International Journal of WILDLAND FIRE

# Poisson mixed models for predicting counts of forest fires

| | |
|---|---|
| Journal: | *International Journal of Wildland Fire* |
| Manuscript ID | WF17037.R4 |
| Manuscript Type: | Research Paper |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Boubeta, Miguel; Universidade da Coruña, Mathematics<br>Lombardía, María; Faculty of Computer Science, Mathematics<br>Marey-Perez, Manuel; Universidad de Santiago de Compostela, Escuela Politecnica Superior<br>Morales, Domingo; Universidad Miguel Hernandez de Elche |
| Keyword: | Fire behaviour: modelling, Fire fighters: management, Fire history, Fire intensity |
| | |

# Poisson mixed models for predicting number of fires[1]

Miguel Boubeta, María José Lombardía, Manuel Marey-Pérez and Domingo Morales

22/12/2018

## Abstract

Wildfires are considered one of the main causes of forest destruction. In recent years, the number of forest fires and burned area in Mediterranean regions have increased. This problem particularly affects Galicia (north-west of Spain). Conventional modelling of the number of forest fires in small areas might have a high error. For this reason, four area-level Poisson mixed models with time effects are proposed. The first two models contain independent time effects, while the random effects of the other models are distributed according to an autoregressive process AR(1). A parametric bootstrap algorithm is given to measure the accuracy of the plug-in predictor of fires number under the temporal models. A significant prediction improvement is observed when using Poisson regression models with random time effects. Analysis of historical data finds significant meteorological and socioeconomic variables explaining the number of forest fires by areas and reveals the presence of a temporal correlation structure captured by the area-level Poisson mixed model with AR(1) time effects.

**Key words:** Bootstrap, empirical best predictor, forest fires, mean squared error, method of moments, Poisson mixed models, plug-in predictor, time dependency.

**Highlights:** • The proposed methodology predicts the number of fires by considering their spatial and temporal structure. • Territorial variables change in space and less in time and climatic variables determine the temporal difference. • The new tools explain how changes in variables affect the number of arson fires.

## 1 Introduction

The size, severity and frequency of forest fires have been increasing in the last decades (North et al. 2015). Forest fires are generally regarded as negative for the environment, but they have a key role for the biodiversity and the ecosystem (Driscoll et al. 2010). When the analysis of wildfires focuses on the temperate zones, fires transcend forest management and have become, in the words of Fischer et al. (2016), a "sociecological pathology". More specifically, in Mediterranean Europe, data indicate that on average there are 45,000 fires, with 0.5 million burned hectares every year (San-Miguel-Ayanz and Camia 2009; Moreira et al. 2011; Krasovskii et al. 2016) affecting mainly Spain and Portugal (Reyer et al. 2017).

[1] Supported by the grants MTM2017-82724-R, MTM2015-64842-P, MTM2014-52876-R and MTM2013-41383P of the Spanish "Ministerio de Ciencia e Innovación" and partial support by the Xunta de Galicia Grupo de Referencia Competitiva Grant ED431C 2016-015.

Forest Administrations have not made substantial changes to regional and national wildfire policies (Moritz et al. 2014) despite the significant economic impact of forest fires (DiFonzo et al. 2015; Mourao et al. 2016) and the loss of human lives, as in the case of the fire in 2017 in Pedrogao (Portugal) with 62 fatalities (Wildfire Today 2017). As a consequence of this, nowadays changes beyond those in fire protection services are necessary such as changes in the forestry and territorial models of the Mediterranean countries.

The number of fire occurrences is one of the most studied variables in the topic of wildfire research. Both in the recent case studies Urbieta et al. (2015); Boubeta et al. (2015); Turco et al. (2016); Zhang and Zhuang (2017); Davis et al. (2017); Fox et al. (2018) like in the review of Costafreda-Aumedes et al. (2017) about the arson fires; the authors highlight the good knowledge of the spatio-temporal distribution of fires is crucial for the design of prevention policies adapted to each region. The good knowledge of the spatio-temporal distribution of fires is crucial for the design of prevention policies adapted to each region.

Within the context of climate and environmental change linked to the exodus of rural population and change of land uses, it is necessary to have models at an operational scale, such as forest areas. These models can be employed to simulate different scenarios related to the explanatory variables and to analize the response of the regressor variable, in our case the number of fires.

Different studies have analyzed how the type of vegetation influences the fire risk. Calviño-Cancela et al. (2016) and Calviño-Cancela et al. (2017) found differences between the risk of ignition inside the wildland-urban interface (WUI) area and outside of it, depending on the type of fuel, so that forest plantations near houses were at higher risk. Molina et al. (2017) identified the relationship between live fuel moisture and flammability. The presence of shrub and grassland is also associated with the increase of fires (Anderson et al. 2015; Wyse et al. 2016), most notable in degraded wooded areas. Botequim et al. (2013); Botequim et al. (2017); Martin et al. (2016) and Mirra et al. (2017) conclude that the control of the shrubland reduces the risk in plantations of *Eucalyptus sp* and *Pinus sp*, which in many cases replace old agricultural areas.

There are three types of variables related to human activity which have had increasing relevance in the explanation of fires (McCaffrey et al. 2013): variables related to population, to landowners and to cadastral parcels. Ganteaume and Jappiot (2013) studied population size and Khabarov et al. (2016) used population density as determining factors for fires throughout Europe. Martínez-Fernández et al. (2013) analyzed how differences in population density affect forest fire behavior differently in Galicia than in other areas of Spain. At a regional level Boubeta et al. (2015) and Barreal and Loureiro (2015) have also used population variables in their models. In both cases, the authors agree that the population decline in rural areas is correlated with many fires. In the case of Barreal and Loureiro (2015), the population variable was not significant, due to the excessive size of the areas considered, according to the authors.

Concerning the number and characteristics of the landowners, Canadas et al. (2016) study new forms and models of joint management by individual owners based on Collective Action (Agrawal 2001; Ostrom 2011) which were created by the Portuguese Forestry Administration to reduce the risk of fires. The authors concluded that it is not possible to establish a general model of joint management for the whole country. In Galicia, Diaz-Balteiro et al. (2016a); Diaz-Balteiro et al. (2016b) have analyzed how privately-owned forest plantations of *Eucalyptus globulus* are more sustainable and reduce the risk of fire. In the same region and in the case of collective owners, Alló and Loureiro (2016) concluded that the application of the Principles of Collective Action postulated by Ostrom (1990) reduces the number of fires.

In Mediterranean Europe, forest fires are related to social conflicts, thus the size, number and distribution of land cadastral parcels explain many of the wildfires fires (Ganteaume and Jappiot 2013).

In Spain (Martínez et al. 2009; Padilla and Vega-García 2011; Vilar et al. 2016; Costafreda-Aumedes et al. 2016), and especially in Galicia, the conflicts over land ownership and management is the cause of numerous fires as noted in Gómez-Vázquez et al. (2009); Marey-Pérez and Gómez-Vázquez (2010a); Comas et al. (2014) and Caballero (2015).

We find some papers in the literature that introduce Poisson models for the prediction of forest fires occurrences. For example, Mandallaz and Ye (1997) presented a general statistical methodology for the prediction of forest fires occurrences and 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. Brillinger et al. (2003) and Preisler et al. (2004) used probability-based models for predicting fire risk. However, the use of Poisson mixed models is new in this field, giving good results as shown in Baltar et al. (2014) and Boubeta et al. (2015).

This paper proposes a methodology that incorporates Poisson regression models for counting events and random effects for taking into account the extra variability between areas and time periods. The first objective is to model and explain the number of fires in forest areas during a given time period, by using auxiliary variables. Taking into account the results achieved by the first objective, the second objective is to predict number of fires by forest area in a near future, based on plausible scenarios.

Through the development of the Poisson mixed model methodology, our aim is to have a tool for anticipating the number of fires in the forest areas, reducing the risk of life losses and organizing the response to wildfires. We review which are the variables that seem to explain better the existence and variability of forest fires. The obtained information will help to take appropriate decisions and preventive actions in each area. The new methodology has general nature, but it is illustrated with datasets from Galicia.

The paper is organized as follows. Section 2 presents the background. Section 3 presents the proposed methodology, introducing the study region, the data, the area-level Poisson mixed model, the plug-in predictor of observed fires, the bootstrap approximation to the MSE and the out-of sample prediction. Section 4 applies the developed methodology to forest fires data of Galicia, by months, in the period 2007-2008. Sections 5 and 6 give some recommendation of operational use, a discussion and some conclusions showing that the proposed methodology is a new and useful contribution for forest engineers and policy makers.

# 2 Background

Poisson regression models are generalized linear models (GLM) that are used for counts, i.e. for response variables counting some events of interest (such as the number of forest fires). Sometimes the GLMs cannot explain the variability of the response variable through the selected auxiliary variables. It may happen that observations from different areas are independent, but observations within the same area are dependent because they share common properties. The generalized linear mixed models (GLMM) are extensions of GLMs that capture the variability between areas by introducing random effects, which are usually assumed to be normally distributed. The normality of the random effects is often assumed because it allows obtaining useful distributional properties for testing hypothesis or for confidence interval estimation. More information about GLMMs can be found in the monographs Demidenko (2004) and McCulloch et al. (2008), among others.

Despite the usefulness of GLMMs, inferences based on these models have some computational difficulties because the likelihood may involve high-dimensional integrals which cannot be evaluated analytically. This paper uses the method of moments (MM) suggested by Jiang (1998) for fitting the proposed area-level Poisson mixed model, which is a GLMM. The novelty is the inclusion of temporal effects extending the area-level Poisson mixed model proposed in Boubeta et al. (2015) and following the methodology introduced in Boubeta et al. (2017) for Poisson models or in Hobza et al. (2018) for logistics regression models. We derive plug-in predictors based on area-level Poisson mixed models for predicting count indicators by time period. We use the mean squared error (MSE) as an accuracy measure of the proposed predictor. For estimating the MSE, we implement a parametric bootstrap approach by following the ideas of González-Manteiga et al. (2007) and González-Manteiga et al. (2008a) in the context of logistic and normal mixed models and later extended by González-Manteiga et al. (2008b) to a multivariate area-level model. This approach allows us to calculate the empirical version of the MSE based on a parametric bootstrap.

In the literature on forest fires, the Poisson mixed models and models with temporal effects are treated separately. For example Baltar et al. (2014) and Boubeta et al. (2015) use the Poisson mixed models, Prestemon et al. (2012) consider autoregressive (AR) processes and Boubeta et al. (2016) apply semiparametric time-series models, among others. Because of the cross-sectional structure, area-level mixed models can be used to deal with few time periods. Simple time correlation structures (AR(1) or Moving Average of order 1, denoted by MA(1)) in mixed models, unlike time series, do not require a long sequence of random variables or vectors. Boubeta et al. (2017) analize the effect of the number of time periods through different simulation experiments. In that paper, the number of time periods are $T$=5,9,12. Here we consider a methodology that takes into account both effects, area and time, by means of a Poisson regression mixed model.

# 3 Methodology

## 3.1 Study region

Galicia is a region in the north-west of Spain (see localization in Figure 1a). Around 251,106 wildfires were recorded in Galicia affecting an estimated area of 1,830,000 ha in the last 50 years (Rios-Pena et al. 2017). Since 1999, the administrative structure of the fire-fighting system has been divided into four levels: region, provinces (4), forest districts (19) and forest areas (63). See the forest areas division in Figure 1(b). Three zones have been established according to mountain orography, climatology, demography and forestry factors: coastal, central diagonal and mountainous.

