Asymptotic models and inference for extremes of spatio-temporal data

**Kamil Feridun Turkman · M. A. Amaral Turkman ·
J. M. Pereira**

Abstract Recently there has been a lot of effort to model extremes of spatially dependent data. These efforts seem to be divided into two distinct groups: the study of max-stable processes, together with the development of statistical models within this framework; the use of more pragmatic, flexible models using Bayesian hierarchical models (BHM) and simulation based inference techniques. Each modeling strategy has its strong and weak points. While max-stable models capture the local behavior of spatial extremes correctly, hierarchical models based on the conditional independence assumption, lack the asymptotic arguments the max-stable models enjoy. On the other hand, they are very flexible in allowing the introduction of physical plausibility into the model. When the objective of the data analysis is to estimate return levels or kriging of extreme values in space, capturing the correct dependence structure between the extremes is crucial and max-stable processes are better suited for these purposes. However when the primary interest is to explain the sources of variation in extreme events Bayesian hierarchical modeling is a very flexible tool due to the ease with which random effects are incorporated in the model. In this paper we model a data set on Portuguese wildfires to show the flexibility of BHM in incorporating spatial dependencies acting at different resolutions.

K. F. Turkman (✉) · M. A. Amaral Turkman
CEAUL and DEIO, FCUL, University of Lisbon, Lisbon, Portugal
e-mail: kfturkman@fc.ul.pt

J. M. Pereira
ISA Technical University of Lisbon, Lisbon, Portugal

🖄 Springer

1 Introduction

Suppose we have observations $x(s_i, t)$, $s_i \in A \subset \mathcal{R}^2$, $t = 1, 2, ..., T$, $i = 1, ..., n_t$ from a non-stationary space-time process $X(s, t)$, continuous in space and discrete in time. Based on such a data set, how can we make inference on the extremal properties of $X(s, t)$? Specifically, we may want to estimate probabilities, such as

- $$P\left(\max_{s \in A} X(s, T+1) > u\right),$$

- $$P\left(\max_{1 \leq t \leq T} \max_{s \in A} X(s, t) > u\right),$$

- $$P\left(X\left(s_1^*, T+1\right) > u_1, ..., X\left(s_p^*, T+1\right) > u_p\right),$$

for some high thresholds $u_1,...,u_p$ and at any locations $s_1^*,...,s_p^*$.

Let us start with the simpler case of extremes of a spatial process $X(s)$, $s \in A \subset \mathcal{R}^2$. As in univariate and multivariate extreme value theory, it makes sense to base inference on asymptotic models. However, since in most applications A is limited, it may not make sense to look for asymptotics by expanding the space indefinitely, as in the case of temporal extremes. Hence, there are two alternatives:

1. Start with n iid replicates $X_i(s)$ and look at the convergence in distribution of $\max_{1 \leq i \leq n} X_i(s)$, as $n \to \infty$, for all $s \in A$ upon suitable normalization

$$Y_n(s) = a_n(s)^{-1}\left(\max_{1 \leq i \leq n} X_i(s) - b_n(s)\right) \to^{\mathcal{D}} Y(s). \tag{1}$$

2. Start with a sufficiently fine grid (lattice) over A, characterize the extremal properties of the process $X(s)$ over this fine grid, and obtain asymptotics by letting the grid sizes go to 0. This method is often called the double sum method of Pickands (see for example, Piterbarg 1996) and is used extensively to get limiting results for extremes of continuous time processes. See Leadbetter et al. (1983) and Albin (1987, 1990). See also Piterbarg (1996) for the extension of this method to obtain limiting results for Gaussian random fields.

The two approaches above can broadly be described respectively as constructive and descriptive studies of extremal characteristics. In constructive studies, the starting point is to give a good, parametric representation of the underlying process, possibly depending on a few parameters to permit statistical inference on the most important characteristics of the process. In descriptive studies, the starting point is a process of a fairly general type specified in terms of characteristics not directly linked to extremal properties. The object is to study what extremal properties such a process can be expected to have, and under what conditions. Although the results may be valuable in showing that for a wide class of processes under mild conditions the standard extreme value distributions are still relevant, they may not be very suitable for modeling purposes.

The approach in this paper will be in the constructive spirit. The structure of the paper is as follows: In Section 2, we look at the relationship between the max-stable models and Bayesian hierarchical models (BHM). In Section 3, we look at the hierarchical modeling strategies for extremal data. In Section 4, we look at a specific data set on Portuguese wildfires. Large fires tend to occur under specific, relatively uncommon synoptic meteorology conditions acting on large spatial scales and typically involving variable combinations of high temperatures, prolonged drought and strong winds. However, small scale spatial variations, such as land topography and fire prone vegetation distribution, are also very important factors on large fire regimes. One of the main objectives of this study is to understand the relative importance of these small scale and large scale variations in producing extreme fires. In order to achieve this objective, we suggest a BHM for fire size excesses.

2 Max-stable processes and hierarchical models

2.1 Max-stable processes

For a fixed s_0 or for a finite collection of locations $s_1, .., s_p$, the conditions for the convergence of (1), as well as the possible asymptotic models are well known and part of the univariate and multivariate extreme value theory. The convergence of the stochastic process in (1) is more complicated, but the results are also complete. See for example, de Haan and Ferreira (2006) or de Haan and Pereira (2006). Convergence is assumed to occur with non-degenerate margins and when (1) converges weakly to a stochastic process $Y(s)$, $s \in A$, (or any compact subset of an Euclidean space) then the limiting process $Y(s)$ is a max-stable process, having non-degenerate, univariate marginal GEV distributions. In order to obtain mathematically tractable results and representations, the process in (1) is transformed to have standard marginal distributions: If the functions $a_n(s)$, $b_n(s)$ are chosen in such a manner that for every s, the marginal limiting distribution have the form

$$\lim_{n \to \infty} P\big(Y_n(s) \le y\big) = \exp\left[-\left(1 + \frac{k(s)}{\sigma(s)}(y - \mu(s)) \right)^{-1/k(s)} \right],$$

then $Y(s)$ is transformed to

$$V(s) = \left(1 + \frac{k(s)}{\sigma(s)}(Y(s) - \mu(s))\right)^{1/k(s)}, \tag{2}$$

so that the limiting transformed max stable process has unit Frechet margins given by

$$P(V(s) \le v) = \exp\left(-v^{-1}\right).$$

Here, $k(s)$ is called the index (shape) function, $\sigma(s)$ and $\mu(s)$ are scale and location functions. In statistical applications, either all these functions are assumed to be known or, if estimated, a plug-in method is used so that their sampling variations are ignored in order to get tractable representations.

