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

Miguel Boubeta[1], María José Lombardía[1]†,
Manuel Francisco Marey-Pérez[2], Domingo Morales[3]

[1] Universidade da Coruña, Spain,

[2]Universidad de Santiago de Compostela, Spain,

[3]Universidad Miguel Hernández de Elche, Spain.

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.

Research highlights: • The methodology predicts the probability fire counts. • Bootstrap estimators are created for mean squared errors of fire predictions. • Bootstrap confidence intervals are used to calculate two-scenario fire variation. • The new tools study the effect of auxiliary variables in fire count reduction.

Key words: Forest fires, Poisson mixed models, prediction, bootstrap.

1 Introduction

The number of human-caused bushfires has increased worlwide 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

*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: María José Lombardía, Departamento de Matemáticas, Universidade da Coruña. Mail address: maria.jose.lombardia@udc.es; telephone: +34 981 167000.

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) allows to model the extra variability between areas by introducing a 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 modelling 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 explains 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 $\{v_d : d = 1, \dots, D\}$ 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

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

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, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$, of the model parameters and the predictors, $\hat{v}_d$, of the random effects. Based on the parameter estimates and on the random effect predictions, we obtain the predictor $\hat{\mu}_d = \hat{\mu}_d(\hat{\boldsymbol{\theta}}, \hat{v}_d) = \exp\{\boldsymbol{x}_d \hat{\boldsymbol{\beta}} + \hat{\phi} \hat{v}_d\}$. Finally, we estimate $MSE(\hat{\mu}_d)$ by applying the following parametric bootstrap algorithm

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

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

3 Case study

The region of Galicia is in the northwest of the Spain (see Figure 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

Figure 1: Geographic location map of Galicia in Spain.

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 rates and an ageing population (Balsa-Barreiro and Hermosilla, 2013; Fuentes-Santos et al., 2013). These municipalities have also been unaffected by recent foreign immigration patterns, which, combined with last century's rural flight has led to strong declines in population. Additionally, economic structures are based on primary sectors (González et al., 2007). In this paper, we use the described methodology for modelling the number of fires by forest areas ($D = 63$) in Galicia during the summer 2007. This summer may be taken as representing the wildfire problems of recent years. We also extend the statistical methodology to the prediction of fire counts and the estimation of the corresponding uncertainties under new related scenarios.

Figure 2 (left) plots the annual totals of forest fires occurrences in Spain, Galicia and rest of Spain during the period 1980-2007. These data are taken from the web of Ministry of Agriculture, Food and Environment of the Spain Government (MAGRAMA, 2015) and show that there have been as many fires in Galicia as in the rest of Spain during the cited period. Figure 2 (right) shows the number of fires per forest areas of Galicia during the summer of 2007. Galicia is divided into 63 forest areas, which are the basic territorial structure of the fight against wildfires. At the same time, these areas are grouped into 19 districts. In that summer there were a high amount of fires, 15 areas had more than 19 fires, 15 areas had between 14 and 19 fires, 16 areas had between 8 and 13 fires and 17 areas had less than 7 fires. Most fires were concentrated in the coastal and south-east regions of Galicia.

For each area d, our data set contain the numerical values of the target variable y_d (number of fires) and of the following auxiliary variables:

- dwr: days without rain. Meteorological stations record the average number of summer days without rain: Without rain / (With rain + Without rain). The variable dwr is computed by averaging over the stations in the same area.
- pop: population. Population size.
- par: cadastral parcel. It gives the number of parcels. Non-urban territory is divided in public and private parcels.
- scrub: scrub area. Percentage of land with scrub or bush vegetation.

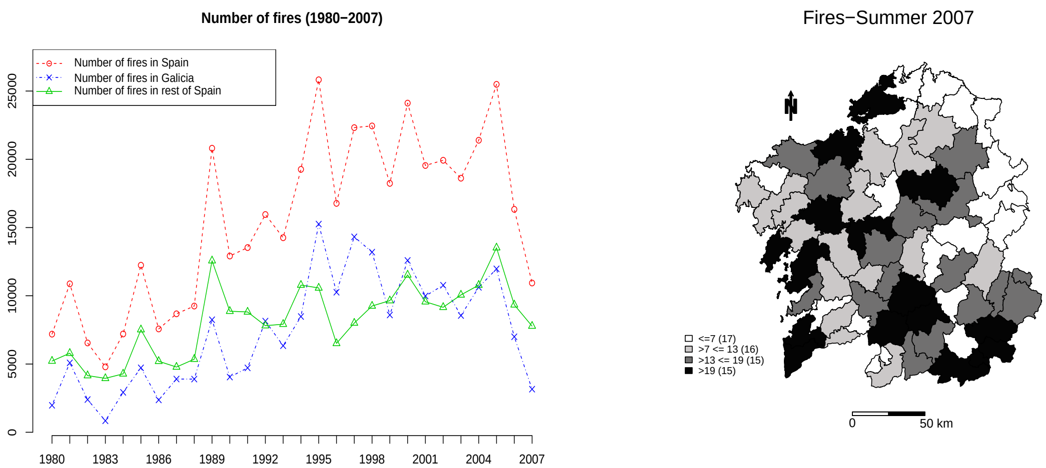

Figure 2: Fires in Spain during 1980-2007 (left) and in Galicia during summer 2007 (right).

- wet: wet area. Percentage of wet lands (rivers, lagoons, swamps...)
- woods: wood area. Percentage of wood lands (forests).
- lu: livestock units. Total number of cattle.

