Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994-2008

Laura Serra[1,2], Marc Saez[2,1], Jorge Mateu[3], Diego Varga[2,4], Pablo Juan[3], Carlos Diaz-Ávalos[5], Håvard Rue[6]

[1] CIBER of Epidemiology and Public Health (CIBERESP)

[2] Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Spain

[3] Department of Mathematics, Campus Riu Sec, University Jaume I of Castellon, Spain

[4] Geographic Information Technologies and Environmental Research Group, University of Girona, Spain

[5] Department of Probability and Statistics, IIMAS, National Autonomous University of Mexico, Mexico DF, Mexico

[6]Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway

Corresponding author:

Prof. Marc Saez, PhD, CStat, CSci

Research Group on Statistics, Econometrics and Health (GRECS)

CIBER of Epidemiology and Public Health (CIBERESP)

University of Girona

Campus de Montilivi, 17071 Girona, Spain

Tel 34-972418338, Fax 34-972418032

http://www.udg.edu/grecs

E-mail: marc.saez@udg.edu

Serra L, Saez M, Mateu J, Varga D, Juan P, Diaz-Ávalos C, Rue H. Spatio-temporal log-Gaussian Coxprocesses for modelling wildfire occurrence: the case of Catalonia,1994-2008.Environmental and Ecological Statistics 2013.

Abstract

Wildfires have become one of the principal environmental problems in theMediterranean basin. While fire plays an important role in most terrestrial plant ecosystems, the potential hazard that it represents for human lives and property has led to the application of fire exclusion policies that, in the long term, have caused severe damage, mainly due to the increase of fuel loadings in forested areas, in some forest systems. The lack of an easy solution to forest fire management highlights theimportance of preventive tasks.

The observed spatio-temporal pattern of wildfire occurrences may be idealised as a realization of some stochastic process. In particular, we may use a spatio-temporal point pattern approach for the analysis and inference process. We studied wildfires in Catalonia, a region in the north-east of the Iberian Peninsula, and we analysed the spatio-temporal patterns produced by those wildfire incidences by considering the influence of covariates on trends in the intensity of wildfire locations. A total of 3,166 wildfires from 1994-2008 have been recorded.

We specified spatio-temporal log-Gaussian Cox process models. Modelswere estimated using Bayesian inference for Gaussian Markov RandomField (GMRF) through the Integrated Nested Laplace Approximation (INLA)algorithm.

The results of our analysis have provided statistical evidence that areas closer to humans have more human induced wildfires, areas farther have more naturally occurring wildfires.

We believe the methods presented in this paper maycontribute to the prevention and management of those wildfires which arenot random in space or time.

Key words: wildfire, spatio-temporal point processes, marks, covariates, log-Gaussian Cox models, GMRF, INLA.

Highlights

- We obtain a model that maps fire risk in Catalonia.
- We have provided clues as to which risk factors are associated with which different causes of wildfires.
- Wildfires started intentionally were associated with low elevation locations.
- With wildfires caused by nature, relative risks were higher for locations far from the coastal plains, and from urban areas, roads and railways.
- Wildfires associated with human activity, are related to the accessibility of the areas.

1.- Introduction

Forest fires are considered dangerous natural hazards around the world (Agee, 1993). After urban and agricultural activities, fire is the most ubiquitous terrestrial disturbance. It plays an important role in the dynamics of many plant communities, accelerating the recycling time of important minerals in the ashes, and allowing the germination of many dormant seeds in the soil. Natural occurring forest fires are ignited by lightings. In the Mediterranean area however, many forest fires are ignited by arsonists or by other human related causes, such as negligence or by machinery in farm land areas.

In recent decades, forest fires have become one of the main environmental problems and one of the most significant causes of forest destruction in Mediterranean countries (Varga 2007). The term *forest fire* comprises any conflagration that might take place in a forest or wild land area, and includes wildfires. A wildfire is defined as an unplanned ignition caused by lightning, volcanoes, or unauthorized or accidental human actions (National Wildfire Coordinating Group (NWCG) Fire Policy Committee 2010). A wildfire differs from other fires in its extensive size, the speed at which it can spread out from its original source, its potential to change direction unexpectedly, and its ability to jump gaps such as roads, rivers and fire breaks (National Interagency Fire Centre 2011).

Wildfires are classified according to the cause of ignition, physical properties such as speed of propagation, the type of combustible material and the effect of weather on the fire (Flannigan et al. 2006). The four major natural causes of wildfire ignitions are lightning, volcanic eruption, sparks from rock falls, and spontaneous combustion (Scott, 2000). However, many wildfires are attributed to human sources directly provoking ignitions deliberately or accidentally (Pyne et al. 1996).

At the beginning of the twentieth century, 10% of Catalonia (a region located in the northeast of the Iberian Peninsula and representing 6.4% of Spanish national territory, see Figure 1), was covered by forests, whereas currently the forest represents about 61% (two million hectares) (Varga 2007; CREAF 1991).This increase in the forested area has been particularly notable in recent years, making wild areas prone to the outbreak of wildfires. However, the re-shaping of the landscape due to the social and economic changes that have occurred in the last fifty years (Díaz-Delgado and Pons 2001; Moreira et al. 2001; Loepfe et al. 2011); together with many features of global climate change in the Mediterranean basin, such as a temperature increase and a reduction in precipitation (Varga 2007), could explain the evolution of wildfires. Accordingly, the worst years of wildfires in Catalonia have been in 1979, 1991, 1994 and 1998, when more than 400,000 Ha burned (Varga 2007).

The aforementioned facts drew the attention of government agencies about the importance of having scientific studies regarding wildfire occurrence, as well as the risk factors associated as the temperature (Dever et al, 2008, Piñol and Lloret, 1998), from different perspectives (Varga, 2007). One such perspective comes from the statistical modelling of the spatial distribution of wildfires, while assessing which factors can be related to their existence. In fact in various locations around the globe, there are now many studies of the spatial patterns of wildfire risk. Without being exhaustive,and referringonly to those more recent studies, we cite works on fires, above all, in North America (Chen 2007; Yang et al.2008; Gedalof 2011; Miranda et al. 2011; Gralewitz et al. 2012), but also in the Mediterranean region (Millington et al. 2009; Millington et al. 2010; Romero-Calcerrada et al. 2010), including Catalonia (Juan et al. 2012; Serra et al.2012), as well as in Asia (Liu et al. 2012) and Oceania (O'Donnell 2011).

Wildfires can be associated to their spatial coordinates (representing, for example, the location of the origin, or the center of a burned area), the temporal instant, and the corresponding covariates. This association facilitates the representation of a wildfire as a realization of a spatio-temporal stochastic process. Spatio-temporal clustering of wildfires might indicate the presence of risk factors which are not evenly distributed in space and time. In fact, what is usually of interest is to assess the association of clustering of wildfires to spatial and seasonal covariates (Serra et al. 2012). Covariate information usually comes in the form of spatial patterns in regular lattices or as regular vector polygons that may be rasterized into lattice images using GIS (Simpson et al. 2011). The right methodological context able to deal with these pieces of information comes from spatio-temporal point processes. In particular, Log Gaussian Cox processes (LGCP) define a class of flexible models that are particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field (Illian et al. 2010; Simpson et al. 2011). These processes provide models for point patterns where the intensity function is supposed to come from a continuous Gaussian random field. In this sense, LGCP are able to mix the two main areas of spatial statistics, point processes and geostatistics. The spatial dependence amongst locations depends on the spatial structure of the underlying random field depicting a nice and clear combination between the two areas of spatial statistics.