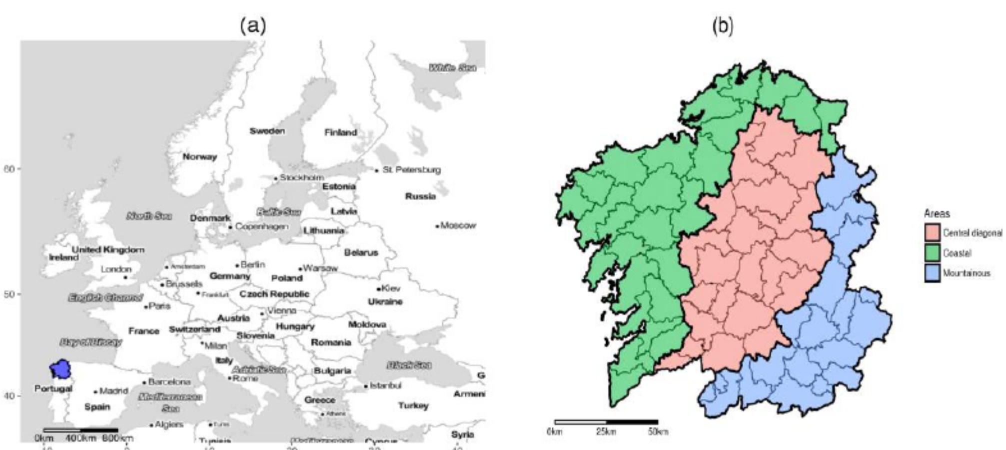


Figure 1: Geographic location map (a) and forest areas (b) of Galicia.

In August 2006, a total of 83,000 hectares (7.5% of the territory or 11% of the forest surface), in the provinces of A Coruña and Pontevedra were affected by wildfires (González-Alonso and Merino-de-Miguel 2009; Balsa-Barreiro and Hermosilla 2013; MMA 2006; Rios-Pena et al. 2017). It was a time of great crisis in a region heavily affected by forest fires (Fernandes 2008; Boubeta et al. 2015; Boubeta et al. 2016), arson for the most part (Román et al. 2013; Fuentes-Santos et al. 2013; Chas-Amil et al. 2015). Consequently, in April 2007, the new Law 3/2007 on "Prevention and defense against forest fires in Galicia" (Consellería de Medio Rural 2007) was passed. This law involved a change of focus in the firefighting to adapt to a new type of arson that mainly affected the WUI interface (Chas-Amil et al. 2012; Modugno et al. 2016). The years 2007 and 2008 were the first to launch this model that changed a tradition of 20 years in firefighting. Here we analize and model the number of forest fires in the community of Galicia by forest areas and months during 2007-2008.

Different authors have studied the causes of fire ignition activity: (1) the disappearance of the traditional agrarian lifestyle (Balsa-Barreiro and Hermosilla 2013), (2) the conflicts over land management and ownership (Marey-Pérez et al. 2010b; Marey-Pérez et al. 2014a; Marey-Pérez et al. 2014b; Caballero 2015), (3) the conflicts in the WUI (Chas-Amil et al. 2012; Chas-Amil et al. 2013), and (4) the socio-economic situation (Alvarez-Díaz et al. 2015; Barreal and Loureiro 2015). Other authors have studied how the fires in the region are distributed and their methodologies are based on: (5) autoregressive processes (Prestemon et al. 2012), (6) intensity functions (Fuentes-Santos et al. 2013; Fuentes-Santos et al. 2015; Comas et al. 2014), (7) Poisson mixed models (Boubeta et al. 2015), and (8) structured additive regression models (Rios-Pena et al. 2017). This paper follows the approach (7) and introduces temporal Poisson mixed models for modelling the number of fires per areas and time periods.

## 3.2 Data

The original forest fires database is provided by the *Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente of the Spain Government* (MAPAMA 2017), and the area-level aggregation is of own elaboration. The response variable, $y_{dt}$, is the *number of forest fires* by forest areas, $d$, and time periods (months), $t$. Galicia is divided into $D$=63 forest areas. For each area, $d$, we observe the change of the response variable by month from 2007 to 2008. Therefore, the number of time periods is $T$=24 months.

Table 1 presents the number of wildfires by month during 2007 and 2008. It suggests that the largest concentrations of wildfires in both years occurs between August and October.

Table 1: Number of wildfires in 2007−2008 by month.

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2007 | 19 | 10 | 167 | 248 | 82 | 56 | 85 | 275 | 648 | 465 | 869 | 129 |
| 2008 | 158 | 663 | 249 | 170 | 37 | 135 | 243 | 356 | 217 | 167 | 24 | 47 |

We assume that the response variable can be explained by some auxiliary variables through an area-level Poisson mixed model with time effects. We consider two sources of auxiliary information depending on their structure. First we include the auxiliary variables that depend only on the areas, i.e. they are constant over time. Second we consider average measurements at meteorological stations for each month and forest area. In the first group we take the number of owners of cadastral parcels, also called number of cadastral holders (*cadHold*). The remaining auxiliary variables are: area of woods (*woods*), shrub (*shrub*) and grassland (*grassland*) per forest area; all of which are given in percentages. In the second group we specifically examine accumulated rain (*acumRain* in $l/m^2$), average air temperature (*averTemp* in °C) and days without rain (*dwr*). Table 2 summarizes the information about the auxiliary variables.

Table 2: Description of the auxiliary information.

| Variables | Data Source | Description | Units |
|---|---|---|---|
| cadHold | Land registry (1:2,000), | Number of cadastral holders, | Num. |
| | Task Office Ministry | Land owners of the plots | |
| woods | Third Spanish Forest, MARM, | Wooded forest land area | % |
| | Inventory cartography (1:50,000) | | |
| shrub | Third Spanish Forest, MARM, | Non wooded forest land area | % |
| | Inventory cartography (1:50,000) | | |
| grassland | Third Spanish Forest, MARM, Inventory cartography (1:50,000) | Grassland area | % |
| pop | Instituto Nacional de Estadística (INE2012) | Population of the plot | Num. |
| acumRain | Climatic Atlas of the Iberian Peninsula | Accumulated water | $l/m^2$ |
| | Spatial resolution 200 m | Monthly data | |
| averTemp | Climatic Atlas of the Iberian Peninsula | Temperature mean | $\circ C$ |
| | Spatial resolution 200 m | Monthly data | |
| dwr | Climatic Atlas of the Iberian Peninsula | Days with rain | Num. |
| | Spatial resolution 200 m | Monthly data | |

In the application to real data, all auxiliary variables were standardized by subtracting its mean value and dividing by its standard deviation. Consequently, all the employed auxiliary variables have mean 0 and standard deviation 1. Table 3 presents some descriptive statistics of the considered auxiliary variables. Specifically, it includes the quartiles, the correlation between each auxiliary variable and the logarithm of the response variable (see model equation (2) in Section 3.3) and the *p*-value for testing null correlation. As the corresponding *p*-values are all lower than 0.05, we conclude that the correlations differs significantly from zero. Table 4 gives the correlations between the response variable (number of fires) and the auxiliary variables. It is interesting to observe that the three meteorological variables (*dwr*, *acumRain* and *averTemp*) are highly correlated. Similarly, the two demographic variables (*pop* and *cadHold*) have also a high correlation. We thus expect that only one in the first group and one in the second group will be selected as auxiliary variables in models explaining the number of fires per forest areas.

Table 3: Description and summary of the auxiliary variables.

| variable | Min. | 1st Qu. | Median | 3rd Qu. | Max. | corr. | p-value |
|---|---|---|---|---|---|---|---|
| cadHold | -1.2946 | -0.6199 | -0.2629 | 0.3376 | 4.1992 | 0.1607 | $< 0.001$ |
| woods | -2.7415 | -0.7738 | 0.1282 | 0.6814 | 2.0548 | -0.2368 | $< 0.001$ |
| shrub | -1.3164 | -0.7855 | -0.2864 | 0.3789 | 2.5773 | 0.1587 | $< 0.001$ |
| grassland | -0.5649 | -0.5509 | -0.3320 | 0.0801 | 5.6597 | 0.0470 | 0.0679 |
| pop | -0.5693 | -0.4434 | -0.3242 | -0.1608 | 4.8678 | 0.1247 | $< 0.001$ |
| acumRain | -1.3163 | -0.7134 | -0.2819 | 0.4995 | 4.4438 | -0.3545 | $< 0.001$ |
| averTemp | -2.2558 | -0.8681 | -0.0261 | 0.8286 | 2.3094 | 0.3277 | $< 0.001$ |
| dwr | -2.4230 | -0.7063 | -0.0815 | 0.7302 | 2.4118 | 0.4969 | $< 0.001$ |

Table 4: Correlations between the response and the auxiliary variables.

| Variable | fires | *cadHold* | woods | scrub | *grassland* | pop | *acumRain* | *averTemp* | dwr |
|---|---|---|---|---|---|---|---|---|---|
| fires | 1.000 | 0.075 | -0.249 | 0.206 | 0.121 | 0.077 | -0.229 | 0.139 | 0.325 |
| cadHold | 0.075 | 1.000 | 0.257 | -0.506 | -0.142 | 0.657 | 0.041 | 0.110 | 0.054 |
| woods | -0.249 | 0.257 | 1.000 | -0.594 | -0.174 | 0.078 | 0.097 | 0.009 | -0.066 |
| shrub | 0.206 | -0.506 | -0.594 | 1.000 | 0.243 | -0.412 | -0.122 | -0.062 | 0.098 |
| grassland | 0.121 | -0.142 | -0.174 | 0.243 | 1.000 | -0.089 | -0.063 | -0.094 | -0.090 |
| pop | 0.077 | 0.657 | 0.078 | -0.412 | -0.089 | 1.000 | 0.077 | 0.067 | 0.015 |
| acumRain | -0.229 | 0.041 | 0.097 | -0.122 | -0.063 | 0.077 | 1.000 | -0.487 | -0.690 |
| averTemp | 0.139 | 0.110 | 0.009 | -0.062 | -0.094 | 0.067 | -0.487 | 1.000 | 0.694 |
| dwr | 0.325 | 0.054 | -0.066 | 0.098 | -0.090 | 0.015 | -0.690 | 0.694 | 1.000 |

## 3.3 The models

The Poisson distribution is usually employed for modeling the number of events of a certain type that can occur in a time period or space interval. This work studies the number of forest fires per month and forest area. The observations within the same area are dependent because they share common properties, but they are assumed to be independent between areas. Consequently, we speak of two sources of variation: between and within areas. Mixed models are well suited for the analysis of this type of data. Here, we introduce two random effects. The first one takes into account the variability between forest areas. The second one deals with the area-time interaction. These random effects complete the classical Poisson model, as they explain the variability that is not included in the fixed part of the model. Our proposal extends the area-level Poisson mixed model given by Boubeta et al. (2015) to the temporal context. We assume that the data are grouped into territorial units (forest areas) and we denote the number of all those areas by $D$. For each forest area $d$ ($d = 1, \dots, D$), a number of interest $y_{dt}, t = 1, \dots, T$, is sequentially recorded along $T$ time periods. In our real data case, $y_{dt}$ denotes the number of forest fires in the area $d$ and time period $t$. Two independent sets of random effects are considered: $\{v_{1,d}: d = 1, \dots, D\}$ depending on the area and $\{v_{2,dt}: d = 1, \dots, D,\ t = 1, \dots, T\}$ depending on the area-time interaction.