When the limit exists, the max-stable process $V(s)$ with unit Frechet margins has the following very useful representation:

$$V(s) =^{\mathcal{D}} \vee_{i=1}^{\infty} U_i W_i(s), \tag{3}$$

where U_i are the points of a point process defined on $(0, \infty)$ with mean measure $r^{-2}dr$ and $W_i(s)$ are i.i.d. replicates of an arbitrary positive spatial process $W(s)$, $s \in A$, satisfying some conditions to ensure that the marginal distribution of $V(s)$ for each s is a unit Frechet distribution. Hence, $W(s)$ is a process such that $E(W(s)) = ||A||$, for every $s \in A$ and $E(\sup_{s \in A} V(s)) < \infty$. Here the mean measure $r^{-2}dr$ corresponds to the exponent measure of the unit Frechet extreme value distribution, quantifying the size of the process, whereas, $W_i(s)$ can be seen as the independent and identical replicates of a spatially dependent latent random effect, quantifying the spatial dependence structure of the large values. Construction of max-stable processes is not restricted to unit Frechet margins. The process

$$\eta(s) = \vee_{i=1}^{\infty} \left(U_i + W_i(s) - \sigma^2(s)/2\right), \tag{4}$$

where $W_i(s)$ are independent copies of an intrinsically stationary Gaussian process with variance $\sigma^2(s)$ and variogram $\gamma(s)$, and U_i are the points of a Poisson point process on the real line with intensity $e^{-y}dy$, independent of $W(s)$, is a stationary max-stable process with standard Gumbel margins (Kabluchko et al. 2009). See also Fougères et al. (2009) for a class of max-stable processes which are constructed by mixing extreme value distributions over positive stable distributions.

Transformation (2) is crucial for the representation (3). To our knowledge, there is no mathematically tractable representation for max stable processes with non-unit Frechet margins. This is due to the fact that the point process U in the representation (3) has to be replaced by a non-homogenous point process on $R^2 \times (0, \infty)$ with mean

measure depending on the spatially varying index, scale and location functions. Necessary conditions on $W(s)$ to render a Frechet marginal distribution to $V(s)$ at every point $s \in A$ would then be intractable. Hence, in order to infer on the representation (3) as a model for extremes of non-stationary spatial data, the following steps are taken (see for example, Buishand et al. 2008),

1. Fit locally $k(s)$, $\sigma(s)$ and $\mu(s)$ using (temporal) block maxima data at each location s, assuming that such temporal data at each location exist.
2. Transform the data using (2), substituting $k(s)$, $\sigma(s)$ and $\mu(s)$ by the estimated values.
3. Assume a specific parametric model for $W(s)$. Buishand et al. (2008) suggest using an exponential martingale, having a single parameter which represents the degree of dependence of large values of the process at two spatial locations. Another alternative is to use a spatial Gaussian process (see Schlather 2002 for details).
4. Estimate the parameters of this latent process (see de Haan and Pereira 2006; Ribatet 2009).

The point process representation for max-stable processes given in (3) permits the use of all the relevant extreme observations in the inference, and therefore, it permits a data efficient inference. See Coles (2001) for the point process representation of extremes of time series and its relation to modeling threshold excesses. However, the representation (3) and the consequent statistical inference turns out to be quite difficult since the joint distributions of spatial extremes beyond bivariate distributions are mathematically intractable. In order to overcome this problem, Padoan et al. (2009) suggest a likelihood-based inference for max-stable processes, derived from a composite-likelihood approach. Moreover the introduction of explanatory variables and other underlying physical characteristics of the process that generates the data, although possible, is not easy.

Ideally, we would like to use more flexible models which overcome at least partially these difficulties, and at the same time are equally data efficient.

2.2 Bayesian hierarchical models

BHM and simulation based inference techniques are now accepted as the main principal tools for modeling spatial, non-stationary data. See for example Banerjee et al. (2004) for the general introduction to BHM. Typically, if we have block maxima data $y(s_i)$, $i = 1, ..., p$, where each $y(s_i)$ represents the maximum of observations at location s_i over a time interval, we may want to model $\{y(s_i), i = 1, ..., p\}$ by a Bayesian hierarchical model, assuming that conditional on the realization of a generic latent spatial process $W(s)$ and a set of explanatory variables chosen in accordance with the physical process governing the extremes, the data are independent and have a $GEV(k(s), \sigma(s), \mu(s))$ distribution. Here the spatially varying (random) model parameters depend on the latent process $W(s)$ as well as the explanatory variables

through properly chosen link functions. Once a proper parametric model for this latent process is chosen (often a spatial Gaussian process) and prior distributions are assigned to the model parameters and hyper-parameters, then simulation based inference techniques can be applied to make inference on the posterior distributions of the model parameters, including the latent process, as well as the on joint predictive distributions of $(Y(s_1^*), ..., Y(s_p^*))$ at any (unobserved) locations $(s_1^*, ..., s_p^*)$. Such models are very flexible and recently their use in modeling extreme data has been increasing. See for example, Fawcett and Walshaw (2006), Cooley et al. (2006, 2007), Mendes et al. (2008). Methodological study of strategies for Bayesian hierarchical models for block maxima data, can be found in Sang (2008). The principal starting point for using these modeling strategies is that we want to use as a model a spatial process whose univariate marginal distributions (conditional on the model parameters) are generalized extreme value distributions with space dependent parameters, having an arbitrary dependence structure. This dependence structure is then modeled by a latent spatial process through the model parameters. Typically, data, conditional on the realizations of this latent process as well as other explanatory variables, are assumed to be independent, but not identical, although more complicated Markov structures can also be assumed (Fawcett and Walshaw 2006). See also Sang (2008) for ways of introducing dependence structure in the likelihood using a Gaussian copula.

However, there are some questions regarding these modeling strategies. First, these models are not max-stable hence they will not be compatible with the asymptotic theory. Second, BHM models based on GEV distribution can be seen as inferior compared to the point process model (3), since not all the relevant data is used in the inference. A third objection which is often raised is the conditional independence. The main assumption behind conditional independence is that the dependence structure of the process is inherited from that of exploratory variables as well as spatially dependent latent factors through the model parameters. How reasonable this assumption is will depend on each data set, and careful judgement is needed to see if this assumption is compatible with the physical mechanism that generates the data. However, conditional independence is a very powerful tool with which physical characteristics of the underlying process can be introduced into the model.

We will argue that, although the Bayesian models explained above are not max-stable, they are mixtures of max-stable processes, due to spatially varying parameters and in that sense are compatible with the asymptotic theory. We will also suggest using Bayesian models with a GPD as a basis to overcome the second problem raised above.

Assume that $\{Y(s)\}$ is a process with non-unit GEV margins and that, conditional on a generic latent process $W(s)$, the components are independent. Then, for any spatial locations $s_1, s_2, ..., s_p$

$$P\big(Y(s_1) \le x_1, .., Y(s_p) \le x_p\big) = \int_{\boldsymbol{w}} P\big(Y(s_1) \le x_1, .., Y(s_p) \le x_p | \boldsymbol{W} = \boldsymbol{w}\big) d P_{\boldsymbol{W}}(\boldsymbol{w})$$

$$= \int_{\boldsymbol{w}} \prod_{j=1}^{p} P\big(Y(s_j) \le x_j | W(s_j) = w_j\big) d P_{\boldsymbol{W}}(\boldsymbol{w}). \quad (5)$$

Marginally, each $P(Y(s_j) \leq x_j | W(s_j) = w_j)$ are GEV and satisfy the max-stability condition, hence the product

$$\prod_{j=1}^{p} P(Y(s_j) \leq x_j | W(s_j) = w_j)$$

will also satisfy the max-stability condition. Therefore, although our (conditionally specified) model is max-stable, the unconditional process will be a mixture of max-stable processes.

