Extremes 1:4, 449–468 (1999)

Spatial Regression Models for Extremes

EDWARD CASSON
Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, U.K.

STUART COLES
Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

[Received March 13, 1998; Revised January 28, 1999; Accepted January 29, 1999]

Abstract. Meteorological data are often recorded at a number of spatial locations. This gives rise to the possibility of pooling data through a spatial model to overcome some of the limitations imposed on an extreme value analysis by a lack of information. In this paper we develop a spatial model for extremes based on a standard representation for site-wise extremal behavior, combined with a spatial latent process for parameter variation over the region. A smooth, but possibly non-linear, spatial structure is an intrinsic feature of the model, and difficulties in computation are solved using Markov chain Monte Carlo inference. A simulation study is carried out to illustrate the potential gain in efficiency achieved by the spatial model. Finally, the model is applied to data generated from a climatological model in order to characterize the hurricane climate of the Gulf and Atlantic coasts of the United States.

Keywords. extreme values, hurricanes, Markov chain Monte Carlo, point process, spatial process

1. Introduction

A common application of extreme value theory is to the quantification and estimation of extreme levels of meteorological processes such as wind speeds, rainfall or sea-levels. The basis of the argument is well-known. The data, $X_1, \ldots, X_n$, representing hourly wind speeds say, are taken to be independent observations from a common distribution F, whose tail is characterized by an extreme value model. A suitable subset of the extreme values of the X_i series is extracted and used to fit this model, which in turn summarizes the tail behavior of F. Simple modifications at the modeling level are generally applied to handle temporal dependence or non-stationarity in the X_i process.

There are many examples of this type of analysis in the literature. In contrast, there are relatively few examples where the spatial characteristics of an extreme value process have been studied, despite the fact that meteorological processes are generally recorded on a spatial scale. A notable exception is in the hydrological literature where the issue of pooling spatial data, whilst allowing for spatial dependence and non-homogeneity, has been given high prominence (N.E.R.C. (1975), Hosking & Wallis (1988), Buishand (1989), Dales & Reed (1989), for example). Explicit spatial regression models for extremes have also been developed (Coles & Tawn (1991, 1996), for example), but only for instances where linear models for the extreme value parameters in terms of specified covariates proved sufficient. In this paper we propose a methodology that is applicable

when there is a spatial cohesion to the variation in extremal behavior, but without the requirement of simple linear forms. The main advantage in spatial modeling is the pooling of information—a key strategy in extreme value analysis to overcome difficulties caused by the sparsity of data—but there are others as well. First, the model can be used for interpolation to sites where little or no data may have been collected. Secondly, the methodology is formally Bayesian, enabling inferences and predictions to incorporate uncertainties in process variation and parameter estimates.

Our methodology is based on recent developments in non-linear spatial statistics. The model structure is simple: at each location, conditional on model parameters, a standard extreme value model is assumed to characterize the data-generating process at extreme levels. Our extra assumption is that across locations the unobservable model parameters are themselves a realization of a smooth stochastic spatial process. Thus, nearby locations are more likely than distant sites to have extreme value data with similar characteristics, to an extent determined by the strength of spatial dependence in the underlying spatial process. Estimating the realized but unobserved parameter process associated with the observed data is the analogue of identifying the regression model in a more classical analysis, without the restriction of simple linear forms.

Inference for this model is considerably less straightforward. The model parameters constitute a latent spatial process, the uncertainty of which must be accounted for in any analysis. Conventional procedures for inference, such as maximum likelihood, are intractable in these circumstances. However, recent advances in sample-based inference, most notably Markov chain Monte Carlo techniques, have been shown to work well for inferring spatial latent process models in other applications (Diggle et al. (1998), for example). The scope of this work is to examine the performance of the latent process model structure for describing spatial variation in the extremal behavior of a process.

The paper is structured as follows. In Section 2 we review a point process characterization for univariate extremal behavior. In our spatial construction, this serves as a conditional model for extremes given a realization of the latent spatial process of extreme value parameters. The complete model specification, together with details of the Markov chain Monte Carlo scheme used for inference, is given in Section 3. Our research into this area was motivated by a requirement to model the extremal characteristics of the hurricane process along the eastern and southeastern United States coastlines. We describe this application in Section 5. Preceding the application, in Section 4, we undertake a simulation study to investigate the viability of our model for inferring spatially coherent extreme value processes and to assess its utility compared with a classical site-by-site analysis.

2. Tail characterization

The archetypal extreme value problem can be expressed in the following way: given independent data $X_1, X_2, \ldots, X_n$, that are, say, hourly observations from a process with marginal distribution F, what inferences can be made about the tail of F? Various characterizations have been proposed, most of which are special cases of a point process

characterization developed by Pickands (1971); see also Leadbetter et al. (1983). For applications this can be interpreted in the following way: the 2-dimensional point process

$$P_n = \left\{ \left(\frac{j}{n+1}, X_j \right), j = 1, \ldots, n \right\}.$$

restricted to the region $A_u = (0,1) \times (u, \infty)$ for large u, approximates a Poisson process with intensity measure in the parametric family

$$\Lambda((a,b) \times (x, \infty)) = (b-a)m \left[1 + \frac{\xi}{\sigma}(x-\mu) \right]_+^{-\frac{1}{\xi}}, \tag{1}$$

where $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ and $\xi \in (-\infty, \infty)$, with the case $\xi = 0$ obtained by taking the limit as $\xi \to 0$. The coefficient m is an arbitrary scaling that determines the time-scale of the parameters; for applications it is convenient to let $m = n/n_y$ where n_y is the number of observations in a year. Thus, m is the number of years of observations. Then, by standard properties of a Poisson process and (1), the annual maximum $M_{n_y} = \max(X_1, \ldots, X_{n_y})$ has distribution function

$$\begin{aligned} \Pr\{M_{n_y} \leq x\} &= \exp\{-\Lambda((0, n_y/n) \times (x, \infty))\} \\ &= \exp\left[-\left\{ 1 + \frac{\xi}{\sigma}(x-\mu) \right\}_+^{-\frac{1}{\xi}} \right]. \end{aligned} \tag{2}$$

This is the distribution function of the generalized extreme value distribution, $\mathcal{G}(\mu, \sigma, \xi)$; thus the connection between the point process characterization and the classical characterization of extremal behavior is immediate. In particular, the parameters (μ, σ, ξ) in the Poisson intensity function (1) are, respectively, the location, scale and shape parameters of the associated generalized extreme value distribution of the annual maximum. In the usual way, it follows that the n–year return level of the annual maxima process is given by

$$q_n = \mu + \frac{\sigma}{\xi} \left[1 - \{ -\log(1 - 1/n) \}^{-\xi} \right]. \tag{3}$$

In a similar way, the standard models for threshold exceedances are also derived immediately from the point process representation. Considering a point (T_i, X_i) of the process, for which $X_i > u$,

$$\begin{aligned} \Pr(X_i > x | X_i > u) &= \frac{\left[1 + \frac{\xi}{\sigma}(x-\mu) \right]_+^{-1/\xi}}{\left[1 + \frac{\xi}{\sigma}(u-\mu) \right]_+^{-1/\xi}} \\ &= \left[1 + \frac{\xi(x-u)}{\sigma + \xi(u-\mu)} \right]_+^{-1/\xi}, \end{aligned} \tag{4}$$

which is the usual representation of the generalized Pareto distribution function on reparameterizing $\{\sigma + \xi(u - \mu)\} \to \sigma$.