For each area-time, the distribution of the discrete response variable, $y_{dt}$, conditioned to the random effects $v_{1,d}$ and $v_{2,dt}$, is

$$y_{dt}|v_{1,d}, v_{2,dt} \sim Poisson\left(\mu_{dt}\right),\ d = 1, \dots, D,\ t = 1, \dots, T, \qquad (1)$$

where the mean of the Poisson distribution, $\mu_{dt}$, is our target parameter since it brings us to the characteristic of interest $y_{dt}$. We assume that the logarithm of $\mu_{dt}$ (natural parameter) can be expressed in terms of a set of auxiliary variables through a regression model, i.e.

$$\log \mu_{dt} = \boldsymbol{x}_{dt}\boldsymbol{\beta} + \emptyset_1 v_{1,d} + \emptyset_2 v_{2,dt},\ d = 1, \dots, D,\ t = 1, \dots, T, \qquad (2)$$

where $\boldsymbol{\beta} = \underset{1\le k\le p}{col}(\beta_k)$ is the vector of regression coefficients, $\boldsymbol{x}_{dt} = \underset{1\le k\le p}{col'}(x_{dtk})$ is the row vector containing the $p$ selected meteorological and socioeconomic auxiliary variables and $\emptyset_1$ and $\emptyset_2$ are the variance parameters. Conditioned to $\boldsymbol{v}_1 = \underset{1\le d\le D}{col}\left(v_{1,d}\right)$ and $\boldsymbol{v}_2 = \underset{1\le d\le D}{col}\left(\boldsymbol{v}_{2,d}\right)$ , where $\boldsymbol{v}_{2,d} = \underset{1\le t\le T}{col}\left(v_{2,dt}\right)$, we assume that the $y_{dt}$'s are independent. Equation (2) employs the random effects for capturing part of the area and time variability and correlation that is not explained by the auxiliary variables. We have that the conditional probability that the response variable tales the value $y_{dt}$ is

$$P(y_{dt}|\boldsymbol{v}_1, \boldsymbol{v}_2) = P\left(y_{dt}|v_{1,d}, v_{2,dt}\right) = \frac{1}{y_{dt}!}\, exp\left\{-\mu_{dt}\right\}\mu_{dt}{}^{y_{dt}}, \qquad (3)$$

where $P$ denotes "probability" and $\mu_{dt} = exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \emptyset_1 v_{1,d} + \emptyset_2 v_{2,dt}\right\}$.

We work with four models depending on the assumed time correlation structure. The first model (Model 1) considers that the two independent sets of random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are independent and identically distributed (i.i.d.) as $N(0, 1)$. As estimation method, we use the MM algorithm based on the method of simulated moments suggested by Jiang (1998). A natural set of equations for applying this method

$$\begin{aligned} 0 = f_k(\boldsymbol{\theta}) &= \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} E_{\boldsymbol{\theta}}[y_{dt}]x_{dtk} - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}x_{dtk}, \qquad k = 1, \dots, p, \\ 0 = f_{p+1}(\boldsymbol{\theta}) &= \frac{1}{D}\sum_{d=1}^{D} E_{\boldsymbol{\theta}}\left[y_{d.}^2\right] - \frac{1}{D}\sum_{d=1}^{D} y_{d.}^2, \qquad (4) \\ 0 = f_{p+2}(\boldsymbol{\theta}) &= \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} E_{\boldsymbol{\theta}}\left[y_{dt}^2\right] - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}^2, \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \emptyset_1, \emptyset_2)$ is the vector of all model parameters. The MM estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\emptyset}_1, \widehat{\emptyset}_2)$, is obtained by solving the system (4) of nonlinear equations.

On the other hand, Model 2 assumes that the random effects in $\boldsymbol{v}_1$ are i.i.d. $N(0,1)$, while in $\boldsymbol{v}_2$ they are AR(1)-correlated within each area $d$ and independent between areas. That is to say, Model 2 assumes $\boldsymbol{v}_1 \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D)$, $\boldsymbol{v}_{2,d} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_d(\varrho))$ and $\boldsymbol{v}_2 \sim N(\boldsymbol{0}, \boldsymbol{\Omega}(\varrho))$. The covariance matrix $\boldsymbol{\Omega}(\varrho)$ of $\boldsymbol{v}_2$ is a block diagonal matrix, where each block $\boldsymbol{\Omega}_d$ is

$$\boldsymbol{\Omega}_d = \boldsymbol{\Omega}_d(\varrho) = \frac{\boldsymbol{A}_d(\varrho)}{1-\varrho^2}, \boldsymbol{A}_d(\varrho) = \begin{pmatrix} 1 & \varrho & \dots & \varrho^{T-2} & \varrho^{T-1} \\ \varrho & 1 & \ddots & \varrho^{T-3} & \varrho^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \varrho^{T-2} & \varrho^{T-3} & \ddots & 1 & \varrho \\ \varrho^{T-1} & \varrho^{T-2} & \dots & \varrho & 1 \end{pmatrix}, d = 1, \dots, D. \quad (5)$$

The system of MM nonlinear equations has the three equations (4) and the new equation associated to the time correlation, i.e.

$$0 = f_{p+3}(\boldsymbol{\theta}) = \frac{1}{D(T-1)} \sum_{d=1}^{D} \sum_{t=2}^{T} E_{\boldsymbol{\theta}}[y_{dt} y_{dt-1}] - \frac{1}{D(T-1)} \sum_{d=1}^{D} \sum_{t=2}^{T} y_{dt} y_{dt-1}, \qquad (6)$$

where the model parameters are now $\boldsymbol{\theta} = (\boldsymbol{\beta}', \emptyset_1, \emptyset_2, \varrho)$. For solving the system of nonlinear equations (4) and (6), we run a Newton-Raphson algorithm and obtain $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\emptyset}_1, \widehat{\emptyset}_2, \widehat{\varrho})$. The theoretical details for calculating the MM estimator, using the Newton-Raphson algorithm under Model 1 and Model 2, can be found in Boubeta et al. (2017). Süli and Mayers (2003) gives extensive information about the properties of this algorithm.

We also consider simplified versions of Model 1 and Model 2, Model $1_2$ and Model $2_2$ respectively, that maintain the expression in Eq. (2) but only with area-time random effects $v_{2,dt}$. Namely, the natural parameters of Model $1_2$ and Model $2_2$ fulfill

$$\log \mu_{dt} = \boldsymbol{x}_{dt} \boldsymbol{\beta} + \emptyset_2 v_{2,dt}, \; d = 1, \dots, D, \; t = 1, \dots, T, \qquad (7)$$

where now the vector of all model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}', \emptyset_2)$ for Model $1_2$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \emptyset_2, \varrho)$ for Model $2_2$.

## 3.4 Plug-in predictors

This section provides plug-in predictors of $\mu_{dt}$ for Model 1, Model 2 and their respective simplified versions, Model $1_2$ and Model $2_2$. For Model 1 and Model 2, the plug-in predictor of $\mu_{dt}$ is

$$\hat{\mu}_{dt} = exp\{\boldsymbol{x}_{dt} \widehat{\boldsymbol{\beta}} + \hat{\phi}_1 \hat{v}_{1,d} + \hat{\phi}_2 \hat{v}_{2,dt}\} \qquad (8)$$

where $\widehat{\boldsymbol{\beta}}$, $\widehat{\phi}_1$ and $\widehat{\phi}_2$ are consistent estimators of the model parameters and $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ are predictors of $v_{1,d}$ and $v_{2,dt}$ respectively. This paper employs the MM estimators of $\boldsymbol{\beta}$, $\emptyset_1$ and $\emptyset_2$. As the MM algorithm does not give direct predictors of the random effects $v_{1,d}$ and $v_{2,dt}$, their empirical best predictors (EBP) are applied. Boubeta et al. (2017) present full technical details for calculating the EBPs $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ under Model 1 and 2. Unlike the present study, that paper treats the Poisson distribution as a limiting case of the binomial. That is to say, Boubeta et al. (2017) assume that the Poisson parameter, $\mu_{dt}$, can be expressed as $\nu_{dt} p_{dt}$ where $\nu_{dt}$ is a known size parameter and $p_{dt}$ is the binomial probability parameter. Their computation approach can be applied here by taking $\nu_{dt} = 1$.

Under Model $1_2$ and Model $2_2$, the plug-in predictor of $\mu_{dt}$ is

$$\hat{\mu}_{dt} = exp\{\boldsymbol{x}_{dt} \widehat{\boldsymbol{\beta}} + \hat{\phi}_2 \hat{v}_{2,dt}\}, \qquad (9)$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}_2$ are consistent estimators of $\boldsymbol{\beta}$ and $\emptyset_2$, and $\hat{v}_{2,dt}$ is a predictor of $v_{2,dt}$. The Model $1_2$, having only interaction time-area random effects, can be treated as the non-temporal Poisson mixed model with area random effects studied by Boubeta et al. (2015). Therefore, the EBP of $v_{2,dt}$ can be calculated by applying the methodology given by these authors. On the other hand, the EBP of $v_{2,dt}$ for Model $2_2$ is

$$\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}}[v_{2,dt}|\boldsymbol{y}_d] = \frac{\int_{R^T} v_{2,dt} P(\boldsymbol{y}_d|v_{2,d}) f(v_{2,d}) dv_{2,d}}{\int_{R^T} P(\boldsymbol{y}_d|v_{2,d}) f(v_{2,d}) dv_{2,d}} = \frac{N_{2,dt}(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}})}{D_d(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}})}, \quad (10)$$

where

$$N_{2,dt}(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}}) = \int_{R^T} \prod_{\tau=1}^{T} I_{2,dt}(\tau) exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}) - exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}\}\right\} f(\boldsymbol{v}_{2,d}) d\boldsymbol{v}_{2,d},$$

$$D_d(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}}) = \int_{R^T} \prod_{\tau=1}^{T} exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}) - exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}\}\right\} f(\boldsymbol{v}_{2,d}) d\boldsymbol{v}_{2,d},$$

and $I_{2,dt}(\tau)$ is 1 if $t \neq \tau$ and $v_{2,dt}$ if $t = \tau$.

As the above ratio involves high-dimensional integrals, we approximate them by using an antithetic Monte Carlo algorithm. The steps are

1. For $s_2 = 1, \dots, S_2$, generate $\left(v_{2,d1}^{(s2)}, \dots, v_{2,dT}^{(s2)}\right) \sim N_T(0, \boldsymbol{\Omega}_d(\hat{\varrho}))$ and calculate $\left(v_{2,d1}^{(S2+s2)}, \dots, v_{2,dT}^{(S2+s2)}\right) = -\left(v_{2,d1}^{(s2)}, \dots, v_{2,dT}^{(s2)}\right)$.

2. Calculate $\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{2,dt}(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}})/\hat{D}_d(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}})$, where

$$\hat{N}_{2,dt}(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}}) = \sum_{s2=1}^{2S2} \prod_{\tau=1}^{T} I_{2,dt}^{(s2)}(\tau) exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s2)}) - exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s2)}\}\right\},$$

$$\hat{D}_d(\boldsymbol{y}_d,\hat{\boldsymbol{\theta}}) = \sum_{s2=1}^{2S2} \prod_{\tau=1}^{T} exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s2)}) - exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s2)}\}\right\}.$$

The mean squared error (MSE) of the plug-in predictors is considered to measure their accuracy. It is defined as

$$MSE(\hat{\mu}_{dt}) = E\left[\left(\hat{\mu}_{dt} - \mu_{dt}\right)^2\right] \quad (11)$$

For estimating the MSE of the plug-in predictor $\hat{\mu}_{dt}$ defined in (8), we adapt the parametric bootstrap procedure given in González-Manteiga et al. (2007). The steps of the bootstrap algorithm are

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}}$. Note that $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2)$ for Model 1 and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi}_1, \hat{\phi}_2, \hat{\varrho})$ for Model 2.

2. For each area $d$ ($d = 1, \dots, D$) and time period $t$ ($t = 1, \dots, T$), repeat B times ($b = 1, \dots, B$):

   (a) Generate the boostrap random effects $v_{1,d}^{*(b)}$ and $v_{2,dt}^{*(b)}$. The area random effects $v_{1,d}^{*(b)}$ are i.i.d. N(0,1) in both models. The area-time random effects $v_{2,dt}^{*(b)}$ are i.i.d. N(0,1) in Model 1 and AR(1)-correlated within each area $d$ in Model 2.

(b) Calculate the theoretical bootstrap plug-in predictor $\mu_{dt}^{*(b)} = exp\left\{\boldsymbol{x}_{dt}\widehat{\boldsymbol{\beta}} + \widehat{\phi}_1 v_{1,d}^{*(b)} + \widehat{\phi}_2 v_{2,dt}^{*(b)}\right\}$.