We select the model covariates by taking into account an exploratory analysis, the Akaike information criterion (AIC) defined in (8), a correlations study and some expert judgments. Table 1 shows the parameter estimates of the selected best Poisson mixed model, in terms of simplicity and performance. All regression parameters, except wood area, have positive slopes and therefore an increase in these covariates causes an increase in the number of fires. This is to say, the number of fires is greater in those forest areas with higher values of the variables *number of summer days without rain, population size, number of cadastral parcels, percentage of wet areas* and *total number of cattle units.*

The coefficient of wood area has a negative sign and hence this covariate protects against the increase of the response variable. Most forested areas experience fewer fires due to the greater economic interest in forestry. It is in scrublands, according to Bajocco and Ricotta (2008); Díaz-Delgado et al. (2004); González et al. (2006); González and Pukkala (2007); Koutsias et al. (2009); Montané et al. (2009); Moreira et al. (2009); Mouillot et al. (2005); Nunes et al. (2005) and Sebastian-Lopez et al. (2008), where greater intensity and number of fires is. The positive coefficient of the wet areas is related to the location and the surface of reservoirs, built in the mid to late 20th century to produce electricity. They are mainly found in the headwaters in the mountains of the interior of the region, where an extensive livestock causes a significant number of fires for obtaining pasture, both in early spring and late summer.

Each forest area d ($d = 1, \dots, 63$) has a random intercept with estimated normal distribution $N(0, \hat{\phi}^2)$, where $\hat{\phi} = 0.329$. The 95% percentile bootstrap confidence interval is $(0.162, 0.383)$. We approximate the distribution of $\hat{\phi}$ by the sampling distribution of $\hat{\phi}^*$ in the $B = 1000$ bootstrap samples obtained by applying the algorithm described at the end of Section 2. This is, $P(\hat{\phi} \leq a) \approx P^*(\hat{\phi}^* \leq a)$. Therefore, if $t^*_{(\alpha)}$ is the α-quantile of $\hat{\phi}^*$, then $IC^* = (t^*_{(\alpha/2)}, t^*_{(1-\alpha/2)})$. Shao and Tu (1995) give mathematical details about the construction of the employed bootstrap confidence intervals. We conclude that ϕ is different from 0 at the 95%-level of confidence and we propose a Poisson GLMM with the random effects v_d's instead of the corresponding Poisson GLM without the random effects.

Parameter	Estimate	Std. Error	z-value	$P(>|z|)$
Intercept	2.510	0.057	43.941	0.000
dwr	0.322	0.071	4.503	0.000
pop	0.449	0.064	7.004	0.000
par	0.250	0.059	4.247	0.000
scrub	0.274	0.078	3.521	0.000
wet	0.244	0.057	4.305	0.000
woods	-0.148	0.063	-2.363	0.018
lu	0.324	0.066	4.896	0.000

Table 1: Parameter estimates and p-values for the Poisson mixed model.

For the set of auxiliary variables appearing in Table 1, Figure 3 plots the Pearson residuals for the models without (left) and with (right) random effects. We observe that the residuals of the Poisson mixed model (right) are closer to zero than the ones of the Poisson model with only fixed effects (left), so that we again prefer the model with random effects. The obtained p-values for Shapiro-Wilk test are 0.054 (GLM) and 0.860 (GLMM), and thus the normality assumption is accepted in both cases.

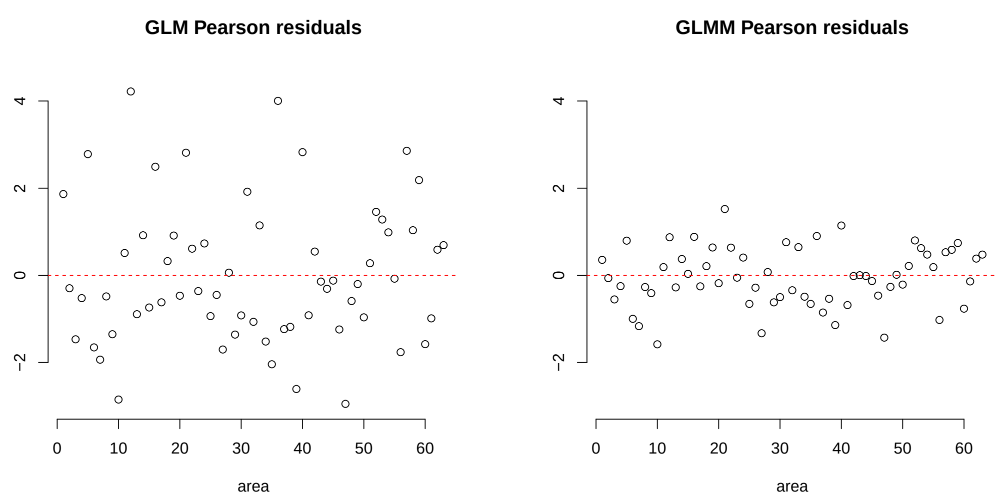

Figure 3: Pearson residuals for the models without (left) and with (right) random effects.

Figure 4 plots the observed number of fires versus the predicted number of fires under the models without (left) and with (right) random effects. We observe that Poisson mixed model has a greater prediction strength. This is why we confirm our selection of the introduced Poisson GLMM.

Figure 5 plots the histogram (left) and the normal qqplot (right) of the predicted random effects, $\hat{v}_d$, $d = 1, \ldots, D$. Based on the two plots and on the p-value of the Shapiro-Wilk test (0.266), we can assume that the normality hypothesis on the model random effects is fulfilled.

