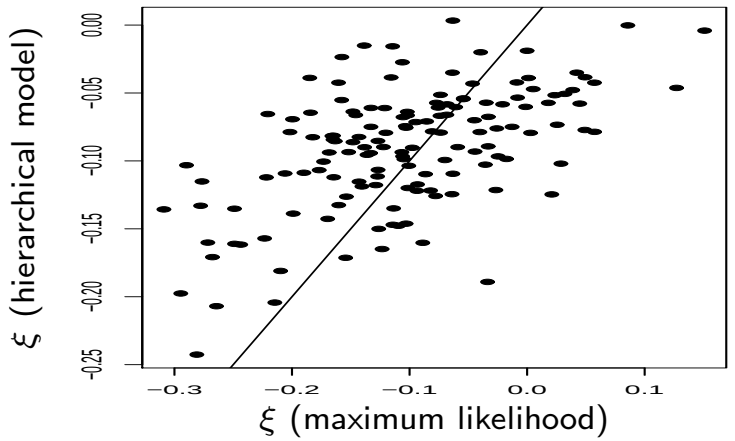


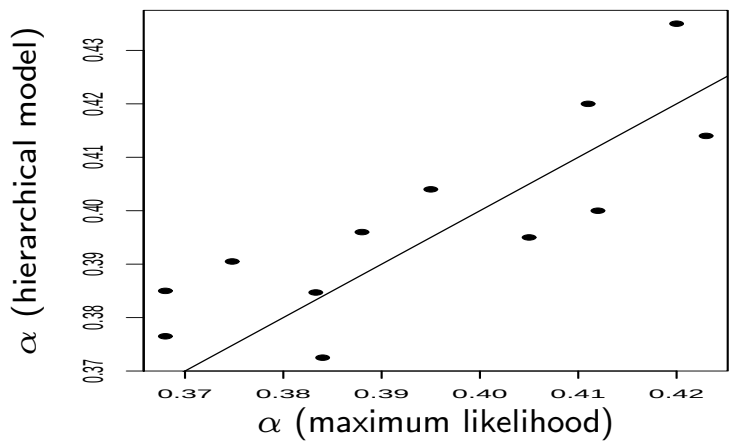
5.3.2 More complex structures: Some results

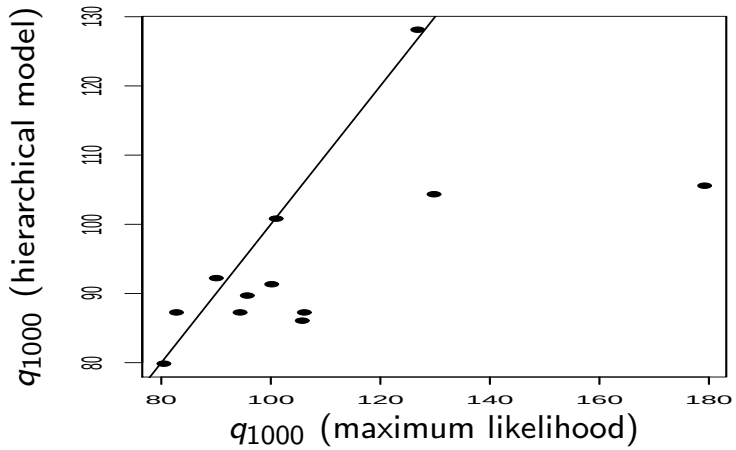
	Bradfield, January	**Nottingham, July**
	Mean (st. dev.) MLE (s.e.)	Mean (st. dev.) MLE (s.e.)
$\gamma_{\tilde{\sigma}}^{(m)}$	1.891 (0.042)	1.294 (0.042)
$\gamma_{\xi}^{(m)}$	0.021 (0.018)	0.002 (0.018)
$\epsilon_{\tilde{\sigma}}^{(j)}$	0.367 (0.044)	–0.121 (0.041)
$\epsilon_{\xi}^{(j)}$	–0.105 (0.020)	–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)

