Universitat de Girona

MIXED MODELS AND POINT PROCESSES

Laura SERRA SAURINA

Dipòsit legal: Gi. 1778-2013
http://hdl.handle.net/10803/127348

Doctoral Thesis

MIXED MODELS AND POINT PROCESSES

Laura Serra Saurina

2013

Doctoral Thesis

MIXED MODELS AND POINT PROCESSES

Laura Serra Saurina

2013

EXPERIMENTAL SCIENCES AND SUSTAINABILITY PhD PROGRAMME

Directed by:

Dr.Marc Saez (GRECS, UdG) and Dr.Jorge Mateu (UJI)

Thesis submitted in fulfilment of the requirements for the degree of Doctor from the University of Girona

LAURA SERRA SAURINA

Universitat de Girona

Dr.Marc Saez, of the University of Girona and Dr.Jorge Mateu, of the University Jaume I, Castelló

We declare:

That the Thesis is entitled “Mixed models and point processes”, presented by Laura Serra Saurina to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, we hereby sign this document.

J.Mateu

Prof. Dr.Marc Saez | Prof. Dr.Jorge Mateu

Girona, 8 of August, 2013

Acknowledgements

First of all I want to thank my supervisors, especially Marc Saez, who was the one who encouraged me to begin this career path and convince me to leave my last job in Barcelona in order to start a new life in Girona. He has helped a lot during all this time and with him I have not only learnt in academics but also helped me grow personally. He has been more than a supervisor. Moreover I want to thank Jorge Mateu because he has also been a person very important in this process. I am very grateful to accept being my second supervisor and invite me to his University in Castellon to work together and receive their mathematical advises. He has also given me the opportunity to meet people from different parts of the world which has been a very good experience.

I do not want either to forget Diego Varga who has been a great partner during these three years. We have shared a lot of moments during this time which I specially remember one of them. It was the first Congress I assisted and the one I did my first oral presentation. I want to highlight his patience and his great help he offered me during those days.

I would also like to thank in particular Håvard Rue and Finn Lindgren for their help during my pre-doctoral stage at the Department of Mathematical Sciences of the Norwegian University of Science and Technology, Trondheim, Norway. And also all the people I met there which made my stay there perfect.

On the other hand, I 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 and to the Environment Department of the Government of Catalonia for the access to the digital map databases. I appreciate the comments of the attendees at the "1st Conference on Spatial Statistics 2011", on March, 23-25, 2011, at the University of Twente, Enschede, The Netherlands; 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.

Moreover, I would also like to thank the department of the Faculty of Business and Economic Sciences in the University of Girona for helping me financially and allowing me to do stages abroad as well as attending courses and Congresses related to my research.

I do not have to forget to mention my colleagues, especially those in my office, with whom I have shared laughter, nerves and even a spinning class looking at “l’Estany de Banyoles”.

Finally, I want to pay special attention to my family for supporting me in everything and understand me at specific times of nerves. Especially I want to thank my mother for giving me helpful advises during all the Thesis process and revise some parts of this work. But even more than my family, I want to thank my partner, David Sabaté, who has lived these three years deeply. He has made me laugh when I was sad, he has encouraged me in times of stress and despair and he has also helped me throughout the development of this work. For everything I really want to thank his patience and understanding.

Published Work

This Thesis is presented as a traditional monograph format. However, during the Thesis some research results have been written. In this chapter we expose the articles published or at least, sent, during this PhD Thesis.

Publications related to this Thesis

Serra L, Saez M, Varga D, Tobías A, Juan P, Mateu J. Spatio-temporal modelling of wildfires in Catalonia, Spain, 1994-2008, through log gaussian Cox processes. In Brebbia CA, Perona G (eds). *Modelling, Monitoring and Management of Forest Fires III*. ISBN: 978-1-84564-584-7. Southampton: WITpress, 2012, pp. 39-49.

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

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.

Serra L, Saez M, Juan P, Varga D, Mateu J. A spatio-temporal Poisson Hurdle point process to model wildfires.Stochastic Environmental Research and Risk Assessment (SERRA) 2013 (under review).

Abbreviations

AIC – Akaike's Information Criterion

Atm – Atmosphere

C – Carbon (C)

°C – Degrees centigrade

C_2H_4 – Ethylene

C_3H_6 – Propane or propylene

C_4H_8 – Butane

CnHm – Unsaturated hydrocarbons unidentified in analysis

CO – Carbon monoxide

CO_2 – Carbon dioxide

CPO – Conditional Predictive Ordinate

CORINE – Coordination of Information on the Environment

CSR – Complete Spatial Randomness

DIC – Deviance Information Criterion

EEA – European Environment Agency

Exp(.) – Exponencial

GF – Gaussian Field

GIF – Big wildfire

GMRF – Gaussian Markov Random Field

H – Heat

H_2 - Hydrogen

H_2O - Water vapour

ICC – Catalonian Cartographic Institute [in Catalan]

IDESCAT – Catalan Statistics Institute [in Catalan]

INE – Spanish Statistical Office [in Spanish]

INLA – Integrated Nested Laplace Approximation

K_{inhom} – Inhomogeneous K-function

$\widehat{K}_{inhom}$ – Inhomogeneous K-functionfor the observed process

kW/m^2 – Energy intensity per unit area

LFL– Lower flammability limit

LGCP – Log-Gaussian Cox Processes

MCMC – Markov Chain Monte Carlo

MMU – Minimum Mapping Unit

N_2 – Inert gases

NWCG – National Wildfire Coordinating Group

O_2 – Oxygen

RR – Relative Risk

RW1 – Random Walk of order 1

SPDE – Stochastic Partial Differential Equation

Sq mi – Square mile

Tpois(.) – Truncated Poisson distribution

UEL– Upper explosive limit

ZIP – Zero inflated Poisson model

ZIP0 – Type 0 Zero-inflated Poisson model

ZIP1 – Type 1 Zero-inflated Poisson model

Index of figures

Chapter 1. Introduction

Chapter 2. Methodology

Index of tables

Chapter 1. Introduction

Index

Resum

Des del punt de vista ambiental, els incendis representen una destrucció de boscos i matolls, una alliberació a l'atmosfera d'una part del carboni i dels nutrients acumulats prèviament a l'ecosistema i importants efectes sobre la fauna. També tenen un efecte directe sobre els processos geomorfològics i hidrològics. D'altra banda, molts estudis mostren alguns efectes positius del foc per a la biodiversitat però la realitat és que els incendis posen en perill els assentaments humans i faciliten l'erosió del terreny.

El risc d'incendis és molt important en la regió mediterrània degut a una marcada estacionalitat, en la qual destaca un període estival caracteritzat per les altes temperatures i una baixa humitat relativa de l'aire. Si en aquesta combinació de factors climàtics s'hi afegeixen episodis de vents secs i càlids, propis d'aquestes regions, es reuneixen totes les condicions perquè es produeixi un escenari d'incendi catastròfic que pot arribar a cremar desenes de milers d'hectàrees. Addicionalment, després de l'estació seca, moment propici pels incendis, succeeix una estació amb pluges torrencials que actúa erosionant els terres desproveïts de tot tipus de coberta vegetal. A més a més, la tendència climàtica es decanta cap a un increment del número de dies estivals, amb altes temperatures i baixa humitat de l'aire i cap a una reducció de les precipitacions, que es tornaran episòdiques i més intenses.

L'objectiu principal d'aquesta tesi és modelitzar l'ocurrència dels incendis i, en particular, analitzar la variabilitat del seu comportament en funció de l'espai i el temps tot coneixent quins són els factors que, amb més o menys intensitat, influeixen en el seu comportament.

La tesi planteja tres grans objectius. En primer lloc s'analitza si les dades, en aquest cas els incendis, segueixen un patró determinat o altrament tenen un comportament aleatori. Analitzant únicament els incendis de Catalunya produïts en el període 2004-2008 i aplicant la metodologia dels processos puntuals basada en la comparativa d'un model estocàsticament independent, es descarta, en primera instància, el comportament aleatori. En segon lloc, s'estudia que la distribució dels incendis és variable en el temps i s'aplica un model que incorpora la component temporal. Aquest segon treball amplia els anys d'estudi considerant els incendis ocorreguts des de l'any 1994 fins al 2008. Finalment, particularitzem l'ocurrència dels incendis i ens interessem únicament en els incendis més grans que una extensió específica fixada (50ha, 100ha o 150ha) ja que, tot i no ser els més abundants en número són els que més extensió i més mal mediambiental ocasionen. D'aquest tercer anàlisi se n'extreu que els grans incendis són provocats majoritàriament per l'acció de l'home, ja sigui per accident o intencionat però es descarta que siguin degut a causes naturals.

Els mètodes presentats en aquesta tesi s'engloben dins la teoria de processos puntuals però, cada un d'ells, té les seves particularitats. El primer mètode analitza el tipus d'interacció entre els punts analitzats (incendis en el nostre cas d'estudi) fent una comparativa gràfica amb la funció K de Ripley, que suposa un comportament completament aleatori. El segon mètode es basa en una classe de models flexibles molt útils per modelitzar punts agregats amb informació subjacent no observada. En particular, es tracta dels processos de Cox, que són capaços de barrejar les dues principals branques de l'estadística espacial, els processos puntuals i la geoestadística. Finalment, per tractar el darrer objectiu, s'utilitza un model economètric en dues parts, concretament, el model Hurdle.

Els resultats obtinguts en aquesta tesi poden contribuir a la prevenció i a la gestió dels incendis forestals. A més, la metodologia utilitzada en aquest treball és útil per conèixer quins són els factors que fan que un incendi es converteixi en un gran incendi forestal.

Resumen

Desde el punto de vista ambiental, los incendios representan una destrucción de bosques y matorrales, una liberación a la atmosfera de una parte del carbono y los nutrientes acumulados previamente en el ecosistema e importantes efectos sobre la fauna. También tienen un efecto directo sobre los procesos geomorfológicos e hidrológicos. Por otra parte, muchos estudios muestran algunos efectos positivos del fuego para la biodiversidad, pero la realidad es que los incendios ponen en peligro los asentamientos humanos y facilitan la erosión del terreno.

El riesgo de incendios es muy importante en la región mediterránea debido a una marcada estacionalidad en la que destaca un período estival caracterizado por las altas temperaturas y una baja humedad relativa del aire. Si a esta combinación de factores climáticos se añaden episodios de vientos secos y cálidos, propios de estas regiones, se reúnen todas las condiciones para que se produzca un escenario de incendio catastrófico que puede llegar a quemar decenas de miles de hectáreas. Además, después de la estación seca, momento propicio para los incendios, sucede una estación con lluvias torrenciales que actúa erosionando los suelos desprovistos de todo tipo de cubierta vegetal. Por otra parte, la tendencia climática se decanta hacia un incremento del número de días estivales, con altas temperaturas y baja humedad del aire y hacia una reducción de las precipitaciones, que se volverán episódicas y más intensas.

El objetivo principal de esta tesis es modelar la ocurrencia de los incendios y, en particular, analizar la variabilidad de su comportamiento en función del espacio y el tiempo conociendo cuáles son los factores que, con mayor o menor intensidad, influyen en su comportamiento.

La tesis plantea tres grandes objetivos. En primer lugar se analiza si los datos, en este caso los incendios, siguen un patrón determinado o de lo contrario tienen un comportamiento aleatorio. Analizando únicamente los incendios de Cataluña producidos en el periodo 2004-2008 y aplicando la metodología de los procesos puntuales basada en la comparativa de un modelo estocástico independiente, se descarta, en primera instancia, el comportamiento aleatorio. En segundo lugar, se estudía que la distribución de los incendios es variable en el tiempo y se aplica un modelo que incorpora la componente temporal. Este segundo trabajo amplía los años de estudio considerando los incendios ocurridos desde el año 1994 hasta el 2008. Finalmente particularizamos la ocurrencia de los incendios y nos interesamos únicamente en los incendios más grandes que una extensión específica fijada (50ha, 100ha o 150ha) ya que, aunque no son los más abundantes en número son los que más extensión y más daño medioambiental producen. De este tercer análisis se extrae que los grandes incendios son provocados mayoritariamente por la acción del hombre, ya sea por accidente o intencionado pero se descarta que sea debido a causas naturales.

Los métodos presentados en esta tesis se engloban dentro de la teoría de procesos puntuales pero cada uno de ellos tiene sus particularidades. El primer método analiza el tipo de interacción entre los puntos analizados (incendios en nuestro caso de estudio) haciendo una comparativa gráfica con la función K de Ripley, que supone un comportamiento completamente aleatorio. El segundo método se basa en una clase de modelos flexibles muy útiles para modelar puntos agregados con información subyacente no observada. En particular, se trata de los procesos de Cox, que son capaces de mezclar las dos principales ramas de la estadística espacial, los procesos puntuales y la geoestadística. Finalmente, para tratar el último objetivo, se utiliza un modelo econométrico en dos partes, concretamente, el modelo Hurdle.

Los resultados obtenidos en esta tesis pueden contribuir a la prevención y a la gestión de los incendios forestales. Además, la metodología utilizada en este trabajo es útil para conocer cuáles son los factores que hacen que un incendio se convierta en un gran incendio forestal.

Summary

From an environmental point of view, fires represent a danger to forests and brush; they release some of the carbon and nutrients previously accumulated in the ecosystem to the atmosphere and seriously affect wildlife. They also have a direct effect on geomorphological and hydrological processes. On the other hand, many studies show some positive effects of fire for the biodiversity but on balance, fires endanger human settlements and facilitate soil erosion.

Fire risk is highly important in the Mediterranean region because of its seasonal nature, with summers of high temperatures and low humidity. Weather is a fundamental component of the fire environment. The prolonged drought and high temperatures of the summer period in the Mediterranean climate are the typical drivers that define the temporal and spatial boundaries of the main fire season. Future trends of wildfire risks in the Mediterranean region, as a consequence of climate change, will lead to an increase of temperature in the East and West of the Mediterranean, with more frequent dry periods and heat waves, facilitating the development of very large fires. Due to the climate change there is an increasing relationship between the number of days of extreme fire hazard weather and the number and size of fires in the Mediterranean coast of Spain.

The main objective of this Thesis is to model the occurrence of wildfires and, in particular, knowing the factors with more influence, to evaluate how they are distributed in space and time.

The Thesis presents three major objectives. Firstly it has been analysed if data - in this case, fires - follows a particular pattern or behaves randomly. Analysing only fires in Catalonia occurred in the period 2004-2008 and applying the methodology of point processes based on the comparison of an independent stochastic model, random behaviour is discarded. Secondly, this study has shown that fire distribution is variable in time, so a model which includes the temporal component is used. This second study extends the database considering fires occurred from 1994 to 2008. Finally, we focus on modelling the occurrence of big wildfires, which are those that burn areas greater than a given extension of hectares (50ha, 100ha or 150ha); even though they only represent a small percentage of all fires, they signify a high percentage of the area burned and cause important environmental damage. The main finding of this third analysis is that big wildfires are mostly caused by human action, either through negligence and accidents or intentionally but not by natural causes.

Methods presented in this Thesis are included in the theory of point processes but each one has its own specific characteristics. The first method explores the nature of interaction between the points analysed (fires in our case of study) applying K Ripley's function, a graphical tool for discarding random behaviour. The second method is based on a flexible class of point processes that is particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field. These processes, which are Cox models, are able to mix the two main areas of spatial statistics, point processes and geostatistics. Finally, to deal with the last objective an adapted two-part econometric model is used, specifically a Hurdle model.

The results presented in this Thesis may contribute to the prevention and management of wildfires. In addition, the methodology used in this work can be useful to determine those factors that help any fire to become a big wildfire.

Hypothesis

Throughout this project we formulate the following hypothesis:

1. The occurrence of wildfires in a given period can be predicted using statistical methods.
2. Wildfires are not randomly distributed in space or time but they are concentrated in certain areas and / or periods.
3. Clustering of wildfires depends on covariates, specifically on the topographic variables (slope, aspect, hill shade and altitude; proximity to anthropic areas such as roads, urban areas and railways), meteorological variables (maximum and minimum temperatures), land use and forest fuels.
4. The probability of occurrence can also be different depending on the initial cause of the wildfire.
5. Assuming separability between spatial and temporal patterns allows include interaction between the two components.
6. Wildfires bigger than a given extension (50ha, 100ha or 150ha) are mostly caused by human action either through negligence and accidents or intentionally but not by natural causes.
7. Because every wildfire can turn into a big wildfire, they are not modelled as structural zeros by a ZIP model but by a Hurdle model.

Objectives

The main objective of this Thesis is to analyse the spatio-temporal patterns produced by wildfire incidences in Catalonia, located in the north-east of the Iberian Peninsula, using spatio-temporal point processes.

Specific objectives:

1. To evaluate how the extent of clustering in wildfires differs across the years they occurred.
2. To analyse the influence of covariates on trends in the intensity of wildfire locations.
3. To analyse the spatio-temporal patterns produced by those wildfire incidences by considering the influence of covariates on trends in the intensity of wildfire locations.
4. To model the occurrence of big wildfires (greater than a given extension of hectares) using an adapted two-part econometric model, specially a Hurdle model.
5. To build maps of wildfire risks, by year and cause of ignition, in order to provide a tool for preventing and managing vulnerability levels.
6. To analyse which factors have more influence in generating wildfires bigger than a given extension (50ha, 100ha or 150ha).
7. To evaluate two different statistical alternatives (ZIP models and Hurdle models) to analyse and estimate the excess of zeros of a stochastic process.

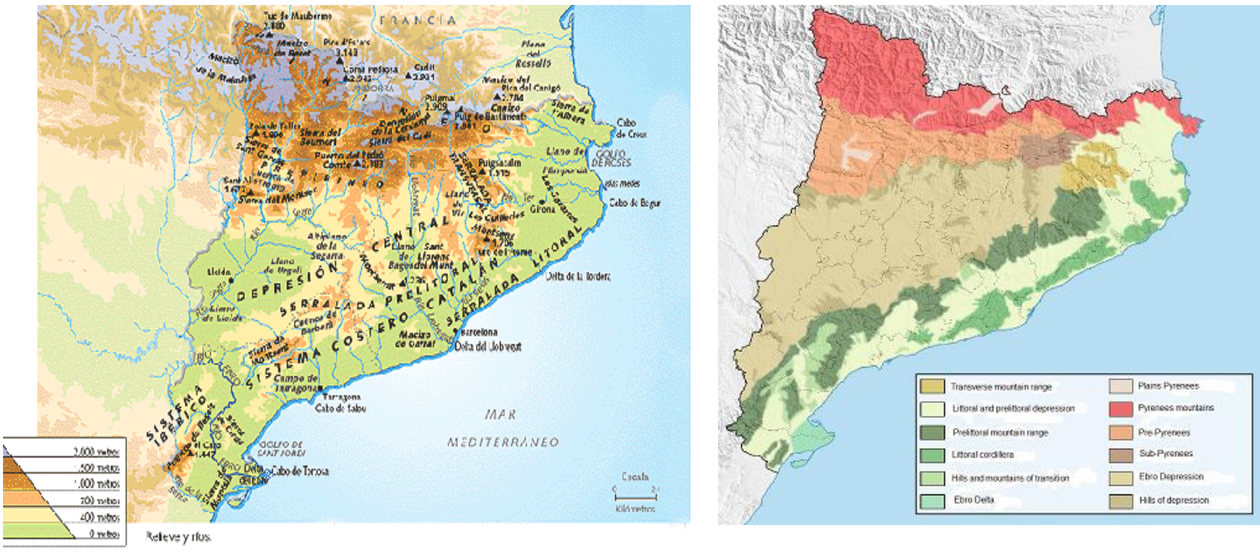
Chapter 1.Introduction

1. Characterisation of the study area

Catalonia, located in the north-east of the Iberian Peninsula, is one of the autonomous communities of Spain. The region is bordered by mountains, with the Pyrenees lying in the north and the Iberian System to the south. The region is further demarcated by the Ebro River to the south and south-west, and the Mediterranean coast to the east. It is a region with a surface area of 30,000 square kilometres (12,355 sq mi), representing 6.4% of the total Spanish national territory. According to the Catalan Statistics Institute (IDESCAT) and Spanish Statistical Office (INE), in 2010 Catalonia was inhabited by 7,512,000 people[1] of whom two thirds lived in the metropolitan area of Barcelona, a very dense and highly industrialised region.

Broadly speaking, Catalonia can be categorised into three main geographical areas: a mountainous region made up by the Pyrenees Mountains, which connect the Iberian Peninsula to continental Europe and are located in the north of Catalonia. Another region is formed by alternating elevations and plains parallel to the Mediterranean coast called the Catalan Mediterranean System, or the coastal Catalan mountain ranges, and a third element located within a flatter area called the Catalan Central Depression. Figure 1 depicts this varied geography which has a variety of landscapes, from the high Pyrenees to the curious geological formations such as the mountains of Montserrat or the now extinct volcanoes of La Garrotxa.

Figure 1: Morphostructural units of Catalonia

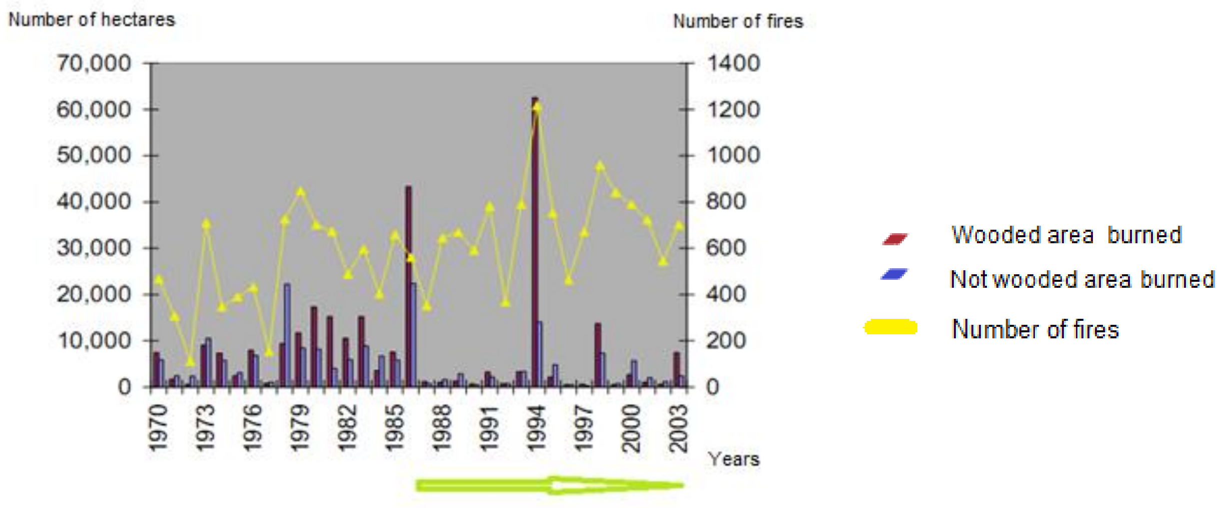


Source: Translated from http://www.zonu.com

The climate in Catalonia is not uniform throughout the region and has significant temperature variations caused by Catalonia's complex relief. This heterogeneity leads to different climate types. The coastline is characterised by a mild climate, warm in winter and very hot in summer, whereas inland Catalonia is noted for its Continental Mediterranean climate characterised by cold winters and hot summers. Finally, the mountainous areas close to the Pyrenees have a typical alpine climate featuring temperatures below zero and high winter snowfall. The annual rainfall is over 1,000 mm and summers are cool[2].

The heterogeneity of the Catalan landscape, both morphological and climatologically, gives rise to a territory of extraordinary diversity, making Catalonia a region rich in a wide variety of landscapes which can be considered as part of the country's environmental, cultural, social and historical heritage influencing the quality of the citizens' life. This wealth is a resource for economic development, particularly in tourism, but also in agriculture, livestock farming and forestry. This diversity contributes to the preservation of the biodiversity and, in particular, plays a positive role in preventing wildfires[3].

Figure 2:Evolution of the number of fires and the wooded and not wooded areas from 1970 to 2003.



Source: Catalan Fire Department and author's own construction

Figure 2 shows a cyclic behaviour with respect to the number of fires, which directly affectsthe number of wooded or not wooded hectares. Taking into account the number of hectares burned, the worst years were from 1978 to1986. However, after a peak in fires in 1994 a decrease in the annual burned area can be noticed, as well as an improvement in wildfire extinction and better climatic conditions, characterised by less harsh and wetter summers. Nevertheless, from 1996 fires continued to occur with high frequency, intensity and extension.

In 1983 the "Bombers de la Generalitat" were created, which are the firefighters of the Governement of Catalonia, and since 1987, as the arrow below the graph in Figure 2 shows, the program "Foc Verd" (Green Fire) has been implemented[4].

2. Theory on fires

2.1 Definition of wildfires

Fire is defined as the rapid oxidation of a material in the exothermic chemical process of combustion, releasing heat, light, and various reaction products. It starts when a flammable and/or a combustible material, in combination with a sufficient quantity of an oxidizer such as oxygen gas or another oxygen-rich compound (although there are non-oxygen oxidizers which can replace oxygen) is exposed to a source of heat or ambient temperature above the flash point for the fuel/oxidizer mix, and is able to sustain a rate of rapid oxidation that produces a chain reaction. The minimum temperature needed to trigger the combustion is called ignition temperature, defined in degrees centigrade (°C) at a pressure of one atmosphere (1 atm), which is the condition in which the vapours generated start to burn[5].

The result of this exothermic process is carbon dioxide (CO_2), water vapour, energy and a solid waste or ashes[5].

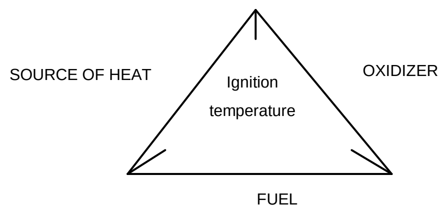
$$\text{Fuel} + O_2 + \text{Heat (H)} = CO_2 + H_2O + \text{Energy} + \text{Waste}$$

The above equation shows that in any combustion there is always a burning element, called fuel, and another which produces the combustion (oxidizer) that it usually is oxygen as gaseous O_2.

Fires cannot exist without the correct combination, in the right proportions of three elements. It requires a fuel, an oxidizer, such as oxygen, plus activation energy or ignition source. For example, a flammable liquid will start burning only if the fuel and oxygen are present in precise proportions. Some fuel-oxygen mixes may require a catalyst, a substance that is not directly involved in any chemical reaction during combustion, but which enables the reactants to combust readily.

There is a model called the fire or combustion triangle which describes these three elements graphically (see Figure 3):

Figure 3: Fire triangle



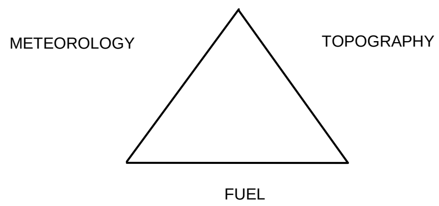
Source: Own construction from “Basic Course on Wildfires [in Spanish]”[5]

The key in preventing or attacking a fire is simply to remove any of these three factors. Thus, if one of these elements in the triangle is absent, fire cannot be generated. This is a key concept in establishing fire prevention methods which are based on the reduction or elimination of one of these elements[5].

Without adequate heat, fire cannot start or propagate, and without fuel the fire stops. This second element can be consumed by the fire itself, or it can be eliminated naturally or artificially, by introducing a retardant chemical to the flame which obstructs the chemical reaction itself until the rate of combustion is too slow to maintain the chain reaction. To prevent the fire from gaining access to the fuel there are also some physical obstacles such as firewalls. Finally, insufficient oxygen, as well the absence of heat, prevents the fire from starting or spreading.

Correspondingly, wildfire behaviour and the severity of the resulting blaze are a combination of factors such as available fuels, physical setting, and weather conditions, all of which make up the fire behaviour triangle (see Figure 4). Some authors suggest that, under extreme weather conditions and on steep slopes, the importance of fuel is, at best, relative[6]. However, from a forest management and fire prevention point of view, fuel is the only factor which can be influenced in order to modify the behaviour of a fire. The methods used to keep the fire away or to change its behaviour are isolation as well as a fuel modification or conversion[7]. The objective is to control the fire in a specific area in order to be able to attack it directly. The forests where the fuel has been modified or converted may be useful in surrounding the fire, but the main objective is to influence the behaviour of the fire.

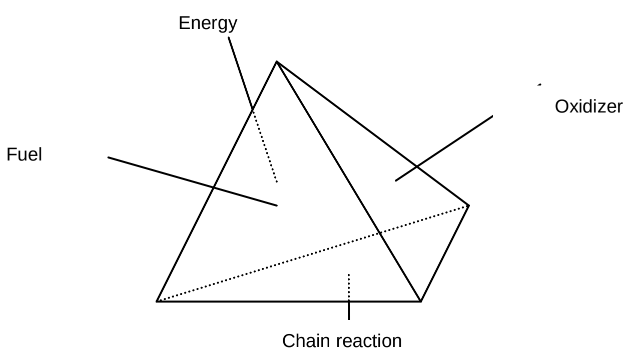
Figure 4: Fire behaviour triangle



Source: Own construction from “Basic Course on Wildfires [in Spanish]”[5]

Apart from the fire triangle, another concept to explain fire is the fire tetrahedron which, unlike the two previously explained triangles, shows the essential elements for a fire to propagate and persist[5]. The fire tetrahedron adds another component, the chemical chain reaction, to the three elements already present in the fire triangle (see Figure 5). Once a fire has started, the resulting exothermic chain reaction sustains the fire and allows it to continue until at least one of the elements of the fire is blocked. Foam can be used to starve the fire of the oxygen it needs. Water can be used to lower the temperature of the fuel to below ignition point or to remove or disperse the fuel. Halo methane can be used to remove free radicals and create a barrier of inert gas in a direct attack on the chemical reaction responsible for the fire

Figure 5: Fire tetrahedron



Source: Own construction from “Basic Course on Wildfires [in Spanish]”[5]

To better understand the types of chain reaction originating from the combination of the three elements, it is worth analyzing the propagation speeds, which are present[5] (see Table 1).

Table 1: Types of chain reaction according to their propagation speed

SPEED OF PROPAGATION	TYPE OF CHAIN REACTION
Very slow	Oxidation
Slow	Combustible
Quick	Conflagration
Brief	Explosion

Source: Own construction from “Basic Course on Wildfires [in Spanish]”[5]

As in the previous case, if any of these four elements is missing, the fire will be extinguished.

A wildfire is an uncontrolled fire that occurs in the countryside or a wilderness area with an area of combustible vegetation. A wildfire differs from other fires by 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. Moreover, wildfires differ from other fires because they occur in areas of grassland, woodlands, bush land, scrubland, peat land, and other wooded areas that act as a source of fuel, or combustible material. Buildings may be affected if a wildfire spreads to adjacent communities. While the causes of wildfires vary and the outcomes are always unique, all wildfires can be characterised in terms of their physical properties, their fuel type, and the effect that weather has on the fire.

Fuel accumulation, due to total fire control but especially because of the abandonment of the rural environment and the agro-forestry-pastoral activities, as well as the progressive alteration of the landscape, generate new more devastating classes of fires; ones which quickly destroy enormous extensions of terrain. Official Spanish statistics call such fires “big wildfires” (GIF from now on) and they are characterised by areas larger than 500 hectares being ablaze. However, GIFs are not strictly large surface fires but rather fires which spread quickly and cannot be suppressed. In other words, GIFs are those fires which cannot be extinguished[8]. Their proliferation is made possible because of the change in the behaviour of wildfires over the years. It is interesting to note that between 1986 and 1997 while GIFs accounted for a mere 0.6% of all registered fires in Catalonia, they represented 80% of the total area burned. In 2007 fires over 100ha represented 94.6 % of the burn surface. This development can be analysed by considering four generations of fires (see Figure 2).

The first generation of fires, which began at the end of the 50-60s, is characterised by having a surface area bursting with possible fuel which would cause large fires. Such amounts of fuel were available because rural areas were being or had been abandoned and were not being maintained. Therefore, the goal here was to increase accessibility to the area and use linear prevention infrastructures (firewalls)[8].

The second generation manifested itself as faster and more intense fires and was a consequence of 10-15 years of large amounts of fuel accumulating after cultivation, and also traditional forests management, ceased[9]. Second generation fires appeared in the 70-80s and it were dealt with by reducing access time to fire control systems (water points, roads, security, fast arrival, etc.) and by increasing the number of resources, particularly airborne, in order to reduce the intensity of these fires. At the same time, linear infrastructures were applied to break the line of continuity between forests and houses.

