

Noname manuscript No.  
(will be inserted by the editor)

---

## Spatial point process modeling applied to the assessment of risk factors associated to forest wildfires incidence in Castellón, Spain.

P. Aragón<sup>1</sup>, P. Juan<sup>1</sup>,  
C. Díaz-Avalos<sup>2</sup> and P.  
Salvador<sup>1</sup>

Received: date / Accepted: date

**Abstract** During the last decades the mediterranean zone in Europe has experienced an increment in the incidence of forest wildfires. This increase is partly explained by higher mean temperature and lower relative humidity, whilst socio-economic change has lead to the abandonment of farms, resulting in an increase in an unusual accumulation of forest fuels, increasing the risk of wildfires. Mapping wildfire risk is highly important because wildfires are known to potentially lead to landscape changes and to modify fire regime by inducing potential changes in vegetation composition. Also, they pose a hazard to human property and life. Maps of wildfire risk based on statistical models are provide a measure of uncertainty for the inferences derived from such risk maps, leaving a quantitative error margin for managers and decision takers. Further, some of the model parameters often have a physical or a biological interpretation which can give ecologists and forest engineers answers

---

Corresponding author. *Email address:* zhangkalo@gmail.com. Phone: +525554404530

P. Aragón  
Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón,  
Spain  
E-mail: jorientr@yahoo.es

P. Juan  
Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón,  
Spain  
E-mail: juan@uji.es

C. Díaz-Avalos  
Departament de Probabilitat i Estadística, Instituto de Investigaciones en Matemáticas Apli-  
cadas y en Sistemas, Universidad Nacional Autónoma de México  
E-mail: zhangkalo@gmail.com

P. Salvador  
Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón,  
Spain  
E-mail: sanipau@gmail.com

about scientific questions of interest. In this paper we analyze the incidence of wildfires in the province of Castellón, in Spain in order to identify risk factors associated to wildfire incidences during the years 2001-2006. We used the discrete nature of wildfire events to build such models using point process theory and methods and included information about elevation, slope, aspect, land use and distance to nearest road as covariates in our modeling process. Our results show that wildfire risk in Castellón is associated to all the covariates considered, and that three land use categories have the highest risk of wildfire incidence. Also, wildfire incidences are not independent and some degree of interaction exists, which indicates that the commonly used Poisson point process models are not applicable in this case, but instead Area-Interaction models should be considered.

**Keywords** Environmental covariates · Risk mapping · Spatial interaction · Spatial point processes · Wildfires

## 1. Introduction

During the last decades, Spain as well as all the mediterranean zone in Europe have experienced an increment in the incidence of forest fires and consequently an increase in the risk of forest fire ignition (Moreira et al., 2011; Wittenberg and Malkinson, 2009). It is worth noting here that we will mean by forest fire risk the probability that a fire ignites at a given location within a study area (Hardy, 2005), and by forest fire incidence the number of fires per unit area. Thus, fire incidence refers to the observed spatial pattern (number and location) of wildfires occurring in a given study area whilst fire risk refers to a probability measure obtained from a mathematical model. In Spain, the increase in forest fires incidence is partly explained by climate change as well as socio-economic transformation in rural areas. Climate change has resulted in higher mean temperature and lower relative humidity, whilst socio-economic change has led to the abandonment of farms, resulting in an increase and an unusual accumulation of forest fuels (Villar del Hoyo et al., 2011). The accumulation of forest fuels and the higher temperature can potentially lead to the outbreak of wildfires. Yet however, the risk of wildfires is not expected to be uniform since it is not only the quantity but also the vegetation type which is directly related to the probability of ignition of a wildfire. Other factors associated to the risk of wildfire ignitions are related to variables such as vegetation types, human activities, land use among others (Moreira et al., 2011; Carmo et al., 2011).

The spatial variation shown by climate and by vegetation composition and coverage and by socio-economic factors suggest strongly that wildfire risk is not evenly distributed over medium to large sized areas. It is known for example that quantity and quality of forest fuels are related to topography and

geomorphology, which also affect relative humidity and rainfall (Caballero et al., 2007).

Wildfire incidence is high in the Mediterranean region of Spain due to the high temperatures and low humidity during summer months, a climatic trend with high seasonality that extends to the last months of spring and the beginning of autumn. In the province of Castellón (northeast of Spain), the process of afforestation in different agricultural areas and the increasing abandonment of rural activities have led to a situation of high vulnerability to fires, particularly in Mediterranean mountainous areas, where the aforementioned factors have led to forests being abandoned and their subsequent expansion, proliferation and fuel continuity (Bastarrika and Chuvieco, 2006). The four major natural causes of wildfire ignitions are lightning, volcanic eruption, sparks from rock falls, and spontaneous combustion (National Interagency Fire Center, 2011). In Spain, a significant proportion of the wildfires is provoked by arsonists (Ministerio de Medio Ambiente, 2006), making this kind of fires difficult to predict using only the weather risk index.

Predicting where a wildfire will occur is not possible due to the many factors involved in the process of wildfire incidence. However, one may construct maps showing the probability of wildfire ignition or risk maps, using data of wildfire incidence aggregated by non overlapping geographical regions with known area or by the use of the coordinates of the starting location of past wildfire events, assuming an adequate statistical model. Risk maps are needed in the process of risk management by government agencies managing natural resources, by fire fighting agencies and by the general public.

Methods to construct maps of wildfire ignition risk should be based on easily measurable variables (covariates) to be useful for forest and risk managers. An approach that has been followed to construct wildfire risk maps is based in logistic regression. For example, Villar del Hoyo et al. (2011) used logistic regression to model human-caused wildfire risk in central Spain. Díaz-Avalos et al. (2001) used a spatial autologistic model with covariates to construct forest fire risk maps in Oregon, using a bayesian approach. Other approaches have used spatial point process models, in which the so called intensity function is proportional to wildfire risk (Juan et al., 2012; Mateu et al., 1998; Møller and Díaz-Avalos, 2010; Turner, 2009; Serra et al, 2014). The inclusion of covariates in the modeling helps to account for the spatial variation associated to the association between the expected mean number of fires per unit area and the covariates, thus reducing the variance of the model parameter estimates (Martínez-Fernández et al., 2013).

Statistical models are usually preferred because they provide a measure of uncertainty for the inferences derived from the risk maps, leaving a quantitative error margin for managers and decision takers. Further, some of the model parameters often have a physical or a biological interpretation which can give ecologists and forest engineers answers about scientific questions of interest.

