Journal of Environmental Management 154 (2015) 151–158

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Prediction of forest fires occurrences with area-level Poisson mixed models☆

Miguel Boubeta [a], María José Lombardía [a,*], Manuel Francisco Marey-Pérez [b], Domingo Morales [c]

[a] Universidade da Coruña, Spain
[b] Universidad de Santiago de Compostela, Spain
[c] Universidad Miguel Hernández de Elche, Spain

ARTICLE INFO

Article history:
Received 6 October 2014
Received in revised form
28 January 2015
Accepted 9 February 2015
Available online 25 February 2015

Keywords:
Forest fires
Poisson mixed models
Prediction
Bootstrap

ABSTRACT

The number of fires in forest areas of Galicia (north-west of Spain) during the summer period is quite high. Local authorities are interested in analyzing the factors that explain this phenomenon. Poisson regression models are good tools for describing and predicting the number of fires per forest areas. This work employs area-level Poisson mixed models for treating real data about fires in forest areas. A parametric bootstrap method is applied for estimating the mean squared errors of fires predictors. The developed methodology and software are applied to a real data set of fires in forest areas of Galicia.

1. Introduction

The number of human-caused bushfires has increased worldwide in the last years (Plucinski et al., 2013; Le Page et al., 2014; Krawchuk and Moritz, 2014). Policy makers and technicians request scientific models to explain the causality of fire and to establishing future scenarios of fire risk conditions, especially in one of the worst hit regions of the world, Mediterranean Europe (Rodrigues and de la Riva, 2014). This need has brought about the development of several prediction models (Martínez et al., 2009; Thompson and Calkin, 2011; Ager et al., 2014) which have focus on explaining spatio-temporal patterns that relate different variables (physiographic, infrastructural, socio-economic and weather) with ignition arson wildfires. The "number of fires" is a count variable, which is not continuous and the use of linear models is not appropriate. For this type of data, McCullagh and Nelder (1989) proposed an extension of linear models called generalized linear models (GLMs). GLMs assume that the distribution of the target variable belongs to the exponential distribution and therefore they can be used for Bernoulli, binomial and Poisson distributions, among others.

Poisson regressions and binomial-logit models are GLMs that are generally used for counts of rare events. This is the case of counting number of fires in a given territory and time period. Some papers introducing statistical models for counts of forest fires can be found in the literature. García et al. (1995) introduced a logit model to predict the number of fire-days in the Whitecourt Forest of Alberta. Mandallaz and Ye (1997) presented a general statistical methodology for the prediction of forest fires occurrences in the context of Poisson models. They applied their methodology to data from France, Italy, Portugal, and Switzerland. Wotton et al. (2003) developed Poisson regression predictive models for the daily number of fires in ecoregions of Ontario. Martínez et al. (2009); Chuvieco et al. (2010) and De Vicente and Crespo (2012) gave models for probability of occurrence of a fire.

The occurrence of the studied events might have some variability between areas that the GLMs cannot explain with the auxiliary variables. Generalized linear mixed models (GLMMs) allow to model the extra variability between areas by introducing a

☆ Supported by the grants MTM2013-41383-P, MTM2012-37077-C02-01, MTM2011-22392 and MTM2008-03010 of the Spanish "Ministerio de Ciencia e Innovación", by "Xunta de Galicia" CN2012/130 and 07MRU035291PR, by "Ministerio del Medio Ambiente, Rural y Marino" PSE-310000-2009-4 and by COST Action/UE COST-OC-2008-1-2124.
* Corresponding author. Departamento de Matemáticas, Universidade da Coruña, Spain.
E-mail address: maria.jose.lombardia@udc.es (M.J. Lombardía).

http://dx.doi.org/10.1016/j.jenvman.2015.02.009
0301-4797/

random effect. The estimation of GLMMs are computationally intensive and requires advanced numerical procedures. Some monographs about GLMMs are Demidenko (2004); McCulloch et al. (2008) and Jiang (2007). Finney et al. (2009) analyzed the containment of large wild fires by using GLMMs. Díaz-Avalos et al. (2001) used GLMMs to study the effect of vegetation cover, elevation, slope, and precipitation on the probability of ignition in the Blue Mountains, Oregon, and to estimate the probability of ignition occurrence at different locations in space and in time.

This paper introduces an area-level Poisson mixed models for modeling the number of fires occurred in the forest areas of Galicia during the summer of 2007. The fitted GLMM is also used for predicting the number of fires under different scenario and therefore it gives a tool for decision making. The mean squared error (MSE) of the fire predictions was estimated by applying the parametric bootstrap method described in González-Manteiga et al. (2007, 2008). The proposed methodology for predicting fire counts and estimating prediction uncertainties is a new and useful contribution for forest engineers and policy makers.