We also note that, the model suggested by Buishand et al. ([2008](#)) is a max-stable process conditional on the values of the estimated parameters. Once the parameters are estimated, their model results in a conditionally specified model given by

$$V(s) | \hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\Theta}$$

where $\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s)$, are the model parameters, estimated marginally at each location and $\hat{\Theta}$ are the estimated parameters of the latent process $W(s)$ in ([2](#)). However, the unconditional distribution $F_{V(s)}(v)$ of $V(s)$, obtained by integrating out

$$V(s) | \hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\Theta}$$

with respect to the sampling variation of the estimators

$$\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\Theta}$$

need not have GEV margins and hence can not be max-stable, but is also a mixture of max-stable processes. This observation is important in understanding that, in a Bayesian context, it is difficult to construct max-stable processes in the sense that their predictive distributions satisfy conditions of max-stability. It would be highly useful to have classes of max-stable processes which are independent conditional on a latent process, having conditional marginal distributions which are extreme value. Such models would then benefit the inferential ease of hierarchical modeling, and at the same time capture correctly the dependence structure of spatial extremes. One such class of max-stable processes is given by Fougères et al. ([2009](#)), which are obtained by mixing extreme value distributions over positive stable distributions. However, further work is needed before using these processes as models for spatial extremes.

3 Bayesian hierarchical modeling strategies for spatial extremes

3.1 Hierarchical GEV model

Sang ([2008](#)) gives alternative strategies for implementing hierarchical GEV models for space-time data sets observed over a discrete grid: Let $Y(s_i, t)$ denote the block

maximum over some time period (say, one year) at some areal units or grid cells (or observation stations) s_i. Then, using the conditional independence assumption, the data can be modeled by

$$Y(s_i, t)|\mu_{s_i,t}, \sigma_{s_i,t}, k_{s_i,t} \sim GEV(\mu_{s_i,t}, \sigma_{s_i,t}, k_{s_i,t}).$$

Sang (2008) assumes a somewhat simplified dependence structure for the parameters:

- $k(s_i, t) = k$,
- $\sigma(s_i, t) = \sigma_i$ are spatial random effects,

so that the space-time dependence interactions are modeled only through the location parameter.

$$\mu(s_i, t)|\boldsymbol{\beta}, W(s_i, t), \tau^2 \sim \mathbf{x}_i^T \boldsymbol{\beta} + W(s_i, t) + \epsilon(s_i, t),$$

where $\mathbf{x}_i$ are site specific explanatory variables, $\boldsymbol{\beta}$ regression coefficients, $W(s_i, t)$ is a spatio-temporal random effect and $\epsilon(s_i, t)$ are iid random variables with variance $1/\tau^2$, representing the nugget effect.

There are many alternative ways of modeling the spatial, temporal variation in $W(s_i, t)$. One possibility is to assume the simple additive model

$$W(s_i, t) = \phi_i + \delta_t,$$
$$\delta_t = \psi \delta_{t-1} + w_t,$$
$$w_t \sim N\left(0, W_0^2\right).$$

Sang (2008) suggests modeling σ_i and ϕ_i jointly by using coregionalization CAR model:

- $\log \sigma_i = \sigma_0 + \lambda_i$,
-

$$(\lambda_i, \phi_i)^t = (\mathbf{A})(V_{1,i}, V_{2,i}),$$

- $(V_{1,i}, V_{2,i})$ are two independent CAR models,
-

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix}. \tag{6}$$

More complicated space-time structures for the scale and location parameters can be given (Sang 2008). However, it is doubtful that the extremal data will be rich enough to support such complicated structures. The choice of prior distributions are discussed in Sang (2008). The choice of these prior distributions is crucial. Choosing

non-informative priors for all the hyper-parameters may not work, resulting in non-convergence of the chain.

If

$$\boldsymbol{y}_{T+1} = \left(Y(s_1, T+1), ..., Y(s_p, T+1) \right)$$

represent the annual maxima of the process at the grid cells for the next time period, then they can be obtained by updating samples from the predictive density

$$p\left(\boldsymbol{y}_{T+1} | \text{Data}\right) = \int p\left(\boldsymbol{y}_{T+1} | \boldsymbol{\Theta}\right) p(\boldsymbol{\Theta} | \text{Data}) d\boldsymbol{\Theta},$$

where $\boldsymbol{\Theta}$ are all the model parameters. Implementation of these models, using Metropolis-Hastings algorithm is given in Sang (2008).

The conditional independence assumption, by its nature, produces non-smooth response surfaces. In order to remedy this situation and to relax the conditional independence assumption, Sang (2008) suggests using the following equation for the first stage of the hierarchy to obtain a smoother response surface:

$$Y(s, t) = \mu(s, t) + \frac{\sigma(s, t)}{k} \left(V(s, t)^k - 1 \right). \tag{7}$$

Here, the conditional independence assumption implies that $V(s, t)$ are i.i.d. unit Frechet random variables. Sang (2008) suggests introducing extra dependence structure in the likelihood through Gaussian copula by assuming that

$$V(s, t) = G^{-1} \left(\Phi\left(Z^*(s, t)\right) \right),$$

where G is the unit Frechet, Φ is the unit Normal distribution function and $Z^*(s, t)$ is a spatio-temporal Gaussian process. Here, $\mu(s, t)$, $\sigma(s, t)$, k are then modeled in the usual manner explained above. Note that $V(s, t)$ are standard Frechet random variables and their dependence structure is completely determined by the Gaussian copula

$$C_{\boldsymbol{v}}\left(u_1, u_2, .., u_p\right) = F_{\boldsymbol{v}} \left(\Phi^{-1}\left(u_1\right), ..., \Phi^{-1}\left(u_p\right) \right),$$

where $F(.)$ is the multivariate Gaussian distribution. This method of introducing dependence is computationally very convenient, as there are many efficient computational algorithms on spatial Gaussian processes. It may also be a convenient way of introducing dependence structure directly in the data, when the conditional independence assumption cannot be justified. However, this Gaussian copula based approach reduces the dependence structure between extreme values to second order moments and has little connection with the asymptotic properties of multivariate extreme value theory.