In this paper we analyze the incidence of wildfires in the province of Castellón, in Spain in order to identify risk factors associated to wildfire incidences during the years 2001-2006 using the locations of the wildfires centroids. Our goals are to construct wildfire risk maps for the province, to identify which factors are relevant to explain the spatial variation of wildfire incidence and to analyze the interaction among wildfires. To attain our goals we need to find and fit statistical models to the observed spatial pattern of wildfire incidence. The resulting models and risk maps can provide aid in tasks such as planing wildfire fighting campaigns, to assess the hazard for the human populations and can also be helpful to plan fire prevention and pre-suppression activities. We use the discrete nature of wildfire events to build such models using point process theory and methods.

The rest of the paper is organized as follows: In section 2 we describe the study area and the data sets. Section 3 gives all the details needed to clarify why and how we fit the models as well as some diagnostics to assess model fit. Section 4 describes and discusses the results and the interpretation and implication of such results. Also, in section 4 we discuss how the maps constructed can be used in planning fire fighting strategies, controlled burns and other related tasks. We finish with conclusions and description of future open research lines in section 5.



Fig. 1: Geographic location of the province of Castellón, Spain.

## 2. Study area and data.

The province of Castellón, is located in the North-Est of the Iberian peninsula (Figure 1). It is delimited by the mountain range called Sistema Iberico to the West, and the Mediterranean sea to the East. It also has borders with the provinces of Catalonia in the North and Valencia in the South. Castellón has a

surface area of 6632 square kilometers, which represent 1.3% of the Spanish national territory. The topography and geology of Castellón allows the existence of a wide variety of vegetation types, which include coniferous forests, oak forests, mixed forests, shrubs and grasslands among others. According to the Coordination for the Information on the Environment (CORINE), there are forty four vegetation types in the province of Castellón. Some of the tree species that compose these vegetation types produce litter that is highly flammable, as is the case of the mediterranean pine (*Pinus halepensis*).

The data base considered in this paper includes information about all the wildfires recorded in the study area during the period 2001-2006 . The information includes the geographic coordinates of the centroid of the fire at its final size, the year, elevation, slope, aspect, land use, distance to nearest road to the wildfire's centroid, isothermality and soil permeability. This last covariate may be taken as a proxy for fuel moisture, as less permeable soils tend to retain water and moisture. However, soil permeability may well relate with fuel moisture of the deeper horizon in the forest floor (humus) for some point after rain. Thus, moisture should not be expected to exert an influence over longer time periods, because the moisture of slow-drying fuels depends essentially of time since rain. Note also that slow-drying fuels contribute very weakly to fire ignition and fire spread. Although other authors such as Koutsias et al (2013) have acknowledged the association between fire incidence and precipitation, we did not have those data available for our study area during the time scope of this study. Except for the geographic coordinates and the year, the rest of the information was obtained from the corresponding digital maps for the province of Castellón. Figure 2 shows the distribution of the wildfires occurring in Castellón by year for the time period considered in our analysis. Except for time, all the other variables were also used as spatial covariates. In particular, four continuous covariates: slope, aspect, distance to nearest road and elevation, and one categorical covariate (land use) were included in the modelling process.

Slope is the steepness or degree of incline of a surface. As slope cannot be directly computed from elevation points; one must first create either a raster or TIN surface. In this paper, the slope for a particular location was computed as the maximum rate of change of elevation between that location and its surrounding pixels. Slope was expressed in degrees.

Regarding land use, we used the CORINE database (Coordination of Information on the Environment). In particular, we used the CORINE land cover map for the year 2006 (European Environment Agency, 2007 and Heymann et al. 1994), on a 1:100.000 scale with a minimum mapping unit (MMU) of 25 hectares; the linear elements listed are those with a width of at least 100 meters.

In this paper we reclassified land use into nine categories (Table 1). We arbitrarily assigned a probability of zero to wildfire risk in category 9, so from now on we will refer exclusively to categories 1 through 8. All the covariates



Fig. 2: Spatial location of wildfires occurring in Castellón from 2001 to 2006. 2001 (*top left*) and 2006 (*down right*)

code	Land use
1	Coniferous forests
2	Dense forests
3	Fruit trees and berries
4	Artificial non-agricultural vegetated areas
5	Transitional woodland Scrub
6	Scrub
7	Natural grassland
8	Mixed forests
9	Urban, beaches, sand, bare rocks and water bodies

Table 1: Land use in Castellón and codes assigned. The categories represent an aggregation of a more diverse categorization for land use published by Heymann, 1994.

were defined on a grid of 360000 points covering Castellón. The digital images for some of the covariates are shown in Figure 3. The main causes of wildfires in Castellón during 2001-2006 were lightning strikes, followed by negligences and arsonism. It is expected that fires due to those causes be associated to different risk factors. For this reason, we made separate statistical models for wildfires caused by lightning strike (natural cause) and for negligence, arsonism and other human related causes.

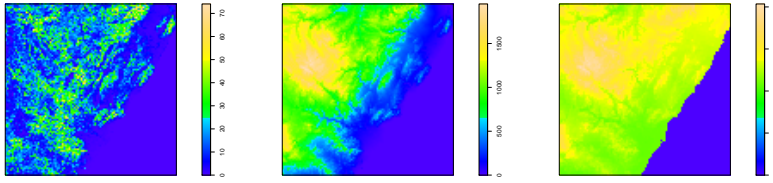


Fig. 3: Spatial variation of the covariates available for this study. (a) Continuous: Slope (left), Elevation (center), Permeability (right)).

### 3. Statistical analysis

The statistical analysis of the wildfire patterns for the years 2001-2006 in Castell3n was done using the theory and methodology of spatial point processes (Cressie, 1993; Diggle, 2003, M3ller and Waagepetersen, 2007). Although we could have followed the most common logistic and auto logistic approaches, those methods cannot be extended easily to different spatial scales. On the other hand, as spatial point process modelling is based on the first and second order properties. The spatial continuity of such properties allows models fitted at a given scale to be adapted to a coarser or finer resolution scale without much difficulty. A full description of such theory is beyond the scope of this paper, and the interested reader is referred to the previously mentioned authors. Here we will only describe the theoretical parts that are relevant to explain our analytical methods.

#### 3.1 First and second order properties.

We will denote the points of a spatial point process by  $x = (\text{long}, \text{lat})$  and we will make a distinction of the theoretical spatial point process and the spatial point pattern observed whenever the context of the text does not make clear to which one of the two we are referring. Point process models can be characterised by their first and second order properties, which are analogous to first and second moments of ordinary random variables. We will denote by  $\mathcal{A}$  the study area, also known as the observation window.

The (first-order) intensity function of a spatial point process is defined as

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\}, \quad (3.1)$$

where  $dx$  denotes a small region containing the point  $x$ , and  $E[\cdot]$  denotes the expected value of the random variable inside the brackets.