Smith (1989) was the first to advocate using the point process characterization as a modeling tool; other applications include Smith (1994), Coles (1994) and Coles & Tawn (1996). The advantage of working with this representation relative to methods based on the annual maxima distribution is that the behavior of all extreme values, in the sense of having exceeded a high threshold, is characterized. This objective is also achieved with more familar threshold techniques, such as the modeling of threshold exceedances by the generalized Pareto distribution (4). However, there are considerable modeling advantages to using the point process representation explicitly, which stem largely from the invariance of the parameters to the threshold choice u, and the fact that the threshold exceedance rate is not marginalized out of the analysis as in the standard threshold models.

As discussed, adopting a likelihood analysis based on (1), all such extreme values contribute to the inference as follows. Letting u be a high threshold and $\mathscr{E} = \{(t_i, x_{(i)}); i = 1, \ldots, N_{A_u}\}$ be the set of points for which $x_{(i)} > u$, the likelihood function is derived as

$$L_{A_u}(\mu, \sigma, \xi; \mathscr{E}) = \exp\{-\Lambda(A_u)\} \prod_{i=1}^{N_{A_u}} d\Lambda(t_i, x_{(i)})$$

$$= \exp\left\{-m\left[1 + \frac{\xi}{\sigma}(u - \mu)\right]^{-\frac{1}{\xi}}\right\} \prod_{i=1}^{N_{A_u}} \left\{\frac{1}{\sigma}\left[1 + \frac{\xi}{\sigma}(x_{(i)} - \mu)\right]^{-\frac{1}{\xi}-1}\right\}. \qquad (5)$$

Equation (5) can be maximized numerically to obtain the maximum likelihood estimate of (μ, σ, ξ) and hence return levels by substitution into (3).

3. Spatial models

3.1. Model structure

The point process model in Section 2 provides a convenient characterization and inferential basis for the tail behavior of a spatial process at each point location. Furthermore, the method can be applied independently to data at a finite set of locations and the estimates examined to obtain an informal impression of the degree of spatial variability or homogeneity in the extreme data (Coles (1994), for example). To identify formally any spatial structure, or to improve inferences by exploiting such relationships, requires the specification of a spatial model. One approach is to postulate simple linear relationships for the parameters (μ, σ, ξ) in terms of spatial location Coles & Tawn (1990, 1996). Ignoring spatial independence, the form of such relationships can be investigated and optimized within a likelihood analysis. A generic procedure for modifying such inferences to account for dependencies in data is described by Liang & Self (1996). This

amounts to maximum likelihood under the assumption of independence, followed by corrections to standard asymptotic formulae for parameter variances and covariances to account for observed dependence in the data. A similar procedure in the specific context of modeling spatially dependent extremes had been suggested earlier by Smith (1991). Often, however, the nature of the spatial variation is too complex to be easily described a priori by simple linear forms, so that other techniques are required.

As an alternative, we propose a model based on a latent spatial process to describe smooth variation in extreme value parameters over the region of interest; conditional on these parameters, threshold exceedances at each location are assumed to follow the limiting Poisson process with intensity given by (1). This type of model formulation is commonplace within the discipline of spatial statistics, but novel within extreme value modeling. Thus, we assume the parameters $(\mu(z), \sigma(z), \xi(z))$, as defined in Section 2 but indexed now by spatial location z, vary smoothly over space according to a stochastic spatial process. Explicitly, we model $\mu(z)$ as

$$h_\mu(\mu(z)) = f_\mu\left(z; \boldsymbol{\beta}_\mu\right) + \mathbf{S}_\mu\left(z; \boldsymbol{\alpha}_\mu\right), \tag{6}$$

with similar expressions for the parameters σ and ξ. In this specification h_μ is a known link function, f_μ is a regression function with unknown parameters, $\boldsymbol{\beta}_\mu$, and $\mathbf{S}_\mu$ is a zero-mean, stationary, spatial stochastic process with unknown parameters $\boldsymbol{\alpha}_\mu$. Moreover, we assume that $\mathbf{S}_\mu, \mathbf{S}_\sigma$ and $\mathbf{S}_\xi$ are Gaussian and mutually independent, though this latter assumption could easily be relaxed. Now, data are observed at a finite number of locations $z_1, \ldots, z_k$ and, with a slight abuse of notation, we denote the extreme value parameters at location z_j by $(\mu_j, \sigma_j, \xi_j) = \left(\mu(z_j), \sigma(z_j), \xi(z_j)\right)$. Similarly, $\mathscr{E}_j$ denotes the set of exceedances of a high threshold u_j at location z_j. Then, conditional on the realized values of (μ_j, σ_j, ξ_j), $\mathscr{E}_j$ is assumed to follow the Poisson process approximation of Section 2, denoted by:

$$\mathscr{E}_j \left| \left(\mu_j, \sigma_j, \xi_j\right) \sim \mathscr{P}\left(\mu_j, \sigma_j, \xi_j\right), \right. \tag{7}$$

independently for each location $z_1, \ldots, z_k$.

The point of this formulation is that spatial variation in the extreme value parameters, over and above that which can easily be modelled through the regression functions f_μ,f_σ and f_ξ, is absorbed into the respective stochastic components $\mathbf{S}_\mu, \mathbf{S}_\sigma$ and $\mathbf{S}_\xi$. There are, however, some limitations imposed by this model construction. First, the asymptotic point process characterization is now assumed only conditionally. After marginalization over the uncertainty in the latent process parameters (μ_z, σ_z, ξ_z), the point process structure is perturbed and, for example, the distribution of the annual maximum is no longer generalized extreme value. The second restriction is the spatial independence of the conditional processes (7). This implies that having allowed for spatial variation in the underlying extreme value parameters, observations from location to location are independent. This is a serious limitation of the model. In practice, the spatial dimension of meteorological events generally induces a spatial dependence beyond that which can be

explained by the spatial structure in (μ_z, σ_z, ξ_z). Further research is required to generalize the model structure to accommodate spatial dependence in the observations conditional on the latent parameter process. However, for the application in Section 5, in which the spatial characteristics at extreme levels of a simulated database of hurricane wind speeds are required, the present model structure is adequate.

