

Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes

Jorge M. Mendes · Patrícia Cortés de Zea Bermudez · José Pereira · K. F. Turkman · M. J. P. Vasconcelos

Received: 2 August 2007 / Revised: 3 March 2008 / Published online: 7 August 2008
© Springer Science+Business Media, LLC 2008

Abstract In Portugal, due to the combination of climatological and ecological factors, large wildfires are a constant threat and due to their economic impact, a big policy issue. In order to organize efficient fire fighting capacity and resource management, correct quantification of the risk of large wildfires are needed. In this paper, we quantify the regional risk of large wildfire sizes, by fitting a Generalized Pareto distribution to excesses over a suitably chosen high threshold. Spatio-temporal variations are introduced into the model through model parameters with suitably chosen link functions. The inference on these models are carried using Bayesian Hierarchical Models and Markov chain Monte Carlo methods.

Keywords Bayesian hierarchical models · Generalized Pareto distribution · Spatial and temporal processes · MCMC

1 Introduction

Most wildfires are extinguished upon ignition and do not have significant effect. However, some wildfires go out of control, causing significant economical and social

J. M. Mendes
CEGI, ISEGI-UNL, Lisbon, Portugal
e-mail: jmm@isegi.unl.pt

P. C. de Zea Bermudez · K. F. Turkman (✉)
CEAUL, Faculty of Sciences, University of Lisbon, Lisbon, Portugal
e-mail: kfturkman@fc.ul.pt

J. Pereira
ISA-UTL, Lisbon, Portugal

M. J. P. Vasconcelos
Tropical Research Institute, Lisbon, Portugal

impact. In Portugal, due to the combination of climatological and ecological factors, large wildfires are a constant threat and are important policy issue, requiring correct quantification of regional fire risk for organizing an efficient fire fighting capacity and resource management. Occurrence and the size of wildfires depend on local and global variations in climatological as well as ecological conditions, showing evident spatio-temporal heterogeneities and dependencies at different time-space resolutions. Therefore, in order to quantify the risk of large wildfires correctly, these variations and dependencies need to be carefully taken into account in the statistical inference. In this paper, our objective is to model the large fire sizes, taking into account spatial and temporal variations.

Extreme value theory is the natural inferential tool to quantify the risk of large fire sizes. Within this theory, the most commonly used tool is the Peaks Over Threshold (POT) method. In this method, the Generalized Pareto distribution (GPD) is used to model excesses over a suitably chosen high threshold and the asymptotical justification is given by [Pickands \(1975\)](#). Hence the tail of the conditional distribution of fire size, which we denote by X , is characterized by the GPD model

$$P(X > x + u | X > u) = \left(1 + k \frac{x}{\sigma}\right)^{-1/k}, \quad (1)$$

where $\sigma > 0$, $k \in (-\infty, \infty)$ and $1 + k \frac{x}{\sigma} > 0$. Here σ is the scale parameter and depends on the chosen threshold u . The k is called the shape parameter and is invariant in the threshold u chosen.

The theory for the POT method is carried under the assumption that the data is independent and identically distributed. When dealing with non-stationary data with strong short term serial dependence, the POT method then has 3 steps: first, one has to remove non stationarity. Several methods have been suggested to deal with non-stationarity (see, for example, [Smith \(1989\)](#) and [Ramesh and Davison \(2002\)](#)). Second, the threshold exceedances need to be de-clustered into independent clusters of exceedances. The GPD model is then fitted to the cluster maxima data, treating them as being independent.

For a general review of extreme value theory and specific statistical methods, see [Embrechts et al. \(1997\)](#).

The techniques dealing with non stationary data and de-clustering techniques for dependent local extremes are well understood and studied for time series. However, extension of these data handling techniques for non-stationary, dependent spatial data sets is an active area of research. [Turkman \(2006\)](#) gives the notion of extremal index for spatial extremes. However, there has been no attempt to estimate this spatial extremal index. [de Haan and Pereira \(2006\)](#) suggest starting from the stationary spatial process $X(\mathbf{s})$ and its independent, identical copies $X(s, t)$ traveling in time and look at the convergence

$$\frac{1}{a_T} \left(\max_{t \in T} X(\mathbf{s}, t) - b_T \right) \rightarrow^W Y(\mathbf{s})$$

for some suitable normalizing constants a_T and b_T . Under reasonable conditions, the limiting process $Y(\mathbf{s})$ is a max-stable process (see [de Haan and Lin \(2001\)](#)). This max-stable process can then be used as an asymptotic model for the extremal values of the spatial process $X(\mathbf{s})$. If one can derive or make inference on the probability structure of $Y(\mathbf{s})$, the problem of extrapolation of large values in space to unobserved sites and/or inference on the largest value $\sup_{\mathbf{s} \in S} X(\mathbf{s})$ can be made. [de Haan and Pereira \(2006\)](#) then look at specific models for the extremal process $Y(\mathbf{s})$ to address the issue of spatial extrapolation. However, their method only permits the calculation of bivariate distributions of $Y(\mathbf{s})$ for specific models. [Davis and Mikosch \(2006\)](#) look at the same problem, but start with a heavy tailed linear space-time process, continuous in space and discrete in time, and then derive the characteristic behavior of the limiting extremal process. Similar models were already suggested by [Coles and Tawn \(1996\)](#) for modeling areal rainfall processes. Although the use of these models are justified from asymptotic results, they are not sufficiently flexible enough to handle non stationary data. One possible and very flexible way of approaching this problem is by using Bayesian hierarchical models and simulation based inference ([Banarjee et al. 2004](#); [Paulino et al. 2003](#)). Few attempts have been made to model extremal data using the hierarchical modeling techniques ([Casson and Coles 1999](#)). However, first fully Bayesian model is suggested by [Cooley et al. \(2007\)](#) for modeling spatial extreme precipitation return levels using the GPD model. See also [Sang and Gelfand \(2007\)](#), [de Zea Bermudez et al. \(2007\)](#), classical extreme value theory and methods were used to quantify the risk of large wildfires in Portugal. Spatial heterogeneity was taken into consideration by fitting different models to 8 different geographical regions. In this paper, we model the spatial variation of extreme fires with Bayesian hierarchical models, thus quantifying the spatial variation of the large fires in a more precise manner. Excesses over a suitably chosen threshold will be modeled by a GPD model, whereas the counts of fires in excess of this threshold will be modelled by a over-dispersed Poisson model. Spatio-temporal variation will be introduced through the model parameters with properly chosen link functions. At present, we do not have information on the climatological and ecological factors that may influence large fires and these factors will be represented in the model through a hidden spatially, temporally colored random effect.

The outline of the paper is as follows. In [Sect. 2](#), data set will be described and some preliminary data analysis results will be represented. In [Sect. 3](#), GPD models are fitted to each of the 30 annual data sets to assess how large fires size risk changes over time. In [Sect. 4](#), several areal GPD models are suggested to quantify the risk of large fires for 18 administrative districts of Portugal, taking into account the spatio-temporal variations.

2 Data

The data consist of records of wildfires observed in Portugal between 1975 and 2004. Fire perimeters were mapped from Landsat Thematic Mapper satellite imagery, with 30m spatial resolution, acquired annually after the end of the summer fire season. The burned areas are mapped with a supervised classification approach and a classification

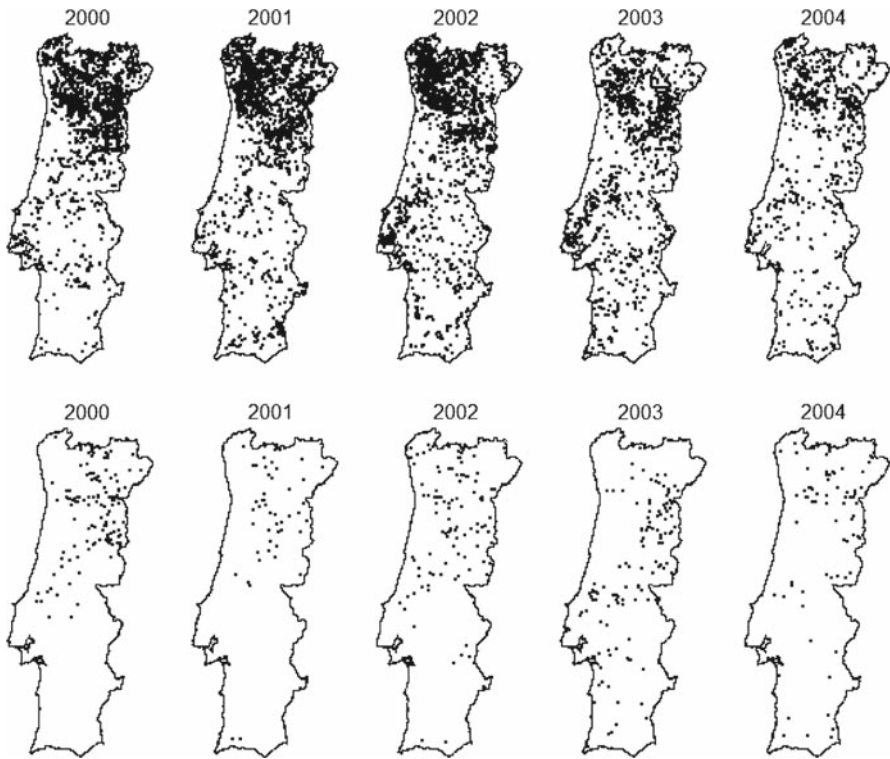


Fig. 1 Locations of fires above 5 hectares (top) and locations of fires above 250 hectares (bottom)

tree algorithm (Breiman et al. 1984). Spatial locations are the centroids of burned areas. Data set 1975–1983 contains fires sizes above 35 hectares of burned land, whereas, due to the improved resolution of the satellite imagery, data set 1984–2004 is on fire sizes above 5 hectares. Since excesses above a high threshold will be modeled (This threshold will be chosen to be at least 100 hectares, for reasons to be explained later), the data set is truncated at 35 hectares for all the period 1975–2004. Hence our data set is of the form

$$\{x(s_i, t), i = 1, \dots, n_t, t = 1, 2, \dots, 30\}, \quad (2)$$

where $\mathbf{s} = (s_1, s_2)$ are the spatial coordinates of the centroids of the recorded fires, n_t is the number of wildfires with sizes above 35 hectares in year t , $t = 1975, \dots, 2004$.