$\lambda(x)$  may be interpreted as the expected number of events occurring in an infinitesimal set of area  $dx$  and is known as the intensity function of the point process. The intensity function is unknown and has to be estimated from the data  $\{x_1, \dots, x_n\}$  from the observed spatial point pattern. Estimation may be done by posting a parametric model  $\lambda_\theta(x)$  for the intensity function or non parametrically, using kernel density estimators (Cressie, 1993; Diggle, 2003)

$$\hat{\lambda}(x) = \frac{1}{n} \sum_{i=1}^n K_n \frac{1}{h^2} \left( \frac{x - x'_i}{h} \right) \quad (3.2)$$

where  $h$  is known as the bandwidth and controls the amount of smoothing in the estimation of  $\lambda$ . Non parametric estimates of  $\lambda(s)$  are commonly used as exploratory tools to gain insight about the type of parametric model that should be fitted to the data and have been used in fire incidence mapping by Koutsias et al. (2004), De la Riva et al. (2004) and Amatulli et al. (2007).

We fitted nonparametric kernel estimators to the observed wildfire patterns for each year considered in our study. The resulting estimates are shown in Figure 4. The maps for  $\hat{\lambda}$  for each year show that wildfires in Castellón tend to occur in some fixed areas forming clusters of different sizes. Part of such clustering may be related to factors such as elevation, land use, or others. Nevertheless, the plots suggest that the class of point process models that should be fitted to the observed patterns need to consider clustering and the possible effect of covariates. Possible choices are the Non Homogeneous Poisson point process or an interaction model that can allow clustering, such as the Area Interaction.

The second-order intensity is a measure of the dependency structure of the events in  $\mathcal{A}$  and is given by

$$\lambda_2(x_i, x_j) = \lim_{|dx_i|, |dx_j| \rightarrow 0} \left\{ \frac{E[N(dx_i)N(dx_j)]}{|dx_i||dx_j|} \right\} \quad (3.3)$$

The second order intensity does not have a biological interpretation as it relates to data. An alternative very useful statistic is the reduced second moment function (Ripley 1976 and 1981)

$$K(h) = \lambda^{-1} E[N_o(h)] \quad (3.4)$$

where  $N_o(h)$  is the number of extra events at a distance  $h$  from an arbitrary event  $x \in \mathcal{A}$ . When the points of the process distribute independently at random within  $\mathcal{A}$  we speak of Complete Spatial Randomness (CSR), which can be modelled with an homogeneous Poisson process (Cressie, 1993, Diggle, 2003). In this case, the intensity function (3.1) is a constant  $\lambda$ , and a non parametric estimator of  $K(h)$  is given by

$$\hat{K}(h) = \frac{|\mathcal{A}|}{n(n-1)} \sum_{x \neq y} I(|x-y| \leq h) w_{\mathcal{A}}(x, y) \quad (3.5)$$



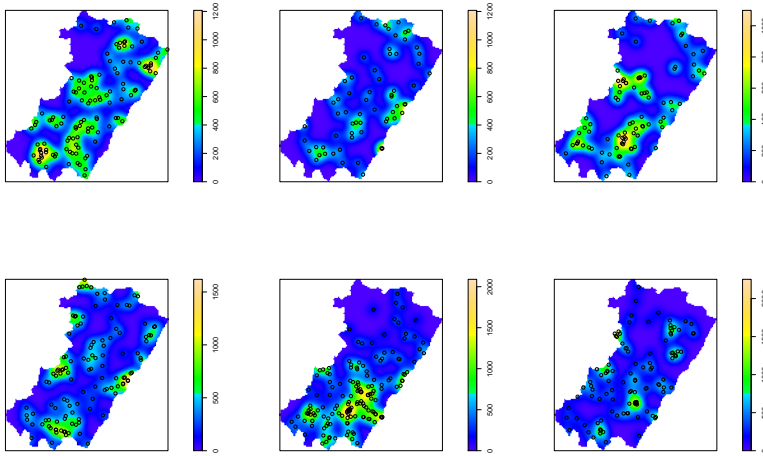


Fig. 4: Nonparametric intensity function estimates for wildfires occurring in Castellón region in the different years considered in this study using kernel density estimator.

where  $|\mathcal{A}|$  is the area or volume of the study area,  $x$  and  $y$  are arbitrary data points and  $w_{\mathcal{A}}(x, y)$  is a border effect correction.

The reduced second moment function is also known as Ripley's K-function or simply as the K-function.

The intensity function provides information related to the homogeneity of the process within  $\mathcal{A}$ , whilst second-order properties provide information related to the interaction between points in a spatial point pattern, and can be used to test the null hypothesis of CSR (Illian et al. 2008). For a point pattern with  $n$  points, non parametric tests for CSR are constructed by simulating  $s$  spatial patterns of size  $n$  from an homogeneous spatial Poisson process, under CSR and using such simulated patterns, compute confidence bands for the K-function

$$U(h) = \max_{i=2, \dots, s} \left\{ \hat{K}_i(h) \right\}$$

$$L(h) = \min_{i=2, \dots, s} \left\{ \hat{K}_i(h) \right\}$$

The hypothesis of CSR is rejected if the empirical K-function obtained from the observed point pattern fall outside the confidence bands for some set of distances  $h$ . Non parametric tests based on second order moments allow to conclude whether or not an observed spatial point pattern follows a random,

clustered or dispersed distribution pattern. Non homogeneity of  $\lambda(s)$  may often be explained through spatial varying covariates such as temperature, litter depth and elevation above sea level among others. When information on spatial covariates is available, a common choice for the form of the intensity function is

$$\log[\lambda(x)] = \mathbf{z}^T \beta \quad (3.6)$$

where  $\mathbf{z}$  is a vector whose entries are the values of the covariates at the point  $x$ , and  $\beta$  is a vector of coefficients. For model identifiability we used sum contrasts for the categorical variables, i.e., the coefficients associated to the satisfy the restriction  $\sum_j \beta_j = 0$ .

Tests of CSR which are constructed from functional summary statistics of an observed pattern such as the  $K$  – *function* are useful for two reasons: when CSR is conclusively rejected, the behavior of such summary statistics provide clues about the kind of model which might provide a reasonable fit to the data. They also suggest preliminary estimates of model parameters. In our case, we tested first the null hypothesis of Complete Spatial randomness (CSR) using the  $K$ -function in order to gain further insight about weather clustering or repulsion was evident in the observed wildfire patterns.