The third generation was in the 90s and it was characterised by high intensity fires due to crown fires, which burn materials at the canopy level. These fires were a result of 30-50 years of poor forest management and the suppression of all low and medium intensity fires, and were impossible to be extinguished in any way[8].

Finally, the fourth generation includes fires that spread over a new fuel: residential areas. These fires spread using the dense vegetation of gardens, as well as the fuel between forests, urban areas and housing[8].

GIFs are mainly deliberately lit and very difficult to control so a significant financial investment in fire extinction equipment may not be enough. Instead it is better to devote more resources and efforts on fire prevention rather than focusing on attacking the fires directly. Extensively analyzing a fire’s history in order to design a good social prevention plan, as well as detecting a fire’s pattern to better understand its behaviour and thus identify adequate fire extinguishing techniques and organization is paramount.

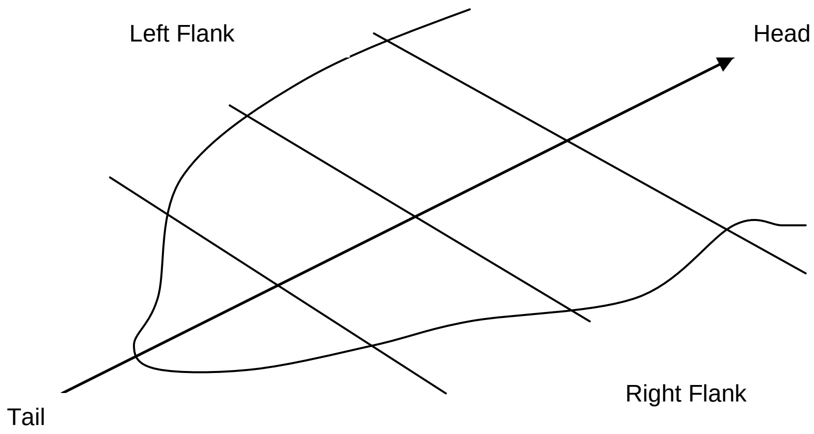
Given the scope of GIFs, fire fighting equipment must have a flexible and dynamic structure, and must include professional well-trained expert fire-fighters who are able to take charge the moment a fire starts. Being ahead of and being able to anticipate any changes in the fire is also vital in being able to predict where it will be possible to suppress the fire; as is knowing the fire’s intensity, where and when it will change its behaviour (critical points), and which of the fires could turn into a GIF (design fire), etc. Having teams of highly experienced fire fighters to manage large fires is vital, however, as building up such experienced teams is time-consuming and costly, it is often preferable to anticipate where fires might occur and devote more effort to fire prevention rather than focus on the direct attack of the wildfire because many times, the resources available are simply inadequate for fire-fighting[10].

Thus, it is important that any local action plan is based on a first-rate study about fire behaviour. Additionally, it is crucial to identify the critical points where one could apply the patterns of propagation analysed and then act according to the design created.

2.2 Wildfire components

A fire is principally composed of three parts: the head, the flanks and the tail. The head of a fire is the most rapidly spreading portion of a fire's perimeter, usually to the leeward or up slope; may have multiple heads if there are separated flanking fires. The flanks are the parts of a fire's spread perimeter that grow to the sides and then run roughly parallel to the main direction of spread. Separated flank heads are extremely dangerous in steep terrain. The tail is the opposite side from the head. This last part corresponds to the portion which burns slower[5] (see Figure 6).

Figure 6: Wildfire components



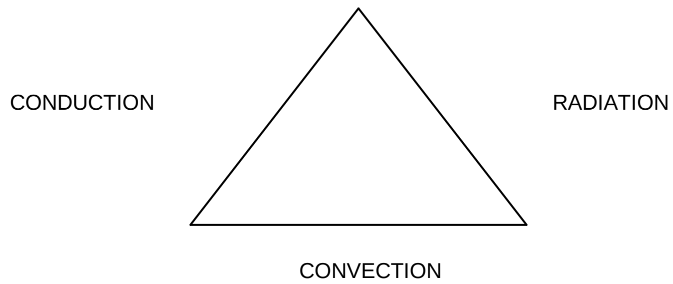
Source: Own construction from "Basic Course on Wildfires [in Spanish]"[5]

There are three phases of a fire: incipient (growth), free burning (fully developed) and smouldering (decay). First there is the initiation phase, the beginning of the fire, which occurs either by natural causes or by human action (negligence, intentional or accidents). Then, there is the spread, which is the extension of the fire to the nearby vegetation. Lastly, there is the extinction phase, the end of the fire, either by natural causes (rain or lack of vegetation) or by human action[5] (work of extinction). Each phase has its own unique characteristics and dangers to fire-fighters and should be understood thoroughly to ensure and improve safety during fire fighting operations.

The first phase includes the ignitability concept, this is to say, the ability of the fuel to start the ignition. A material burns when it reaches its ignition temperature (flashover). In particular, ignition is defined as the time (t) required for ignition divided by the energy intensity per unit area (kW/m^2) supplied[11].

The second phase depends on the weather conditions, the topography and the vegetation present. At any rate, the basic forms of fire spread may be categorised in a new triangle (see Figure 7):

Figure 7: Triangle of possible pathways



Source: Own construction from "Basic Course on Wildfires [in Spanish]"[5]

Radiation is that the heat transmitted by any material without requiring physical contact and is one of the most common causes for a fire spreading. Heat radiation occurs especially in urban areas where the proximity to other structures and the generation of a large amount of heat originates the ignition of neighbouring buildings[5].

Conduction is the heat transfer via direct contact between objects. In the case of forest fuels conduction is not decisive as these fuels are very poor thermal conductors.

Finally, convection is the most dangerous way of transmission as this is what causes major problems. Fire generates its own stream of overheated air that moves through the air surrounding us reaching temperatures high enough to ignite combustible materials on its way[5].

The last phase of a fire is the extinction which includes two possible ways to end the fire. One is a natural way, for example, the end of the fuel, while the other can considered as human action which will try to act on either side of the fire triangle (fuel, oxidizer and heat)[5].

In general, the variables that influence a wildfire behaviour are the weather (speed, direction, temperature and relative humidity), topography (slope, aspect, altitude and relief) and fuel (quantity, moisture-delay time, distribution and compaction)[5].

2.2.1 Fuel

Fuel is defined as a substance that, under certain conditions, is able to burn. All that is required is the presence of an oxidizer (oxygen, mostly) and the contribution of a certain activation energy.

The general composition of fuel is essentially carbon (C) and hydrogen (H_2), either in free form or combined in the form of hydrocarbons. It also contains sulphur, even if only in small percentages, due to the detrimental effects of the oxygen compounds. Another component is oxygen, which can be either fixed to carbon and hydrogen, or in a free state in the fuel. Finally, fuel contains inert elements, such as moisture, ash, CO_2 and nitrogen. Also fuel can be defined as any material that stores potential energy in a form that can be practicably released and used as heat energy[12].

Chemical fuels can be divided in two ways. First, by their physical properties, they can be considered as a solid, liquid or gas. Secondly, on the basis of their occurrence they can be considered as either primary (natural fuel) or secondary (artificial fuel).

Solid fuels are characterised by the ash they produce when they burn. Combustion can be by flame or incandescent and it depends mainly on the moisture content of the solid, the heat conductivity, the ignition temperature, the degree of combustion and the spread speed. Solid fuels include coal, wood, corn, wheat, rye, peat and other grains. Coal was the fuel source which powered the industrial revolution, from firing furnaces, to running steam engines. Wood was also extensively used to run steam locomotives. Both peat and coal are still used to generate electricity today[12].

Liquid fuels act differently from solid fuels because it is the fumes of liquid fuels, rather than the fluid, that are flammable. In this case, one must take into account the flash point which is the lowest temperatureat which the fumes can vaporize to form an ignitable mixture in the air. Measuring a flash point requires an ignition source. At the flash point, the vapour may cease to burn when the source of ignition is removed.Every liquidhas a vapour pressure, which is a functionof that liquid's temperature. As the temperature increases, the vapour pressure increases. As the vapour pressure increases, the concentration of vapour of the flammable liquid in the air increases. Hence, temperature determines the concentration of vapour of the flammable liquid in the air. Among liquid fuels there are natural gas and liquefied petroleum gas, non-petroleum fossil fuels, alcohols, biodiesel, ethanol, hydrogen, ammonia and petroleum such as gasoline, diesel or kerosene[13].

Fuel gas is the most used over the previous two (solid and liquid). Fuel gas is contrasted with liquid fuels and from solid fuels, though some fuel gases are liquefied for storage or transport. While their gaseous nature has advantages, avoiding the difficulty of transporting solid fuel and the dangers of spillage inherent in liquid fuels, it also has limitations. It is possible for a fuel gas to be undetected and collected in certain areas, leading to the risk of a gas explosion. This is the reason why odorizes are added to most fuel gases so that they may be detected by a distinct smell. The combustion of a fuel gas requires the presence of combustion air (pure oxygen combustion is not considered). When considering all possible mixtures characterised by the content of gas compared to the homogeneous mixture, for example from 0% (pure air) to 100% (pure gas), it is observed that combustion can only occur and propagate within a zone between these extremes. This area is known as the flammability zone. The lower limit is regarded as the value below which there is too much air in the mixture to make the combustion possible, and the upper limit the value above which there is insufficient combustion air to produce the combustion. This type of fuel is mainly composed of hydrogen, carbon monoxide (CO), saturated hydrocarbons (methane, ethane, propane, butane and isobutane, pentane and hexane vapour exceptionally), unsaturated hydrocarbons such as ethylene (C_2H_4), butane (C_4H_8), propane or propylene (C_3H_6) and unsaturated hydrocarbons unidentified in analysis (CnHm). Eventually also contain oxygen oxidizer and inert gases (CO_2, N_2) in small proportions[16]. Its main properties are density, important in respect to local ventilation, the calorific value and the ignition temperature, which represents the minimum value at which a point of a flammable mixture of fuel gas and oxidizer must be taken for combustion to begin and spread. The most common type of fuel gas in current use is natural gas[13].

The forest system's ability to maintain and extend fire defines its combustibility. Moreover, combustibility is defined as the speed at which the fuels are burned.

For each type of vegetation, its flammability and combustibility are determined, which vary depending on the type and quantity of biomass and its spatial distribution or stratification[13].

2.2.2 Flammability

Flammability represents how easily something will burn or ignite, causingfireor combustion.The degree of difficulty required to cause the combustion of a substance is quantified through fire testing. Internationally, a variety of test protocols exist to quantify flammability[14].

Materials present two physical properties that indicate their flammability: flash point and volatility, which is determined by the boiling point[14].

On one hand the material's flash point is the lowest temperature at which a liquid (or a volatile solid) can vaporize to form an ignitable mixture in the air. When there is an external source of ignition (for example, electric sparks, flames) a material can ignite at temperatures equal or above its flash point[5].

The flash points of some products are:

Gasoline	Ethyl	Benzene	Hexane	Diesel	Diesel oil
-43ºC	12ºC	20ºC	-28ºC	52ºC a 96ºC	150ºC

Flammable gases have no flash point as they already are in the vapour phase.

On the other hand, the volatility of a material indicates the ease with which a liquid or a solid turns into steam. Volatility is measured by the boiling point of the material (the temperature at which the vapour pressure of the material is equal to the atmospheric pressure).There are some materials which are not volatile but rather are flammable, such as water, chloroform and mercury.

In the case of a gas mixture, such as the gases present in a fire, there are a number of different molecules, each subjected to the action of heat. This heat, as a primary form of energy, transfers a movement to these molecules, which is added to their own movement. In this state, the lighter gas molecules move more quickly than the heavier ones, causing collisions between them which increase the internal energy of the gas, both for light molecules as well as heavy ones. As the heat increases, the molecules increase their motion and gradually multiply the number of collisions between them and therefore their energy level. As this process continues, it leads to a state in which the energy accumulated by the gas is greater than the energy which joins the molecules and these molecules may eventually be broken by the shock effect, i.e., they disintegrate. If there is enough oxygen in the surroundings the activated fuel will ignite[14].

The presence of oxygen in the fuel (oxidation) generates a reaction which, thanks to the energy (heat) provided by the mechanism described above, releases heat (exothermic). It can be said that the flammability of a gas is a mechanical consequence aided by an energy source, i.e., heat. However, there are other sources such as shock waves, or the combination of heat and shock waves[14].

It is important to keep in mind that the disintegration of molecules is not enough to start the ignition. A significant number of molecules together with oxygen in the air are needed. The mixtures of gaseous fuels and air will only burn if the fuel concentration lies within well-defined limits. These limits are determined experimentally.The flammability range is delineated by the upper and lower flammability limits. On one hand, the lower flammability limit(LFL), usually expressed in volume percentage, is the lower end of the concentration range over which a flammable mixture of gas or vapour in air can ignite at a given temperature and pressure. Outside this range of air/vapour mixtures, the mixture will not ignite (unless the temperature and pressure are increased).The LFL decreases with increasing temperature; thus, a mixture that is below its LFL at a given temperature may ignite if heated sufficiently. On the other hand, the upper explosive limit (UEL) is the highest concentration (percentage) of a gas or a vapour in air capable of producing a flash of fire in presence of an ignition source (arc, flame, heat). Concentrations higher than UFL or UEL are "too rich" to burn[14].

The flammability limits depend primarily on three factors: the temperature, the pressure and the concentration of the oxidizer.

Temperature is very important because it affects both the fuel and the oxidizer. Thus, if the temperature is increased it will have an influence on two factors. On one hand, it will affect the contribution of the heat energy to the fuel, whereby it will be close to the flash point and consequently insignificant amounts of this it may be flammable. On the other hand, it will reduce the cooling effect of the excess of air in the enclosure. Along these lines, higher temperature results in lower LFL and higher UFL, while greater pressure increases both values. On the other hand, oxygen enriched atmospheres lower the LFL and increase the UFL. An atmosphere devoid of an oxidizer is neither flammable nor explosive, regardless of the fuel gas concentration. Increasing the fraction of inert gases in an air mixture raises the LFL and decreases the UFL[14].

Some materials are pyrophoric, i.e., they can burn spontaneously without any external ignition source. For example, metallic sodium can react with atmospheric moisture. This reaction produces hydrogen gas and the heat generated by the reaction may be sufficient to ignite the hydrogen and oxygen.

2.3 Classes of wildfires

There are different criteria to separate and classify wildfires. Some of them are: by fuel type, where the fire spreads or what governs it[5].

According to the type of fuel, wildfires can be classified into 3 different groups. The first group corresponds to solid material fires, usually organic material fires, where combustion takes place by the creation of embers (wood, cloth, rubber or some plastics). The second refers to fires of liquids or liquefiable solids (gasoline, grease, etc.) and the last includes gas fires as butane or natural gas[12].

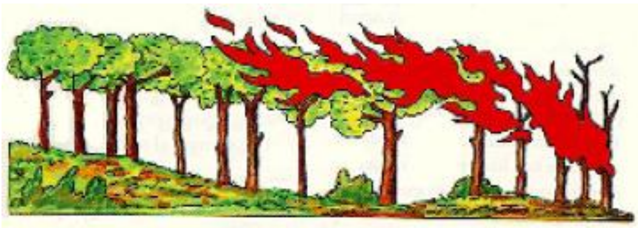
In the second criterion, there are three different classes of wildfires: surface wildfires, smoldering wildfires and crown wildfires. A surface wildfireis the most common type and burns along the floor of a forest, moving slowly and killing or damaging trees (see Figure 8a). Such fires can also start other fires because they can become crown fires[5].

Figure 8: Classes of wildfires

a. Surface fire

b.Smoldering fire

c. Crown fire



Source:
http://www.proteccioncivil.org/catalogo/carpeta02/carpeta24/vademecum12/vdm010.htm

Asmoldering fireis usually started by lightning and burns on or below the forest floor in the dark earth made of organic material such as decayed leaves and plants (see Figure 8b). Such fires are less common and are characterised by burning with little or no flame due to the little oxygen available. For this reason its propagation is very slow compared with other types of wildfires. However these types of wildfires can be more destructive as they are able to eliminate the underground systems of vegetation[5].

Finally, there is the crown fire which usually represents the greatest threat to the fire fighting system as it generates high intensities, massive generation of secondary outbreaks, high flame length and propagation speeds which are double those produced by surface fires. It spreads rapidly by wind and moves quickly by jumping along the tops of trees. The ignition of a crown fire is dependent on the density of the suspended material, canopy height, canopy continuity, and sufficient surface and ladder fires in order to reach the tree crowns (see Figure 8c).These types of wildfires usually begin as surface fires[5].

Crown fires can be classified into three different categories: torching, passive or active. Torching is the movement of a surface fire up into tree crowns, the precursor to an active crown fire. Passive crown fires involve the torching of individual trees or groups of trees. Crown fires become active when enough heat is released to preheat and combust fuel above the surface, followed by active spreading of fires from one tree crown to the next though the canopy. Crown fires are usually intense and are strongly influenced by wind, topography, and tree (crown) density[5].

Finally, considering the last criterion, what governs the flames, fires can be classified as[15]:

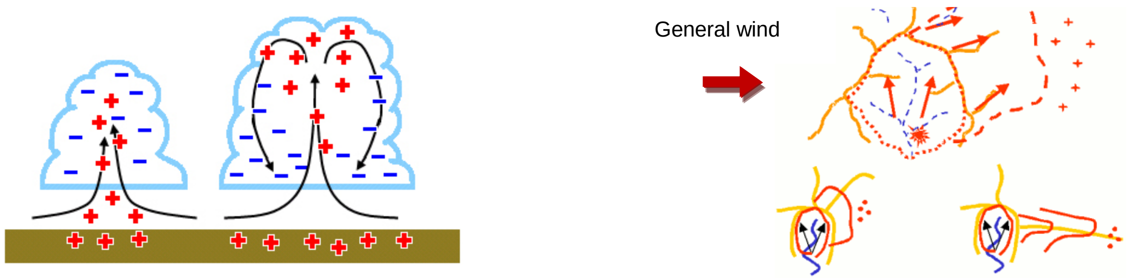
1) Convection fires or fuel fires: the large accumulation of forest fuel is responsible of the developed intensity (see Figure 9).
2) Topographic fires: topography of the terrain causes fire to develop in complex orography being influenced by the slope, sun exposure (daytime) and roughness. The driving force is the convective wind produced by the heating of the surface and its interaction with the relief. These fires usually follow the valleys and ravines (see Figure 10).
3) Wind fires: the weather plays a very important role here. The direction from where the wind comes, its intensity and velocity, provides oxygen and dries fuel in general and more importantly, it quickly dries the ‘death fuel’. These fires tend to spread linearly in the wind direction and adapt more or less, to the morphology of the ground (see Figure 11).
4) Hungry fires: they are called big wildfires and are characterised by creating their own weather conditions (temperature, relative humidity and wind speed) that make only an indirect attack feasible.

The first type is subdivided according to whether the fuel is underground, on the surface or in the air resulting, in each case, in different intensities and fire propagation velocities. These fires are characterised by spreading by convection and not by radiation, developing extreme behaviours and advancing thanks to a massive generation of secondary outbreaks. These fires are detected by analyzing the fuel that is burning and the way it spreads. Their extinction is usually achieved if the fuel can be moved on to a less favourable place to burn or by changing its structure[5].

Figure 9: Convection fires

Standard: the accumulation and availability of fuel generate enough intensity to create a fire

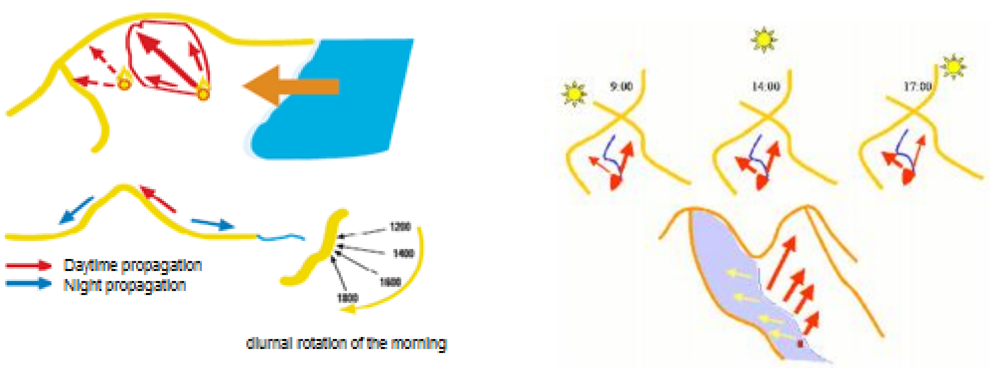
Convection with wind: convection dominates the fire and secondary centres follow the general wind axis.



Source: Integrating risk of big wildfires (GIF) in forest management[16]

Topographic fires are quite devastating and are characterised by having the same behaviour in both the head and the flanks. The extinction of this type of fire has to take into account orientation (sun exposure), roughness and, above all, slope[5]. In general they are characterised by having a high diurnal intensity and a low night intensity[16].

Figure 10: Topographic fires



Source: Integrating risk of big wildfires (GIF) in forest management[16].

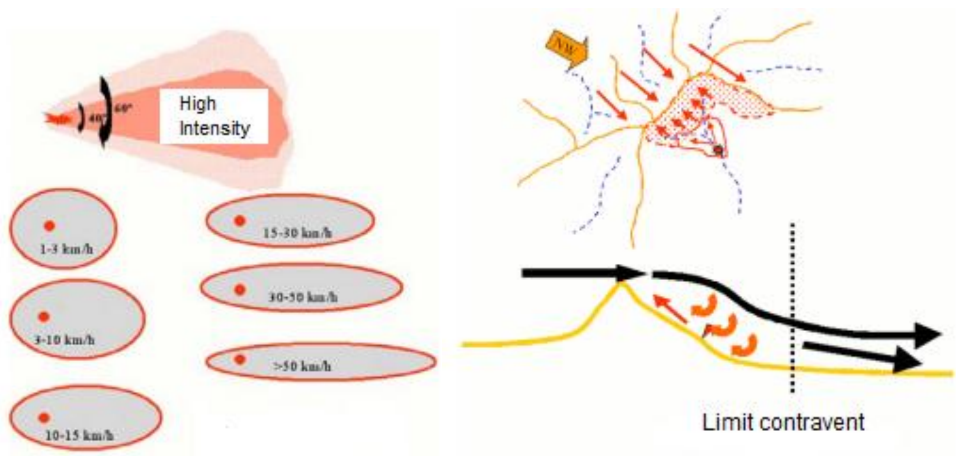
Figure 10 shows that this type of fire changes direction by following the sunny slopes (thin arrows point out a lower spread intensity)[16].

Fires driven by wind, as its name suggests, are those whose strength and speed are determined by the wind which brings oxygen and dries the fuel in the areas that is susceptible to burning. These fires are detected by observing the status of the plume and the presence of strong winds on the surface. The characteristics of convective columns (colour, size or slope of the column) provide a lot of information about the type of fire that is being generated. A white column of smoke will show a low-intensity fire while a grey-black colour will indicate a high intensity fire. On the other hand, a vertical column suggest a topographic fire with atmospheric instability, one lying prone will warn wind and a parting plume, produced by a topographic fire with upper wind, will represent a column which generates secondary fires [5].

The extinction of these types of fires is based on waiting for the fire in areas without any wind so that they can treated as if they were a topographic or a fuel fire.

Figure 11: Wind fires

In the plains With relief



Source: Integrating risk of big wildfires (GIF) in forest management[16].

Finally, as we have already mentioned, hungry fires are a particular case of GIF's and the major factors influencing their occurrence are the weather conditions (drought), large amount of vegetation (fuel) and above all, the determining factor for their spread is the extreme weather conditions (low ambient relative humidity and high wind speed).

For an active fire to become a GIF certain weather conditions, that are described from the synoptic situations which generate them, are necessary. However, assigning a synoptic situation only makes sense in the case of wind fires and convection fires. With topographic fires, there is not a clear synoptic situation and therefore other variables are analysed[16]. These variables are shown in Table 2.

Table 2: Meteorological conditions and weather station data for the topographic analysis of fires[16].

Meteorological conditions	Weather station data
Presence of meteorological or geographical elements that can modify local weather conditions and adjust fire behaviour.	Type of wind, whether general, topographic, topographic from a valley, marine, offshore or erratic, or sudden changes in wind speed and direction in the day and night.

So, the assignment of a specific synoptic situation to wind fires and convection fires is carried out by consulting the historical daily synoptic available online at www.wetter3.de. These maps are the following[16]:

- Geopotential Height Map at 500hPa and surface pressurefrom 01/01/1948.
- Temperature mapat 850 hPafrom01/01/1948.
- Air pressure map with fronts[23]from01/27/1998.

In Catalonia, the synoptic situations which can generate GIF are identified in Table 3. The other synoptic situations that may occur are not considered here because they are deemed as not capable of generating GIF[16].

Table 3: Synoptic situations that can generate GIFs in Catalonia, grouped by common features.

Entries of south and west	**Wind fromnortheast to northwest**	**Instability and storm with front step**
General situation of south	Synoptic situation of wind from north	Synoptic situation of instability with front subsequent step
General situation of southwith out west	Synoptic situation of wind from northeast	
Synoptic situation of wind from west	Synoptic situation of wind from northwest	

The south synoptic situation or south entries annotated in Table 3 reference the input of Saharan air mass. In addition, the other synoptic situations present in this Table explain most of the burned area in Catalonia during the last decades. However, other specific synoptic situations cannot be ruled out as GIF generators[16].

Then, the study of fires allows us to observe that in the same topography and meteorological conditions, fire spreads along similar propagation schemes,which mainly depends on the water stress accumulated and the amount of fuel and its structure,and changes its intensity according to fuel availability.

In this sense, the same type of fire does not involve the same fire behaviour. Differences in the structure of the fuel, the land use or the ignition points determine fire behaviour, although propagation scheme remains constant. Therefore, for the same type of fire, the points where it changes its behaviour with respect to the orography and the opportunities for extinction are similar[16].

2.4 The concept of wildfire risk

The term "risk" applied to wildfires, includes many definitions and interpretations and meanings can vary. In this sense, the risk can be defined exclusively as the probability of ignition[17], while the danger, according to other authors, is an abstract concept defined by the social perception and the evaluation of the factors which are considered harmful[18]. Moreover, the English term 'hazard' refers to the vulnerability of a forest to suffer a fire when considering only fuel[16].

In the field of engineering, wildfire risk is defined as the probability of occurrence in a specific space and period of time, and the potential damage of the fire in that area[18]. From this definition one creates a wildfire risk model that incorporates the likelihood of occurrence and the impact or the potential damage of the wildfire. Thus, the factors which influence the probability of occurrence are the cause of ignition (human / natural) and the pre-fire conditions (type of fuel and moisture content). On the other hand, the factors which influence the impact or potential damage of the fire are the impact probability, whose expression is related to a gradual scale obtained from the difficulty of extinction, fire behaviour (type of fuel, moisture content, wind, terrain) and the element impact, expressed as the value of this and the fire behavior[16,18].

In Catalonia, the current tools that characterise the risk of wildfires are the daily risk map[19] and the basic hazard map[20]. The daily risk map of wildfires is calculated every day according to the integration of a series of information into a single map: fine fuel moisture, maps of variables and risk indexes of meteorological components, percentile calculations of the basic variables, historical information about wildfires and situations that have occurred in the recent years, graphics of specific risk monitoring sensors, tracking synoptic weather patterns related to the risk of fire, and static maps (forest fuels, flammability, altimetry). The daily risk map computes every day the risk according to momentary situations, and is then used for the activation of the alpha plan, the movement of brigades, the coordination with fire-fighters and civil protection entities, the warning of local authorities, fire authorizations, danger warnings to the population, etc[16].

The basic hazard map defines the probability of static occurrences of wildfires based on vegetation, historical, orographic and climatic factors. This map integrates the concept "ignition hazard", i.e., the ease with which a forest fire can start and the risk of spreading or ease with which it can spread. This map is quantitative, i.e., each point of the territory has a numeric risk value ranging from zero to ten assigned to it, of which is the result of the combination of different factors that determine the risk of fire. The basic fire hazard map is a map created by management and planning to help establish territorial priorities in preventive actions; rationalize and optimize the performance of management and define areas of planning and intervention. Historical factors are frequency of ignitions (number of fires in a period of time divided by this period) and the frequency of ignitions' consequences, in which the weight of each is weighted according to the affected area. Vegetation is included in a fire hazard because of its flammability, and its combustibility, or ability to ignite, which affects fire behaviour. Orographic factors taken into consideration in the risk of fire are the slope and the isolation. One of the factors considered most important in determining the risk of wildfires is the weather. Wind and water deficit together with adverse situations (extreme conditions which are low in frequency or duration but have great impact on the occurrence of fires) are included in this map. The wind is analysed on the climate field, not the episodic. The basic hazard map analyses the ease with which a wildfire starts and spreads[16].

The standard fire risk map provides information on the most vulnerable areas, from the point of view of the odds of having a certain type of GIF, and it is used as a basis to identify key areas where it is more important to establish wildfire prevention as the preferred management tool (see Figure 12).

The Standard fire risk map applied in Catalonia allows us, by providing information about the most likely type of fire as well as its main characteristics in terms of its behaviour, pattern of spread and control possibilities, to identify those areas with a higher risk of having a standard wildfire. This is a map that responds to current land and forest landscapes and that may be adapted in the future, if there are significant changes in the landscape and the structural configuration of the forests in Catalonia[16].

Figure 12. Development of the Standard fire risk map

STANDARD FIRE RISK MAP

⇧

Standard fire risk calculation based ona number of factors

⇧

Factor 1: Standard fire (difficulty of extinguishing and complicating fire behaviour)

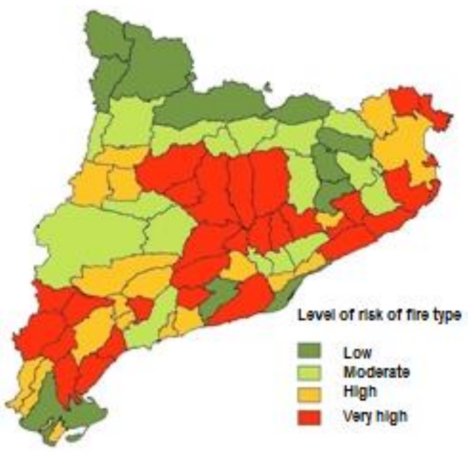
Factor 2: Frequency of fires.

Factor 3: Land, landscape, climate and vegetation features.

Source: Integrating risk of big wildfires (GIF) in forest management[16].

Thus, especially in recent decades, convection fires, characterised by their speed of spread and high front intensities and with a potential determined by the continuity of fuel and the duration of the synopsis episode, are concentrated in areas of high continuity of forests and forest structures which have accumulated vertical fuel continuity. In these cases, forest management can focus on landscapes that are more resistant to wildfires and change the level of risk of fire type in areas currently affected by convection fires (see Figure 13)[16,21].

Figure 13: Standard fire risk map of Catalonia



Source: Departament d'Interior. Generalitat de Catalunya i Centre Tecnològic Forestal de Catalunya[21].

Then, some of the challenges that must be addressed jointly in the fields of prevention and suppression of forest fires, are in anticipating big wildfires, reducing the spreading ability of latent GIFs, as well as reducing the damage that fires can cause to people, property and landscape uses[22]. Therefore, standard risk maps are a useful tool in determining those areas of very high and high risk, which will be denoted as the Catalan areas with the highest fire prevention priority and where forest management models that provide a greater degree of prevention against a GIF should be used, while areas of minor fire risk levels would not be prioritised. On the other hand, there are some areas in Catalonia where it is difficult to determine a standard risk of fire (these are the zones that lie between areas of high and moderate risk) because while their fire history may not be very abundant, they have been identified as areas with a possible risk of powerful convection fires.

3. Wildfires in Catalonia

Since the early twentieth century until today, in European countries and particularly in Catalonia, there has been remarkable social change thanks to rural to urban migration. In particular, after the land confiscation by the State of the eighteenth century, the Mediterranean landscape underwent an important demographic change which in the twentieth century led to intenseindustrialization accompanied by the abandonment of farming and the movement of the population from rural areas to towns[23]. The process of giving up farming practises because of the loss of job profitability, coupled with an aging rural population, has accelerated the difficulties of the environmental change in recent years. This change is of great consequence in the analysis of fires since forest fire behaviour is related to the state of forests and rural areas in general.