(c) Generate the responses variables $y_{dt}^{*(b)} \sim Poiss(\mu_{dt}^{*(b)})$.

(d) Calculate $\widehat{\boldsymbol{\theta}}^{*(b)}$ and the plug-in predictor $\hat{\mu}_{dt}^{*(b)} = \hat{\mu}_{dt}^{*(b)}(\widehat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_{1,d}^{*(b)}, \hat{v}_{2,dt}^{*(b)})$ given in (8).

3. Output:

$$mse^*(\hat{\mu}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}_{dt}^{*(b)} - \mu_{dt}^{*(b)}\right)^2 \quad (12)$$

Similarly, one can get an approximation of the MSE of (9) under Model $1_2$ and Model $2_2$.

## 3.5 Out-of-sample prediction

The plug-in predictors (8) and (9) and the MSE estimator (12) can be used as diagnosis tools for analyzing how the introduced Poisson mixed models fit to data $(y_{dt}, \boldsymbol{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, of the period under investigation. These predictors can also be employed to predict the values of the target variable $y_{dt}$, $t = T + 1, \ldots, T + t_0$. In the application to real data, this is predicting the number of fires per forest area and month during a near future period, like one year.

From the point of view of time variability, the auxiliary variables can be divided in two sets. The first one contains the variables having small or null changes across time, like *cadHold*, *shrub*, *grassland* and *woods* appearing in Table 5. The second set contains the time-dependent variables, like *dwr* in the model fitted to the data.

It is hard to predict the number of days without rain per month and forest areas in a near future right after the studied time interval. Nevertheless, by looking into the past, those applying the proposed Poisson mixed models may select *dwr* data from several time periods that corresponds to scenarios depending of the amount of recorded rain. We thus assume that a set of auxiliary variables $\boldsymbol{x}_{dt}, d = 1, \ldots, D$, can be constructed for the period $t = T + 1, \ldots, T + t_0$. Under this assumption, the predictors

(8) and (9) can be adapted to predict the values of the target variable $y_{dt}$ (number of fires) per month, forest area and scenario. For Model 2 (or Model 1), this can be done by applying the following prediction algorithm.

1. Fit the model to the data $\left(y_{dt}, \boldsymbol{x}_{dt}\right), d = 1, \ldots, D, t = 1, \ldots, T$. Calculate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\phi}_1, \widehat{\phi}_2, \widehat{\varrho})$. Obtain the preliminary predictions $\tilde{\mu}_{dt} = exp\{\boldsymbol{x}_{dt}\widehat{\boldsymbol{\beta}}\}, d = 1, \ldots, D, t = T + 1, \ldots, T + t_0$.

2. Run the Monte Carlo algorithms that calculate $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ in the period $t = 1, \ldots, T + t_0$. Apply the algorithm formulas with $\widehat{\boldsymbol{\theta}}$ and with the target variable values $y_{dt}$ (true) if $1 \le t \le T$ and $\tilde{\mu}_{dt}$ (predicted) if $T + 1 \le t \le T + t_0$.

3. Apply formulas (8) and (9) with $\widehat{\boldsymbol{\theta}}$ and with the outputs $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ of the Monte Carlo algorithms of Step 2. Obtain the predictors $\hat{\mu}_{dt}, d = 1, \ldots, D,\ t = T + 1, \ldots, T + t_0$.

For estimating the MSEs of the out-of-sample predictors, we propose the following parametric bootstrap algorithm for Model 2 (similarly, for Model 1).

1. Fit the model to the data $\left(y_{dt}, \boldsymbol{x}_{dt}\right)$, $d = 1, \ldots, D$, $t = 1, \ldots, T$. Calculate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\phi}_1, \widehat{\phi}_2, \widehat{\varrho})$.

2. Repeat $B$ times ($b = 1, \dots, B$):

   (a) Generate $v_{1,d}^{*(b)}$ i.i.d. $N(0, 1)$, $d = 1, \dots, D$. Within each area $d$, $d = 1, \dots, D$, generate $v_{2,dt}^{*(b)}$ AR(1) $\hat{\varrho}$-correlated in the time interval $\{1, \dots, T + t_0\}$.

   (b) Calculate the theoretical bootstrap means $\mu_{dt}^{*(b)} = exp\left\{\boldsymbol{x}_{dt}\widehat{\boldsymbol{\beta}} + \widehat{\phi}_1 v_{1,d}^{*(b)} + \widehat{\phi}_2 v_{2,dt}^{*(b)}\right\}$, $d = 1, \dots, D$, t=$1, \dots, T + t_0$.

   (c) Generate the response variable $y_{dt}^{*(b)} \sim Poiss\left(\mu_{dt}^{*(b)}\right)$, $d = 1, \dots, D, t = 1, \dots, T$.

   (d) Fit the model to the data $\left(y_{dt}^{*(b)}, \boldsymbol{x}_{dt}\right)$, $d = 1, \dots, D$, $t = 1, \dots, T$, and calculate $\widehat{\boldsymbol{\theta}}^{*(b)}$.

   (e) Obtain the plug-in predictors $\hat{\mu}_{dt}^{*(b)}$, $d = 1, \dots, D, t = T + 1, \dots, T + t_0$, by applying the prediction algorithm with input data $\widehat{\boldsymbol{\theta}}^{*(b)}$ and $\left(y_{dt}^{*(b)}, \boldsymbol{x}_{dt}\right)$, $d = 1, \dots, D$, $t = 1, \dots, T$.

3. Output:

$$mse^*(\hat{\mu}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}_{dt}^{*(b)} - \mu_{dt}^{*(b)}\right)^2, d = 1, \dots, D, t = T + 1, \dots, T + t_0. \quad (13)$$

# 4 Results

Table 5 presents the significant MM estimates ($p$-value $< 0.05$) of the fixed effect coefficients for the two models with correlated time effects (Model 2 and Model $2_2$). We select the same set of covariates to make fair comparisons between the two models. Estimates suggest that *dwr*, *cadHold*, *shrub* and *grassland* are directly related to the response variable, given that an increase in those variables causes an increase in the response variable if *woods* remains fixed. By contrast, the relationship between *woods* and $y_{dt}$ is inverse since an increase in this variable causes a decrease in the response variable. We take the level of significance $\alpha = 5\%$ for selecting the variables in the final model.

Table 5: Significant MM estimates under Model 2 and Model $2_2$ ($\alpha = 5\%$).

| | Model 2 | | | | Model $2_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| variable | coef. | s.e. | $z$-val | P(> \|z\|) | coef. | s.e. | $z$-val | P(> \|z\|) |
| *Intercept* | 0.4799 | 0.1034 | 4.6391 | <0.001 | 0.5417 | 0.1011 | 5.3561 | <0.001 |
| *dwr* | 0.6204 | 0.0526 | 11.7951 | <0.001 | 0.5915 | 0.0518 | 11.4116 | <0.001 |
| *cadHold* | 0.3275 | 0.0738 | 4.4375 | <0.001 | 0.3416 | 0.0705 | 4.8491 | <0.001 |
| *woods* | -0.3725 | 0.0833 | -4.4728 | <0.001 | -0.3673 | 0.0742 | -4.9484 | <0.001 |
| *shrub* | 0.1789 | 0.0894 | 2.0002 | 0.0455 | 0.1958 | 0.0757 | 2.5858 | 0.0097 |
| *grassland* | 0.1278 | 0.0644 | 1.9852 | 0.0471 | 0.1230 | 0.0593 | 2.0745 | 0.0380 |

The variance parameter estimates of Model 2 are $\widehat{\phi}_1 = 0.0002$ and $\widehat{\phi}_2 = 0.7474$. Their 95% percentile bootstrap confidence intervals are [0, 0.258) and (0.619, 0.887), respectively. See Shao and Tu (1995) for the mathematical details on the construction of this bootstrap confidence intervals. The random effects related to the areas in Model 2 are not significant since the confidence interval of $\phi_1$ contains 0. The estimated correlation parameter is $\hat{\varrho} = 0.5841$ and its 95% percentile bootstrap confidence interval is (0.328, 0.729). In this way, the results suggest a temporal correlation structure and moreover $\phi_1$ is not significant. Therefore, we consider the simplified version of Model 2 with only area-time effects, i.e. Model $2_2$. The fixed effect estimates for Model $2_2$ can be interpreted analogously to Model 2. The estimate of the variance parameter is $\hat{\phi}_2 = 0.7465$ and its 95% bootstrap confidence interval is (0.595, 0.889). The estimated correlation parameter is 0.5571 and its 95% bootstrap confidence interval is (0.329, 0.692). Taking such results into account, we select Model $2_2$ to fit the data of the Galician forest fires since all the components are significant.

Table 6 presents the estimates of the regression coefficients and the corresponding $p$-values for the sequence of type $2_2$ models that lead to the finally chosen model. By taking out the auxiliary variable with the largest $p$-value each time, the Model $2_2$ of Table 5 is selected. We recall that Table 4 shows that variables *acumRain* and *averTemp* are highly correlated with *dwr* and similarly with *pop* and *cadHold*. This fact explains why *acumRain*, *averTemp* and *pop* are not in the final selected Model $2_2$.

Table 6: Coefficient estimates and $p$-values under a sequence of type $2_2$ models.

| variable | coeff. | $p$-val. | coeff. | $p$-val. | coeff. | $p$-val. | coeff. | $p$-val. |
|---|---|---|---|---|---|---|---|---|
| Intercept | 0.5366 | <0.001 | 0.5341 | <0.001 | 0.5391 | <0.001 | 0.5417 | <0.001 |
| acumRain | 0.0586 | 0.0948 | 0.0597 | 0.1186 | 0.0578 | 0.1239 | | |
| dwr | 0.6035 | <0.001 | 0.6010 | <0.001 | 0.5989 | <0.001 | 0.5915 | <0.001 |
| pop | 0.1087 | 0.1449 | 0.1054 | 0.1682 | | | | |
| cadHold | 0.2792 | <0.001 | 0.2888 | <0.001 | 0.3409 | <0.001 | 0.3416 | <0.001 |
| averTemp | -0.0381 | 0.4600 | | | | | | |
| woods | -0.3339 | <0.001 | -0.3323 | <0.001 | -0.3628 | <0.001 | -0.3673 | <0.001 |
| shrub | 0.2202 | 0.0138 | 0.2335 | 0.0052 | 0.1967 | 0.0181 | 0.1958 | 0.0097 |
| grassland | 0.1125 | 0.0631 | 0.1104 | 0.0888 | 0.1103 | 0.0659 | 0.1230 | 0.0380 |

Figure 2 presents scatter plots of the Pearson residuals of Model 0 with only fixed effects (a) and of the proposed Model $2_2$ (b). Figure 3 shows the corresponding histograms. The Pearson residuals of the model with random time effects show a clear improvement since they are closer to 0. In addition, its behaviour is basically the one expected under the normal distribution. These figures gives a practical illustration of the extra flexibility that models with random effects, like the selected Model $2_2$, have for fitting real data in comparison with their counterparts based only on fixed effects.

Figures 4 and 5 map the obtained plug-in predictions by using the simplified area-level Poisson mixed model with AR(1)-correlated time effects, Model $2_2$. The results are presented for August, September and October. Figure 4 contains the results for 2007 and Figure 5 presents the results for 2008. We use these months as they have the most fires (see Table 1) in both years. In August 2007 (first map), there were 23 areas with up to 2 predicted wildfires, 31 between 3 and 6 (including both values), 4 between 7 and 10 (including both values) and 5 areas with more than 10 predicted wildfires. When counting the number of forest areas in each subset, we take into account that some of them do not form connected territories.