Recently, Illian et al. (2010) have proposed a flexible framework, using integrated nested Laplace approximations (INLA), for fitting complicated LGCPs (Rue et al. 2009). However, this approach is still based on a regular lattice, and although this leads to consistent estimates if the lattice is fine enough and appropriately discretized (Waagepetersen 2004), this approach could be highly inefficient, especially when the intensity of the process is high or the observation window is large or, as in the case of wildfires, typically oddly shaped (Simpson et al. 2011).

To bypass the problem of inefficiency in the estimation under a general INLA approximation, we have tried another more computationally tractable approach based on stochastic partial differential equation (SPDE) models (Lindgren et al. 2011). On one hand, we used SPDE to transform the initial Gaussian Field (GF) to a Gaussian Markov Random Field (GMRF). GMRFs are defined by sparse matrices that allow for computationally effective numerical methods.

Furthermore, by using Bayesian inference for GMRFs in combination to the INLA algorithm, we take advantage of the many significant computational improvements (Rue et al. 2009). If, in addition, we follow the approach suggested by Simpson et al. (2011), in which the specification of the Gaussian random field is completely separated from the approximation of the Cox process likelihood, we gain far greater flexibility.

We present here, the results of analyzing data for wildfires in Catalonia for the years 1994 to 2008. The objective of this study was two-fold: (a) to evaluate which factors were associated with the presence of wildfires and their spatial distribution; and (b) to evaluate in time, the spatial variation of fire risk across Catalonia. We used two different kinds of log-linear models: Poisson regression and zero-inflated Poisson regression. In addition to the above, we were also interested in assessing the possible existence of interaction between space and time, in order to improve the quality of our models.

The paper is structured as follows. Section 2 presents the dataset and the statistical approach. The results of the statistical analysis are presented in Section 3, and the paper ends with some discussion in Section 4.

2.- Material and Methods

2.1.- Data setting

In this paper we analyzed the spatio-temporal pattern observed in the wildfires that occurred in Catalonia between 1994 and 2008. The study area encompasses 32,000 square kilometres and represents about 6.4% of the total Spanish national territory (see Figure 1). We consider a wildfire to be a fire that burns forested areas larger than 0.5 hectares, or a fire bigger than 1 hectare in mixed and non-forested areas. The total number of fires recorded in the analysis was 3,166, representing 126,989.44 hectares burned.

In Catalonia, it is the Forest Fire Prevention Service (Government of Catalonia) who is the agency responsible for identifying, in each fire, the coordinates of the origin of the fire, the starting time and the cause of the fire. In addition, they record the ending time of the fire, the hectares (and their type) affected and the perimeter of the fire. The data used in this article were provided directly by the Service, and are definitive, once tested and approved.

We distinguished between the numerouspotential causes of wildfire ignition. In particular, we considered: (i) natural causes; (ii) negligence and accidents; (iii) intentional fires or arson; and (iv) unknown causes and rekindled.

The first category includes lightning strikes or heat from the sun. The second, takes into account that human carelessness can also start a wildfire, for instance with campfires, smoking, fireworks or improper burning of trash. Negligence and accidents also includes those wildfires caused purely by chance. The third cause considers those wildfires that are started deliberately. Finally, the fourth set includes unknown causes and rekindled fires. Table 1 depicts the fires and some of their features.

In the Mediterranean region we find episodes with high temperature and low moisture for many days. These episodes, added to the increase of forest mass in the last 50 years, lack of forest management and the lack of a fire prevention policy makes this territory very vulnerable. So any cigarette, unauthorized grass burning or barbecue may produce a wildfire. It is true that until now no one has been arrested for this crime.

Many arsonist wildfires in Spain are caused for economic interests (payment of compensations, burnt wood, land price speculation, quarrels between hunters, landowners and tenants). It seems obvious that Spain needs to enact some more drastic anti-fire policing strategies.

In addition to the locations of the fire centroids, several covariates were considered. Spatial covariates were also considered Spatial covariates were also considered. Specifically, eight continuous covariates (i.e. topographic variables – slope, aspect, hill shade and elevation; proximity to anthropic areas – roads, urban areas and railways; and meteorological variables – maximum and minimum temperatures) and one categorical variable (land use).

Land use will obviously affect fire incidence, but moreover, topographic variables (slope, aspect and hill shade) affect not only fuel and its availability for combustion (Ordóñez et al. 2012), but also affect the weather, inducing diverselocal wind conditions, which include slope and valley winds. In fact, Dillon et al. (2011) point out that those topographic variables were relatively more important predictors of severe fire occurrence, than either climate or weather variables. The proximity to anthropic areas could be considered a factor explaining not only the incidence of fires in the intentional fires and arson category, but also why natural cause fires do not occur. As climatic variables, feasiblyimportant for natural cause fires and perhaps rekindled fires, we use the maximum and minimum temperatures (further details can be found in Serra et al. 2012).

In this paper, slope was the steepness or degree of incline of a surface. Slope cannot be directly computed from elevation points; one must first create either a raster or a TIN surface. In this article, the slope for a particular location was computed as the maximum rate of change in elevation between the location and its surroundings. Slope was expressed in degrees. Aspect was the orientation of the slope where the wildfire occurred, and was measured clockwise in degrees from 0 to 360. Given the circular nature of this covariate, it was transformed into four categories: 0 (north facing), 1 (east facing); 2 (south facing) and 3 (west facing).Hill shading is a technique used to visualize terrain as shaded relief by illuminating it with a hypothetical light source. Here, the illumination value for each raster cell was determined by its orientation to the light source, which, in turn, was based on slope and aspect and was also measured in degrees, from 0 to 360. Finally, elevation was considered as elevation above sea level and expressed in meters. To obtain topographic variables (DTM) we have used the MET-15 model, which is a regular grid containing orthometric heights distributed according to a 15 m grid side, and has been created for the Cartographic Institute of Catalonia.

The distances, in meters, from the location of the wildfire to urban areas, roads and railroads, were constructed by considering a geographical layer in each case. The urban area and road layers were obtained from the Department of Territory and Sustainability of the Catalan Government, through the Cartographic Institute of Catalonia (ICC) (http://www.icc.cat).

We also used the land use in Catalonia maps (1:250 000), with classification techniques applied on existing LANDSAT MSS images for 1992, 1997 and 2002(Chuvieco et al. 2010; García et al. 2008; Röder 2008). Additionally, we used orthophotomaps (1:5000) 2005-2007, to create the land use map for 2010. Specifically, we assigned the land use map just before the date of each wildfire. We assigned,asthe land use for eachbuffer, only the percentagevalue corresponding to the principal land use of thebufferwithin. In this paper, we transformed the twenty-two categories, obtained from the Catalonian Cartographic Institute (ICC) cover map of Catalonia, into eight categories: coniferous forests; dense forests; fruit trees and berries; artificial non-agricultural vegetated areas; transitional woodland scrub; natural grassland; mixed forests; and urban, i.e., beaches, sand, bare rocks, burnt areas, and water bodies. Figure 2 provides a graphic distribution of the wildfires over time and with this categorical covariate. In general, they are spread out over the eight land use categories. However, wildfires caused by negligence and accidents are mainly concentrated in four specific categories: dense forests (2); fruit trees and berries (4);natural grassland (8) and mixed forests (10).