The remainder of the paper is organized as follows. Section 2 introduces the area-level Poisson mixed model in a general context and the model-based predictors of number of fires in a general context. It also gives the bootstrap procedure for estimating the MSEs. Section 3 describes the situation and the data under study. The main idea is to look for the factors that explain the number of large fires in rural areas of Galicia. This section presents the statistical analysis and shows how the fitted model can be used for predicting number of fires under scenarios that are close to the one observed in Galicia during the summer of 2007. Section 4 gives some conclusions. The Appendix gives some mathematical details about the Laplace approximation fitting algorithm.

2. The model

This section introduces the Poisson mixed model employed in the data analysis of the case study. First we introduce some notation and assumptions. Let us assume that the region under study can be partitioned into D forest areas (domains). Let $\{v_d:\ d=1,\ldots,D\}$ be a set of random effects that are i.i.d. $N(0,1)$. The distribution of the target variable y_d (number of fires in the forest area d), conditioned to the random effect v_d, is

$$y_d\Big|_{v_d} \sim \text{Poiss}(\mu_d), \quad d=1,\ldots,D, \tag{1}$$

where $\mu_d>0$ is the mean of the Poisson distribution. For the natural parameter, we assume

$$\eta_d = \log\mu_d = \boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d, \quad d=1,\ldots,D, \tag{2}$$

where $\phi>0$ is a variance component parameter, $\boldsymbol{\beta} = \underset{1\le k\le p}{\text{col}}(\beta_k)$ is a vector of fixed regression coefficients and $\boldsymbol{x}_d = \underset{1\le k\le p}{\text{col}'}(x_{dk})$ is the vector of auxiliary variables. Further, we assume that the y_d's are independent conditioned to $\boldsymbol{v}$. Therefore

$$P(y_d|\boldsymbol{v}) = P(y_d|v_d) = \frac{1}{y_d!}\exp\{-\mu_d\}\mu_d^{y_d},$$

where $\mu_d = \exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}$. Let $\boldsymbol{\theta}=(\boldsymbol{\beta},\phi)$ be the vector of all unknown parameters. To fit the model we employ the Laplace approximation algorithm described in the Appendix. This algorithm calculates the maximum likelihood (ML) estimators, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}},\widehat{\phi})$, of the model parameters and the predictors, $\widehat{v}_d$, of the random effects. Based on the parameter estimates and on the random effect predictions, we obtain the predictor $\widehat{\mu}_d = \widehat{\mu}_d(\widehat{\boldsymbol{\theta}},\widehat{v}_d) = \exp\{\boldsymbol{x}_d\widehat{\boldsymbol{\beta}}+\widehat{\phi}\widehat{v}_d\}$. Finally, we estimate $MSE(\widehat{\mu}_d)$ by applying the following parametric bootstrap algorithm

1. Fit the model to the sample and calculate the estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}},\widehat{\phi})$.
2. Repeat B times ($b=1,\ldots,B$)
   (a) Do $v_d^{*(b)} \sim N(0,1)$, $\mu_d^{*(b)} = \exp\{\boldsymbol{x}_d\widehat{\boldsymbol{\beta}}+\widehat{\phi}v_d^{*(b)}\}$, $y_d^{*(b)} \sim \text{Poiss}(\mu_d^{*(b)})$, $d=1,\ldots,D$.
   (b) Calculate $\widehat{\boldsymbol{\theta}}^{*(b)}$, $\widehat{v}_d^{*(b)}$, $\widehat{\mu}_d^{*(b)} = \widehat{\mu}_d^{*}(\widehat{\boldsymbol{\theta}}^{*(b)},\widehat{v}_d^{*(b)})$, $d=1,\ldots,D$.
3. Output:

$$mse^*(\widehat{\mu}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\mu}_d^{*(b)} - \mu_d^{*(b)}\right)^2. \tag{3}$$

3. Case study

The region of Galicia is in the northwest of the Spain (see Fig. 1). Galicia has a population of 2,795,422 (5.9% of the Spanish population) and a surface area of 29,574 km^2 (5.8% of Spain). Galicia has the sixth greatest absolute forest areas among Autonomous Communities, with more than 1,424,094 ha of forest wooded representing 51% of Galician land cover and 7.7% of forest cover in Spain. Ratio of the forest surface is notably higher than the national average (55.2%) (MAGRAMA, 2014). Regarding the property regime of the land, most Galician forest is private (97.2%), and this percentage is much higher than the national average at 67.7% (Rodríguez-Vicente and Marey-Pérez, 2009a). Following the study of Marey-Pérez and Gómez-Vázquez (2010), private forest ownership is subdivided into either particular ownership or communal ownership in collective woodlands (Montes Vecinales en Mano Común, MVMC), an ownership typology almost exclusive to Galicia.

Wildfires in Galicia are a recurrent problem and show increasing levels of severity. There were 251,106 wildfires in Galicia since 1961, the year in which forest fire statistics started, until December 2013 (MAGRAMA, 2014). These fires burned an area of 1,829,330 ha, equivalent to 65.4% of the geographical area of Galicia and almost 128.5% of total wooded forest area. Wildfires mainly affect rural municipalities in the south of the region with low population densities and regressive demographic dynamics due to low birth

Fig. 1. Geographic location map of Galicia in Spain.