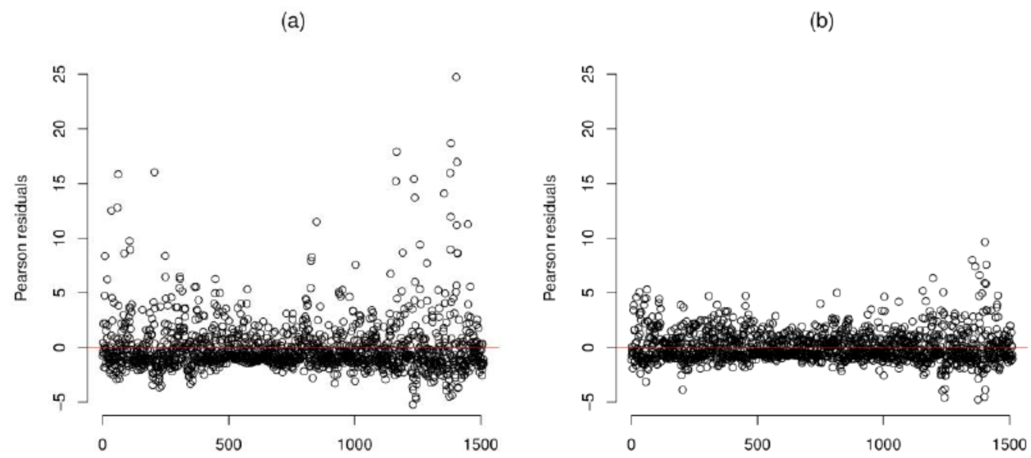


Figure 2: Scatter plots of Pearson residuals of Model 0 with fixed effects (a) and Model $2_2$ (b).

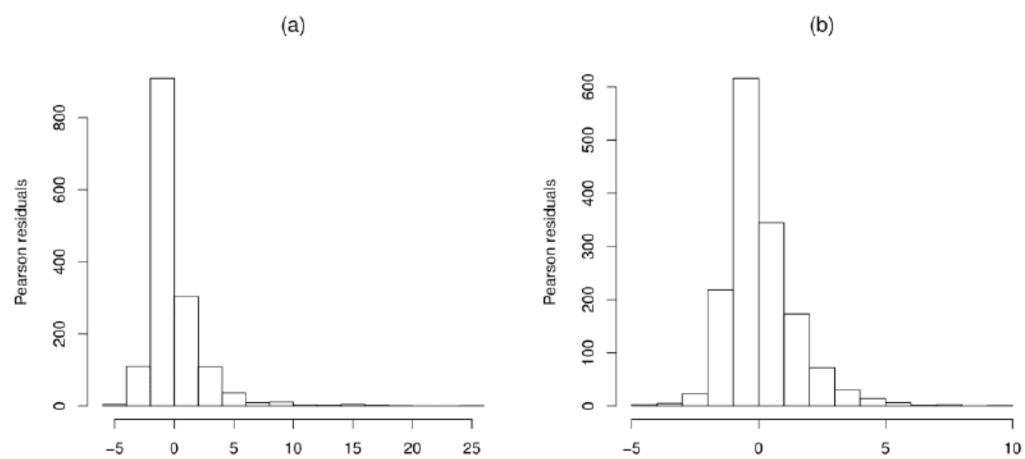


Figure 3: Histograms of Pearson residuals of Model 0 with fixed effects (a) and Model $2_2$ (b).

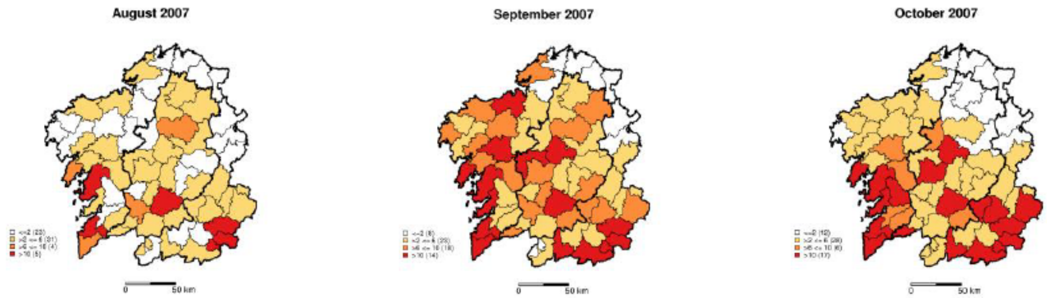


Figure 4: Number of predicted wildfires between August and October in 2007. The number of forest areas by intervals of fire numbers is presented in brackets.

Figures 4 and 5 also suggest that the areas with the greatest number of fires are the South-West coast, the South-East region and some parts of central Galicia. On the other hand, the North-East region has the lowest number of fires.

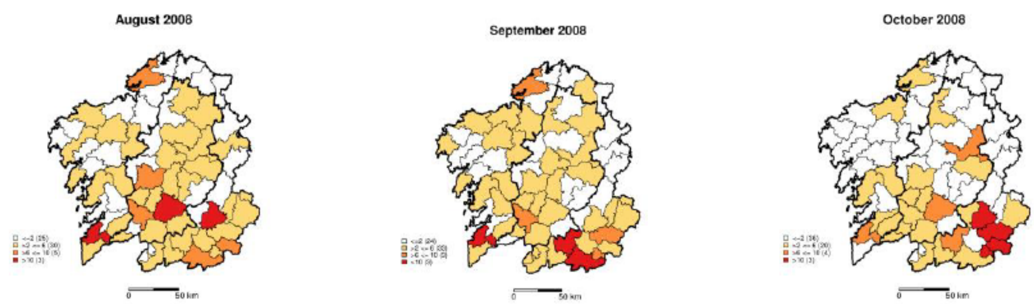


Figure 5: Number of predicted wildfires between August and October in 2008. The number of forest areas by intervals of fire numbers is presented in brackets.

By using Eq. (12), Figure 6 plots the monthly bootstrap predictions of the MSEs for Model 0 and Model $2_2$. We take $B = 500$ bootstrap resamples. The estimated MSEs are plotted for the three areas with highest number of fires: Viana 1 (total fires 311), Terra de Tribes (total fires 329) and Viana 2 (total fires 347). These three areas belong to the mountainous zone. The average MSE for the three areas is 80.02 in Model 0 and 33.73 in Model $2_2$. A clear increase of accuracy is achieved when we use Model $2_2$ since its MSE is much lower.

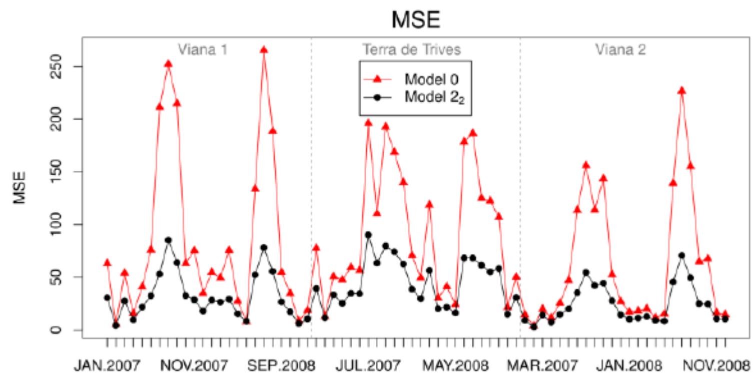


Figure 6: Bootstrap MSE estimates for the three areas with highest fires.

Finally, the behaviour of the proposed model (Model $2_2$) is analized in a context of out-of-sample data. We predict the number of wildfires in 2009 by using the model developed for the period 2007−2008. For that, we could select scenarios of low, medium and high values of the variable *dwr*. However, we have preferred to illustrate the methodology by using the real observed values of *dwr* in the prediction period 2009. For the remaining auxiliary variables we consider the same values as those obtained in 2008, since in an hypothetical future scenario we could assume that they (almost) do not depend on time.

By taking into account the recorded number of fires, $y_d^{(09)}$, in forest area $d$ during 2009, $d \in S = \{1, \dots, D\}$, we divide the set forest areas in the subsets $S1 = \left\{d \in S: y_d^{(09)} \leq 39\right\}$, $S2 = \left\{d \in S: 40 \leq y_d^{(09)} \leq 69\right\}$, and $S3 = \left\{d \in S: y_d^{(09)} \geq 70\right\}$ respectively. Figures 7-9 plot the prediction errors (predicted minus observed number of fires) for subsets S1, S2 and S3. The figures are divided in three parts. The left and central parts contain the boxplots of observed errors by forest areas and months respectively. The right part contains a dispersion graph of observed errors versus number observed of fires.

The boxplots of prediction errors by forest areas are centered around zero in many of the areas of subset S1. However they tend to be centered below zero in forest areas of subsets S2 and S3. Therefore the EBPs derived under Model $2_2$ tend to underestimate the number of fires when the observed number of fires are too large.

The boxplots of prediction errors by months show that the predictions of fire numbers are quite reliable in months 1, 4, 5, 6, 7, 10, 11 and 12 of 2009. In contrast, the methodology proposed under-predicted the number of fires observed in months 2, 3, 8 and 9 of 2009, where there were an unusually high number of fires.

The dispersion graphs of prediction errors versus number of observed fires shows that the prediction errors tend to be negative (under-prediction) when the number of observed fires increases. The prediction methodology works best in counties of subset S1, where the number of observed fires is small.

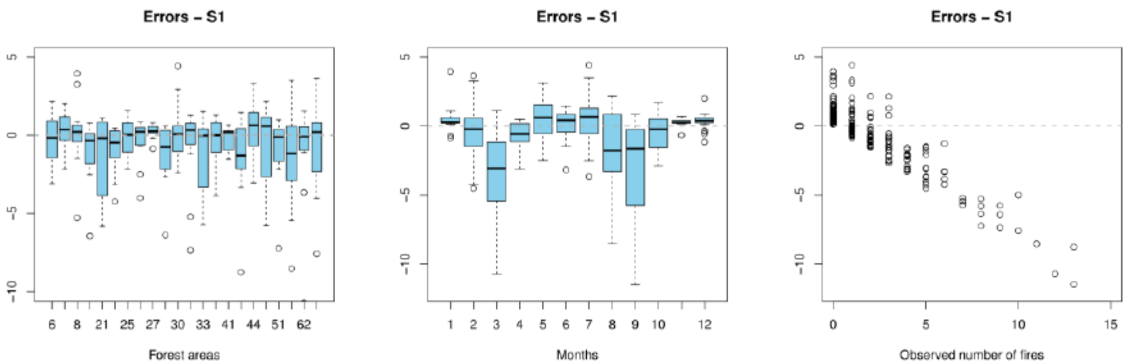


Figure 7: Boxplots and dispersion graphs of predictions errors for forest areas of subset S1.

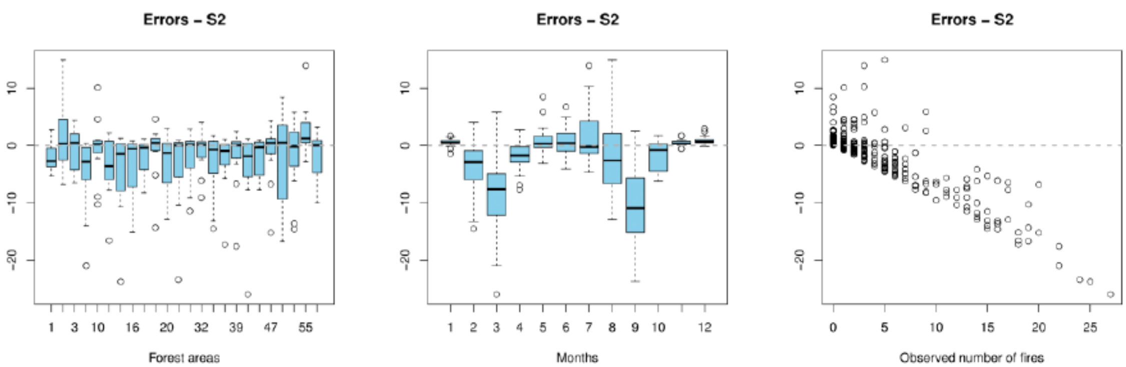


Figure 8: Boxplots and dispersion graphs of predictions errors for forest areas of subset S2.