We also included the temperatures (maximum and minimum) from up to seven days before the occurrence of the fire, in the location of the wildfire (Note that meteorological data were provided by the Area of Climatology and Meteorological Service of Catalonia). The temperatures at the point of the occurrence of the wildfire, along with the temperatures from the previous day and up to a week before, were estimated by means of a two-step Bayesian model. Further details can be found in Saez et al. (2012). In Table 2 we specify covariates and their source ordered by their importance on fire hazard generation.

Rather than constructing a fine regular lattice, we constructed a very irregular grid using buffers. The reason being, that an irregular lattice avoids the arbitrariness of assigning the summary for the whole cell (i.e. the sum of the wildfires) to the centroid of the regular cell, and instead assigns the centre where wildfires occurred. We first built a buffer of some 1,500 meters (diameter) around each of the wildfires, with the centre being defined by its geographic coordinates. Then, we merged those buffers to form an intersection. Now, we had not only buffers (those without any intersection with other buffers), but also groups of (merged) buffers that, in turn, could form intersections with other groups of (merged) buffers. We remerged those groups of buffers that showed any intersection with other groups and/or with another single buffer. We ended the process when any group of buffers (and/or single buffer) did not intersect with another group (and/or single buffer). At the end of the process, we had a grid of 'cells', i.e. each final group of buffers and/or single buffer. Specifically, we had 1,516 cells, each cell with a median of 2 wildfires, first quartile 1, and third quartile 5 wildfires. Since we follow the usual assumption that the point pattern observed is a realization of a point process defined in a space that contains the study area as a proper subset (Baddeley and Turner, 2000, Møller and Díaz-Avalos, 2010), the system of buffer cells that surround the study area is necessary to avoid the bias in the estimation of the intensity function. Since the behaviour of the intensity function outside the study area does not have an effect in the estimation process, the form of the buffer system is irrelevant (Møller and Díaz-Avalos, 2010). The partition of the study area in a system of cells in spatial point process inference is necessary to compute the approximation of the pseudo-likelihood function and obtain the estimates of the model parameters. In our study, the cell system is based on a tessellation and was built such that every point within the study belongs to a lattice cell.

2.2.-Methods

Spatio-temporal data can be idealised as realizations of a stochastic process indexed by a space and a time dimension

$$Y(s,t) \equiv \{y(s,t) | (s,t) \in D \times T \in \mathbb{R}^2 \times \mathbb{R}\}$$

where$D$ is a (fixed) subset of $\mathbb{R}^2$ and T is a subset of $\mathbb{R}$. The data can then be represented by a collection of observations $y = \{y(s_1,t_1), \dots, y(s_n,t_n)\}$, where the set $(s_1, \dots, s_n)$ indicates the spatial units, at which the measurements are taken, and $(t_1, \dots, t_n)$ the time points.

The mathematical theory of point processes on a general space is now well-established (Bremaud 1981; Daley and Vere-Jones 1988). However, most models for specific applications are restricted either to point processes in time or to the two-dimensional space. Cox processes are widely used as models for point patterns which are thought to reflect underlying environmental heterogeneity.

In the general spatial point process context, intensity stands for the number of events (fires in our case) per unit area. When writing total intensity in each cell, we refer to the number of fires per cell area. A particular problem in our wildfire dataset is that the total intensity in each cell,Λ_{jt}, was difficult to compute, and so we used the approximation $\Lambda_{jt} \approx |s_j| \exp(\eta_{jt}(s_j))$; where s_j is a point within the jth cell and $\exp(\eta_{jt}(s_j))$, is the estimated intensity function within such cell. Note that here we assume that Λ_{jt} is constant or has small spatial variation within the jth cell, so s_j could be any point inside the cell. The approximation allows the use of a GLM structure for the likelihood and therefore the computation of the estimate for $\eta_{jt}(s_j)$ is straightforward using numerical methods (Simpson et al., 2011).

This approximation allowed us to describe the log-intensity of the Poisson processes by a linear predictor (Illian et al. 2012) of the form

$$\eta_{ijtk}\left(s_j\right) = \beta_j + \log\left(Esp_{jtk}\right) + \sum_\alpha \beta_\alpha z_{\alpha,it} + S_j + \tau_t + \upsilon_{jt} \tag{1}$$

where β_j represents the heterogeneity, Esp_{jtk} the expected number of wildfires, of cause k, in cell j and year t, $z_{\alpha,it}$ the spatial covariates, β_α the parameters associated with covariates, S_j the spatial dependence, τ_t the temporal dependence and υ_{jt} the spatio-temporal interaction.

A full and detailed explanation of the role and meaning of each term in (1) will be given in section 2.4.

Log Gaussian Cox processes (LGCP) are a particular case of a flexible class of point processes known as Cox processes, and which are characterised by their intensity surface being modelled as

$$\log\left(\lambda(s)\right) = Z(s)$$

where$Z(s)$ is a Gaussian random field.

Conditional on a realization of $Z(s)$, a log-Gaussian Cox process is an inhomogeneous Poisson process. Considering a bounded region $\Omega \subset \mathbb{R}^2$, it follows that the likelihood for an LGCP is of the form

$$\pi(Y|\lambda) = \exp(|\Omega| - \int_\Omega \lambda(s)ds) \prod_{s_i \in Y} \lambda(s_i)$$

where the integral is complicated by the stochastic nature of $\lambda(s)$. However, this integral can be numerically computed using fairly traditional methods. We note that, the log-Gaussian Cox process fits naturally within the Bayesian hierarchical modelling framework. Furthermore, it is a latent Gaussian model, which allows us to embed it within the INLA framework. This embedding paves the way for extending the LGCP to include covariates, marks and non-standard observation processes, while still allowing for computationally efficient inference (Illian et al. 2012).

The basic idea is that, from a Gaussian Field (GF) with Matérn covariance function, we will use a SPDE approach to transform the initial Gaussian Field to a Gaussian Markov Random Field (GMRF), which, in turn, has very good computational properties. In fact, GMRFs are defined by sparse matrices that allow for computationally effective numerical methods. Furthermore, by using Bayesian inference for GMRFs, it is possible to adopt the Integrated Nested Laplace Approximation (INLA) algorithm, which, subsequently, provides significant computational advantages over MCMC.

Although Gaussian Fields are defined directly by their first and second order moments, their implementation is costly and provokes the so-called "*big n problem*" which is due to the computational costs of $O(n^3)$ to perform a matrix algebra operation with nxn dense covariance matrices, which is notablybigger when the data increases in space and time. To solve this problem, we analyse an approximation that relates a continuously indexed Gaussian field with Matérn covariance functions, to a discretely indexed spatial random process, i.e., a Gaussian Markov random field (GMRF). The idea is to construct a finite representation of a Matérn field by using a linear combination of basis functions defined in a triangulation of a given domain D. This representation gives rise to the stochastic partial differential equation (SPDE) approach, which is a link between the GF and the GMRF, and allows replacement of the spatio-temporal covariance function and the dense covariance matrix of a GF with a neighbourhood structure and a sparse precision matrix, respectively, typical elements that define a GMRF. This, in turn, produces substantial computational advantages (Lindgren et al. 2011).

