

# A Bayesian spatio-temporal analysis of forest fires in Portugal

## Un análisis espacio-temporal Bayesiana de incendios forestales en Portugal

GIOVANI L. SILVA<sup>a</sup>, MARIA INÊS DIAS<sup>b</sup>

CEAUL & DM, INSTITUTO SUPERIOR TÉCNICO, UNIVERSIDADE TÉCNICA DE LISBOA, LISBOA, PORTUGAL  
CIMA, DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDADE DE ÉVORA, ÉVORA, PORTUGAL

---

### Abstract

In the last decade, forest fires have become a natural disaster in Portugal, causing great forest devastation, leading to both economic and environmental losses and putting at risk populations and the livelihoods of the forest itself. In this work, we present Bayesian hierarchical models to analyze spatio-temporal fire data on the proportion of burned area in Portugal, by municipalities and over three decades. Mixture of distributions was employed to model jointly the proportion of area burned and the excess of no burned area for early years. For getting estimates of the model parameters, we used Monte Carlo Markov chain methods.

**Key words:** Forest fires, Spatio-temporal data, Bayesian analysis.

### Resumen

En la década pasada, los incendios forestales se han convertido en un desastre natural en Portugal, causando la gran devastación del bosque, llevando a las pérdidas económicas y ambientales y poniendo las poblaciones en peligro y los sustentos del bosque sí mismo. En este trabajo, presentamos modelos jerárquicos Bayesianos para analizar datos de incendios forestales espacio-temporales sobre la proporción de área quemada en Portugal, por los municipios y durante tres décadas. Mezcla de distribuciones fue empleada para modelar en común la proporción de área quemada y el exceso de ninguna área quemada por años. Para conseguir estimaciones de los parámetros del modelo, utilizamos métodos de la cadena de Markov Monte Carlo.

**Palabras clave:** Incendios forestales, Datos espacio-temporales, Análisis Bayesiana.

## 1. Introducción

According to the National Forestry Authority (*Direcção Geral dos Recursos Florestais*), Portugal has the largest number of forest fires among five Mediterranean countries (Portugal, Spain, France, Italy and Greece). In order to look for spatio-temporal patterns of fires, we can model the proportion of burned area ( $Y$ ), which is a (0,1)-restricted continuous variable, assuming naturally a beta distribution (Ferrari & Cribari-Neto 2004) or Gaussian distribution and a Skew-Normal (Azzalini & Dalla Valle 1996) distributions after a *logit* transformation, *i.e.*  $\log(Y/(1-Y))$ . In addition, we can use Bayesian hierarchical models to take into account spatially correlated random effects (Silva et al. 2008) and excess zeros in the proportion of burnt area by municipalities and years (Amaral-Turkman et al. 2011). Our aim is to present a spatio-temporal analysis of forest fires in 278 Portuguese municipalities between 1980 and 2006, from a Bayesian perspective and using Monte Carlo Markov chain (MCMC) methods to make inference on the parameters of interest.

---

<sup>a</sup>Assistant professor. E-mail: gsilva@math.ist.utl.pt

<sup>b</sup>Teaching assistant. E-mail: misd@uevora.pt

## 2. Spatio-temporal modeling

Let  $Y_{it}$  the proportion of burned area in municipality  $i$  and year  $t$ ,  $i=1, \dots, n$ ,  $t=1, \dots, T$ . Assume  $Y_{it}$  or  $\log(Y_{it}/(1-Y_{it}))$  has a probability distribution with mean  $\mu_{it}$  and variance  $\sigma^2$ . Silva et al. (2008) suggest that  $\mu_{it}$  can be expressed by

$$\mu_{it} = \alpha + S_0(t) + S_i(t) + \phi_i, \quad (1)$$

where  $S_0(t)$  can represent a nonlinear temporal effect,  $S_i(t)$  is the temporal effect by region  $i$  and  $\phi_i$  a random effect of the spatial variation associated with region  $i$ . If  $\phi_i = b_i + h_i$ , component  $h_i$  represents the unstructured spatial random effect with Gaussian priori distribution, *i.e.*,

$$h_i \sim N(0, \sigma_h^2 \equiv \tau_h^{-1}), \quad (2)$$

and  $b_i$  the spatially correlated random effect with priori distribution,  $p(b_i | \tau_b = \sigma_b^{-2})$ , chosen in terms of a conditional autoregressive model (CAR) (Besag et al. 1991), *i.e.*,

$$b_i | \mathbf{b}_{-i}, \sigma_b^2 \sim N(\bar{b}_i, \sigma_b^2/m_i), \quad (3)$$

where  $\bar{b}_i$  is the mean of the random effects related to the “neighbors” of the region  $i$ ,  $m_i$  the number of adjacent regions to region  $i$  and  $\sigma_b^2$  the variance component.