For those years for which the CSR hypothesis was rejected and the evidence was for clustering, we fitted first non homogeneous Poisson point process (NHPP) models, with intensity function of the form 3.6. NHPP are a natural alternative to CSR, since they preserve the assumption of independence between events of Poisson point processes. Although the NHPP models may not be good to capture strong clustering in the observed patterns, they are useful as a first step to check if the spatial variability in the observed spatial pattern may be explained by the spatial variation of the covariates. Also, in case the NHPP gives a good fit to the data, the significance of the covariate effect may be done using the fact that the coefficient estimators are asymptotically normal if the number of data is large enough (Waagepetersen, 2007), as is our case. Because the intensity function is not constant on  $\mathcal{A}$ , we used the corrected version for the  $K$  – *function* (Møller and Waagepertsen, 2007) to test goodness of fit, namely

$$\hat{K}_{inhom}(r) = |A|^{-1} \sum_{i=1}^n \sum_{j \neq i} \frac{I(0 < \|x_i - x_j\| \leq r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)} w_{ij}^{-1}. \quad (3.7)$$

For an inhomogeneous random Poisson process with intensity function  $\lambda(u)$ , the inhomogeneous  $K$ -function is  $K_{inhom}(r) = \pi r^2$ , exactly as for the homogeneous case. We can then define the  $L_{inhom}$ -function as

$$L_{inhom}(r) = \sqrt{K_{inhom}(r)/\pi}$$

Note that the inhomogeneous  $K$ -function depends on the first-order properties of the point process.

Wildfires burn areas that occupy a wide surface. Although the shape of the burned area is highly irregular, it is expected that the incidence of a wildfire in a given zone affects the probability of another wildfire starting within the burned area. This results in negative interaction between wildfires at least for some short term. Therefore, we also explored the goodness of fit of area interaction models (Baddeley and van Lieshout, 1995). For a bounded number of events  $s_1, \dots, s_n$ , the area interaction model has a density of the form

$$f(s_1, \dots, s_n) = \alpha \beta^{n(s)} \gamma^{-A(s)} \quad (3.8)$$

where  $\alpha$  is a proportionality constant,  $\beta$  is the intensity function of the process,  $\gamma$  is the interaction parameter,  $n(s)$  is the number of points of the process and  $A(s)$  is the area of the region formed by the intersection of balls of radius  $r$  centered at the points of the process. For the intensity  $\beta$  we choose a log linear form as in (3.6). The interaction parameter  $\gamma$  can be any positive number. If  $\gamma = 1$  then the model reduces to a Poisson process with intensity  $\kappa$ . If  $\gamma < 1$  then the process is regular, while if  $\gamma > 1$  the process is clustered. Thus, an Area-Interaction process can be used to model either clustered or regular point patterns. Two points interact if the distance between them is less than  $2r$ . If  $\gamma = 0$ , then a hardcore process is encountered in which there are no points within the hardcore distance ( $r$ ) of any point in the point process; that is, there is total inhibition.

Goodness of fit for the NHPP and for the area-interaction models was done by simulating patterns under the fitted models with the same number of wildfires for each year. Each of the simulated patterns was used to compute the empirical non homogeneous  $K$ -function to obtain the envelopes. We used the rank test (Grabarnik et al., 2011) to compute the 95% envelopes to get a better approximation to the true p-value for the test. To fit the models we used the  $R$  free software (R Development Core Team, 2010), and the contributed package *spatstat* (Baddeley and Turner, 2000 and 2005), which can be obtained from the  $R$  project website [www.r-project.org](http://www.r-project.org).

In all cases, the intensity function was standardized to integrate to 1.0 in the study area, thus becoming a probability density function. Therefore, the intensity function  $\lambda(x)$  is proportional to the probability of ignition of a wildfire in  $x$ , and thus to the risk of wildfire at  $x$ .

#### 4. Results

The number of wildfires occurring on each year considered in this study is shown in Table 2. The number of fires per year showed high variability during

Year	Number of wildfires
2001	120
2002	65
2003	88
2004	120
2005	160
2006	110
All years	663

Table 2: Annual distribution of wildfires occurring in Castellón by year.

the time span of this study, with an average of 110.5 fires per year. Years 2002, 2003 and 2005 were those that deviated more from the average, due to high precipitation in both years during the summer months, resulting in an increase in fuel moisture.

The spatial distribution of the wildfires (Figure 2) and the kernel estimators of the intensity function (Figure 4) indicated the presence of multiple hot spots, where wildfires tend to occur at a significant higher rate than in the rest of the province. The majority of the wildfires observed in Castellón between 2001 and 2006 occurred at distances closer than 400 meters from the nearest road. This last covariate is a proxy of human caused wildfires, as most of human activities tend to occur near roads. The empirical odds of wildfire for the different vegetation classes is shown in Figure 6, where one can see that the risk of wildfires standardised by the area covered by each land use category was higher for classes one, six and seven, which correspond to coniferous forests, scrub and natural grasslands. Areas covered by that kind of vegetation have been recognised as fire prone (Gonzalez et al., 2006).

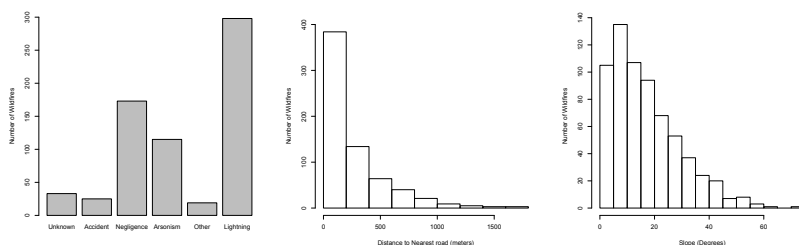


Fig. 5: Histograms of wildfire incidences by covariate (cause, distance and slope) in Castellón, for the years 2001-2006.

The multiplicative structure of the area-interaction model permits to analyze the results of the spatial and the temporal components of the model sepa-

rately. The intensity parameter was given a log linear structure dependent on the covariates, as in (3.6), where  $Z$  is a matrix whose entries are the values covariates. The parameter estimates of the significant covariates for each year are shown in table 3. The absolute value of the coefficient estimates related to elevation, slope and distance to nearest road are small due to the variation range of those continuous covariates. When multiplied by the values of the corresponding covariate and exponentiating the result gives either the increase or the decrease in the intensity function associated to the covariate. Comparing the values of  $\exp \beta_j(Z_{jk} - Z_{jl})$  provides a measure of the relative change in the probability of wildfire caused by a change  $Z_{jk} - Z_{jl}$  in the covariate. Thus for example, in 2001, a change of 100m in elevation resulted in a decrease of 10% in the probability of wildfire, keeping the rest of the covariates at fixed values. Analogous results are obtained for slope and distance to nearest road.