2.3.- Zero Inflated Poisson

Data were taken for several causes of fire and when we worked with just one single cause we found some buffers without any wildfire, which led to the data having numerous zero counts. In many areas of interest, including public health, epidemiology, sociology, psychology, engineering, agriculture, among others, count data analysis is of primary interest. Typically, a Poisson model is assumed for modelling the distribution of the count observation or, at least, approximating its distribution. However, it has been observed in various applications that, the dispersion of the Poisson model underestimates the observed dispersion. This phenomenon, also called overdispersion, occurs because a single Poisson parameter is often insufficient to describe the population. In fact, in many cases it may be suspected that population heterogeneity, which has not been accounted for, is causing this overdispersion. This population heterogeneity is unobserved; in other words, the population consists of several subpopulations, in this case of the Poisson type, but subpopulation membership is not observed in the sample. Mixed-distribution models, such as the zero-inflated Poisson (ZIP), are often used in such cases. In particular, the zero-inflated Poisson distribution (ZIP) regression model might be used to model count data for which the proportion of zero counts is greater than expected on the basis of the mean of the non-zero counts (Breslow 1984; Broek 1995).

Therefore, we can also consider that N_{jt} follows a zero-inflated Poisson model, thus providing a way of modelling the excess of zeros, in addition to allowing for overdispersion.

In this paper, we analysed the two most common types of ZIP models, namely ZIP0 and ZIP1. Considering Λ_{jt} as the total intensity per cell, we can thus define the observed number of wildfires in a specific cell as

$$N_{jt} \sim \left\{ \begin{array}{c} Poisson(\Lambda_{jt}) \\ ZIP \begin{pmatrix} ZIP0(\Lambda_{jt}) \\ ZIP1(\Lambda_{jt}) \end{pmatrix} \end{array} \right\}$$

The different types of the zero-inflated Poisson models differ from the others in terms of the form of their likelihood functions (Lambert 1992).

Firstly, Type 0 (ZIP0) likelihood is in the form of

$$f(y;\theta;p) = \begin{cases} p, & if\ y = 0 \\ (1-p)\ Po(y,\theta|y>0), & if\ y > 0 \end{cases}$$

wherePo denotes the Poisson density, p is a hyperparameter given by

$$p = \frac{\exp(\theta)}{1+\exp(\theta)}$$

andθ is the internal representation of p, meaning that the initial value and prior is given for θ.

Type 1 zero-inflated Poisson model (ZIP1) is a mixture of a point mass at 0 and a regular Poisson distribution, whereas Type 0 is a mixture of a truncated Poisson (the y>0 bit) and a point mass at 0, so that the probability at zero is governed directly by p.

This means, for instance, that Type 0 can have a lower probability at 0 than a pure Poisson, (relative to the probability at 1), whereas Type 1 can only increase the relative probability for 0.

Therefore, Type 1 likelihood has the form

$$f(y;\theta;p) = \begin{cases} p + (1-p)Po(y,\theta), & if\ y = 0 \\ (1-p)Po(y,\theta), & if\ y \neq 0 \end{cases}$$

where p is a hyperparameter defined as in Type 0 and θ is the internal representation of p.

Note that, the only difference between Type 0 and Type 1 is the conditioning on y>0 for Type 0, which means that for Type 0 the probability that y=0 is p, while for Type 1, the probability is $p + (1-p)Po(y,\theta)$.

2.4.- Model specification

Let N_{jt} denote the observed number of wildfires in a specific cell s_j, j=1,…,1,516 and year t (t= 1994,…,2008). As a consequence of the definition of the LGCP, N_{jt} may be considered as an independent Poisson random variable (Simpson et al. 2011). Summing up, we specified our LGCP defined in (1) with four explicitfeatures.

1) We specified a spatio-temporal mixed model with two levels, the wildfire, with subscript *i* (*i=1,…,3,166*); and the cell to which the wildfire belonged, with subscript *j* (*j=1,…,1,516*). In addition, subscript *t* (*t=1994,…,2008*) denoted the year the wildfire occurred, and subscript k (*k=1,…,4*) denoted the cause.
2) We included in the model (1), as an offset, the expected number of wildfires, of cause k, in cell *j* (and year *t*), Esp_{jtk}.We constructed this variable as a sample (one per cell) from a Poisson distribution with mean equal to the average of wildfires (per cause) per cell in the year *t*. In fact, we were not interested in the (predicted) number of wildfires per cell and year, or in the effect of covariates on the (predicted) number of wildfires. Rather, our interest was in the relative risk (RR) of wildfires per cell and year, as well as the effect of covariates on such relative risk. Directly analyzing the number of wildfires per cell does not give us a reference for determining whether the occurrence of wildfires is higher or lower than expected. Relative risk is a ratio of the observed number of wildfires, of cause k, in cell *j*, divided by the expected number of wildfires, of cause k, in cell *j*. It is the risk of an event relative to exposure. That is to say, if the risk of a wildfire occurring was higher than (RR>1), equal to (RR=1) or less than (RR<1) expected.
3) Note that, we included only spatial covariates $z_{\alpha,it}$ as explanatory variables of the relative risk of a wildfire. That is, all covariates were included at the level of the wildfire, not the cell (the subscript was *i*). β_j denoted (unknown) parameters associated with covariates. With the exception of temperatures (both maximum and minimum), we categorised all continuous covariates. Thus, we approached a possible non-linear relationship between the covariate and relative risk parametrically. The finer the categorization, the closer it is to the possible nonlinear relationship. In fact, we preliminarily tested directly with continuous variables and other categorizations (seventh percentile, quartiles and thirds), but it provided a better fit were the quintiles.In addition, the categorization of a continuous variable allows for a better interpretation, because the relative risk associated with the quintile (in our case) is interpreted in relation to the reference quintile (the first, in our case).
4) We introduced four random effects in (1): (i) heterogeneity, i.e. β_j accounting for variation in relative risk across different cells; (ii) spatial dependence, S_j; (iii) temporal dependence, τ_t and (iv) spatio-temporal interaction, υ_{jt}. Note that, we assume

separability between spatial and temporal patterns and allow interaction between the two components.

The heterogeneity was specified as a vector of independent and Gaussian distributed random variables on *j*, with constant precision (R-INLA, 2012).

When spatio-temporal geostatistical data are considered, we need to define a valid spatio-temporal covariance function. For the spatial covariance structure we used the Matérn family, which specifies the covariance function as $\Sigma_{ij} = Cov(\theta_{it}, \theta_{ju}) = \sigma_C^2 M(s_i, s_j | \nu, \kappa)$ where $\sigma_C^2 > 0$ is the variance component and

$$M(h|\nu,\kappa) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa\|h\|)^{\nu}\mathrm{K}_{\nu}(\kappa\|h\|) \quad (2)$$

controls the spatial correlation at distance $\|h\| = \|s_i - s_j\|$. Here, K_ν is a modified Bessel function of the second kind and $\kappa > 0$ is a spatial scale parameter whose inverse, $1/\kappa$ is sometimes referred to as a correlation length. The smoothness parameter $\nu > 0$ defines the Hausdorff dimension and the differentiability of the sample paths (Gneiting et al. 2010). Specifically, we tried υ=1,2,3) (Plummer, 2008). When $\nu + d/2$ is an integer, a computationally efficient piecewise linear representation can be constructed by using a different representation of the Matérn field $x(s)$, namely as the stationary solution to the stochastic partial differential equation (SPDE) (Simpson et al. 2011)

$$(\kappa^2 - \Delta)^{\alpha/2} x(s) = W(s)$$

where $\alpha = \nu + d/2$ is an integer, $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is spatial white noise.

