Appl. Statist. (2006)
55, *Part* 5, *pp.* 631–646

A hierarchical model for extreme wind speeds

Lee Fawcett and David Walshaw

University of Newcastle, UK

[Received April 2005. Final revision June 2006]

Summary. A typical extreme value analysis is often carried out on the basis of simplistic inferential procedures, though the data being analysed may be structurally complex. Here we develop a hierarchical model for hourly gust maximum wind speed data, which attempts to identify site and seasonal effects for the marginal densities of hourly maxima, as well as for the serial dependence at each location. A Gaussian model for the random effects exploits the meteorological structure in the data, enabling increased precision for inferences at individual sites and in individual seasons. The Bayesian framework that is adopted is also exploited to obtain predictive return level estimates at each site, which incorporate uncertainty due to model estimation, as well as the randomness that is inherent in the processes that are involved.

Keywords: Extreme value theory; Generalized Pareto distribution; Hierarchical model; Markov chain; Wind speeds

1. Introduction

In this paper we develop a hierarchical model for hourly maximum wind speeds over a region of central and northern England. The data that are used consist of hourly gust maximum wind speeds recorded for the British Meteorological Office at 12 locations (Fig. 1). 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 that is inherent in the time series of hourly maximum speeds obtained at each site.

For 10 of the 12 sites, data are recorded over a period of 18 years, from January 1st, 1974, to December 31st, 1991, inclusively, constituting approximately 157000 observations at each site. For two of the sites, Sheffield and Bradfield, data are available for 10 years, from January 1st, 1975, to December 31st, 1984, inclusively, constituting approximately 86000 observations at each site. The 12 sites represent a variety of geographical locations—both urban and rural, high and low altitudes, easterly and westerly positions etc. Fig. 2 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 3 years of data only are used in each case, to illustrate the relevant data characteristics best.

The stations at Bradfield and Nottingham are chosen to represent contrasting climate characteristics in terms of the wind behaviour. Bradfield is at a relatively high altitude (380 m above sea-level), rural and exposed. The Nottingham station, in contrast, is low lying (50 m above sea-level), urban and shielded by some nearby buildings, being a city centre site. The effects of these differences are clearly seen in Fig. 2. Although both sites have stronger winds in the

Address for correspondence: David Walshaw, School of Mathematics and Statistics, Merz Court, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK.
E-mail: David.Walshaw@ncl.ac.uk

© 2006 Royal Statistical Society 0035–9254/06/55631

Fig. 1. Location of the wind speed stations

winter, the seasonal pattern is more obviously defined at Bradfield. The distribution of hourly maxima at Bradfield is shifted considerably to the right of that at Nottingham, with the mean, median and quartiles (which are not shown) all being substantially greater. We also observe a stronger serial correlation at lag 1 at Bradfield. This is likely to arise as a consequence of the fact that urban sites such as Nottingham generally exhibit more *turbulence*, leading to a greater variance in the distribution of the hourly maximum speed if we condition, say, on the value of the hourly mean. The greater variation that is observed in the hourly maxima at such a location is likely to weaken the relationship between consecutive hourly maxima. For a detailed description of the behaviour of the wind climate in the UK, see for example Shellard (1976) and Smith (1983).

The aim of our hierarchical model is to exploit the complex structure that is inherent in the data to improve over simplistic inferential procedures, which would more typically be applied to sites individually. Coles (2002) achieved a similar aim for annual maximum wind speeds that were obtained from about 100 locations in the USA. In the present paper the much more detailed (hourly) records require a different extremal model, the incorporation of the seasonal random effects, and a model for the serial dependence. The Bayesian paradigm transfers naturally to the problems of inference that are inherent in our analysis, and we agree with arguments that were presented by Coles and Powell (1996) which support the Bayesian approach as a good means of exploiting data on extreme events, which are, by definition, rare. We employ Markov chain Monte Carlo methods to address the analytically intractable problems of inference.

Section 2 focuses on the construction of the hierarchical model. In Section 3 we present results for parameter estimation, together with some diagnostic assessments. Finally, Section 4 discusses results for return level inference, both estimative and predictive. The wind speed data (threshold exceedances only) for each of the 12 sites that are studied in this paper, as well as the R code that was used and instructions on how to use the code can be obtained from `www.mas.ncl.ac.uk/~nlf8/c5776R/home.html`.

2. A hierarchical model

2.1. A model for threshold exceedances

A natural way of modelling extremes of time series such as the hourly maximum wind speeds is to use the generalized Pareto distribution (GPD) as a model for excesses over a high threshold.

