

# Modelling and analysis of forest fire data in Portugal<sup>1</sup>

Giovani Silva<sup>2</sup>, Manuela Oliveira and José Borges

Universidade Técnica de Lisboa (IST) and CEAUL, Universidade de Évora and CIMA,  
Universidade Técnica de Lisboa (ISA) and Centro de Estudos Florestais,  
e-mail: gsilva@math.ist.utl.pt, mmo@uevora.pt, joseborges@isa.utl.pt

**Abstract:** In the last decade forest fires became a serious problem in Portugal due to different issues such as climatic characteristics and nature of Portuguese forest. In order to analyse forest fire data, we use generalized linear models for modeling the proportion of burned forest area. Our goal is to find out fire risk factors that influence that proportion of burned area and what may make a forest type susceptible or resistant to fire. Then, we analyse forest fire data in Portugal during 1990-1994 through frequentist and Bayesian approaches.

**Keywords:** Fire management, Burned area proportion, Forest fire data, Generalized linear model.

## 1. Introduction

In Portugal, forest fires and the related burned area have been increasing in the last years, as opposed to other southern European countries. Fire is indeed an important issue in Mediterranean region affecting the ecological and economic aspects of forest areas and causing loss of human life. Many factors have contributed to the increasing number of forest fires, *e.g.*, climate change. González and Pukkala (2007) identified changes in the number of fires, burned area and fire size distribution depending on topographical variables and vegetation type in a Spanish region.

The motivating application is the analysis of forest fire data in the entire Portuguese mainland between 1990 and 1994. There were 5,706 fires in Portugal during that period burning about 4.97% of the country area. Figure 1 displays the distribution of these fires identifying high and critical fire risk zones that are specially located in the northern and central interior of Portugal. Gomes (2006) pointed out many causes and consequences of forest fires in Portugal. Namely, Portuguese forest is mainly based on a monoculture of pine and eucalyptus, which are highly combustible species due to their essential oils.

For analyzing forest fire data, generalized linear models (GLM) are usually adopted (Amaral-Turkman and Silva, 2000). The goal of this work is to model the proportion of burned area analyzing high dimensioned forest fire data in Portugal. The rest of the article is organized as follows. Section 2 succinctly describes a few GLM for the burned area proportion. In Section 3 we present the corresponding analysis, results and discussion related to model selection for the burned area in Portugal from the point-of-view of classical and Bayesian methods.

---

<sup>1</sup>The present paper is partially supported by Fundação para a Ciência e a Tecnologia (FCT).

<sup>2</sup>Address of correspondence: Dep. Matemática - IST, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal.

**Figure 1:** *Distribution of forest fires in Portugal (1990-1994).*



## 2. Statistical model

Portuguese mainland covers about 90,000 km<sup>2</sup> in southern Europe. Most of the country is included in the Mediterranean region and the altitudes range from sea level to about 2,000 *m*. For the analysis of the forest fires in Portugal, the forest was initially divided into classes of altitude (*m*), slope (%), slope orientation or aspect, population (hab/km<sup>2</sup>), proximity to roads, number of days with precipitation greater than 1 *mm*, number of days with maximum temperature higher than 25°C, and fuel (Table 1).

| Proximity                                                                                                                                                                                                          | Population | Slope | Altitude | Aspect | Precipitation | Temperature |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-------|----------|--------|---------------|-------------|
| ≥ 1000                                                                                                                                                                                                             | < 25       | 0–10  | < 200    | plane  | 0–6           | 0–3         |
| < 1000                                                                                                                                                                                                             | 25–100     | 10–20 | 200–400  | north  | 7–13          | 4–48        |
|                                                                                                                                                                                                                    |            | 20–30 | 400–700  | east   | 14–18         | 49–71       |
|                                                                                                                                                                                                                    | > 100      | > 30  | > 700    | south  | 19–22         | 72–92       |
|                                                                                                                                                                                                                    |            |       |          | west   | 23–26         | 93–112      |
|                                                                                                                                                                                                                    |            |       |          |        | ≥ 27          | ≥ 113       |
| Fuel: no fuel, annual crop, eucalyptus, hardwoods, hardwoods and softwoods mixed with eucalyptus (HSME), agro-forestry, permanent crop, shrubs, resinous or softwoods (RS), softwoods mixed with eucalyptus (SME). |            |       |          |        |               |             |