3.2. Inference

The motivation for our proposed model is that it provides a convenient representation for non-linear spatial variations in the extreme value parameters. However, its complexity renders impractical standard estimation procedures such as maximum likelihood, so the utility of the representation depends on the availability of an alternative inference procedure. Conveniently, the latent process framework is especially amenable to analysis by Markov chain Monte Carlo methodology. This requires a Bayesian specification of the model, giving, at least, formal priors on each of the parameters. Then, a Markov chain is designed for which the equilibrium distribution is the objective posterior distribution, $\pi(\cdot)$. Simulation from the chain then leads to a series which, after convergence, can be interrogated to estimate $\pi(\cdot)$ and, consequently, summary features of the model parameters. This is now a standard procedure in statistical modeling; see Smith & Roberts (1993) for a general overview, and Coles & Powell (1996) for applications in extreme value modeling.

The most general Markov chain Monte Carlo formulation is based on the Metropolis–Hastings algorithm. This requires the specification of a proposal transition density $q(t' \mid t)$ from which, given a current value of the series $t_i = t$, a proposal value t' for t_{i+1} can be simulated. To obtain π as the equilibrium distibution of the simulated chain a rejection step is included:

$$t_{i+1} = \begin{cases} t' & \text{with probability} \quad \Delta(t_i, t') \\ t_i & \text{with probability} \quad 1 - \Delta(t_i, t') \end{cases}$$

where

$$\Delta(t, t') = \min\left\{1, \frac{\pi(t')q(t|t')}{\pi(t)q(t'|t)}\right\}. \tag{8}$$

Under mild conditions on q, which, in particular, ensure the chain is recurrent, the sequence $\{t_i\}$ is bound to converge to the target posterior distribution π. In slightly greater generality, if $\mathbf{t}$ is a vector of components, $\mathbf{t} = (t^{(1)}, \ldots, t^{(d)})$, then the above algorithm can be applied in cycles to the t_j by simply replacing $\pi(\boldsymbol{t})$ in (8) with the conditional density $\pi_j(t_j|\mathbf{t}_{-j})$, where $\mathbf{t}_{-j} = (t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_d)$ and iterating over j; this is the single-component Metropolis–Hastings algorithm.

In principle, subject to the regularity conditions, the transition density q, or densities q_j

in the case of the single-component Metropolis–Hastings algorithm, can be chosen arbitrarily. In practice, choices must be made that facilitate easy simulation and lead to reasonable convergence properties of the chain. Computations can also be simplified by exploiting conditional independencies within the model structure. The structure of our model is most clearly seen from Figures 1(a) and 1(b) which give, respectively, the directional acyclic graph and conditional independence graph of the model. By adopting uniform and independent priors, denoted generically by $p(\cdot)$, for each of the parameters we obtain an analysis that is closest in spirit to a conventional likelihood analysis, though there would be little additional complexity in adopting a fully Bayesian approach by adding extra nodes to the model to serve as hyperparameters for any subset of the regression or correlation parameters. Combining the specification of uniform priors with the model format identified in Figures 1(a) and 1(b) leads to the following algorithm for recursive simulation of parameter values.

Given a current parameter value $\boldsymbol{\theta}_i = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu, \boldsymbol{\alpha}_\sigma, \boldsymbol{\beta}_\sigma, \boldsymbol{\alpha}_\xi, \boldsymbol{\beta}_\xi)$, the next point in the chain, $\boldsymbol{\theta}_{i+1}$, is generated in the following way.

Step 1: *Set* $q(\boldsymbol{\alpha}'_\mu|\boldsymbol{\alpha}_\mu) \propto 1$ *on* $[\boldsymbol{\alpha}_\mu - \boldsymbol{\varepsilon}_\alpha, \boldsymbol{\alpha}_\mu + \boldsymbol{\varepsilon}_\alpha]$, *a random walk process. Then*

$$\begin{aligned}\Delta(\boldsymbol{\alpha}_\mu, \boldsymbol{\alpha}'_\mu) &= \min\left\{1, \frac{f(\boldsymbol{\alpha}'_\mu|\ldots)q(\boldsymbol{\alpha}_\mu|\boldsymbol{\alpha}'_\mu)}{f(\boldsymbol{\alpha}_\mu|\ldots)q(\boldsymbol{\alpha}'_\mu|\boldsymbol{\alpha}_\mu)}\right\} \\ &= \min\left\{1, \frac{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}'_\mu, \boldsymbol{\beta}_\mu\right)p(\boldsymbol{\alpha}'_\mu)}{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)p(\boldsymbol{\alpha}_\mu)}\right\} \\ &= \min\left\{1, \frac{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}'_\mu, \boldsymbol{\beta}_\mu\right)}{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)}\right\},\end{aligned}$$

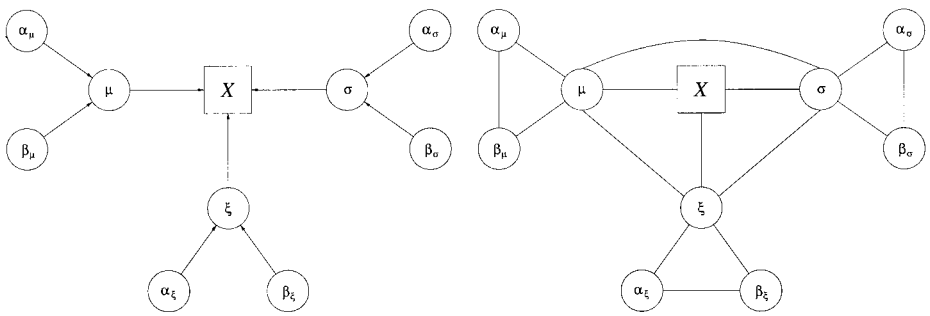

Figure 1. The directional acyclic graph and conditional independence graph of the model used to describe the non-linear spatial variation of the GEV parameters $\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\xi}$.

a ratio of multivariate Normal densities. With probability $\Delta(\boldsymbol{\alpha}_\mu, \boldsymbol{\alpha}'_\mu)$, *the* $\boldsymbol{\alpha}_\mu$ *component of* $\boldsymbol{\theta}_{i+1}$ *is set at* $\boldsymbol{\alpha}'_\mu$; *otherwise it remains at* $\boldsymbol{\alpha}_\mu$.