3.2 Hierarchical GPD model

The fundamental drawback of using the conditionally specified GEV model is that rarely there is available time-block maxima data. Often, we do not have sufficient time replicate data at each location to construct block maxima data for inference on the GEV model on each observation site. Also, BHM based on the GEV model are not data efficient in the sense that they do not use all the relevant extremal data coming from the tails. However, there is a very useful duality between modeling peaks over a high threshold and block maxima, equivalently between the generalized Pareto distribution (GPD) and the generalized extreme value distribution (GEV) in classical extreme value theory. This duality extends to spatial extremes, permitting more efficient use of spatial data, as Buishand et al. (2008) suggest: Consider the following GPD-process

$$V_1(s) = Y W(s), \tag{8}$$

where Y is a random variable with Pareto distribution function

$$P(Y \le y) = 1 - \frac{1}{y}, \ y > 1.$$

Writing this distribution in the form

$$P(Y \le y) = 1 - (1 + (y - 1))^{-1}, \ y > 1$$

we see that this is GPD distribution with $k = 1$, $\sigma = 1$ and and a location parameter $\mu = 1$. Buishand et al. (2008) call this random variable, the unit GPD random variable. Now, let $W(s)$ be the latent spatial process in the representation (3). $V_1(s)$ is in the domain of attraction of the unit max stable process given in (3) (Buishand et al. 2008), hence the asymptotic dependence structure of the max-stable processes $V(s)$ in (3) and $V_1(s)$ in (8) are identical and either process can be used for inference.

Note that the distribution of $V_1(s)$ conditional on $W(s) = w(s)$, is given by

$$P(V_1(s) \le v | W(s) = w(s)) = 1 - \left[1 + \frac{1}{w(s)}(v - w(s))\right]^{-1}, \ v > w(s),$$

a GPD with $k = 1$, $\sigma = w(s)$ and location parameter $\mu = w(s)$. Hence, $W_1(s)$ introduces dependence as well as spatial variability to $V_1(s)$ through the scale and the location (threshold) parameter of the GPD.

Note also that, for any s_1 and s_2 in A,

$$P(V_1(s_1) \le v_1, V_2(s_2) \le v_2 | W(s_1) = w_1, W(s_2) = w_2) = P(Y \le \min(v_1/w_1, v_2/w_2)),$$

collapsing to the total dependence case. The dependence structure of the bivariate random variable $(V_1(s_1), V_1(s_2))$ is then introduced through integration of this degenerate distribution with respect to the joint distribution of $W(s_1), W(s_2)$ by

$$P(V_1(s_1) \leq v_1, V_2(s_2) \leq v_2) = \int P(Y \leq \min(v_1/w_1, v_2/w_2))$$
$$\times dP_{W(s_1), W(s_2)}(w_1, w_2).$$

We would prefer not to work with the unit max-stable process $V(s)$ or the process $V_1(s)$ given in (8) which belong to the domain of attraction of this unit max-stable process, but to work on a conditionally independent, hierarchical GPD process as given below.

There will be many variations on how one can construct hierarchical models based on excesses as will be shown for a specific data set in Section 4, but the following hierarchical model can serve as a basis for other variations.

Suppose that $X(s)$ is a spatial process and $Z(s)$ is the excess of $X(s)$ over a sufficiently high but fixed threshold u, with

$$P(Z(s) \leq z) = 1 - \left(1 + \frac{k(s)}{\sigma(s)}z\right)^{-1/k(s)}, \qquad (9)$$

Assume that the shape parameter $k(s)$ in (9) is random but spatially static, whereas the scale parameter is a spatially varying random function, resulting in the following hierarchical model:

1.

$$p\left(z(s_1), ..., z(s_p)|k, \sigma(s_i), i=1, ..., p\right)$$
$$= \prod_{i=1}^{p} \frac{1}{\sigma(s_i)}\left(1 + \frac{k}{\sigma(s_i)}z(s_i)\right)^{-(1/k)-1}, \qquad (10)$$

2. Assume that we have some prior information for k, given in the form of the density function $p(k|\boldsymbol{\alpha})$ for $-\infty < k < \infty$, where $\boldsymbol{\alpha}$ are the hyper-parameters of this prior distribution.

3. We model the scale parameter in (10) by

$$\log \sigma(s_i) = \sigma_0 + \mathbf{x}(s_i)\boldsymbol{\beta}' + W_0(s_i),$$

where, $W_0(s_i)$ is a Gaussian process $N(0, \Sigma)$ and $\mathbf{x}(s_i)$ are explanatory variables and $\boldsymbol{\beta}$ are random regression coefficients having some specified prior distributions. Σ is a covariance matrix based on a valid, parametric covariance structure. See for example, Banerjee et al. (2004) for a list of these valid covariance structures.

Once prior distributions for the hyper-parameters of the model are specified, the above model becomes a fully specified hierarchical model, and well established simulation based inference techniques can be used to estimate the model parameters.

In model (10), spatial dependence and variation are only introduced through the scale parameter. However, it is possible to introduce extra spatial dependence and heterogeneity to the model through the shape parameter by adding another link function of the form

$$k(s_i) = k_0 + \boldsymbol{\alpha}\mathbf{x}'(s_i) + W_1(s_i),$$

where $W_1(s)$ is another latent, Gaussian spatial process which may or may not be independent of the latent process $W_0(s)$. One possible dependence structure for $W_0(s)$ and $W_1(s)$ is to start with two independent Gaussian processes $G_1(s)$ and $G_2(s)$ and write

$$W_0(s) = G_1(s),$$

and

$$W_1(S) = a_1 G_1(s) + a_2 G_2(s),$$

for some parameters a_1 and a_2. This is the coregionalized CAR model (Banerjee et al. 2004).

Note that we are modeling the excesses over a high threshold, that is, we model

$$P\left(X(s_1) > z_1 + u, ..., X(s_p) > z_p + u | X(s_1) > u, ..., X(s_p) > u\right),$$

for any collection of locations $s_1, ..., s_p$ and for some high, fixed threshold u. In general, we may want to make the inference conditional on at least one component exceeding the fixed threshold. However, implementation of the hierarchical model, due to the complications in the likelihood (see Coles 2001) can be difficult. One alternative is to start with a conditional approach to multivariate extremes suggested by Heffernan and Tawn (2004) and adapt these methods to spatial extremes.

Often, one needs to make inference on the joint probability $P(X(s_1^*) > u + z_1, .., X(s_p^*) > u + z_p)$ for arbitrary locations $s_1^*,...,s_p^*$ and in this case, apart from estimating the joint distribution of the excesses, one will need an estimate of the joint probability

$$P\left(X(s_1) > u, ..., X(s_p) > u\right).$$

Empirical methods for estimating this probability may not be straightforward, thus further extension of conditional independence assumption is needed to implement simpler modeling strategies.