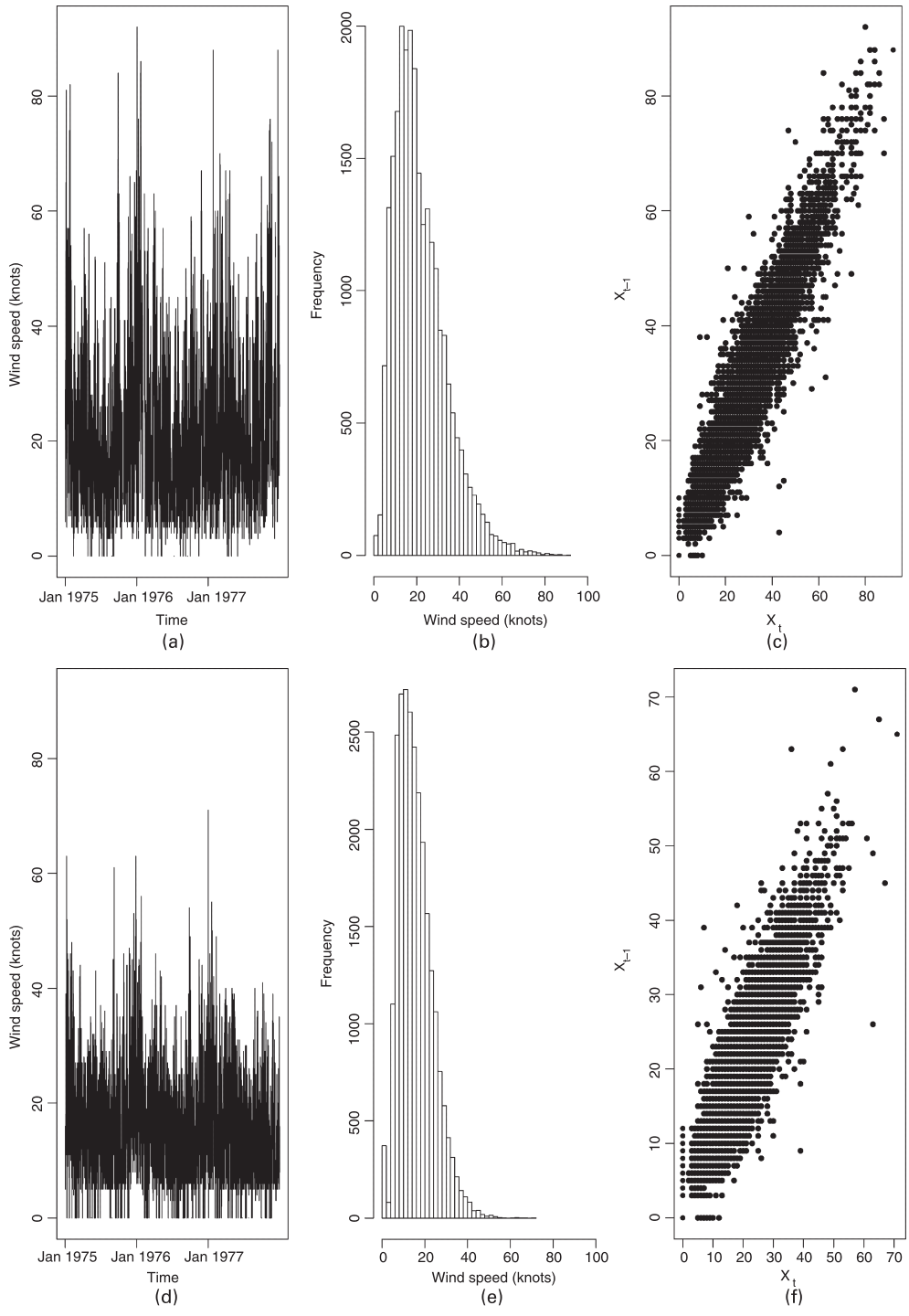

Fig. 2. Time series plots and histograms of hourly gusts observed at (a), (b) Bradfield and (d), (e) Nottingham over a 3-year period (1975–1977 inclusively): also shown are (c), (f) plots of the time series against the lagged series

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common distribution function F. Then, for a high value of the threshold u, the conditional distribution $(X - u | X > u)$ will be approximately of GPD form, with cumulative distribution function

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

where $a_+ = \max(0, a)$ and σ and ξ are scale and shape parameters respectively. This arises because, if a limiting distribution exists for any normalization of $(X - u | X > u)$, then this limit will be exactly of GPD form (Davison and Smith, 1990). The limit is taken as $u \uparrow u_F$, where u_F is the finite upper end point of F if it exists, or ∞ otherwise. This basic theory sets up the model for independent and identically distributed random variables; additional model structure to handle non-stationarity and dependence is considered in Sections 2.2 and 2.3 (respectively).

2.2. The marginal distribution for exceedances at site j, season m

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 12 '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 that is used for excesses modelled by equation (1) to vary, since different criteria about what constitutes an extreme value will be in play for each combination of season and site. We shall denote by $u_{m,j}$ the value of the exceedance threshold for month m and site j.

2.3. Modelling temporal dependence

By far the most common approach that is used to handle the serial correlation inherent in extremes of environmental variables is to identify 'clusters' of extremes, and to model only the cluster peaks, which are taken to be a set of independent exceedances; see, for example, Davison and Smith (1990) and Walshaw (1994). However, Fawcett and Walshaw (2006a) make a strong case against this procedure, demonstrating that the discarding of the majority of exceedances is not only wasteful but also often appears to introduce bias into return level estimation, which would not otherwise be present (simulation results are presented in Fawcett (2005) and Fawcett and Walshaw (2006a), with some theoretical results also available in Fawcett (2005)). For this reason we prefer to retain all exceedances over a high threshold for modelling. This leaves us with the options of

(a) somehow adjusting the inferences to take account of the dependence in the data that are used for modelling or
(b) explicitly modelling the dependence.

Here we adopt option (b), since the data support a simple first-order Markov chain model for the serial dependence. The partial autocorrelation functions for all sites (which are not shown) indicate a very large value at lag 1, with all lags greater than 1 showing negligible partial autocorrelation. Although this does not necessarily imply that the temporal dependence is entirely explained by the first-order dependence for the extreme values, it does at least suggest this as

a reasonable assumption for modelling, which can be assessed later. The stochastic properties of such a chain are completely determined by the joint distribution of successive pairs. Given a model $f(x_i, x_{i+1}; \psi)$ specified by parameter vector ψ, it follows that the likelihood for ψ is given by

$$L(\psi) = f(x_1; \psi) \prod_{i=1}^{n-1} f(x_i, x_{i+1}; \psi) \Big/ \prod_{i=1}^{n-1} f(x_i; \psi). \tag{2}$$

Contributions to the numerator can be modelled by using an appropriate bivariate extreme value model, thereby capturing the first-order Markov structure as suggested above. One of the most flexible and accessible of such models is the logistic model (for example see Tawn (1988) and Coles (2001)). For consecutive threshold exceedances, the appropriate form of this model is

$$F(x_i, x_{i+1}) = 1 - \{Z(x_i)^{-1/\alpha} + Z(x_{i+1})^{-1/\alpha}\}^{\alpha}, \qquad x_i, x_{i+1} > u, \tag{3}$$

where the transformation Z, for generic threshold u, GPD parameters σ and ξ, and exceedance rate per observation, λ, is given by

$$Z(x) = \lambda^{-1}\{1 + \xi(x-u)/\sigma\}_{+}^{1/\xi} \tag{4}$$

and ensures that the margins are also of GPD form. The parameter $\alpha \in (0, 1]$ measures the strength of dependence between consecutive extremes, smaller values indicating stronger dependence. Independence and complete dependence are obtained when $\alpha = 1$ and in the limit as $\alpha \downarrow 0$ respectively. We take a pragmatic approach to the parameter λ, using the empirical proportion of threshold exceedances as a fixed point estimate for each location and each month. Although it would be more consistent with our overall approach to incorporate exceedance rate parameters into our hierarchical model fully, the experience of both authors is that to allow for inference on such parameters has negligible influence on the inferences for other parameters, including return levels, while significantly increasing the computational load (see Walshaw (1991) and Fawcett (2005)).