At present, no data on climatological, ecological as well as topological covariates are available for the analysis. A latent spatially and temporally colored random effect will be used to describe the spatial-temporal variations due to these covariates.

In Fig. 1, the spatial locations of all recorded fires above 5 hectares as well as the location of fires above 250 hectares are given for the years 2000 and 2004.

In de Zea Bermudez et al. (2007), the preliminary data analysis carried showed that GPD stability condition holds very well for any threshold chosen above 100 hectares

and the choice of the threshold at $u = 250$ is taken as a good choice. Therefore, for all the models fitted in this paper, we use $u = 250$. This choice which is based on asymptotical considerations is also confirmed by the expert opinion that in Portugal, a wildfire in excess of 250 hectares is considered extreme.

3 Temporal variation of wildfire sizes

In order to assess the temporal variation in the risk of large fire sizes, we first ignore the spatial variation that may exist in the data and fit the following model to the excesses of fire sizes above the 250 hectares threshold for the data set between 1975 and 2004.

3.1 Temporal model I

Let $Z_i(t)$ be the excesses of fire sizes above 250 hectares:

$$Z_i(t) = X_i(t) - 250, \quad \forall X_i(t) > 250,$$

and let N_t be the number of these fires in year t . The likelihood can be written as

$$\begin{aligned} p(z_i(t), n_t, i = 1, 2, \dots, n_t, t = 1, \dots, 30) \\ = P(z_i(t), i = 1, \dots, n_t, t = 1, 2, \dots, 30 | n_t, t = 1, \dots, 30) \\ p(n_t, t = 1, \dots, 30), \end{aligned} \tag{3}$$

so that the inference on the excesses conditional on the observed number of excesses and the counting process $N_t, t = 1, \dots, 30$, can be handled separately. Our primary interest is the inference on the tail of the excess size distribution, and therefore our efforts will concentrate on the first part of the likelihood in (3).

- **Excesses:**

1. **Level 1: likelihood**

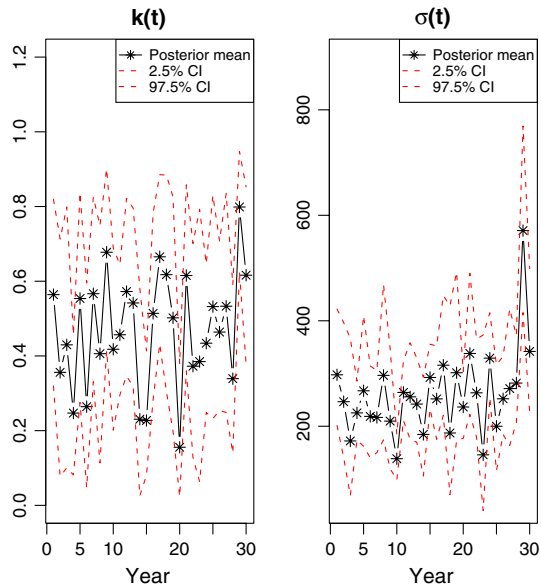
$$p(z_i(t), i = 1, 2, \dots, n_t, t = 1, \dots, 30, | \Theta) = \prod_{t=1}^{30} \prod_{i=1}^{n_t} p(z_i(t) | \Theta), \tag{4}$$

where,

$$\begin{aligned} p(z_i(t) | \Theta) &= GPD(k(t), \sigma(t)) \\ &= \frac{1}{\sigma(t)} \left(1 + \frac{k(t)}{\sigma(t)} z \right)^{-1-1/k(t)} \end{aligned} \tag{5}$$

where $\sigma(t) > 0, k(t) \in (-\infty, \infty), 1 + \frac{k(t)}{\sigma(t)} x > 0$ and n_t is the number of excesses over the threshold $u = 250$ in year t .

Fig. 2 Posterior means and credible intervals of the shape and scale parameters (1975–2004), for temporal model I



2. Level 2: parameters

- (a) $\{\sigma(1), \sigma(2), \dots, \sigma(30)\}$ in log scale are independent and identically distributed random variables with

$$\log(\sigma(t)) \sim \text{Uniform}(-\infty, \infty), \quad (6)$$

- (b) $\{k(1), k(2), \dots, k(30)\}$ are independent and identically distributed random variables with

$$k(t) \sim \text{Uniform}(0, 1). \quad (7)$$

Here, the prior for $\sigma(i)$ is a vague prior. Prior for $k(t)$ is informative, which is obtained by eliciting information from the experts. This information corresponded to the fact that the fire size distribution should have finite mean, but at the same time heavy tailed, which is then conveniently transformed to $\text{Uniform}(0, 1)$ prior for k . Note that, having independent prior distributions for each of the 30 years, this model is equivalent to fitting separately one GPD model for each of the 30 years.

The posterior mean of k and σ for each year together with their credible intervals are given in Fig. 2.

The figure suggests a two-three year cyclical behavior in both of the parameters. This is consistent with the growth cycle of fire prone vegetation and hence this may indicate the effect of the availability of fire prone vegetation on producing large fires. This point needs further study and needs to be backed up by proper spectral analysis. At present, with only 30 years of information, such analysis will not produce reliable conclusions. Both the shape and the scale parameters for the year 2003 are high and indeed this year was exceptional in terms of the fire sizes. In that year, there are 4 very large,

influential observations and it is suspected that these observations may correspond to clusters of contiguous fire scars, resulting in artificially large burned areas. Although these observations are kept in the analysis, further work is should be carried to split these composite fire scars into their individual components. Estimated autocorrelation and partial autocorrelation functions, albeit with few degrees of freedom, indicate that there is no evident serial dependence structure for both of the estimated parameters, suggesting temporal independence.

- **The counting process:** we will assume the following simple Poisson structure for the counting process:

$$\begin{aligned}
 p(n_t, t = 1, 2, \dots, 30) &= \prod_{i=1}^{30} p(n_t), \\
 p(n_t) &\sim \text{Poisson}(\lambda(t)), \\
 \lambda(t) &\sim \text{Gamma}(0.0001, 0.0001).
 \end{aligned}
 \tag{8}$$

Poisson character of the number of excesses over 250 hectares is justified by the well known asymptotic theory for extremes. Preliminary data analysis using the mean excess functions (see for example, Embrechts et al. (1997)) not reported here, justifies the use of the Poisson model for the numbers and the GPD model for the excesses of fire over 250 hectares (de Zea Bermudez et al. 2007).

Assumption of independence for the annual counts is over simplification. More complicated models representing time dependence can be introduced. However, our primary interest is on the excess size distribution and therefore, more complicated models for the counting process will not be pursued.

Figure 3 below gives the mean of the posterior distribution of $\lambda(t)$, as well as the credible intervals. Note that λ corresponding to 2003 is not relatively large, contrary to the estimated shape parameter of the GPD model for that year. This observation confirms the fact that in 2003, few but very large fires caused a lot of damage. The model fully captures this fact.

3.2 Temporal model II

We suggest a more complicated model for the excesses, taking temporal dependence into considerations.

1. Level 1: likelihood

The level 1 likelihood for this model is the same as that for the previous model.

2. Level 2: link Functions

(a)

$$\log(\sigma(t)) = \sigma_0 + \alpha_1 N_1(t - 1) + \alpha_2 N_2(t - 1), \tag{9}$$

$$\log\left(\frac{k(t)}{1 - k(t)}\right) = k_0 + \beta_1 N_1(t - 1) + \beta_2 N_2(t - 1). \tag{10}$$

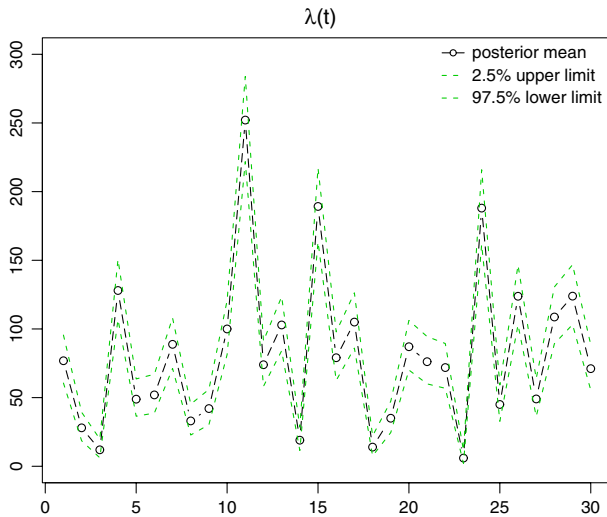


Fig. 3 Posterior mean and credible intervals of the mean of Poisson counting model 8

Here $N_1(t-1)$ and $N_2(t-1)$ are covariates denoting, respectively, the (normalized) average burned area and (normalized) average number of fires above 35 hectares in $t-3$, $t-2$ and $t-1$ whereas α_1 , α_2 , β_1 and β_2 are scalar regression parameters. It is known that availability of fire prone vegetation is an important factor for fire ignition, as well as for fire sizes. However, we do not have data on this covariate. Instead, we use $N_1(t-1)$ and $N_2(t-1)$ as proxy.