It is interesting to note that the auxiliary variable *par* has a regression parameter with positive sign. This fact is related to the small size of the rural parcels in Galicia. Formerly these parcels were cultivated and were the livelihood of families living from agriculture and livestock. Currently many of these parcels are no longer worked, as young people preferred to migrate to cities in search of better opportunities. Many parcels are then abandoned and overgrown allow fast expansion of summer fires. Therefore, the authorities would be interested in promoting the

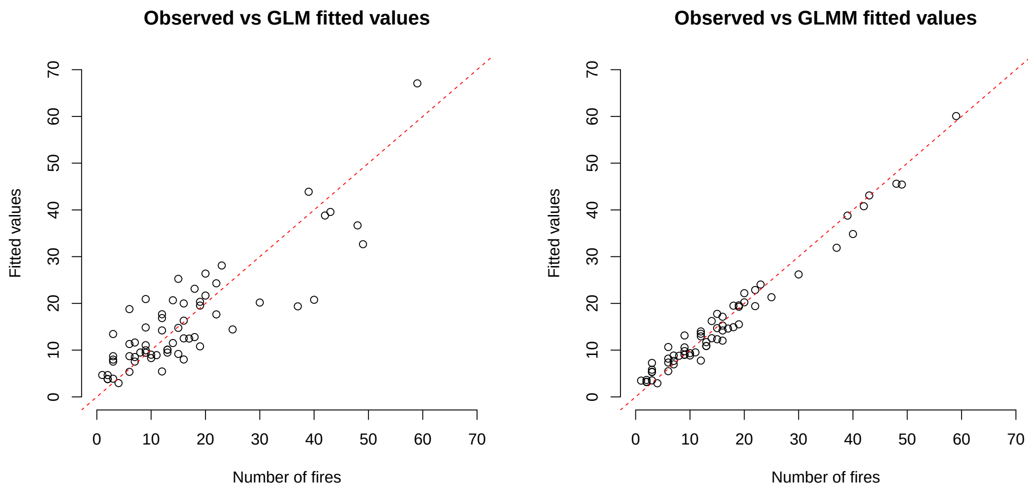

Figure 4: Observed versus predicted number of fires for the models without (left) and with (right) random effects.

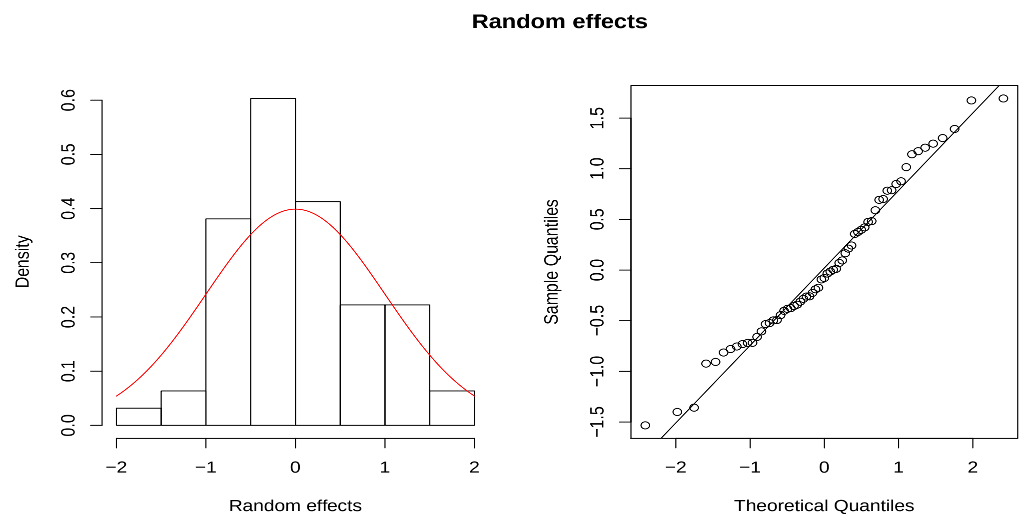

Figure 5: Histogram (left) and normal qqplot (right) of predicted random effects.

creation of cooperatives or land consolidation to get more out of the land and to reduce the number of fires. Below we illustrate the reduction that would occur in the number of fires if we reduce by 5% the number of parcels and we let the remaining variables as they were (scenario $0.95 \times par$).

Figure 6 (left) shows the predicted number of fires per forest areas of Galicia under the scenario $0.95 \times par$. Reducing by 5% the number of cadastral parcels, the model predicts 15 areas with $y_d > 19$, 14 areas with $13 < y_d \leq 19$, 23 areas with $7 < y_d \leq 13$ and 11 areas with $y_d \leq 7$, where y_d is the number of fires in the forest area d. Therefore, if the number of cadastral parcels were reduced by 5% then an important reduction of the number of fires might occur. The Poisson mixed model predicts a reduction from 1001 to 970 fires in summers with similar environmental conditions than the one of 2007. This is because today one of the main problems of the Galician forest sector is the large number of plots (5 plots by ha) (Marey-Pérez et al., 2006; Rodríguez-Vicente and Marey-Pérez, 2009b) and the large number of forest owners (average of 4 ha for owner) (Marey-Pérez and Rodríguez-Vicente, 2009; Rodríguez-Vicente and

Marey-Pérez, 2009a), that is at the origin of a percentage of fires by unprofitability (Rodríguez-Vicente and Marey-Pérez, 2010; Barreal et al., 2011), conflict of ownership and parcel boundaries (Gómez-Vázquez et al., 2009; Bruña García and Marey-Pérez, 2014). Reducing the number of plots contributes to a increasing the profitability (Rodríguez et al., 2013) and decreased much of those fires specially in the most conflictive areas.