Step 2: *Set* $q\left(\boldsymbol{\beta}'_\mu|\boldsymbol{\beta}_\mu\right) \propto 1$ *on* $\left[\boldsymbol{\beta}_\mu - \boldsymbol{\varepsilon}_\beta, \boldsymbol{\beta}_\mu + \boldsymbol{\varepsilon}_\beta\right]$, *a random walk process. Then*

$$\begin{aligned}\Delta\left(\boldsymbol{\beta}_\mu, \boldsymbol{\beta}'_\mu\right) &= \min\left\{1, \frac{f\left(\boldsymbol{\beta}'_\mu|\ldots\right)q\left(\boldsymbol{\beta}_\mu|\boldsymbol{\beta}'_\mu\right)}{f\left(\boldsymbol{\beta}_\mu|\ldots\right)q\left(\boldsymbol{\beta}'_\mu|\boldsymbol{\beta}_\mu\right)}\right\} \\ &= \min\left\{1, \frac{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}'_\mu\right)p\left(\boldsymbol{\beta}'_\mu\right)}{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)p\left(\boldsymbol{\beta}_\mu\right)}\right\} \\ &= \min\left\{1, \frac{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}'_\mu\right)}{f\left(\boldsymbol{\mu}|\boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)}\right\},\end{aligned}$$

again, a ratio of multivariate Normal densities. With probability $\Delta\left(\boldsymbol{\beta}_\mu, \boldsymbol{\beta}'_\mu\right)$, *the* $\boldsymbol{\beta}_\mu$ *component of* $\boldsymbol{\theta}_{i+1}$ *is set at* $\boldsymbol{\beta}'_\mu$; *otherwise it remains at* $\boldsymbol{\beta}_\mu$.

Step 3: *The components of* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ *are updated singly according to a transition density* $q_j(\mu'_j|\mu_j)$, *to be specified below.*

$$\begin{aligned}\Delta\left(\mu_j, \mu'_j\right) &= \min\left\{1, \frac{f\left(\mu'_j|\ldots\right)q\left(\mu_j|\mu'_j\right)}{f\left(\mu_j|\ldots\right)q\left(\mu'_j|\mu_j\right)}\right\} \\ &= \min\left\{1, \frac{f\left(\mathscr{E}_j|\mu'_j, \sigma_j, \xi_j\right)f\left(\mu'_j|\boldsymbol{\mu}_{-j}, \boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)q_j\left(\mu_j|\mu'_j\right)}{f\left(\mathscr{E}_j|\mu_j, \sigma_j, \xi_j\right)f\left(\mu_j|\boldsymbol{\mu}_{-j}, \boldsymbol{\alpha}_\mu, \boldsymbol{\beta}_\mu\right)q_j\left(\mu'_j|\mu_j\right)}\right\}.\end{aligned}$$

Now set

$$q_j(\mu'_j|\mu_j) = f(\mu'_j|\boldsymbol{\mu}_{-j}, \boldsymbol{\alpha}, \boldsymbol{\beta}_\mu), \tag{9}$$

corresponding to the simulation of μ'_j *from the univariate conditional distribution of one component of a* k*-dimensional multivariate Normal in accordance with standard*

multivariate Normal theory (Theorem 3.2.4 of Mardia (1972), for example). In this case we obtain the simplification,

$$\Delta(\mu_j, \mu_j') = \min\left\{1, \frac{f\left(\mathscr{E}_j|\mu_j', \sigma_j, \xi_j\right)}{f\left(\mathscr{E}|\mu_j, \sigma_j, \xi_j\right)}\right\}, \tag{10}$$

a likelihood ratio based on the exceedances at z_j only, with likelihoods of the form (5).

Step 4: *Repeat steps 1–3 interchanging, respectively,* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\xi}$ *with* $\boldsymbol{\mu}$*, so that all components of* $\boldsymbol{\theta}_i$ *are updated in one complete iteration.*

4. Application to simulated data

In Section 5 we apply the algorithm of Section 3 to model data which, notionally, represent 999 hurricane wind speed measurements at each of 55 equally spaced locations along the eastern and southeastern coastlines of the United States. Our approach there is to linearize the coastline so that the data locations $\{z_1, \ldots, z_{55}\}$ can be represented by the points $\{1, \ldots, 55\}$. on the real line. In this section we mimic the application by simulating data according to the model

$$X_{i,j} \sim \mathscr{G}(\mu_j, \sigma_j, \xi_j); \qquad i = 1, \ldots, 999;\ j = 1, \ldots, 55, \tag{11}$$

where the parameter sets (μ_j, σ_j, ξ_j) are generated from models having the form of (6) for each parameter. The choice of the $\mathscr{G}(\mu_j, \sigma_j, \xi_j)$ distribution for the marginal distribution at each site, conditional on parameter values, guarantees fast convergence to the limiting Poisson process. Moreover, nominally setting $m = n$ in equation (1), ensures that the extreme value parameters of $\mathscr{E}_j$ are also (μ_j, σ_j, ξ_j).

Complete determination of the model requires, for each parameter, a specification of the link function, the regression equation and the latent process parameters. We look at a range of model formulations, though in each case we set:

(i) $\mu_j = \alpha + \gamma j + \mathbf{S}_\mu(j)$

(ii) $\log(\sigma_j) = \delta + \mathbf{S}_\sigma(j)$,

where $\mathbf{S}_\mu(j)$ and $\mathbf{S}_\sigma(j)$ for $j = 1, \ldots, 55$ are independent point realizations of a zero mean Gaussian process with respective variances a_μ^2 and a_σ^2, and correlation functions of the form

$$\rho(i,j) = \exp\{-(bd(i,j))^c\}, \tag{13}$$

where $d(\cdot, \cdot)$ is a metric and (a, b, c) are specified parameters. Thus, the model is linear in distance for $\boldsymbol{\mu}$ with spatial noise determined by $\mathbf{S}_\mu$, but admits no spatial regression other than noise on a logarithmic scale for $\boldsymbol{\sigma}$. The parameter values $(\alpha, \gamma, \delta) = (10, 0.3, 2.45)$ were chosen for reasonable consistency with the data to be introduced in Section 5.

For the correlation functions of both $\mathbf{S}_\mu$ and $\mathbf{S}_\sigma$ we set $c = 1.8$, thus guaranteeing continuous sample paths. A range of possible values for both the variance and correlation parameters of $\mathbf{S}_\mu$ and $\mathbf{S}_\sigma$ were considered, including combinations of high(H), moderate(M) and low(L) values of the variance and correlation of both $\mathbf{S}_\mu$ and $\mathbf{S}_\sigma$. In each case, the corresponding values of a_μ, a_σ, b_μ and b_σ are summarized in Table 1.

A number of different formulations for the shape parameter $\boldsymbol{\xi}$ were also considered; results are presented separately for each case in the subsequent sections.

Having generated data according to the latent spatial process model, our interest focuses on the efficiency of the inferential scheme described in Section 3 for parameter and return level estimation, both point-wise and globally. To assess the gain in efficiency by accounting for the spatial structure we compare results obtained against those of a separate-site likelihood analysis. In all cases the spatial analysis was carried out using 100,000 iterations of the Metropolis–Hastings algorithm described in Section 3, of which the first 20,000 terms were discarded. After this point, at which both informal and formal assessments of chain convergence seemed satisfactory, every 10th point of the series was retained and the resulting series interrogated to estimate the posterior distribution. The separate-site likelihood analysis was carried out by maximizing (5) based on site-specific data, but sometimes constrained to ensure a constant estimate of $\boldsymbol{\xi}$ across sites.