3. Level 3: parameters, priors, hyper-parameters and hyper-priors

(a) $\Theta = (\sigma_0, k_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ are the model hyper-parameters all having independent prior distributions, as follows:

- i. The baseline value for the shape and scale parameters in the log and logit scales, k_0 and σ_0 follow vague priors:

$$k_0 \sim \text{Uniform}(-\infty, +\infty), \quad (11)$$

$$\sigma_0 \sim \text{Uniform}(-\infty, +\infty). \quad (12)$$

- ii. Regression parameters α_l and β_l , $l = 1, 2$ have vague independent Gaussian priors:

$$\alpha_l \sim \text{Gau}(0, 0.0001), \quad l = 1, 2 \quad (13)$$

and

$$\beta_l \sim \text{Gau}(0, 0.0001), \quad l = 1, 2. \quad (14)$$

Posterior means of k and σ over the 30 years are given in Fig. 4. Note that, as expected, the temporal variation in model parameters are relatively lower as compared

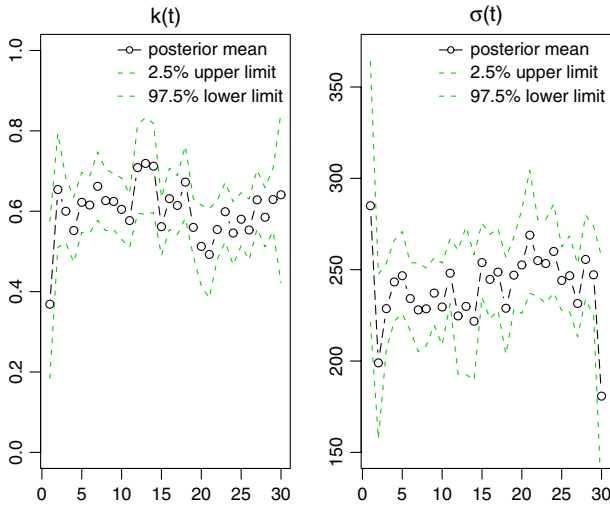


Fig. 4 Posterior means and credible intervals of the shape and scale parameters (1975–2004), for temporal model II

to the individual models for each year. Note also the effect of large fire sizes in 2003 on the scale parameter in 2004. Similar edge effect is seen on the shape and scale parameters of 1976, when $N_1(1975)$ and $N_2(1975)$ enter in the model.

Marginal prior and posterior distributions of the model parameters are given in Fig. 5. Note that β_1 is negative valued random variable, agreeing with the fact that the total area burned in the previous years are negatively correlated with the shape parameter, due to reduced availability of fire prone vegetation for the next year. Behavior of β_2 is less clear. β_2 and α_1 have posterior means very near 0, indicating that average number of fires above 35 hectares in the previous 3 years does not influence the shape parameter much. Posterior distribution of k_0 is not concentrated, but this may be explained by the fact that the fire size distributions in separate years have varying tails.

4 Spatial variation of large fire risk

4.1 Areal model I

In order to account for the spatial variation of the large fire sizes, we divide up the data into 18 administrative regions \mathbf{D}_j , $j = 1, 2, \dots, 18$, called *districts* (see Fig. 6). This geographical stratification of the study area represents a compromise between the Portuguese Forest Service standard procedure of organizing statistical fire data by administrative units, and the natural regions classification of Portugal.

Let $N_j(t)$ denote the number of fires with sizes over 250 hectares, observed in region j , during year t . We suggest the following model for the excesses over the threshold of 250 hectares conditional on $(n_j(t), j = 1, \dots, 18, t = 1, \dots, 30)$, the observed number of fires in excess of 250 hectares in each region and year to account for the spatial variation in the excess data:

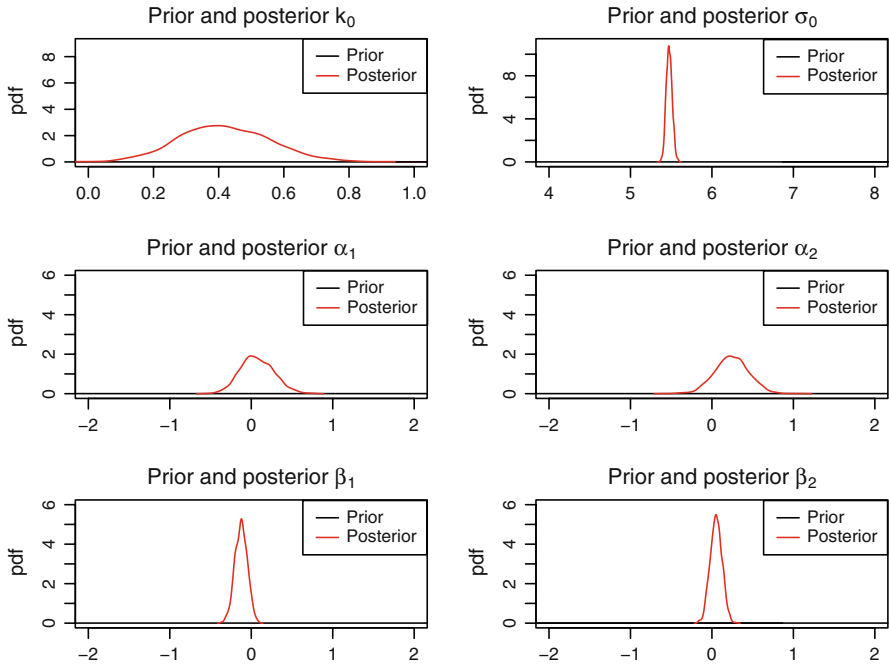


Fig. 5 Marginal prior and posterior densities of the model parameters (temporal model II)

Fig. 6 Administrative regions at district level, in Portugal



1. **Level 1: likelihood**

$$\begin{aligned}
 & p(z(\mathbf{s}_i, t), i = 1, 2, \dots, n_j(t), t = 1, \dots, 30; \\
 & j = 1, \dots, 18 | (n_j(t), j = 1, \dots, 18, t = 1, \dots, 30), \eta_\sigma, \Theta) \\
 & = \prod_{t=1}^{30} \prod_{j=1}^{18} \prod_{i=1}^{n_j(t)} p(z(\mathbf{s}_i, t) | \eta_\sigma, \Theta), \tag{15}
 \end{aligned}$$

where,

$$\begin{aligned}
 p(z(\mathbf{s}_i, t)) & = GPD(k, \sigma(\mathbf{s}_i, t)) \\
 & = \frac{1}{\sigma(\mathbf{s}_i, t)} \left(1 + \frac{k}{\sigma(\mathbf{s}_i, t)} z \right)^{-1-1/k} \tag{16}
 \end{aligned}$$

in year t . Here the location \mathbf{s}_i belongs to region j . Correct notation should be \mathbf{s}_{ij} , indicating the i th location belonging to the j th region, but we will avoid notational complication. η_σ is a latent process with regional resolution and

$$\eta_\sigma = (\eta_\sigma(s_{ij}, t), i = 1, \dots, n_j(t), j = 1, \dots, 18, t = 1, \dots, 30),$$

where $\eta_\sigma(s_{ij}, t) = \eta_\sigma(j, t)$, for all s_i belonging to the j th region. This relation will be conveniently represented in level 2 below.

2. **Level 2: link functions**

(a)

$$\log(\sigma(\mathbf{s}_i, t)) = \sigma_0 + \mathbf{h}'(\mathbf{s}_i, t) [\eta_\sigma + \alpha_1 \mathbf{N}_1(t - 1) + \alpha_2 \mathbf{N}_2(t - 1)]. \tag{17}$$

Here η_σ is a spatial process representing a *district* effect at location (\mathbf{s}_i, t) , $\mathbf{N}_1(t - 1)$ and $\mathbf{N}_2(t - 1)$ are (18×1) are (non random) vector of observations, denoting respectively the (normalized) average burned area and (normalized) average number of fires in $t - 3, t - 2$ and $t - 1$ in *district* j , whereas α_1, α_2 are scalar regression parameters. $\mathbf{h}(\mathbf{s}_i, t)$ is a (18×1) non-stochastic vector that relates the log of the scale parameter σ at observation location (\mathbf{s}_i, t) to a *district* effect; thus $\mathbf{h}(\mathbf{s}_i, t)$ is a vector of zeros and ones, with one in the j -th position indicating that the j *district* is associated with the (\mathbf{s}_i, t) observation location.

(b) For mathematical simplicity, we will assume that the shape parameter has no spatial structure, although this is not realistic based on the information we have from the historical data. Theoretically, the shape parameter k can take values in $(-\infty, +\infty)$. However, preliminary data analysis of historical data indicated that the shape parameter has a mode around 0.47. Hence we use as prior for k a vague beta distribution, restricting it to the interval $[0, 1]$. This prior distribution matches with the expert opinion that the fire size distribution should have finite mean but heavy tailed. Hence, we assume

$$k \sim \text{beta}(0.5, 0.5). \tag{18}$$

3. Level 3: parameters, priors, hyper-parameters and hyper-priors

- (a) η_σ is an *Intrinsically* conditional, nearest neighbor Auto Regressive (IAR) model (Besag et al. 1995) defined over the eighteen *districts* (see 6):

$$\eta_\sigma \sim IAR(\tau_\eta). \quad (19)$$

- (b) $\Theta = (\sigma_0, \alpha_1, \alpha_2, \tau_\eta)$ are the model hyper-parameters all having independent prior distributions, as follows:

- i. IAR prior is a pairwise difference prior (Besag et al. 1995) that is identified only up to an additive constant. Thus to identify and intercept term σ_0 in the scale parameter, we must add the constraint $\sum_{j=1}^{18} \eta_\sigma(j, t) = 0$, for all t . Note that in implementing a Gibbs sampler to fit this model, this constraint can be imposed numerically by re-centering each sampled η_σ vector around its own mean following each Gibbs iteration.
- ii. The baseline value for the scale parameter in the log-scale, σ_0 , follows a vague prior:

$$\sigma_0 \sim \text{Uniform}(-\infty, +\infty) \quad (20)$$

- iii. Regression parameters α_l , $l = 1, 2$ have vague independent Gaussian priors:

$$\alpha_l \sim \text{Gau}(0, \hat{\tau}_\alpha), \quad l = 1, 2 \quad (21)$$

where the precision parameter is given by $\hat{\tau}_\alpha = 0.0001$

- iv. In order to maintain the model identifiable, the precision parameter that controls the clustering due to η cannot be chosen to be arbitrarily large. Hence, we suggest

$$\tau_{\eta_\sigma} \sim \text{Ga}(0.0001, 0.0001) \quad (22)$$

We now give some justification for this model:

1. The areal model above reduces the spatial variability to a regional (*district*) level. Hence it is computationally more feasible, at the cost of reducing resolution. In any case, in forestry management, the high resolution to the point of predicting a large fire at any location s_0 is not realistic.
2. We assume that the meteorological and other environmental variables such as spatial distribution of fire-prone vegetation, organizational ability of each district to fight fires etc. are expressed in terms of a hidden, unobserved, spatially colored random effect only on the *district* resolution. Hence although the σ is allowed to vary in time and space, they are assumed to have no variation within each region. The quality of the predictions should increase if part of this random effect is explained by covariates.