**Table 1:** *Description of the classes used in the forest fire study.*

Secondly, we record the observed proportion of burned forest area, denoted by  $r_i$  that is the burned area out of total area for the  $i$ th combination of levels for the covariates in study,  $i = 1, \dots, k$ . There are several GLM to model the proportion of burned area from these eight underlying covariates. For instance, using the *logit* transformation of  $r_i$ , one

has

$$y_i \equiv \log(r_i/(1 - r_i)) = \mathbf{z}_i' \boldsymbol{\beta} + \epsilon_i, \quad (1)$$

where the random errors  $\epsilon_i$ ,  $i = 1, \dots, k$ , may be considered (independent) Gaussian random variables with mean zero and variance  $\sigma^2$ , and  $\boldsymbol{\beta}$  is the regression coefficient vector associated with the observed covariate vector  $\mathbf{z}_i$ . Notice that the eight covariates in Table 1 give rise to thirty two dummy variables.

Other transformations may be adopted for  $r_i$ , such as,  $\arcsin(\sqrt{r_i})$  and Box-Cox transformations. One can make inference on regression parameter  $\boldsymbol{\beta}$  in (1) based on classical (Amaral-Turkman and Silva, 2000) and Bayesian (Banerjee *et al.*, 2004) methods for GLM. For the latter, the prior distributions may be flat but proper Gaussian and inverse gamma priors for regression parameters and variance, respectively.

### 3. Results and discussion

Let  $M_1$  denote model (1) with all covariates showed in Table 1, whereas  $M_2$  and  $M_3$  represent the corresponding models with  $\arcsin(\sqrt{r_i})$  and Box-Cox transformations, respectively. The Akaike Information Criterion (AIC) values for the models  $M_1$ ,  $M_2$  and  $M_3$  are respectively 50350, 58815 and 57613, pointing to the  $M_1$  model. We also fitted and compared other reduced models for example the model  $M_4$ , which is the model  $M_1$  without the covariates slope orientation, precipitation and maximum temperature.

In addition, we explore a Bayesian approach for model  $M_1$ , assuming Gaussian prior with mean zero and variance  $10^6$  for the regression parameters and inverse gamma with shape and scale parameters equal to 0.001 for the variance  $\sigma^2$ . Based on Deviance Information Criterion (DIC), the best model is  $M_1$  (DIC=149011) in comparison with  $M_4$  (DIC=152481). Then, we decided to select model  $M_1$  because DIC is a generalization of AIC that handles hierarchical models of any degree of complexity (Banerjee *et al.*, 2004). For simplicity, Table 2 only displays the model parameter estimates for  $M_1$  from a Bayesian perspective: posterior mean and 95% credible intervals (CI) for the regression parameters and variance  $\sigma^2$ . Note that Markov chain Monte Carlo (MCMC) samples of size 10,000 were obtained for  $M_1$ , taking every 10th iteration of the simulated sequence, after 5,000 iterations of burn-in (Spiegelhalter *et al.*, 2007). A study of convergence of the samples was carried out using several diagnostic methods and none of them showed any worrying features.

According to model  $M_1$  (Table 2), there is only no significant effect of annual and permanent crop fuels and precipitation 14–18 on the proportion of burned forest area in Portugal during 1990-1994. The proportion tends to increase with slope, altitude, road proximity and even precipitation, whereas population, aspect and temperature display a decreasing effect in the posterior mean of  $y_i$ . The areas with larger likelihood to have forest fires are (in increasing order) softwoods mixed with eucalyptus (SME), eucalyptus, shrubs, resinous or softwoods (RS), hardwoods and softwoods mixed with eucalyptus (HSME), hardwoods, and agro-forestry.