Assume that there is a latent process $W(s)$ such that conditional on $W(s)$, large values of $X(s)$ over the threshold are independent, so that

$$
\begin{aligned}
P\left(X(s_1) > u + z_1, ..., X(s_p) > u + z_p | W(s_1) = w(s_1), ..., W(s_p) = w(s_p)\right) \\
= P\left(Z(s_1) > z_1, ..., Z(s_p) > z_p | W(s_1) = w(s_1), ..., W(s_p) = w(s_p)\right) \\
\times P\left(X(s_1) > u, ..., X(s_p) > u | W(s_1) = w(s_1), ..., W(s_p) = w(s_p)\right) \\
= \prod_{i=1}^{p} P\left(Z(s_i) > z_i | W(s_i) = w(s_i)\right) P\left(X(s_i) > u | W(s_i) = w(s_i)\right).
\end{aligned}
\tag{11}
$$

Writing $\eta(s_i) = P\left(X(s_i) > u | W(s_i) = w(s_i)\right)$ as a spatially varying random parameter, we suggest the following hierarchical model for making inferences on the joint posterior probability $P\left(X(s_1) > u + z_1, ..., X(s_p) > u + z_p\right)$:

1. Likelihood:

$$
\begin{aligned}
p\left(x(s_1), ..., x(s_p) | k, \sigma(s_i), \eta(s_i), i = 1, ..., p\right) \\
= \prod_{i=1}^{p} \eta(s_i) \frac{1}{\sigma(s_i)} \left(1 + \frac{k}{\sigma(s_i)} z(s_i)\right)^{-(1/k)-1},
\end{aligned}
\tag{12}
$$

2. Since $\eta(s_i) \in (0, 1)$, $\eta(s_i)$ can be connected to the latent process $W(s_i)$ through the logit function:

$$
\log \frac{\eta(s_i)}{1 - \eta(s_i)} = \eta_0 + W(s_i).
$$

Specification of the other parameters would be the same as in the previous model. Note that for any s_i, in principle, $\eta(s_i)$ is a small positive number, hence prior distributions of the parameters in the link function need to be finely tuned, otherwise it is very unlikely that the chain would converge. In general, we would recommend simplification by assuming that $\eta(s_i) = \eta$ and $\eta \sim Beta(0.1, 10)$.

The choice of prior distributions may have a big impact on convergence and it may not be possible to use vague priors for all the parameters and hyper-parameters. Another possible complication is the large dimension of the data set, although, in principle, this should not be a problem for data sets appearing in extreme value problems. However, data sets having more than 150 observations will involve inversion of matrices more than 150×150 in each updating iteration, which often results in failure in convergence. Banerjee et al. (2004) discuss ways of overcoming this dimension problem.

4 A case study

In many fire regimes, a small number of very large fires is responsible for the vast majority of the area burned and of the social and environmental damage caused. Therefore, large wildfires are a relevant public policy issue, especially considering

that the frequency of extremely severe fire weather may increase as a consequence of global warming. A review of the study of wildfire extremes can be found in de Zea Bermudez et al. (2009) and Mendes et al. (2008), where Bayesian hierarchical modeling strategies are discussed.

The data consist of 36180 records of wildfires larger than 5ha, observed in Portugal between 1975 and 2005. Fire perimeters were mapped from Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper satellite imagery, with 30 m spatial resolution, to a geographical scale of 1:100.000. About 170 satellite images, acquired annually after the end of the summer fire season, were analyzed over the 31-year period.

The study area corresponds to mainland Portugal (see Fig. 1) located between 37°N and 42°N latitude and between 6°W and 10°W longitude. The study area is stratified into 18 regions, representing a compromise between the Portuguese Forest Service standard procedure of organizing statistical fire data by administrative units and the natural regions classification of Portugal.

A similar data set was previously analyzed in Mendes et al. (2008). However, we revisit the data set and extend this study in different directions:

1. During the exceptional fire season of 2003, several large fires coalesced into huge, continuous fire scars, corresponding to multiple individual fire events. It is

Fig. 1 18 Administrative regions of Portugal.

Map of Portugal

unfeasible, using post-fire season Landsat imagery, to split these composite fire scars into their individual components. Field data from both the Forest Service and the Civil Protection Service were also found to be inadequate to accomplish this task, due to geographical and/or temporal inaccuracies and missing information. Thus, extra work was carried on using daily NOAA/AVHRR imagery, at 1km spatial resolution to unravel these cases. All fire scars larger than 1000 ha in the database were checked, and these very influential outliers were split. We believe that this improved data set will give more reliable results in estimating the tails. Also data for 2005 became available and is now introduced into to the model.

2. In Mendes et al. (2008), it is assumed that the shape parameter k should lie in (0, 1). This prior information was introduced to the model by using a logit link function for the spatially varying shape parameter in the hierarchy. This prior information was due to the expert opinion that the fire size distribution is heavy tailed but a fire size distribution with infinite mean would not be compatible with their beliefs. However, there is very strong empirical evidence that in Portugal the fire size distributions have heavy tails (de Zea Bermudez et al. 2009), which are compatible with infinite mean fire size distributions. Reed and McKelvey (2002) also give theoretical justification for possible heavy tailed behavior for fire sizes. They show that as $x \to \infty$

$$P(X(s) \geq x | X(s) \text{ is finite}) \sim x^{-\alpha},$$

if and only if as $x \to \infty$

$$x \frac{\nu(x)}{\mu(x)} \to x^{-\alpha},$$

where, $\nu(x)$ and $\mu(x)$ are respectively the extinguishment and growth rates of a fire with size x. It is clear from this necessary and sufficient condition that even fire regimes whose extinguishment rates are larger that their growth rate can produce heavy tailed fire size distributions. When the problem is introduced in this form, experts agree on models which accommodate heavy tailed fire sizes, possibly with infinite means. Hence, in this paper, we substitute the logit link function for the shape parameter by the logarithmic link function.

3. An areal model (typically a spatial Markov chain) based on 18 regions would be sufficient to capture the spatial variation due to synoptic meteorology conditions (Mendes et al. 2008). We now provide a more complicated hierarchical model with extra latent spatial structure to take into account smooth changes within each region representing small scale variations such as land topography, land use and fire prone vegetation.

4. There is strong empirical evidence that the extreme fire sizes change over time and show temporal (annual) dependence. It is believed that this is due basically to the availability of fire-prone vegetation throughout the years. In this version, we also add to the model, latent temporal random effects to account for this temporal variation.

4.1 Model

Let $N_j(t)$ denote the number of fires with sizes over 250 hectares, observed in region j, during year t. The threshold of 250 hectares is chosen after careful analysis as the best compromise between threshold stability and number of data points for inference. In this paper we will not attempt to model the point process of fires above 250 hectares but only the excess fire sizes over the threshold of 250 hectares conditional on the observed counts $(N_j(t) = n_j(t), j = 1, ..., 18, t = 1, ..., 31)$, the number of fires in excess of 250 hectares in each region between 1975–2005. See Mendes et al. (2008) for Poisson-GPD models.