3. Assuming that the fire sizes are independent of time is the weakness of this model, since given a fire at one location \mathbf{s} at year t , it is very unlikely that there will be large fire the consecutive year at the same location. However, we expect to capture part of this space-time interaction through the proxy covariates N_1 and N_2 and their respective coefficients. Other weakness of this areal model is that within each region, the excesses are assumed to be independent. We address this weakness later in Sect. 6
4. The IAR model given in (19) results in a improper prior distribution, However, its use is still justified for various reasons:
 - It parameterizes the precision matrix, instead of the covariance matrix, reducing the computational effort.
 - We are only using the IAR model as prior; the posterior will typically emerge as proper, so the inference may still proceed.
 - The IAR model is applied to the spatial random effects and as a consequence, the improper prior may actually enable wider scope for posterior spatial pattern (Banarjee et al. 2004).

The posterior distribution of all the parameters is given by

$$\begin{aligned}
 & p(k, \boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_{30}, \boldsymbol{\eta}_\sigma, \sigma_0, \tau_\eta, \alpha_1, \alpha_2 | \mathbf{Z}, \mathbf{N}_1, \mathbf{N}_2) \\
 & \propto \left\{ \prod_{t=1}^{30} \prod_{j=1}^{18} \prod_{i=1}^{n_j(t)} p(z(\mathbf{s}_i, t) | \sigma(\mathbf{s}_i, t), k) \right. \\
 & \quad \times p(\sigma(\mathbf{s}_i, t) | \sigma_0, \boldsymbol{\eta}_\sigma, \alpha_1, \mathbf{N}_1(t-1), \alpha_2, \mathbf{N}_2(t-1), \mathbf{h}(\mathbf{s}_i, t)) \left. \right\} \\
 & \quad \times p(k) p(\boldsymbol{\eta}_\sigma | \tau_{\eta_\sigma}) p(\tau_{\eta_\sigma}) p(\sigma_0) p(\alpha_1) p(\alpha_2).
 \end{aligned} \tag{23}$$

where \mathbf{Z} represents the observed excesses over u hectares.

The counting process $N_j(t)$, $j = 1, \dots, 18, t = 1, \dots, 30$ will be modeled using the following hierarchy:

$$\begin{aligned}
 & p(n_j(t), t = 1, \dots, 30, j = 1, \dots, 18 | \lambda(j, t), \\
 & j = 1, \dots, 18, t = 1, \dots, 30) = \prod_{j=1}^{18} \prod_{t=1}^{30} \text{Poisson}(\lambda(j, t)), \\
 & \lambda(j, t) \sim \text{Gamma}\left(\frac{1}{v}, \frac{1}{\omega(j, t)v}\right).
 \end{aligned} \tag{24}$$

The mean value of $\lambda(j, t)$ is given by $\omega(j, t)$, and is modeled through the link function:

$$\log(\omega(j, t)) = \omega_0 + \gamma_1 N_{1j}(t-1) + \gamma_2 N_{2j}(t-1) + W(j), \tag{25}$$

where $N_{1j}(t-1)$ and $N_{2j}(t-1)$ are observed covariates in *district* j , as defined previously, whereas γ_1 and γ_2 are scalar regression parameters with independent Gaussian prior distributions:

$$\gamma_l \sim \text{Gau}(0, 0.0001), l = 1, 2,$$

\mathbf{W} is a nearest neighbor IAR model, as defined previously, independent of η_σ , γ_1 and γ_2 :

$$\mathbf{W} \sim IAR(\tau_W), \quad (26)$$

and the baseline value for $\omega(j, t)$ in the log-scale, ω_0 , follows a vague prior:

$$\omega_0 \sim \text{Uniform}(-\infty, +\infty), \quad (27)$$

and the precision parameter τ_W follows a vague Gamma distribution:

$$\tau_W \sim \text{Ga}(0.0001, 0.0001) \quad (28)$$

We assume a vague prior *Gamma*(8, 1.5) for the hyperparameter ν in (24).

The Poisson model given above is over-dispersed (Mendes et al. 2007). We will not give the full report on the posterior distributions of the model parameters for the counting process. However, we report only the latent random factors. Figures 7 and 8 give respectively the posterior means of $\omega(j, t)$ and \mathbf{W} , for $t = 2000$ and $t = 2003$ as well as their posterior standard deviations. We stress again that more complicated models for the counting process, capturing the temporal variation can also be given, but we will not pursue this.

4.2 Areal model II

We also propose a more complex model with spatially varying shape parameter. We add in the second level of the previous model the following link function, assuming that a spatially colored latent processes η_k , independent of η_σ drives the shape parameter:

$$\text{logit}(k(\mathbf{s}_i)) = k_0 + \mathbf{h}'(\mathbf{s}_i, t)\eta_k, \quad (29)$$

where η_k is a nearest neighbor IAR model, as defined previously for η_σ :

$$\eta_k \sim IAR(\tau_{\eta_k}). \quad (30)$$

The baseline value for the scale parameter in the logit scale, k_0 follows a vague prior:

$$k_0 \sim \text{Uniform}(-\infty, +\infty), \quad (31)$$

and the precision parameter τ_{η_k} follows a vague Gamma distribution:

$$\tau_{\eta_k} \sim \text{Ga}(0.0001, 0.0001). \quad (32)$$

Note that, we do not have a temporal variation component in the link function for the scale parameter k in terms of $N_1(t)$ and $N_2(t)$, as we have for the scale parameter. Such model was experimented, but did not converge.

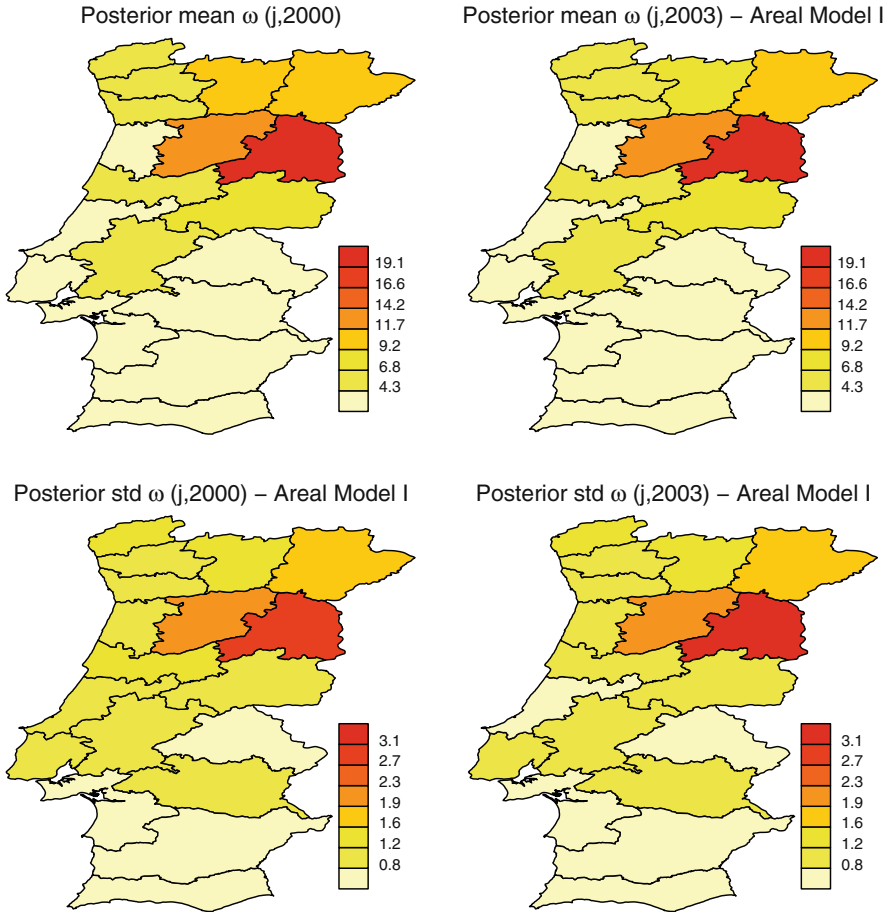


Fig. 7 Posterior mean and std of the $\omega(j, t)$ for $j = 1, \dots, 18$ and $t = 2000, 2003$.