5.3.2 More complex structures: Some results

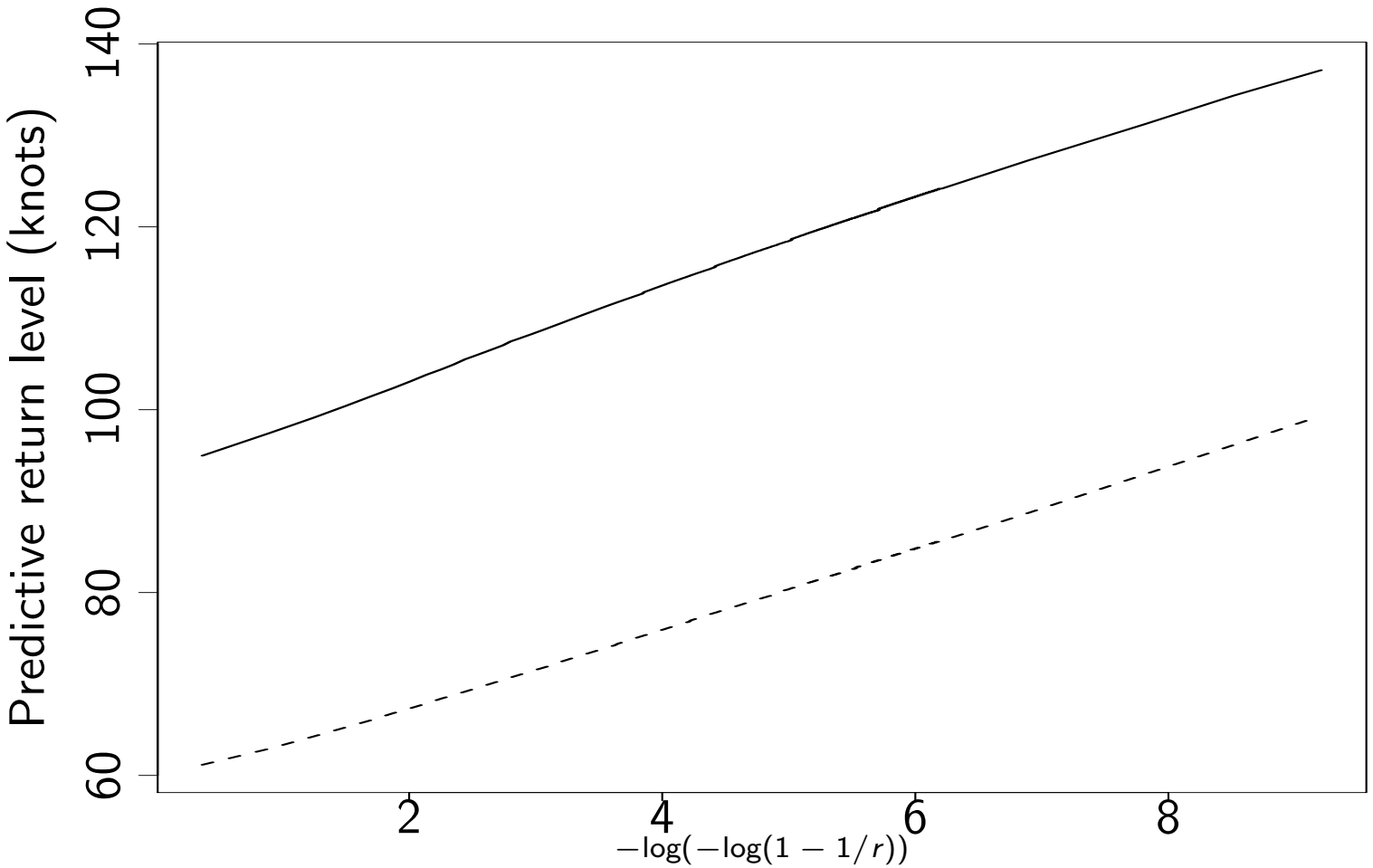








5.3.2 More complex structures: Some results



5.3.2 More complex structures: Some results

The main take–home points are:

- A reduction in sampling variation under the Bayesian hierarchical model
  - posterior standard deviations substantially smaller than the corresponding standard errors...
  - ... probably due to the pooling of information across sites and seasons
  - This is also evident in the "shrinkage plots"
- Estimates of return levels using maximum likelihood estimation can be very unstable – the hierarchical model achieves a greater degree of stability
- The Bayesian paradigm offers an extension to **predictive return levels**, which cannot be achieved under the classical approach to inference