The first visible consequence has been the change of a rural mountain lifestyle (pasture, cultivated land, forestry exploitation and hamlets) to that of a lifestyle of entertainment in the mountains with the advent of food and beverage outlets, hiking trails, adventure tourism, second homes etc., but with the added distinction of very few houses. Unfortunately, in recent decades, Catalan landscapes have experienced a progressive social change influenced mainly by changes in the economic structure of the region and the movement of society from rural to industrial areas[23].

On one hand, changes in the economic structure have led to very poor forest management which does not invest in the mass quality improvement and forgets the long term objectives of persistence: ignoring the fire, the return of nutrients and the erosion as part of the system[23]. On the other hand, urban areas have increased without order or control; there has been the impact of certain infrastructures, the abandonment of farming, forestry and ranching, and the degradation of some urban areas or the saturation of other places. All this has directly influenced the state of environmental, cultural and historical values of these landscapes and has increased geological risks, along with other environmental hazards.

In particular, the depopulation of rural areas and the consequent abandonment of agricultural activity is an immediate cause for the generation of wildfires as it has facilitated the rapid increase of scrub and woodland causing the modification of the territory in irregular structures with a dense undergrowth. All this means that the evolution of the Catalan region represents an increased risk of wildfires as the historically open spaces are transformed into highly flammable ones and thus have become far more vulnerable to possible wildfires. Other factors such as the increase of second homes in forest areas, the proliferation of roads and power grids and the increase in recreation use have made Catalonia increasingly susceptible to forest fires, because the combination of all these factors simply serves to aid wildfire ignition. These changes in land use have had not only an impact on the natural vegetation, but also on the risk of wildfires and the loss of cultural, biological and landscape diversity[23].

All of these changes can be reflected in numbers. In Catalonia in the early twentieth century, 10% of the area was forested surfaces, whereas now they represent about 61%, according to the Ecological and Forest Inventory of Catalonia (CBEFIs, 1991)[23]. With only these data the figure representing the increased amount of fuel is justified. Therefore, it would be in the countryside's best interest to stop the process of depopulation and abandonment and attempt to secure the population in the territory by maintaining traditional rural activities (agriculture, ranching, forestry, etc.) because this would help to restore and maintain the rich cultural heritage, the quality of life and the environmental sustainability throughout Catalonia.

It can be seen that over the years the population of Catalonia has experienced a very significant increase[24] (see Table 4). Specifically, the percentage of the population trends of the mid-twentieth century, with respect to the beginning of the twenty-first century, represents an increase of 56'87%.

Table 4: Evolution of the population of Catalonia

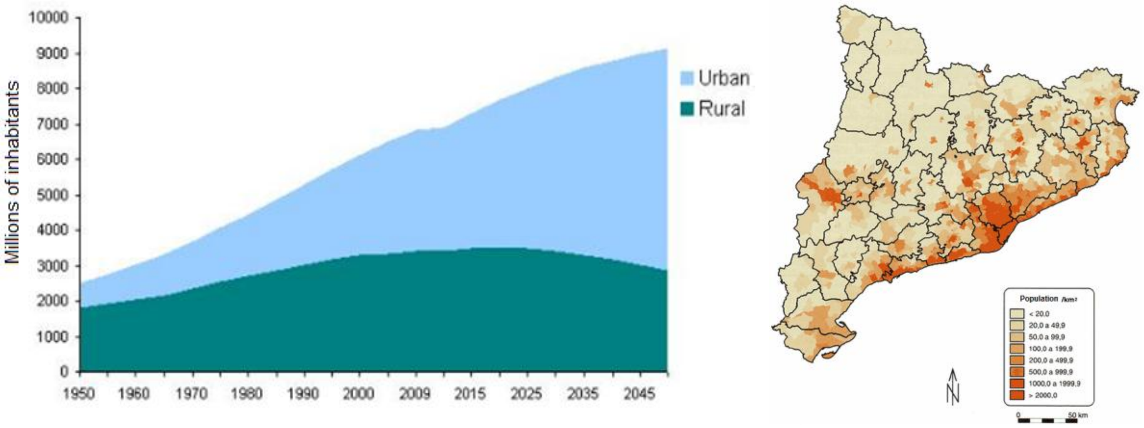
POPULATION	1950	2000	2004	2005
Catalonia	3.240.313	6.261.999	6.813.319	6.995.206

POPULATION	2006	2007	2008	2009	2010
Catalonia	7.134.697	7.210.508	7.364.078	7.475.420	7.512.381

Source: Own elaboration from IDESCAT. Padró Continu (2011).

Graphically we can see a global prediction for the evolution of the rural and urban population from 1950 to 2050 and a map with the distribution of the population in Catalonia (see Figure 14).

Figure 14: Prediction of the evolution of the rural and urban population



Source:

Left: Own construction from United Nations Population Division

Right: http://blocs.xtec.cat/legosocials/2009/01/29/la-poblacio-a-catalunya/

A further significant factor related to wildfire evolution, is the effort in recent years (2007-2008) to have total extinction. The choice of these types of models assumes a negative selection of fires as authorities choose to fight those with minor or moderate intensity while letting fires with a more extreme behaviour burn[8]. Those fires which burn with a low intensity burn off quickly and burn very small surfaces, while those of high intensities devastate large areas, escaping the control of the systems set up to extinguish them.

Although there are studies showing that fire has some positive effects for biodiversity, the fact is that fires threaten not only human lives and settlements but they also facilitate soil erosion.

In Catalonia climatological trends interact with the landscape dynamics. Fire risk is interconnected with the Mediterranean climate and its distinct seasons. In general, we can speak about a summer period with high temperatures and low relative humidity combined with episodes of hot and dry winds which are typical of these regions. Such factors create the perfect setting for the occurrence of a large fire[22].

It would also seem that the climate is shifting towards more intense extreme conditions and an increase in the number of summer days with high temperatures and low humidity, together with a decrease in rainfall which seems to become more episodic and intense favouring soil erosion devoid of any vegetation. This will increase the frequency of fires and their consequences[17,22,25].

Catalonia's socioeconomic change is visible in much of the Catalan territory and has caused significant changes in the behaviour of fires. The abandonment of rural areas has led to an evolution of a kind of symbiosis between the forest and urban spaces which requires a much more complex management of the risk of fires: special training for the fire-fighting services in order to work in these areas[26], campaigns to increase the resident population's awareness[27], preventative actions to reduce the biomass fuel[28] and a more careful distribution of residential developments in forested areas with high fire risk. Proper characterisation of these areas against fire behaviour is a first step towards an effective fire management, both from the point of view of extinction as primarily prevention.

All of these changes have forced society to adapt to the new situation by taking new measures of prevention and prohibition. It is important to understand, and above all to learn how to anticipate the fire behaviour, in order to identify the strength and power of each fire and to be able to anticipate and improve the ways to extinguish them. Work has to be done before, during and after the fire, so adopting a change in the intervention policy both from the perspective of emergency management and the social perception of the effects and uses of fire is essential. A dynamic and flexible structure of the fire-fighting services is required, based on the anticipation of fire behaviour, to the dynamic decisions taken at the fire line and the integrated management of fire as an emergency[23].

It has been found that work done prior to any fire is much more effective than immediate action once the fire has started to burn because late action usually leads to the fire exceeding extinguishing capabilities and thus making the situation difficult to control and consequently the extinction of the fire[5].

4. References

1. IDESCAT. *Basic Statistics of Catalonia. Demography and Quality of Life* [in Catalan]- Institut d'Estadística de Catalunya (IDESCAT) 2011 [Available at http://www.idescat.cat/dequavi/Dequavi?TC=444&V0=1&V1=1, accessed on October 23, 2011].

2. Equip Editorial Cruïlla. *Coneixement del medi (projecte 3.16). Medi social I cultural. 5 Primària*. Ed. Cruïlla, ISBN: 9788466122863

3. Boada M. Catalunya, land of forests without knowing [in Catalan]. *Món Sostenible* July 2012 [Available at: http://www.monsostenible.net/catala/opinions/catalunya-pais-de-boscos-sense-saber-ho/, accessed on November 23, 2010].

4. Castellnou M, Arilla E, López M. Fire management in Catalonia: first steps. *International conference on prevention strategies of fires in southern Europe* Barcelona 2005; 1-9.

5. Blanco J, García D, Castellnou M, Molina D, Grillo F, Pous E. *Basic Course on Wildfires* [in Spanish]. Universitat de Lleida, Cabildo de Gran Canaria, Generalitat de Catalunya, Spain, 2008 [Available at: http://www.slideshare.net/gtfsaltominho/curso-basico-de-incendios-forestales, accessed on May 13, 2011].

6. Carey H, Schumann M. *Modifying wildfire behaviour: The effectiveness of fuel treatments: the status of our knowledge*. Santa Fe, Southwest Region: National Community Forestry Centre, 2003.

7. Graham RT, McCaffrey S, Jain TB. *Science basis for changing forest structure to modify wildfire behavior and severity*. US Department of Agriculture Forest Service, 2004 [Available at: http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1161&context=barkbeetles,accessed on February 7, 2011].

8. Castellnou M, Nebot E, Miralles M. The role of fire in landscape management [in Spanish].*IV International Wildfire Fire Conference* Sevilla 2007 [Available at: http://www.fire.uni-freiburg.de/sevilla-2007/contributions/doc/SESIONES_TEMATICAS/ST1/Castellnou_et_al__2_SPAIN_Cat.pdf, last accessed August 1, 2013].

9. Vélez R. *Thefire management. Foundationsand experiences* [in Spanish]. Madrid: M^{c}Graw Hill. Interamericana de España S.L., 2000.

10. Castellou M, Miralles M. The changing face of wildfires. *Crisis Response Journal* 2009; 5(4):56-57.

11. Anderson HE. Forest fuel ignitability *Fire Tech* 1970; 6:312-319.

12. Boulandier J, Esparza F, Garayoa J, Orta C, Anitua P. Fire fighting Manual [in Spanish]. Pamplona: Gobierno de Navarra, 2001 [Available at: http://www.bomberosdenavarra.com/documentos/ficheros documentos/indices.pdf, last accessed August 1, 2013].

13. Types of fuels and their characteristics (Unit 3) [Available at: http://www.ignou.ac.in/upload/unit-3.pdf, accessed on January 20, 2011].

14. Basset JM. Flashover: development and control [in Spanish] [Available at: http://www.kume.cl/KFiles/File/Fuego_y_flashover.pdf, accessed on January23, 2011].

15. Rothermel RC. *A mathematical model for predicting fire spread in wildland fuels*. Research paper INT-115. Ogden, UT: U.D. Department of Agriculture, Intermountain Forest and Range Experiment Station, 1972.

16. Piqué M, Castellnou M, Valor T, Pagés J, Larrañaga A, Miralles M, Cervera T. *Integrationof the riskof large firesForestals(GIF) in forest management: Firetype andvulnerability of structuresheadersforestfires*. Series:Sustainable Forest ManagementGuidelinesfor Catalonia (ORGEST). [in Catalan] Centre de la Propietat Forestal. Departament d'Agricultura, Ramaderia, Pesca, Alimentació i Medi Natural. Generalitat de Catalunya 2011.

17. Hardy CC. Wildland fire hazard and risk: Problems, definitions, and context. *Forest Ecology and Management* 2005; 211 (1-2):73-82.

18. Bachmann A, Allgower B. The need for a consistent wildfire risk terminology. Neuenschwander L, Ryan K, Golberg G. (Eds*.). Crossing the Millenium:Integrating SpatialTechnologies and Ecological Principles for a New Age in Fire Management*, The University of Idaho and the International Association of Wildland Fire, Moscow, ID 2000; 67-77.

19. DMAH, 2005b. *Fire Risk Map On-line* [in Catalan]. Departament de Medi Ambient i Habitatge, Generalitat de Catalunya [Available at http://www.gencat.cat/medinatural/incendis/mapaperill/, last accessed August 1, 2013].

20. DMAH, 2002. *Map Basic Wildfire Danger* [in Catalan]. Departament de Medi Ambient Habitatge, Generalitat de Catalunya.[Available at http://www.gencat.cat/medinatural/incendis/mapaperill/, last accessed August 1, 2013].

21. Castellnou M, Pagés J, Larrañaga A, Piqué M. *Standard fire risk map of Catalonia* [in Catalan]. GRAF-Bombers. Departament d'Interior. Generalitat de Catalunya i Centre Tecnològic Forestal de Catalunya, 2010.

22. Rifà A, Castellnou M. *The Catalan model of forest fires extinction* [in Spanish]. IV International Wildfire Fire Conference, Sevilla (Spain), 2007.

23. Varga D. *Landscape and Agricultural Abandonment in the Mediterranean Mountains. An approximation to the Hortmoier and Sant Aniol Valleys ('Alta Garrotxa') from the Landscape Ecology.* PhD dissertation. University of Girona, Spain: 2007.

24. The population in Catalonia, 2011 [in Catalan]: [Available at: http://blocs.xtec.cat/legosocials/2009/01/29/la-poblacio-a-catalunya/, accessed on March 15, 2012).

25. Beringer J, Hutley LB, Tapper NJ, Coutts A, Kerley A, O'Grady AP. Fire impacts on surface heat, moisture and carbon fluxes from a tropical savanna in northern Australia. *International Journal of Wildland Fire* 2003; 12(4):333–340.

26. Davis IB. The wildland/urban interface. Paradise or battleground? *Journal of Forestry* 1990: 26-31.

27. Beringer J. 2000. Community fire safety at the urban/rural interface: The bushfire risk. *Fire Safety Journal* 2000; 35:1-23.

28. Winter G, Fried JS. Homeowner perspectives on fire hazard, responsibility, and management strategies at the wildland-urban interface. *Society and Natural Resources* 2000; 13(1):33-49.

Chapter 2.Methodology

1. Introduction

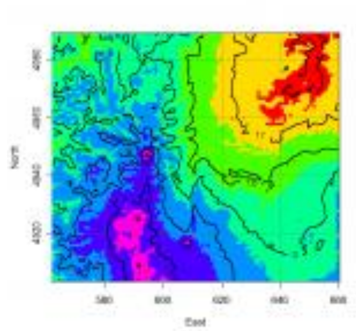
1.1 Spatial statistics

Spatial statistics is a general discipline that includes a set of appropriate methodologies for data analysis which corresponds to the measurement of random variables in different locations (points in space or spatial clusters) of a given region. In other words, spatial statistics analyse the elements of a stochastic process $\{Z(s): s \in D\}$, where $s \in \mathbb{R}^d$ represents the location in a Euclidean space of dimension d, $Z(s)$ is a random variable in the location s and s varies on a set of indexes $D \subset \mathbb{R}^d$. The methodology used differs depending on the features of the set D and allows spatial data to be classified into three large groups: geostatistical data, lattice data and point processes[1].

Geostatistics is concerned with spatial data, that is, each data value is associated with a location in space and there is at least an implicit connection between the location and the data value. "Location" has at least two meanings; one is simply a point in space (which only exists in an abstract mathematical sense) and the other is an area or volume in space. Therefore, geostatistical data are measurements taken at fixed points but defined anywhere in the space so their locations spatially define a continuous surface. The idea is to extend the spatial distribution of the values taken at fixed sampling points of a particular attribute to the entire study region.

From a mathematical point of view, for this type of data, D is a continuous fixed subset of $\mathbb{R}^d$ while $Z(s)$ is a random vector in the location $s \in D$[1].

Figure 1: Representation of spatially continuous data

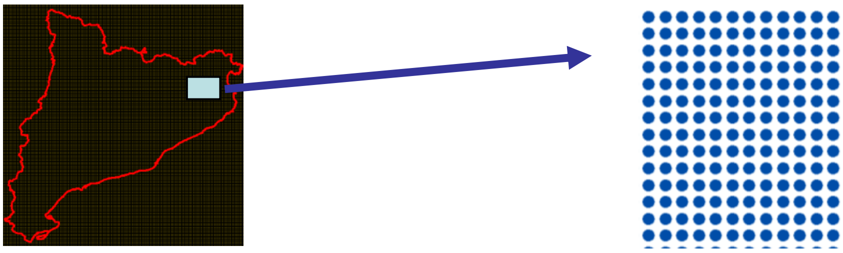


Source: CartoEduca.cl Geography, TICs and Education (digital library)

Lattice data are observations from a random process (observed over a countable collection of spatial regions which may be regularly or irregularly distributed), supplemented with information on neighbouring regions. Because this type of data is defined in spatial regions, the explicit locations specified by the vector s usually refer to the centroid of the region and do not form a surface but rather a set of connected nodes.

From a mathematical point of view, D is considered as a discrete fixed subset of $\mathbb{R}^d$, and $Z(s)$ a random vector in the location $s \in D$[1].

Figure 2: Lattice data representation



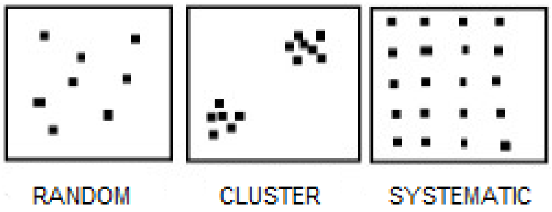
Source: Own construction

Point processes are characterised because their locations are the variables of interest. One considers a finite number of observed locations in a specific region and observes whether the distribution of individuals within the region is random, aggregated or uniform, i.e., if the intensity of the events varies over the region of study. Moreover, its goal is to look for models that explain or help to understand the phenomenon[1]. If one observes a variable of interest or a mark in each location, then it is said that the events have associated measures, or marks, and is called a "marked point process" or "with marks"[2].

On the other hand, spatial covariates provide additional relevant information that is needed to create a more comprehensive framework for the analysis of the study.

From a mathematical point of view, in spatial processes, the observations belong to a random subset $D \subset \mathbb{R}^d$ which can be discrete or continuous.

Figure 3: Outline of point processes



Source: Own construction

Hereafter, this Thesis will focus only on point processes with marks.In particular, within the field of fires, variables analysed only at the site of the fire are called marks. In this work marks include the year the fire occurred and also the causes of ignition. In particular, we consider: (i) natural causes, (ii) negligence and accidents, (iii) intentional fires or arson and (iv) unknown causes and rekindled fires. Spatial covariates are also considered, specifically, eight continuous covariates (topographic variables: slope, aspect, hill shade and altitude; proximity to anthropic areas-roads, urban areas and railways; and meteorological variables-maximum and minimum temperatures) along with one categorical variable (land use).

1.2 Point processes

Point processes are a type of discrete stochastic processes whose importance, is mainly due to their ability to model a wide variety of phenomena in physics, biology, economics and engineering. They correspond to the mathematical abstraction that arises when considering such phenomena as if they were a population randomly located in a space of parameters or as a random sequence of events in time[3].

A point process can be specified by its joint distributions of the number of points in arbitrary sets or by its joint distributions of the time intervals between successive points, starting at an appropriate origin. However, it is better to give the formal definition of a point process in terms of counting properties[3].

A point process can easily be defined as a stochastic model that generates a finite number of events $\{x_i,\ i = 1, \ldots, n\}$, which represent spatial locations in a set X. Since this work focuses on spatial point processes, X will be a bounded region of $\mathbb{R}^d$or a torus (“donut”) or, more generally,

X can be a locally compact Hausdorff space with a second countable topology[4]. Mathematically, it is considered appropriate to define a point process by a measure Λin X.

Let Λ be the space of all non-negative measures $\Lambda(.)$, with values in the integers, defined in the σ-algebra $\mathrm{B}(\mathbb{R})$ of the Borel sets on the real numbers $\mathbb{R}$, such as $\Lambda(A) < \infty$ for all boundedBorel sets $A \in B$. Let ζ be the σ-algebra generated by subsets such as, $\{\Lambda: \Lambda(A) \leq k\}$, $k \in \{0, 1, \ldots\}$and $A \in B$[3].

According to this nomenclature, a point process can be defined as a measurable function of a probability space (Ω, $\mathfrak{I}$, P) in (Λ, ζ). In particular, any probability measure defined on (Λ, ζ) produces a point process. The main characteristics are[3]:

1. $\Lambda(t_1, t_2)$represents the number of points which occurred at the interval (t_1, t_2). The stochastic process $\{\Lambda(t, u); u \geq t\}$ identifies the point process as a counting process.

2. $L_n(t)$is the required time, i.e., the length of time required for the n-th point after time t to occur. When n varies, $L_n(t)$ identifies the process in terms of the length of intervals between successive events.

3. $L - n(t)$is the required time for the nth point previous to t to occur.

1.2.1 Point process properties

Let $N(A)$ be the random variables which represent the number of events in a region $A \subset \mathbb{R}^2$

$$N(A) = \#(x_i \in A)$$

Some of the properties that can verify spatial point processes are:

a) The process is stationary if for any integer k and region $A_i, i = 1, \ldots, k$,the joint distribution of $N(A_1)$, $N(A_2), \ldots, N(A_k)$is invariant bytranslationsof A_i, for any set x.

This means that, for a time point process, the process depends only on the separation between the different moments considered but not on the shifts in time.

The concept of stationality is very useful in modelling time series[5]. In this case, the interpretation is straightforward because there is only one direction of variation (time). In the space field there are multiple directions and therefore one has to assume that in all of them, the phenomenon is stationary.

b) The process is isotropic[2] if the joint distribution is invariant by rotation of the union of A_i, where $i = 1, \dots k,\ \forall\, k \in Z$. In addition, a process is isotropic if the correlation between the data does not depend on the direction in which it is calculated. Mathematically, the isotropy is studied by calculating autocovariance functions or sample semivariance such that if they have significantly different forms the assumption of isotropy can be rejected.

The functions that model the dependence on isotropic processes are easier to interpret.

c) The process is orderlywhen there are no coincident events[2], that is to say:

$$\lim_{|dx| \to 0} \frac{P(N(dx) > 1)}{|dx|} = 0$$

Likewise we can say a process is orderly[2] if:

$E[N(dx)] \sim P[N(dx) = 1]$, because their ratio tends to 1 when $|dx| \to 0$.

d) It can be verified that a process is second order orderly when, for any pair of events[2] x and y:

$$\lim_{\substack{|dx| \to 0 \\ |dy| \to 0}} \frac{P(N(dx) > 1)P(N(dy) > 1)}{|dx||dy|} = 0$$

Assuming stationarity (invariant process by translation) and isotropy (invariant process by rotation) point processes are characterised by two basic properties[2]:

- **First order properties**

  These describe the intensity or the expected number of points by unit area in any location.

  Given the random variable $N(A)$, for $A \subseteq \mathbb{R}^d$, the first order characteristics are specified by the measure of intensity defined as:

$$\lambda(A) = E[N(A)], \qquad A \subseteq \mathbb{R}^d$$

  In certain cases this can also be expressed as follows:

$$\lambda(A) = \int_A \lambda(\zeta) d\zeta$$

  where λ is a non-negative function and represents the intensity function.

- **Second order properties**

These describe the relationship between arbitrary pairs of points. In the case of uniform or regular patterns, the probability of finding a point in the neighbourhood of the other is lower than it would be in a random pattern, while in a cluster pattern the probability is higher.

The factorial moment of order 2 describes the characteristics of the second order of the random variable $N(A)$, for $A \subseteq \mathbb{R}^d$ and it is represented by the following expression:

$$\alpha^{(2)}(C) = E[\, \sum_{\zeta \neq \eta \in X} I((\zeta, \eta)) \in C)],\ C \subseteq \mathbb{R}^d \times \mathbb{R}^d$$

The most commonly used estimate of the second-order properties is Ripley's K function, which estimates on all scales[2].

From a mathematical point of view, a spatial point process is characterized as:

- The first order intensity function:

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

- The second order intensity function:

$$\lambda_2(x, y) = \lim_{\substack{|dx| \to 0 \\ |dy| \to 0}} \left\{ \frac{\mathrm{E}[N(dx)N(dy)]}{|dx||dy|} \right\}$$

- The density of covariance:

$$\gamma(x, y) = \lambda_2(x, y) - \lambda(x)\lambda(y)$$

In particular, if you consider a stationary and isotropic point process, you can define $u = \|x - y\|$ and the above functions can be expressed as follows[2]:

1. $\lambda(x) = \ \lambda \ = E[N(A)]/|A|$ (constant for all A)
2. $\lambda_2(x, y) = \lambda_2(\|x - y\|)$
3. $\gamma(u) = \lambda_2(u) - \lambda^2$

Another interesting definition for the development of the analysis is the conditional intensity function expressed by:

$$\lambda_c(x/y) = \lim_{|dx|\to 0}\left\{\frac{E[N(dx)/event\ in\ y]}{|dx|}\right\}$$

The reduced second moment function ofan stationary and isotropic point process is expressed by the K-Ripley function which simplifies its interpretation[2]:

$$K(r) = \lambda^{-1}\ E[N_0(r)]$$

where$N_0(r)$ represents the number of events located at a distance smaller than$r$ to an arbitrary event (randomly chosen), λ the individual density and $\lambda K(r)$ the average number of events within a circle of radius r around a certain event of the pattern. Sometimes, it is also interesting to consider$\frac{\lambda^2 K(r)}{2}$, which is interpreted as the average number of different pairs of points with a distance smaller than or equal to r, and where one of the points belongs to a fixed surface unit subset A.

The practical importance of this feature is that it can be expressed as the mean of an observable amount, which suggests it is a good tool for estimating this function.

In order to relate $K(r)$and $\lambda_2(r)$, one assumes that the process is orderly, i.e, $E[N(dx)] \sim P[N(dx) = 1]$ because their ratio tends to 1 when $|dx| \to 0$. On the same way, one assumes also the equivalence relation $[E(dx)N(dy)] \sim P[N(dx) = N(dy) = 1]$.Thus, one affirms that the expected number of events at a distance less than r to an arbitrary event can be calculated by integrating the conditional intensity in the disk of the centre of origin and radius r:

$$\lambda K(r) = \int_0^{2\pi}\int_0^r \lambda_c(x/0)x dx d\theta$$

Knowing that $\lambda_c(x/0) = \lambda_2(x)/\lambda$, a new expression for function $K(.)$, which is easier to manipulate, can be computed as[2]:

$$K(r) = 2\pi\lambda^{-2}\int_0^r \lambda_2(x)x dx \tag{1}$$

Or, conversely as:

$$\lambda_2(r) = \lambda^2(2\pi r)^{-1}K'(r)$$

Although graphically $K(r)$ has a more intuitive interpretation than $\lambda_2(r)$, from the theoretical point of view, it is better to work with $\lambda_2(r)$ as it is easier to manipulate analytically[2].

2. Methods

2.1 Features and properties of spatial point processes

2.1.1 Homogenous spatial point processes

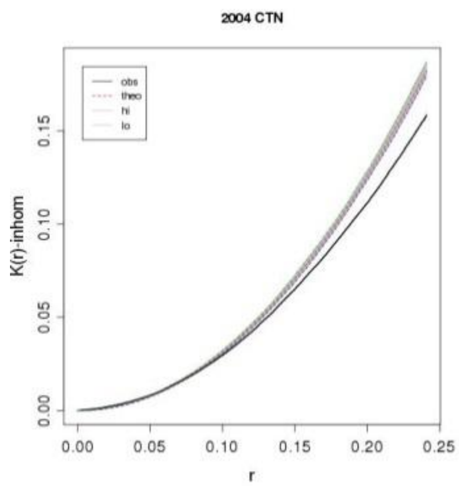
Spatial point processes are governed by a particular law of process which describes the spatial structure of these points: completely random distribution, regular or clustered.

To analyse the spatial structure of a pattern of points, firstly a test of complete spatial randomness (CSR, Complete Spatial Randomness) is required. This test attempts to detect if there is any data structure, i.e., if there is any interaction between the points of the process[6].

A complete random process is related to white noise and is characterised by its random variables which are not correlated. This is why these types of processes are identified by the homogeneous Poisson processes.

A test, in which the null hypothesis states that the process follows a pattern of a homogeneous Poisson process, i.e., that it follows a complete random distribution, is made to rule out completely random behaviour. The test consists of computing the K-function of the pattern of the observed points and comparing it with the theoretical K-function of a Poisson pattern of the same intensity[6]. In practice one constructs a joint graph with the observed values and the theoretical values and a visual comparison between the two resulting curves is made (see Figure 4).

Figure 4: Inhomogeneous K-function representation



Source: Own construction

In Figure 4, the bottom line represents the K-function of the observed points' patternand the upper band represents the confidence interval of the theory K-function. As the bottom line appears outside the upper band, one interprets that the observed data do not follow a random distribution, although depending on the graph displayed above or below the band there is some interaction between the points via attraction or repulsion.

As it has been depicted in Section 1.2.1, the K-function is the reduced second moment function or Ripley's K function and it is expressed as:

$$K(r) = \lambda^{-1}\, E[N_0(r)]$$

The function $K(r)$ describes the dependence between the pairs of points in the process[2].

For stationary processes, an easy way to estimate λ and $K(r)$ is:

$$\hat{\lambda} = \frac{N}{|A|} \qquad \hat{K}(r) = \frac{1}{\hat{\lambda}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} I(d_{ij} < r)$$

where$N$ is the number of points of the pattern, $|A|$ the surface of the study area, $\hat{\lambda}$ represents the observed number of events per unit area and $I(d_{ij} < r)$ the indicator function which is defined as:

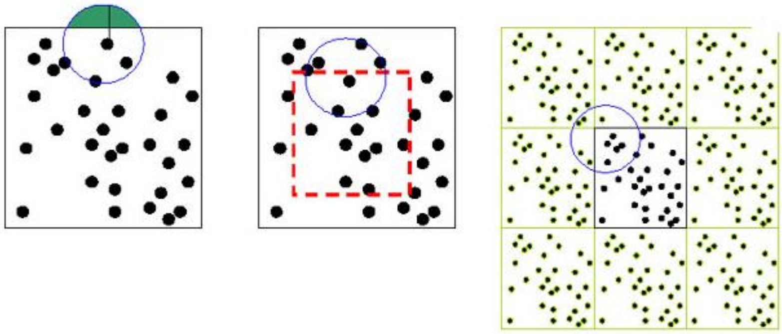
$$I = \begin{cases} 1 \text{ if } d_{ij} < r \\ 0 \text{ otherwise} \end{cases}$$

To estimate Ripley's K function, we must take into account that in many applications of spatial point processes the boundary of thestudy area is arbitrary and what is called the "border effect" may appear. This effect refers to the points that lie outside the analysed surface and are not considered to estimate the Ripley's K function, although they are at a distance less than r from a point located within the region. By ignoring this effect, biased estimations of the function K, especially for large values of r, are obtained[6].

There are different ways and estimators to correct this effect, such as, weighted counts around points near the edge[7] or to replicate the pattern around the study area[8]. However, as the potential solutions are not perfect, it is recommended not to calculate K(r) beyond $r < 1/3$ of the length of the shorter side of the study area[9] or, in the case of non-rectangular areas[10] not further than$r < (\frac{|A|}{2})^{1/2}$.

Figure 5 shows some of the methods used to correct the boundary effect graphically. From left to right, Ripley's Method, the buffer area method and the translation method.

Figure 5: Method to correct border effect



Source: Environmental information systems[6]

When considering the border effect, the K-function estimation for an observed point pattern in a region A, may take a slightly more complicated expression:

$$\widehat{K_1}(t) = \frac{|A|}{n(n-1)} \sum_{x \neq y} I(|y-x| \leq t) w_a(x,y)$$

where$w_a(x,y)$ is a corrector of the border effect[2].

Nevertheless, assuming CSR (a specific case of a homogenous Poisson process in $\mathbb{R}^2$) it follows that the value of$K(r)$is πr^2. Thus, the visual or numerical comparison between the theoretical and the observed K-function give the following classification[6]:

$K(r) > \pi r^2$ ⇒Aggregated process

$K(r) < \pi r^2$ ⇒Regular or uniform process

The graphic representation of the estimator $\widehat{K}(r)$, together with the upper and lower covers calculated by the Monte Carlo method, provide a graphical CSR test.

2.1.2 Extension to the inhomogeneous case

In the particular case of fires (the focus of this study) the intensity is clearly not constant, as the number of fires depends on the year and it will also be necessary to mention the inhomogeneous case. To extend the second order analysis of the process to the inhomogeneous case, we need to introduce the inhomogeneous K-function.

To define the inhomogeneous K-functioncertain preconditionsare needed. If$\lambda(x)$, $x \in \mathbb{R}^2$ is the first order intensity of the point process$X$, it is defined:

$$M(A,B) = E\left[\sum_{x_i \in X \cap A} \sum_{x_j \in X \cap B} \frac{1}{\lambda(x_i)\lambda(x_j)}\right]$$

And it is assumed that it is finite for every pairofBorel sets.Then$M$is thesecond order momentof the randommeasurement $\mathcal{F}$ which associate the weight $\frac{1}{\lambda(x_i)}$ to each event, that is to say,

$$\mathcal{F} = \sum_{x_i \epsilon X} \frac{1}{\lambda(x_i)}$$

It is said that a point process is second-order reweighted stationary when the random measure $\mathcal{F}$ is second-order stationary.