Comparison of the spatial and non-spatial analyses is complicated by the large number of parameters involved. Since most extreme value analyses are aimed at the estimation of extreme return levels, we focus mainly on global measures of the accuracy of return level estimation. Two such measures are:

$$Q_1 = \sum_j \frac{(\hat{q}_j - q_j)^2}{q_j} \text{ and } Q_2 = \max |\hat{q}_j - q_j|$$

where, conditional on parameter values, q_j is a specified annual maximum quantile at location j (see equation (3)), and $\hat{q}_j$ is its estimate. To compare the spatial analysis with that

Table 1. Parameter settings in simulation example.

	Variance			Correlation		
	L	M	H	L	M	H
		a^2			b	
μ	10	40	60	36	18	10
σ	0.01	0.03	0.06	36	18	10

of a maximum likelihood analysis we calculate the ratio of the corresponding values of Q_1 and Q_2 obtained under the respective inference procedures. Thus, we define

$$E_1 = \frac{Q_1(\text{mle})}{Q_1(\text{spatial})} \text{ and } E_2 = \frac{Q_2(\text{mle})}{Q_2(\text{spatial})}. \tag{14}$$

4.1. *Homogeneous* $\xi = -0.2$

Spatial homogeneity in the shape parameter is a phenomenon that has often been observed in analyses of environmental data (Dales & Read (1989); Coles & Tawn (1996), for example). In this example we fix $\xi = -0.2$, corresponding to bounded but reasonably long-tailed marginal distributions, and assume ξ to be constant but unknown in the modelling process. A summary of the comparison between maximum likelihood estimation and the spatial model, as determined by the measures E_1 and E_2 based on three different return periods, is given in Table 2. To make the comparison fair, the maximum likelihood estimates were also constrained so that ξ was held constant across all locations; without this constraint the spatial model overwhelmingly outperforms maximum likelihood.

The actual values in Table 2 should be treated with some caution as each value is based on a single simulated spatial dataset. Consequently, the values have some degree of sampling variability. This is presumably why there is no uniformity in the efficiency values as either the spatial variance or correlation increases. Nonetheless, taking the results as a whole, some patterns emerge. The spatial model generally outperforms maximum likelihood in both efficiency measures, and increasingly so at high return periods. This conclusion holds for almost all parameter combinations but the precise effect of both the scale and correlation of spatial variation is unclear. Unfortunately, the computing time required to undertake this analysis precludes a more systematic study.

Table 2. Efficiency comparisons in simulated example with $\xi = -0.2$. Efficiencies are computed for $p_1 = 0.01$, $p_2 = 0.0005$ and $p_3 = 0.0001$ corresponding to the 100-, 2000- and 10,000-year return level estimates respectively.

		E_1			E_2		
Var	Cor	p_1	p_2	p_3	p_1	p_2	p_3
L	M	1.17	1.59	1.68	1.51	1.83	2.09
M	L	1.21	1.59	1.54	1.54	2.06	2.02
M	M	1.30	1.55	2.11	1.20	1.43	1.94
M	H	0.82	0.96	1.10	1.06	1.03	0.97
H	M	1.32	1.96	2.10	1.05	1.31	1.49

The example with the moderate values of correlation and scale parameters is also illustrated in Figure 2. This shows the tendency for the spatial estimates of both parameters and return levels to follow the spatial variation in the simulated values with less erratic variation than the corresponding site-by-site maximum likelihood estimates.

4.2. *Homogeneous* $\xi = 0.2$

We now repeat the analysis of Section 4.1 but with $\xi = 0.2$, corresponding to marginal distributions with unbounded upper tails. In this case the results are summarized in Table 3.

Despite the change in tail form, the conclusions are much the same as for the bounded tail case. The spatial method is almost uniformly more efficient than maximum likelihood for estimation at all return levels, though the precise dependence of the efficiency

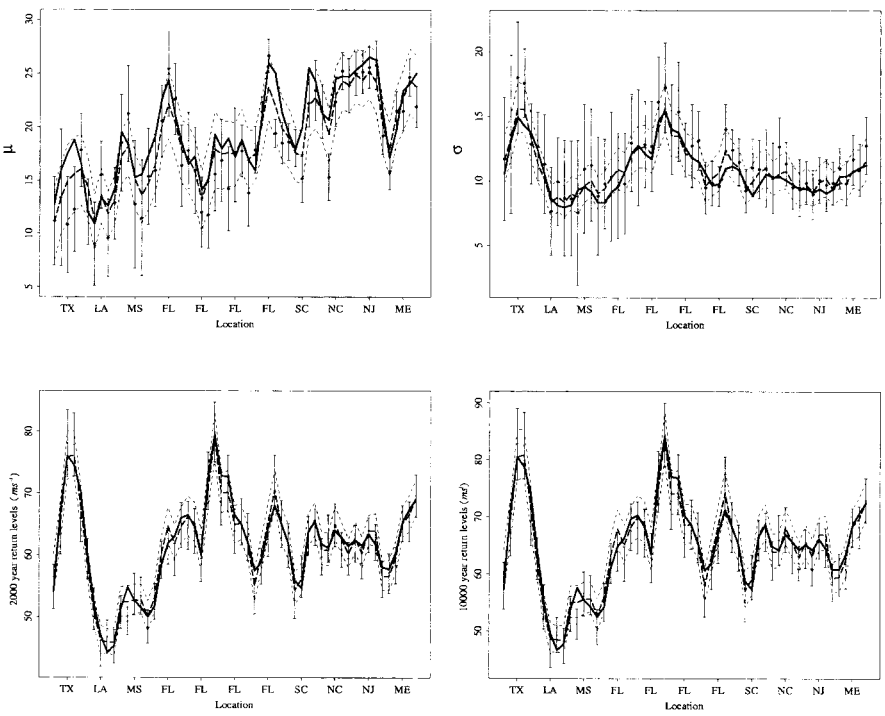

Figure 2. Spatial behavior, with 95% confidence intervals for simulated GEV parameters (a) $\boldsymbol{\mu}$, (b) $\boldsymbol{\alpha}$ and (c, d) n-year return levels, for $n = 2000$, 10,000 years, estimated by non-linear models (dotted lines) and maximum likelihood (vertical bars) respectively.

Table 3. Efficiency comparisons in simulated example with $\xi = 0.2$. Efficiencies are computed for $p_1 = 0.01$, $p_2 = 0.0005$ and $p_3 = 0.0001$ to produce the 100-, 2000- and 10,000-year return level estimates respectively.

		E_1			E_2		
Var	Cor	p_1	p_2	p_3	p_1	p_2	p_3
L	M	1.42	1.32	1.39	1.88	2.06	2.29
M	L	1.61	1.01	1.03	1.54	1.43	1.36
M	M	1.18	1.66	1.89	1.17	1.25	1.40
M	H	1.10	0.93	0.87	2.98	1.34	0.86
H	M	0.87	1.06	1.13	1.89	1.76	1.74

improvement on the scale and correlation of the spatial variation is not apparent from such a limited study.