Inference for this model is complicated by the fact that a bivariate pair may exceed a specified threshold in just one of its components. Let

$$\begin{aligned} R_{0,0} &= (0, u) \times (0, u), \\ R_{1,0} &= [u, \infty) \times (0, u), \\ R_{0,1} &= (0, u) \times [u, \infty), \\ R_{1,1} &= [u, \infty) \times [u, \infty), \end{aligned}$$

so that, for example, a point $(x_i, x_{i+1}) \in R_{1,0}$ if x_i exceeds the threshold but x_{i+1} does not. For points in $R_{1,1}$, model (3) applies, the density of which gives the appropriate likelihood component. However, for a point in $R_{1,0}$ or $R_{0,1}$, the corresponding likelihood component is given by

$$\frac{\partial F}{\partial x_i}\Big|_{(x_i, u)}$$

or

$$\frac{\partial F}{\partial x_{i+1}}\Big|_{(u,x_{i+1})}$$

respectively. If both x_i and x_{i+1} fall below the threshold u, the contribution to the numerator in equation (2) is given by the distribution function evaluated at the threshold u. The denominator in equation (2) is simply replaced with the corresponding univariate densities by using equation (1).

2.4. Construction of the hierarchical model

We assume that the GPD model (1) is valid for exceedances over a high threshold, allowing this threshold and the GPD parameters to vary across sites and seasons. We also make the assumption that extremes between sites and between seasons are independent, but that successive extremes *within* seasons have a first-order Markov dependence structure and so can be modelled by using equation (2). Although in reality there will clearly be some spatial dependence in the weather observed at locations that are separated by the distances involved, this does not translate to strong correlations in extreme values across sites. Further, the temporal dependence in extremes of wind speed is typically short lived, so independence between months is a reasonable assumption. This also justifies the use of a first-order Markov chain model for extremes within seasons, though Fawcett and Walshaw (2006b) examine models which allow for longer-range dependence.

We also make the assumption that there is no interaction between seasonal and site effects. Again, the mild contrast in seasonal behaviour that is illustrated by Bradfield and Nottingham warns that this is, at best, an approximation, but we shall see from an assessment of performance that our model appears to be robust to any moderate departures from this and the other assumptions. A more detailed discussion can be found in Fawcett (2005). Finally, in our model we assume that both spatial effects and seasonal (monthly) effects are exchangeable *a priori*. It would be reasonable to suggest that this last assumption may seem oversimplistic: for example we would, *a priori*, expect there to be clear temporal structure in the data, with wind speeds in January and February being more similar than those in January and July, for instance. However, we have investigated more complex model structures which exploit such temporal structure and have found little benefit for the added complexity. We return to this issue in Section 3.2.

We denote by $(\sigma_{m,j}, \xi_{m,j})$ the parameters of the GPD model (1) which is assumed to be valid for threshold excesses in season (month) m and site j. 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 (3) describes the dependence over all seasons at site j, with a logistic dependence parameter α_j substituted for α. 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 parameterization, 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 (for example 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 assumptions above lead us to specify the following random-effects model:

$$\log(\tilde{\sigma}_{m,j}) = \gamma_{\tilde{\sigma}}^{(m)} + \varepsilon_{\tilde{\sigma}}^{(j)}, \tag{5}$$

$$\xi_{m,j} = \gamma_{\xi}^{(m)} + \varepsilon_{\xi}^{(j)}, \tag{6}$$

$$\alpha_j = \varepsilon_{\alpha}^{(j)}, \tag{7}$$

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:

$$\gamma_{\tilde{\sigma}}^{(m)} \sim N_0(0, \tau_{\tilde{\sigma}}), \tag{8}$$

$$\gamma_{\xi}^{(m)} \sim N_0(0, \tau_{\xi}), \qquad m = 1, \ldots, 12, \tag{9}$$

for the seasonal effects, and

$$\varepsilon_{\tilde{\sigma}}^{(j)} \sim N_0(a_{\tilde{\sigma}}, \zeta_{\tilde{\sigma}}), \tag{10}$$

$$\varepsilon_{\xi}^{(j)} \sim N_0(a_{\xi}, \zeta_{\xi}), \qquad j = 1, \ldots, 12, \tag{11}$$

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 distributions (8) and (9) to avoid overparameterization 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

$$\varepsilon_{\alpha}^{(j)} \sim U(0, 1). \tag{12}$$

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

$$a_{\tilde{\sigma}} \sim N_0(b_{\tilde{\sigma}}, c_{\tilde{\sigma}}), \qquad a_{\xi} \sim N_0(b_{\xi}, c_{\xi}), \tag{13}$$

$$\tau_{\tilde{\sigma}} \sim \mathrm{Ga}(d_{\tilde{\sigma}}, e_{\tilde{\sigma}}), \qquad \tau_{\xi} \sim \mathrm{Ga}(d_{\xi}, e_{\xi}), \tag{14}$$

$$\zeta_{\tilde{\sigma}} \sim \mathrm{Ga}(f_{\tilde{\sigma}}, g_{\tilde{\sigma}}), \qquad \zeta_{\xi} \sim \mathrm{Ga}(f_{\xi}, g_{\xi}), \tag{15}$$

subject to the choice of arguments for these functions, i.e. the hyperparameters which determine the precise normal and gamma distributions.

3. Analysis of the wind speed data

The first step of the analysis involves setting the thresholds $u_{m,j}$ to select the set of exceedances for modelling. This is achieved here by adopting the standard approach of using mean residual life plots to identify an appropriate threshold for each month m and location j. The details for this can be found in Fawcett (2005), with the method itself being described well by Davison and Smith (1990).