In this framework the hypothesis of constant intensity is removed but stationarity and isotropy remain. In particular, the process must be stationary second-order reweighted[2].

A second-order stationary process is also second-order reweighted stationary[2].

Calculating K_{inhom}requires a previous estimation of the intensity at each event. There are two possible estimation methods: parametric and non-parametric.

The first method consists of finding a fitting model that explains both the spatial trend and the interactions between events. If there is no interaction between events, parametric models (where the logarithm of the intensity is a polynomial), can explain the intensity of the processes belonging to all fires and also to those related to each type or cause. The formal expression of the intensity associated with these models is[2]:

$$\lambda(x) = \exp\{\beta^T F(x)\}$$

Where $F(x) = \{f_0 = 1, f_1 = x, f_2 = x^2, \dots, f_m = x^m\}$ is an m-order polynomial in $\mathbb{R}^2$ and β^T is the vector of the coefficients associated with the polynomial. These coefficients will be obtained by the method of fitting models by maximum pseudo-likelihood[9].

The second method for estimating the variable intensity requires constructing a kernel type estimator of λ. It has the following expression:

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^{n} k_h(x - X_i) = \frac{1}{p_h(x) h^2} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right)$$

where k is the kernel function, h the smoothing parameter and $p_h(x) = \int_A h^2 k\left(\frac{x-u}{h}\right) du$ is the corrector of the border effect. Similarly, a Gaussian kernel is used, where σ acts as a parameter window[9]. In this way, large values of σ carry on smoothing and approach a constant intensity, whereas excessively small values introduce too much variability and reflect a local trend rather than an overall one[2].

The interpretation of the inhomogeneous K-function is the same as it was in the homogeneous case but now the intensity is not constant but rather depends on the location of the events. In this case the intensity is represented by the function $\lambda(x_i)$, which is variable in x_i. The inhomogeneous K-function is defined as[2]:

$$K_{inhom}(r) = \frac{1}{|A|} E\left(\sum_{x_i \epsilon X \cap A} \sum_{x_j \epsilon (X \cap A) \setminus \{x_i\}} \frac{I(\left\| x_i - x_j \right\| \leq r)}{\lambda(x_i)\lambda(x_j)} \right)$$

where A is a bounded Borel set in $\mathbb{R}^2$, $I(.)$ is the indicator function, X is the point process and r the maximum distance between pairs of events x_i, x_j.

As an estimator, the following unbiased punctual estimator of the inhomogeneous K-function[11] can be considered:

$$\widehat{K}_{inhom}(t) = \frac{1}{|A|} \left(\sum_{x_i \epsilon X \cap A} \sum_{x_j \epsilon (X \cap A) \setminus \{x_i\}} \frac{I(\left\| x_i - x_j \right\| \leq r)}{\hat{\lambda}(x_i)\hat{\lambda}(x_j) w_{ij}} \right)$$

where w_{ij} is the corrector of the border effect[2].

As in the homogeneous case, once the inhomogeneous K-function for the observed process, represented by $\widehat{K}_{inhom}$, is estimated by following the same steps as in the homogeneous case, we can apply a CSR contrast based on this function. Then, the K-function $\widehat{K}_i(t), i = 1, \dots, s$, is computed for s-1 independent simulations of a process with estimated intensity $\hat{\lambda}(x)$ and the upper and lower covers are defined by the Monte Carlo method[2].

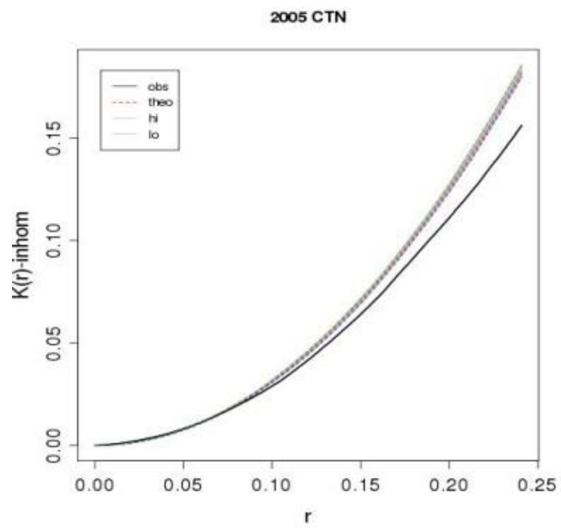
$$U(r) = \max_{i=2\ldots s}\{\widehat{K}_{inhom,i}(r)\}$$

$$L(r) = \min_{i=2\ldots s}\{\widehat{K}_{inhom,i}(r)\}$$

From these results we can graphically represent the observed data $\widehat{K}_{inhom,1}$(t), the covers and the empiric K-function. The result is a graphic test of CSR, which is interpreted similarly to the description given for the homogeneous case.

Figure 6 shows an example of a non-homogeneous K-test for all the fires in Catalonia in 2005. It may show that the black curve (representing the observed data) is not within the limits represented by the confidence interval of the theory K-function, so we can reject the null hypothesis and say that there is some interaction between the analysed data.

Figure 6: Example of a non-homogeneous K-test using the fire pattern of Catalonia of 2005



Source: Own construction using the free software R[12].

To improve data interpretation we often transform the $K(r)$ function and we use:

$$L(r) = \sqrt{\frac{K(r)}{\pi}}$$

in order to linearize the function and stabilize the variance. This new expression is interpreted representing the function $L(r) - r$with which the null hypothesis is rejected from the zero line.In this way, if$L(r) - r$ is significantly greater than zero, points follow a cluster distribution, whereas if it is less than zero it tends to follow a regular pattern[2].

We can verify directly that for a homogeneous Poisson process $L(r) = r$, by slightly simplifying the value that is obtained by considering the function $K(r)$, $K(r) = \pi r^2$.

In addition, we also can transform the inhomogeneous K-function by the expression

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

And likewise, we can consider the test based on the non-homogeneous L-function[2].

Under CSR, this function, as in the homogeneous case, verifies that $L_{inhom}(r) = r$.

2.2 Models for spatial point processes

2.2.1 Poisson processes

The homogeneous Poisson processes are the simplest stochastic models for a planar point pattern and are frequently referred to as the model of complete spatial randomness (CSR). They represent the base from which one constructs the theory of spatial point processes and they are characterised because their points are stochastically independent, and behave independently, which it is not a realistic option with natural phenomenon.

A point process is a flat homogeneous Poisson process of intensityλ if[2]:

1) The number of events in a flat region A, represented by $N(A)$, follows a Poisson distribution with mean $\lambda|A|$ where $|A|$ represents the area A and λ is the process intensity, i.e., the expected number of events per unit area.
2) Given n events $\{x_i\}_{i=1}^{n}$ in the region A, $x_i$form a random sample of a uniform distribution on A. Therefore, there is no interaction between the events.
3) For two disjoint regions Ay B, random variables $N(A)$ and $N(B)$ are independent.

Assuming $N(A)$ as a fixed number of events, the simulation of a partial realization of a Poisson process in A consists of generating uniform and independent events in A. If the shape of region A is complex, the simulation process is performed in a larger region with a simpler form, such as a rectangle or a disc, and one considers only the events inside A.

If you want $N(A)$ to vary randomly, one uses the above process preceded by the $N(A)$ simulation according to the corresponding Poisson distribution. In some implementations, direct simulation of $N(A)$ has a high computational cost. In 1979, an alternative method that can be used when A is a rectangle[13], for example $(0, a) \times (0, b)$ was proposed. This method is based on the fact that the location of the coordinate x of each event in the band $0 \leq y \leq b$, form a Poisson process with intensity λb. Therefore, the differences between successive x coordinates are independent realizations of an exponential random variable with distribution function[2]:

$$F(v) = 1 - \exp(-\lambda b v)\,;\; v \geq 0$$

The K-function of the homogeneous Poisson processes is represented by the following expression[2]:

$$K(r) = \frac{1}{A}\sum_{I=1}\sum_{j\neq 1}\frac{\omega_{ij}\, I(d_{ij}\ \leq r)}{\lambda^2}$$

where$A$ represents the area of the study region; λ is the intensity, ω_{ij} is the edge correction term, d_{ij} represents the distance between two points and I is the piecewise function such that:

$$I = \begin{cases} 1\ if\ d_{ij}\ < r \\ 0\ otherwise \end{cases}$$

The intensity (event number density) is the parameter that can be estimated in this model.

Within Poisson processes, inhomogeneous processes are more realistic than the last one, as they consider that the intensity is not constant but rather it is a heterogeneous function that includes the space component. They are the simplest models when it comes to non-stationary processes.

It is said that a point process is a non-homogeneous Poisson process if[2]:

1) The number of events in a region A, $N(A)$, follows a Poisson distribution with mean$\int_A \lambda(x)dx$, given any non-negative function $\lambda(x)$.
2) Given n events $\{x_i\}_{i=1}^n$ in the region A, x_i form a random sample from the distribution in A with probability distribution function proportional to $\lambda(x)$.
3) Given two disjoint regions Ay B, random variables $N(A)$ and $N(B)$ are independent.

Inhomogeneous Poisson processes can incorporate covariates, which provide additional information about the point pattern behaviour, thus improving the modelling[2]. Covariates are included in the intensity function$\lambda(x) = \lambda\left(z_1(x),\ z_2(x), \ldots, z_p(x)\right)$.

The K-function of these types of processes is defined as:

$$K(r) = \frac{1}{A}\sum_{I=1}\sum_{j\neq 1}\frac{\omega_{ij}\, I(d_{ij} \le r)}{\lambda(x_i)\lambda(x_j)}$$

where $A, \lambda, \omega_{ij}, d_{ij}$ and I represent the same parameters as in the homogeneous Poisson process and $\lambda(x_j)$ and $\lambda(x_i)$ are the values of the intensity function in x_j and x_i, respectively. In particular, the intensity function $\lambda(x)$ is modelled as a polynomial regression with logarithm:

$$\lambda(x) = \exp(\beta^T F(x))$$

where $F(x)$ is a variable vector and β^T is the regression parameters vector[2].

2.2.2 Thomas Processes

Homogeneous Thomas processes describe processes of dispersal, in which "offspring" are limited to aggregate around their "parent". Therefore, they model the effect of dispersal limitation. They are a particular class of Poisson cluster processes and can be used to model a series of clustered patterns[14,15].

Homogeneous Thomas processes are modelled in two steps. First, locations of parents are generated by a homogeneous Poisson process with a density j. Second, a group of offspring are produced around each parent. Their locations are assumed to be independent of one another and isotropically distributed around each parent with a Gaussian dispersal kernel, $N(0, r)$. The number of offspring is determined by a Poisson distribution with the mean being I[9,16].

It is said that a point process is a Poisson cluster process if:

1) The main events form a homogeneous Poisson process with intensity ρ.
2) Each main event produces a random number, S, of offspring, generated independently and identically distributed for each main event, according to the probability distributions ρ_s, s =1,2,…
3) The offspring locations, with respect to their predecessor, are independent and identically distributed according to a bivariate probability distribution function $h(.)$.

By agreement, final design is made only by overlapping the offspring of all the main events.

According to this definition, Poisson processes with clusters are stationary with intensity $\lambda = \rho\mu$where $\mu = E[S]$, and they are isotropic when in the last three properties it is considered a probability distribution function radially symmetric[2].

The K-function of the homogenous Thomas process is given by:

$$K(r) = \pi r^2 + \frac{1 - e^{\left(-r^2/4\sigma^2\right)}}{k}$$

where$r$ is distance, k represents the intensity of parents in a Poisson distribution and σ is the standard deviation of distance from offspring to the parent[14,15].

Inhomogeneous Thomas processes are the most complicated models of the four described thus far. They are used to evaluate the joint effects of covariates on the behaviour of events analysed[14,15]. This model is the same as a homogeneous Thomas process, except that the number of offspring per parent, I, is no longer constant and must be modeled by a spatially heterogeneous intensity function. As with the inhomogeneous Poisson process above, intensity functions are modelled by means of log-polynomial regressions.

2.2.3 Gibbs process: Area-Interaction

Gibbs processes are a fundamental class of point processes which emerged from Statistical Physics. They are able to capture the interaction structure of the generating spatio-temporal process, whose parameters can be estimated by maximum likelihood or pseudo-maximum likelihood[4].

Its general form is given by the expression:

$$p(x) = exp\left\{-v_0 - \sum_i v_1(x_i) - \sum_{i<j} v_2(x_i, x_j) - \sum_{i<j<k} v_3(x_i, x_j, x_k) - \cdots\right\}$$

where $x = \{x_i\},\ i = 1 \dots n(x)$, v_0 is constant and $v_k : W^k \to \mathbb{R} \cap \{-\infty\}$are symmetric functions for k=1, 2,… That is, the possible interactions between points can be decomposed.

Functions v_k are called interaction potentials and point processes can be classified by the order of interaction between points. It is said that a point process has a K-order interaction between points if $k = max\{j \in N: v_j \not\equiv 0\}$.

Gibbs point processes belong to the family of Markov processes[17] and they are characterised because there is a symmetric neighbourhood relationship $\sim$ and interactions are null expect for sets of points that are neighbours of each other (called cliques).

The Gibbs process used in this study is the area-interaction for its good properties and better modelling. It is a Markov point process consisting of a generalization of pair interaction of point processes, obtained by giving freedom to the order of interaction between points.

The probability density of a homogeneous area-interaction process in a compact region $A \subset \mathbb{R}^d$ with discsof radius r, intensity parameter k and interaction parameter γ is given by[18,19,20]:

$$f(x_1, \dots, x_n) = \alpha k^{n(x)} \gamma^{(-A(x))}$$

where $x_1, \dots, x_n$ represent the points of the pattern, $n(x)$ is the number of points in the pattern, and $A(x)$ is the area of the region formed by the union of discs of radius $r$centered at the points x_i. Here, α is a normalizing constant. The interaction parameter γ can be any positive number. If $\gamma = 1$, then the model is reduced to a Poisson process with intensity k. 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 $2r$.

These kinds of models compute the likelihood function by neighbourhood. Each environment is determined by a radius that maximizes the likelihood function. Given the shape and size of A, the radius is defined by the expression

$$R = \frac{1}{2} sup \{||x - y||: x, y \in A\}$$

The area-interaction is very convenient because it creates slightly aggregated or regular patterns. In addition to computing areas compact sets less standard and more general than disks can be used.

The probability density function initially described, can be slightly modified, parameterizing the model into a different form easier to interpret[9]. In canonical scale-free form, the probability density is rewritten as

$$f(x_1, \dots, x_n) = \alpha \beta^{n(x)} \eta^{(-C(x))}$$

whereβ is the new intensity parameter, η is the new interaction parameter and $C(x) = B(x) - n(x)$ is the interaction potential, $B(x) = A(x)/(\pi r^2)$ is the normalised area (so that the discs have unit area).

The inhomogeneous area interaction process is similar, except that the contribution of each individual point x_i is a function $\beta(x_i)$ of location rather than a constant beta.

2.3. Models for spatio-temporal point processes

Returning to the definition of point process with marks, spatio-temporal point processes can be introduced as a series of observations of a point process with marks at instants $(t_1, t_2, \dots, t_n) \in T$.It is assumed that events are distributed in a certain spatial region $D \subset \mathbb{R}^d$ and occur at a specific temporal interval$(0, T)$. Following the above notation, these processes are interpreted as a point process[2] in $\mathbb{R}^d \times \Psi \times T$.

The spatio-temporal modellingof spatial processes is a recent field of research and is presented as an extension of the spatial case. Their study is distinguished by the three types of data that are in spatial statistics; geostatistical data, lattice data and point processes. They indicate data collected in space and evolve in time.

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 $(s_1, \dots, s_n)$ indicate the spatial units, at which the measurements are taken, and $(t_1, \dots, t_n)$ the time points.

To analyse spacetime data it is important to distinguish whether individual events are developed in a continuous spatio-temporal or it is considered that the time scale is either naturally discreet or it is discretised only considering spatial pattern events aggregated over a sequence of discrete time of periods. This distinction is essential when deciding the analysis method as it differs in each case.

**Hereinafter s will denote the space component*

2.3.1 Mixed models

Mixed models are a generalization of the classical linear regression models and are characterised by considering two or more dimensions of analysis simultaneously. The term mixed model refers to the use of both fixed and random effects in the same analysis. Fixed effects have levels that are of primary interest and would be used again if the experiment were repeated. Random effects have levels that are not of primary interest, but rather are thought of as a random selection from a much larger set of levels. Subject effects are almost always random effects, while treatment levels are almost always fixed effects[21].

Mixed models allow solving issues of complex experimental design study, based on the simultaneous modelling of the response's expected value and its variability. Such models include multi-level designs or multi-level, also called hierarchical models, and longitudinal studies, or repeated measures[21].

Multi-level studies have a hierarchical structure where observations are grouped into clusters, and the distribution of an observation is determined not only by common structure among all clusters, but also by the specific structure of the cluster where this observation belongs[21]. In general, and considering a lineal response, a hierarchical model can be specified by the following equation[1]:

$$Y_{ij} = \beta_1 z1_{ij} + \beta_2 z2_{ij} + \cdots + \beta_p zp_{ij} + \varepsilon_{ij}$$

which defines the observations of the dependent variable Y_{ij} as being determined by p observed explanatory variables in j, zp_{ij}. Some of the explanatory variables are fixed ($zp_{ij} = zp_i \ \forall j$) and others are variable, depending on the subscript i. $\beta$are$p$ unknown parameters. Finally, ε_{ij}, are independent random variables with a zero mean and have met the following requirements[1]:

$$E(\varepsilon_{ij}) = 0$$

$$E\left(\varepsilon_{ij}^2\right) = \sigma_\varepsilon^2 \ \ constant$$

$$E(\varepsilon_{it}\varepsilon_{js}) = 0 \ \forall \, t \neq s \ \ y \ \ \forall i \neq j$$

Using matrix notation and considering the entire sample, the model is represented by the expression[1]:

$$Y = Z\beta + \epsilon$$

And it is specified assuming $E(Y) = \mu, \ Var(Y) = V$ and that V follows the scheme:

$$V = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & & \ddots & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V_n \end{bmatrix}$$

It is important to point out that V must always be a block diagonal, i.e., it is assumed that areas (or clusters) are mutually independent.

On the other hand, longitudinal studies or repeated measures involve repeated observations of the same variables over long periods of time. The structure of these models can be considered mixed, with observations (repeated) grouped within each individual and time, which can be considered as another explanatory variable within each group. This type of analysis is the only one which can distinguish between the variance between individuals (interindividual) and variation within the individual (intraindividual)[21]. The general expression is[1]:

$$Y_{ij} = \beta_0 + \beta_1 z2_{ij} + \cdots + \beta_p zp_{ij} + \varepsilon_{ij}$$

As an advantage in comparison to cross-sectional, longitudinal analyses allow the temporal order of interest events to be studied. In particular, they make it possible to determine if the risk factors precede the possible effects of these factors on the variations of the variable of interest; a feature called temporality. Longitudinal analysis can be approximated marginally or conditionally[1,21].

On one hand, the marginal approach describes variation in population means of subgroups, averaged over all individuals. They attempt to explain the relationship between the dependent variable and explanatory variables independently of the intraindividual variability. This approach implies that both, the intercept (β_0) and the coefficients associated to the explanatory variables, are common to all individuals. There is not individual heterogeneity, i.e., all the effects of the explanatory variables, including the intercept, are fixed. The random effect (ε_{ij}) has a constant variance and is correlated. Covariance parameters, i.e., autocorrelation and/or heteroscedasticity, are not of interest so that the marginal approach controls them but they are not estimated[1,21].

On the other hand, the conditional approach makes individual inferences modelling simultaneouslythe mean of the dependent variable (interindividual variability) and the covariance or correlation structure (intraindividual variability). In this approach, parameters defining the correlation have the same or even more interest that those corresponding to the average. The best known conditional approach is given by random effects models which assume that the effects of some (or all) explanatory variables (regression coefficients) are specific to individuals (not common to all of them). There is individual heterogeneity, which is due to unobservable factors (or omitted variables) common to some individuals. In this sense, the representation of the model considers the variations of the parameters according to each individual[1,21]:

$$Y_{ij} = \beta_{0i} + \beta_{1i} z2_{ij} + \cdots + \beta_{pi} zp_{ij} + \varepsilon_{ij}$$

The Markov transition models, autoregressive or with a covariance structure, are another type of conditional approach. They model the conditional expectation of the response and the dependence (correlation) between the observations within each group in a single equation. It can be considered a first order autoregressive model in which the conditional expectation of the response variable depends not only on the explanatory variables, but also on the prior behaviour itself. We therefore introduce another explanatory variable, corresponding to a response variable delayed one period[1,21].

$$Y_{1ij} = \beta_0 + \beta_1 z2_{ij} + \cdots + \beta_p zp_{ij} + \gamma Y_{1ij-1} + \varepsilon_{ij}$$

Moreover, the random effect (ε_{ij}) has no constant variance and it is autocorrelated.

Autoregressive models, unlike other conditional models, assume that the dependence between repeated observations, represented by the coefficient associated with the lagged response (γ), is a fixed effect, that is, common to all individuals. Moreover, the dependence structure by autoregressive models of order one implies that the greatest influence on the dependent variable is produced by the value immediately preceding, with the influence declining exponentially as we move away in time[1,21].

Depending on the response variable to be analysed, either a statistical model or another model will be used. Thus, if the dependent variable is a continuous quantitative variable, distributed normally then, linear regression models, known as linear mixed models will be used, otherwise nonlinear mixed models will serve. When the response variable is discrete, quantitative Poisson regressions (mixed) will be used, and the analysis will be based on binomial logistic or multinomial regressions (mixed). Finally, when the response variable is dichotomous, binomial distribution will be used, and when it is polytomous, a multinomial probability distribution will be used[21].

From a mathematical point of view the notation in matrix format of the mixed models can be represented as:

$$y = Z_f \beta + Z_r u + \varepsilon$$

where y is the vector of observations with mean $E(y) = Z_f \beta$, β is a vector of fixed effects, u is an independent and identically distributed vector (iid) of random effects with mean $E(u) = 0$ and variance $var(u) = \sigma_u^2$ and ε is a vector of random error terms iid with mean $E(\varepsilon) = 0$ and variance $var(\varepsilon) = \sigma_\varepsilon^2$. Z_f and Z_r are matrices of explanatory variables on the observations β and u, corresponding to the fixed and random effects, respectively.

The variance of the random effect reflects the variability between individuals, while the variance of the error collects the variability which is not explained by the model within each individual. If the variance of random effects was null, the model would be equal to the fixed effects model or lineal regression.

2.4 Spatio-temporal mixed models

Part of this Thesis focuses on longitudinal studies in which the time dependence is modelled in a spatial pattern of points considering the variation depending on time.

The mathematical theory of spatial point processes is well defined[22,23]. However, most models for specific applications are restricted either to point processes in time or to the two-dimensional space.

There are more adaptable types of models which allow solving some of the problems presented by point processes in fitting models. In particular, Cox processes are widely used as models for point patterns which are thought to reflect underlying environmental heterogeneity.These models are useful when the observed data presents a complex structure, one that would be impossible to represent in a regular lattice.

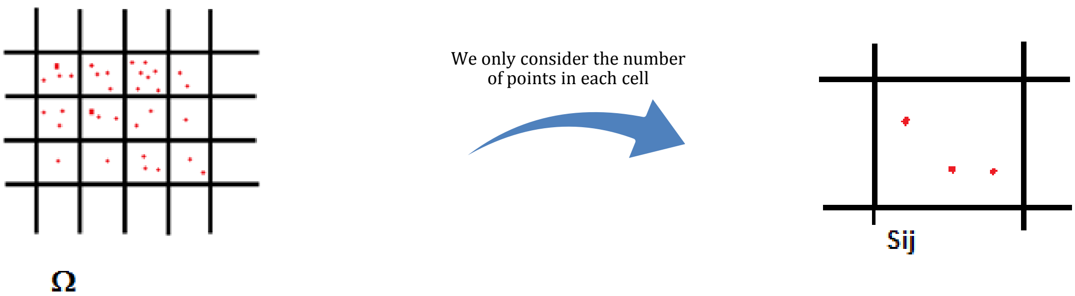
Moreover, when the analysed data contain a significant number of zeros, Cox models are not adequate and it is necessary to use another type of mixed models. There are various alternatives. On one hand, there are zero inflated Poisson models (ZIP) which might be used to model count data for which the proportion of zero counts is greater than expected on the basis of the mean of the non-zero counts[24,25]. On the other hand, there are Hurdle models[26,27] which are modified count models with two processes, one generating the zeros and one generating the positive values. The main difference between the two models is that ZIP model distinguish between two kinds of zeros, "true zeros" and excess zeros, whereas Hurdle models analyse those zeros that are zero at this moment but can be non-zeros in the future.

2.4.1 Log-Gaussian Cox processes (LGCPs)

Log-Gaussian Cox models (LGCP) are a particular case of Cox processes and are particularly interesting as models for point patterns which are thought to reflect underlying environmental heterogeneity. However, standard methods to fit Cox processes have a high computational cost and those methods which use Markov chains by Monte Carlo methods (MCMC) are very difficult to fit this problem.

Recently, a flexible framework using integrated nested Laplace approximations[28] for fitting complicated LGCPs has been proposed (INLA)[29]. This approach is based on finding a Poisson approximation to the likelihood function of the LGCP and uses this to make the inference. This approach is done by replacing the concept of regular lattice, created on the observed points, to consider the number of points in each cell (see Figure 7).

Figure 7: Scheme how a complicated LGCP is fitted



Source: Own construction

Although this approach is still based on a regular lattice, it can be shown that if the lattice is fine enough and appropriately discretised[30], this approach leads to consistent estimates. However, this approach could be highly inefficient, especially when the intensity of the process is high, the observation window is large or, as in the case of wildfires, typically oddly shaped[31].

Consider a bounded region$\Omega \in \mathbb{R}^2$. As has been described in previous sections, the simplest model, and one of the most commonly used in the context of point processes, is the inhomogeneous Poisson process in which the number of points within a region $D \subset \Omega$ is distributed as a Poisson with mean $\Lambda(\mathrm{D}) = \int_D \lambda(s)ds$, where $\lambda(s)$ is the surface intensity of a point process. Given the intensity surface and a point of the model Y, the likelihood for an LGCP is of the form

$$\pi(Y|\lambda) = exp\left(|\Omega| - \int_{\Omega} \lambda(s)ds\right) \prod_{s_i \in Y} \lambda(s_i) \quad (2)$$

where the integral is complicated by the stochastic nature of $\lambda(s)$. However, this integral can be numerically computed using fairly traditional methods.

Considering the intensity surface as a realization of a random field $\lambda(s)$ a type of point process, called Cox process, is obtained[32]. These types of processes are particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field[31,33].

The Log Gaussian Cox intensity surface is modelled as

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

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 because its likelihood function follows the expression (2) which, as it is said, due to the stochastic structure of $\lambda(s)$, includes an integral which is difficult to solve.

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[28].

2.4.2 Zero inflated Poisson

As discussed in previous sections, a Poisson model is assumed for modelling the distribution of the count observation or, at least, approximating its distribution. However, in various applications it has been observed 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[24,25].

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.

Considering Λ_{jt} as the total intensity per cell, we can thus define the number of observations in a specific cell as

$$N_{jt} \sim \begin{Bmatrix} ZIP0(\Lambda_{jt}) \\ ZIP1(\Lambda_{jt}) \end{Bmatrix}$$

Different types of zero-inflated Poisson models differ from the others in terms of the form of their likelihood functions[34].

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 denote 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 $p + (1-p)Po(y,\theta)$.

2.4.3 Poisson Hurdle models

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 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, which generates counts, some of which can be equal to zero.

A Hurdle model consists of two stages:

The first part of the decision process can be modelled using a logistic regression, that models the probability of a specific event happening:

$$p_{ikt} = Prob(y_{itk} > A|Z, \beta)$$

$$\log\left(\frac{p_{itk}}{1-p_{itk}}\right) = Z^{'}\beta + S_i + \tau_t + \upsilon_{it}$$

In accordance with that proposed by[23], in the second part of the model the distribution of an event happening is a truncated Poisson that models the number of events that there are per spatial unit, introducing covariates and spatial random effects:

$$p(y_{itk}|S_i) = (1-p_{itk})1_{(y_{itk}<A)} + p_{itk}Tpois(y_{itk};\mu_{itk})1_{(y_{itk}>A)}$$

$$log(\mu_{itk}) = \eta(p_{itk})$$

$$\eta(p_{itk}) = \sum_{\alpha}\beta_\alpha z_{\alpha,it} + S_i + \tau_t + \upsilon_{it}$$

where$Tpois(y_{itk};\mu_{itk})$ denotes a truncated Poisson distribution with parameterμ_{itk}; η denotes a link function such as the logit link; $z_{\alpha,it}$represents the same spatial covariates used in the first part; and β_α denotes the parameters associated with covariates. We also introduce three random effects: (i) spatial dependence, S_i; (ii) temporal dependence, τ_t and (iii) spatio-temporal interaction, υ_{it}.

This particular estimation process has 2 steps. In the first step we use a binomial link to estimate the probability of occurrence of an event. The probabilities of occurrence obtained from this first step are used in the second step as interim priors. In the second step the link is a truncated Poisson distribution. In any case, the likelihood of each part is introduced multiplicatively in only one equation.

3 References

1. Saez M, Saurina C. *Spatial Statistics and Epidemiology* [in Spanish]. Girona: Documenta Universitaria, 2007.

2. Fuentes Santos I. *Statistical inferenceforspatialpoint processes. Application to the analysisof forest fires inGalicia* [in Spanish]. Master's degree dissertation. University of Coruña, Spain: 2008-2009.

3. Arunachalam V.Point processes, product densities and celular biology [in Spanish]. *Revista Colombiana de Estadística* 2005; 28(1):1-16.

4. Mateu J, Montes F.Modelling the spatial distribution of cysts in the porpoise stomach by Gibbs process [in Spanish]. *Qüestiió* 1998; 22(1):175-194.

5. Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*.New Jersey: Holden-Day, 1976.

6. De la Cruz Rot M. Introduction to mapped data analysis or some of the (many) things that I can do if I have coordinates [in Spanish]. *Ecosistemas* 2006; 15(3):19-39.

7. Ripley BD. *Statistical inference for spatial processes*. Cambridge CB2 1RP: Cambridge University Press, 1988.

8. Ohser J. On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik, series Statistics* 1983; 14.

9. Baddeley AJ, Turner R. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 2005; 12:1-42.

10. Dixon PM. Ripley's K function. *Encyclopedia of Environmetrics* 2002; 3:1796-1803.

11. Baddeley AJ, Moller J, Waagepetersen R..Non and semi- parametric estimation of interactions in inhomogeneous point patterns. *Statistica Neerlandica* 2000; 54:329-350.

12. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [Available at http://www.r-project.org/, 2010, accessed on September 24, 2011].

13. Lewis PAW, Shedler GS. Simulation of nonhomogenous Poisson processes by thinning. *Naval Research Logistics Quarterly* 1979; 26(3):403-413.

14. Diggle PJ. *Statistical Analysis of Spatial Point Patterns (2nd Edition)*. London UK: Arnold, 2003.

15. Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical analysis and modelling of spatial point patterns*. New York: Wiley, 2008.

16. Møller J, Waagepetersen RP. *Statistical inference and simulation for spatial point processes*. New York: CRC Press, 2004.

17. Ripley BD, Kelly FP. Markov point processes. *Journal of the London Mathematical Society* 1977; 15:188-192.

18. Widom B, Rowlinson JS. New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics* 1970; 52:1670–1684.

19. Baddeley AJ, van Lieshout MNM. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* 1995; 47:601-619.

20. Van Lieshout MNM*. Markov Point Processes and Their Applications.* London: Imperial College Press, 2000*.*

21. Saez M. The problem of repeated measures. Longitudinal epidemiologycal analysis [in Spanish]. *Gaceta Sanitaria* 2001; 15(4):347-352

22. Bremaud P. *Point Processes and Queues: Martingale Dynamics*. New York: *Springer*, 1981.

23. Daley D, Vere Jones D. *An Introduction to the Theory of Point Processes*. New York: Springer Volume 2, 2003.

24. Breslow NE. Extra-Poisson variation in log-linear models. *Applied Statistics* 1984; 33(1):38-44.

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

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

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

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

29. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B* 2009; 71(2):319-392 [Available at http://www.math.ntnu.no/~hrue/r-inla.org/papers/inla-rss.pdf, accessed on May 21, 2011].