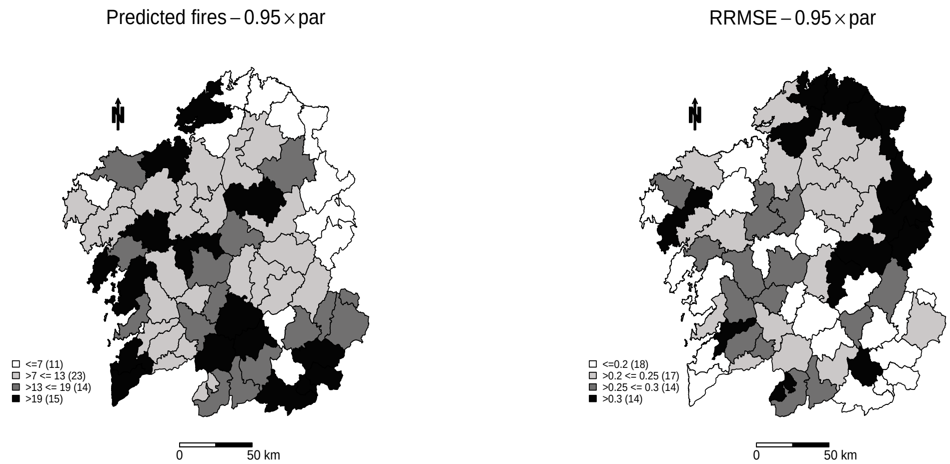

Figure 6: Predicted number of fires and relative root-MSEs under the scenario $0.95 \times par$.

Figure 6 (right) plots the estimated relative root-MSEs (RRMSE), which are obtained as the ratio of the square root of the MSE estimators $mse^*(\hat{\mu}_d)$ and the model-based predictors $\hat{\mu}_d$. In this framework, we take the MSE estimator based on a parametric bootstrap (3) with $B = 1000$. The average RRMSE across all the forest areas is 24.66% and its behaviour is more satisfactory in the western region.

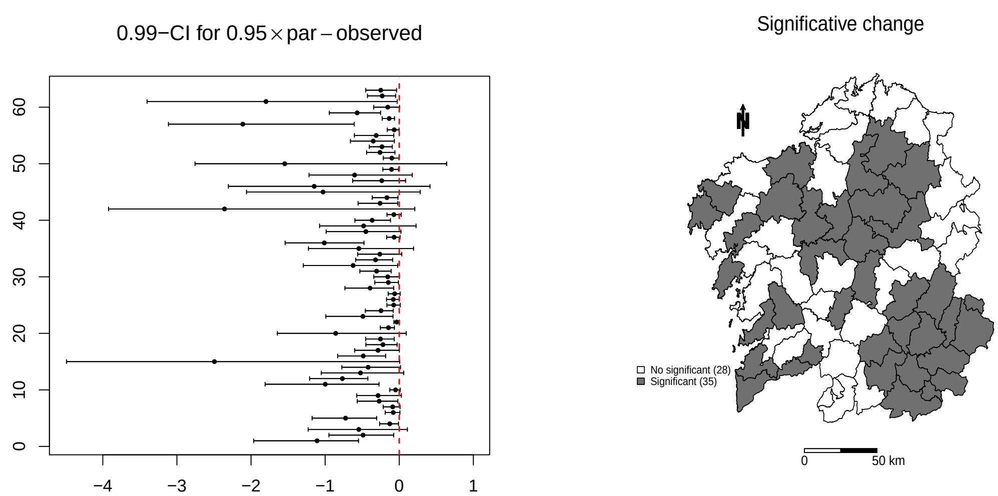

Figure 7: Confidence interval for the variation of the number of fires if we reduce by 5% the number of parcels (left) and its significance (right), with 99% confidence.

Figure 7 presents the basic bootstrap confidence intervals for the estimates of the difference between the number of fires in the scenarios $0.95 \times par$ ($\tilde{\mu}_d$) and observed (μ_d, summer 2007), with a 99% confidence. Now the construction of the bootstrap confidence intervals is a little different. It follows the classic idea Davison and Hinkley (2007) but we consider the random effect in the bootstrap world to estimate μ_d. So, the $(1-\alpha)\%$ confidence limits are $\hat{L} = (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t_{(1-\alpha/2)}$ and $\hat{U} = (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t_{(\alpha/2)}$, being $\widehat{\tilde{\mu}}_d$ the predictor in the new scenario. We approximate the $t_{(\alpha)}$ quantiles by bootstrap following the steps of the previous algorithm in Section 2:

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$.
2. Repeat B times ($b = 1, \ldots, B$)
   (a) Do $v_d^{*(b)} \sim N(0,1)$, $\mu_d^{*(b)} = \exp\{\boldsymbol{x}_d \hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$, $y_d^{*(b)} \sim \text{Poiss}(\mu_d^{*(b)})$, $d = 1, \ldots, D$.
   (b) From $\{\boldsymbol{x}_d, y_d^{*(b)}\}$, calculate $\hat{\boldsymbol{\theta}}^{*(b)}$, $\hat{v}_d^{*(b)}$, $\hat{\mu}_d^{*(b)} = \hat{\mu}_d^*(\boldsymbol{x}_d, \hat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_d^{*(b)})$, $d = 1, \ldots, D$.
   (c) For the new scenario $\{\tilde{\boldsymbol{x}}_d, y_d^{*(b)}\}$, calculate $\tilde{\mu}_d^{*(b)} = \exp\{\tilde{\boldsymbol{x}}_d \hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$ and $\widehat{\tilde{\mu}}_d^{*(b)} = \hat{\mu}_d^*(\tilde{\boldsymbol{x}}_d, \hat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_d^{*(b)})$, $d = 1, \ldots, D$. Note that, the model is fitted from the bootstrap initial sample and then we study the effect of changing the values of the auxiliary variable $\boldsymbol{x}_d$ for $\tilde{\boldsymbol{x}}_d$, as the real world.
3. Calculate the quantiles $t^*_{(\alpha/2)}$ and $t^*_{(1-\alpha/2)}$ as the $(\alpha/2)$-quantile and $(1-\alpha/2)$-quantile of $\left\{(\widehat{\tilde{\mu}}_d^{*(b)} - \hat{\mu}_d^{*(b)}) - (\tilde{\mu}_d^{*(b)} - \mu_d^{*(b)})\right\}_{b=1}^{B}$, respectively.
4. Finally, the bootstrap confidence interval is

$$\left((\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t^*_{(1-\alpha/2)},\ (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t^*_{(\alpha/2)}\right).$$

We have a significant reduction in the number of fires in the forest regions where the confidence interval does not cut the dashed line at the origin. For example, in area 61 (located on the south-west coast), a significant decrease in the number of fires is obtained. In this case we can achieve a reduction of up to 3 fires with a confidence level of 99%. It also corresponds to one of the biggest cities in the community, so a reduction in the number of fires there would be positive for its socio-economic impact and risk of casualties. Figure 6 (right) shows that the above difference is not significant in 28 forest areas and significant in 35. The greatest changes are observed in the interior and southeast zones of Galicia.

Another important auxiliary variable for the occurrence of forest fires is *scrub area*. If we reduce this covariate in the same percentage as in the previous case (5%), the Poisson mixed model predict a reduction of 17 fires. Although this alternative involves a smaller reduction in the number of fires than in the previous case, its economic impact is also lower.

4 Discussion and Conclusions

In recent years, several authors have studied the situation of wildfires in Galicia from an environmental perspective (Vega et al., 2011; Lombao et al., 2014), their financial consequences (Barreal et al., 2011) and the causality in conflictive areas (Fuentes-Santos et al., 2013) or in the whole region (Chas-Amil et al., 2015). There are also different studies on the causes and consequences of extraordinary situations, such as the case of the first half of August, 2006 (González-Alonso and Merino-de-Miguel, 2009; Balsa-Barreiro and Hermosilla, 2013). This paper poses three differences in relation to previous works. First, we select as an analysis unit what happened in the whole region in a representative period of the situation faced by planning and wildfire fight resources, such as summer, 2007. The second difference is using territorial units as modelization units in which fight against fire is developed in three stages: planning, detection and fighting. Third, there is a difference in the modelization method used and its predictive capacity. We use a Poisson mixed model with area random effects because incorporate forest-area random effects that explain the additional between area variations in the data that is not explained by the fixed part of the model. The fitted model can be used for investigating the target variable under modified scenarios. In our case, the number of fires is greater in those forest areas with higher values of the variables *number of summer days without rain, population size, number cadastral parcels, percentage of wet areas* and *total number of cattle units.*

This paper employs Poisson mixed models for quantitative risk-prediction in Mediterranean areas. The paper presents two major contributions. The first one is the methodology for predicting fire counts and for estimating the corresponding efficiency measures approximated by bootstrap. The second one is the construction of bootstrap confidence intervals for the variation of number of fires between observed data and data coming from new scenarios. This tool allows us to study whether the small changes in some auxiliary variables may produce significative reduction of fires in some domains of interest. The introduced statistical methodology gives an useful decision-making tool for policy makers. It can be also extended to target variables with distributions from the exponential family and, therefore, it can be used under Binomial and Poisson distributions, among others. The use of these Poisson mixed models is the first stage towards the integration of a higher number of variables and data about a longer wildfire period. These improvements can increase the predictive capacity which explains the presence of arson wildfires in a conflictive area, answering thus the demands of policy makers and technicians.

5 Appendix

To fit the area-level Poisson mixed model given by (1) and (2), we maximize the Laplace approximation of the joint marginal log-likelihood of the target vector $\boldsymbol{y} = (y_1, \ldots, y_D)$. For this sake, let $h : R \mapsto R$ be a continuously twice differentiable function with a global maximum at x_0. This is to say, let us assume that $\dot{h}(x_0) = 0$ and $\ddot{h}(x_0) < 0$. A Taylor series expansion of $h(x)$ around x_0 yields to

$$h(x) = h(x_0) + \dot{h}(x_0)(x - x_0) + \frac{1}{2}\ddot{h}(x_0)(x - x_0)^2 + o\big(|x - x_0|^2\big) \approx h(x_0) + \frac{1}{2}\ddot{h}(x_0)(x - x_0)^2.$$

The univariate Laplace approximation is

$$\int_{-\infty}^{\infty} e^{h(x)}\, dx \quad \approx \quad \int_{-\infty}^{\infty} e^{h(x_0)} \exp\Big\{ -\frac{1}{2}\big(-\ddot{h}(x_0)\big)(x - x_0)^2\Big\}\, dx = (2\pi)^{1/2}\big(-\ddot{h}(x_0)\big)^{-1/2} e^{h(x_0)}.$$