4.3. Non-homogeneous ξ

Our simulations so far have been based on models in which the shape parameter ξ is constant across locations, and this homogeneity assumption has been incorporated into both the maximum likelihood and spatial analyses. We now consider two further scenarios: first, the situation where ξ is genuinely homogeneous, but this constraint is not explicitly modelled; and second, the situation where ξ is also spatially varying, with spatial variation also of the form (6).

In the first of these cases, for both $\xi = -0.2$ and $\xi = 0.2$, the results are summarized in Table 4. In the spatial analysis we assume only a model of the form (6) for ξ, and make the comparison with maximum likelihood analyses where, respectively, homogeneity and non-homogeneity for ξ is assumed. Again, at all return levels, the results give convincing evidence for the superiority in efficiency of the spatial analysis, especially with respect to

Table 4. Efficiency comparisons in simulated example with constant ξ not assumed in spatial model. Efficiencies are computed for $p_1 = 0.01$, $p_2 = 0.0005$ and $p_3 = 0.0001$ to obtain the 100-, 2000- and 10,000-year return level estimates respectively. In the case where data are simulated with ξ varying $\xi(j) = -0.2 + \boldsymbol{S}_{\xi}$.

		E_1			E_2		
ξ	MLE model for ξ	p_1	p_2	p_3	p_1	p_2	p_3
-0.2	Homogeneous	1.28	1.44	1.40	1.45	1.63	1.60
	Non-Homogeneous	1.27	1.91	2.17	1.52	2.94	4.56
0.2	Homogeneous	1.49	2.03	1.79	2.36	2.87	2.48
	Non-Homogeneous	1.47	5.80	8.90	2.59	19.9	42.2
varying	Non-Homogeneous	1.01	1.28	1.53	1.22	1.37	3.85

the non-homogeneous maximum likelihood analysis, and most notably in the $\xi = 0.2$ case.

For the non-homogeneous ξ case we simulated data with $\xi(j) = -0.2 + \boldsymbol{S}_{\xi}(j)$, where $\mathbf{S}_{\xi}$ is a Gaussian process with correlation function (13) and parameters $(a_{\xi}, b_{\xi}, c_{\xi}) = (18, 0.1, 1.6)$. Over the 55 locations, this gave rise to values of ξ in the range $[-0.36, 0.03]$, so a comparison with the homogeneous $\xi = -0.2$ case is the most appropriate. The results, also shown in Table 4, show that the spatial analysis has improved efficiency in this situation also, although it is perhaps surprising that the gains are apparently no greater than in the homogeneous ξ case.

4.4. Extrapolation and interpolation

Whilst the extrapolation of tail estimates to rare events is the usual objective of an extreme value analysis, the spatial modelling approach adopted here also raises questions about aspects of spatial interpolation. We look at two issues: the first is the utility of the model when data are too sparse at any location to enable a conventional extreme value analysis; the second concerns the predictive capability of the model at locations for which no data are recorded.

4.4.1. Sparse data. Extreme value analyses invariably require a delicate balance between bias and variance—analyses based on low thresholds incur bias due to invalidity of asymptotic model arguments, while high thresholds lead to large variance because of the lack of data. A well-specified spatial model enables thresholds to be set at higher levels than would normally be possible because of the information transfer across sites. To examine this, we have repeated the simulation study with $\xi = -0.2$ and the moderate scale and correlation values, but using thresholds that yield just 5 exceedances at each location. Efficiencies are now calculated using (14), but where the values of Q_1 (mle) and Q_2 (mle) are those obtained from the earlier analysis based on lower thresholds—thus, we are comparing the efficiency of the spatial model relative to a separate-site analysis that has the advantage of many more data.

The simulated values of E_1 and E_2 were found to be (0.16, 0.32, 1.26) and (0.55, 0.74, 1.56) respectively, for $p = 0.01, 0.0005, 0.0001$. Therefore, at short return periods there is a substantial reduction in precision of return level estimates due to the loss of information. However, for long extrapolations, the spatial modeling compensates for the reduction of information to the extent that the efficiency of the spatial analysis exceeds that of the earlier likelihood analysis based on the lower threshold. This lends further support to the spatial modeling approach when the amount of data is very limited, provided that the assumptions about spatial smoothness in the parameter values are not unreasonable.

4.4.2. Prediction. Another advantage of the spatial model, in combination with the Monte Carlo inference, is the facility to obtain the posterior distribution of parameter values associated with any site in the region. This simply requires the application of step 3 of the Metropolis–Hastings algorithm for each of the parameters μ_j, σ_j and ξ_j at

any new location j. Adopting the same updating scheme as (9) for μ_j and similarly for σ_j and ξ_j, leads to acceptance probabilities equal to one in equation (10) since there are no data at this location.

In Figure 3 we plot kernel density estimates of parameters and return levels at site $j = 3$ where data were simulated with moderate correlation and scale parameters defined in Section 3. The plots compare the predictive distribution of parameter values and return levels at this location in analyses that respectively include and exclude the site's own data. In the latter case, estimates based on posterior modes are seen to be reasonably consistent with those of the fuller analysis, but the reduction of information manifests itself in a substantial increase in posterior variance. This opportunity to quantify prediction uncertainty in interpolation represents a substantial advantage of the methodology we have adopted in preference to more naive methods of parameter interpolation.