The posterior distribution of all processes and parameters, given the observed data (For this model, the part of the data corresponding to 2004 is excluded from the analysis for model verification purpose):

$$\begin{aligned}
 & p(\mathbf{k}, \sigma_1, \dots, \sigma_{29}, \eta_\sigma, \eta_k, k_0, \sigma_0, \alpha_1, \alpha_2, \tau_{\eta_\sigma}, \tau_{\eta_k} | \mathbf{Z}, \mathbf{N}_1, \mathbf{N}_2) \\
 & \propto \left\{ \prod_{t=1}^{29} \left[\prod_{j=1}^{18} \prod_{i=1}^{n_j(t)} p(z(\mathbf{s}_i, t) | \sigma(\mathbf{s}_i, t), k(\mathbf{s}_i)) \times p(k(\mathbf{s}_i) | k_0, \eta_k, \mathbf{h}(\mathbf{s}_i, t)) \right] \right\} \\
 & \times p(\sigma(\mathbf{s}_i, t) | \sigma_0, \eta_\sigma, \alpha_1, \mathbf{N}_1(t-1), \alpha_2, \mathbf{N}_2(t-1), \mathbf{h}(\mathbf{s}_i, t)) \} \\
 & \times p(k_0) p(\eta_k | \tau_{\eta_k}) p(\tau_{\eta_k}) p(\eta_\sigma | \tau_{\eta_\sigma}) p(\tau_{\eta_\sigma}) p(\sigma_0) p(\alpha_1) p(\alpha_2) \tag{33}
 \end{aligned}$$

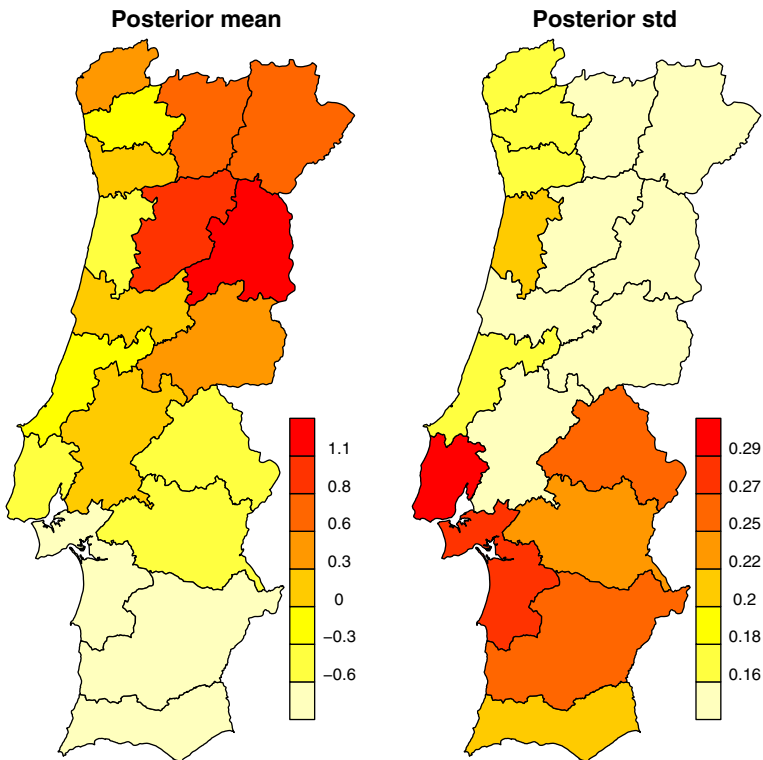


Fig. 8 Posterior mean and std of the W process for $j = 1, \dots, 18$

4.3 Results

One cannot evaluate the previous posterior distributions analytically and must resort to numeric simulation methods. We use the special case of MCMC known as Gibbs sampling (see, for example, [Gilks et al. \(1996\)](#)). The simulation were performed in WinBugs. We got convergence for both models and the main results are reported below.

We run the chain for 50000 iterations and discard the first 5000 interactions where we think the chain has achieved the convergence. Tables 1 and 2 presents the most important Gibbs sampling results for both models. The chain has passed on the Geweke ([Geweke 1992](#)) and [Raftery and Lewis \(1992\)](#) convergence diagnostics and [Heidberger and Welch Stationary test \(Heidberger and Welch 1983\)](#). The [Raftery and Lewis convergence diagnostic](#) suggest to use a thin of 20; hence, we sampled the 45000 available iterations every 20-th samples and Tables 1 and 2 show the marginal posterior results for the parameters of interest.

Table 3 indicates that model II performs better than model I ([Spiegelhalter et al. 2002](#)) as expected.

Figures 9 and 10 show the prior and posterior distribution of univariate parameters of model I and II.

Table 1 Summary of results of Gibbs sampling—areal model I

Parameter	Prior mean (std)	Posterior mean (std)	2.5% CI	97.5% CI
α_1	0 (100)	-0.124 (0.04)	-0.198	-0.052
α_2	0 (100)	0.085 (0.03)	0.02	0.152
k	0.5 (0.35)	0.505 (0.04)	0.431	0.582
σ_0	-	5.565 (0.06)	5.444	5.673
τ_η	1 (100)	3.414 (1.91)	1.167	8.369

Table 2 Summary of results of Gibbs sampling—areal model II

Parameter	Prior mean (std)	Posterior mean (std)	2.5% CI	97.5% CI
α_1	0 (100)	-0.125(0.04)	-0.2	-0.048
α_2	0 (100)	0.082 (0.04)	0.015	0.152
k_0	-	-0.01(0.15)	-0.3	0.319
σ_0	-	5.602 (0.06)	5.499	5.713
$\tau_{\eta\sigma}$	1 (100)	4.267 (2.16)	1.39	10.06
$\tau_{\eta k}$	1 (100)	0.608 (0.62)	0.103	1.945
$k(1)$	-	0.621 (0.15)	0.338	0.933
$k(2)$	-	0.75 (0.22)	0.185	0.994
$k(3)$	-	0.379 (0.13)	0.145	0.674
$k(4)$	-	0.348 (0.11)	0.162	0.588
$k(5)$	-	0.776 (0.1)	0.564	0.956
$k(6)$	-	0.701 (0.12)	0.458	0.926
$k(7)$	-	0.805 (0.14)	0.467	0.988
$k(8)$	-	0.914 (0.09)	0.687	0.999
$k(9)$	-	0.344 (0.08)	0.196	0.505
$k(10)$	-	0.714 (0.14)	0.423	0.959
$k(11)$	-	0.672 (0.21)	0.207	0.973
$k(12)$	-	0.876 (0.09)	0.655	0.991
$k(13)$	-	0.46 (0.12)	0.237	0.748
$k(14)$	-	0.795 (0.11)	0.57	0.972
$k(15)$	-	0.788 (0.17)	0.378	0.992
$k(16)$	-	0.54 (0.15)	0.278	0.865
$k(17)$	-	0.313 (0.09)	0.141	0.507
$k(18)$	-	0.387 (0.1)	0.202	0.603

Table 3 Deviance information criterion (DIC)

Model	Deviance	pD	DIC
Model I	33 260.00	21.38	33 281.38
Model II	33 220.00	26.53	33 246.53

Figures 11 and 12 show the posterior mean and posterior standard deviation of the scale parameter in 2000 and 2003, for model I and II, respectively. It is clear that in both models, although more clear in model II, the posterior mean of the scale parameter increased from 2000 to 2003. This is due, in part, to the heat wave that took place in 2003 and in part to four influential observations which were possibly clusters of fires rather than individual fires. Figure 13 shows the spatial distribution of the posterior mean and posterior standard deviation of the shape parameter, for

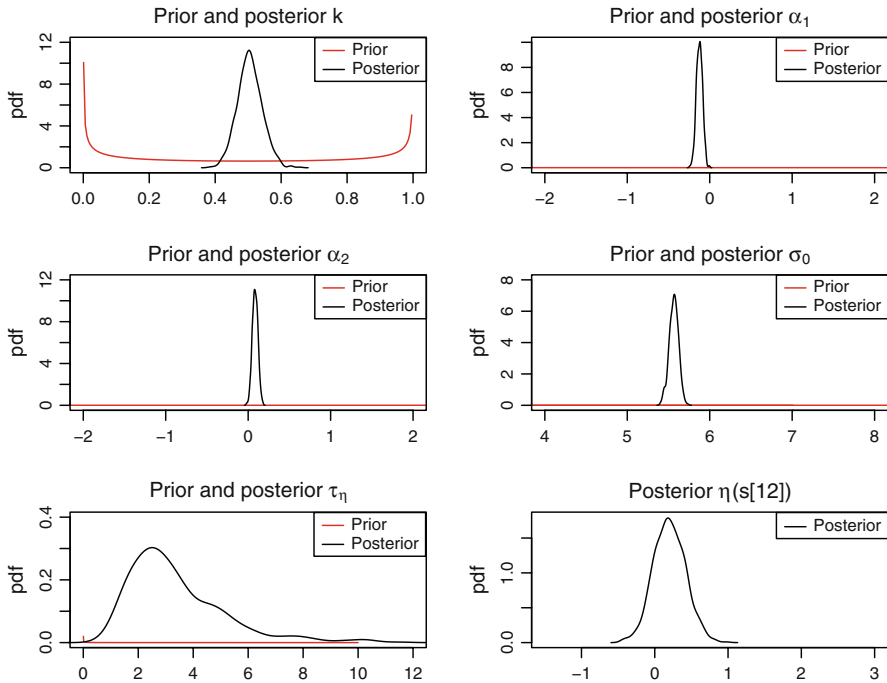


Fig. 9 Priors and posteriors distributions of areal model I

model II. As expected, the *districts* in the center of the country have higher values of k (posterior mean) with lower variability. This is not surprising since these *districts* have the highest fire prone vegetation density as well as great homogeneity of its landscape. We also suspect, but due to lack of climatological data can not prove, that there is a strong effect of ecological-climatological variable interaction on large fire sizes. North of Portugal has the highest density of fire prone vegetation, whereas south of Portugal has the worst combination of climatological factors for wildfires. However, it is the center of Portugal which combines the worst of these two factors to produce extreme fire sizes, as Fig. 13 shows. We believe that later as data on covariates become available, the model and the simulations would make these statements more precise.