Let $z(s_{ij}, t)$ denote the excess over 250 hectares of the ith fire in the jth region in year t, located at s_{ij}, which is the centroid of the observed fire scar. Let $\mathbf{z}(\mathbf{s}) = \{z(s_{ij}, t), i = 1, 2, ..., n_j(t), t = 1, ..., 31; j = 1, ..., 18\}$,

1. **Level 1: likelihood**

$$p\left(\mathbf{z}(\mathbf{s})|(n_j(t), k(s_{ij}, t), \sigma(s_{ij}, t), j = 1, ..., 18, t = 1, ..., 31\right)$$

$$= \prod_{t=1}^{31} \prod_{j=1}^{18} \prod_{i=1}^{n_j(t)} p\left(z(s_{ij}, t)|k(s_{ij}, t), \sigma(s_{ij}, t)\right), \tag{13}$$

where,

$$p\left(z(s_{ij}, t)|k(s_{ij}, t), \sigma(s_{ij}, t)\right) = GPD\left(k(s_{ij}t), \sigma(s_{ij}, t)\right)$$

$$= \frac{1}{\sigma(s_{ij}, t)}\left(1 + \frac{k(s_{ij}, t)}{\sigma(s_{ij}, t)}z(s_{ij}, t)\right)^{-(1/k(s_{ij}, t))-1}. \tag{14}$$

2. **Level 2: link Functions**

    (a)

$$\log\left(\sigma(s_{ij}, t)\right) = \sigma_0 + \eta_\sigma(j) + \delta_\sigma(t) + W_\sigma(s_{ij}, t) \tag{15}$$

    (b)

$$\log\left(k(s_{ij}, t)\right) = k_0 + \eta_k(j) + \delta_k(t), \tag{16}$$

where

- k_0 and σ_0 are independent a priori with flat priors, that is, they are uniformly distributed over the real line.
- $\eta_\sigma(j), \eta_k(j)$ are independent, nearest neighbor CAR models (Banerjee et al. 2004) with precision hyper-parameters τ_σ and τ_k, representing the large scale spatial variations.

- $\delta_\sigma(t)$ and $\delta_k(t)$ are independent latent processes representing temporal (annual) variations, having precision hyper-parameters τ_{δ_σ} and τ_{δ_k} respectively. We use a second order random walk prior given as car.normal distribution in GeoBUGS (Thomas et al. 2004). Temporal dependence is assumed to exist due basically to the annual availability of fire prone vegetation. For example, one may expect that a year with a high rate of large fires would be followed by a year with smaller fires due to the unavailability of fire prone vegetation. Since in Portugal the cycle of growth for fire prone vegetation is around 2–3 years, a second order random walk is preferred to first order random walk.

- $W_\sigma(s_{ij}, t)$ is a latent spatial moving average process representing the small scale variation, such as the local land topography and concentration of fire prone vegetation in each year, given by

$$W_\sigma(s, t) = \sum_{k=1}^{\ell} \varphi(s, s_{k,t}) \epsilon(s_{k,t}), \qquad (17)$$

where, $\epsilon(s_{k,t})$ are iid $N(0, \tau_\epsilon)$ and $s_{k,t}, k = 1, 2, .., \ell$ are ℓ excess fire locations (possibly from different regions) nearest to the excess fire location s in year t. For computational convenience, we chose $\ell = 5$. Alternatively, one could choose all fires within a fixed distance from the fire located at s, but this would unnecessarily complicate the analysis. The weights are Gaussian kernels given by

$$\varphi(s, s_{k,t}) = \frac{1}{b^2} \exp\left[-\frac{1}{b^2} \|s - s_{k,t}\|^2\right].$$

Note that for each year, the location of excess fires, namely s as well as the nearest five neighbors $s_{k,t}$, are different and are generated by a random point process in space. For simplicity, we call $W_\sigma(s, t)$ the moving average kernel.
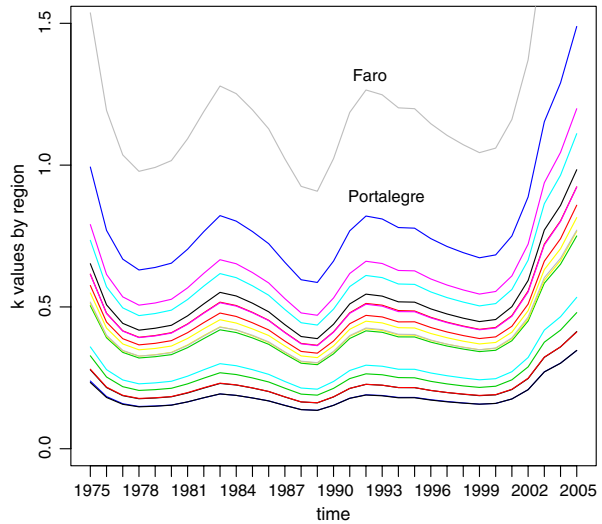
3. **Level 3: hyper-parameters**

τ_σ τ_k, τ_{δ_σ}, τ_{δ_k} and τ_ϵ are all independent precision hyper-parameters. We assume diffuse Gamma priors for all these parameters except for τ_ϵ. For $(\tau_\epsilon)^{-1}$, we use an informative uniform distribution over $(0.01, 0.20)$. A non-informative prior on this parameter causes non-convergence of the chain. This prior distribution allows $W_\sigma(s, t)$ to have a scale of variability in the same order as the other random effects $\eta_\sigma(s)$ and $\delta_\sigma(s)$. For the band-width parameter b, we assume a uniform distribution between 1 and 12 kilometers. This is a reasonable bandwidth for small scale variation.

More complicated space time dependence structures can be given in the link functions See for example, Sang (2008). However, we will not pursue these structures.

The choice of priors evidently affects the convergence of the MCMC method. Banerjee et al. (2004) suggest also that assuming noninformative priors for all the

Fig. 2 Estimated shape parameter (posterior means) by region and by time.

hyper-parameters may result in non-convergence of the chain. They suggest informative priors for the precision hyper-parameters. Chavez-Demoulin and Davison (2005) suggest re-parametrization of the GPD parameters by $\tau = \sigma(1 + k)$, whereas Castellanos and Cabras (2006) look at the effect of this re-parametrization on the MCMC method used. A small sensitivity analysis showed that changes in priors for the hyper-parameters do not have an impact on the estimated values of k. However, careful choice of the prior for τ_ϵ is needed since it affects the variability in the posterior distribution of σ.

4.2 Results

Apart from the model given in Section 4.1, we have implemented several other alternative models to evaluate the impact of the small scale and the temporal latent factors, as well as the log-link function versus logit function for the shape parameter $k(s, t)$. The benchmark model given in Section 4.1 performed better than other alternatives.[1] We will not represent results on comparative studies; however, they are available upon request.

Yearly estimated shape parameters (means of the corresponding marginal posterior distributions) for the 18 regions show that Region 8 (Faro) stands alone with its very heavy tailed behavior. Consistently it has values of k above 1. Also region 12 (Portalegre) has estimated values of k between 0.5 and 1. There are clusters of regions according to the values of k; see Fig. 2 which shows the variation of the posterior means of k (Bayesian estimated values) for each region during 31 years. The analysis of these 31-year time series of the estimated shape (k) parameters does

---

[1]A more detailed and complete analysis of this data set can be found in an unpublished research report by the authors available by request.

not seem to indicate annual, trend like movements. However, it suggests a cyclical behavior of extreme fire sizes, with a frequency of 2 to 3 years, although a longer time series may be needed to assess the robustness of this pattern by using formal statistical procedures. If it really exists, it is more likely to result from post-fire vegetation response dynamics than from meteorological or anthropogenic factors. See de Zea Bermudez et al. (2009) for similar findings.