Estimation for the model that was developed in Section 2 can then be carried out following an appropriate choice of hyperparameters in equations (13)–(15), with the inference being based on the implementation of an appropriate Markov chain Monte Carlo scheme. The algorithm that was employed is a Metropolis within Gibbs algorithm, i.e. we update each component singly by using a Gibbs sampler where the conjugacy allows straightforward sampling from the full conditionals, and a Metropolis step elsewhere. The full conditionals for the Gibbs sampling are

$$a_{\cdot}|\ldots \sim N\left(\frac{b_{\cdot}c_{\cdot} + \zeta_{\cdot}\sum \varepsilon_{\cdot}^{(j)}}{c_{\cdot} + n_s\zeta_{\cdot}}, c_{\cdot} + n_s\zeta_{\cdot}\right), \tag{16}$$

$$\zeta_{\cdot}|\ldots \sim \mathrm{Ga}\left\{ f_{\cdot} + \frac{n_s}{2}, g_{\cdot} + \frac{1}{2}\sum(\varepsilon_{\cdot}^{(j)} - a_{\cdot})^2 \right\}, \tag{17}$$

$$\tau_{\cdot}|\ldots \sim \mathrm{Ga}\left(d_{\cdot} + \frac{n_m}{2}, e_{\cdot} + \frac{1}{2}\sum \gamma_{\cdot}^{(m)2} \right), \tag{18}$$

where n_m is the number of months, 12, and n_s is number of sites, 12, and here the notation $\zeta_{\cdot}$, for example, is used generically to denote either $\zeta_{\tilde{\sigma}}$ or ζ_{ξ}. The complexity of the likelihood which is derived from model (1) means that conjugacy is unattainable for the random-effect parameters, and a Metropolis step is used to update each of these.

3.1. Markov chain Monte Carlo simulations

In the absence of expert prior knowledge, prior parameters were chosen to give a highly non-informative specification:

$$b_{\cdot} = 0, \qquad c_{\cdot} = 10^{-6}, \qquad d_{\cdot} = e_{\cdot} = f_{\cdot} = g_{\cdot} = 10^{-2}. \tag{19}$$

The implementation of the Markov chain Monte Carlo scheme yields samples from the approximate posterior distributions for the 12 site effect parameters and the 12 seasonal effect parameters for each of $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$, and for the 12 site effect parameters which constitute the only variation that is allowed for the dependence parameter α_j. This gives a total of 60 samples. The chain was allowed to run for 500 000 iterations. Serial correlation in the samples is removed by thinning to every 100th observation, and the resulting trace plots (which are not shown) demonstrate good mixing and apparent good convergence. Convergence was checked by multiple runs of the chain from various starting-points, demonstrating multiple convergence to the same limit for all the random-effect parameters. Table 1 summarizes the results in terms of the posterior means and standard deviations of the random-effects parameters for Bradfield in January. Also included are the equivalent results for Nottingham in July, presented to illustrate the contrasting site and season characteristics.

For the GPD and logistic dependence parameters, Table 2 provides a comparison of the posterior means and standard deviations from the hierarchical model with maximum likelihood estimates and standard errors obtained from a maximum likelihood analysis applied separately to each site and season. This illustrates that one apparent advantage of the hierarchical model over a standard likelihood analysis is a reduction in sampling variation: Table 2 shows that for the GPD parameters $\tilde{\sigma}_{m,j}$ and $\xi_{m,j}$, for the locations and seasons that were chosen, the posterior standard deviations are substantially smaller than the asymptotic standard errors of the corresponding maximum likelihood estimates. This effect is repeated over all sites and seasons;

Table 1. Posterior means (and standard deviations in parentheses) of some site and seasonal effects

Effects		*Mean for Bradfield, January*	*Mean for Nottingham, July*
Seasonal	$\gamma_{\tilde{\sigma}}^{(m)}$	1.891 (0.042)	1.294 (0.042)
	$\gamma_{\xi}^{(m)}$	0.021 (0.018)	0.002 (0.018)
Site	$\varepsilon_{\tilde{\sigma}}^{(j)}$	0.367 (0.044)	−0.121 (0.041)
	$\varepsilon_{\xi}^{(j)}$	−0.105 (0.020)	−0.059 (0.017)
	$\varepsilon_{\alpha}^{(j)}$	0.385 (0.009)	0.300 (0.011)

Table 2. Posterior means (and standard deviations) with corresponding maximum likelihood estimates (and asymptotic standard errors) of the GPD parameters and the logistic dependence parameter

Parameter	*Results for Bradfield, January*		*Results for Nottingham, July*	
	Mean (standard deviation)	*Maximum likelihood estimate (asymptotic standard error)*	*Mean (standard deviation)*	*Maximum likelihood estimate (asymptotic standard error)*
$\tilde{\sigma}_{m,j}$	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)

hence it appears that the pooling of information is seen to add considerably to the precision of the analysis. It is worth noting that the same effect is present to a lesser extent for the logistic dependence parameters α_j, where the pooling of information occurs across seasons only, without there being any link across locations.

Fig. 3, which shows the posterior means from the hierarchical model plotted against the maximum likelihood estimates obtained separately from each site, and, in the case of $\tilde{\sigma}$ and ξ, each season, is a further illustration of the possible gains that are achieved from the pooling of information through the hierarchical model. The line of equality is plotted on each graph, enabling an easy assessment of the degree to which the variability in estimates is reduced in the hierarchical model. In our case, this reduction in variability, which is sometimes termed *shrinkage* (e.g. Coles (2002)) is very apparent for the GPD shape parameter ξ, but less so for the scale parameter $\tilde{\sigma}$ and the logistic dependence parameter α.

One cautionary point to observe here is that all of the comparisons that were outlined above relate to *Bayesian inference* applied to the hierarchical model, *versus classical inference* applied to the separate sites and seasons model. Hence any gain in performance that is attributed to the hierarchical model is confounded with possible differences due to the inference paradigm that is employed. Given the non-informative priors that are used, we would argue that any such differences are not likely to be responsible for a major part of the contrasts in methods observed, but further work here may shed more light.