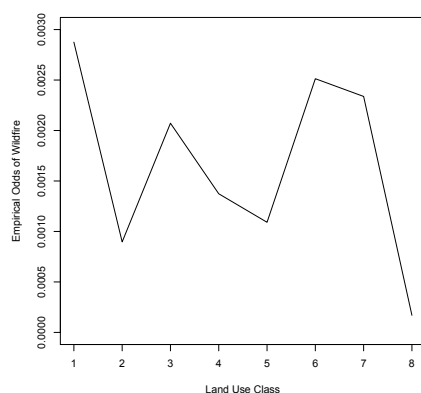


Fig. 6: The odds of wildfire incidence for different land use classes.

Coefficient Estimate	Year					
	2001	2002	2003	2004	2005	2006
$\beta_0$	4.481	5.329	6.726	-1.639	6.726	-9.749
Elevation	4.49E-4	-8.65E-4	0.002	6.99E-4	0.001	-0.001
Slope	0.022	0.002	0.030	0.023	0.028	0.048
Distance to Nearest Road	-5.90E-4	8.87E-5	-0.002	5.97E-4	3.28E-4	-0.001
Isothermality	0.045	-0.180	-0.048	0.149	0.095	0.370
Permeability	-0.003	0.011	-0.005	1.04E-5	-0.013	0.001
Interaction ( $\log[\gamma]$ )	0.621	-12.645	4.108	1.058	-27.613	0.732

Table 3: Parameter estimates for the inhomogeneous Area-Interaction models fitted to the wildfire incidences in Castellón, Spain, for the years 2001 to 2006 for natural causes. Only significant coefficients are shown.

Table 3 shows the coefficient estimates for the logarithm of the intensity function (equation 3.6) for naturally-caused wildfires. The sign and value for the different coefficients for the years considered is not constant and indicates that the importance and association of the covariates to wildfire risk is not constant over time. The coefficients for the log intensity function for human caused wildfires are shown in Table 4. Unlike the coefficients for the naturally-caused wildfires, the sign of the coefficients for most of the covariates remains constant for the years considered in this study.

Distance to nearest road has a negative or neutral effect on the probability of wildfire. This covariate has been considered a proxy of human activity, which is another important cause of wildfires in the study area as well as in other parts of the Mediterranean basin (Wittenberg and Malkinson, 2009; Villar del Hoyo et al., 2011). The negative coefficients for this covariate indicate that human caused wildfires are more likely to occur near the roads, where arsonists and other human activities are often observed.

Coefficient Estimate	Year					
	2001	2002	2003	2004	2005	2006
$\beta_0$	-1.035	4.375	-9.717	-5.453	7.426	5.739
Elevation	-0.002	-7.85E-4	-5.47E-4	-0.001	-5.59E-4	-0.001
Slope	0.027	0.0024	0.07	0.013	0.004	-0.017
Distance to Nearest Road	-0.001	-0.002	6.02E-4	-0.001	-7.27E-4	4.54E-5
Isothermality	0.202	0.163	0.457	0.333	0.185	0.090
Permeability	-1.80E-4	-0.011	-0.004	-1.003	-0.017	-0.007
Interaction ( $\log[\gamma]$ )	-1.256	-70.960	3.025	1.479	2.492	4.193

Table 4: Parameter estimates for the inhomogeneous Area-Interaction models fitted to the wildfire incidences in Castellón, Spain, for the years 2001 to 2006 for human related causes. Only significant coefficients are shown.

The parameter  $\beta_0$  can be thought as the basal risk for each year if the other  $\beta$  coefficients are kept equal to zero and the interaction parameter  $\gamma$  is equal to unity. Under this consideration, we see that the variation of  $\beta_0$  along the years followed the same trend as the yearly number of wildfires shown in table 2. Thus, we may think of  $\beta_0$  as a parameter related to the average annual risk in the whole province and in consequence to the number of wildfires per year  $\{n_t, t = 2001, \dots, 2006\}$  and that the remaining  $\beta$  coefficients increase or decrease the wildfire risk according to the spatial variation shown by the linear combination of the covariates. The table also shows that for areas covered with coniferous forests the risk of wildfire was high, in particular during 2001 and 2002. Except for areas with land use 2, 4 and 7, the fire risk associated to land use followed a similar trend along the years considered in our study, with higher wildfire risk during 2001 and 2002 and moderate risk the rest of the years.

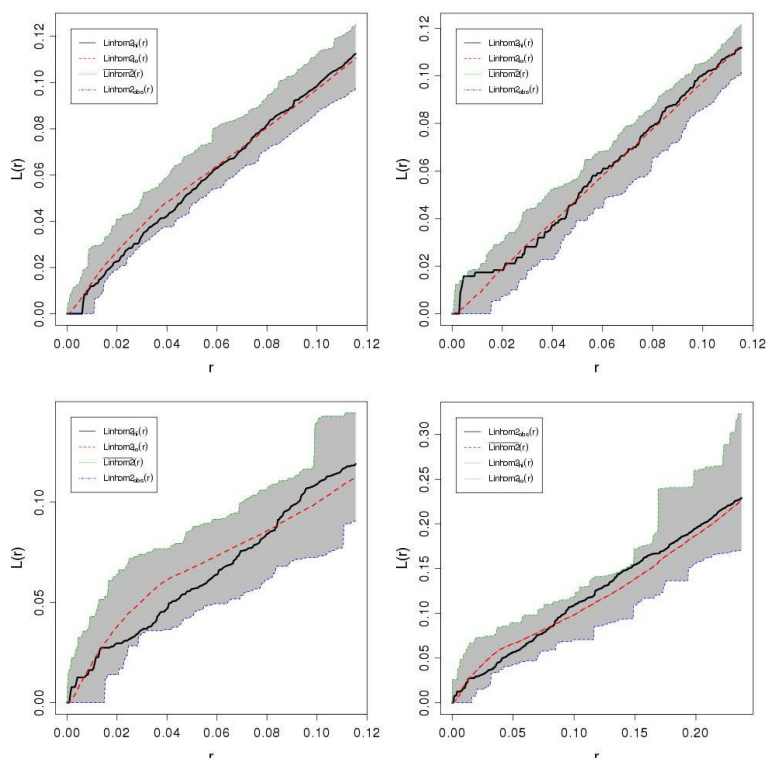


Fig. 7: The inhomogeneous L-function for Area Interaction models fitted to the wildfire incidence data in Castell3n including all the covariates, for the years 2001-2004.

The estimates of the interaction parameter  $\gamma$  for naturally-caused wildfires took values below 1.0 for the years 2002 and 2005 (Table 4), and for the years 2001 and 2002 for human caused wildfires (Table 4) indicating the presence of repulsion among the wildfires for those years and causes. This results in a more regular pattern of wildfire spatial distribution than under complete spatial randomness. For the rest of the years in this study, the positive value of  $\gamma$  indicates the presence of clustering of wildfire incidences. The fact that the parameter  $\gamma$  is not closer to 1.0 provides evidence that the incidences of wildfires in Castell3n are not independent events, but the result of a complex process where the incidence of a wildfire in a given location somehow affects the probability of ignition of other wildfires.