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

31. Simpson D, Illian J, Lindgren F, Sørbye SH, Rue H. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Technical Report Trondheim University* 2011. [Available at: http://www.math.ntnu.no/~daniesi/S10-2011.pdf, accessed on August 7, 2012].

32. Møller J, Waagepetersen RP. *Statistical inference and simulation for spatial point processes*. New York: CRC Press, 2003.

33. Illian JB, Hendrichsen DK. Gibbs point process models with mixed effects. *Environmetrics* 2010; 21:341-353.

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

Chapter 3.Results

Spatial pattern modelling of wildfires in Catalonia, Spain 2004-2008

Laura Serra[1,2], Pablo Juan[3], Diego Varga[2,4], Jorge Mateu[3], Marc Saez[2,1],

[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

Corresponding author:

Prof. Marc Saez, PhD, CStat, CSci
Research Group on Statistics, Econometrics and Health (GRECS) and CIBER of Epidemiology and Public Health (CIBERESP)
University of Girona
Campus de Montilivi, 17071 Girona
Tel 34-972-418338, Fax 34-972-418032
http://www3.udg.edu/fcee/economia/english/grecs.htm

e-mail: marc.saez@udg.edu

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

Abstract

The paper has three objectives: firstly, to evaluate how the extent of clustering in wildfires differs across the years they occurred; secondly, to analyse the influence of covariates on trends in the intensity of wildfire locations; and thirdly, to build maps of wildfire risks, by year and cause of ignition, in order to provide a tool for preventing and managing vulnerability levels. For these objectives we analysed the spatio-temporal patterns produced by wildfire incidences in Catalonia, located in the north-east of the Iberian Peninsula. The methodology used has allowed us to quantify and assess possible spatial relationships between the distribution of risk of ignition and causes. These results may be useful in fire management decision-making and planning. The methods shown in this paper may contribute to the prevention and management of wildfires, which are not random in space or time, as we have shown here.

Key words: wildfire, spatial point processes, marks, covariates, Area-interaction processes

1.- Introduction

A wildfire is any uncontrolled fire in combustible vegetation that occurs in the countryside or a wilderness area. 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 characterised in terms of the cause of ignition, their physical properties such as speed of propagation, the combustible material present, 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 (National Wildfire Coordinating Group, 1998; Scott, 2000). McRae (1992), among others, suggests that lightning ignitions do not occur anywhere, but favour locations satisfying certain terrain conditions. However, many wildfires are attributed to human sources (Pyne *et al.*, 1996). First, there are human actions that directly cause ignitions deliberately or accidentally. In addition to this, great social upheaval in the last century has led much of the population to move from rural to urban areas. Many areas have witnessed an abandonment of farming and livestock practices, which leads to an accumulation of fuel for fires to feed on. The abandonment of rural areas has also reduced the capacity among the population for noticing fires and taking action when they first begin (William *et al.*, 2000; Badia *et al.,* 2002).

Fire risk is very high in the Mediterranean region due to its marked seasonality, with high temperatures and low humidity in summer, and these climatic trends interact with landscape dynamics. In the case of Catalonia, the process of afforestation in different agricultural areas and the increasing abandonment of rural activities have led to a situation of extreme vulnerability to fires, especially in Mediterranean mountain areas, where the aforementioned factors have led to forests being abandoned and their subsequent expansion, proliferation and interconnection (Loepfe et al., 2011). In addition to this, other factors, such as constructing second homes in these forest areas, the proliferation of roads and electricity networks and an increased flow of people make this region more susceptible to the ignition of large forest fires (Díaz-Delgado and Pons, 2001; Moreira et al., 2001).

Given that fire is a naturally occurring element in the Mediterranean ecosystem, the prevention and suppression of forest fires needs to be addressed so as to minimize risk and vulnerability of society.

In fact, there are by now many studies of the spatial patterns of wildfire risk in various locations around the globe. Without being exhaustive, and referring only to more recent 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 2005; Millington *et* *al.*, 2009, 2010; Romero-Calcerrada *et al.*, 2010), Asia (Liu *et al.*, 2012) and Oceania (O'Donnell, 2011), among many others.

If we associate wildfires with their spatial coordinates, the longitude and latitude of the centroid of the burned area or the place where they were detected, along with other variables such as size or cause, and we know the time at which they started, it is possible to identify them by means of a spatio-temporal stochastic process. Such processes, called spatial point processes, often present dependences between the spatial positions and time instants, as well as interdependence between one another.

Spatial point processes are complex stochastic models that describe the location of events of interest and occasionally some information on these events. The most common models are those where the locations are given in two dimensions. Univariate point processes include only the location of events; point processes with marks (or marked point processes) include additional information about each event. These data sets can be used to respond to a variety of questions. The scientific context of these questions depends on the area of application, but they can be classified into three broad groups. First, one might be interested in whether the spatial pattern for the observed data is grouped, distributed regularly or random. A second group of questions would refer to the relationship between different types of events in a marked process or process with marks (variables measured only at fire locations, such as size of area burned, cause of fire or year of occurrence). The third and last group of questions focuses on density (number of events per unit area).

What is usually of more interest is to detect trends in the intensity of fire locations (i.e. probability of occurrence), and to determine how (or whether) such trends are influenced by covariates, observable at each location of the spatial window. These covariates might include vegetation or land use, other descriptors of terrain (such as elevation, slope and orientation), and others such as proximity to human populations or to concomitants of human activity (roads and railroads). Interaction between points may generally be of some interest in its own right. More important is the impact of the presence of interaction on statistical inference concerning trends and its dependence on covariates. Temporal clustering of wildfires, whether deriving from multiple ignition lightning events, arson (Butry and Prestemon, 2005), or other sources, combined with favourable fuel and weather conditions, can force suppression resource rationing across space. Spatial clustering can also indicate the presence of risk factors.

This paper has three objectives. Firstly, to evaluate how the extent of clustering in wildfires differs across marks, in particular years and causes of wildfire ignition. Secondly, to analyse the influence of covariates such as land use, slope, aspect and hill shade on trends in the intensity of wildfire locations. And finally, to build maps of wildfire risk, by year and cause of ignition, in order to provide a tool for managing vulnerability levels.

2.- Methods

2.1.- Data set

We have analysed the spatio-temporal patterns produced by wildfire incidences in Catalonia, located in the north-east of the Iberian Peninsula. The region is bordered by mountain landscapes, the Pyrenees in the north and the Iberian System in the south. The region is further delimited by the Ebro river to the south and south-west, and the Mediterranean coast to the east. It is a region with a surface area of 30,000 square kilometres (12,355 sq mi), representing 6.4% of the Spanish national territory.

The total number of fires recorded in the area studied during the period 2004-2008 was 3,083. In addition to the locations of the fire centroids, measured in Cartesian coordinates (Mercator transversal projections, UTM, Datum ETRS89, zone 31-N), several marks and covariates were considered.

Variables measured only at fire locations are called marks. In this paper, marks included the year the wildfire occurred (from 2004 to 2008) and also the cause of ignition of each wildfire (classified as: natural causes; negligence and accidents; intentional or arson; and unknown causes and revived).

Spatial covariates were also considered. In particular, three continuous covariates: slope, aspect and hill shade; 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 their availability for combustion (Ordóñez *et al.*, 2012) but also have effects on weather, inducing several local wind conditions including slope and valley winds. In fact, Dillon *et al.* (2011) point out that topographic variables were relatively more important predictors of severe fire occurrence than either climate or weather variables.

Slope is the steepness or degree of incline of a 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 surroundings. Slope was expressed in degrees. Aspect is the orientation of the slope, measured clockwise in degrees from 0 to 360, where 0 is north-facing, 90 is east-facing, 180 is south-facing, and 270 is west-facing. Hill shading is a technique used to visualize terrain as shaded relief, 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. With respect to land use variables, we used the CORINE database (Coordination of Information on the Environment). The CORINE program was initiated in 1985 by the European Commission and was adopted by the European Environment Agency (EEA) in 1994. The main objective of the CORINE program is to capture numerical data and geographical order to create a European database on environment for certain priority topics such as land cover and biotopes (habitats), through the interpretation of images collected by the Landsat series of satellites and SPOT. Although this is based on remote sensing images as a data source it is actually a photo interpretation project and not automated classification. In this paper we have used the CORINE land cover map for the year 2006 (European Environment Agency, 2007). Data are gathered 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. The database includes forty-four categories, in accordance with a standard European nomenclature, organised into five large groups: artificial surfaces, agricultural areas, forest and semi-natural areas, wetlands and water bodies (Heymann *et al.*, 1994). In this paper we reclassified land use into ten categories: coniferous forests; dense forests; pastures; fruit trees and berries; artificial non-agricultural vegetated areas; transitional woodland scrub; scrub; natural grassland; mixed forests; and urban, beaches, sand, bare rocks, burnt areas and water bodies.

In order to model the dependence of a point pattern on a spatial covariate, there are several requirements. Firstly, the covariate must be a quantity observable at each location in the window (e.g. slope, aspect and hill shade). Such covariates may be continuous values or factors (e.g. land use). Secondly, the values of the covariate at each point of the data point pattern must be available. Thirdly, the values at some other points in the window must also be available.

2.2.- Statistical methods

The simplest of all possible point process models is the constant intensity Poisson process, frequently referred to as the model of complete spatial randomness (CSR). In this model, the points of a spatial pattern are stochastically independent. The nature of the phenomenon under study or a casual glance at a plot of the data usually makes it obvious when CSR is not a realistic option.

The nature or behavior of a point pattern may be thought of as comprising two components, trend and dependence, or interaction between the points of the patterns. The simplest manifestation of such interaction consists of either attraction (aggregation or clustering) or repulsion (regularity) in the pattern. A useful step in analyzing a point pattern is to apply graphical tools which reveal information as to the nature of the interaction. A widely used tool for exploring the nature of interaction is Ripley's K-function (Ripley, 1976; Ripley, 1977; Cressie, 1993; Diggle, 2003).

The basic idea in interpreting the K-function is that a constant intensity Poisson process (a process exhibiting CSR) has a K-function equal to $K(r) = \pi r^2$. If there is attraction (with impact at distance *r*, the bounding radius of the spatial domain), then *K(r)* is larger than it would be under CSR. Conversely, if there is repulsion, then *K(r) i*s smaller than it would be under CSR.

2.2.1.- Spatial models

Poisson processes

The homogeneous Poisson process is the simplest point process that represents no underlying process, corresponding in our case to complete randomness in wildfire distribution. The K-function of the homogenous Poisson process is defined as:

$$K(r)=\frac{1}{A}\sum_{i=1}\sum_{j\neq 1}\frac{\omega_{ij}I(d_{ij}\leq r)}{\lambda^2}$$

In this function, *A* denotes the area of the plot; *λ* is wildfire density, $\omega_{ij}$is an edge correction term, d_{ij} represents the distance between two points, and *I* is an index

function where $I = 1$, if $d_{ij} \leq r$, and $I = 0$ otherwise (Ripley, 1976). Wildfire density, *λ*, is the parameter to be estimated in this model.

The inhomogeneous Poisson process can be used to model heterogeneous association in wildfires. In this model, relationships between density and heterogeneity are included via a spatially heterogeneous intensity function, *λ(s)* (Diggle, 2003; Illian *et al.*, 2008). The *K*-function of the inhomogeneous Poisson process is defined as

$$K_{inh}(r)=\frac{1}{A}\sum_{i=1}\sum_{j\neq 1}\frac{\omega_{ij}I(d_{ij}\leq r)}{\lambda(s_i)\lambda(s_j)}$$

where*A*, *λ,* ω_{ij}, d_{ij}, and *I* are the same as in the homogeneous Poisson process; and $\lambda(s_i)$ and $\lambda(s_j)$ are the values of the intensity function at points s_i and s_j, respectively (Diggle, 2003; Illian *et al.*, 2008).

Specifically, the intensity function, *λ(s)*, is modelled as a log-polynomial regression:

$$\lambda(s)=\exp\left(\beta^T X(s)\right)$$

where*X(s)* is a vector of variables and β^T is a vector of regression parameters. Two different types of log-polynomial regressions were used in this study: log-linear regressions with covariates and log-quadratic regressions with the coordinates of the wildfire.

Thomas processes

The homogeneous Thomas process is a particular class of Poisson cluster process and can be used to model a series of clustered patterns (Diggle, 2003; Illian *et al.*, 2008). This model describes processes of dispersal, in which 'offspring' are limited to aggregate around their 'parent'. Therefore, it models the effect of dispersal limitation. The homogeneous Thomas process is modelled by two steps. First, locations of parents are generated by a homogeneous Poisson process with a density, *j*. Second, a group of offspring are produced around each parent. Their locations are assumed to be independent of one another and isotropically distributed around each parent with a Gaussian dispersal kernel, *N(0,r)*. The number of offspring is determined by a Poisson distribution with mean being *l* (Moller and Waagepetersen, 2004; Baddeley and Turner, 2005).

The K-function of the homogenous Thomas process is given by:

$$K(r) = \pi r^2 + \frac{1 - e^{\left(-r^2/4\sigma^2\right)}}{\kappa}$$

where$r$ is distance, κ represents the intensity of parents in a Poisson distribution and σ is standard deviation of distance from offspring to the parent (Diggle, 2003; Illian *et al.*, 2008). Mean number of offspring per parent in a Poisson distribution, *l*, can be inferred from estimated intensity $\lambda$and κ.

The inhomogeneous Thomas process is used to evaluate the joint effects of covariates (Diggle, 2003; Illian *et al.*, 2008). This model is the same as a homogeneous Thomas process, except that the number of offspring per parent, I, is no longer constant and must be estimated by a spatially heterogeneous intensity function. As with the inhomogeneous Poisson process above, intensity functions were modelled by means of log-polynomial regressions.

Area-interaction processes

The homogeneous area-interaction process (Widom and Rowlinson, 1970; Baddeley and van Lieshout, 1995) with disc radius r, intensity parameter κ and interaction parameter γ is a point process with probability density:

$$f\left(s_1,\ldots,s_n\right) = \alpha\kappa^{n(s)}\gamma^{(-A(s))}$$

where $S_1, \ldots S_n$ represent the points of the pattern, *n(s)* is the number of points in the pattern, and *A(s)* is the area of the region formed by the union of discs of radius *r*centred at the points $S_1, \ldots S_n$. Here, *α* is a normalizing constant.

The interaction parameter *γ* can be any positive number. If *γ = 1* then the model reduces to a Poisson process with intensity *κ*. If *γ < 1* then the process is regular, while if *γ > 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*.

Here, we parameterised the model in a different form (Baddeley and Turner, 2005), which is easier to interpret. In canonical scale-free form, the probability density is rewritten as

$$f(s_1 \ldots, s_n) = \alpha \beta^{n(s)} \eta^{(-C(s))}$$

where*β* is the new intensity parameter, *η* is the new interaction parameter, and *C(s) = B(s) - n(s)* is the interaction potential. Here

$$B(s) = A(s) / (\pi \cdot r^2)$$

is the normalised area (so that the discs have unit area). In this formula, each isolated point of the pattern contributes a factor *β* to the probability density (so the first order trend is *β*). The quantity *C(s)* is a true interaction potential, in the sense that *C(s) = 0* if the point pattern *x* does not contain any points that lie close together (closer than *2r* units apart).

The old parameters *κ* and *γ* of the standard form are related to the new parameters *β* and *η,* of the canonical scale-free form, by

$$\beta = \kappa \gamma^{(-\pi \cdot r^2)} = \kappa / \eta$$

and

$$\eta = \gamma^{(-\pi \cdot r^2)}$$

provided*κ* and *γ* are positive and finite (Baddeley and Turner, 2005).

In the canonical scale-free form, the parameter *η* can take any non-negative value. The value *η=1* again corresponds to a Poisson process, with intensity *β*. If *η<1* then γ>1 for any value of *r*, and the process is clustered. If *η>1* then γ<1, and the process is regular. The value *η=0* corresponds to a hard core process with hard core radius *r* (interaction distance *2r*).

The inhomogeneous area interaction process is similar, except that the contribution of each individual point s_i is a function $\beta(s_i)$ of location, rather than a constant beta.

2.2.2.- Statistical inference

To fit Poisson cluster models, such as the Thomas model, and given a user-specified maximum distance h_k, the model parameters can be estimated by minimizing a 'discrepancy measure' of the empirical K-function $\hat{K}(h)$ and the expected value K(r) (Diggle, 1983):

$$\int_0^{h_k} (\hat{K}(h)^{0.25} - K(r)^{0.25})^2 \, dh$$

And for the inhomogeneous case:

$$\int_0^{h_k} (\hat{K}_{in\,\mathrm{hom}}(h)^{0.25} - K(r)^{0.25})^2 \, dh$$

These are known as minimum contrast procedures.

Because the choice of h_k is arbitrary, Diggle (2003) recommended that h_k should be considerably smaller than the dimensions of the observation plot.

When spatial interactions exist, and a likelihood is obtained in closed form, the model parameters can then be estimated via the maximum pseudo-likelihood method. Goodness-of-fit of the models was evaluated using Akaike's information criterion (AIC) and Monte Carlo simulations. The AIC was used to select the best-fit model for each set of wildfires and was calculated using the following formula:

$$AIC = n \ln(R) + 2k$$

where n is the number of observations, R is the sum of residual squares and k is the number of parameters (Moller and Waagepetersen, 2004; Shen *et al.*, 2009). Number of parameters ranged from 1 to 12 for the various models. The AIC was used because in this study parameters were not estimated using standard maximum likelihood methods. Goodness-of-fit was further evaluated by means of Monte Carlo simulations, which were used to generate 95% confidence intervals of the K(r) for different models.

To fit an area-interaction process we made use of maximum pseudo-likelihood based-methods (Baddeley and Turner, 2005). The pseudo-likelihood estimation approach provides an alternative to likelihood methods when the normalizing constant is no longer to be determined. This pseudo-likelihood function is based on the conditional intensity of a point s_i given realization in a bounded region A. For pairwise interaction processes, the conditional intensity (Papangelou, 1974; Daley and Vere-Jones, 2003) of ϕ at a location s_i may be loosely interpreted as giving the conditional probability that ϕ has a point at x given that the rest of the process coincides with ϕ. For $s_i \,/\!\in \phi$ and $f(\phi) > 0$, the conditional intensity equals:

$$\lambda(s_i;\phi) = \frac{f(\phi \cup \{s_i\})}{f(\phi)}$$

while for $s_i \in \phi$ we have $\lambda(s_i;\phi) = f(\phi)/f(\phi - \{s_i\})$. If η (i.e. pair potential function) is bounded then for any finite point configuration ϕ with $f(\phi)$, the pseudo-likelihood of pairwise interaction processes is defined by (Besag, 1974; Jensen and Moller, 1991),

$$PL(\phi;\theta) = \exp\left(-\int_A \lambda(s_i;\phi)ds\right)\prod_{i=1}^{n} \lambda(\mathrm{s_i};\phi)$$

Usually, this is re-cast in terms of its logarithm. Maximization of the PL function with respect to the parameter set yields the maximum pseudo-likelihood estimators.

2.3.- Analysis of spatial segregation

With the double aim to assess our first objective and to provide additional evidence of the deviation of our data from the null hypothesis of CSR, we follow Diggle *et al.* (2005) and investigate the occurrence of spatial segregation. Spatial segregation occurs if, within a region of interest, some types of points dominate in some sub regions, and this dominance is determined by circumstances other than random.

Data are now represented as a set of multinomial outcomes Y_{it}, $i=1,\ldots,n$, $t=2004, \ldots, 2008$ ($n=3083$ wildfires from 2004 to 2008; 563 in the year 2004; 893 in 2005; 628 in 2006; 578 in 2007; and 421 in the year 2008), where, for each of $k=1,..,m$ ($m=4$, 1= natural causes; 2= negligence and accidents; 3= intentional or arson; 4= unknown causes and revived), and year t, the outcome $Y_{it}=k$ denotes the occurrence of a wildfire for the cause k at the location $x_i$the year t, and the corresponding multinomial cell probabilities are the cause-specific probabilities $p_k(x)$,

$$p_k(x) = \frac{\lambda_k(x)}{\sum_{j=1}^{m} \lambda_j(x)}$$

where$\lambda_k(x)$ denotes the intensity function of the independent Poisson process corresponding to cause k.

The cause-specific probabilities, $p_k(x)$, were estimated through a multivariate adaptation of a kernel smoothing. In particular, the cause-specific probability surfaces were estimated by means of a kernel regression estimator,

$$\hat{p}(x) = \sum_{i=1}^{n} w_{ik}(x) I(Y_i = k)$$

where*I(.)* was an indicator function; and, for each *k=1,2,..,m*,

$$w_{ik}(x) = \frac{w_k(x - x_i)}{\sum_{j=1}^{n} w_k(x - x_j)}$$

$w_k(.)$ was the kernel function with bandwith $h_k>0$,

$$w_k(x) = \frac{e^{\left(-\|x\|^2/2\right)}}{h_k^2}$$

where we used the Gaussian kernel as the standardised form of the kernel function in the numerator; and $\|x\|$ denotes the Euclidean distance of the point *x* to the origin.

Estimation of the cause-specific probabilities was done by means of cross-validated log-likelihood (Diggle *et al.*, 2005). For the testing of both, the null hypotheses of no spatial variation in the probability surfaces between wildfire causes (i.e. no spatial segregation); and of no change over time of the cause-specific probability surfaces (i.e. no temporal changes in spatial segregation) we used Monte Carlo sampling (Diggle et al., 2005).

The analyses were carried out using the R freeware statistical package (versions 2.12.1 and 2.14.2) (R Development Core Team, 2010);and the *Spatstat* (Baddeley and Turner, 2005); *splancs* (Rowlingson and Diggle, 1993) and *spatialkernel* (Zheng *et al.*, 2012) packages. The free environment *gvSIG (version 1.10)* was used for representing maps.

3.- Results

Table 1 and Figure 1 show the distribution of the 3,083 wildfires occurring in Catalonia by year (from 2004 to 2008) and cause (natural causes; negligence and accidents; intentional or arson; and unknown causes and revived).

The category of negligence and accidents was the most frequent cause (54.91%), followed by intentional or arson (20.27%), unknown (12.88%) and natural causes (11.94%). Note also that after a dramatic increase in the number of wildfires from 2004 to 2005 (58.61%), there was then a decrease (29.68% from 2005 to 2006; 7.62% from 2006 to 2007; and 27.16% from 2007 to 2008). In fact, 2008 was the year with the lowest number of wildfires during the study period. Note that the behaviour of each of the causes by year was more or less the same, with a slight difference in the case of unknown and revived, where the year with the lowest number of wildfires was 2004. The spatial distribution of the wildfires was very similar for all causes, perhaps with the exception of natural causes, with very few fires in coastal and urban areas (which are located mainly on the coasts). Note also, that intentional and unknown causes behaved more like each other than like other causes.

Although it would be theoretically possible that the use of topographic variables would produce high collinearity, this problem did not occur in practice. Correlation between hill shade and aspect was equal to 0.402 ($p<0.001$); between hill shade and slope -0.360 ($p<0.001$); and -0.025 ($p=0.158$), between aspect and slope. Nevertheless, we repeated all analyses using: i) slope and aspect separately (without hill shade); and ii) hill shade alone (without slope and aspect). However, the results vary very little. In neither model parameter estimators suffer a significant change and, in any case, vary the parameters of statistical significance.

To model the behaviour of the wildfires we have used three different models for each year and cause. Initially, we individually entered each covariate (slope, aspect, hill shade and land use, see Figure 2) into each of the models. However, due to the large number of resulting models for all possible combinations, we decided to focus the study solely on the four covariates together, as this already covered individual results and gave other general information.

Discarding the homogeneous models, since intensity is always clearly variable in the case of modelling wildfire, we started with the simplest model, the inhomogeneous Poisson process. We followed this with a cluster process model, specifically the Thomas process, and finished with the area-interaction point process. In total sixty models were fitted to the data. Among them, and according to the fit of the model to the observed spatial distribution of the wildfires, we selected the twenty best models (see Table 2). It is interesting to note that the Thomas process was only well modeled for natural causes for practically all years (except 2007). Note that the area-interaction model shows a clear improvement over the other two models. It properly adjusted most cases including, unlike the other two, the cause intentional or arson. The results of estimating the twenty best models are shown in Table 3 and the simulated envelopes of the fit are shown in Figure 3.

Models were selected using a second filtration taking into account the AIC. As is known, this criterion measures the relative goodness-of-fit of the model in question and is therefore an effective tool for comparing models. The lower the AIC, the better the model (in terms of goodness-of-fit). AIC can be computed in terms of close likelihoods in which the number of parameters playing a role in the specific model are considered. For some cluster models, such as the Thomas one, the close likelihood can not be computed and thus AIC could also be calculated in an approximate way, making the comparison with Poisson or Area-Interaction models unfair. In this sense, we only used AIC criteria for these latter two models (see Table 4). In these cases, there was generally little difference between the two models analysed, except for the cause negligence and accidents in the year 2007, where the Poisson model had the lowest AIC.

The analysis of spatial segregation is shown in Table 5. All four cause-specific probabilities show wide spatial variation for all years. Using 999 simulated random relabellings of the wildfire causes among all case in the Monte Carlo tests for both, spatial segregation and the temporal changes of the spatial segregation, we rejected the null hypotheses each one of the years and causes, respectively.

Note, that there was a great variability in the estimated effects of the variables on the intensities for all causes, years and spatial point process models finally fitted (see Table 4). As a result, conditional intensities should vary accordingly.

Finally, we computed the conditional intensity of all fitted spatial point process models in Table 2, evaluated at each spatial location used for model fitting. Our interest was to use these intensities to build maps of wildfire risks, by year and cause of ignition (Figure 4), in order to provide a tool for prevention and management of vulnerability levels. Here we show only those maps corresponding to the best models (in terms of goodness-of-fit), i.e. area-interaction with two exceptions, natural causes in 2006, where the only model was the Thomas process (see Table 2), and negligence and accidents in 2007, built using the Poisson model (see Table 4) (results not shown can be requested from the authors).

With some exceptions (unknown causes and revived in 2004 and 2005, and natural causes in 2005 and 2007) the risks for each cause varied over the years. Note that, generally speaking, risks for natural causes in particular and, to a lesser extent, unknown causes and revived, were not very high, whereas for negligence and accidents (2004 and 2007) and intentional or arson (2006 and 2008) risks reached figures close to the maximum in some areas, mainly coastal and most urbanised areas.

Wildfires attributed to natural causes are where we find more variability: each year is completely different from the rest. This is due to the fact that these fires are basically caused by dry lightning storms (no rain); this phenomenon can occur anywhere: mountains, sea, etc. With anthropogenic causes, on the other hand, we see a clear pattern in the coastal areas (especially Barcelona and the Costa Brava), where fires occur in July and August, when more tourists can be found in these places.

4.- Discussion and Conclusions

This study has three main findings. Firstly, the extent of clustering in wildfires differed across marks, i.e. years and cause of wildfire ignitions. Secondly, covariates such as land use, slope, aspect and hill shade influenced trends in the intensity of wildfire locations. Finally, we have built maps of wildfire risks, by year and cause of ignition, from the estimated models. These could prove very useful as tools in preventing and managing vulnerability levels.

As regards the first finding, we conclude that, despite the variability found among marks, especially over time, the model that best fits the behaviour of fires for most years and causes is the area-interaction point process model. This result is interesting because it will allow us to expand the study to spatio-temporalmodelling and will lead to improved prediction of fire risk.

Regarding the second finding, the analysis should be completed using other covariates such as fuel (present at the location of fire), flammability, proximity of the wildfire to roads and towns, and economic factors such as land prices. On the other hand, it would also be interesting to introduce some meteorological covariates such as temperature, precipitation and wind.

Note that we have restricted our attention to inhomogeneous spatial models where we have fixed the temporal scale and modelled only the spatial component. A necessary and further improvement of our modelling efforts might incorporate time into the model itself, considering thus spatio-temporal point process models. One good thing about this latter approach is that we could model and evaluate the corresponding spatio-temporal interaction, something that we have not considered in the present approach. We note that some approaches have already been considered in this respect, but they consider independent spatial replication in time, and this is not realistic. In the context of spatio-temporalmodelling, one useful approach is to try and model the spatio-temporal intensity function as an additive or multiplicative form of the spatial and temporal intensities, and then adding a spatio-temporal residual component (for instance Diggle et al., 2005b). This would probably provide more insight into the problem of modelling wildfires.

With respect to the third finding, we constructed maps that could: i) help authorities to develop more reliable plans for forest fire prevention; ii) predict areas of forest fire risk more efficiently; iii) design awareness campaigns more efficiently; and iv) reduce the budget designed of fire management. In fact, if we understand prevention as prediction of scenarios on where and when forest fires could be stronger in critical moments, our maps could provide the necessary land-use planning measures, both in terms of preventive management practices and stakeholders' responsibility in managing fires; and identify priority actions for each component of the fire management system throughout the year.

As we said, hill shade was defined as a function of slope and aspect (along with angle of illumination). In fact, the correlation between hill shade-aspect and hill shade-slope were actually quite high (about 0.4 in each case). However, it seems that there is some property of the stochastic models which makes them insensitive to inter-correlation[1]. In fact, when we repeated all analyses using: i) slope and aspect separately (without hill shade); and ii) hill shade alone (without slope and aspect) the results vary very little.

The methodology used has allowed us to quantify and assess possible spatial relationships between the distribution of risk of ignition and causes. These results may be useful in fire management decision-making and planning. We believe the methods shown in this paper may contribute to the prevention and management of wildfires, which are not random in space or time, as we have shown here.

[1]We thank an anonymous reviewer for pointing out this point.

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

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

We would like to thank the Forest Fire Prevention Service (*Servei de Prevenció d'Incendis Forestals)*, Government of Catalonia (*Generalitat de Catalunya*) for providing wildfire data and to the Environment Department of the Government of Catalonia for the access to the digital map databases. We appreciate the comments of the attendees at the '1st Conference on Spatial Statistics 2011', on March, 23-25, 2011, at the University of Twente, Enschede, The Netherlands, where a preliminary version of this work was presented. We are indebted to Pau Aragó, Universitat Jaume I, Castelló, because, among other things, his continuous help with GIS processing. Finally, we would like to think to the editor and two anonymous referees for their valuable comments that, without doubt, improve the paper.

References

Baddeley AJ, van Lieshout MNM. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* 1995; 47:601-619.

Baddeley AJ, Turner R. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 2005; 12:1-42.

Badia A, Saurí D, Cerdan R, Llurdés KC. Causality and management of forest fires in Mediterranean environments: an example from Catalonia. *Global Environmental Change Part B: Environmental Hazards* 2002; 4(1): 23-32

Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion).*Journal of the Royal Statistical Society, Series B* 1974; 36:192-236.

Butry DT, Prestemon JP. Spatio-temporal wildland arson crime functions. American Agricultural Economics Association Annual Meeting, 26–29 July 2005, Providence, Rhode Island, USA (available in: http://ageconsearch.umn.edu/handle/19197, accessed on March 7, 2011).

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

Cressie NAC *Statistics for Spatial Data (Revised ed.)*. New York: Wiley, 1993.

Daley D, Vere-Jones D.*An Introduction to the Theory of Point Processes*. Vol. I. 2. New York, Springer, 2003.

Díaz-Delgado R, Pons X. 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* 2001; 147(1):67-74.

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

Diggle, PJ. *Statistical Analysis of Spatial Point Patterns*. London: Academic Press, 1983.

Diggle PJ. *Statistical Analysis of Spatial Point Patter*ns.2nd Edition. London: Arnold, 2003.

Diggle P, Zheng P, Durr P. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Journal of the Royal Statistical Society, Series C, Applied Statistics* 2005; 54(3):645-658.

Diggle P, Rowlingson B, Su TL. Point process methodology for on-line spatio-temporal disease surveillance.*Environmetrics* 2005b; 16:423-434.

European Environment Agency.*Technical report No 17/2007. CLC2006 technical guidelines*. Luxembourg: Office for Official. Publications of the European Communities, 2007.

Flannigan MD, Amiro BD, Logan KA, Stocks BJ, Wotton BM. Forest fires and climate change in the 21st century. *Mitigation and Adaptation Strategies for Global Change* 2006; 11(4):847-859.