This preliminary analysis of forest fire data in Portugal helps us to figure out the influence of the observed combinations of risk factors ( $k = 25, 388$ ) on the proportion of burned forest area. However, further research is being developed for capturing the spatio-temporal effect on the proportion (Amaral-Turkman *et al.*, 2008) or using more proper distributions and link functions, *e.g.*, Beta regression (Ferrari and Cribari-Neto, 2004).

| Parameter  | Mean   | 95% CI |        | Parameter      | Mean   | 95% CI |        |
|------------|--------|--------|--------|----------------|--------|--------|--------|
| intercept  | 10.120 | 9.524  | 10.72  | Precipitation  |        |        |        |
| Road prox. |        |        |        | 7–13           | -0.225 | -0.409 | -0.031 |
| < 1000     | 0.133  | 0.020  | 0.247  | 14–18          | 0.116  | -0.075 | 0.302  |
| Population |        |        |        | 19–22          | 0.784  | 0.586  | 0.978  |
| 25–100     | -0.311 | -0.575 | -0.263 | 23–26          | 0.590  | 0.379  | 0.796  |
| ≥ 100      | -0.423 | -0.436 | -0.179 | ≥ 27           | 0.811  | 0.570  | 1.039  |
| Slope      |        |        |        | Temperature    |        |        |        |
| 10–20      | 1.025  | 0.902  | 1.148  | 4–48           | -9.837 | -10.2  | -9.456 |
| 20–30      | 2.746  | 2.571  | 2.925  | 49–71          | -10.08 | -10.45 | -9.706 |
| ≥ 30       | 8.362  | 7.947  | 8.769  | 72–92          | -10.07 | -10.45 | -9.684 |
| Altitude   |        |        |        | 93–112         | -8.361 | -8.89  | -7.885 |
| 200–400    | 0.611  | 0.450  | 0.768  | ≥ 113          | -10.22 | -10.89 | -9.533 |
| 400–700    | 0.796  | 0.630  | 0.960  | Fuel           |        |        |        |
| ≥ 700      | 1.796  | 1.613  | 1.969  | annual crop    | -0.231 | -0.502 | 0.036  |
| Aspect     |        |        |        | eucalyptus     | 1.953  | 1.664  | 2.252  |
| north      | -5.806 | -6.198 | -5.408 | hardwoods      | 0.729  | 0.449  | 1.013  |
| west       | -5.945 | -6.332 | -5.554 | HSME           | 1.353  | 1.068  | 1.640  |
| east       | -5.932 | -6.319 | -5.541 | agro-forestry  | 0.302  | 0.011  | 0.584  |
| south      | -5.999 | -6.395 | -5.603 | permanent crop | -0.107 | -0.415 | 0.197  |
| $\sigma^2$ | 20.7   | 20.35  | 21.05  | shrubs         | 1.860  | 1.610  | 2.126  |
|            |        |        |        | RS             | 1.767  | 1.498  | 2.032  |
|            |        |        |        | SME            | 2.232  | 1.923  | 2.537  |

**Table 2:** Estimates of the regression parameters and variance for model  $M_1$ .

## References

- Amaral-Turkman M.A. and Silva G.L. (2000) *Modelos Lineares Generalizados - da teoria à prática*, Portuguese Statistical Society, Peniche, Portugal.
- Amaral-Turkman M.A., Turkman K.F., Yannick L.P. and Pereira J. (2008) Hierarchical space-time models for high dimensioned fire data, Technical Report 10/08, Center of Statistics and Applications, University of Lisbon.
- Banerjee S., Carlin B. and Gelfand A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC, Boca Raton, Florida.
- Ferrari S.P. and Cribari-Neto F. (2004) Beta regression for modelling rates and proportions, *Journal of Applied Statistics*, 31, 799–815.
- Gomes J.F.P. (2006) Forest fires in Portugal: how they happen and why they happen, *International Journal of Environmental Studies*, 63, 109–119.
- González J.R. and Pukkala T. (2007) Characterization of forest fires in Catalonia (north-east Spain), *European Journal of Forest Research*, 126, 421–429.
- Spiegelhalter D., Thomas A., Best N.G. and Lunn D. (2007) *Bayesian inference using Gibbs sampling for Windows (WinBUGS), Version 1.4.3*, MRC Biostatistics Unit, Cambridge, <http://www.mrc-bsu.cam.ac.uk/bugs/>.