Figure 8 shows the intensity function estimates for the naturally-caused wildfires. The areas of high intensity show different location for the different years. This is a result of the variability of the covariate effects described in previous paragraphs where we described the results of the models fitted to the

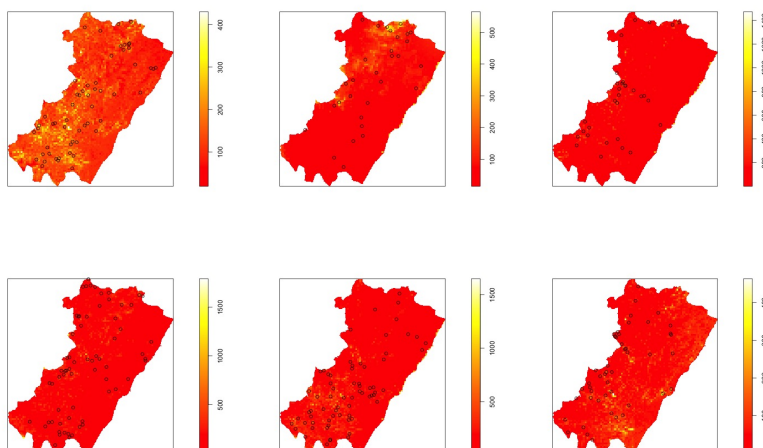


Fig. 8: Esimated intensity function obtained from the area-Interation models for the years 2001 to 2006 in Castellón, natural causes.

intensity function for the observed wildfire patterns. The map for the intensity function of human caused wildfires (Figure 9) shows that the higher risk values occurred in a centered band going from south west to north east, where the elevation ranges between 0 and 500m, where the highest risk was located in the south east part of the province, not in high elevation areas where the Iberian Range is located in Castellón. Note that the maps do not resemble the spatial pattern of the covariates (Figure 3), as what we show in the maps in Figure 9 is the combination of all the components in model 3.8.

The risk maps also provide insight about the spatial variation of wildfire risk, and for non overlapping areas  $A$  and  $B$  within the study area, the ratio ( $q$ ).

$$q = \frac{\int_A \lambda(u) du}{\int_B \lambda(u) du}$$

may be evaluated to get a measure of the relative risk of wildfire in those areas or to compare the estimated annual changes in wildfire risk in the same area if we take  $A = A_t$  and  $B = A_{t+k}$  with  $k = 1, 2, \dots$ . This information is useful for risk managers in the process of risk evaluation and as an input for risk management (Caballero et al., 2007; Hanewinkel et al., 2011). Government agencies and forest managers can also use relative risks to decide where to put the resources for fire control and fighting and provide a more efficient response in case of wildfire occurrence.



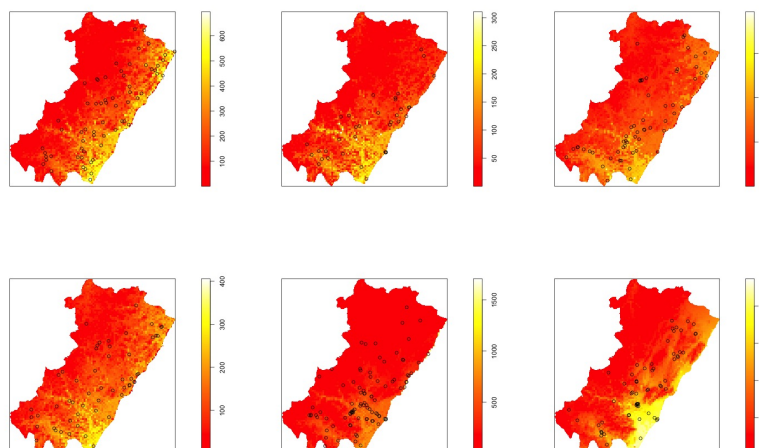


Fig. 9: Estimated intensity function obtained from the area-Interaction models for the years 2001 to 2006 in Castellón, human causes.

## 5. Discussion

The leading cause of wildfires in Castellón during 2001-2006 were lightning strikes, followed by negligences and arsonism (Figure 5, upper). These three causes accounted for over 75% of the wildfires during the time span of our study. This supports the assertion that climate change and changes in land use are among the main factors associated to wildfire incidence (Pausas, 2004). For the particular case of Castellón, land use changes are the consequence of rural abandonment. Terrains with slopes lower than 30 degrees and closer than 400 m to the nearest road are relatively easy to access for humans. Our results show that most of the wildfires occurred in hill sides facing South or East, in slopes with less than 30 degrees. Hillsides facing East and South receive a higher amount of solar energy along the year, and thus the moisture content in the fuel in those hillsides is expected to be lower than in hillsides facing North or West, resulting in different risk of wildfire ignition. This was confirmed through formal statistical analysis, although factors such as relative humidity and wind speed are likely to influence the risk of fires, particularly for West facing slopes. Note however that our analysis is on an annual basis and that variables such as relative humidity and wind speed vary at an hourly basis, and that the average annual variability due to those and other non observed covariates is included in the residual point process (Baddeley et al., 2005) in an analogous way to generalized linear models. Higher temperatures are perhaps associated to the highest number of wildfires per year during 2005. Wildfires

tend to be less selective as their size increase, so factors such as vegetation type are less significant for bigger wildfires (Barros and Pereira, 2014). This explains in part why land use was not a significant covariate in the models.

For 2004 all the covariates showed a positive association to wildfire incidence in Castellón. For year 2001 on the other hand the elevation, slope and isothermality were associated to increase in the wildfire incidence whilst distance to the nearest road and soil permeability were associated to lower incidence. For the rest of the years there was also variation of the signs of the coefficients, indicating that the effect of most of the covariates on the wildfire incidence is not constant along time. Slope was the only covariate that showed positive association for the whole study period for naturally-caused wildfires. This variability in the effects of the covariates is indicative that naturally-caused wildfires may occur in the whole province of Castellón but show a slight association to physical and climatic factors, which show variation along time.