Figure 3 shows the spatial variation of the estimated shape parameters which reports the posterior means and posterior standard deviations of k, averaged over 31 years.
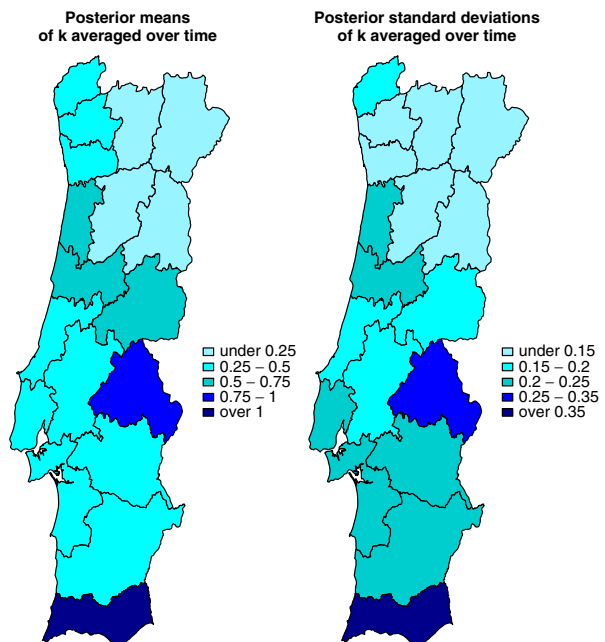
Posterior means and standard deviations for σ are higher in the regions of the center and costal areas between region 11 (Lisbon) and 13 (Porto). Figure 4 shows the spatial variation of the estimated scale parameter (posterior means) σ, averaged over the 31 years.

Year 29 (2003) had many large fires above 250 hectares with large variability in size, particularly in regions 5, 12 and 14. This variability is captured by the latent moving average kernel, explaining the small-scale variation. Note that the shape parameter is common to all fires within a region for a given year, and without the moving average kernel in the model there would have been no internal variation in the estimated quantiles.
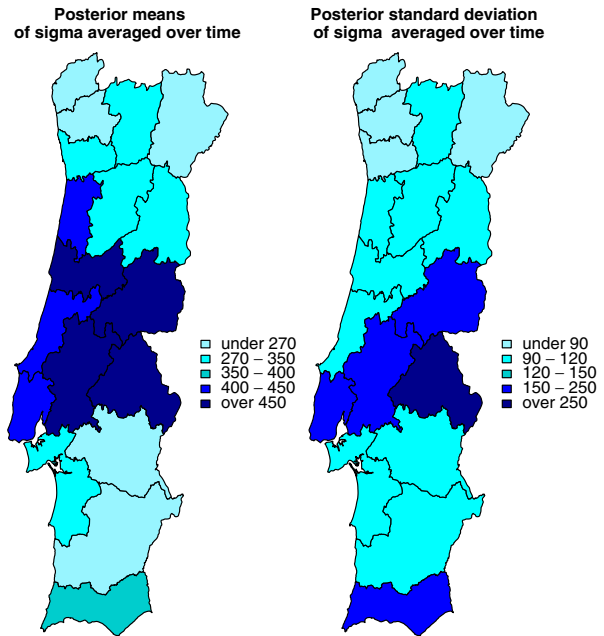
Figure 5 reports on the posterior means of latent temporal components $(\delta_k, \delta_\sigma)$.

We observe clearly the cyclical temporal effect both for k and σ. The increasing temporal effect on the values both for k and σ during the recent years may be due to the fact that both in 2003 and 2005, there were many large fires. Note also that there is an apparent negative relation between the temporal effects of k and σ. This

**Fig. 3** Spatial variation of the estimated shape parameter.

Fig. 4 Spatial variation of the estimated scale parameters.

Posterior means
of sigma averaged over time

Posterior standard deviation
of sigma averaged over time

under 270
270 – 350
350 – 400
400 – 450
over 450

under 90
90 – 120
120 – 150
150 – 250
over 250

is expected since, increasing the scale parameter lengthens the tail as does increasing the shape parameter and so if one parameter is increased the other needs to decrease to roughly unchange the upper tail, consequently they will be negatively dependent (See Coles and Tawn 1996). The latent temporal effects clearly capture this dependence.

Fig. 5 Latent temporal effects.

temporal effects for sigma and k
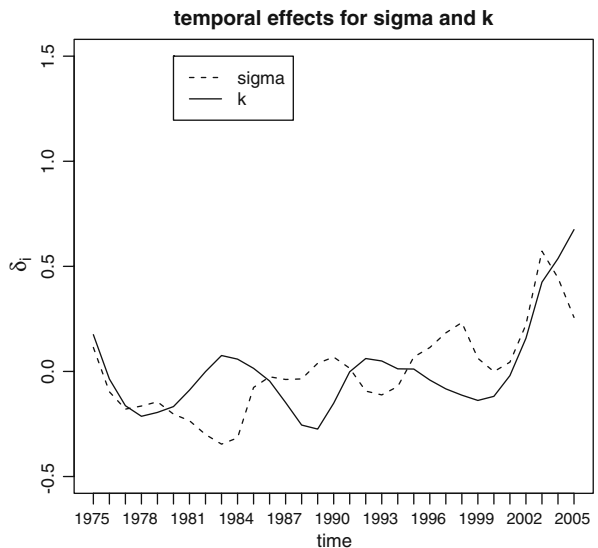
4.2.1 Characterization of the sources of variability

Different sources of variability in the large fire sizes can be studied by a strategy similar to that proposed by Gelman and Pardoe (2005). See also Cocchi et al. (2007). The decomposition of the variance according to the different sources of variability can be studied by decomposing the expected variation in each of the state equations k and σ in the second hierarchy:

$$E\big[var\big(\log k|\text{Data}\big) = E\big[var\big(k_0|\text{Data}\big)\big] + E\big[(var\big(\eta_k|\text{Data}\big)\big] + E\big[var\big(\delta_k|\text{Data}\big)\big]$$
$$+ \text{Respective covariances}, \tag{18}$$

and

$$E\big[var\big(\log \sigma|\text{Data}\big) = E\big[var\big(\sigma_0|\text{Data}\big)\big] + E\big[(var\big(\eta_\sigma|\text{Data}\big)\big] +$$
$$= E\big[var\big(\delta_\sigma|\text{Data}\big)\big] + E\big[var\big(W_\sigma|\text{Data}\big)\big]$$
$$+ \text{Respective covariances}. \tag{19}$$

Note that, while the respective terms *a priori* are independent, they may be dependent *a posteriori*. However, careful analysis of the samples from these posterior distributions showed that these covariances are negligible, so that the variances in (18) and (19) help us to evaluate the percentage explained by the spatial and temporal effects. Calculation of the posterior variances revealed that

1. The percentage of variation in $\log k$ explained by k_0 is 11.7%, the percentage of variation explained by the spatial latent factor η_k is 55.3%, whereas the percentage of variation explained by the temporal latent factor δ_t is 33%.
2. The percentage of variation in $\log \sigma$ explained by σ_0 is 0.49%, whereas the influence of the temporal latent factor δ_σ is 2.7%. The influence of the large scale variation CAR model for the regions is 4.4%, whereas the influence of the small scale moving average kernel is 92.4%.