Figure 14 shows the posterior mean of the η_σ process for areal model II. Areal models I and II produce similar posterior mean surfaces for the respective η_σ processes. Basically, they divide the country in areas which are consistent with the soil occupation and forest diversity. Figures 9 and 10 clearly show that the data update prior distributions well. Note that the posterior distribution of the parameter α_1 is on the negative quadrant, as expected, but it is surprising to see that α_2 is positive. One would expect that average over the last 3 years of total area burned as well as the total number of fires would have negative correlations with the fire sizes. From Table 2, it is clear that there are regional variability in the shape parameter, hence the size of extreme fires varies from one district to another. Hence, model II is clearly better in explaining the spatial variation of fire sizes.

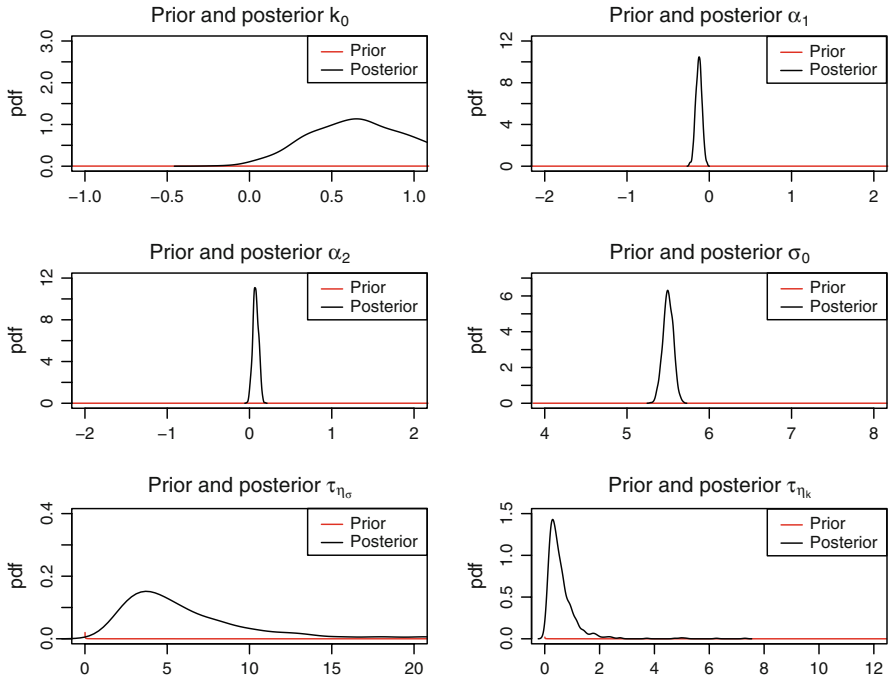


Fig. 10 Priors and posteriors distributions of areal model II

Figure 15 shows the posterior mean and standard deviation of the η_k process. Note that η_k and η_σ processes have different structures, indicating that unobserved meteorological and topological conditions effect the shape and scale parameters in a different fashion.

5 Model assessment

In order to assess the predictive performance of the proposed models, we follow the diagnostic approach of Gneiting et al. (2007), where they suggest performing probabilistic and marginal calibration, as well as assessing the sharpness of the predictive distribution.

5.1 Probabilistic calibration

Probabilistic calibration is done by the probability integral transform (PIT). The PIT value is obtained from $p = F(x)$, where x is a future observation and $F(\cdot)$ is the predictive distribution. If a model is calibrated, then the PIT sequence should be uniformly distributed in $(0, 1)$. However, the converse is not true.

Uniformity is usually assessed in a exploratory sense, and one way of doing this is plotting the histograms of the PIT values for which 10 or 20 bins histogram bins seem to be adequate (Gneiting et al. 2007).

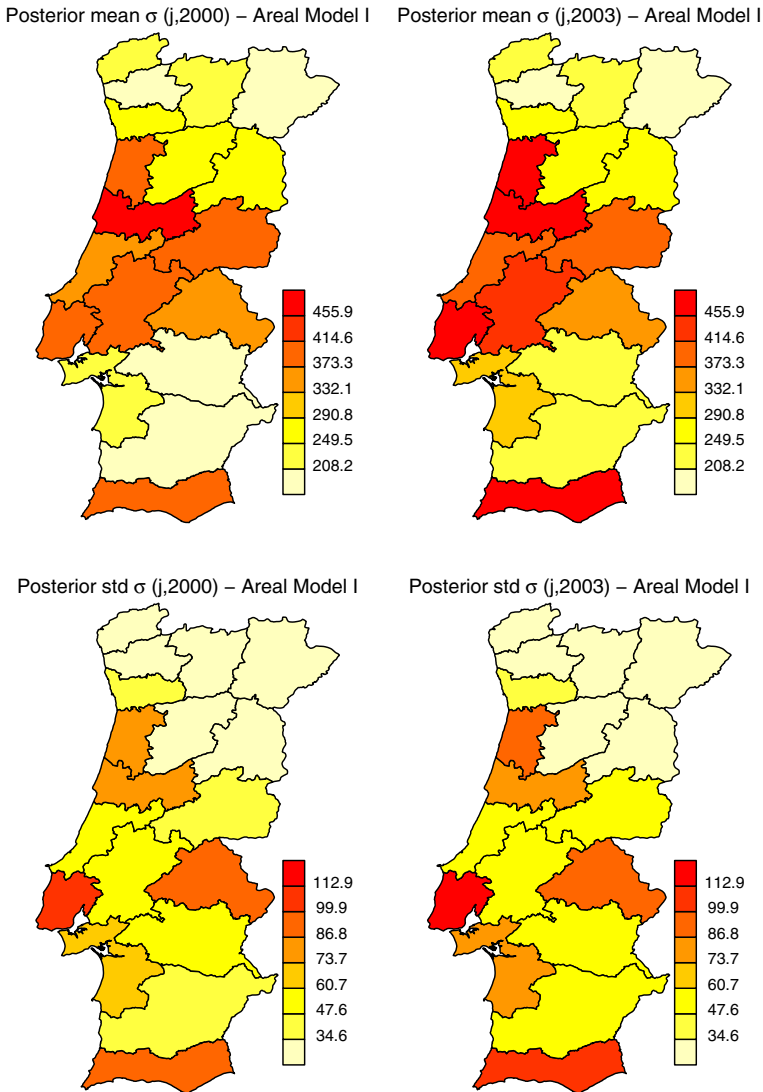


Fig. 11 Posterior mean and std of the $\sigma(j, t)$ for $j = 1, \dots, 18$ and $t = 2000, 2003$, for areal model I

Visual inspection of a PIT histogram can provide hints to the reasons for forecast deficiency. Hump-shaped histograms indicate overdispersed predictive distributions with predictive intervals that are too wide on average. U-shaped histograms often correspond to predictive distributions that are too narrow. Triangle-shaped histograms are seen when the predictive distributions are biased.

In 2004, 71 excesses were observed in 15 (out of 18) districts. Ideally, uniformity of the PIT values should be checked for each of the 18 predictive distributions obtained for each of the 18 regions using the areal model II. However, some regions have very

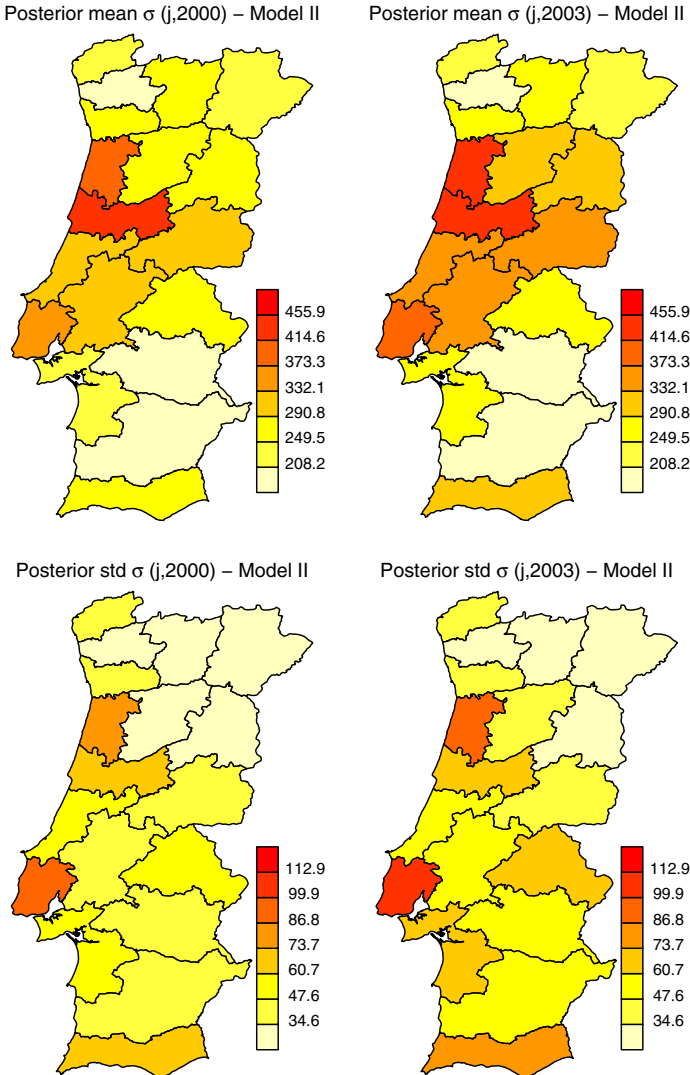


Fig. 12 Posterior mean and std of the $\sigma(j, t)$ for $j = 1, \dots, 18$ and $t = 2000, 2003$, for areal model II

few observed fires with sizes above 250 hectares, making this assessment unreliable. Instead, we will plot all the PIT values $p_{ij}, i = 1, \dots, 18, j = 1, \dots, n_i$ in one graph to obtain a reasonable histogram with sufficient bins, at the cost of introducing bias.