3.2. Diagnostics

The performance of the hierarchical model in terms of its fit to the data can be assessed via probability plots and quantile–quantile plots (plots of the fitted distribution function *versus* the empirical distribution function, and of fitted quantile *versus* empirical quantile). By using the fitted parameter values (here we use posterior means) to transform each monthly set of excesses at each site to a common margin (say unit exponential), the fit of the GPD (1) to excesses in all seasons can be assessed simultaneously. Fig. 4 shows probability plots and quantile–quantile plots for Bradfield and Nottingham. The very good fit that is demonstrated here is repeated across the other 10 sites and is sustained when separate months are analysed individually, supporting the overall adequacy of the model, and demonstrating some robustness to any deviations from the modelling assumptions. Though not used here, methods for checking model goodness of fit specific to complex Bayesian models have been developed by, for example, Gelfand and Trevisani (2002).

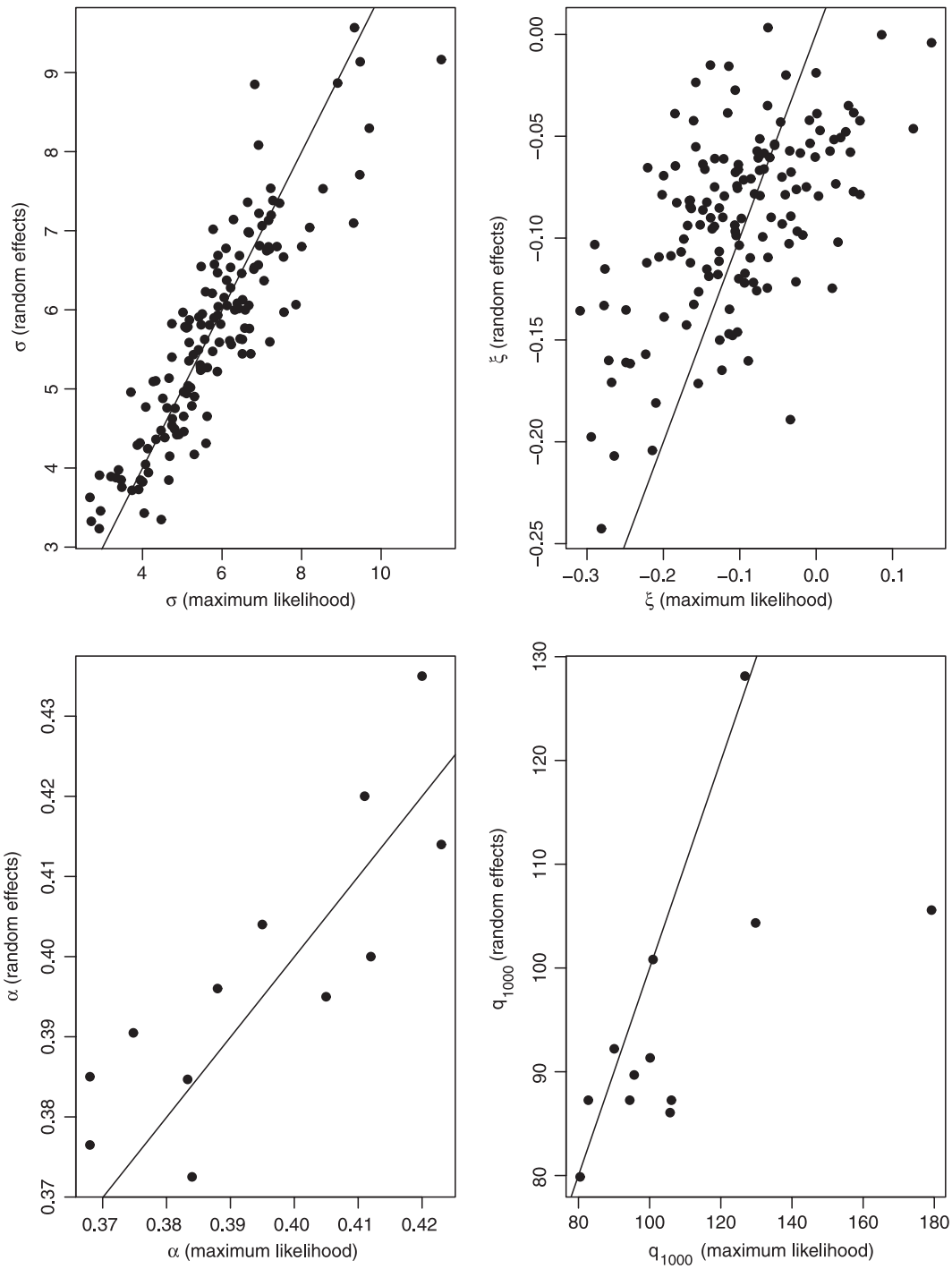

Fig. 3. Plots of posterior means against the maximum likelihood estimates of GPD parameters, the logistic dependence and the 1000-year return level