High value of percentage explained by the kernel is due to the fact that the shape parameter is fixed in each region and hence the only way the model can adapt to the data is by varying the scale parameter. Note that the model without the local moving average kernel would attribute the same risk for sites in the same region, regardless of their location, since not only k, as well as σ estimates would be only region dependent. For instance, for the year 2003, the probabilities $P(X > 1000|X \geq 250)$, using the model without the moving average kernel were estimated as

$$P\big(X > 1000|X \geq 250\big) = \begin{cases} 0.412, & \text{Region 5;} \\ 0.414, & \text{Region 12;} \\ 0.381, & \text{Region 14.} \end{cases}$$

The effect of the kernel is visible in the estimation of the probabilities. Without this small scale random effect in the model, the risk of a large fires would have been underestimated.

4.2.2 Final conclusions

Conditional independence which is the basis of hierarchical modeling is a very strong assumption which may not be easily verified, and it may not represent correctly the local extremal behaviour. When the objective of the data analysis is to estimate return levels or kriging of extreme values in space, capturing the correct dependence structure between the extremes is crucial and max-stable processes are better suited for these purposes. However in this work, our primary interest is to explain the sources of variation in extreme fires and in this case, Bayesian hierarchical modeling is a very flexible tool due to the ease with which random effects are incorporated in the model. Therefore, the choice of the modeling strategy will depend on specific data set and the purpose of the study and the benefits of using each of the strategies must be carefully weighted. For example, since in our case study, the data are not generated at fixed observation locations, the model parameters can not be estimated locally to transform the data. On the other hand, spatial heterogeneity is too strong to ignore. Hence, it is not clear how one can make inference on a max-stable process based on this data set. The conditional independence assumption on which we base our hierarchical model cannot be verified, but overall it is seen as being fairly consistent with the mechanism that generates the data. Therefore, it is our belief that the hierarchical modeling strategy is more suitable for the data set that we analyze.

In our model, we used simple, separable space-time dependence structures, although more complicated structures can be used. The important advantage of the use of these simple structures is that the models can be run by using the WinBUGS software (Lunn et al. 2000). Ideally, one should use a point referenced local model in dealing with small-scale variations. However, this would bring a very heavy computational burden on the model. The purpose of the model, we suggest, is to explain the sources of variation in extreme fire events; it is not particularly suited for predictive purposes. For prediction, one should have information on explanatory variables such as land topology, land use, spatial and temporal variation of fire prone vegetation as well as on meteorological variables such as temperature, wind speed, wind direction, humidity etc. In principle, introduction of these explanatory variables in the model through the link functions would not create extra difficulty. However, at present no such information was available for the study.

References

Albin, P.: On extremal theory for non differentiable stationary processes. Ph.D. thesis, Dept. of Mathematical Statistics, University of Lund (1987)

Albin, P.: On extremal theory for stationary processes. Ann. Probab. **18**, 92–128 (1990)

Banerjee, S., Carlin, B.P., Gelfand, A.: Hierarchical Modelling and Analysis for Spatial Data. Chapman and Hall (2004)

Buishand, A., de Haan, L., Zhou, C.: On spatial extremes: with application to a rainfall problem. Ann. Appl. Stat. **2**, 624–642 (2008)

Castellanos, M.E., Cabras, S.: A default Bayesian procedure for the generalized Pareto distribution. J. Stat. Plan. Inference **137**, 473–483 (2006)

Chavez-Demoulin, V., Davison, A.: Generalized additive modelling of sample extremes. Appl. Stat. **54**, 207–222 (2005)

Cocchi, D., Greco, F., Trivisano, C.: Hierarchical space-time modelling of $PM_1 0$ pollution. Atmos. Environ. **41**, 532–542 (2007)

Coles, S.: An Introduction to Statistical Modelling of Extreme Values. Springer, New York (2001)

Coles, S., Tawn, J.: Modelling extremes of the areal rainfall process. J. R. Stat. Soc. B **58**, 329–347 (1996)

Cooley, D., Naveau, P., Joneli, V., Rababtel, A., Grancher, D.: A Bayesian hierarchical extreme value model for lichenometry. Environmetrics **17**, 555–574 (2006)

Cooley, D., Nychka, D., Naveau, P.: Bayesian spatial model of extreme precipitation return levels. JASA **102**, 824–840 (2007)

de Haan, L., Ferreira, A.: Extreme Value Theory, An Introduction. Springer, New York (2006)

de Haan, L., Pereira, T.: Spatial extremes: models for the stationary case. Ann. Stat. **34**, 146–168 (2006)

de Zea Bermudez, P., Mendes, J., Pereira, J.M.C., Turkman, K.F., Vasconcelos, M.J.P.: Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). Int. J. Wildland Fire (2009, in press)

Fawcett, L., Walshaw, D.: A hierarchical model for extreme wind speeds. Appl. Stat. **55**, 631–646 (2006)

Fougères, A.L., Nolan, J.P., Rootzén, H.: Models for dependent extremes using stable mixtures. Scand. J. Statist. **36**, 42–59 (2009)

Gelman, A., Pardoe, I.: Bayesian measures of explained variance and pooling in multilevel models. Technometrics **40**, 241–251 (2005)

Heffernan, J.E., Tawn, J.: A conditional approach for multivariate extreme values. J. R. Stat. Soc. B **66**, 497–546 (2004)

Kabluchko, Z., Schlather, M., de Haan, L.: Stationary Max-stable fields associated to negative definite functions. Ann. Probab. www.arxiv.org (2009)

Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983)

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D.: WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. **10**, 325–337 (2000)

Mendes J.M., de Zea Bermudez P., Pereira J.M.C., Turkman K.F., Vasconcelos, M.J.P.: Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. Environ. Ecol. Stat. doi:10.1007/s10651-008-0099-3 (2008)

Padoan, S.A., Ribatet, M., Sisson, S.A.: Likelihood-based inference for max-stable processes. JASA (Theory and Methods) (2009, submitted)

Piterbarg, V.: Asymptotic Methods in the Theory of Gaussian Processes and Fields. AMS Monographs (1996)

Reed, W.J., McKelvey, K.S.: Power-law behaviour and parametric models for the size-distribution of forest fires. Ecol. Model. **150**, 239–254 (2002)

Ribatet, M.: A User's Guide to the SpatialExtremes Package. École Polytechnique Fédérale de Lausanne, Switzerland (2009)

Sang, H.: Extreme value modeling for space-time data with meteorological applications. Ph.D. thesis, Duke University (2008)

Schlather, M.: Models for stationary max-stable random fields. Extremes **5**, 33–44 (2002)

Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D.: GeoBUGS User Manual Version 1.2 (2004)