Elevation and permeability showed negative coefficients, something that could be expected because the higher the elevation the lesser human presence in the province. Because soil permeability can be considered a proxy of fuel moisture, the negative coefficients for this covariate suggest that human caused wildfires tend to occur in areas where fuel moisture is lower. On the other hand, slope and isothermality showed positive coefficients for the six years. The positive coefficients for isothermality indicate that human caused wildfires are more likely to occur in areas where temporal variation of temperature is low, perhaps due to the higher population density in those areas. In Castellón such areas are found mostly in coastal areas, where dry and hot weather conditions prevail most of the year. Regarding slope, the coefficients have the same sign as for naturally-caused wildfires, which suggests that the effect of slope is the same regardless of the cause, and the positive values may be due to the fact that fire propagates easier in steep terrains, facilitating a small fire to become a wildfire. The relative risks between the different land use categories were not constant across time. This may be because in a given year wildfires may occur at a slightly higher rate in areas with a particular land use or have larger wildfires, suggesting that for the next year the fuel loads are lower and therefore the expected number of wildfires will be smaller. This is in fact the idea behind the use of controlled fires to reduce wildfire incidence (Hutchinson et al., 2005; Clarck et al., 2014).

The presence of interaction between the wildfires is probably a consequence of the burned area in previous years, which left a patchy arrangement of unburned areas where wildfires can occur. It is not clear the reason for the change from repulsion to clustering in the observed wildfire pattern, although a possible explanation is that the repulsion during years 2001 and 2002 for human related causes and 2002 and 2005 for natural causes left wide unburned areas that became fire prone the last two years. Another possible explanation is that the number of wildfires and the burned area are not too high and there-

fore the change from a regular distribution pattern to a clustered one is due to the variation in the number of wildfires along the years under study. The area interaction model covers a wide range of possible scenarios, going from regular patterns with strong repulsion to point processes with moderate clustering. The fit to the wildfire data of Castellón was acceptable, as no empirical  $L - function$  lied outside the confidence bands obtained by simulating point patterns using the parameter estimates for each year.

For human caused wildfires the high intensity areas are located near coastal areas, in particular in the southern part of the province, where the capital city Castellón is located (Figure 9). This Figure shows a well defined pattern, associated to the areas where most of the human population (about 80%) lives, making those areas more likely to have a human caused ignition, which combined to the dry weather prevailing in the coastal areas makes the presence of a wildfire to have a higher probability of ignition. The expansion of the low risk zone is in part because of the increased amount of rain, which has a direct association to elevation and to the reduced human presence due to farm abandonment. Overall, we see from Figures 8 9 and that in terms of wildfire risk the province of Castellón may be roughly divided in three zones: A high risk zone comprising the coastal plains where most of the human activities take place, a medium risk zone is distributed in a central strip in the NE-SW direction, which contains areas where farms and agricultural activities used to take place and where field abandonment has increased fuel loads in the recent years, and a low risk area, located in the NW part of the province, where most of the forest of the province are located.

Other authors have approached the problem of producing forest fire risk maps using methods such as GIS (Vázquez and Moreno, 1998), logistic regression (Presley and Westerling, 2007), autologistic regression (Koutsias, 2003) and empirical generalized linear models (Barbero et al., 2014) among others. Some of such approaches are either based on pure descriptive methods or ignore the presence of spatial dependence observed in the data. The use of inferences based on pure descriptive methods may be misleading because the variability of empirical data sets, and therefore posing a model for the observed data as we have done, has the advantage of providing external validation to the inferences obtained from such models. Our approach also acknowledges the presence of spatial dependence through the incorporation of spatially varying covariates and through the inherent spatial association of non Poisson point process models like the ones we have fitted to the wildfire data in Castellón, allowing the assessment of the effect of changes in the covariate values on the risk of wildfires for different locations within the study area. Although posing and fitting models may be difficult in some cases, it is a clear improvement as compared to pure descriptive or non spatial methods of analysis of fire risk.

## 6. Conclusions

The first step in risk assessment is the estimation of the probability of wildfire. The standard tool to model the probability of fires in forest ecosystems has been logistic regression as in Witenberg and Malkinson (2009) or in Vilar del Hoyo et al. (2011). Here we have proposed the use of spatial point process models with interactions, a well known methodology that acknowledges the non independence between wildfire events. The general form of the model we used allows its application in wider areas as long as information about wildfire incidence and covariate information is available. It is worth pointing that the significance of the covariates used for modelling wildfire risk in a given area is likely to change when the model is applied in a different area, or if the study area becomes broader than the original. This however, is not a disadvantage of the model used here, as the general form allows to adapt the model in case drastic spatial changes in vegetation composition or in topography may occur. Our results show that the spatial variation of forest wildfire risk is associated to the environmental covariates and to human factors considered in our analysis. Distance to nearest road was used as a proxy for human activity in the study area, but the model fit may improve if information about population density, socioeconomic activities and other related information becomes available. Castellón's province is a mountainous area and forest are restricted to the interior part of the province. Unlike what happens in other provinces of Spain, such as Catalonia, where forest wildfire risk is associated to abandonment of agricultural land (Serra et al., 2014), wildfire presence in Castellón is higher in forest covered areas. However, for the last two years analysed the higher risk zone occurred in the flat area close to the sea. This change in the spatial distribution of wildfire risk is likely associated to non observed human related covariates, as the vegetation cover and geographical features of the province did not change during the time span of this study.

We have compared the incidence of wildfires separated in two wide classes, which has showed that both classes have a different association with the climatic and physical factors included as covariates. Naturally-caused wildfires do not show a specific spatial pattern for the different years but are nevertheless slightly associated to the covariates we have included in our study. On the other hand, human caused wildfires show a well defined pattern with higher incidence associated to the covariate values observed in coastal areas.

Although kernel estimates have been used previously to study the spatial distribution of forest fires incidence (Díaz-Avalos and Alvarado, 1998; Koutsias et al., 2004; de la Riva et al., 2004; Amatulli et al; 2004), those are only nonparametric estimators and are not useful for spatial prediction nor to asses the existence of association between wildfire incidence to suspected risk factors. Parametric models such as the ones fitted in our study besides providing predictive estimators for the association between the intensity function

of wildfire incidence and are also useful to assess the effect of changes of those factors on such intensity function.

The statistical modeling approach to forest wildfire risk mapping provides to forest fire managers a valuable tool for planning forest wildfire prevention task and surveillance. These maps show not only areas with high wildfire risk but they also may be used to distinguish between those forest areas with high risk of wildfire by natural causes and those with high risk of wildfire because of human activities. Point process methods are a sensible approach to model the probability of wildfire ignition. However, this approach works better if reliable data bases including historical records of forest fire locations as well as digitalized maps of spatial varying variables are available to the modeler. Such data bases are becoming available due to the increasing quantity of data acquired with remote sensing technologies and the increasing accessibility to such information. Thus, it is reasonable to expect that the number of applications of spatial point process models to forest fires data will increase in the near future and that improved models and fire risk maps will become available to risk managers and to the general public. Predictive models for the final size of wildfires and the factors influencing such size are yet to be developed. Better management of wildfire risk is therefore a goal that is being attained as research continues.