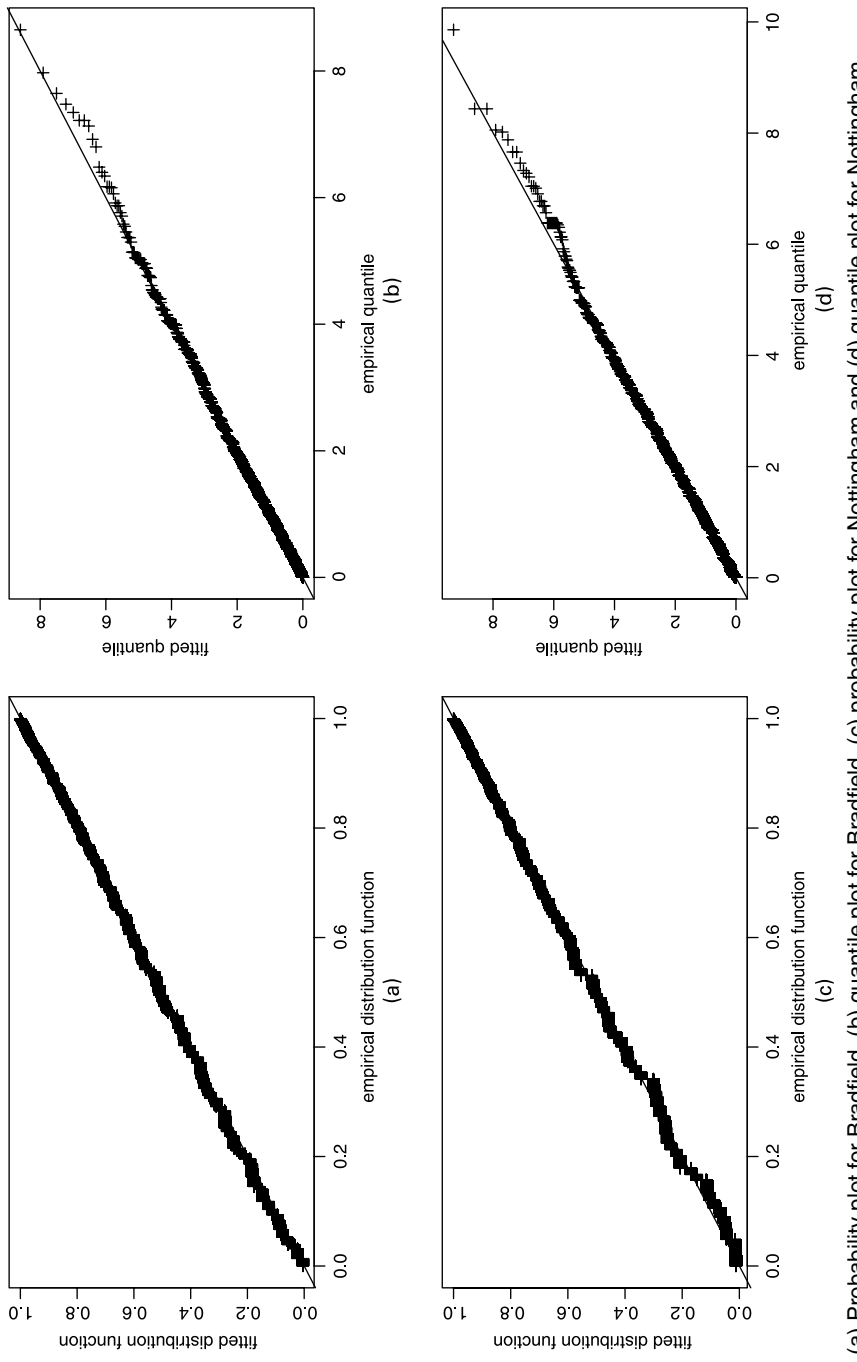

Fig. 4. (a) Probability plot for Bradfield, (b) quantile plot for Bradfield, (c) probability plot for Nottingham and (d) quantile plot for Nottingham

Table 3. Posterior means (and standard deviations in parentheses) of some seasonal and site effects based on the conditional autoregressive model

Effects		*Mean for Bradfield, January*	*Mean for Nottingham, July*
Seasonal	$\gamma_{\tilde{\sigma}}^{(m)}$	1.888 (0.031)	1.286 (0.032)
	$\gamma_{\xi}^{(m)}$	0.022 (0.011)	0.004 (0.014)
Site	$\varepsilon_{\tilde{\sigma}}^{(j)}$	0.371 (0.043)	−0.120 (0.039)
	$\varepsilon_{\xi}^{(j)}$	−0.106 (0.019)	−0.058 (0.017)
	$\varepsilon_{\alpha}^{(j)}$	0.376 (0.010)	0.302 (0.009)

We now investigate our *a priori* assumption of exchangeable month and site effects. Although this assumption may seem simplistic, our experience (Walshaw, 1991, 1994) suggests that attempts to link the month effects are not liable to yield much change in terms of inference, and the spatial structure is even more difficult to exploit, especially with so few sites. Nevertheless, we investigate whether the same applies here by adopting a conditional autoregressive model (see, for example, Sun *et al*. (1999)) for the seasonal effects, which incorporates the most obvious aspect of the temporal structure, namely the fact that we would expect adjacent months to be more similar than non-adjacent months.

We adopt the following (conditional) prior distributions for the seasonal effects for $\log(\tilde{\sigma}_{m,j})$ and $\xi_{m,j}$, which replace equations (8) and (9):

$$\gamma_{\cdot}^{(m)} | (\gamma_{\cdot}^{(m-1)}, \gamma_{\cdot}^{(m+1)}) \sim N_0\{\tfrac{1}{2}(\gamma_{\cdot}^{(m-1)} + \gamma_{\cdot}^{(m+1)}), \tau_{\cdot}\}, \tag{20}$$

where once again $\gamma_{\cdot}^{(m)}$, for example, can denote either $\gamma_{\tilde{\sigma}}^{(m)}$ or $\gamma_{\xi}^{(m)}$. The cyclical nature of the monthly sequence is taken care of here by replacing the values 0 and 13 for $m-1$ and $m+1$ respectively by the values 12 and 1, ensuring that the months 12 and 1 are correctly identified as being adjacent in our model.

This model conditions each month effect on those of the months on either side of it. It could be argued that this is a more reasonable assumption than that of full exchangeability under the original model. However, the inferences are not altered greatly. Table 3 shows the results for the seasonal and site effects, and is directly comparable with Table 1. There are some improvements in the precision of estimation, illustrated by the reduced posterior standard deviations. However, although the additional structure may prove to be useful for other problems, here we feel that the changes in inference are not sufficient to justify the additional complexity of the model, and we retain the original model for making further inferences on return levels. It is worth noting here that the inferences for the GPD parameters themselves, and the return levels, are similarly insensitive to the choice between these two models.

3.3. Return levels and predictive inference

In a practical setting, we are often more interested in estimating extreme quantiles, expressed as *return levels*, than the actual GPD parameters themselves. Let $X_1, X_2, \ldots, X_n$ be the first n observations from a non-independent stationary sequence with marginal distribution function F. Standard arguments in Leadbetter *et al*. (1983), chapter 3, show that, for large n and x, typically

$$\Pr\{\max(X_1, X_2, \ldots, X_n) \leqslant x\} \approx F(x)^{n\theta}, \tag{21}$$

where $\theta \in (0, 1]$ is the extremal index and is a measure of the degree of extremal dependence in the series. It can be interpreted as the reciprocal of the mean cluster size; for a more detailed discussion of the extremal index, see, for example, Coles (2001). Setting $x = q_r$ in approximation (21), equating this to $1 - r^{-1}$ and solving for q_r, gives, to a good approximation, the r-year *return level* of the process, or the value which is exceeded (on average) once every r years. In the context of extreme wind speeds, the 50-year return level is used by the British Standards Institution (1997) to produce contour maps displaying the strength requirements for buildings and other large structures so that this level of wind speed can be withstood. Thus the accuracy and precision of return level estimation is an important design consideration.