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

Gralewicz NJ, Nelson TA, Wulder MA.Spatial and temporal patterns of wildfire ignitions in Canada from 1980 to 2006.*International Journal of Wildland Fire* 2012 (in press, http://dx.doi.org/10.1071/WF10095)

Heymann Y, Steenmans CH, Croissille G, Bossard M. *CORINE Land Cover. Technical Guide*. Luxembourg: Office for Official. Publications of the European Communities, 1994.

Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical analysis and modelling of spatial point patterns*. New York: Wiley, 2008.

Jensen JL, Moller J. Pseudo-likelihood for exponential family models of spatial point processes. *Annals of Applied Probability* 1991; 3:445-461.

Liu Z, Yang J, Chang Y, Weisberg PJ, He HS.Spatial patterns and drivers of fire occurrence and its future trend under climate change in a boreal forest of Northeast China. *Global Change Biology* 2012 (in press, DOI: 10.1111/j.1365-2486.2012.02649.x)

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

McRae RHD. Prediction of areas prone to lightning ignition.*International Journal of Wildland Fire* 1992; 2:123-130.

Millington, JDA. Wildfire risk mapping: considering environmental change in space and time. *Journal of Mediterranean Ecology* 2005; 6(1):33-44.

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

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

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

Moller J, Waagepetersen RP.*Statistical inference and simulation for spatial point processes*. New York: CRC Press, 2004.

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

National Interagency Fire Centre.*The Science of Windland* Fire (available in: http://www.nifc.gov/, accessed on February 14, 2011).

National Wildfire Coordinating Group.*Wildfire Prevention Strategies* March 1998 (available in: http://www.nwcg.gov/pms/docs/wfprevnttrat.pdf, accessed on February 14, 2011).

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

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

Papangelou F. The conditional intensity of general point processes and an application to line processes. Z. Wahrscheinlichkeitsth 1974; 28:207-226.

Pyne SJ, Andrews PL, Laven RD. *Introduction to Wildland fire. 2nd ed*. John Wiley and Sons, 1996.

R Development Core Team.*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/, 2010.

Ripley BD. The second-order analysis of stationary point processes. *Journal of Applied Probability* 1976; 13:256-266.

Ripley BD. Modelling spatial patterns.*Journal of the Royal Statistical Society, Series B* 1977; 39:172-212.

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

Rowlingson B, Diggle P. Splancs: spatial point pattern analysis code in S-Plus. *Computers and Geosciences* 1993; 19: 627-655; (the original sources can be accessed at: http://www.maths.lancs.ac.uk/~rowlings/Splancs/.

Scott A. The pre-Quaternary history of fire.*Palaeogeography, Palaeoclimatology, Palaeoecology* 2000; 164(1-4):281-329.

Shen G, Mingjian Y, Xin-Sheng H, Xiangcheng M, Haibao R, I-Fang S, Ma K. Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology* 2009; 90:3033-3041

Widom B, Rowlinson JS. New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics* 1970; 52:1670–1684.

William A. Thompson, Ilan Vertinsky, Hans Schreier, Bruce A. Blackwell. Using forest fire hazard modelling in multiple use forest management planning.*Forest Ecology and Management* 2000; *134(1-3)*:163-176.

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

Zheng P, Diggle P, Gómez-Rubio V. *Package 'spatialkernel'*, version 0.4-10, February 15, 2012 [available in: http://cran.r-project.org/web/packages/spatialkernel/spatialkernel.pdf, accessed on March 19, 2012]

Table 1.- Distribution of wildfires occurring in Catalonia by year and cause

Number of fires	Cause 1 (Naturals)	Cause 2 (Negligence)	Cause 3 (Intentional)	Cause 4 (Unknown)	All causes
2004	65	336	103	59	563
2005	115	525	151	102	893
2006	98	297	155	78	628
2007	52	308	122	96	578
2008	38	227	94	62	421
All years	368	1693	625	397	3083

Table 2.- The twenty models with best fit to the observed spatial distribution of the wildfires

Year	2004				2005				2006				2007				2008			
Cause	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
Poisson					X								X	X			X			
Thomas	X				X				X								X			
Area-Interaction	X	X	X	X	X			X		X	X		X	X			X		X	

Table 3.- Results of estimating the twenty models with best fit to the observed spatial distribution of the wildfires

Thomas							
2004	β_0	(Aspect)β_1	(Hill shade)β_2	(Slope)β_3	(Land Use)β_4	κ	σ
Cause 1	4.2783	-0.0004	0.0026	0.0235	0.0735	53.3747	0.0432
2005							
Cause 1	5.4771	-4.848e-06	-0.0011	0.0119	0.0463	18.4070	0.0549
2006							
Cause 1	4.7536	0.0001	0.0013	0.0025	0.0879	60.5309	0.0338
2008							
Cause 1	4.7579	-0.0009	-0.0003	0.0192	0.0411	45.2217	0.0543

Areal-interaction							
2004	β_0	(Aspect)β_1	(Hill shade)β_2	(Slope)β_3	(Land Use)β_4	η	Interaction
Cause 1	4.2446	-0.0003	0.0024	0.0145	0.0571	31.563	3.4520
Cause 2	7.2342	-0.0014	-0.0002	-0.0068	-0.0135	7.0907	1.9588
Cause 3	5.2597	-0.0024	0.0009	0.0251	0.0018	254.80	5.5405
Cause 4	6.1766	-0.0028	-0.0015	0.0011	0.0128	4.0229	1.3920
2005							
Cause 1	5.5155	-0.0002	-0.0011	0.0249	0.0456	0.8201	-0.1983
Cause 4	5.6711	-0.0006	0.0008	0.0073	-0.0671	70.431	4.2546
2006							
Cause 2	7.2418	-0.0023	-0.0006	0.0168	0.0025	8.3818	2.1261
Cause 3	5.5181	-0.0011	-0.0001	-0.0002	-0.0331	313.024	5.7463
2007							
Cause 1	4.8859	-0.0007	-0.0005	0.0330	-0.0041	48.078	3.8728
Cause 2	6.9511	-0.0018	0.0009	0.0010	0.0230	6.0967	1.8077
2008							
Cause 1	4.6428	-0.0008	-0.0007	0.0188	0.0336	32.852	3.4920
Cause 3	5.7116	-0.0027	0.0007	0.0038	-0.0785	201.55	5.3060

Table 4.- AIC of best Poisson and area-interaction models (common years and causes)

	AIC	
	Poisson	**Area-Interaction**
2005 cause 1	-1023.863	-1025.464
2007 cause 1	-386.7365	-386.2792
2007 cause 2	-3512.036	-735.2788
2008 cause 1	-255.3777	-254.4718

Table 5.- Analysis of the spatial segregation

	2004	2005	2006	2007	2008	Monte Carlo test for temporal changes of spatial segregation
Cause-specific probabilities[1]						
Natural causes	0.44982-5.353e-07	0.44781-1.262e-07	0.27130-0.00245	0.42920-0.00523	0.34484-0.05102	0.034
Negligence and accidents	0.52385-2.719e-05	0.91912-0.01334	0.89029-0.00814	0.26738-0.00135	0.37881-0.00833	0.002
Intentional or arson	0.88786-0.31920	0.69422-0.00056	0.81789-0.01535	0.43960-0.00011	0.81901-0.17470	0.021
Unknown causes and revived	0.47607-9.673e-06	0.46596-0.00674	0.55427-0.00300	0.88825-0.35331	0.39663-0.01364	0.044
Monte Carlo test for spatial segregation	0.002	0.012	0.003	0.002	0.020	

[1] Maximum-minimum

Figure 1.- Distribution of wildfires occurring in Catalonia by year and cause (natural; negligence or accident; intentional or arson; unknown causes and revived)

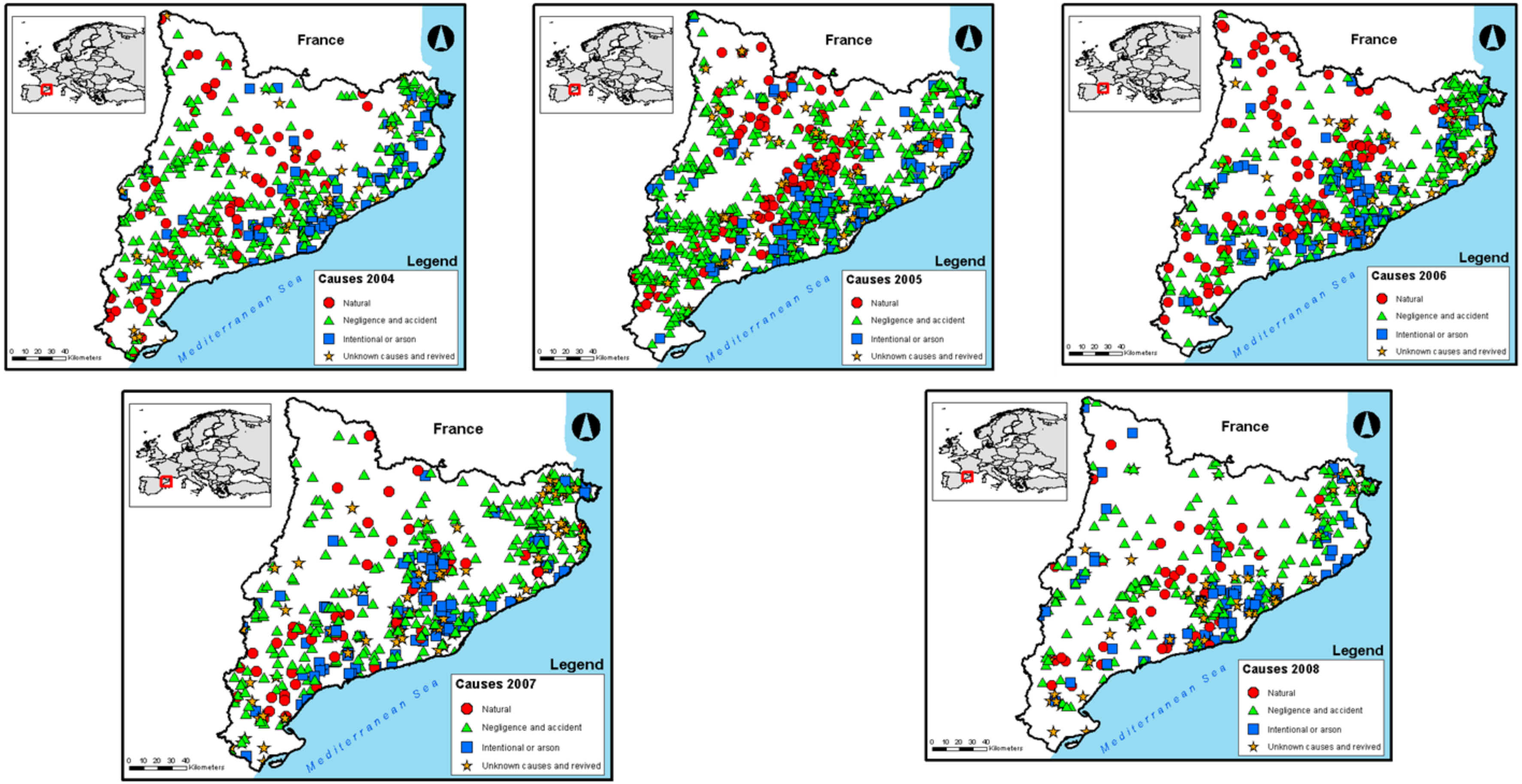
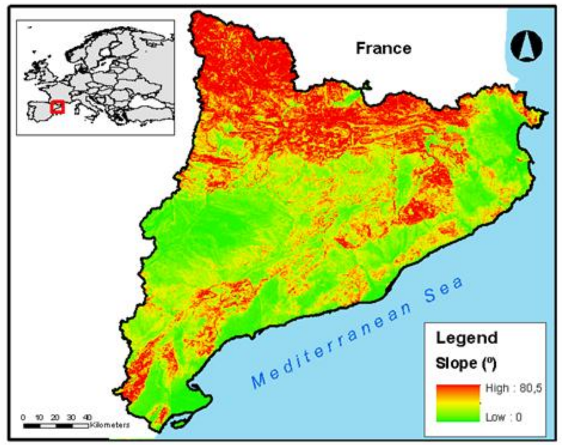
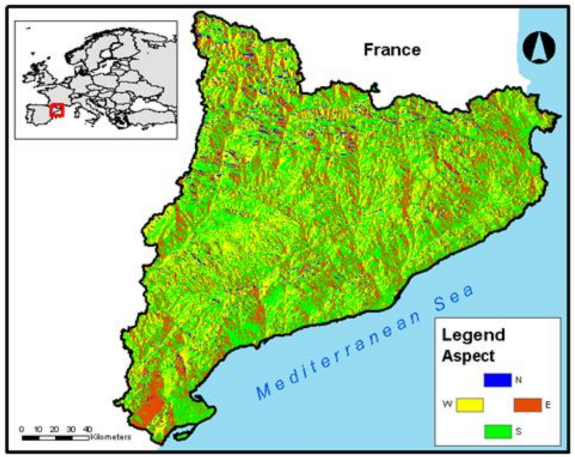


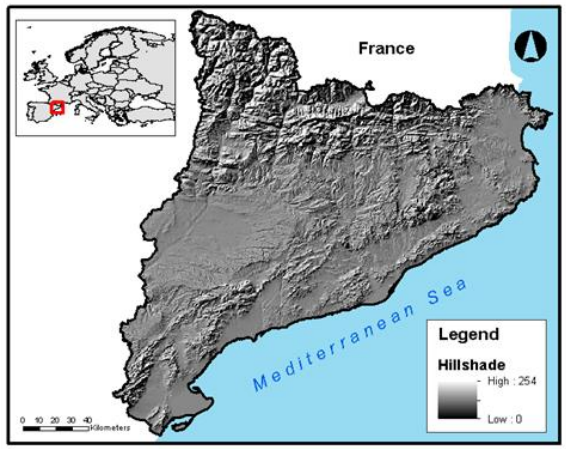
Figure 2.- Spatial covariates, both continuous (slope, aspect and hill shade) and categorical (land use)



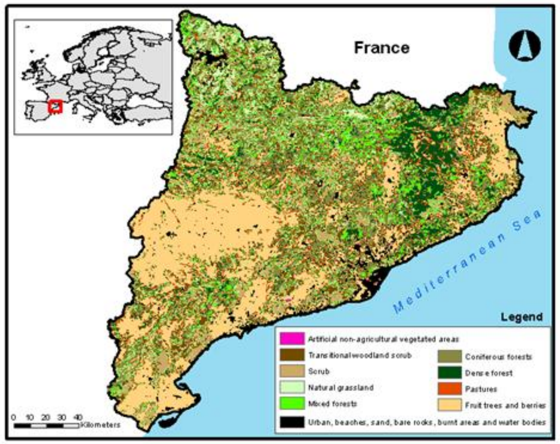
Slope



Aspect



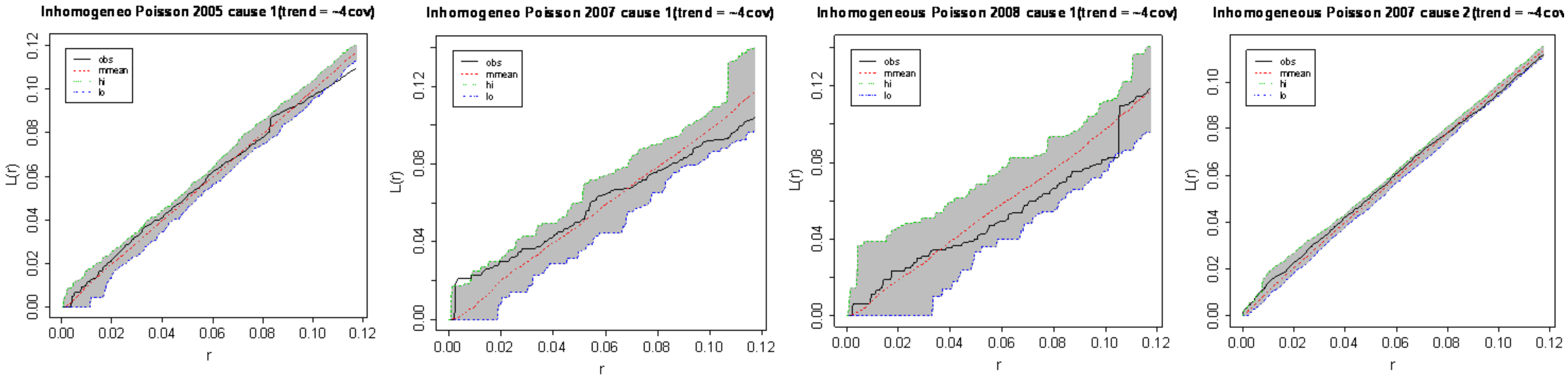
Hill shade



Land use

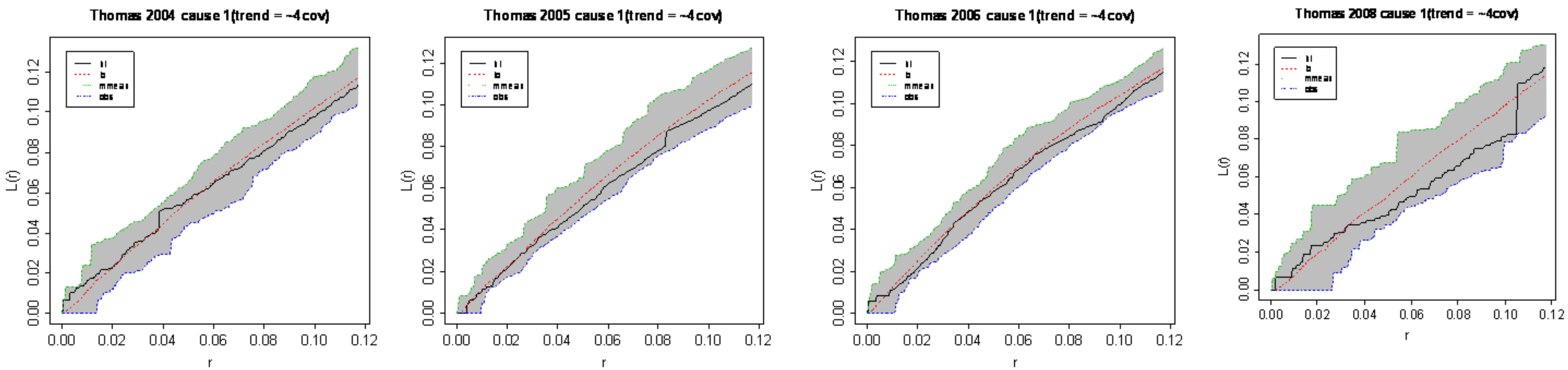
Figure 3a.- Simulated envelopes of the twenty models with best fit to the observed spatial distribution of the wildfires

Inhomogeneous Poisson



2005 – Natural causes | 2007 – Natural causes | 2008 – Natural causes | 2007 – Negligence and accidents

Thomas



2004 – Natural causes | 2005 – Natural causes | 2006 – Natural causes | 2008 – Natural causes

Figure 3b.- Simulated envelopes of the twenty models with best fit to the observed spatial distribution of the wildfires

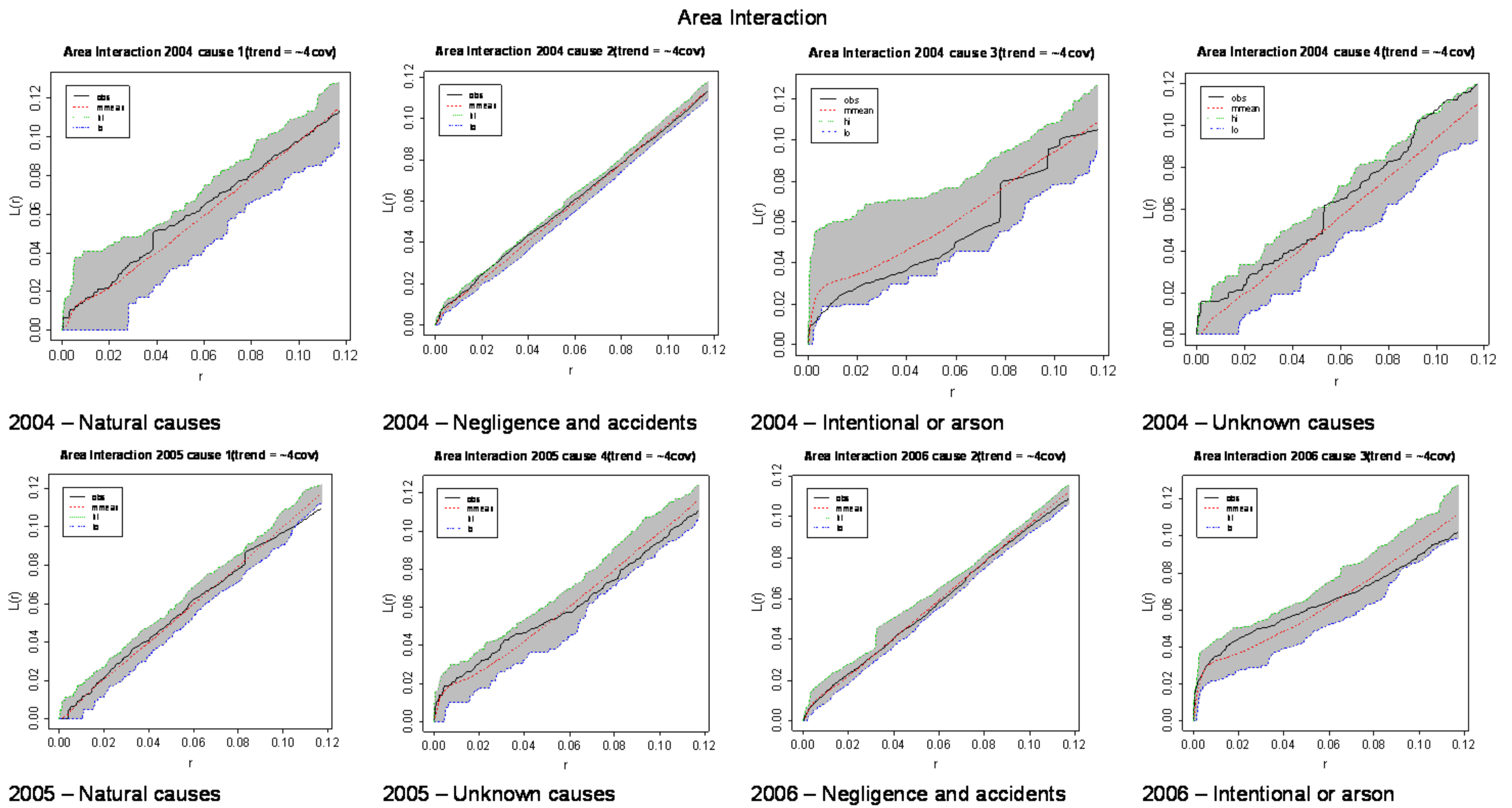
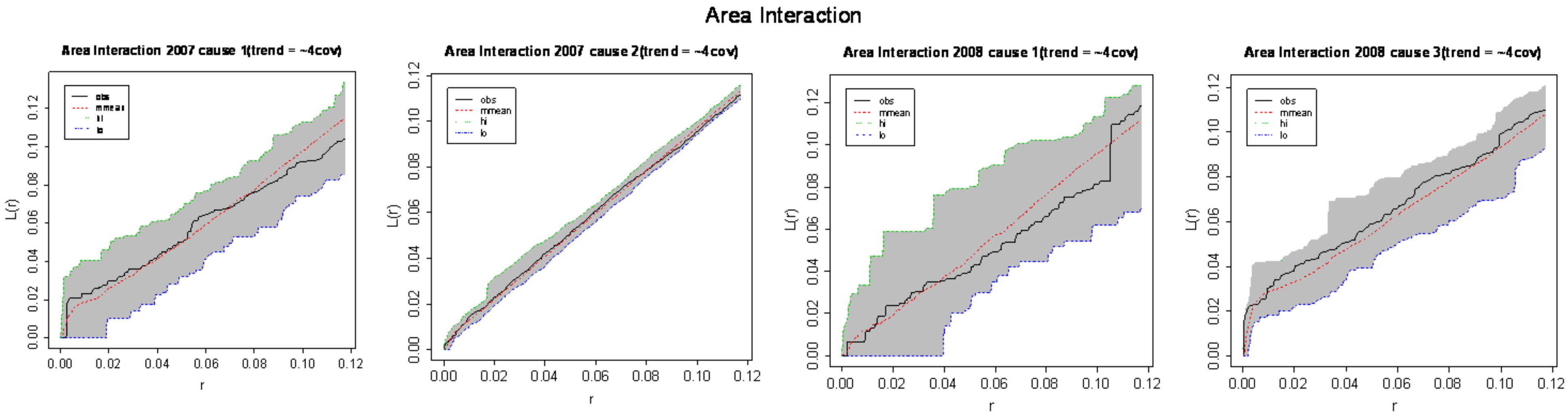


Figure 3c.- Simulated envelopes of the twenty models with best fit to the observed spatial distribution of the wildfires



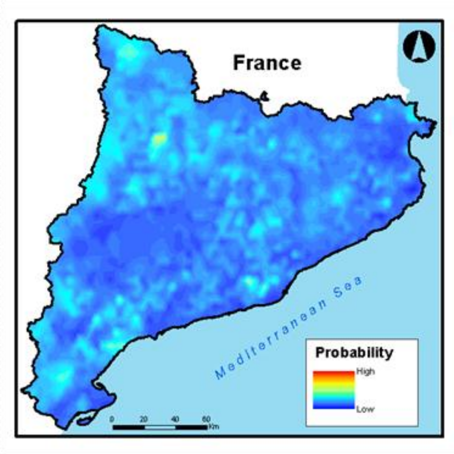
2005 – Natural causes

2005 – Unknown causes

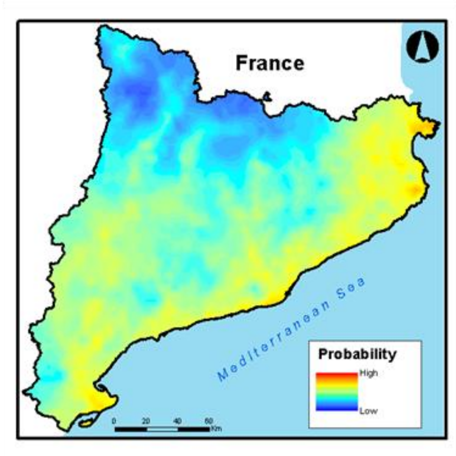
2006 – Negligence and accidents

2006 – Intentional or arson

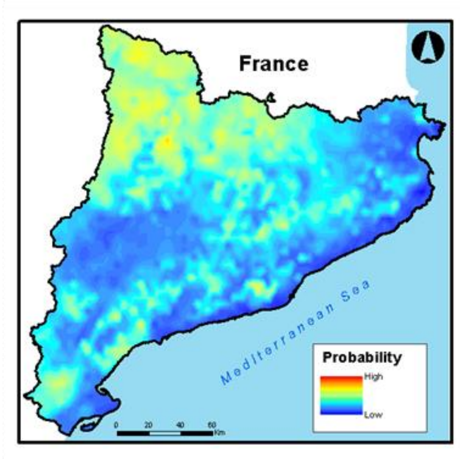
Figure 4a.- Conditional intensity of the fitted spatial point process models. Natural causes



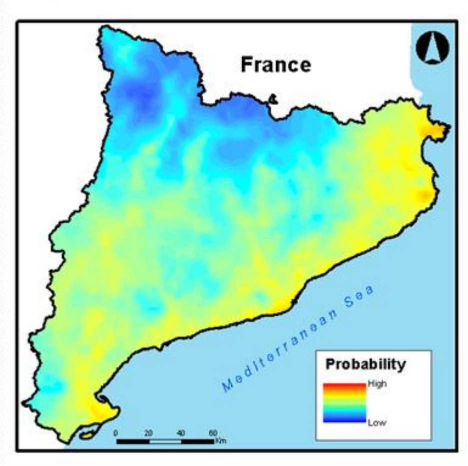
2004 (Area-Interaction)



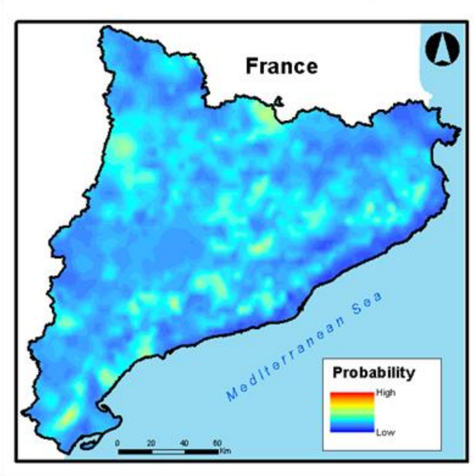
2005 (Area-Interaction)



2006 (Thomas)



2007 (Area-Interaction)



2008 (Area-Interaction)

Figure 4b.- Conditional intensity of the fitted spatial point process models. Negligence and accidents

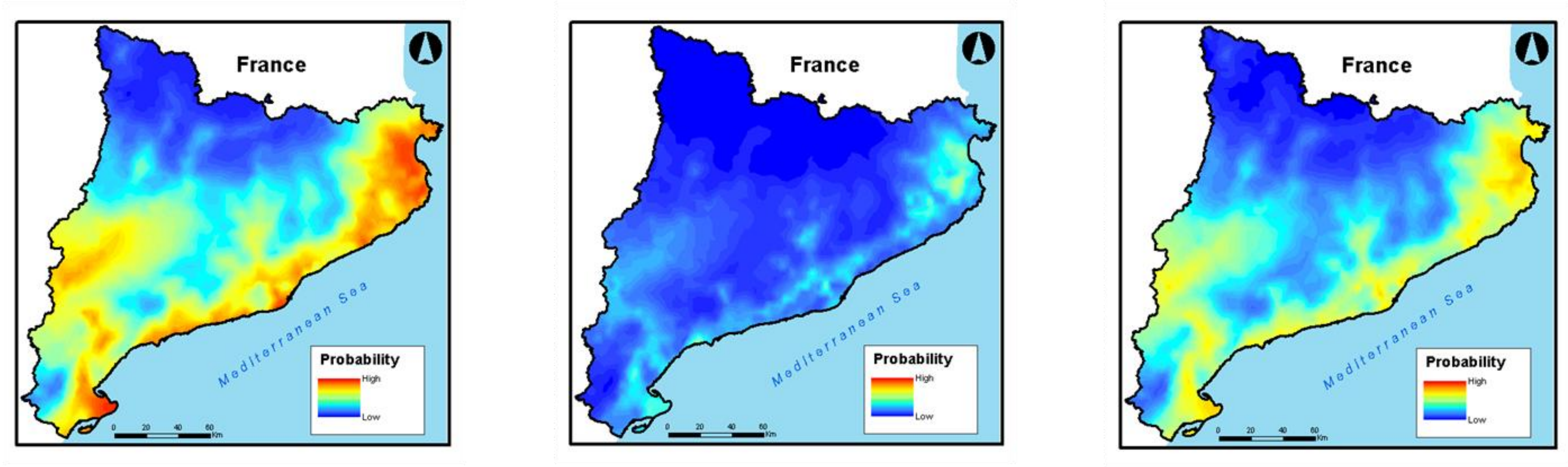
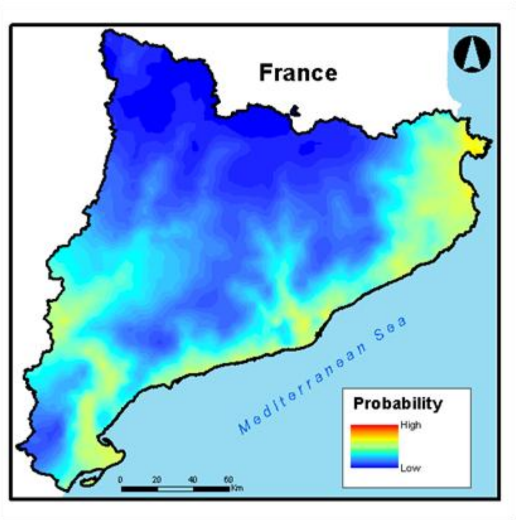
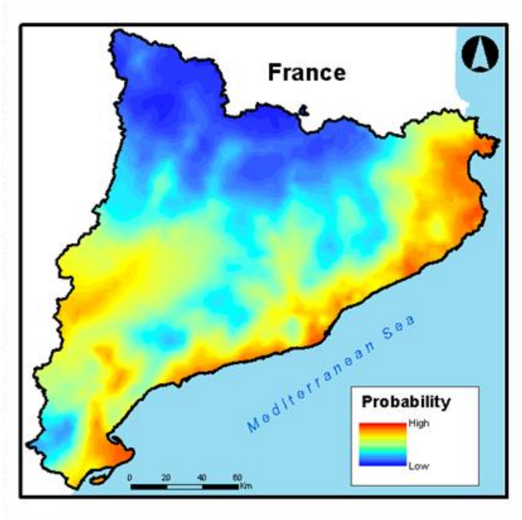