Figure 16 employs 20 bins and shows the PIT histogram for the 71 wildfires sizes (above 250) observed in 2004, but not included in inference.

The histogram (Fig. 16) suggests that the model is underestimating the probability of large fires, as the relative frequency of PIT values above 0.8 is high. Note that by choosing smaller number of bins (5–6 bins instead of 20) it is possible to get a relatively more uniform histogram.

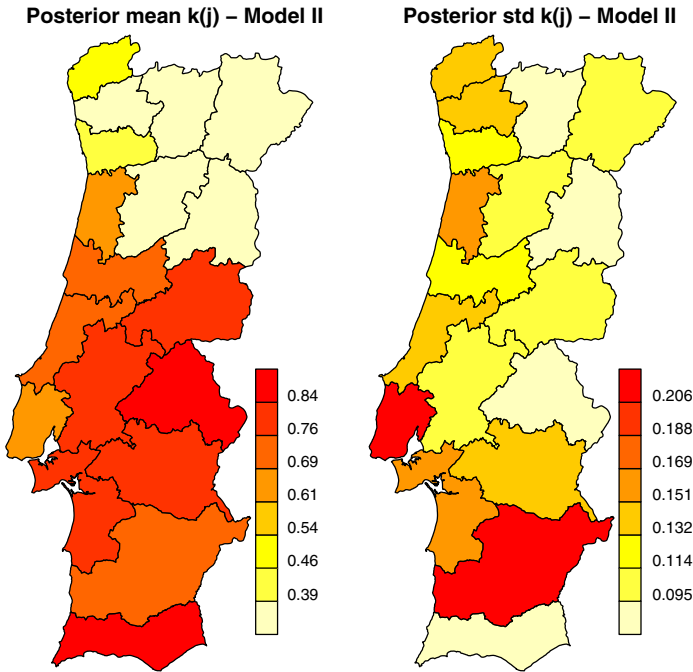


Fig. 13 Posterior mean and std of the $k(j)$ for $j = 1, \dots, 18$, for areal model II

5.2 Marginal calibration

Marginal calibration compares the equality of the predicted versus the observed values. [Gneiting et al. \(2007\)](#) suggest the comparison of the predictive and empirical cumulative distribution functions (CDF). Again, ideally, one should do this comparison based on each of the 18 regions. However, due to the lack of sufficient number of fires with sizes above 250 hectares, we suggest using a weighted predictive and empirical CDF's defined by

- $\bar{F}_t(x)$ is the average predictive CDF at year t :

$$\bar{F}_t(x) = \sum_{j=1}^{18} w_j F_t^j(x), \tag{34}$$

where $F_t^j(x)$ is the predictive CDF at year t in district j . $\bar{F}_t(x)$ is a mixture of the eighteen district individual CDF and w_j is the weight that is proportional to the number of observations at each district in 2004.

- $\hat{G}_t(x)$ is the average empirical CDF of the observations:

$$\hat{G}_t = \frac{1}{18} \sum_{j=1}^{18} \frac{1}{n_j} \mathbf{1}(x_{ij} \leq x), \tag{35}$$

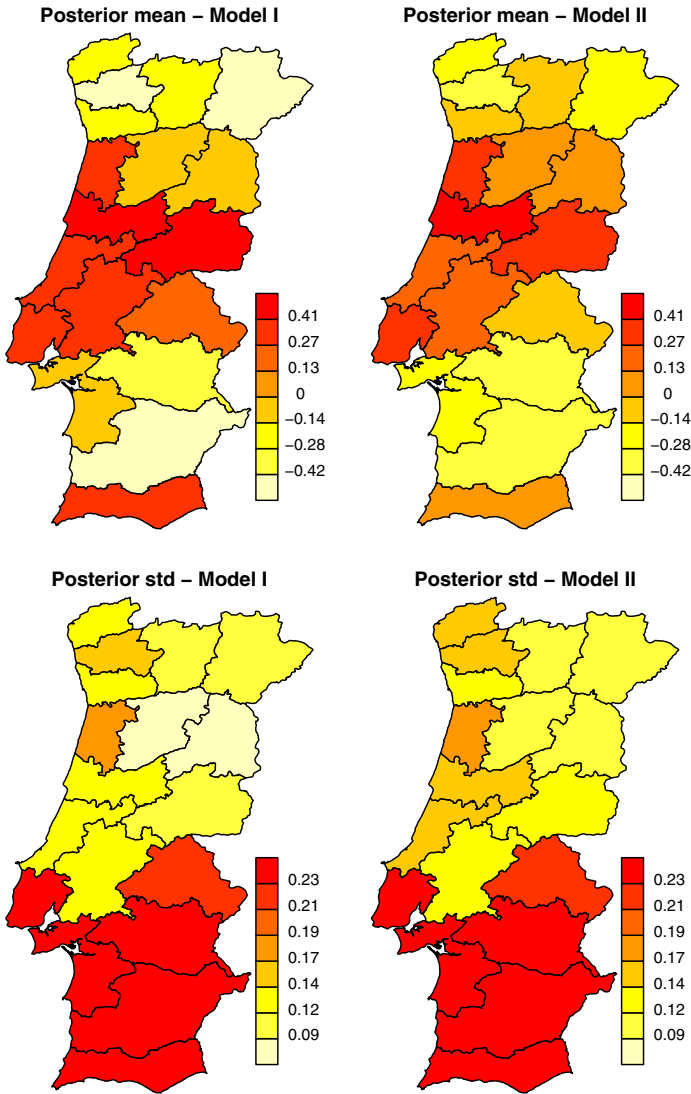


Fig. 14 Posterior mean and std of the η_σ process for $j = 1, \dots, 18$, for areal models I and II

where $\mathbf{1}(x_{ij} \leq x)$ is the indicator function, x_{ij} are the observed excesses in district j , in year t , x is such that belongs to a sequence of values that explores the empirical CDF and n_j is the number of excesses in district j in year t . Table 4 shows the values of \hat{G} for $t = 2004$

The natural graph device in the assessment of marginal calibration is a plot of $\hat{G}_t(x)$ versus $\bar{F}_t(x)$. However is more intuitive to plot the difference of simulations the two CDFs, as in Fig. 17:

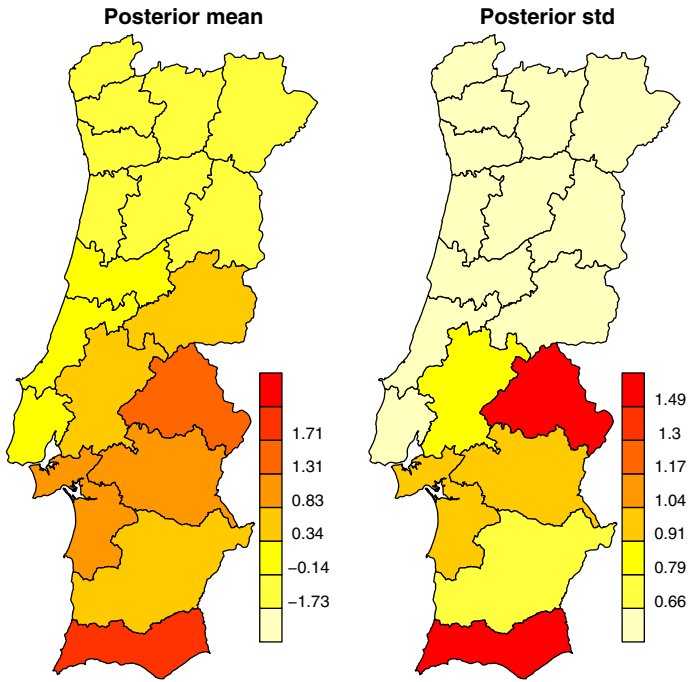


Fig. 15 Posterior mean and std of the η_k process for $j = 1, \dots, 18$, for areal model II

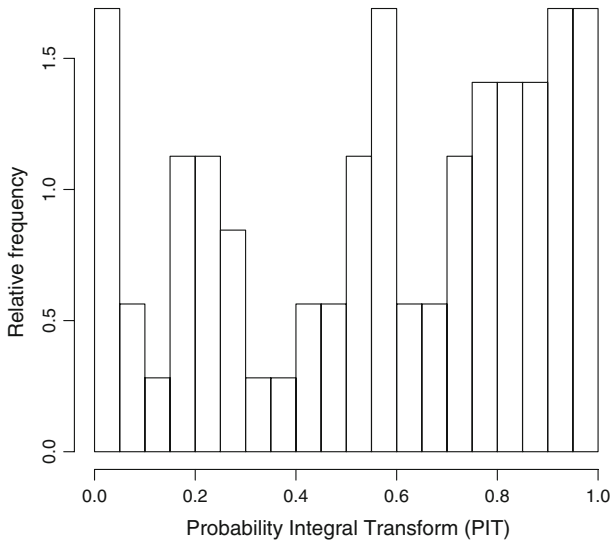


Fig. 16 Probabilistic calibration

Table 4 Values of \hat{G} for $t = 2004$

x	$\hat{G}(x)$
10	0.070
50	0.155
100	0.296
200	0.380
250	0.451
300	0.507
400	0.563
500	0.620
750	0.732
1000	0.817
1500	0.915
2000	0.944
2500	0.944
5000	0.972
10000	0.972
20000	0.986
30000	1.000

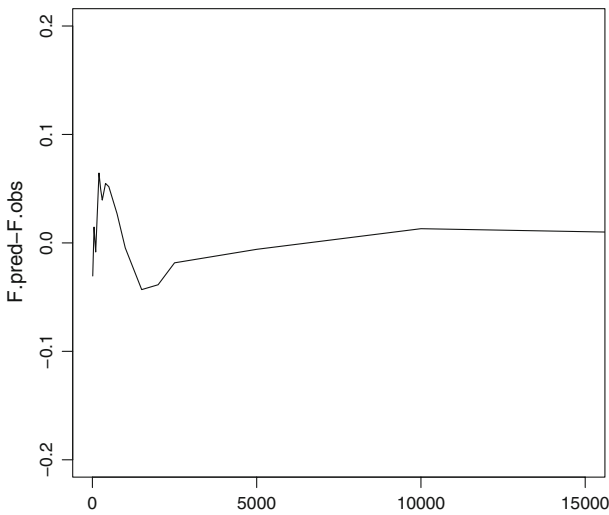


Fig. 17 Marginal calibration

$$\bar{F}_t(x) - \hat{G}_t(x). \tag{36}$$

Under the hypothesis of marginal calibration, we expect minor fluctuations about 0.