Upon the occurrence of zeros, the distribution of the proportion of area burned ( $Y_{it}$  is considered a mixture of distributions with probability function  $f(y_{it})$ , denoting  $f_1(y_{it}) = f(y_{it} | y_{it} \neq 0)$ ,  $i=1, \dots, n$ ,  $t=1, \dots, T$ . Define  $V_{it}$  as a Bernoulli random variable such that,  $V_{it} = 0$ , with probability  $p_{it_0}$ , and 1, with probability  $p_{it_1} \equiv 1 - p_{it_0}$ , where  $p_{it_0}$  represents the probability of non-burned area in the region  $i$  in the year  $t$ .  $V_{it}$  indicates the existence of the burnt area in the region  $i$  in the year  $t$ . Thus,

$$f(y_{it}) = f_1(y_{it})^{V_{it}} (1 - p_{it_0})^{V_{it}} p_{it_0}^{1-V_{it}}. \quad (4)$$

The probability of no burned area in the region  $i$  at time  $t$  is modeled as,

$$\log\left(\frac{p_{it_0}}{1 - p_{it_0}}\right) = \beta_0 + \beta_1 t + \psi_i, \quad (5)$$

where  $\psi_i$  is also a CAR model. We use assigned highly dispersed but proper priors. In fact, one typically assumes independent normal prior for the regression coefficients. For the variance component hyperparameters, one usually assigns an inverse gamma prior, *e.g.*,  $\sigma^2 \sim IG(r_1, s_1)$ ,  $\sigma_b^2 \sim (r_2, s_2)$ ,  $\sigma_h^2 \sim IG(r_3, s_3)$  and  $\sigma_\psi^2 \sim IG(r_4, s_4)$  with kernel density given for  $x^{-(r+1)} \exp(-s/x)$ ,  $x > 0$ . Consequently, we can construct the related joint posteriori distribution and use MCMC methods because the corresponding marginal posteriors are not easy to get explicitly. Notice that these methods are implemented *e.g.* in WinBUGS (Lunn et al. 2000).

## 3. Forest fires data analysis

Based on the models in section 2, we analyze the proportion  $Y_{it}$  of burnt area due to forest fires in 278 municipalities (mainland Portugal) and over 27 years (1980-2006). Data were collected by Portuguese National Forestry Authority. Three scenarios were considered for the data modeling:

- A) Gaussian probability model:  $\text{logit}(Y) \sim N(\mu, \sigma^2)$ ;
- B) Skew-normal model:  $\text{logit}(Y) \sim SN(\mu, \sigma^2, \lambda)$ , where  $\lambda$  is a shape parameter;
- C) Beta model:  $Y \sim \text{Beta}(a, b)$ , with  $E[Y] = \mu$ ,  $\text{Var}(Y) = \frac{\mu(1-\mu)}{\gamma+1}$  and  $\gamma = a+b$ .

By using MCMC methods via WinBUGS, we used 15,000 iterations for all fitted models, taking every 10th iteration of the simulated sequence, after 5000 iterations of burn-in. The model comparison can be based on the Deviance Information Criterion ( $DIC$ ), which handles hierarchical Bayesian models of any degree of complexity, and is computed as the sum of two components: the expected posterior deviance ( $\overline{D}$ ) and the effective number of parameters ( $p_D$ ), measuring the goodness of fit and complexity of the model, respectively (Spiegelhalter et al. 2002). It is often expressed as

$$DIC = 2\overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}), \quad (6)$$

where  $\overline{D(\boldsymbol{\theta})}$  and  $\overline{\boldsymbol{\theta}}$  denote the posterior mean of the deviance and the model parameter vector  $\boldsymbol{\theta}$ , respectively. Though we rely principally on this measure for assessing models in our application, the other measures are also computed for comparison. In table 1, one can be observed some fitted models and, based on (6), the selected model is model  $M_4$ . Note that  $S_0(t) = \eta_t$ , in model  $M_4$ , represents a second order random walk.

TABLE 1: Model selection based on DIC.

	Model	$p_D$	DIC ( $\times 10^6$ )
$M_1(A)$	$\mu_{it} = \beta_0 + \beta_1 t + \phi_i t + b_i + h_i$ $\text{logit}(p_{it}) = \delta_0 + \delta_1 t + a_i$	521	150.150
$M_2(B)$	$\mu_{it} = \beta_0 + \beta_i t + \phi_i t + b_i + h_i$ $\text{logit}(p_{it}) = \delta_0 + \delta_1 t + a_i$	509	150.150
$M_3(C)$	$\text{logit}(\mu_{it}) = \beta_0 + \beta_1 t + \phi_i t + b_i + h_i$ $\text{logit}(p_{it}) = \delta_0 + \delta_1 t + a_i$	581	149.996
$M_4(C)$	$\text{logit}(\mu_{it}) = \beta_0 + \eta_t + b_i$ $\text{logit}(p_{it}) = \delta_0 + \delta_1 t + a_i$	411	149.995

For selected model ( $M_4$ ), the posteriori mean, standard deviation (SD) and 95% highest posterior density (HPD) credible intervals (CI) of some parameters of interest are in table 2. Based on model  $M_4$ , the spatio-temporal risks of burned area, defined here by  $\exp(\eta_t + b_i)$  for municipality  $i$ , were used to produce maps in 1985, 1994 and 2001 in figure 1.