The main idea of the SPDE approach consists in defining the continuously indexed Matérn GF X(s) as a discrete indexed GMRF by means of a basis function representation defined on a triangulation of the domain D,

$$X(s) = \sum_{l=1}^{n} \varphi_l(s)\omega_l \quad (3)$$

where n is the total number of vertices in the triangulation, $\{\varphi_l(s)\}$ is the set of basis function and $\{\omega_l\}$ are zero-mean Gaussian distributed weights. The basis functions are not random, but rather were chosen to be piecewise linear on each triangle;

$$\varphi_l(s) = \begin{cases} 1 & \quad at\ vertice\ l \\ 0 & \quad elsewhere \end{cases}$$

The key is to calculate $\{\omega_l\}$, which reports on the value of the spatial field at each vertex of the triangle. The values inside the triangle will be determined by linear interpolation (Simpson et al. 2011).

Thus, the expression (3) is an explicit link between the Gaussian field X(s) and the Gaussian Markov random field, and defined by the Gaussian weights $\{\omega_l\}$ that can be given by a Markovian structure.

Both the temporal dependence (on t) and the spatio-temporal interaction (on j and t) were assumed smoothed functions, in particular random walks of order 1 (R-INLA, 2012). Thus, the random walk model of order 1 (RW1) for the Gaussian vector $x = (x_1, \ldots, x_n)$ is constructed assuming independent increments:

$$\Delta x_i = x_i - x_{i+1} \sim N(0, \tau^{-1})$$

The density for x is derived from its $n-1$ increments as

$$\pi(x|\tau) \propto \tau^{(n-1)/2} exp\left\{-\frac{\tau}{2}\sum(\Delta x_i)^2\right\} = \tau^{(n-1)/2} exp\left\{-\frac{1}{2}x^T Q x\right\}$$

where$Q = \tau R$ and R is the structure matrix reflecting the neighbourhood structure of the model.

Given the specification in (1), the vector of parameters is represented by $\theta_j = \{\beta, \beta_\alpha, S, \tau_t, \upsilon_{jt}\}$ where we can consider $X_i = (S, \tau_t, \upsilon_{jt})$ as the i-th realization of the latent GF $X(s)$ with the Matérn spatial covariance function defined in (2). We can assume a GMRF prior on θ, with mean 0 and a precision matrix Q. In addition, because of the conditional independence relationship implied by the GMRF, the vector of the hyper-parameters $\psi = (\psi_S, \psi_\tau, \psi_\upsilon)$ will typically have a dimension of order 4 and thus will be much smaller than θ.

Table 3 shows the results after analyzing the wildfire data with the four different kinds of LGCP. A natural way to compare models is to use a criterion based on a trade-off between the fit of the data to the model and the corresponding complexity of the model. The Bayesian model comparison criterion based on this principle is called Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002):

$$\text{DIC} = \text{‘goodness of fit’} + \text{‘complexity’} = D(\bar{\theta}) + 2p_D$$

where $D(\bar{\theta})$ is the deviance evaluated at the posterior mean of the parameters and p_D denotes the ‘effective number of parameters’ which measures the complexity of the model (Spiegelhalter et al. 2002). When the model is true, $D(\bar{\theta})$ should be approximately equal to the ‘effective degrees of freedom’, $n - p_D$. Alternatively, because DIC may underpenalise complex models with many random effects (Plummer 2008; Riebler et al. 2013), Table 3 also shows the conditional predictive ordinate (CPO) (Pettit 1990; Geisser 1993, Held *et al.*, 2009), which expresses the posterior probability of observing the value (or set of values) of y_i when the model is fit to all data except y_i.

$$CPO_i = \pi\left(y_i^{obs} \middle| y_{-i}\right)$$

Here, y_{-i} denotes the observations y with the i-th component omitted. This facilitates computation of the cross-validated log-score (Gneiting and Raftery 2007) for model choice (-(mean(log(cpo)))). Therefore, both the lower DIC and the lower (-(mean(log(cpo)))) suggest the best model. Table 3 shows that Poisson regression proved the best method for modelling both the natural, and unknown and rekindled causes, and a zero-inflated Poisson regression was better for modelling the second and third causes. Finally, the last line in Table 3 shows the effective number of parameters of the model. The larger this is, the worse the data fit for the model. A high number of parameters mean more complexity. The best models are those with a lower level of complexity and high goodness of fit.

All analyses were carried out using the R freeware statistical package (version 2.14.1) (R-Development Core Team 2011) and the R-INLA package (R-INLA 2012).

3.- Results

Table 2 shows the evolution of wildfires (1994-2008) and distinguished by cause. In general, the table shows a decreasing trend with regards to the number of wildfires over the years. Specifically, it shows a decrease in the number of wildfires from 1994 onwards, coinciding with the development of better extinction methods and favourable weather conditions. The number of fires also differs greatly between causes.

Table 4 provides total number of wildfires distinguishing by cause (natural causes; negligence and accidents; intentional fires or arson; and unknown causes and rekindled) and the numberof wildfires by buffer. The number of buffers differs between causes and depends on the number of wildfires; i.e., more fires mean more buffers. Table 4 also shows that there are a large number of buffers without wildfires. Specifically, natural causes have 94.40% zeros per buffer, followed by unknown causes and rekindled with 85.8%. The second and the third causes have fewer zeros: 41.60% and 78.20%, respectively. We can see that generally there are not many wildfires per buffer. For all causes, the percentage of buffers with more than three wildfires per buffer is below 2%.

Tables 5 to 8 show the relationship between relative risks (RR), according to the associate covariates and depending on the cause analysed. We have marked the estimated fixed effects that proved statistically significant. The RR>1 (i.e. risk factor) is highlighted, and the RR<1 (i.e. protective factor) is depicted in bold cursive.

In the category of natural causes, it seems that the higher the elevation the greater number of wildfires. In this same category, relative risk increases with the distance to urban areas, roads and railways; this is clearly because we are dealing with natural causes. That is to say, concentrations of fires by natural causes are usually in zones without human presence and zones with more difficult access. On the other hand, low values of hill shade (i.e. the presence of shadow) were associated with a smaller number of fires, although with the exception of the third quintile.

With reference torandom effects, we see not only a weak association between buffers and interaction dependence, but also an insignificant temporal association. In relation to negligence and accidents, a greater distance from both urban areas and roads and railways (from 0.72km to 10.49km) is associated with a decrease in the number of wildfires. Regarding topographic variables, high hill shade values are associated with an increase in the number of wildfires, and the higher the elevation the fewer fires. With respect to random effects, it is worth noting the presence of a significant spatial association and significant values with regard to heterogeneity. As for intentional causes or arson, a low elevation (90%-179%) increases the number of fires, and with respect to aspect, the relative risk of a wildfire was 23.54% in the fourth quintile, which is higher than other quintiles. Considering random effects, spatial dependency is even more important than in negligence and accidents, whereas heterogeneity is less significant. In the final category, topographic variables, with the exception of elevation, are generally associated with a reduction in the number of fires. In relation to random effects, the spatial and heterogeneity terms of the model are also very significant. Compared to the other terms, interaction dependency is also significant.