Ideally, the graph should be a straight line. Note that the difference is mostly positive again indicating that the model is underestimating the probability of large fires.

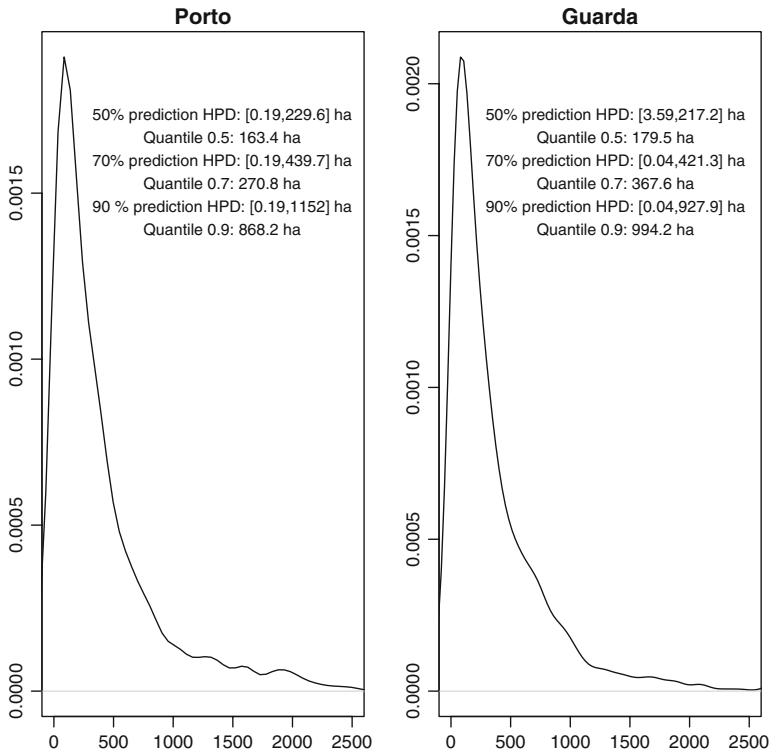


Fig. 18 Sharpness of the predictive density for two different observations in 2004

5.3 Sharpness

Sharpness is the degree of the concentration of the predictive distribution and more concentrated predictive distributions give sharper predictions. To assess sharpness we use graphical summaries and the width of the central 50%, 70% and 90% prediction intervals. Figure 18 shows the kernel estimates of the predictive densities for two different regions, namely Porto and Guarda in 2004, as well as the 50%, 70% and 90% prediction intervals and quantiles. Predictive distributions of other regions are similar. Note that, as expected, the predictive distributions of excess fire sizes are heavy tailed and sharpness (or lack of it) should be judged within this context. We believe that the predictive distributions are sharp considering the heavy tailed behaviour of the fire excesses.

6 Discussion

In this paper, we made an attempt to quantify the risk of large fire sizes, taking into account the spatial heterogeneity. Bayesian hierarchical models were used to quantify the spatial variation by introducing a spatially colored random effects in the model parameters. However, these models can be improved in several directions:

1. Ideally, the data should be considered as geo-statistical, and point referenced models should be used to account for local dependencies and variations in a much efficient manner. The areal model we use can not take into consideration local spatial dependencies within each region which certainly exist. Ideally, one should include in the model continuous latent processes $W_i(s)$, specific to each region, to take into consideration local topology and other local factors to explain dependency and heterogeneity in higher resolutions. This is work will be reported elsewhere.
2. Outliers and influential observations should be checked very carefully to assess their role in the model. It is believed that in some cases, several fires with different origins are reported as a large merged fire cluster. Four observations from 2003 are an example of this situation. These observations are highly influential and there is evidence to suggest that they are indeed fire clusters.
3. Study of annual fires sizes indicate that there is no clear trend like structure in the large fire size risks over time. However, there is an indication of a cyclical movement of 3 years, possibly corresponding to the regeneration cycle of fire prone vegetation. It may be interesting to look at this point with more attention, in order to understand how regeneration cycle of fire prone vegetation effects extreme fires.

Acknowledgements We thank the referees for their careful reading of the manuscript and their valuable comments. This work is partially supported by FCT, POCTI and Research project PTDC/MAT/64353/2006.

References

- Banarjee S, Carlin BP, Gelfand A (2004) Hierarchical modelling and analysis for spatial data. Chapman and Hall
- Besag J, Green P, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems (with discussion). *Stat Sci* 10:3–66
- Breiman L, Friedman JH, Olsen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Pacific Grove, California
- Casson E, Coles S (1999) Spatial regression models for extremes. *Extremes* 1:449–468
- Coles S, Tawn J (1996) Modelling extremes of the areal rainfall process. *J Roy Stat Soc B* 58(2):329–347
- Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. *J Am Stat Assoc* 102(479):824–840
- Davis RA, Mikosch T (2006) Extreme value theory for space-time processes with heavy tailed distributions, Preprint towards a geostatistics of extremes unpublished report, presented at the workshop on statistical extremes and environmental risk. Lisbon, Portugal
- de Haan L, Lin T (2001) On convergence toward an extreme value distribution in $C[0, 1]$. *Ann Probab* 29:467–483
- de Haan L, Pereira T (2006) Spatial extremes: models for the stationary case. *Ann Stat* 34:146–168
- de Zea Bermudez P, Kotz S (2006) Parameter estimation of the generalized pareto distribution. Technical report No.6, CEAUL, University of Lisbon
- de Zea Bermudez P, Mendes JM, Pereira J, Turkman KF (2007) Spatial extremes of wildfire sizes I. Research Report, No.2, CEAUL, University of Lisbon
- Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, Netherlands
- Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo JM, Berger JO, David AP, Smith AFM (eds) Bayesian statistics 4. Oxford University Press, Oxford

- Gilks WR, Richardson S, Spiegelhalter DJ (eds) (1996) Markov chain Monte Carlo in practice. Interdisciplinary statistics. Chapman & Hall, London
- Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *J Roy Stat Soc B* 69:243–268
- Heidelberger P, Welch P (1983) Simulation run length control in the presence of an initial transient. *Oper Res* 31:1109–1144
- Ledford AW, Tawn J (2003) Diagnostics for dependence within time series extremes. *J Roy Stat Soc B* 65(2):521–543
- Mendes JM, Turkman KF, Jardim E (2007) A Bayesian hierarchical model for over-dispersed count data: a case study for abundance of hake recruits. *Environmetrics* 18:27–53
- Paulino CD, Turkman MAA, Murteira B (2003) *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisbon
- Pickands J (1975) Statistical Inference using extreme order statistics. *Ann Stat* 3:119–131
- Raftery AL, Lewis S (1992) How many iterations in the Gibbs sampler? In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) *Bayesian statistics 4*. Oxford University Press, Oxford
- Ramesh NI, Davison A (2002) Local models for exploratory analysis of hydrological extremes. *J Hydrol* 256:106–119
- Sang H, Gelfand AE (2007) Hierarchical modeling for extreme values observed over space and time. Research Report, Duke University
- Smith R (1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat Sci* 4:367–393
- Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde (2002) Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 64:583–640
- Turkman KF (2006) A note on the extremal index for space-time processes. *J Appl Probab* 43:114–126
- Tzeng SL, Huang HC, Cressie N (2005) A fast optimal spatial prediction method for massive data sets. *J Am Stat Assoc* 472:1343–1357

Author Biographies

Jorge M. Mendes is currently Assistant Professor at ISEGI of the New University of Lisbon. Research interests include Environmental Statistics and Bayesian hierarchical modelling. He is a Doctor of Philosophy in Statistics and Operational Research (Ph.D.). Besides his teaching activities is Director of the Methodological Department at Statistics Portugal.

Patrícia Cortés de Zea Bermudez is currently Assistant Professor and got her degree in Applied Mathematics in 1990, did her M.Sc. in Probability and Statistics-Time Series Analysis in 1994 and the Ph.D. in Statistics and Operations Research—Bayesian Statistics and Extreme Value Theory in 2003. Her research interests focus on Extreme Value Applications, Bayesian Statistics and Epidemiology.

José Pereira is full Professor, Department of Forestry, Institute of Agronomy, Technical University of Lisbon. Has an undergraduate degree in Forestry and a Ph.D. in Renewable Natural Resources (University of Arizona). His main research interests are in the macroecology of vegetation fires, and remote sensing of global biomass burning.

K. F. Turkman is full Professor, Department of Statistics and Operations Research, University of Lisbon. Has an undergraduate degree in Mathematics and Ph.D. in Probability (University of Sheffield, UK). His main interests are in time series analysis and extreme value theory.

M. J. P. Vasconcelos has an undergraduate degree in Forestry, (Instituto Superior de Agronomia, Lisboa) and a Ph.D. in Natural Resources Studies (University of Arizona, Tucson). She is an Auxiliary Researcher at the Tropical Research Institute (Lisbon), working on remote sensing of tropical land cover change and its impacts on carbon sequestration.