Let us now approximate the loglikelihood of the considered Poisson mixed model. We recall that $v_1, \ldots, v_D$ are i.i.d N(0,1) and that

$$y_d|_{v_d} \underset{ind}{\sim} \text{Poiss}(\mu_d), \quad \mu_d = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}, \quad d = 1, \ldots, D.$$

It holds that $y_1,\dots,y_D$ are unconditionally independent with marginal probability distribution function

$$\begin{aligned} P(y_d) &= \int_{-\infty}^{\infty} P(y_d|v_d)f(v_d)\,dv_d = \int_{-\infty}^{\infty} \frac{1}{y_d!}\exp\{-\mu_d\}\mu_d^{y_d}(2\pi)^{-1/2}\exp\{-\frac{1}{2}v_d^2\}\,dv_d \\ &= \frac{1}{(2\pi)^{1/2}y_d!}\int_{-\infty}^{\infty}\exp\big\{-\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}+y_d(\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d)-\frac{1}{2}v_d^2\big\}\,dv_d \\ &= \frac{1}{(2\pi)^{1/2}y_d!}\exp\{y_d\boldsymbol{x}_d\boldsymbol{\beta}\}\int_{-\infty}^{\infty}\exp\big\{h(v_d)\big\}\,dv_d, \end{aligned}$$

where

$$\begin{aligned} h(v_d) &= -\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}+\phi y_d v_d-\frac{1}{2}v_d^2, \\ \dot{h}(v_d) &= -\phi\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}+\phi y_d-v_d, \quad \dot{h}(v_{0d})=0, \\ \ddot{h}(v_d) &= -(1+\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}), \quad \ddot{h}(v_{0d})<0. \end{aligned} \tag{4}$$

By applying (4) in $v_d=v_{0d}$, we get

$$P(y_d)\approx\frac{1}{y_d!}\exp\{y_d\boldsymbol{x}_d\boldsymbol{\beta}\}(1+\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\})^{-1/2}\exp\big\{-\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}+\phi y_d v_{0d}-\frac{1}{2}v_{0d}^2\big\}.$$

The loglikelihood is

$$\begin{aligned} \ell &= \sum_{d=1}^{D}\log P(y_d)\approx\sum_{d=1}^{D}\bigg\{-\log y_d!+y_d\boldsymbol{x}_d\boldsymbol{\beta}-\frac{1}{2}\log(1+\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}) \\ &- \exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}+\phi y_d v_{0d}-\frac{1}{2}v_{0d}^2\bigg\}\doteq\ell_L(\boldsymbol{\beta},\phi,v_{01},\dots,v_{0D})=\ell_L. \end{aligned} \tag{5}$$

The components of the score vector and the Hessian matrix are

$$\begin{aligned} U_{0r} &= \frac{\partial\ell_L}{\partial\beta_r},\ U_{0p+1}=\frac{\partial\ell_L}{\partial\phi}, \\ H_{0rs} &= H_{0sr}\frac{\partial^2\ell_L}{\partial\beta_s\partial\beta_r},\ H_{rp+1}=H_{p+1r}=\frac{\partial^2\ell_L}{\partial\phi\partial\beta_r},\ H_{0p+1p+1}=\frac{\partial^2\ell_L}{\partial\phi^2}. \end{aligned}$$

$$\boldsymbol{U}_0=\boldsymbol{U}_0(\boldsymbol{\theta})=\underset{1\le r\le p+1}{\mathrm{col}}(U_{0rs}),\quad \boldsymbol{H}_0=\boldsymbol{H}_0(\boldsymbol{\theta})=(H_{0rs})_{r,s=1,\dots,p+1}.$$

The Newton-Raphson algorithm maximizes $\ell_L(\boldsymbol{\theta})$, with $v_d=v_{0d}$ fixed, $d=1,\dots,D$.

$$\boldsymbol{\theta}^{(k+1)}=\boldsymbol{\theta}^{(k)}-\boldsymbol{H}_0^{-1}(\boldsymbol{\theta}^{(k)})\boldsymbol{U}_0(\boldsymbol{\theta}^{(k)}). \tag{6}$$

For $d=1,\dots,D$, the Newton-Raphson algorithm also maximizes $h(v_d)=h(v_d,\boldsymbol{\theta})$, defined in (4), with $\boldsymbol{\theta}=(\boldsymbol{\beta}',\phi)=\boldsymbol{\theta}_0$ fixed. The updating equation is

$$v_d^{(k+1)}=v_d^{(k)}-\frac{\dot{h}(v_d^{(k)},\boldsymbol{\theta}_0)}{\ddot{h}(v_d^{(k)},\boldsymbol{\theta}_0)}. \tag{7}$$

The Laplace-ML algorithm combines the updating equations (6) and (7), i.e.

1. Set the initial values $k=0$, $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(-1)}=\boldsymbol{\theta}^{(0)}+\boldsymbol{1}$, $v_d^{(0)}=0$, $v_d^{(-1)}=1$, $d=1,\dots,D$.

2. Until $\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k-1)}\|_2 < \varepsilon_1$, $|v_d^{(k)} - v_d^{(k-1)}| < \varepsilon_2$, $d = 1, \ldots, D$, do

   (a) Apply algorithm (7) with seeds $v_d^{(k)}$, $d = 1, \ldots, D$, convergence tolerance ε_2 and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ fixed. Output: $v_d^{(k+1)}$, $d = 1, \ldots, D$.