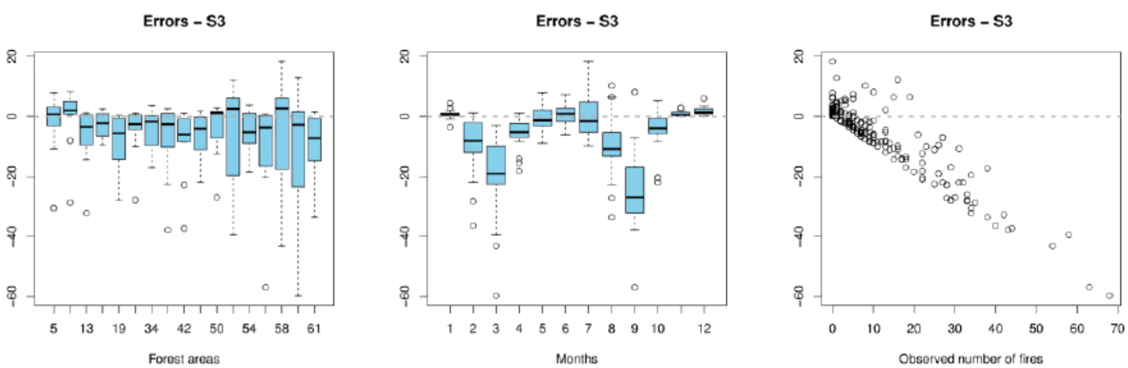


Figure 9: Boxplots and dispersion graphs of predictions errors for forest areas of subset S3.

Figure 10 maps the predicted number of wildfires obtained under Model $2_2$ for 2009. The results are presented for the same months as those shown in Figure 4. For August 2009, Model $2_2$ predicts 18 areas with up to 2 predicted wildfires, 29 with predictions between 3 and 6 (including both values), 8 between 7 and 10 (including both values), and 8 with more than 10 predicted wildfires.

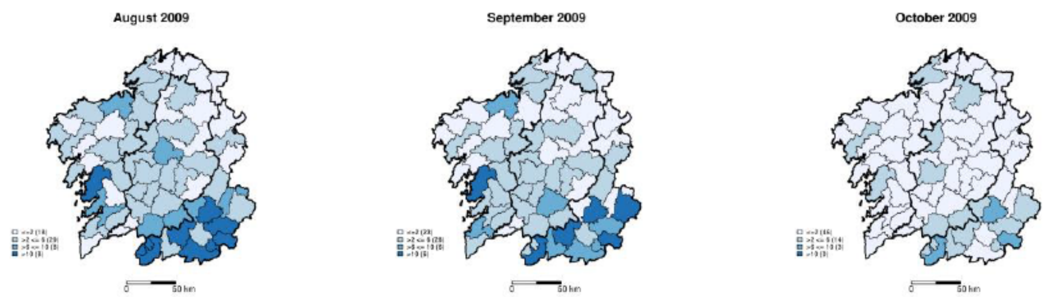


Figure 10: Predicted number of wildfires between August and October in 2009. The number of forest areas within each interval is presented in brackets.

Figure 11 maps the observed number of wildfires for 2009. The results are presented for the same months and intervals as Figure 10. We observe that predictions and observations have a similar spatial distribution.

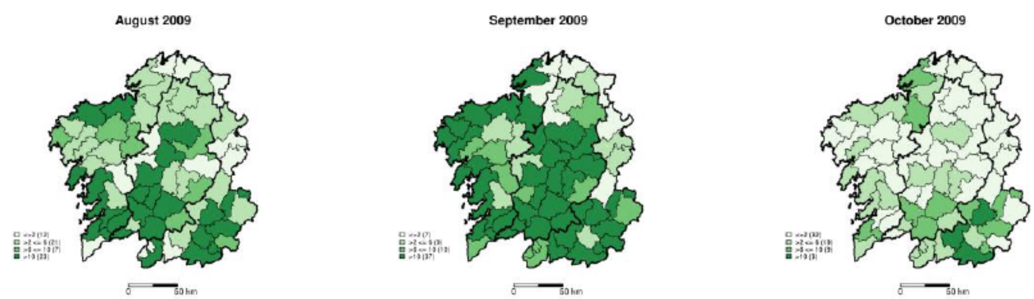


Figure 11: Observed number of wildfires between August and October in 2009. The number of forest areas within each interval is presented in brackets.

Figure 12 presents the bootstrap root-MSEs of the out-of-sample predictor. We remark that we can calculated the prediction errors when the out-of-sample period have finished. This is to say, right after 2009. However, we can calculate the root-MSEs at the same time as the predictions. This is to say, right before the out-of-sample period (year 2009) starts. The root-MSEs gives a measure of the expected reliability of the predictions. In addition to giving predictions of fire numbers, those applying the proposed prediction methodology should give a measure of how reliable they are. Figure 12 shows that the average of the bootstrap root-MSEs for the three months is 6.649 wildfires per month and forest area.

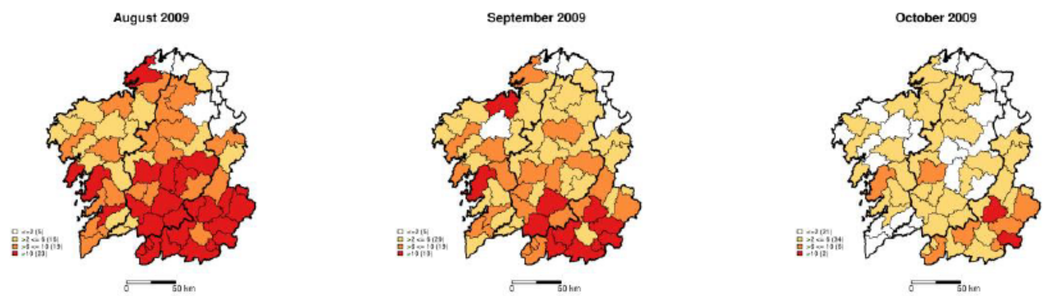


Figure 12: Bootstrap root-MSE estimates (bottom) between August and October in 2009. The number of forest areas within each interval is presented in brackets.

# 5 Operational use

The improvement in the capacity to predict the number of fires, in those territories with arson fires, allows political decision makers to act on the variables of vegetation, ownership and land use. Climatological variables act as a necessary condition to cause fire, when a "window of opportunity" opens. The introduced models allow us to predict the level of fire risk for each forest area in different social, territorial and environmental conditions, establishing the appropriate control measures in a preventive manner. In what follows, we provide some discussion of the limitations of the existing model for operational purposes and what issues need to be addressed for the model to be used effectively in a decision support or policy context.

The introduced methodology for predicting the number of arson fires by forest areas and months has two phases. Phase 1 selects and fits a monthly forest-area-level Poisson mixed model to the sample data. It also makes the corresponding model diagnostics. The sample data file contains the target variable (observed number of arson fires) and the auxiliary variables (related to environ, climate, human activities and social conditions among others) by forest areas and months. It is recommended to fit the model to a period of at least two complete years, so that every month appears the same number of times (at least two) in the sample. After one year, new sample data (12 months) is available and the model can be updated by repeating the steps of Phase 1. The new updated model can be fitted to the enlarged sample. Alternatively, we can update the sample file by entering the most recent 12 months and removing the oldest 12 months. In this last case, the updated model is always fitted to a time period of the same size.

We remind that the introduced prediction methodology assumes that the out-of-sample period will have the same or similar behavior as the sample period. Given the short-medium term trends in auxiliary variables, like *acumRain*, *averTemp* or *dwr*, this fact should be taken into account for deciding the length of the sample period before initiating a new Phase 1. For computational reasons, it is not recommended to use periods of more than five years.

Phase 2 uses the fitted model, during the next 12 months, for predicting future arson fires by forest areas and months. The auxiliary variable can be classified in two types depending on having (type 1) or not (type 2) a sensible time dependency. In the study case, *cadHold*, *woods*, *shrub* and *grassland* are of type 1, so that we can take the same values for the prediction year as those appearing in the last sample year. However, the variable *dwr* is of type 2. Therefore, we recommend considering scenarios of low, medium and high rainfall and calculating the corresponding sets of predictions. The medium rain scenario can be based on long-term weather forecasts.

The introduced methodology depends on the selected set of auxiliary variables. As there are many other factors that influence the occurrence of forest fires, the introduced methodology works well when those factors are not relevant, but it fails when they are. For example, the methodology does not predict well the number of forest fires when there are social conflicts or when very active arsonists appear.

# 6 Discussion

The behavior of people in wildfires is particularly hard to predict (Salvati et al. 2015). The employed methodology appears to be suitable for identifying differentiated spatio-temporal patterns in zones with a great amount of forest fires. The development of new methodologies, especially those contrasted by the evidence of the data, allows a more efficient organization and planning of firefighting, which will result in a lower burnt area and a lower risk for lives.

The spatial and temporal patterns of wildfires in Galicia have been characterized. Differences in climatic conditions within the region is a proposed explanation (Bisquert et al. 2012; Trigo et al. 2016; Fernández-Alonso et al. 2017). Other potential influences include (1) fuel load and continuity (Martín-Martín et al. 2013; González-Ferreiro et al. 2014; Anderson et al. 2015), (2) increasing WUI areas (Chas-Amil et al. 2013; Calviño-Cancela et al. 2014; Calviño-Cancela et al. 2016; Calviño-Cancela et al. 2017), (3) new patterns for agricultural and forest land management (González-Gómez et al. 2013; Fernández-Alonso et al. 2017), (4) agricultural abandonment (Castedo-Dorado et al. 2012; Alló and Loureiro 2016), (5) socioeconomic changes (Chas-Amil et al. 2010; Soliño et al. 2010; Balsa-Barreiro and Hermosilla 2013; Román et al. 2013; Barreal et al. 2014; Barreal and Loureiro 2015; Rodrigues et al. 2016) and (6) Ignition points in each area (Prestemon et al. 2012; Comas et al. 2014; Fuentes-Santos et al. 2013; Fuentes-Santos et al. 2015; Rios-Pena et al. 2015; Rios-Pena et al. 2017; Boubeta et al. 2015; Costafreda-Aumedes et al. 2016).

The results obtained by our model, in terms of the considered meteorological variables, coincide with those obtained by Trigo et al. (2016) and Russo et al. (2017) in that the periods of previous drought are a necessary condition for the presence of fires.

Socioeconomic changes, related to the decline in agricultural activity (Riveiro et al. 2010) and the rural population (Marey-Pérez et al. 2010b) without any changes in the ownership structure (Rodríguez-Vicente and Marey-Pérez 2009) explain many of the conflicts behind a lot of caused fires (Gómez-Vázquez et al. 2009). Model $2_2$ establishes that cadastral holders is a good predictor of the number of wildfires; but this fact cannot be used to draw any conclusions about relationships between conflict, grouped forest management, and the risk of fires cited by the above authors. Our results are similar to those of Alló and Loureiro (2016) in which an increase in the number of owners is related to a higher number of fires. Canadas et al. (2016) showed that the new methodologies of grouped forest management decreased risk of wildfires. We differentiate the three Galician zones.

The link between cadastral data and conflict, or group or collective management (mentioned in the literature) is not supported by Model $2_2$, which does not shed any light on whether conflict or collective management plays a role in the association between number of cadastral holders and the number of fires. Given the correlation between population and cadastral holders, and the fact that population is omitted from the final model, it is likely that cadastral holders is also a proxy for other variables (and inference about this parameter would suffer from omitted variable bias). This is the risk of dropping variables from the model based on $p$-values, so that the resulting model is useful for prediction and fit the data well, but cannot be used for fully explaining the ocurrence of arson fires in the region of Galicia.