For each site j, $j = 1, \ldots, 12$, the annual exceedance rate of q_r is given by

$$\sum_{m=1}^{12} \{1 - F_{m,j}(q_r)^{h_{m,j}\theta_j}\}, \qquad m = 1, \ldots, 12, \tag{22}$$

where $1 - F_{m,j}(q_r)^{h_{m,j}\theta_j}$ is the annual exceedance rate of q_r in month m (obtained from approximation (21)), $F_{m,j}$ is the GPD function in month m with parameters $\tilde{\sigma}_{m,j}$ and $\xi_{m,j}$ (from model (1)) and $h_{m,j}$ is the number of hours in month m. The extremal index θ_j in expression (22) ensures that we incorporate the temporal dependence of the series into the estimation of q_r and is implicitly defined through the value of the logistic dependence parameter α_j at site j. Thus, for each posterior draw for α_j we obtain a corresponding posterior draw for θ_j through simulation; for a more detailed discussion about how this is done, see Fawcett (2005). Though far from automatic, and potentially very computationally expensive, methods due to Smith (1992) could also be employed to *calculate* the extremal index θ_j given a particular value of logistic dependence α_j. Expression (22) can then be set equal to r^{-1} and solved for q_r for each posterior draw of $\tilde{\sigma}_{m,j}$, $\xi_{m,j}$ and θ_j to obtain posterior draws for the r-year return level q_r.

The objective of most extreme value analyses is to obtain an estimate of the probability that future events will reach extreme levels. Conveniently, *prediction* is neatly handled within the Bayesian setting. We have seen that estimation of the GPD and logistic dependence parameters $\boldsymbol{\psi} = (\tilde{\sigma}_{m,j}, \xi_{m,j}, \alpha_j)$ can be made on the basis of previous hourly observations $\mathbf{x} = (x_1, \ldots, x_n)$ by using a Markov chain model, resulting in a posterior distribution $\pi(\boldsymbol{\psi}|\mathbf{x})$. Through the Bayesian methodology, we then have

$$\Pr(X \leqslant x|x_1, \ldots, x_n) = \int_{\boldsymbol{\psi}} \Pr(X \leqslant x|\boldsymbol{\psi})\, \pi(\boldsymbol{\psi}|\mathbf{x})\, \mathrm{d}\boldsymbol{\psi} \tag{23}$$

(where X is a randomly chosen threshold exceedance), which gives the distribution of future large values of the process, allowing both for parameter uncertainty and the randomness in future observations when we condition on any particular values of the parameters. Solving

$$\Pr(X \leqslant q_{r,\mathrm{pred}}|x_1, \ldots, x_n) = 1 - r^{-1} \tag{24}$$

for $q_{r,\mathrm{pred}}$ therefore gives an analogue of the r-year return level which additionally incorporates uncertainty due to model estimation. Although equation (23) is analytically intractable, it is easily approximated through the estimated posterior distribution (which is found by simulation). After deletion of burn-in, we have a sample $\boldsymbol{\psi}^{[1]}, \ldots, \boldsymbol{\psi}^{[B]}$ that may be regarded as observations from the stationary distribution $\pi(\boldsymbol{\psi}|\mathbf{x})$. Thus,

$$\Pr(X \leqslant q_{r,\mathrm{pred}}|x_1, \ldots, x_n) \approx \frac{1}{B} \sum_{k=1}^{B} \Pr(X \leqslant q_{r,\mathrm{pred}}|\boldsymbol{\psi}^{[k]}), \tag{25}$$

Table 4. Return levels for Bradfield and Nottingham†

Model	*Results (knots) for Bradfield and the following return periods (years):*				*Results (knots) for Nottingham and the following return periods (years):*			
	10	*50*	*200*	*1000*	*10*	*50*	*200*	*1000*
Hierarchical	96.887	103.463	112.518	128.128	66.533	73.532	79.515	86.307
	(0.982)	(1.333)	(2.023)	(2.691)	(0.949)	(1.455)	(2.024)	(2.742)
Maximum	96.745	103.236	108.152	113.306	71.362	83.056	94.010	107.718
likelihood	(2.864)	(5.930)	(8.786)	(12.219)	(2.565)	(4.890)	(11.239)	(14.564)
Predictive	104.392	113.089	119.957	127.338	68.389	75.504	81.655	88.782

†Posterior standard deviations, and asymptotic standard errors for the maximum likelihood analyses, are shown in parentheses.

and we can solve approximation (25) by using a standard numerical method. See Coles and Tawn (1996) for details of these ideas.

Table 4 shows the posterior means and standard deviations of the return levels q_r for return periods $r = 10$, 50, 200, 1000 years, for both Bradfield and Nottingham. Again, the posterior means and standard deviations for q_r are compared with maximum likelihood estimates and their associated standard errors. The main contrast here is the much greater degree of precision that is apparent from the hierarchical model, although the *caveat* that was mentioned at the end of Section 3.1, relating to the two different inference paradigms, is also relevant here. Fig. 3 also shows a 'shrinkage' plot for the 1000-year return level, where one possible benefit of the hierarchical model is revealed: the results of maximum likelihood estimation are very unstable for such extreme quantiles and can provide unrealistic estimates, whereas the hierarchical model achieves a degree of stability through the pooling of information across sites, enabling sensible estimates to be obtained at all locations.

Predictive return levels for the periods 10, 50, 200 and 1000 years are also shown in Table 4 for both Bradfield and Nottingham, and a plot of the predictive distribution of a future exceedance is shown in Fig. 5 with the x-axis transformed to the scale indicated, which is commonly chosen as being convenient to show a range of return levels.

4. Discussion

The hierarchical model for the UK hourly wind speed data has been employed here as a natural framework for modelling the variation that is inherent in the data under consideration. The underlying structure of the data, in terms of seasonality and spatial variation, has been estimated through the model, allowing the sharing of information between sites and seasons. This provides apparent benefits in terms of a greatly improved precision in estimation of the model parameters. Estimative inferences for the quantities that are of most interest to engineers, return levels, have also improved in precision as a consequence.