   (b) Apply algorithm (6) with seed $\boldsymbol{\theta}^{(k)}$, convergence tolerance ε_1 and $v_{0d} = v_d^{(k+1)}$ fixed, $d = 1, \ldots, D$. Output: $\boldsymbol{\theta}^{(k+1)}$.

   (c) $k \leftarrow k + 1$.

3. Output: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k)}$, $\hat{v}_d = v_d^{(k)}$, $d = 1, \ldots, D$.

A seed for the Newton-Raphson algorithm is $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the maximum likelihood estimator under the model without random effects. Concerning the variance parameter, we use

$$\phi^{(0)} = \left(\frac{1}{D} \sum_{d=1}^{D} (\tilde{\eta}_d - \hat{\eta}_d^{dir})^2 \right)^{1/2},$$

where $\tilde{\eta}_d = \boldsymbol{x}_d \tilde{\boldsymbol{\beta}}$, $\hat{\eta}_d^{dir} = \log \hat{\mu}_d^{dir}$ and $\hat{\mu}_d^{dir} = \frac{y_d + 1}{2}$. The ML-Laplace AIC is

$$AIC = 2(p+1) - 2\ell_L(\hat{\boldsymbol{\theta}}, \hat{v}_1, \ldots, \hat{v}_D), \tag{8}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$ and $\hat{v}_d$, $d = 1, \ldots, D$, are taken from the output of the Laplace-ML algorithm. The asymptotic variance of the Laplace-ML estimators can be obtained from the diagonal of the matrix $\boldsymbol{H}_0^{-1}(\hat{\boldsymbol{\theta}})$.

References

Ager, A. A., Preisler, H. K., Arca, B., Spano, D., and Salis, M. (2014). Wildfire risk estimation in the Mediterranean area. *Environmetrics*, 25:384–396.

Bajocco, S. and Ricotta, C. (2008). Evidence of selective burning in Sardinia (Italy): which land-cover classes do wildfires prefer? *Landscape Ecology*, 23:241–248.

Balsa-Barreiro, J. and Hermosilla, T. (2013). Socio-geographic analysis of wildland fires: causes of the 2006's wildfires in Galicia (Spain). *Forest Systems*, 22:497–509.

Barreal, J., Loureiro, M., and Picos, J. (2011). Estudo da incidencia dos incendios en Galicia: Unha perspectiva socioeconomica. *Revista Galega de Economía*, 20:227–246.

Bruña García, X. and Marey-Pérez, M. (2014). Public participation: a need of forest planning. *IForest*, 7:215–225.

Chas-Amil, M. L., Prestemon, J. P., McClean, C. J., and Touza, J. (2015). Human-ignited wildfire patterns and responses to policy shifts. *Applied Geography*, 56:164–176.

Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M., Vilar, L., Martínez, J., Martin, S., Ibarra, P., de la Riva, J., Baeza, J., Rodríguez, F., Molina, J., M.A, H., and Zamora, R. (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling*, 221:46–58.

Davison, A. and Hinkley, D. (2007). *Bootstrap methods and their applications*. Cambridge University Press.

De Vicente, J. and Crespo, F. (2012). A new wildland fire danger index for a Mediterranean region and some validation aspects. *International Journal of Wildland Fire*, 21:1030–1041.

Demidenko, E. (2004). *Mixed models: theory and applications*. J. Wiley.

Díaz-Avalos, C., Peterson, D. L., Alvarado, E., Ferguson, S. A., and Besag, J. E. (2001). Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canadian Journal of Forest Research*, 31:1579–1593.

Díaz-Delgado, R., Lloret, F., and Pons, X. (2004). Spatial patterns of fire occurrence in Catalonia, NE Spain. *Landscape Ecology*, 19:731–745.

Finney, M., Grenfell, I., and McHugh, C. (2009). Modeling containment of large wildfires using generalized linear mixed-model analysis. *Forest Science*, 55:249–255.

Fuentes-Santos, I., Marey-Pérez, M., and González-Manteiga, W. (2013). Forest fire spatial pattern analysis in Galicia (NW Spain). *Journal of Environmental Management*, 128:30–42.

García, C., Woodard, P., Titus, S., Adamowicz, W., and Lee, B. (1995). A logit model for predicting the daily occurrence of human caused forest-fires. *International Journal of Wildland Fire*, 5:101–111.

Gómez-Vázquez, I., Álvarez-Álvarez, P., and Marey-Pérez, M. (2009). Conflicts as enhancers or barriers to the management of privately owned common land: A method to analyze the role of conflicts on a regional basis. *Forest Policy and Economics*, 11:617–627.

González, J., Palahí, M., Tasobares, A., and Pukkala, T. (2006). A fire probability model for forest stands in Catalonia (north-east Spain). *Annals of Forest Science*, 63:169–176.

González, J. and Pukkala, T. (2007). Characterization of forest fires in Catalonia (northeast Spain). *European Journal of Forest Research*, 126:421–429.

González, J., Trasobares, A., Palahí, M., and Pukkala, T. (2007). Predicting stand damage and tree survival in burned forests in Catalonia (North-East Spain). *Annals of Forest Science*, 64:733–742.