The proposed model presents, in comparative terms, an evolution on the work of Prestemon et al. (2012), both in terms of the spatial component (we moved from 19 districts to 63 forest areas) and in the fitting to the data. The performance MSE measure is always below 5% for the areas with the highest number of fires. With respect to the model proposed by Boubeta et al. (2015), there is a significant improvement in the obtained residuals as well as in the MSE bootstrap values. Further, the dispersion graphs of model residuals shows that the selected Poisson mixed model has a better fit to data than the model without random effects.

As established in Boubeta et al. (2015), the improvement in statistical modeling can increase the predictive capacity for explaining the presence of wildfires in a conflictive area. This paper advances in this direction since it gives to policy makers an accurate tool to assist with fire fighting according to the forecast of a phenomenon characterized by high spatial variability and changing human causality.

We introduce four area-level Poisson mixed models with time random effects. The first one, Model 1, assumes that the time effects are independent while the second one, Model 2, assumes that they are AR(1)-correlated within the areas. Simplified versions of Model 1 and Model 2, Model $1_2$ and Model $2_2$, with only area-time random effects are also considered. The MM algorithm is employed for estimating the model parameters. Plug-in predictors of the Poisson parameter, $\mu_{dt}$, are proposed in both contexts: independence and AR(1)-correlation. The empirical best predictors for the area-time random effects under Model $2_2$ (10) are provided. The new statistical methodology is adapted to obtain predictions for out-of-sample data. The method is of a general nature and is demonstrated against the Galician datasets.

With regard to the application to real data, the first step is to select appropriate variables for applying the statistical methodology to predicting the number of forest fires by areas in Galicia. The performance of the plug-in predictors in the area-level Poisson mixed models with time effects is studied and compared against the corresponding predictor obtained from the fixed effects model. A clear improvement is achieved when mixed model is used. A temporal correlation structure is uncovered by the auxiliary data and therefore Model 2 or Model $2_2$ are more appropriate in this context. As the area effects are not significant, it is recommended to use the simplified version, i.e. the Model $2_2$.

From the analysis of forest fires in Galicia by month during 2007-2008, the plots of Pearson residuals and the testing of hypotheses on the model parameters show that the selected model fits well to the observed data. For measuring the accuracy of the proposed predictor, a bootstrap MSE based on a parametric bootstrap is considered.

An application to predict the number of fires in 2009 is also given. As the meteorological variables change over time, different scenarios for predicting the number of fires could be assumed. The auxiliary variables related to type of vegetation, human activities and land ownership does not vary too much over time and depend only on the forest areas. The values of these variable in a near future are easy to establish. As an example of application, we took the true meteorologic variables of 2009.

The performance of the prediction methodology was quite reliable in eight months. However, it gave under-predictions of numbers of fires in months 2, 3, 8 and 9 of 2009, where there was an unusually high number of fires. This fact confirms that predictions will be acceptable if future behaves as past and that the model can only take into account circumstances or situations that can be somehow explained by the employed auxiliary variables.

Finally, we consider that model development and data analysis are interesting tools to make a preventive policy and to support the design of more effective measures against fires. In those regions affected by wildfires, it is very important to include these predictions in the planning for sustainable forest management and in the minimizing of risk factors.

## Acknowledgements

The authors are grateful to the editor and to three unknown reviewers for their valuable suggestions and recommendations. Thanks to these recommendations, the final manuscript was greatly improved.

## References

Agrawal A (2001) Common property institutions and sustainable governance of resources. *World development* **29**, 1649-1672

Alló M, Loureiro ML (2016) Evaluating the fulfillment of the principles of collective action in practice: A case study from Galicia (NW Spain). *Forest Policy and Economics* **73**, 1-9

Alvarez-Díaz M, González-Gómez M, Otero-Giráldez MS (2015) Detecting the socioeconomic driving forces of the fire catastrophe in NW Spain. *European Journal of Forest Research* **134**, 1087-1094

Anderson WR, Cruz MG, Fernandes PM, McCaw L, Vega JA, Bradstock RA, Fogarty L, Gould J, McCarthy G, Marsden-Smedley JB, Matthews S, Mattingley G, Grant H, van Wilgen BW (2015) A generic, empirical-based model for predicting rate of fire spread in shrublands. *International Journal of Wildland Fire* **24**, 443-460

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

Baltar, M., Schoenberg, F.P., and Keeley, J. (2014). Countylevel analysis of the impact of temperature and population increases on California wildfire. *Environmetrics* **25**(6), 397-405

Barreal J, Loureiro ML (2015) Modelling spatial patterns and temporal trends of wildfires in Galicia (NW Spain). *Forest System* **24**, e022

Barreal J, Loureiro ML, Picos J (2014) On insurance as a tool for securing forest restoration after wildfires. *Forest Policy and Economics* **42**, 15-23

Bisquert M, Caselles E, Sánchez JM, Caselles V (2012) Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *International Journal of Wildland Fire* **21**, 1025-1029

Botequim B, Fernandes PM, Garcia-Gonzalo J, Silva A, Borges JG (2017) Coupling fire behaviour modelling and stand characteristics to assess and mitigate fire hazard in a maritime pine landscape in Portugal. *European Journal of Forest Research* **136**, 527-542

Botequim B, Garcia-Gonzalo J, Marques S, Ricardo A, Borges JG, Tomé M, Oliveira MM (2013) Developing wildfire risk probability models for Eucalyptus globulus stands in Portugal. *iForest-Biogeosciences and Forestry* **6**, 217-227

Boubeta M, Lombardía MJ, González-Manteiga W, Marey-Pérez M (2016) Burned area prediction with semiparametric models. *International Journal of Wildland Fire* **25**, 669-678

Boubeta M, Lombardía MJ, Marey-Pérez MF, Morales D (2015) Prediction of forest fires occurrences with area-level Poisson mixed models. *Journal of Environmental Management* **154**, 151-158

Boubeta M, Lombardía MJ, Morales D (2017) Poisson mixed models for studying the poverty in small areas. *Computational Statistics and Data Analysis* **107**, 32-47

Brillinger, D.R., Preisler, H.K., and Benoit, J. (2003). Risk assessment: a forest fire example. *Science and Statistics, Lecture Notes in Statistics* 40, **177**-196, IMS.

Caballero G (2015) Community-based forest management institutions in the Galician communal forests: A new institutional approach. *Forest Policy and Economics* **50**, 347-356

Calviño-Cancela M, Chas-Amil ML, García-Martínez ED, Touza J (2017) Interacting effects of topography, vegetation, human activities and wildland-urban interfaces on wildfire ignition risk. *Forest Ecology and Management* **397**, 10-17

Calviño-Cancela M, Chas-Amil ML, García-Martínez ED, Touza J (2016) Wildfire risk associated with different vegetation types within and outside wildland-urban interfaces. *Forest Ecology and Management* **372**, 1-9

Calviño-Cancela M, Chas-Amil ML, Touza JM (2014) Assessment of fire risk in relation to land cover in WUI areas. In: Viegas, Domingos Xavier, (ed.) Advances in forest fire research. Imprensa da Universidade de Coimbra, pp. 657-664

Canadas MJ, Novais A, Marques M (2016) Wildfires, forest management and landowners'collective action: A comparative approach at the local level. *Land Use Policy* **56**, 179-188

Castedo-Dorado F, Gómez-Vázquez I, Fernandes PM, Crecente-Campo F (2012) Shrub fuel characteristics estimated from overstory variables in NW Spain pine stands. *Forest Ecology and Management* **275**, 130-141

Catry FX, Rego FC, Bação FL, Moreira F (2010) Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* **18**, 921-931

Chas-Amil ML, García-Martínez E, Touza J (2012) Fire risk at the wildland-urban interface: A case study of a galician county. *WIT Transactions on Ecology and the Environment* **158**, 177-188

Chas-Amil ML, Prestemon JP, McClean CJ, Touza J (2015) Human-ignited wildfire patterns and responses to policy shifts. *Applied Geography* **56**, 164-176

Chas-Amil ML, Touza J, García-Martínez E (2013) Forest fires in the wildland-urban interface: A spatial analysis of forest fragmentation and human impacts. *Applied Geography*, **43**, 127-137

Chas-Amil ML, Touza J, Prestemon JP (2010) Spatial distribution of human-caused forest fires in Galicia (NW Spain). *WIT Transactions on Ecology and the Environment* **137**, 247-258

Comas C, Costafreda-Aumedes S, Vega-García C (2014) Characterizing configurations of fire ignition points through spatiotemporal point processes. *Natural Hazards and Earth System Sciences* **2**, 2891-2911

Consellería de Medio Rural (2007) Ley 3/2007, de 9 de Abril, de prevención y defensa contra los incendios forestales de Galicia. http://www.xunta.gal/dog/Publicados/2007/20070417/AnuncioD58A_es.html. [Accessed December 3, 2016]

Costafreda-Aumedes S, Comas C, Vega-Garcia C (2016) Spatio-temporal configurations of human-caused fires in Spain through point patterns. *Forests* **7**, 185.

Costafreda-Aumedes, S, Comas, C, Vega-Garcia, C (2017) Human-caused fire occurrence modelling in perspective: a review. International Journal of Wildland Fire 26, 983-998.

Davis R, Yang Z, Yost A, Belongie C, Cohen W (2017) The normal fire environment—Modeling environmental suitability for large forest wildfires using past, present, and future climate normals. *Forest ecology and management*, **390**, 173-186

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

Diaz-Balteiro L, Alfranca O, Bertomeu M, Ezquerro M, Giménez JC, González-Pachón J, Romero C (2016) Using quantitative techniques to evaluate and explain the sustainability of forest plantations. *Canadian Journal of Forest Research* **46**, 1157-1166

Diaz-Balteiro L, Alfranca O, González-Pachón J, Romero C (2016) Ranking of industrial forest plantations in terms of sustainability: A multicriteria approach. *Journal of environmental management* **180**, 123-132

Diaz-Balteiro L, Bertomeu M, Bertomeu M (2009) Optimal harvest scheduling in Eucalyptus plantations: A case study in Galicia (Spain). *Forest Policy and Economics* **11**, 548-554

Díaz-Varela E, Álvarez-López CJ, Marey-Pérez MF (2009) Multiscale delineation of landscape planning units based on spatial variation of land-use patterns in Galicia, NW Spain. *Landscape and Ecological Engineering* **5**, 1-10

Di Fonzo M, Falcone PM, Germani AR, Imbriani C, Morone P, Reganati F (2015) The quantitative and monetary impacts of forest fire crimes. Report compiled as part of the EFFACE project, University of Rome "La Sapienza", www.efface.eu

Driscoll DA, Lindenmayer, DB, Bennett AF, Bode M, Bradstock RA, Cary GJ, Gill M (2010) Fire management for biodiversity conservation: key research questions and our capacity to answer them. *Biological conservation* **143**(9), 1928-1939.