TABLE 2: Estimates of the model parameters ( $M_4$ ).

Parameter	Mean	SD	95% HPD CI
$\delta_1$	-0.169	0.007	(-0.183, -0.156)
$\gamma$	24.82	0.449	(24.02, 25.69)
$\sigma_b^2$	0.334	0.051	(0.237, 0.437)
$\sigma_\eta^2$	3.357	0.508	(2.424, 4.379)
$\sigma_a^2$	0.194	0.060	(0.098, 0.313)
$p_{it_0}$	0.143	0.003	(0.137, 0.150)

## 4. Concluding remarks

The spatio-temporal analysis of the burned area proportion in 278 municipalities of mainland Portugal between 1980 and 2006 reveals an increasing trend in the proportion of burned area, whereas the number of municipalities without burned area trend to decrease. The space-time models studied here have smoothed estimates used in the production of maps that are useful in the interpretation of spatio-temporal data. This analysis of the Portuguese forest fires may isolate trends in small areas of administrative knowledge for promoting an appropriate policy interventions to reduce that national catastrophe.

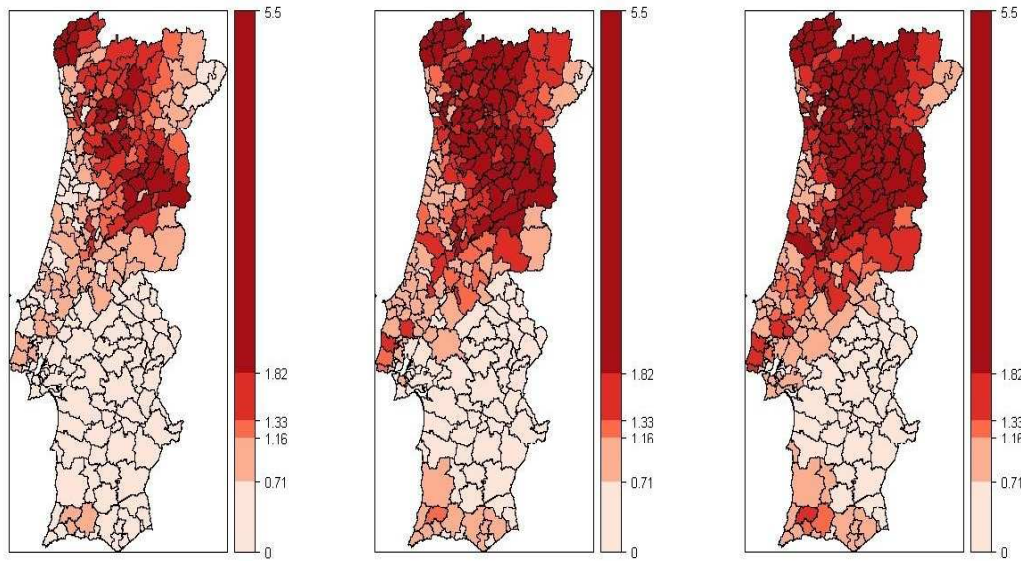


FIGURE 1: Spatio-temporal risks in 1985 (left), 1994 (middle) and 2001 (right)

## Acknowledgements

This paper was partially supported by Pest-OE/MAT/UI0006/2011.

## References

- Amaral-Turkman, M. A., Turkman, K. F., Le Page, Y. & Pereira, J. M. C. (2011), ‘Hierarchical space-time models for fire ignition and percentage of land burned by wildfires’, *Environmental and Ecological Statistics* **18**(4), 601–617.
- Azzalini, A. & Dalla Valle, A. (1996), ‘The multivariate skew-normal distribution’, *Biometrika* **83**(4), 715–726.
- Besag, J., York, J. C. & Mollié, A. (1991), ‘Bayesian image restoration, with two applications in spatial statistics (with discussion)’, *Annals of the Institute of Statistical Mathematics* **43**(1), 1–59.
- Ferrari, S. L. P. & Cribari-Neto, F. (2004), ‘Beta regression for modeling rates and proportions’, *Journal of Applied Statistics* **31**(7), 799–815.
- Lunn, D., Thomas, A., Best, N. & Spiegelhalter, D. (2000), ‘WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility’, *Statistics and Computing* **10**(4), 325–337.  
\*<http://www.mrc-bsu.cam.ac.uk/bugs/>
- Silva, G. L., Dean, C. B., Niyonsenga, T. & Vanasse, A. (2008), ‘Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines’, *Statistics in Medicine* **27**(13), 2381–2401.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002), ‘Bayesian measures of model complexity and fit’, *Journal of the Royal Statistical Society B* **64**(4), 583–639.