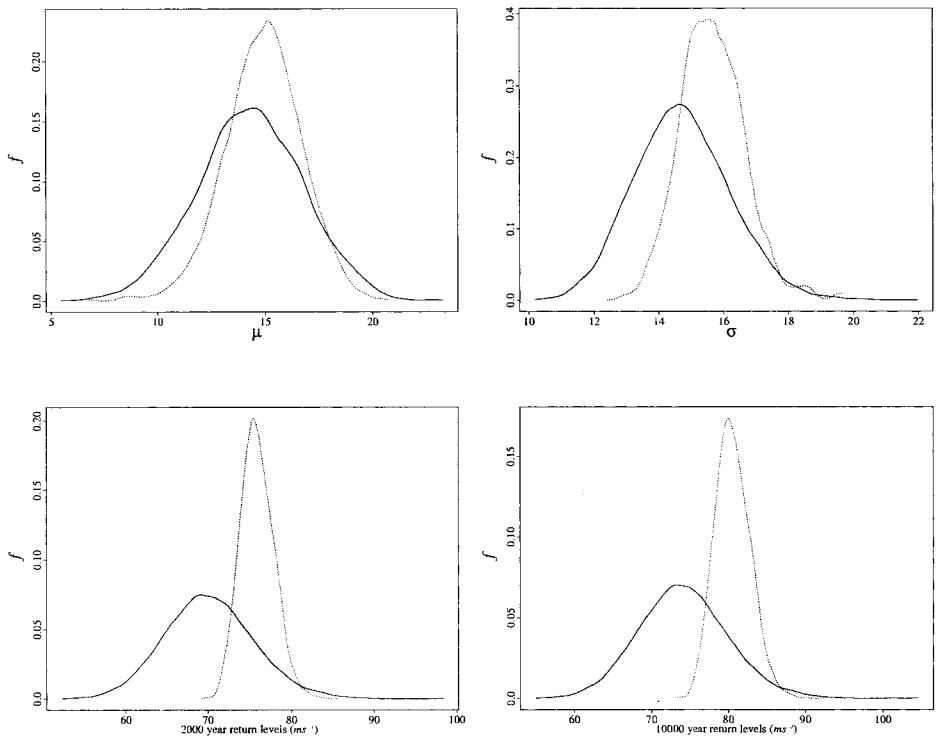

Figure 3. Kernel density ploys of the marginal posterior distributions for simulated GEV parameters (a) $\boldsymbol{\mu}$, (b) $\boldsymbol{\sigma}$ and (c, d) n-year return levels, for $n = 2000$, 10,000 years, at site 3, as modelled by non-linear spatial statistics applied to the 55 sites where data were available (dotted lines) and all sites except locations 3, 27 and 55 (solid curve).

4.5. *Summary*

Across all model configurations, our simulations point to a definite advantage, at least in terms of the efficiency measures E_1 and E_2, for the spatial model. It might be anticipated that the advantage is likely to be greatest when the strength of spatial dependence is greatest; unfortunately, the noise in our simulations prevent us from being able to confirm such a relationship. Moreover, computing time limits our capacity to undertake a more detailed simulation study. Nonetheless, taken as a whole, the results suggest there is sufficient statistical benefit to outweigh the extra computational demands required by the spatial analysis. Furthermore, the options to model with very sparse datasets and to interpolate to spatial locations without data represent substantial benefits of the spatial model. The cost of the analysis is the increased computing time and the assumption of an underlying smoothness in the spatial variation of the extreme value parameters.

5. Application to simulated hurricane data

Understanding the spatial climate of the hurricane process in regions that are prone to tropical storms is a major scientific and meteorological issue. The modelling of extreme wind speeds generated by hurricanes is, however, complicated by the difficulty of measuring wind speeds during such events. On the other hand, relatively accurate measurements of hurricane pressure fields can be made and historical records are fairly complete for the last hundred years or so. Consequently, an approach for modeling extreme hurricane wind speeds has developed based on the fitting of tail models to simulated hurricane wind speed data. Historical records of wind pressure fields are used for localized calibration of meteorological wind-field models. Climatological models are then used to convert the simulated wind-fields into a notional wind speed, and these simulated data then form the basis of an extreme value analysis. In an attempt to characterize hurricane behavior in the United States, Batts et al. (1980) applied this technique independently at each of 55 locations along the United States southern and eastern coastlines. In this section we examine the spatial coherency at extreme levels of these simulated data using the techniques of Section 3.

The data correspond notionally to the maximum wind speed for each of 999 independent hurricanes in the vicinity of each of 55 coastal locations. There are inevitably limitations to the analysis: the wind-field model used in the simulations is quite crude; the number of events simulated at each location is small; and the spatial dimension of hurricane events is ignored. Nonetheless, there is still scientific benefit in understanding how localized input variations to the climatological model lead to regional variations in output winds at extreme levels.

The analysis has the same form as the simulation study of Section 4. A comparison of spatial estimates and maximum likelihood estimates for each extreme value parameter, under the assumption of a constant shape parameter, ξ, is shown in Figure 4. Overall, there is a general decrease in $\boldsymbol{\mu}$ moving from Texas to Maine, a corresponding slight increase in $\boldsymbol{\sigma}$, while the constant shape parameter is estimated at around -0.2. As in Section 4, the

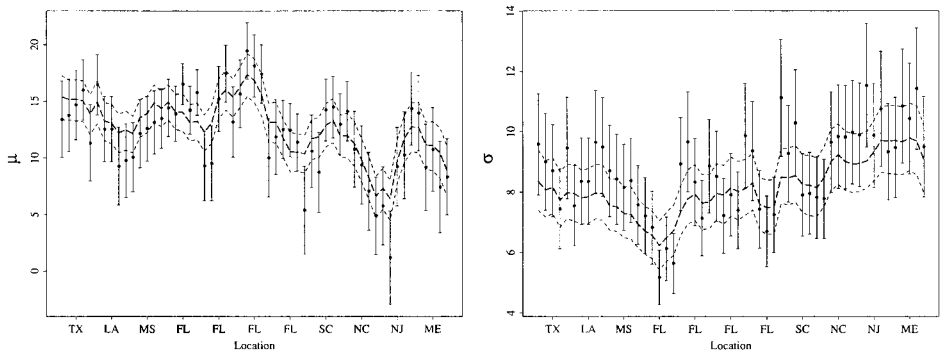

Figure 4. Spatial behavior of non-linear model for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ when fitted to simulated hurricane data (Batts et al., 1980) with maximum likelihood estimates, under the assumption that $\boldsymbol{\xi}$ is held constant at all locations compared with maximum likelihood estimates. Symbols follow the convention set in Figure 2.

spatial estimates are less variable and more precise than the maximum likelihood estimates.

Since the hurricane data are simulated, calculation of quantiles of, say, the annual maxima distribution additionally require an estimate of the yearly rate of occurrence of hurricanes, α. Batt et al. (1980) obtained such estimates for each location by extraction from empirical records—a plot of α against location is given in Figure 5(a). Combining the hurricane rate with the estimated GEV parameters leads to an estimate of the n-year return level, that is the level exceeded in any year with probability $1/n$, as

$$q_n = \mu + \frac{\sigma}{\xi}\left\{\left(\frac{\alpha n}{N}\right)^{\xi} - 1\right\}. \tag{15}$$

Estimates of the 50-, 100- and 2000-year return levels, for consistency with previous studies (e.g. Batts et al. (1980); Georgiou et al. (1983); Simiu et al. (1995)), are given in Figures 5(b)–(d). Again, the main feature of this plot is the substantial reduction in variability obtained in the spatial analysis relative to the site-by-site maximum likelihood analysis. The homogeneity of $\boldsymbol{\xi}$ at a level of around -0.2, corresponding to bounded tail estimates, is also consistent with current meteorological understanding Simiu et al. (1995).

6. Discussion

The idea of modelling parameter variation as a latent spatial process is familiar in spatial statistics. For most applications, having so much data at each location would diminish the advantage of a spatial model; for tail estimation, we have shown that spatial pooling of information can still be extremely advantageous.