There are also several reasons for preferring a Bayesian analysis of extreme data over the likelihood approach. First, although we have used non-informative priors, the facility to extend this model to incorporate other sources of information through prior distributions is there and has obvious appeal, since extreme data are, by their very nature, scarce. Second, the Bayesian framework has enabled a natural expression of predictive quantiles, which incorporate both the randomness of the process and model uncertainty, to produce a single value for design specification.

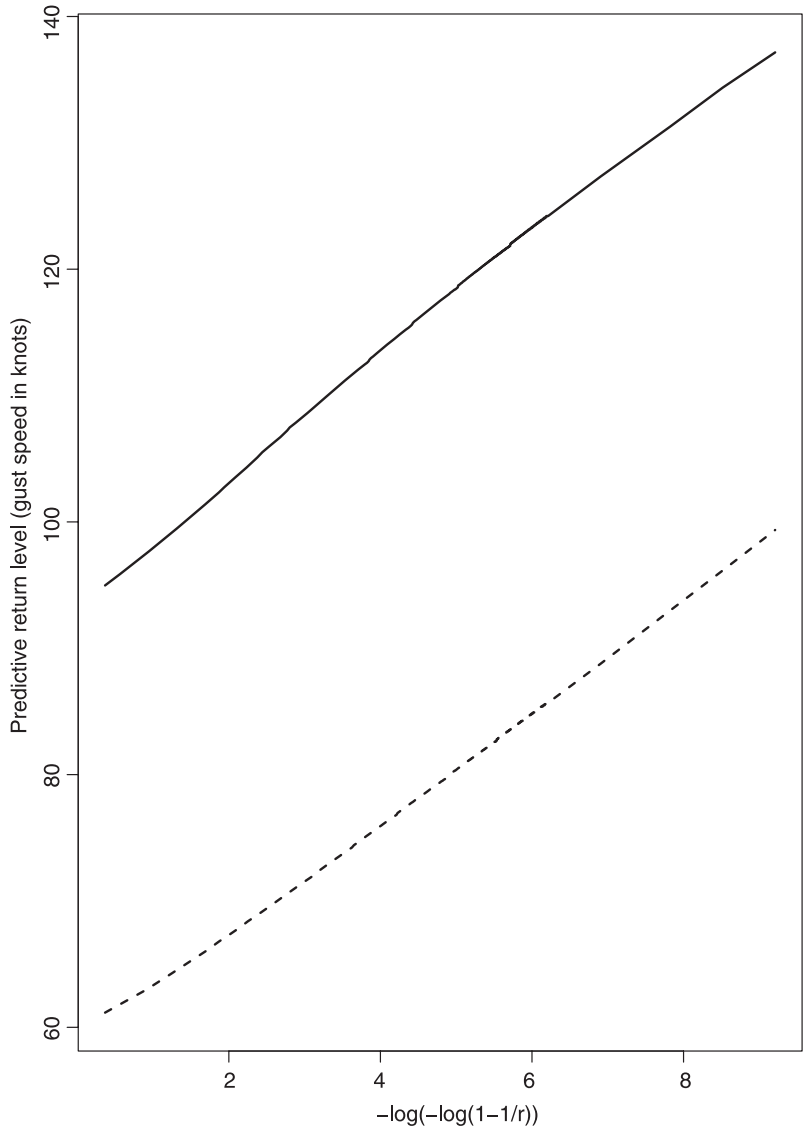

Fig. 5. Predictive return level curves for Bradfield (———) and Nottingham (-------)

The value of information in extreme value analyses is especially high, and the improvements in efficiency that are found here through appropriate structural modelling are likely to have benefits in many practical applications, and the model framework developed is transferable to other environmental variables for which measurements are available over a spatial network of locations.

Acknowledgements

We thank the UK Meteorological Office for supplying the data that are analysed in this paper, and Peter Smithson for his help with the data from Sheffield and Bradfield. Lee Fawcett's work

was largely carried out under an Engineering and Physical Sciences Research Council studentship. Thanks are due to Jonathan Tawn, Stuart Coles, John Matthews, Richard Boys and Darren Wilkinson for helpful comments. We are also indebted to two referees for suggesting considerable improvements.

References

British Standards Institution (1997) *Code of Basic Data for the Design of Buildings; CP 3*, chapter V, ' Loading', part 2, 'Wind loads'. London: British Standards Institution.
Coles, S. G. (2001) *Introduction to Statistical Modeling of Extreme Values*. New York: Springer.
Coles, S. G. (2002) A random effects model in extreme value analyses. *Preprint*. University of Padua, Padua.
Coles, S. G. and Powell, E. (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int. Statist. Rev.*, **64**, 119–136.
Coles, S. G. and Tawn, J. A. (1996) A Bayesian analysis of extreme rainfall data. *Appl. Statist.*, **45**, 463–478.
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. (2005) Statistical methodology for the estimation of environmental extremes. *PhD Thesis*. University of Newcastle, Newcastle upon Tyne.
Fawcett, L. and Walshaw, D. (2006a) Improved estimation for temporally clustered extremes. *Environmetrics*, to be published.
Fawcett, L. and Walshaw, D. (2006b) Markov chain models for extreme wind speeds. *Environmetrics*, to be published.
Gelfand, A. E. and Trevisani, M. (2002) Discussion on 'Bayesian measures of model complexity and fit' (by D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde). *J. R. Statist. Soc.* B, **64**, 631.
Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Series*. New York: Springer.
Shellard, H. C. (1976) Wind. In *The Climate of the British Isles* (eds T. J. Chandler and S. Gregory), pp. 39–73. London: Longman.
Smith, R. L. (1992) The extremal index for a Markov chain. *J. Appl. Probab.*, **29**, 37–45.
Smith, S. G. (1983) Seasonal variation of wind speed in the United Kingdom. *Weather*, **38**, 98–103.
Sun, D. C., Tsutakawa, R. K. and Speckman, P. L. (1999) Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika*, **86**, 341–350.
Tawn, J. A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
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 speed data: a step by step guide. *J. Res. Natn. Inst. Stand. Technol.*, **99**, 399–411.