We have used the conjugate prior to the Poisson likelihood which is a Gamma distribution function. Indeed, with the aim of checking the robustness of our methodological choice we have used several other (non-conjugate) priors for the precision parameters (in particular Gaussian and flat priors) and the posterior distribution for the precision hyper-parameters has not changed significantly. We have thus preferred using in the paper the corresponding Gamma conjugate priors. Clearly, as used generically in INLA for the hyper-parameters, the distribution of the fixed parameters is Normal for the Intercept, as we see in the Figure 3a, and Gamma for the random effects, as we see in the Figure 3b.

With regard to the effect of temporal dependency on the relative risk of wildfire, Figure 4 shows its evolution graphically. In the first cause considered, natural causes, there is a notable temporal association. In fact, this effect decreased until 1998 and increased slowly thereafter. In relation to negligence and accidents, we see that the effect of the temporal association on the relative risk of wildfire starts to increase in 1997, but it is not until after 2002 that the tendency increases significantly. As for intentional causes or arson, the temporal effect oscillated significantly until 2006, decreasing thereafter. In the final category, Figure 4 shows that the temporal effect decreased throughout the period analysed.

On the other hand, causes 1 to 4 correspond to natural causes, negligence and accidents, intentional causes or arson, unknown causes and rekindled, respectively, and Figures 5a and 5b provide a more visual view of the different distribution of fires according to time, space and cause. Looking at the top of Figure 4, we notice that fires produced by natural causes have an important spatial and temporal variability. The intensity of the fires shows a clear spatial and temporal variation. However, in all cases, the highest risk is concentrated in the centre of Catalonia, coinciding with the most rural areas. The relative risk of negligence and accidents, even if its distribution pattern varies over the years, is in general higher in the west. However, the maps at the bottom of Figure 5a do not present large areas with high relative risk, except for the year 2008 where there is a significant focal point around the area of Lleida, which is a city in the west of Catalonia. Intentional causes or arson is that with the least change over the years. Nevertheless, it is interesting to observe that the higher relative risks are concentrated around urban areas, especially the areas of Barcelona and Girona, cities located in the central area of the coastline and in the northeast of Catalonia, respectively. Finally, fires produced by unknown causes and rekindled fires do not follow a specific pattern

4.- Discussion and conclusions

The analysis of wildfire incidence in Catalonia has provided important clues as to which risk factors are associated with which different causes. In the time frame of our study, wildfires started intentionally were associated with low elevation locations, which are easily accessible to most people, particularly arsonists. Although the relative risk of fires in this class indicates that the number of fires observed is 23% higher than the number of fires for hills facing southwest and 13% higher for lag 6 of maximum temperature, it is not easy to find an associated probable cause. The number of wildfires caused by negligence and accidents was, on average, 38% higher than the mean number of fires for hills facing southeast. The nature of this association is not clear. On the other hand, the relative risk for the covariate hill shade indicates that one must expect an incidence of wildfires between 66.9% and 284% higher than the mean in locations with hill shade values ranging between 172 and 251 degrees. The probable reason for this is that these locations have a high chance of small fires spreading quickly and becoming a wildfire. By contrast, the relative risk of wildfire caused by negligence or accident is lower than 1 for high elevations and locations far from urban areas, roads and railways, due to the lower human presence and activities in such locations. Although minimum temperature was also a significant factor for negligence and accidental wildfires, we cannot find a reasonable explanation for this.

For wildfires caused by nature, the relative risk is higher than 1.0 for locations far from the coastal plains and those locations distant from urban areas, roads and railways. For both covariates there is a clear gradient in the relative risk as these covariates increase, because the greater their value, the higher the importance of meteorological factors, such as lightning strikes or sun irradiance, in causing a fire. This, added to the lower human presence in such locations, facilitates the spreading of fire without control. An increased gradient in the relative risk was also observed for lags 1 and 4 of maximum temperature, in this case perhaps associated with a lower humidity of plant material, making it prone to becoming fuel. High temperatures combined with other effects, such as wind, increase fire danger. A slope exposed to the sun will have not only higher air and fuel temperatures, but also lower relative humidity. The lower relative humidity (<30%) rapidly dries out the fine dead fuels, and so a fire's spread rate and intensity will increase. When a fuel has more moisture, it is harder to ignite and burn. Although hills facing south receive higher sun irradiance and consequently tend to be drier, for naturally-caused wildfires the relative risk was below 1.0. Finally, for wildfires with unknown causes or rekindled fires, all covariates (with the exception of elevation) showed a significant association with the relative risk, some higher and some lower than 1. It must be said however, that elevation and distance from urban areas should be correlated, which may make it difficult to attribute single factors to fire occurrence. This complex model structure is most likely due to the fact that here we have a mix of fires from all of the different causes.

The results of our analysis have provided a deeper insight into factors associated with wildfire incidence in Catalonia, Spain, than previous studies on this subject (Serra et al. 2012). We have statistical evidence that wildfires associated with humans could be related to the accessibility of the areas at risk, whilst naturally-caused wildfires show the opposite behaviour. This does not, of course, mean that naturally-caused wildfires are unlikely in areas near urban areas or roads, for example, we simply mean that the relative importance of humans being responsible for starting a wildfire, either intentionally or not, decreases as locations become more difficult to reach. Although the model considers both spatial and temporal structure, the results do not show the superiority of such consideration. Climatic variables (maximum and minimum temperature) could explain the spatial structure but we are not sure what drives the temporal variation of wildfires occurrences on time. However, we can note that land use varies with time and it has an effect on the temporal variation of the wildfire counts[2].

Models for forest fire occurrence have been studied using different approaches (Serra et al. 2012; Juan et al. 2012). We chose the spatio-temporal point process because the nature of our data and the aim of our study suggested that this was the most sensible approach. For a wide class of point process models, the problem of evaluating the likelihood function has been solved using tessellations (Baddeley and Turner 2005). Instead, we have proposed a modification to the INLA method (Rue et al. 2009) by building a grid based on the intersection of buffers around the data points. The advantage of our approach is that it can be easily implemented within the INLA R package, using the computational advantages of INLA. The methodology we used in our analysis has allowed us to find a class of models that best fits the occurrence of wildfires distinguished by cause. In addition, we have proved that there is a spatio-temporal interaction and clearly different characteristics between the distributions of the wildfires, depending on each cause, exist. This leads to an improved predictive capability of fire risk and may contribute to the prevention and management of those wildfires which are not random in space and time, as we have shown here. It is worth noting that, fire is a natural component of all plant ecosystems on Earth, and its role is to accelerate the recycling of minerals, promote the germination of dormant seeds and open areas, and modify the composition of the forest in small areas, thus promoting biodiversity. For this reason, information such as that we have produced here, must be used with care by those agencies responsible for fire control and land management (Carmo et al, 2011, Cardille et al, 2001, Chuvieco et al, 2010).

There is at least one alternative to the ZIP model we have employed to estimate event count models in which the data result in a larger number of zero counts than would be expected. The hurdle Poisson model (Mullahy 1986; King 1989) is a modified count model with two processes, one generating the zeros and one generating the positive values. The two models are not constrained to be the same. The concept underlying the hurdle model is that a binomial probability model governs the binary outcome of whether a count variable has a zero or a positive value. If the value is positive, the "hurdle is crossed," and the conditional distribution of

[2]We acknowledge this comment to one of the anonymous reviewers.

the positive values is governed by a zero-truncated count model. The ZIP model on the other hand is a mix of two models. One is a binomial process which generates structural zeros, and the second component a Poisson model with mean Λ_{jt} , which generates counts, some of which can be equal to zero. The ZIP model then combines both components through a factor p_i that represents the probability of the zero counts coming from the binomial component, and (1- p_i) the probability that a zero comes from the Poisson component. Zero counts coming from the binomial component are also known as structural or excess zeros. Although the practical results are very similar in both approaches, ZIP models are most appropriate in our case, since there are areas in which it is not possible for a wildfire to occur, either because they are urban, aquatic or do not have sufficient forest mass to make a wildfire possible.