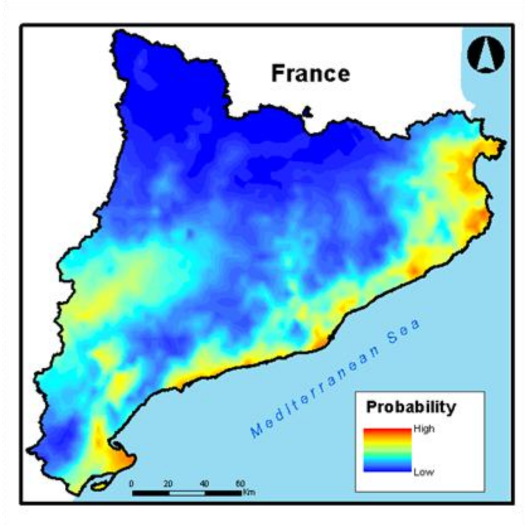
Figure 4c.- Conditional intensity of the fitted spatial point process models. Intentional or arson



2004 (Area-Interaction)

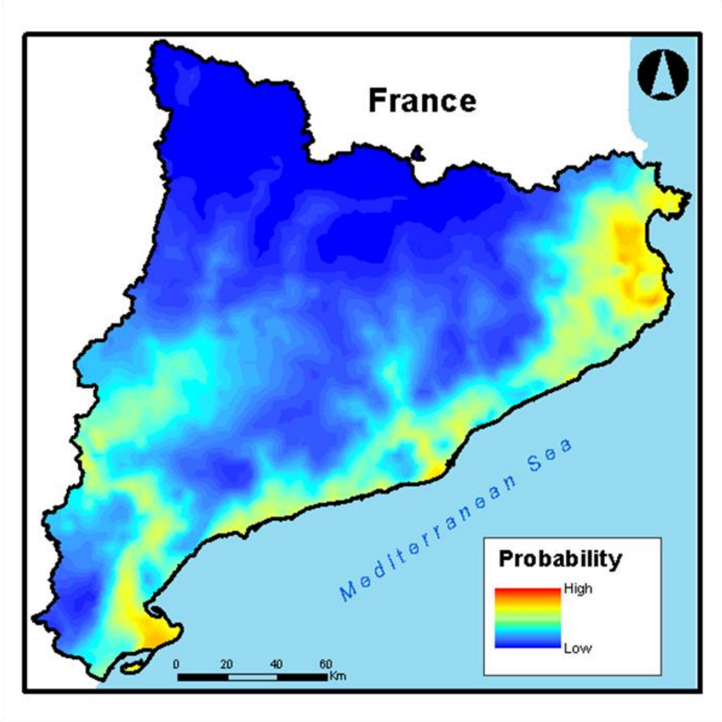


2006 (Area-Interaction)

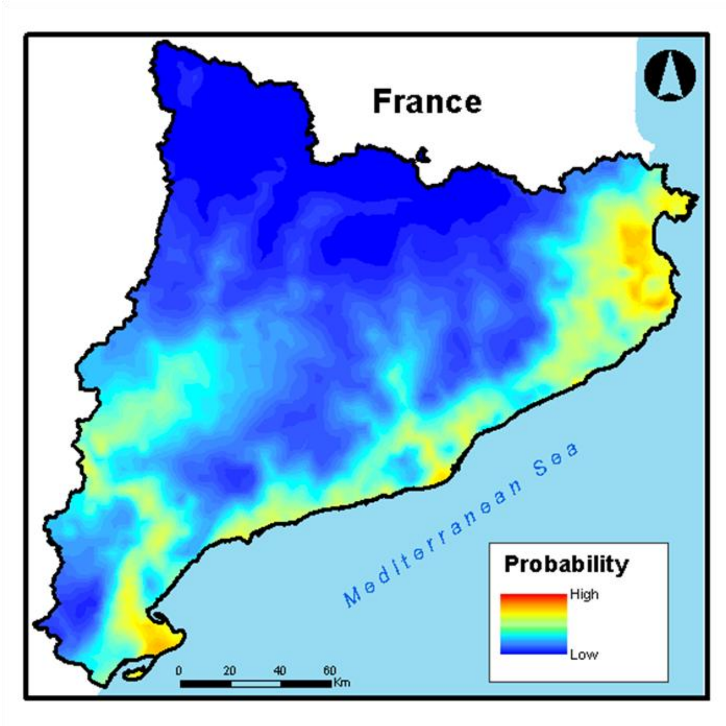


2008 (Area-Interaction)

Figure 4d.- Conditional intensity of the fitted spatial point process models. Unknown causes and revived



2004 (Area-Interaction)



2005 (Area-Interaction)

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

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

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

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

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

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

where Po denotes the Poisson density, p is a hyperparameter given by

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

and θ 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 cells_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, $\tau_t$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

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

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,

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

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 $x = (x_1, \dots, x_n)$ 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 21st 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].

Table 1: Evolution of wildfires by year and cause

Year	Natural causes	Negligence and accidents	Intentional or arson	Unknown causes and rekindled
1994	69	508	374	222
1995	26	104	182	43
1996	12	70	73	17
1997	5	101	178	16
1998	7	182	214	69
1999	7	152	168	45
2000	8	106	96	28
2001	6	128	113	27
2002	8	91	36	18
2003	13	174	110	39
2004	6	125	42	24
2005	11	237	71	69
2006	5	109	63	31
2007	6	78	51	37
2008	2	69	20	10
Sum=Nº fires	191	2234	1791	695
% OF FIRES	3.89%	45.49%	36.47%	14.15%
TOTAL HA BURNED	6,250.13	69,543.03	26,197.16	24,999.12

Table 2: Covariates and their source ordered by their importance on fire hazar generation

Covariates	Source
Land uses	Territory and Sustainability Department (Catalonia Government)
Slope	Own construction
Isolation	Own construction
Aspect	Own construction
Altitude	Cartographic Institute of Catalonia
Distance to antrophic areas	Own construction

Table 3.- Results after analyzing the wildfire data using the three different kinds of LGCP

	Natural Causes			Negligence & accidents			Intentional or arson			Unknown causes and rekindled		
	Poisson	**ZIP0**	**ZIP1**	**Poisson**	**ZIP0**	**ZIP1**	**Poisson**	**ZIP0**	**ZIP1**	**Poisson**	**ZIP0**	**ZIP1**
DIC												
tmin	1163.81	---	---	11124.79	**9358.76**	10168.72	7664.59	---	**5031.78**	4442.66	---	---
tmax	**1155.09**	---	---	11132.21	9369.10	10250.69	7723.35	---	5032.86	**4437.04**	---	---
CPO												
tmin	0.0698	0.0443	---	0.8685	0.5666	0.7381	0.6299	0.2145	0.2845	0.3637	0.1183	---
tmax	0.0680	---	---	0.8692	0.5765	0.7433	0.6335	0.2150	0.2950	0.3643	0.1474	---
nEFF												
tmin	215.71	---	---	638.59	329.32	429.52	560.03	---	293.84	438.47	---	---
tmax	218.35	---	---	644.54	329.33	409.08	544.59	---	287.65	448.82	---	---

In bold, the best model

DIC: Deviance Information Criterion, CPO: Conditional Predictive Ordinate; nEFF: Effective degrees of freedom; tmin: Minimum Temperature; tmax: Maximum Temperature

Table 4: Wildfires distinguished by causes and percentage of buffer

	Natural causes	Negligence and accidents	Intentional or arson	Unknown causes and rekindled
Number of wildfires	191	2234	1791	695
Number of buffers	128	1,035	367	284
% of buffers with no wildfires	94.40%	41.60%	78.20%	85.80%
% of buffers with one wildfires	4.2%	41.9%	10.5%	8.2%
% of buffers with two wildfires	0.8%	10.5%	4.2%	3.2%
% of buffers with three wildfires	0.5%	2.8%	1.8%	1.1%
% of buffers with more than three wildfires	0.1%	1.1%	1.2%	0.5%

Table 5.- Natural causes – Poisson

	RR (95% credible interval)
Slope (<3%)	
Q2 (3%-5%)	1.7707 (0.8437, 3.4655)
Q3 (5%-8%)	1.5483 (0.7398, 3.0778)
Q4 (8%-13%)	1.4331 (0.6626, 2.9337)
Q5 (13%-66%)	1.5835 (0.6746, 3.4364)
Aspect_Orientac (<84º)	
Q2 (84º-147º)	1.1002 (0.5397, 2.0285)
Q3 (147º-202º)	0.5974 (0.2766, 1.1303)
Q4 (202º-264º)	0.7311 (0.3733, 1.2825)
Q5 (264º-360º)	2.0580 (0.9502, 3.9926)
Hill shade(24º-159º)	
Q2 (159º-172º)	***0.4679 (0.2307, 0.8395)***
Q3 (172º-180º)	***0.2796 (0.0931, 0.6444)***
Q4 (180º-189º)	0.7195 (0.2561, 1.6436)
Q5 (189º-251º)	***0.3001 (0.0997, 0.6969)***
Elevation (<90m)	
Q2 (90m-179m)	1.8926 (0.3653, 6.4064)
Q3 (179m-318m)	6.2610 (1.2971, 21.0708)
Q4 (318m-521m)	13.3041 (2.8918, 44.2008)
Q5 (521m-2532m)	25.5195 (5.4483, 85.6278)
Land use (urban, beaches, sand, bare rocks, burnt areas, and water bodies)	
Coniferous forests	0.4961 (0.1446, 1.3340)
Dense forests	0.7850 (0.1136, 2.6617)
Fruit trees and berries	0.7069 (0.2130, 1.8805)
Artificial non-agricultural vegetated areas	0.5155 (0.0200, 2.1723)
Transitional woodland scrub	0.9243 (0.3079, 2.3375)
Natural grassland	0.6216 (0.0044, 3.3089)
Mixed forests	0.5238 (0.0913, 1.6991)
Distance to urban areas, roads and railways (<60m)	
Q2 (60m-169.7056m)	0.9293 (0.4571, 1.6840)
Q3 (169.7056m-361.2478m)	1.1888 (0.6322, 2.0743)
Q4 (361.2478m-724.9828m)	2.0274 (1.1744, 3.3803)
Q5 (724.9828m-10494.5557m)	6.8247 (4.0303, 11.4032)
Minimum temperature	
Lag 1	1.2205 (0.9044, 1.6163)
Lag 2	0.9591 (0.6760, 1.3507)
Lag 3	1.2682 (0.9240, 1.6970)
Lag 4	0.8333 (0.6237, 1.1207)
Lag 5	0.8427 (0.6258, 1.1143)
Lag 6	1.1333 (0.8856, 1.4546)
Lag 7	1.1913 (0.8799, 1.5877)
Maximum temperature	
Lag 1	***0.7073 (0.5063, 0.9804)***
Lag 2	0.9475 (0.6193, 1.4276)
Lag 3	1.5840 (0.9929, 2.4295)
Lag 4	1.7570 (1.0365, 2.8421)
Lag 5	0.9637 (0.6161, 1.4592)
Lag 6	1.0310 (0.7458, 1.4141)
Lag 7	0.7573 (0.5256, 1.0831)

Random effects	Mean (standard deviation)
Heterogeneity	0.0101 (0.0058)
Temporal	0.0103 (0.0062)
Spatial	1.3165 (0.2019)
Range (mean – 95% credible interval)	1317.414 (1179.225, 1458.656)
Interaction	0.0102 (0.0043)
DIC	1163.81
Effective number of parameters	282.80(16.67)
Log(mean(cpo))	0.0699

Reference values in brackets
Highlighted RR>1. In bold cursive RR<1

Table 6.- Negligence and accidents – ZIP0

	RR (95% credible interval)
Slope (<3%)	
Q2 (3%-5%)	1.0116 (0.7916, 1.2702)
Q3 (5%-8%)	0.8074 (0.6257, 1.0226)
Q4 (8%-13%)	0.9579 (0.7601, 1.1913)
Q5 (13%-66%)	1.1705 (0.8962, 1.5016)
Aspect_Orientac (<84º)	
Q2 (84º-147º)	1.3429 (0.9363, 1.8729)
Q3 (147º-202º)	1.3879 (1.0078, 1.8720)
Q4 (202º-264º)	1.1926 (0.8937, 1.5639)
Q5 (264º-360º)	0.8726 (0.6134, 1.2102)
Hill shade(24º-159º)	
Q2 (159º-172º)	1.4129 (0.9936, 1.9518)
Q3 (172º-180º)	1.6212 (1.0697, 2.3583)
Q4 (180º-189º)	2.2423 (1.3901, 3.4325)
Q5 (189º-251º)	1.8630 (1.1523, 2.8496)
Elevation (<90m)	
Q2 (90m-179m)	1.0447 (0.7849, 1.3627)
Q3 (179m-318m)	0.7468 (0.5439, 1.0007)
Q4 (318m-521m)	0.8506 (0.6232, 1.1353)
Q5 (521m-2532m)	***0.4681 (0.3330, 0.6393)***
Land use (urban, beaches, sand, bare rocks, burnt areas, and water bodies)	
Coniferous forests	1.0406 (0.6734, 1.5380)
Dense forests	0.7895 (0.3009, 1.6189)
Fruit trees and berries	0.9012 (0.5922, 1.3208)
Artificial non-agricultural vegetated areas	1.3821 (0.7929, 2.2372)
Transitional woodland scrub	1.4072 (0.9631, 1.9965)
Natural grassland	3.5427 (0.3619, 12.1879)
Mixed forests	1.3519 (0.7038, 2.3191)
Distance to urban areas, roads and railways (<60m)	
Q2 (60m-169.7056m)	1.0917 (0.9044, 1.3033)
Q3 (169.7056m-361.2478m)	1.2781 (1.0544, 1.5312)
Q4 (361.2478m-724.9828m)	0.9857 (0.8092, 1.1864)
Q5 (724.9828m-10494.5557m)	***0.6830 (0.5440, 0.8438)***
Minimum temperature	
Lag 1	1.0079 (0.8792, 1.1495)
Lag 2	1.1690 (1.0047, 1.3573)
Lag 3	0.8865 (0.7588, 1.0307)
Lag 4	1.0326 (0.8894, 1.1943)
Lag 5	***0.8483 (0.7402, 0.9663)***
Lag 6	1.0160 (0.8766, 1.1731)
Lag 7	1.0918 (0.9447, 1.2546)
Maximum temperature	
Lag 1	0.8344 (0.6878, 1.0032)
Lag 2	1.0885 (0.8810, 1.3357)
Lag 3	1.2047 (0.9263, 1.5491)
Lag 4	0.9873 (0.7796, 1.2475)
Lag 5	0.8240 (0.6319, 1.0568)
Lag 6	1.2329 (0.9662, 1.5550)
Lag 7	0.9381 (0.7676, 1.1346)

Random effects	Mean (standard deviation)
Heterogeneity	0.9199 (0.1258)
Temporal	0.0101 (0.0063)
Spatial	0.2907 (0.6514)
Range (mean – 95% credible interval)	1782.846 (1433.903, 2125.503)
Interaction	0.0199 (0.0063)
DIC	9358.76
Effective number of parameters	365.28(16.27)
log(mean(cpo))	0.5666

Reference values in brackets
Highlighted RR>1. In bold cursive RR<1

Table 7.- Intentional or arson – ZIP1

	RR (95% credible interval)
Slope (<3%)	
Q2 (3%-5%)	0.9478 (0.7940, 1.1220)
Q3 (5%-8%)	0.9407 (0.7899, 1.1121)
Q4 (8%-13%)	1.1283 (0.9144, 1.3783)
Q5 (13%-66%)	1.1735 (0.9115, 1.4889)
Aspect_Orientac (<84º)	
Q2 (84º-147º)	1.1349 (0.8936, 1.4210)
Q3 (147º-202º)	1.0906 (0.8622, 1.3599)
Q4 (202º-264º)	1.2354 (1.0231, 1.4787)
Q5 (264º-360º)	1.0516 (0.8156, 1.3361)
Hill shade(24º-159º)	
Q2 (159º-172º)	1.2330 (0.9444, 1.5813)
Q3 (172º-180º)	1.0621 (0.7542, 1.4535)
Q4 (180º-189º)	1.1047 (0.7337, 1.5972)
Q5 (189º-251º)	1.1243 (0.7658, 1.5922)
Elevation (<90m)	
Q2 (90m-179m)	1.3461 (1.0773, 1.6600)
Q3 (179m-318m)	1.1821 (0.8928, 1.5309)
Q4 (318m-521m)	0.6471 (0.4561, 0.8944)
Q5 (521m-2532m)	0.5212 (0.3517, 0.7403)
Land use (urban, beaches, sand, bare rocks, burnt areas, andwater bodies)	
Coniferous forests	1.1873 (0.7845, 1.7276)
Dense forests	0.5230 (0.1974, 1.1036)
Fruit trees and berries	0.9842 (0.6443, 1.4427)
Artificial non-agricultural vegetated areas	1.3178 (0.8482, 1.9574)
Transitional woodland scrub	1.2454 (0.8527, 1.7621)
Natural grassland	1.0098 (0.0294, 4.5954)
Mixed forests	1.1650 (0.6197, 1.9950)
Distance to urban areas, roads and railways (<60m)	
Q2 (60m-169.7056m)	0.9863 (0.8491, 1.1382)
Q3 (169.7056m-361.2478m)	1.0430 (0.9088, 1.1902)
Q4 (361.2478m-724.9828m)	0.9724 (0.8482, 1.1089)
Q5 (724.9828m-10494.5557m)	0.9357 (0.7792, 1.1127)
Minimum temperature	
Lag 1	1.0064 (0.9024, 1.1187)
Lag 2	0.9357 (0.8224, 1.0597)
Lag 3	1.0165 (0.9013, 1.1424)
Lag 4	1.0194 (0.8884, 1.1646)
Lag 5	1.0086 (0.8807, 1.1490)
Lag 6	1.0093 (0.8712, 1.1611)
Lag 7	1.0311 (0.9212, 1.1501)
Maximum temperature	
Lag 1	0.9843 (0.8943, 1.0805)
Lag 2	1.0401 (0.9401 1.1484)
Lag 3	0.9171 (0.8209 1.0214)
Lag 4	1.0291 (0.9243, 1.1434)
Lag 5	0.9827 (0.8765, 1.0991)
Lag 6	1.1336 (1.0164, 1.2612)
Lag 7	0.9411 (0.8651, 1.0220)

Random effects	Mean (standard deviation)
Heterogeneity	0.8180 (0.0605)
Temporal	0.0091 (0.0049)
Spatial	1.5223 (0.7053)
Range (mean – 95% credible interval)	797.6238 (315.573, 1289.779)
Interaction	0.0370 (0.0067)
DIC	5031.78
Effective number of parameters	322.24 (15.50)
log(mean(cpo))	0.2845

Reference values in brackets
Highlighted RR>1. In bold cursive RR<1

Table 8.- Unknown causes and rekindled– Poisson

	RR (95% credible interval)
Slope (<3%)	
Q2 (3%-5%)	0.9276 (0.7433, 1.1401)
Q3 (5%-8%)	***0.8125 (0.6535, 0.9960)***
Q4 (8%-13%)	***0.6267 (0.4964, 0.7792)***
Q5 (13%-66%)	0.7977 (0.6195, 1.0107)
Aspect_Orientac (<84º)	
Q2 (84º-147º)	***0.7598 (0.5697, 0.9946)***
Q3 (147º-202º)	0.9764 (0.7440, 1.2608)
Q4 (202º-264º)	0.7919 (0.6050, 1.0160)
Q5 (264º-360º)	0.8938 (0.6517, 1.1977)
Hill shade(24º-159º)	
Q2 (159º-172º)	0.8042 (0.5875, 1.0736)
Q3 (172º-180º)	***0.5485 (0.3752, 0.7730)***
Q4 (180º-189º)	***0.5640 (0.3660, 0.8303)***
Q5 (189º-251º)	0.6943 (0.4401, 1.0404)
Elevation (<90m)	
Q2 (90m-179m)	1.2445 (0.9622, 1.5840)
Q3 (179m-318m)	1.1526 (0.8615, 1.5095)
Q4 (318m-521m)	1.0155 (0.7147, 1.3965)
Q5 (521m-2532m)	0.8577 (0.5828, 1.2086)
Land use (urban, beaches, sand, bare rocks, burnt areas and, water bodies)	
Coniferous forests	***0.6536 (0.4553, 0.9117)***
Dense forests	***0.1965 (0.0574, 0.4468)***
Fruit trees and berries	***0.6162 (0.4255, 0.8659)***
Artificial non-agricultural vegetated areas	***0.4234 (0.2409, 0.6765)***
Transitional woodland scrub	***0.6226 (0.4502, 0.8432)***
Natural grassland	1.8478 (0.3546, 5.1774)
Mixed forests	0.8200 (0.4594, 1.3380)
Distance to urban areas, roads and railways (<60m)	
Q2 (60m-169.7056m)	0.8704 (0.7138, 1.0463)
Q3 (169.7056m-361.2478m)	0.9654 (0.7997, 1.1506)
Q4 (361.2478m-724.9828m)	0.9458 (0.7811, 1.1319)
Q5 (724.9828m-10494.5557m)	1.4568 (1.1857, 1.7667)
Minimum temperature	
Lag 1	1.0431 (0.9114, 1.1884)
Lag 2	1.1450 (0.9836, 1.3251)
Lag 3	1.1164 (0.9654, 1.2847)
Lag 4	***0.8209 (0.7207, 0.9345)***
Lag 5	0.9414 (0.8201, 1.0769)
Lag 6	1.0223 (0.8847, 1.1774)
Lag 7	0.9802 (0.8552, 1.1196)
Maximum temperature	
Lag 1	1.1456 (0.9394, 1.3853)
Lag 2	0.8938 (0.7183, 1.1019)
Lag 3	0.8889 (0.7256, 1.0860)
Lag 4	1.3642 (1.0767, 1.7095)
Lag 5	***0.8020 (0.6422, 0.9944)***
Lag 6	1.1036 (0.8708, 1.3870)
Lag 7	0.9504 (0.7779, 1.1520)

Random effects	Mean (standard deviation)
Heterogeneity	1.2458 (0.1420)
Temporal	0.0114 (0.0073)
Spatial	0.2624 (0.6582)
Range (mean – 95% credible interval)	1455.749 (845.9864, 1897.105)
Interaction	0.0745 (0.0171)
DIC	4437.04
Effective number of parameters	538.28(31.14)
log(mean(cpo))	0.3643

Reference values in brackets
Highlighted RR>1. In bold cursive RR<1

Figure 1: Location of Catalonia in Europe. The zoom shows the study area in more detail pointing out the regions and provinces in which Catalonia is divided.

Figure 2: In the abscissa we discernthe eight categories of land use listed in the ordinate. On the vertical axis, we showed the number of fires distinguished by cause. From top-left to bottom right, we graphed wildfires triggeredby natural causes; those caused by negligence and accidents; intentional or arson wildfires and unknown causes and rekindled.

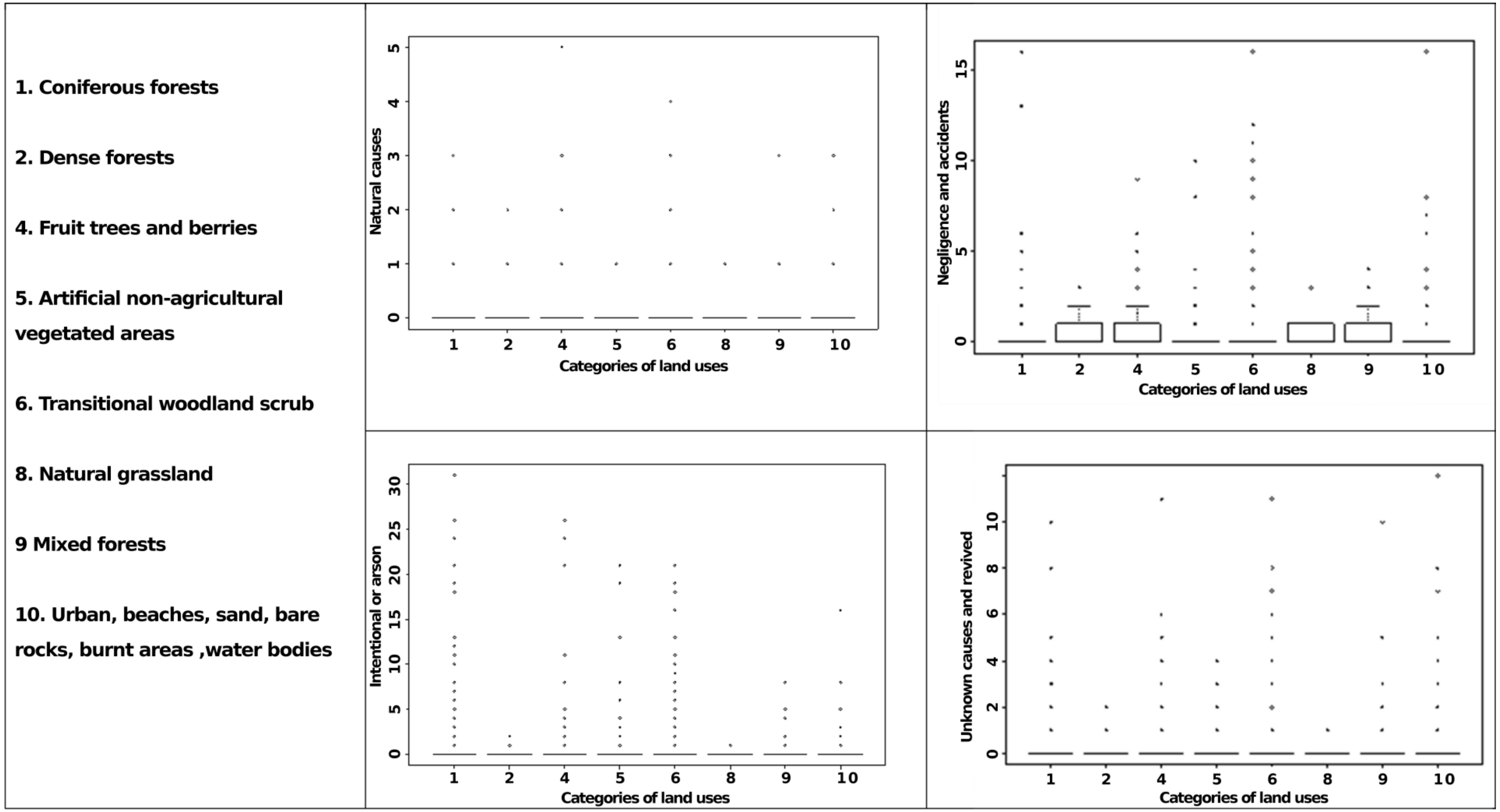


Figure 3a: The intercept distribution.

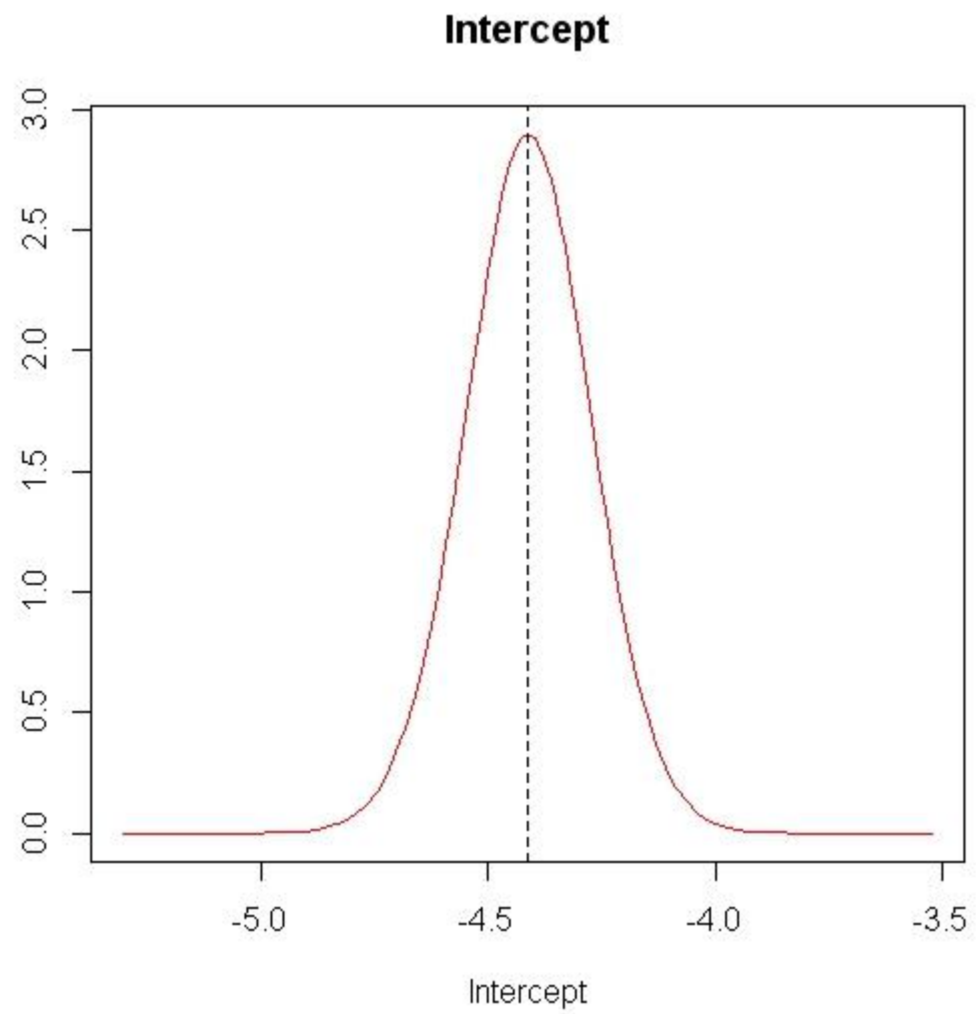


Figure 3b: From top-left to bottom right it is showed the random effects distributions: the marginal posterior distribution for buffer, time, tau and kappa.

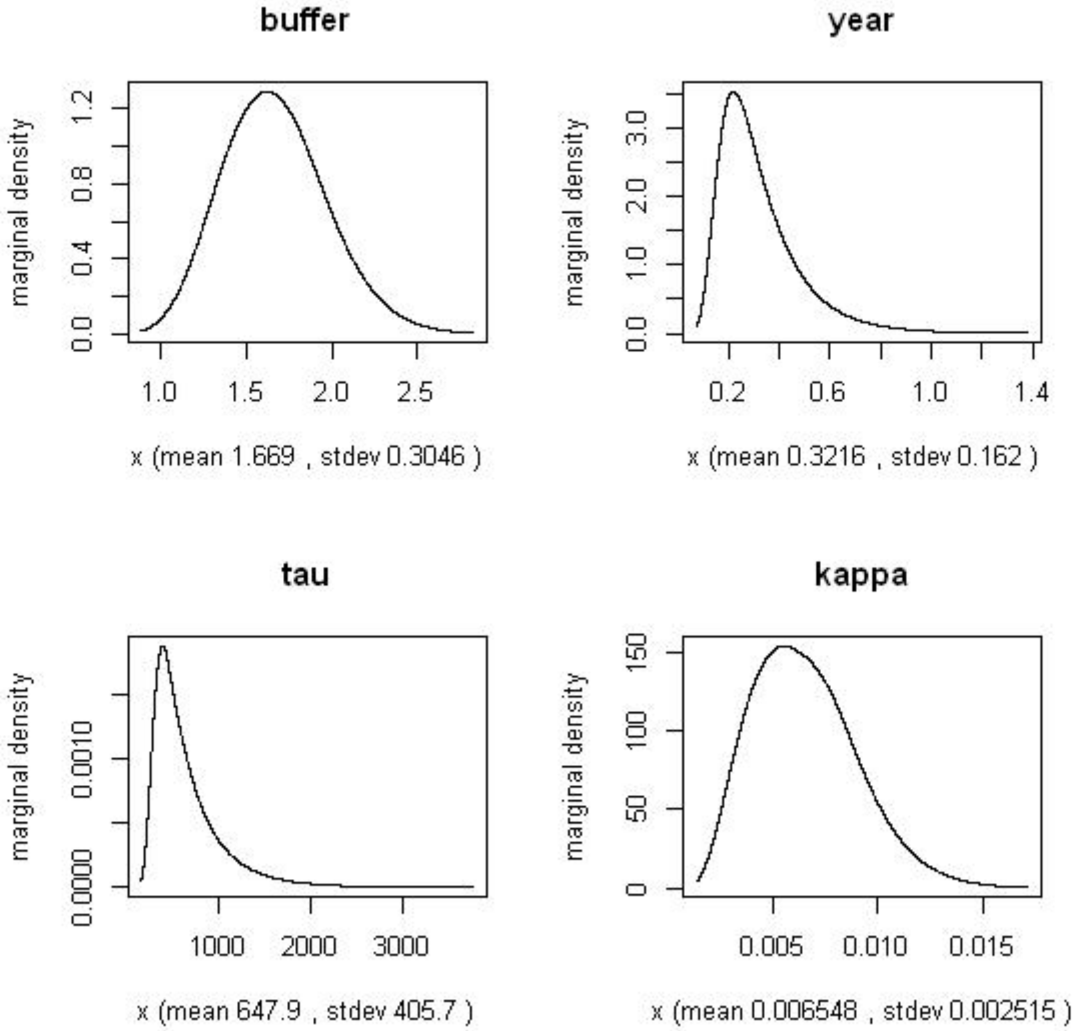


Figure 4: Effect of temporal dependency on relative risk of wildfires distinguishing by causes. From top-left to bottom-right it is showed those wildfires corresponding to natural causes, negligence and accidents, intentional or arson and those caused by unknown causes or rekindled.