We see the specific model proposed in this paper as something of a prototype. What we

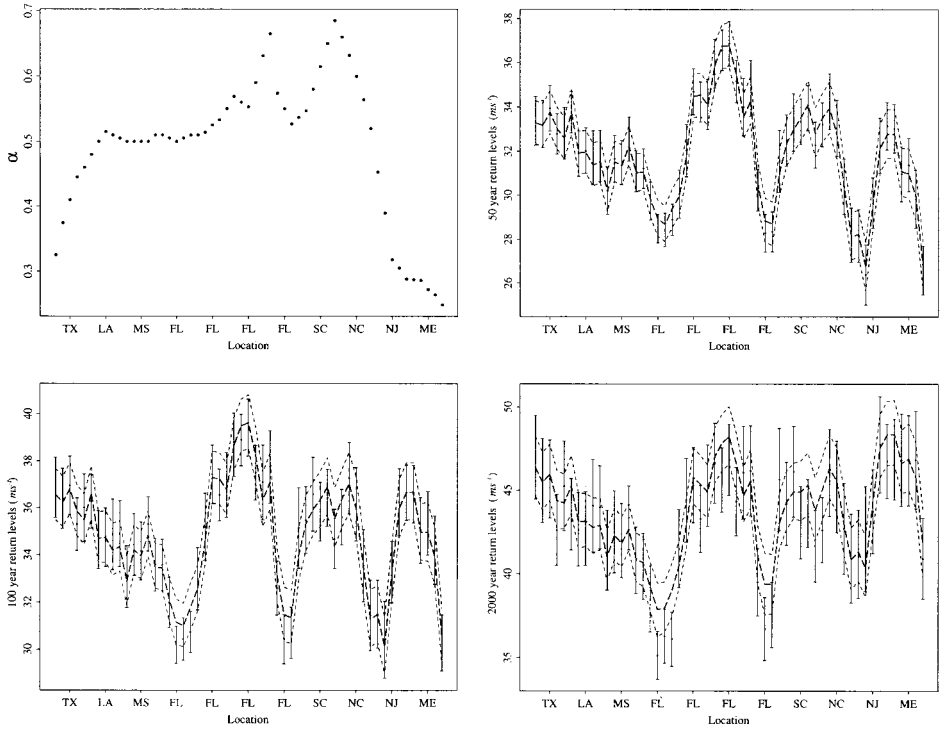

Figure 5. (a) Rate of hurricane occurrence parameter α.(b)–(d) Spatial behavior with 95% confidence intervals for wind speed data simulated by Batts et al. (1980) n-year return levels, for $n = 50$, 100, and 2000 years, estimated by non-linear model (dotted lines) and maximum likelihood (vertical bars) respectively.

hope to have shown is that fairly simple, though admittedly computer intensive, models and inference can lead to substantial gains in efficiency relative to a model which ignores spatial structure in the data. Any number of refinements could be made: other correlation functions could be considered—Kent's discussion of Diggle et al. (1998) suggests this may indeed be beneficial; correlated Gaussian processes for the different extreme value parameters could be used; greater attention could be given to the specification of prior distributions; the linearization of the coastline could be avoided; re-parameterization to speed up convergence of the Markov chain could be explored. In each case some slight gain in modelling flexibility or in inferential efficiency is likely to accrue.

The biggest limitation of our model is the assumption of conditional independence of data given the latent extreme value parameters. Most real-life examples would require a more detailed consideration of spatial dependence, and the development of a spatial regression model which can handle data that are spatially dependent after allowance for parameter variation remains an important research objective.

Appendix

The data used in Section 5 are available by FTP from `ftp.nist.gov` as follows:

```
username: anonymous
password: your e-mail address
cd /pub/bfrl/emil/datasets/hurricane
```

this directory consisting of a separate file for the simulated data corresponding to each location.

Acknowledgments

We are grateful to Emil Simiu of National Institute of Standards and Technology, Gaithersburg for many helpful discussions and for providing the simulated hurricane data. E. Casson was supported by an E.P.S.R.C. studentship. We thank also three referees and the Editor for comments on an earlier version of the paper.

References

Batts, M.E., Cordes, M.R., Russell, L.R., Shaver, J.R., and Simiu, E., ''Hurricane wind speeds in the United States,'' *Nat. Bureau of Standards Rep. No. BSS 124*, U.S. Dept. of Commerce, Washington, D.C., 1980.

Buishand, T.A., ''Statistics of extremes in climatology,'' *Statistica Neerlandica* 43(1), 1–30, (1989).

Coles, S.G., ''Directional modelling of extreme wind speeds,'' *Appl. Statist.* 43(1), 139–157, (1994).

Coles, S.G. and Powell, E.A., ''Bayesian Methods in Extreme Value Modelling: A Review and New Developments,'' *Int. Statist. Rev.* 64(1), 119–136, (1996).

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

Coles, S.G. and Tawn, J.A., ''Modelling extreme multivariate events,'' *J. R. Statist. Soc. B* 53, 377–392, (1991).

Coles, S.G. and Tawn, J.A., ''Modelling extremes of the areal rainfall process,'' *J. R. Statist. Soc. B* 58(2), 329–347, (1996).

Dales, M.Y. and Reed, D.W., ''Regional flood and storm hazard assessment,'' *Technical Report 102,* Institute of Hydrology, Wallingford, (1989).

Diggle, P.J., Tawn, J.A., and Moyeed, R.A., ''Model-based geostatistics,'' *Appl. Statist.* 47, 299–350, (1998).

Georgiou, P.N., Davenport, A.G., and Vickery, P.J., ''Design wind speeds in regions dominated by tropical cyclones,'' *J. Wind Eng. Ind. Aerod.* 13, 139–152, (1983).

Hosking, J.R.M. and Wallis, J.R., ''The effect of intersite dependence on regional flood frequency analysis,'' *Water Resour. Res.* 24, 588–600, (1988).

Leadbetter, M.R., Lindgren, G., and Rootzén, H., *Extremes and Related Properties of Random Sequences and Series*, Springer-Verlag, New York, 1983.

Liang, K.Y. and Self, S.G., ''On the asymptotic behavior of the pseudolikelihood ratio test statistic,'' *J. R. Statist. Soc. B* 58(4), 785–796, (1996).

Mardia, K.V., *Statistics of Directional Data*, Academic Press, London, 1972.

N.E.R.C., *Flood Studies Report, Vol. 1*, National Environment Research Council, London, 1975.

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

Simiu, E., Heckert, N.A., and Whalen, T., ''Estimates of hurricane wind speeds by the 'Peaks Over Threshold' method,'' *Technical Report 1416,* National Institute of Standards and Technology, Gaithersburg, (1995).

Smith, A.F.M. and Roberts, G.O., ''Bayesian computation via the Gibbs Sampler and related Markov Chain Monte Carlo Methods,'' *J. R. Statist. Soc. B* 55(1), 3–25, (1993).

Smith, R.L., ''Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone,'' *Statist. Sci.* 4(4), 367–393, (1989).

Smith, R.L., ''Regional estimation from spatially dependent data,'' *Technical Report* University of Chapel Hill, (1991).

Smith, R.L., ''Multivariate threshold models,'' in J. Galambos, J. Lechner, and E. Simiu (eds), *Extreme Value Theory and Applications*, Dordrecht: Kluwer, pp. 225–248, 1994.