Our approach has some similarities to the model presented (Ramis et al. 2012) in the sense of both fitting a model based in a Poisson regression with an unstructured random effect and using a spatial random effect to account for the spatial structures of the data. However, we also consider the time component and the interaction between space and time, and we do not consider any element that follows a CAR model. Finally, our goal was to obtain a model that allows fire risk mapping and prediction in Catalonia.

The comparison between MCMC and INLA approach has already been done. Most of them use simulations and conclude the superiority of INLA against MCMC alternatives (Held et al, 2009, Wilhelmsen et al, 2009, Martino et al, 2010 and Eidsvik et al, 2012). However, recently Taylor and Diggle, 2013, point out that the INLA approach is not as faster as MALA within a MCMC strategy. It is worth noting that the version of INLA they used is earlier than 2011 and they do not take advantage of the current SPDE approach (Krainski, 2013).

Efforts to suppress wildfires have become an important problematic in last years. Current wildfire management policy is focused in suppressing almost all wildfires. Indirect costs of this achievement include the increase of dense vegetation in absence of wildfires and increasingly more intense wildfires. Furthermore, some results on climate changes argue that fire season comes earlier, stays longer each year and fires burn with more intensity. These fires could cause catastrophic damages as human lives, economics and environmental losses.

For this reason, knowledge of wildfire occurrence (space/time) and wildfire ignition causes should be considered an important part of sustainable forest management and it is essential for effective risk assessment and policy formulation. This study can help to improve current prevention fire policy. Moreover economic benefits include reduced suppression and fuel treatment costs over long term.

Conflicts of Interest

There are no conflicts of interest for any of the authors. All authors disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work.

Acknowledgements

Jorge Mateu and Carlos Díaz-Ávalos were partly funded in this work by the grant from the 'Towards Excellence' doctoral program for visiting professors, MHE2011-00258, ECD/3628/2011, from the Ministry of Education, Culture and Sport, Spain. We would like to thank the Forest Fire Prevention Service (*Servei de Prevenció d'Incendis Forestals)* of the Government of Catalonia (*Generalitat de Catalunya*) for providing wildfire data. We also thank the Environment Department of the Government of Catalonia for access to the digital map databases. We appreciate the comments of the attendees at the '3rd International Conference on Modelling, Monitoring and Management of Forest Fires', on May 22-24, 2012, at the Wessex Institute, New Forest, United Kingdom; at the Royal Statistical Society 2012 International Conference, September, 3-6, 2012, in Telford, United Kingdom; and at the 'VI International Workshop on Spatio-Temporal Modelling', METMA VI, on September, 12-14, 2012, at the Research Centre of Mathematics - CMAT of Minho University, Guimaraes, Portugal, where a preliminary version of this work was presented. We are indebted to Dr. Mark Olson by dedicating one weekend to review our manuscript. Last, but not the least we acknowledged the comments of three anonymous reviewers that, without doubt, help us improve our work.

References

Agee, J.K. (1993). Fire Ecology of Pacific Northwest Forests.Island Press, Washington, D.C.

Baddeley, A. and Turner, R (2005). Spatstat: An R package for analyzing spatial point pattern. Journal of Statistical; 12:1-42.

Bremaud, P (1981). Point Processes and Queues: Martingale Dynamics. New York: Springer.

Breslow NE (1984). Extra-Poisson variation in log-linear models.Applied Statistics; 33:38-44.

Broek J (1995). A score test for zero inflation in a Poisson distribution. Biometrics; 51:731-743.

Cardille, J. A., Ventura, S. J., and Turner, M. G. (2001).Environmental andsocial factors influencing wildfires in the Upper Midwest, UnitedStates, Ecological Applications, 11, 111–127.

Carmo, M., Moreira, F., Casimiro, P., and Vaz, P. (2011). Land use and topography influence on wildfire occurrence in northern Portugal,Landscape Urban Plan., 100, 169–176.

Chen X (2007).Spatial pattern of wildfire occurrences in Alabama, USA. International Journal of Environmental Studies; 64(2): 229-242.

Chuvieco, E., Aguado, I., Yebra, Y., Nieto, H., Salas, J., Martín, M.P., Vilar, L., Martínez, J., Martín, S., Ibarra, P., Riva, J., Baeza, J., Rodríguez, F., Molina, J.R., Herrera, M.A. and Zamora, R (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecological Modelling; 221(1):46-58.

CREAF.Ecological and Forest Inventory of Catalonia [in Catalan]. Bellaterra: CREAF, Universitat Autònoma de Barcelona, 1991.

Daley, D. J. and Vere-Jones, D (1988).An Introduction to the Theory of Point Processes. London: Springer.

Díaz-Delgado R, Pons X (2001). Spatial patterns of forest fires in Catalonia (NE of Spain) along the period 1975–1995: Analysis of vegetation recovery after fire. Forest Ecology and Management; 147(1):67-74.

Dillon GK, Holden ZA, Morgan P, Crimmins MA, Heyerdahl EK, Luce CH (2011). Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. Ecosphere; 2(12):130.

Drever, C. R., Drever, M. C., Messier, C., Bergeron, Y., and Flannigan, M.(2008). Fire and the relative roles of weather, climate and landscape characteristics in the Great Lakes-St. Lawrence forest of Canada, J. Veg. Sci., 19, 57–66.

Eidsvik J, Finley AO, Banerjee S, Rue H. Approximate Bayesian inference for large spatial datasets using predictive process models. Computational Statistics and Data Analysis 2012; 56(6):1362-1380 [available at http://www.sph.umn.edu/faculty1/wp-content/uploads/2012/11/rr2010-016.pdf

Flannigan MD, Amiro BD, Logan KA, Stocks BJ, Wotton BM (2006).Forest Fires and Climate Change in the 21^{st} Century. Mitigation and Adaptation Strategies for Global Change; 11(4):847-859.

García, M., Chuvieco, E., Nieto, H. and Aguado, I (2008).Combining AVHRR and meteorological data for estimating live fuel moisture. Remote Sensing of Environment; 112(9):3618-3627.

Gedalof Z (2011). Climate and Spatial Patterns of Wildfire in North America. Ecological Studies; 2:89-115.

Geisser S (1993). Predictive Inference: An Introduction. London: Chapman & Hall.

Gneiting T, Raftery AE (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association; 102:359–378.

Gneiting T, Kleiber W, Schlather M (2010). Matérn Cross-Covariance Functions for Multivariate Random Fields. Journal of the American Statistical Association: JASA; 105, 491: 1167-1177.

Gralewicz NJ, Nelson TA, Wulder MA (2012).Spatial and temporal patterns of wildfire ignitions in Canada from 1980 to 2006. International Journal of Wildland Fire; 21(3):230-242.

Held L., Schrödle B. and Rue H., Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA (2009), a chapter in Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir. Editors: Tutz, G. and Kneib, T. Physica-Verlag, Heidelberg [Available at http://www.r-inla.org/papers, accessed on July 18, 2012].