Fernandes PAM (2008) Forest fires in Galicia (Spain): The outcome of unbalanced fire management. *Journal of Forest Economics* **14**, 155-157

Fernández-Alonso JM, Vega JA, Jiménez E, Ruiz-González AD, Álvarez-González JG (2017) Spatially modeling wildland fire severity in pine forests of Galicia, Spain. *European Journal of Forest Research* **136**, 105-121

Fischer AP, Spies TA, Steelman TA, Moseley C, Johnson BR, Bailey JD, Ager AA, Bourgeron P, Charnley S, Collins BM, Kline JD, Leahy JE, Littell JS, Millington JDA, Nielsen-Pincus M, Olsen CS, Paveglio TB, Roos CI, Steen-Adams MM, Stevens FR, Vukomanovic J, White EM, Bowman DM (2016) Wildfire risk as a socioecological pathology. *Frontiers in Ecology and the Environment* **14**, 276-284

Fox DM, Carrega P, Ren Y, Caillouet P, Bouillon C, Robert S (2018) How wildfire risk is related to urban planning and Fire Weather Index in SE France (1990–2013). *Science of the Total Environment*, **621**, 120-129

Fuentes-Santos I, González-Manteiga W, Mateu J (2015) Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial Poisson point processes. *Scandinavian Journal of Statistics* **43**, 416-435

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

Ganteaume A, Jappiot M (2013) What causes large fires in southern France. *Forest Ecology and Management* **294**, 76-85

Gómez-Vázquez I, Álvarez-Álvarez P, 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-Alonso F, 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-Ferreiro E, Diéguez-Aranda U, Crecente-Campo F, Barreiro-Fernández L, Miranda D, Castedo-Dorado F (2014) Modelling canopy fuel variables for Pinus radiata D. Don in NW Spain with low-density LiDAR data. *International journal of wildland fire* **23**, 350-362

González-Gómez M, Álvarez-Díaz M, Otero-Giráldez MS (2013) Estimating the long-run impact of forest fires on the eucalyptus timber supply in Galicia, Spain. *Journal of Forest Economics* **19**, 149-161

González-Manteiga W, Lombardía MJ, Molina I, Morales D, 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 MJ, Molina I, Morales D, Santamaría L (2008) Bootstrap mean squared error of small-area EBLUP. *Journal of Statistical Computation and Simulation* **78**, 443-462

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008) Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis* **52**, 5242-5252

INE (2012) Cifras de población del Censo 2011. Instituto Nacional de Estadística. Madrid, Spain. http://ine.es. [Accessed March 17, 2017]

Hobza T, Morales D, Santamaría L (2018) Small area estimation of poverty proportions under unit-level temporal binomial-logit mixed models. *TEST* **27**, 270-294.

Jiang J (1998) Consistent estimators in generalized linear models. *Journal of the American Statistical Association* **93**, 720-729

Khabarov N, Krasovskii A, Obersteiner M, Swart R, Dosio A, San-Miguel-Ayanz J, Durrant T, Camia A, Migliavacca M (2016) Forest fires and adaptation options in Europe. *Regional Environmental Change* **16**, 21-30

Krasoskii A, Khabarov N, Migliavacca M, Kraxner F, Obersteiner M (2016) Regional aspects of modelling burned areas in Europe. *International Journal of Wildland Fire* **25**, 811-818

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

MAPAMA (2017) Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. http://www.mapama.gob.es/es/estadistica/temas/estadisticas-ambientales/ [Online; access 2017/07/14]

Marey-Pérez MF, Calvo-Gonzalez A, Domínguez i Torres G (2014) Are the communal forest owners involved in the management of their lands? A qualitative analysis for the case of Galicia (Spain). *Bosque* **35**, 207-215

Marey-Pérez MF, Díaz-Varela ER, Calvo-Gonzalez A (2014) Does higher owner participation increase conflicts over common land? An analysis of communal forests in Galicia (Spain). *iForest Biogeosciences and Forestry* **7**, 507-517

Marey-Pérez MF, 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 MF, Gómez-Vázquez I, Díaz-Varela E (2010) Different approaches to the social vision of communal land management: the case of Galicia (Spain). *Spanish Journal of Agricultural Research* **8**, 848-863

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

Martin A, Botequim B, Oliveira TM, Ager A, Pirotti F (2016) Resource Communication. Temporal optimization of fuel treatment design in blue gum (Eucalyptus globulus) plantations. *Forest Systems*, **25**(2), eRC09

Martín-Martín C, Bunce RG, Saura S, Elena-Rosselló R (2013) Changes and interactions between forest landscape connectivity and burnt area in Spain. *Ecological indicators* **33**, 129-138

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

Martínez-Fernández J, Chuvieco E, Koutsias N (2013) Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. *Natural Hazards and Earth System Sciences* **13**, 311-327

McCaffrey S, Toman E, Stidham M, Shindler B (2013) Social science research related to wildfire management: an overview of recent findings and future research needs. *International Journal of Wildland Fire* **22**, 15-24

McCulloch CE, Searle SR, Neuhaus JM (2008) 'Generalized, linear, and mixed models.' (J. Wiley)

Mirra IM, Oliveira TM, Barros AM, Fernandes PM (2017) Fuel dynamics following fire hazard reduction treatments in blue gum (Eucalyptus globulus) plantations in Portugal. *Forest Ecology and Management* **398**, 185-195

MMA (2006) Ministerio de Medio Ambiente: Incendios forestales en España. [Accessed: December 3, 2016]

Modugno S, Balzter H, Cole B, Borrelli P (2016) Mapping regional patterns of large forest fires in wildland-urban interface areas in Europe. *Journal of Environmental Management* **172**, 112-126

Molina JR, Martín T, Silva FRY, Herrera MA (2017) The ignition index based on flammability of vegetation improvesplanning in the wildland-urban interface: A case study in Southern Spain. *Landscape and Urban Planning* **158**, 129-138

Moreira F, Viedma O, Arianoutsou M, Curt T, Koutsias N, Rigolot E, Barbati A, Corona P, Vaz P, Xanthopoulos G, Mouillot F, Bilgili E (2011) Landscape-wildfire interactions in southern Europe: implications for landscape management. *Journal of Environmental Management* **92**, 2389-2402

Moritz MA, Batllori E, Bradstock RA, Malcolm A, Handmer J, Hessburg PF (2014) Learning to coexist with wildfire. *Nature* **515**, 58-66

Mourao PR, Martinho VD (2016) Discussing structural breaks in the Portuguese regulation on forest fires-An economic approach. *Land Use Policy* **54**, 460-478

North MP, Stephens SL, Collins BM, Agee JK, Aplet G, Franklin JF, Fulé PZ (2015) Reform forest fire management. *Science* **349**, 1280-1281

Ostrom E (1990) 'The evolution of Institutions for Collective Action'. (University Press, Cambridge)

Ostrom E (2011) Background on institutional analysis and development framework. *Policy Studies Journal* **39**, 7-27

Padilla M, Vega-García C (2011) On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human caused fire occurrences in Spain. *International Journal of Wildland Fire* **20**, 46-58

Preisler, H.K., Brillinger, D.R., Burgan, R.E. and Benoit, J.W. Probability based models for estimation of wildfire risk. (2004). *Int. J. Wildland Fire* **13**, 133-142.

Prestemon JP, Chas-Amil ML, Touza JM, Goodrick SL (2012) Forecasting intentional wildfires using temporal and spatiotemporal autocorrelations. *International Journal of Wildland Fire* **21**, 743-754

Reyer CPO, Bathgate S, Blennow K, Borges JG, Bugmann H, Delzon S, Faias SP, Garcia-Gonzalo J, Gardiner B, Gonzalez-Olabarria JR, Gracia C, Hernández JG, Kellomäki S, Kramer K, Lexer MJ, Lindner M, van der Maaten E, Maroschek M, Muys B, Nicoll B, Palahi M, Palma JHN, Paulo JA, Peltola H, Pukkala T, Rammer W, Ray D, Sabaté S, Schelhaas MJ, Seidl R, Temperli C, Tomé M, Yousefpour R, Zimmermann NE, Hanewinkel M (2017) Are forest disturbances amplifying or canceling out climate change-induced productivity changes in European forests? *Environmental Research Letters* **12**. doi:10.1088/1748-9326/aa5ef1

Rios-Pena L, Kneib T, Cadarso-Suárez C, Marey-Pérez M (2017) Predicting the occurrence of wildfires with binary structured additive regression models. *Journal of Environmental Management* **187**, 154-165

Rios-Pena L, Kneib T, Cadarso-Suárez C, Marey-Pérez M (2015) Applying Binary Structured Additive Regression (STAR) for Predicting Wildfire in Galicia, Spain. *Procedia Environmental Sciences* **27**, 123-126

Riveiro JA, Marey-Pérez MF, Díaz-Varela ER, Álvarez CJ (2010) A methodology for the analysis of the relationships between farms and their physical environment. *The Journal of Agricultural Science* **148**, 101-116

Rodrigues M, Jiménez A, de la Riva J (2016) Analysis of recent spatial-temporal evolution of human driving factors of wildfires in Spain. *Natural Hazards* **84**, 2049-2070

Rodríguez-Vicente V, Marey-Pérez M (2009) Land-use and land-base patterns in non- industrial private forests: factors affecting forest management in Northern Spain. *Forest Policy and Economics* **11**, 475-490

Román MV, Azqueta D, Rodrigues M (2013) Methodological approach to assess the socio-economic vulnerability to wildfires in Spain. *Forest Ecology and Management* **294**, 158-165

Russo A, Gouveia CM, Páscoa P, DaCamara CC, Sousa PM, Trigo RM (2017) Assessing the role of drought events on wildfires in the Iberian Peninsula. *Agricultural and Forest Meteorology* **237**, 50-59

Salvati L, Ferrara A, Mancino G, Kelly C, Chianucci F, Corona P (2015) A multidimensional statistical framework to explore seasonal profile, severity and land-use preferences of wildfires in a Mediterranean country. *International Forestry Review* **17**, 485-497

San-Miguel-Ayanz J, Camia A (2009) Forest fires at a glance: facts, figures and trends in the UE. In Yves Birot, Editor. Living with wildfires: what science can tell us? A Contribution to the Science-Policy Dialogue. EFI, Joensuu, Finland

San-Miguel-Ayanz J, Moreno JM, Camia A (2013) Analysis of large fires in European Mediterranean landscapes: lessons learned and perspectives. *Forest Ecology and Management* **294**, 11-22

Shao J, Tu D (1995) 'The Jackknife and bootstrap.' (Springer Series in Statistics)

Soliño M, Prada A, Vázquez MX (2010) Designing a forest-energy policy to reduce forest fires in Galicia (Spain): a contingent valuation application. *Journal of Forest Economics* **16**, 217-233

Süli E, Mayers F (2003) An introduction to Numerical Analysis. Cambridge University Press

Trigo RM, Sousa PM, Pereira MG, Rasilla D, Gouveia CM (2016) Modelling wildfire activity in Iberia with different atmospheric circulation weather types. *International Journal of Climatology* **36**, 2761-2778

Turco M, Bedia J, Di Liberto F, Fiorucci P, von Hardenberg J, Koutsias N, Provenzale A (2016) Decreasing Fires in Mediterranean Europe. *PLoS one*, **11**(3), e0150663

Turco M, Von Hardenberg J, AghaKouchak A, Llasat MC, Provenzale A, Trigo RM (2017) On the key role of droughts in the dynamics of summer fires in Mediterranean Europe. *Scientific Reports* **7** (81). doi:10.1038/s41598-017-00116-9

Urbieta IR, Zabala G, Bedia J, Gutiérrez JM, San-Miguel-Ayanz J, Camia A, Keeley JE, Moreno JM (2015) Fire activity as a function of fire-weather seasonal severity and antecedent climate across spatial scales in southern Europe and Pacific western USA. *Environment Research Letter* **10**, 114013

Vilar L, Camia A, San-Miguel-Ayanz J, Martín MP (2016) Modeling temporal changes in human-caused wildfires in Mediterranean Europe based on land use-land cover interfaces. *Forest Ecology and Management* **378**, 68-78

Wildfire Today (2017) http://wildfiretoday.com/?s=pedrogao. Accessed: 07-07-2017]

Wotton B, Martell D, Logan KA (2003) Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change* **60**, 275-295

Wyse SV, Perry GL, O'Connell DM, Holland PS, Wright MJ, Hosted CL, Whitelock SL, Geary IJ, Maurin KJ, Curran TJ (2016) A quantitative assessment of shoot flammability for 60 tree and shrub species supports rankings based on expert opinion. *International Journal of Wildland Fire* **25**, 466-477

Zhang T, Zhuang R (2017) Testing proportionality between the first-order intensity functions of spatial point processes. *Journal of Multivariate Analysis*, **155**, 72-82