## Acknowledgements

We would like to thank the Department of Infrastructure Territory and Environment of the Generalitat Valenciana, Section of Forest Fire Prevention for the assignment of geographic data.

## References

1. Amatulli, G., Perz-Cabello, F. and De la Riva, J. (2007) Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling* 200(3-4): 321-333.
2. Baddeley, A. and van Lieshout, M.N.M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.*, **47**: 601-619.
3. Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**:1-42.
4. Baddeley, A., Turner, R., Møller, J. and Hazelton, M. (2005). Residual analysis for spatial point processes. *J. R. Statist. Soc. B*, **67**(5):617-666.
5. Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point processes. *New Zealand Journal of Statistics*, **42**:283-322.
6. Barbero, R., Abatzoglou, J. T., Steel, E. A. and Larkin N. K. (2014). Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environmental Research Letters*, **9**: 1-11.
7. Barros, A.M.G. and Pereira, J.M.C. (2014) Wildfire selectivity for land cover type : Does size matter?. *PlosOne* 9(1): e84760. doi:10.1371/journal.pone.0084760.

8. Bastarrika Izagirre, A. and Chuvieco Salinero, E. (2006). Cartografía del rea quemada mediante crecimiento de regiones: aplicación en entornos mediterráneos con imágenes tm y etm+, *GeoFocus*, **6**: 182-204,
9. Caballero, D., Beltrán, I. and Velasco, A. (2007). Forest fires and wildland-urban interface in Spain: types and risk distribution. *In proc. of the IV International Wildfire Conference*, Seville, Spain.
10. Cressie, N.A.C. (1993). *Statistics for Spatial Data (Revised ed.)*. New York: Wiley.
11. de la Riva, J., Prez-Cabello, F., Lana-Renault, N. and Koutsias, N. (2004) Mapping forest fire occurrence at a regional scale. *Remote Sensing of Environment* 92(3): 363-369.
12. Diggle, P.J. (2003). *Statistical Analysis of Spatial Point Patterns*. Arnold, London, 2nd edition.
13. European Environment Agency, (2007). *Technical report No 17/2007. CLC2006 technical guidelines*. Luxembourg: Office for Official. Publications of the European Communities.
14. González, J. R., Palahi, M., Trasobares, A. and Pukkala, T. (2006). A fire probability model for forest stands in Catalonia (north-east Spain). *Annals of Forest Science*, **63**: 169-176.
15. Grabarnik, P., Myllymäki, M. and Stoyan, D. (2011). Correct testing of mark independence for marked point patterns. *Ecological Modelling*, **222**: 3888-3894.
16. Hanewinkel, M., Hummel, S. and Albretch, A. (2006). Assessing natural hazards in forestry for risk management: a review. *European Journal of Forest Research*, 130:329-351.
17. Heymann, Y., Steenmans, C.H., Croissille, G., Bossard, M. (1994). *CORINE Land Cover. Technical Guide*. Luxembourg: Office for Official. Publications of the European Communities.
18. Hutchinson, T.F., Sutherland E.K. and Yaussy D.A. (2005). Effects of repeated prescribed fires on the structure, composition, and regeneration of mixed-oak forests in Ohio. *Forest ecology and management*, 218(1-3): 210-228.
19. Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley, West Sussex.
20. Juan, P., Mateu, J. and Saez, M. (2012). Pinpointing spatio-temporal interactions in wildfire patterns. *Stochastic Environmental Research and Risk Assessment*. DOI 10.1007/S00477-012-0568-y.
21. Koutsias, N. (2003). An autologistic regression model for increasing the accuracy of burned surface mapping using Landsat Thematic Mapper data. *International Journal of Remote Sensing*, **24**(10):2199-2204.
22. Koutsias, N., Kalabokidis, K.D. and Allgwer, B. (2004) Fire occurrence patterns at landscape level: beyond positional accuracy of ignition points with kernel density estimation methods. *Natural Resource Modeling* 17(4): 359-375.
23. Koutsias, N., Xanthopoulos, G., Founda, D., Xystrakis, F., Nioti, F., Pleniou, M., Mallinis, G. and Arianoutsou, M. (2013) On the relationships between forest fires and weather conditions in Greece from long-term national observations (1894-2010). *International Journal of Wildland Fire* 22(4): 493-507.
24. Martínez-Fernández, J., Chuvieco, E. and 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**(2):311-327.
25. Mateu, J., Uso, J.L. and Montes, F. (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling*, **108**: 163-174.
26. Ministerio de Medio Ambiente (2006). Los incendios forestales en España. Decenio 1996-2005.
27. Møller, J. and Waagepetersen, R.P. (2007). Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics*, **34**: 643-711.
28. Møller, J. and Díaz-Avalos, C. (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, **37**: 2-15.
29. Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F. and Bilgili, E. (2011) Landscape

- wildfire interactions in Southern Europe: implications for landscape management. *Journal of Environmental Management* 92: 2389-2402.
30. National Interagency Fire Center (2011). The Science of Wildland Fire (available in: <http://www.nifc.gov/>).
  31. Pausas JG (2004) Changes in fire and climate in the eastern Iberian Peninsula (Mediterranean basin). *Climatic Change* 63, 337-350.
  32. Preisler, H. K. and Westerling, A. L. (2007). Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States. *Journal of Applied Meteorology and Climatology*, **46**: 1020-1030.
  33. R Development Core Team. (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, <http://www.r-project.org/>.
  34. Ripley, B.D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**: 256-266.
  35. Ripley, B.D. (1981). *Spatial Statistics*. Wiley, New York.
  36. Serra, L. Varga, D, Juan, P. Saez, M. and Mateu, J. (2014). Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994-2008. *Environmental and Ecological Statistics*, **21-3**: 531-563.
  37. Turner, R. (2009). Point patterns of forest fire locations. *Environmental and Ecological Statistics*, **16**: 197-223.
  38. Vázquez, A., and Moreno, J. M. (1998). Patterns of lightning-, and people-caused fires in peninsular Spain. *International Journal of Wildland Fire*, **8(2)**, 103-115.
  39. Vilar del Hoyo, L., Martin Isabel, M.P. and Matinez Vega, F.J. (2011). Logistic regression models for human-caused wildfire risk estimation: analyzing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research* , **130**: 983-996.
  40. Waagepetersen, R.P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes with Appendix. *Biometrics* , **63**: 252-258.
  41. Wittenberg, L., Malkinson, D. (2009). Spatio-temporal perspectives of forest fires regimes in a maturing Mediterranean mixed pine landscape. *European Journal of Forest Research* , **128**: 297-304.