Illian JB, Sørbye SH, Rue H, Hendrichsen D (2010). Fitting a log Gaussian Cox process with temporally varying effects - a case study. Preprint Statistics No 17. [Available at http://www.math.ntnu.no/inla/r-inla.org/papers/S17-2010.pdf,accessed on August 13, 2012].

Illian JB, Sørbye SH, Rue H (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Annals of Applied Statistics; 6(4):1499-1530.

Juan, P., Mateu, J. and Saez, M (2012).Pinpointing spatio-temporal interactions in wildfire patterns. Stochastic Environmental Research and Risk Assessment; 26(8):1131-1150.

King G (1989). Event count models for international relations: generalizations and applications. International Studies Quarterly; 33:123-147.

Krainski E.T. (2013). The R-INLA tutorial: SPDE models. http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf. Accessed on May 13, 2013.

Lambert D (1992).Zero-inflated Poisson regression, with application to defects in manufacturing. Technometrics: 34:1-14.

Latent models.The R INLA project.http://www.r-inla.org/models/latentmodels. Accessed on August 13, 2012.

Lindgren F, Rue H, Lindstrom J (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). Journal of the Royal Statistical Society, Series B; 73(4):423-498.

Liu Z, Yang J, Chang Y, Weisberg PJ, He HS (2012).Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of Northeast China. Global Change Biology; 18(6): 2041-2056.

Loepfe L, Martínez-Vilalta J, Piñol J (2011). An integrative model of human-influenced fire regimes and landscape dynamics. Environmental Modelling& Software; 26(8):1028-1040.

Martino S. and Rue H. Case Studies in Bayesian Computation using INLA (2010) to appear in "Complex data modelling and computationally intensive statistical methods" (R-code) [available at http://www.r-inla.org/papers].

Millington JDA, Wainwright J, Perry GLW, Romero-Calcerrada R, Malamud BD (2009).Modelling Mediterranean landscape succession-disturbance dynamics: A landscape fire-succession model. Environmental Modelling and Software: 24:1196-1208.

Millington JDA, Walters MB, Matonis MS, Liu J (2010).Effects of local and regional landscape characteristics on wildlife distribution across managed forests. Forest Ecology and Management; 259:1102-1110.

Miranda BR, Sturtevant BR, Stewart SI, Hammer RB (2011). Spatial and temporal drivers of wildfire occurrence in the context of rural development in northern Wisconsin, USA.International Journal of Wildland Fire; 21(2):141-154.

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.

Moreira F, Rego F, Ferrera P (2001). Temporal (1958-1995) pattern of change in a cultural landscape of north-western Portugal: implications for fire occurrence. Landscape Ecology; 101:111-115.

Mullahy, J (1986). Specification and testing of some modified count data models. Journal of Econometrics; 33:341-365.

National Interagency Fire Centre.The Science of Windland Fire.http://www.nifc.gov/. Accessed on February 9, 2013.

National Wildfire Coordinating Group (NWCG) Fire Policy Committee. 2010. Terminology updates resulting from release of the Guidance for the Implementation of Federal Wildland Fire Management Policy (2009). NWCG#024-2010 Memorandum. Boise, ID: National Wildfire Coordinating Group. 3 p. (+ Attachment A: Terminology updates list April 30, 2010; 8 p.).

O'Donnell AJ (2011). Spatial and temporal patterns of wildfires in semi-arid south-western Australia.Thesis (Ph.D.) - University of Western Australia.

Ordóñez C, Saavedra A, Rodríguez-Pérez JR, Castedo-Dorado F, Covián E (2012). Using model-based geostatistics to predict lightning-caused wildfires. Environmental Modelling& Software; 29(1):44-50.

Pettit LI (1990).The conditional predictive ordinate for the normal distribution. Journal of the Royal Statistical Society: Series B; 52:175–184.

Piñol, J., Terradas, J. and Lloret, F.(1998): Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern spain.climatic change 38:345–357.

Plummer M (2008). Penalized loss functions for Bayesian model comparison. Biostatistics; 9(3):523–539.

Pyne SJ, Andrew PL, Laven RD (1996).Introduction to Wildland Fire.Second edition. New York: John Wiley & Sons.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.ISBN 3-900051-07-0. URL http://www.R-project.org/

R-INLA project.http://www.r-inla.org/home. Accessed on August 13, 2012.

Ramis R, Diggle P, Boldo E, Garcia-Perez J, Fernandez-Navarro P, Lopez-Abente G (2012). Analysis of matched geographical areas to study potential links between environmental exposure to oil refineries and non-Hodgkin lymphoma mortality in Spain. International Journal of Health Geographics, 11:4.

Riebler A., Held L. and Rue H (2013). Estimation and extrapolation of time trends in registry data - Borrowing strength from related populations.Annals of Applied Statistics [in press].

Röder, A., Hill, J., Duguy, B., Alloza, J.A. and Vallejo, R (2008). Using long time series of Landsat data to monitor fire events and post-fire dynamics and identify driving factors. A case study in the Ayora region (eastern Spain). Remote Sensing of Environment; 112(1):259-273.

Romero-Calcerrada R, Barrio-Parra F, Millington JDA, Novillo CJ (2010). Spatial modelling of socioeconomic data to understand patterns of human-caused wildfire ignition risk in the SW of Madrid (central Spain). Ecological Modelling; 221(1):34-45.

Rue H, Martino S, Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B; 71:319- 392.

Saez M, Barceló MA, Tobias A, Varga D, Ocaña-Riola R, Juan P, Mateu J (2012). Space-time interpolation of daily air temperatures.Journal of Environmental Statistics: 3(5), July.

Scott A (2000). The Pre-Quaternary history of fire. Palaeogeography, Palaeoclimatology, Palaeoecology; 164:281.

Serra, L., Juan, P., Varga, D., Mateu, J. and Saez, M (2012). Spatial pattern modelling of wildfires in Catalonia, Spain 2004–2008.Environmental Modelling& Software; 40:235-244

Simpson D, Illian J, Lindgren F, Sørbye SH, Rue H (2011). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. NTNU Technical report 10. http://www.math.ntnu.no/~daniesi/S10-2011.pdf. Accessed on August 13, 2012.

Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society: Series B; 64: 583–639.

Taylor, BM and Diggle, PJ (2013).INLA or MCMC?A Tutorial and Comparative Evaluation for Spatial Prediction in log-Gaussian Cox Processes.To appear in Journal of Statistical Computation and Simulation

Varga D (2007). Landscape and Agricultural Abandonment in the Mediterranean Mountains.An approximation to the Hortmoier and Sant Aniol Valleys (‘Alta Garrotxa’) from the Landscape Ecology.Doctoral Dissertation. Girona, Spain: University of Girona, ISBN: 978-84-691-1538-1.

Waagepetersen R (2004). Convergence of posteriors for discretized log Gaussian Cox processes. Statistics & Probability Letters: 66(3):229–235.

Yang J, He HS, Shifley SR (2008). Spatial controls of occurrence and spread of wildfires in the Missouri Ozark highlands. Ecological Applications; 18:1212–1225.

Wilhelmsen M, Dimakos XK, Husebø T, Fiskaaen M. Bayesian Modelling of Credit Risk Using Integrated Nested Laplace Approximations. SAMBA/47/08, Norsk Regnesentral, 2009 [available at http://publications.nr.no/BayesianCreditRiskUsingINLA.pdf].