González-Alonso, F. and Merino-de-Miguel, S. (2009). Integration of AWiFS and MODIS active fire data for burn mapping at regional level using the Burned Area Synergic Algorithm (BASA). *International Journal of Wildland Fire*, 18:404–414.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational Statistics and Data Analysis*, 51:2720–2733.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate fay-herriot model. *Computational Statistics and Data Analysis*, 52:5242–5252.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Series in Statistics.

Koutsias, N., Arianoutsou, M., Kallimanis, A., and Dimopoulos, P. (2009). Is there any special pattern of the extreme wildland fires occurred in Greece in the summer of 2007? In *52nd International Symposium of the International Association for Vegetation Science, Vegetation Processes and Human Impact in a Changing World, Chania, Greece.*

Krawchuk, M. A. and Moritz, M. A. (2014). Burning issues: statistical analyses of global fire data to inform assessments of environmental change. *Environmetrics*, 25:472–481.

Le Page, Y., Morton, D., Bond-Lamberty, B., Pereira, J. M. C., and Hurt, G. (2014). HESFIRE: an explicit fire model for projections in the coupled Human-Earth System. *Biogeosciences Discuss*, 11:10779–10826.

Lombao, A., Barreiro, A., Carballas, T., Fontúrbel, M. T., Martín, A., Vega, J. A., Fernández, C., and Díaz-Raviña, M. (2014). Changes in soil properties after a wildfire in Fragas do Eume Natural Park (Galicia, NW Spain). *Catena Available online 15 September 2014*.

Mandallaz, D. and Ye, R. (1997). Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research*, 27:1685–1694.

Marey-Pérez, M. F. and Gómez-Vázquez, I. (2010). Modelo para la caracterización del nivel de participación social y la conflictividad en los Montes Vecinales en Mano Común (MVMC) en Galicia. *Spanish Journal of Rural Development*, 1:85–102.

Marey-Pérez, M. F. and Rodríguez-Vicente, V. (2009). Forest transition in Northern Spain: local responses on large-scale programmes of field-afforestation. *Land Use and Policy*, 26:139–156.

Marey-Pérez, M. F., Rodríguez-Vicente, V., and Crecente-Maseda, R. (2006). Using GIS to measure changes in temporal and spatial dynamics of forestland: experiences from north-west Spain. *Forestry*, 79:409–423.

Martínez, J., Vega-García, C., and Chivieco, E. (2009). Human-caused wildfire risk rating for prevention planning in Spain. *Journal of Environmental Management*, 90:1241–1252.

MAGRAMA (2014). Ministry of Agriculture, Food and Environment of the Spain Government. IV National Forest Inventory. *http://magrama.gob.es*, 15/07/2014.

MAGRAMA (2015). Ministry of Agriculture, Food and Environment of the Spain Government. Forest fires. *http://magrama.gob.es*, 07/01/2015.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. J. Wiley.

Montané, F., Casals, P., Taull, M., Lambert, B., and Dale, M. (2009). Spatial patterns of shrub cover after different fire disturbances in the Pyrenees. *Annals of Forest Science*, 66:612–620.

Moreira, F., Vaz, P., Catry, F., and Silva, J. (2009). Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. *International Journal of Wildland Fire*, 18:563–574.

Mouillot, F., Ratte, J., Joffre, R., Mouillot, D., and Rambal, S. (2005). Long-term forest dynamic after land abandonment in a fire prone Mediterranean landscape (central Corsica, France). *Landscape Ecology*, 20:101–112.

Nunes, M., Vasconcelos, M., Pereira, J., Dasgupta, N., Alldredge, R., and Rego, F. (2005). Land cover type and fire in Portugal: do fires burn land cover selectively. *Landscape Ecology*, 20:661–673.

Plucinski, M., McCaw, W., Gould, J., and Wotton, B. (2013). Predicting the number of daily human-caused bushfires to assist suppression planning in south-west Western Australia. *International Journal of Wildland Fire*, 23:520–531.

Rodrigues, M. and de la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling and Software*, 57:192–201.

Rodríguez, P., Anca, A., Diéguez, A., Arenas, S., and Amann, G. (2013). A necesidade dunha xestión eficiente do monte galego. *Grial Dialogos*, 3:26–35.

Rodríguez-Vicente, V. and Marey-Pérez, M. F. (2009a). Characterization of nonindustrial private forest owners and their influence on forest management aims and practices in northern Spain. *Small-Scale Forestry*, 8:479–513.

Rodríguez-Vicente, V. and Marey-Pérez, M. F. (2009b). Land-use and land-base patterns in non-industrial private forests: Factors affecting forest management in Northern Spain. *Forest Policy and Economics*, 7:475–490.

Rodríguez-Vicente, V. and Marey-Pérez, M. F. (2010). Analysis of individual private forestry in northern Spain according to economic factors related to management. *Journal of Forest Economics*, 16:269–295.

Sebastian-Lopez, A., Salvador-Civil, R., Gonzalo-Jimenez, J., and SanMiguel-Ayanz, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in southern Europe. *European Journal of Forest Research*, 127:149–163.

Shao, J. and Tu, D. (1995). *The* J*ackknife and* B*ootstrap*. Springer Series in Statistics.

Thompson, M. P. and Calkin, D. E. (2011). Uncertainty and risk in wildland fire management: A review. *Journal of Environmental Management*, 92:1895–1909.

Vega, J., Jimenez, E., Vega, D., Ortiz, L., and Pérez, J. R. (2011). Pinus pinaster Ait. tree mortality following wildfire in Spain. *Forest Ecology and Management*, 261:2232–2242.

Wotton, B., Martell, D., and K.A., L. (2003). Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change*, 60:275–295.