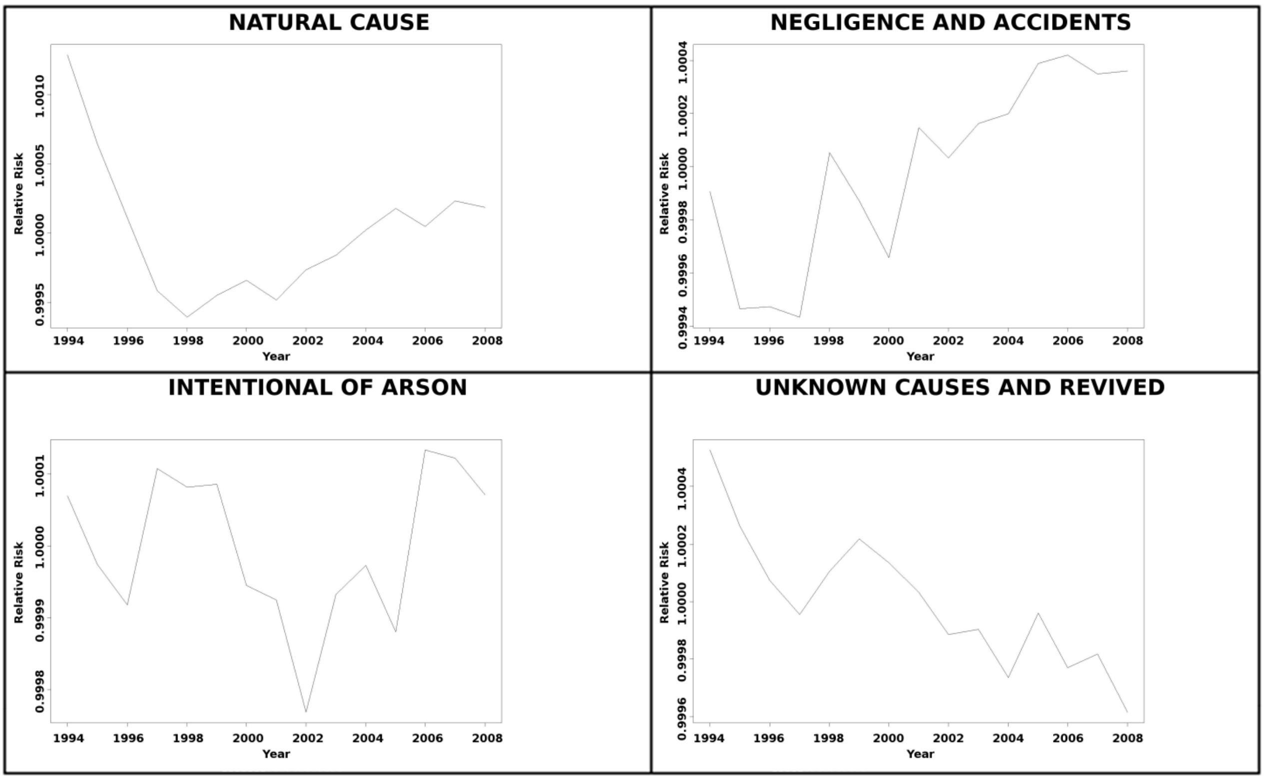


Figure 5a: Effect of spatial dependency on relative risk of wildfire. On the top it is showed the results from wildfires caused by natural causes and on the bottom those caused by negligence and accidents. From left to right the results are specified in 4 years: 1994, 1999, 2004 and 2008 respectively.

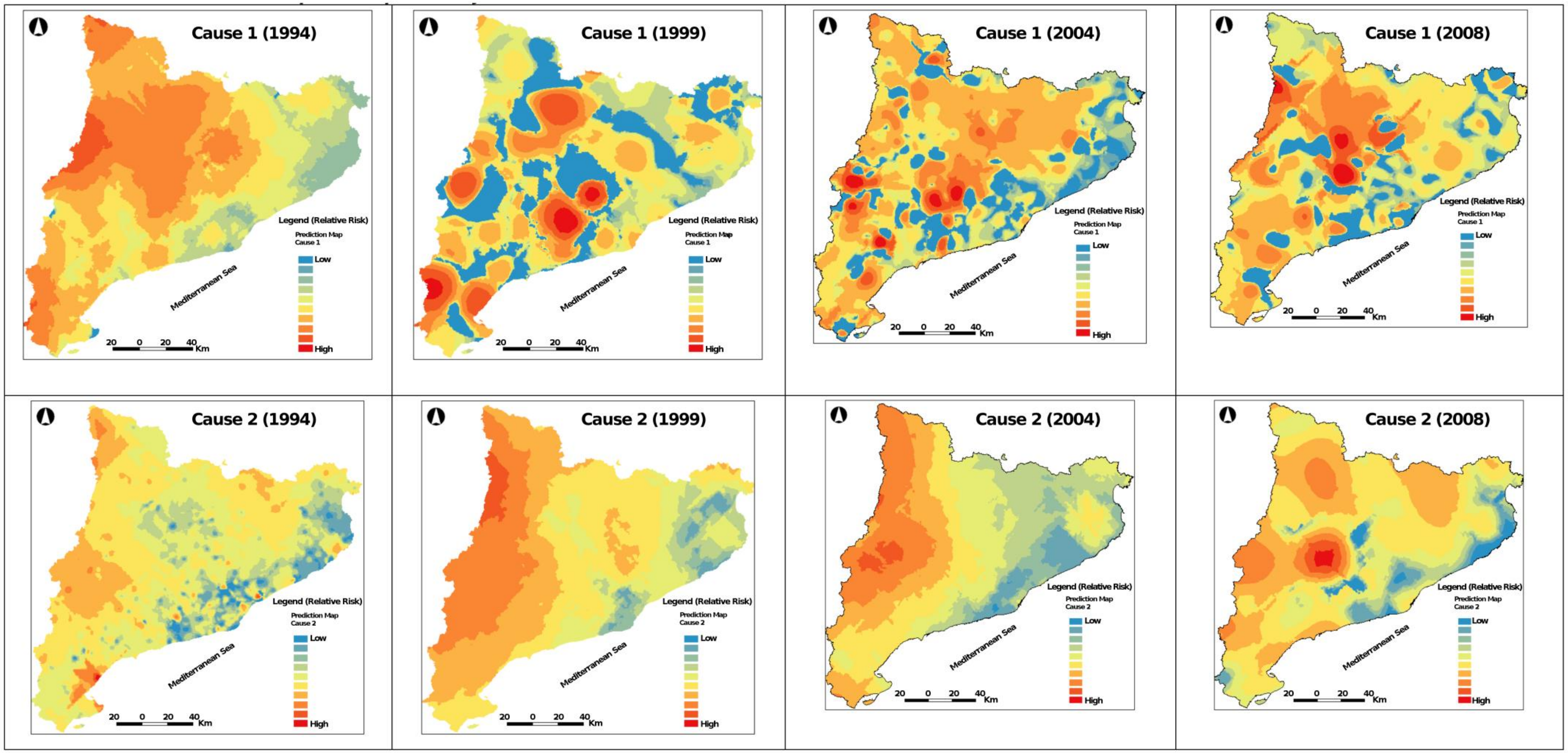
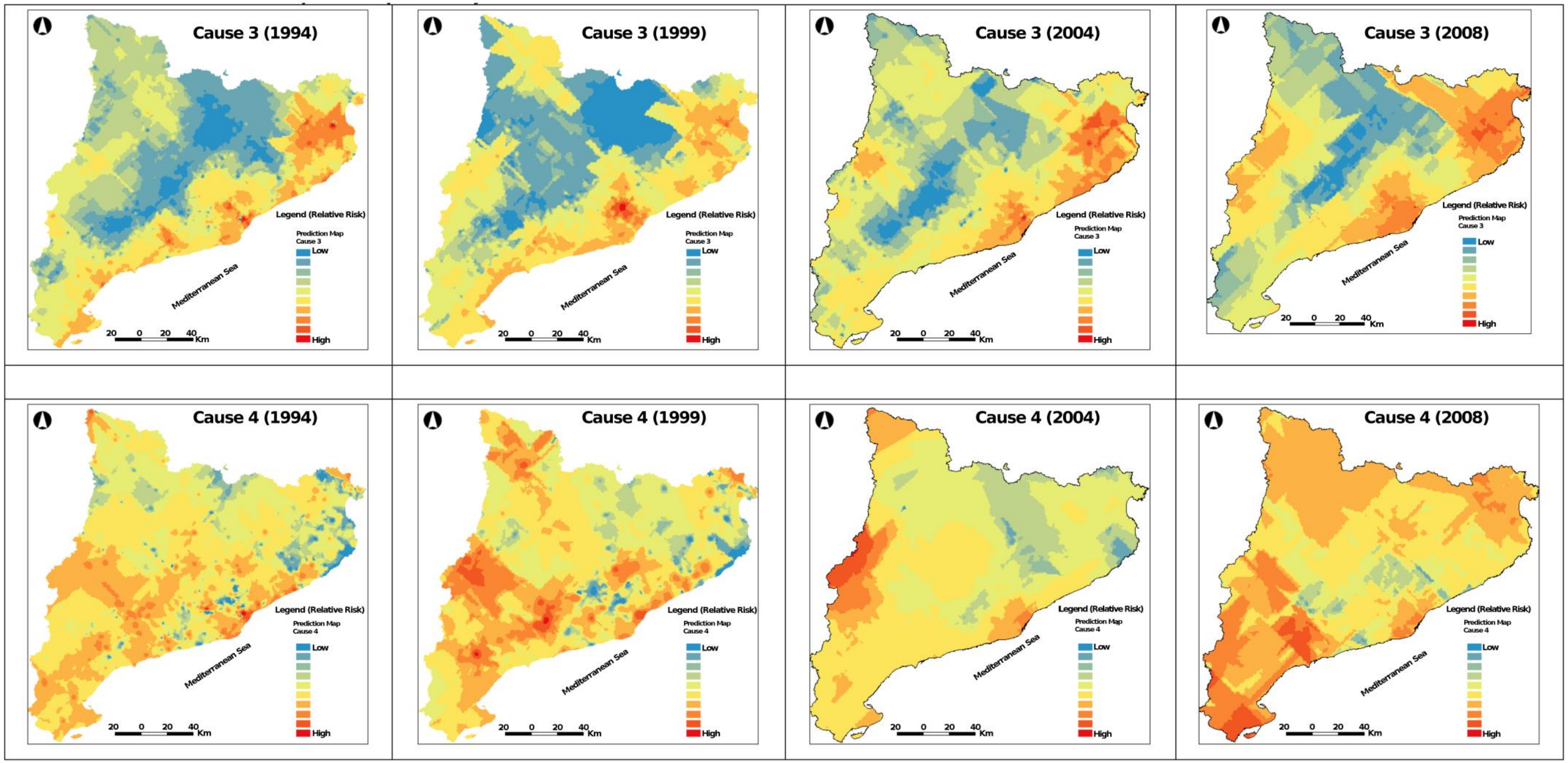


Figure 5b: Effect of spatial dependency on relative risk of wildfire. On the top it is showed the results from intentional wildfires or arson and on the bottom those caused by unknown causes or rekindled. From left to right the results are specified in 4 years: 1994, 1999, 2004 and 2008 respectively.



A spatio-temporal Poisson Hurdle point process to model wildfires

Laura Serra[1,2]*, Marc Saez[2,1] , Pablo Juan[3] , Diego Varga[2,4], Jorge Mateu[3]

Abstract. Wildfires have been studied in many ways, for instance as a spatial point pattern or through modelling the size of fires or the relative risk of big fires. Lately a large variety of complex statistical models can be fitted routinely to complex data sets, in particular wildfires, as a result of widely accessible high-level statistical software, such as R. The objective in this paper is to model the occurrence of big wildfires (greater than a given extension of hectares) using an adapted two-part econometric model, specifically a hurdle model. The methodology used in this paper is useful to determine those factors that help any fire to become a big wildfire. Our proposal and methodology can be routinely used to contribute to the management of big wildfires.

Key words and phrases. Hurdle model, INLA, Spatio-temporal point processes, SPDE, Wildfire.

1. Introduction

Fire risk can be defined as a product of fire occurrence probability and expected impacts [3]. An area can be considered to have high wildfire risk if the probability of fire is high and the expected impacts of fire are large. Furthermore, fires are getting larger, more destructive, and more economically expensive due to fuel accumulations, shifting land management practices, and climate change. Wildfires have negative effects on human life and health, human property and wellbeing, cultural and natural heritage, employment, recreation, economic and social infrastructures and activities. It is worth noting that some fire episodes have caused catastrophic damages as loss of human lives and very significant economic and environmental losses.

The European Mediterranean is a highly populated region. Approximately 65,000 fires occur in the European Mediterranean region every year. Wildfires destroy around 500,000 hectares every year in the European Union, 0.7 to 1 million hectares in the Mediterranean basin. This has a serious impact on the environment and on socio-economic activities, especially in southern Europe. Over 95% of the fires in Europe are due to human causes. An analysis of fire causes show that the most common cause of fires comes from agricultural practices, followed by

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

negligence and arson ([34]). These wildfires are relatively frequent events with recurrence time of 23 years ([42]).

Wildfires also destroy biodiversity, increase desertification, affect air quality, the balance of greenhouse gases and water resources. During recent years the increasing extension of urban areas mixed with rural or forest areas associated with a marked increase of fire activity make this impact even greater. The intense urbanization of our societies, the abandonment of rural lands and rural activities such as forest management along with the rapidly expanding of urban/forest interface are key drivers for wildfires in Europe and in the Mediterranean region.

Weather is a fundamental component of the fire environment. The prolonged drought and high temperatures of the summer period in the Mediterranean climate are the typical drivers that demarcate the temporal and spatial boundaries of the main fire season. Future trends of wildfire risks in the Mediterranean region, as a consequence of climate change, will lead to the increase of temperature in the East and West of the Mediterranean, with more frequent dryness periods and heat waves facilitating the development of very large fires. Future scenarios of climate change should affect locally fire regimes, and therefore local analyses need to be performed by adapting global climatic models to regional conditions. Many factors have been considered to explain the temporal variation in fire regime in recent decades in Spain: Climate change is one factor, with a clear relationship between increasing number of days with extreme fire hazard weather and the number and size of fires in the Mediterranean coast of Spain.

Earlier detection often leads to smaller fire size, and therefore reduces the probability of fire escape ([21]), final fire size, cost and risks to fire response crews. Wildfire prevention should be considered as an important part of sustainable forest management and should integrate a landscape approach taking into account different land uses. Knowledge of short and long-term impacts of wildfire is essential for effective risk assessment, policy formulation, and wildfire management.

Spain is one of the most affected countries in Europe, both considering number of fires and area burned. Between 1980 and 2004 nearly 380.000 fires have occurred in Spain, and more than 4.7 millions hectares have been burned (roughly 10% of the country). Extreme fires (>500ha) are relatively frequent events with recurrence time of 2-3 years, causing large human, economic and environmental damage altogether. Their ignition and spread occur under favorable weather conditions, often following drought periods, in areas where fuel accumulation helps quick fire spread and high fire intensity, they usually burn out of control and can only be stopped when meteorological conditions support aerial and ground fire fighting ([39]). In Catalonia these fires only represent 1.4% of all fires and 79% of burned area. In this study we have included wildfires larger than 50ha because in the Mediterranean region represent more than 75% of the area burned, although they represent only 2.6% of the total number of wildfires ([19] and [30]). Over the last few years, the occurrence of large wildfire episodis with extreme fire behavior has affected different regions of Europe: Portugal, south-eastern France, Spain and Greece.

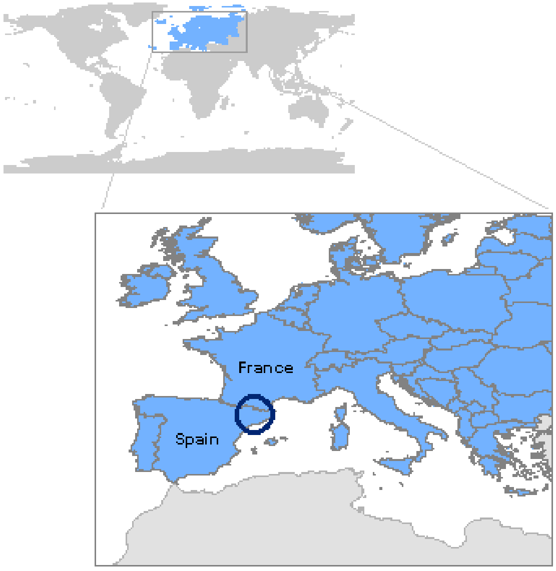


Figure 1.Catalonia location in Europe.

Wildfires have been studied in many ways, for instance as a spatial point pattern ([8], [9], [24], [42] and [44]) or through modelling the size of fires ([1]) or the relative risk of the big fires ([45]). Lately a large variety of complex statistical models can be fitted routinely to complex data sets, in particular wildfires, as a result of widely accessible high-level statistical software, such as R ([32]). Researchers from many different disciplines are now able to analyse their data with sufficiently complex methods rather than resorting to simpler yet non-appropriate methods. In this case, the objective in this paper is to model the occurrence of big wildfires, and to determine those factors which are significative in helping any fire to become a big wildfire.

We analyse the occurrence of big wildfires in Catalonia between 1994 and 2011, and consider a big wildfire to be a fire that burns areas larger than a fixed extension of hectares. Specifically we consider three sizes of areas; 50ha, 100ha and 150ha. Moreover, we distinguish between the numerous potential causes of wildfire ignition. In particular, we consider: (i) natural causes; (ii) negligence and accidents; (iii) intentional fires or arson; and (iv) unknown causes and rekindled. The study area encompasses 32,000 square kilometers and represents about 6.4% of the total Spanish national territory (1).

In addition to the locations of the fire centroids, several marks and covariates are considered. The year the wildfire occurred is the unique mark considered. The spatial covariates are also considered, specifically, eight continuous covariates (i.e. topographic variables – slope, aspect, hill shade and altitude, proximity to anthropic areas – roads, urban areas and railways, and meteorological variables – maximum and minimum temperatures) and one categorical variable (land use).

The methodology for fitting spatial point process models to complex data sets has seen previous advances in facilitating routine model fitting for spatial point processes. For instance, the work by [4] has facilitated the routine fitting of point processes based on an approximation of the pseudolikelihood to avoid the issue of intractable normalizing constants ([5]) through the use of the library spatstat for R ([4]). In the same way, ([22]) consider hierarchical models able to analyse a wide variety of point process models, for example those appearing in fire problems.

In our case, spatio-temporal data can be idealised as realizations of a stochastic process indexed by spatial and temporal coordinates. 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 ([42]). Covariate information usually comes in the form of spatial patterns in regular lattices or as regular vector polygons that may be rasterised into lattice images using GIS ([41]). The right methodological context able to deal with these pieces of information comes from spatio-temporal point processes. To bypass the problem of inefficiency in the estimation under a general integrated nested Laplace approximation (INLA)([36]), we have tried a computationally tractable approach based on stochastic partial differential equation (SPDE) models ([25]). On one hand, we use 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 ([36]). 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.

The proposed method in this paper is an adapted two-part econometric model, specifically a Hurdle model. It consists of two stages and it is specified in such a way as to gather together the two processes theoretically involved in the presence of wildfires, that is, the fact to be a big wildfire (greater than a given extension of hectares) and the frequency of big wildfires per spatial unit. Specifically, the Poisson hurdle model consists of a point mass at zero followed by a truncated Poisson distribution for the non-zero observations.

This paper addresses two issues. We develop complex joint models for big wildfires and, at the same time, we provide methods facilitating the routine for the fitting of these models, using a Bayesian approach. The approach is based on the INLA, which speeds up parameter estimation substantially so that particular models can be fitted within feasible time.

This paper is organised as follows: the following section describes the data. Section 3 presents the methodology used, including the statistical framework, the description of the Poisson Hurdle model and the statistical inference explanation. Section 4 presents the results. Finally, the paper ends with a discussion and future coming steps.

2. DATA SETTING

In this paper we analyse the occurrence of big wildfires in Catalonia between 1994 and 2011. The total number of fires recorded in the analysis is 3,283, which are distributed as follows: 206 wildfires bigger than 50ha, 141 wildfires bigger than 100ha, and 112 wildfires bigger than 150ha. In Figure 2, on the left, we can see all wildfires and wildfires bigger that 50ha.

In Catalonia, the agency responsible for identifying the coordinates of the origin of the fire, the starting time and the cause of the fire is the Forest Fire Prevention Service (Government of Catalonia). 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 are provided directly by the Service, and have been tested and polished before handling.

We distinguish between the numerous potential causes of wildfire ignition. In particular, we consider: (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. In Figure 2, on the right, we show the spatial distribution of wildfires bigger than 50ha distinguishing by causes.

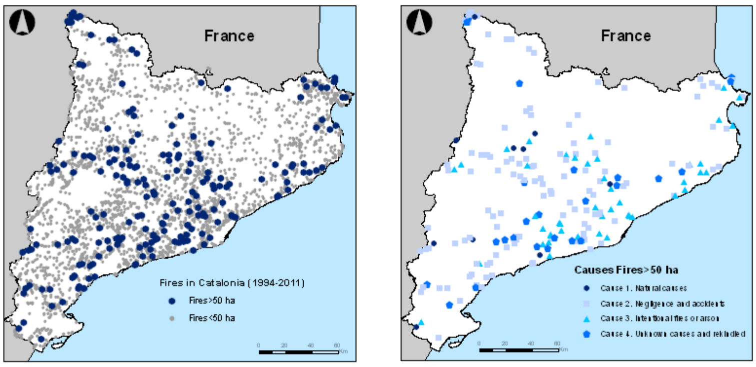


Figure 2. Left: All wildfires (1994-2011) and big wildfires. Right: Big wildfires distinguishing by causes.

In addition to the locations of the fire centroids, measured in Cartesian coordinates (Mercator transversal projections, UTM, Datum ETRS89, zone 31-N), several covariates are considered. Specifically, eight continuous covariates (i.e. topographic variables – slope, aspect, hill shade and altitude; 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 ([29]), but also the weather, inducing diverse local wind conditions, which include slope and valley winds. In fact, [15] point out that those topographic variables are relatively more important predictors of severe fire occurrence, than either climate or weather variables. The proximity to anthropic areas can 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 are feasibly important for natural cause fires and perhaps rekindled fires, we use the maximum and minimum temperatures (further details can be found in [42]).

In this paper, slope is 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 is computed as the maximum rate of change in elevation between the location and its surroundings. Slope is expressed in degrees. Aspect is the orientation of the slope and it is measured clockwise in degrees from 0 to 360, where 0 is north-facing, 90 is east-facing, 180 is south-facing, and 270 is west-facing. Hill shading is a technique used to visualise terrain as shaded relief by illuminating it with a hypothetical light source. Here, the illumination value for each raster cell is determined by its orientation to the light source, which, in turn, is based on slope and aspect and is also measured in degrees, from 0 to 360. Finally, altitude is considered as elevation above sea level and it is expressed in meters. To obtain topographic variables (DTM) we use the MET-15 model, which is a regular grid containing orthometric heights distributed according to a metricconverterProductID15 m15 m grid side, and is created for the Cartographic Institute of Catalonia. We also use the surface analysis tools included in the ArcGis10 application Spatial Analyst ([42]).

The distances, in meters, from the location of the wildfire to urban areas, roads and railroads, are constructed by considering a geographical layer in each case. The urban area and road layers are obtained from the Department of Territory and Sustainability of the Catalan Government, through the Cartographic Institute of Catalonia (ICC) (http://www.icc.cat). To obtain the two new raster layers we use the Euclidean distance function, included in the ArcGis10 application Spatial Analyst. Then, we use the merge function of ArcGis10 Geoprocessing module, to combine those two layers (urban areas and roads and railroads) into one single layer. The layers are continuous and defined as a raster layer (details can be found in [42]).

We also use the land use in Catalonia maps (1:250,000), with classification techniques applied on existing LANDSAT MSS images for 1992, 1997 and 2002 ([7], [17] and [35]). Additionally, we use orthophotomaps (1:5000) 2005-2007, to create the land use map for 2010. Specifically, we assign the land use map just before the date of each wildfire. We assign, as the land use, only the percentage value corresponding to the principal land use of the spatial units. In this paper, we transform 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.

We also consider the temperatures (maximum and minimum) and up to seven days before the occurrence of the fire, at the location of the wildfire (note that meteorological data are 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, are estimated by means of a two-step Bayesian model. Further details can be found in [37].

3. METHODS

3.1. **Statistical framework**. Spatio-temporal data can be idealised as realizations of a stochastic process indexed by a spatial and a temporal dimension

$$(3.1) \qquad Y(s,t) \equiv \{y(s,t) | (s,t) \in D \times T \in \mathbb{R}^2 \times \mathbb{R}\}$$

where$D$ is a (fixed) subset of $\mathbb{R}^2$ and T is a temporal 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 locations, at which the measurements are taken, and $(t_1,\dots,t_n)$ the temporal instants.

In our case we assume separability in the sense that we model the spatial correlation by the Matérn spatial covariance function defined in (3.7) and the temporal correlation using a Random Walk model of order 1 (RW1). We introduce also the interaction effect between the space and time using another RW1 structure. Nevertheless, this inclusion does not change the separability structure. This temporal structure can be justified by the apparent randomness as shown in Figure 3. In fact, the dispersion of big wildfires varies between the periods considered. In particular, there is a reduction considering the number of them, specifically in the period 2008-2011.

3.2. **The Poisson hurdle model**. The model used in this paper is an adapted two-stage econometric model proposed by [13], specifically a hurdle model. It consists of two stages and specified in a way to gather together the two processes theoretically involved in the presence of wildfires, that is, the occurrence of being a big wildfire (greater than a given extension of hectares) and the frequency of big wildfires per spatial unit ([28]). Specifically, the Poisson hurdle model consists of a point mass at zero followed by a truncated Poisson distribution for the non-zero observations.

In the first stage, we predict the probability that any wildfire becomes larger than 50ha, 100ha and 150ha. In the second part, we model the number of these big wildfires per spatial unit.

The first part of the process can be modeled using a logistic regression that models the probability that any wildfire becomes larger than a fixed area

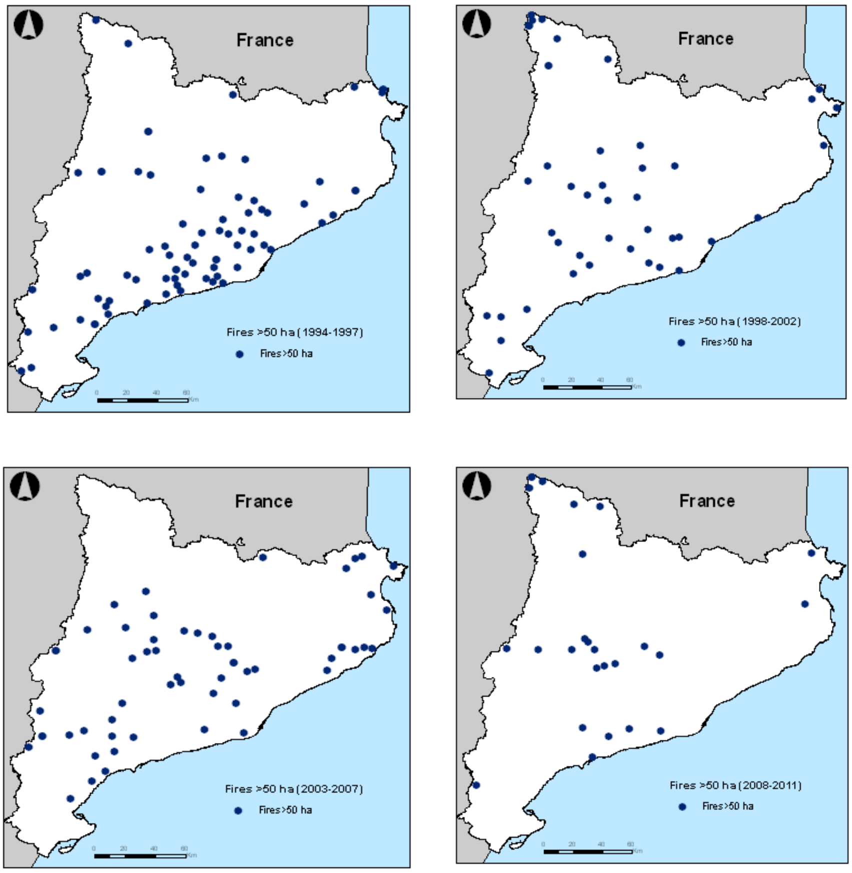


Figure 3.Big wildfires in Catalonia in 1994 to 2011. Left-Up: 1994-1997; Right-Up: 1998-2002; Left-Down: 2003-2007 and Right-Down: 2008-2011.

$$p_{itk} = Prob(y_{itk} > A|Z,\beta)$$

$$log\left(\frac{p_{itk}}{1-p_{itk}}\right) = Z'\beta + S_i + \tau_t + \upsilon_{it} \quad (3.2)$$

where A denotes one of the fixed area's values (50ha, 100ha or 150ha), y is the response variable (in this case, each wildfire), Z a matrix of explanatory spatial covariates (containing the intercept), β is the vector of unknown parameters associated with the covariates, the subscript i denotes the wildfire, the subscript t (t=1994,..., 2011) the year of occurrence of the wildfire, and the subscript k (k =1,..., 4) the cause of occurrence. We also introduced three random effects: (i) spatial dependence, S_i, (ii) temporal dependence, τ_t and (iii) spatio-temporal interaction, υ_{it}.

In accordance with that proposed by [27], in the second stage of the model the distribution of being a big wildfire follows a truncated Poisson that models the number of big wildfires per spatial unit, introducing covariates and spatial random effects ([28])

(3.3)
$$p(y_{itk}\,|S_i) = (1 - p_{itk})1_{(y_{itk}<A)} + p_{itk}\,Tpois(y_{itk}\,;\mu_{itk})1_{(y_{itk}>A)}$$
$$\log(\mu_{itk}) = \eta(p_{itk})$$
$$\eta(p_{itk}) = \sum_m \beta_m Z_{m,it} + S_i + \tau_t + \upsilon_{it}$$

where $Tpois(y_{itk}\,;\mu_{itk})$ denotes a truncated Poisson distribution with parameter μ_{itk}, η denotes a link function such as the logit link, $Z_{m,it}$ represents the same spatial covariates used in the first stage, and β_m denotes the parameters associated with these covariates.

The particular estimation process has two steps. In the first step we use a binomial link in order to estimate the occurrence of a big wildfire. The probabilities of occurrence obtained from this first step are used in the second step as interim priors. In the second step the link is a truncated Poisson distribution. In any case, the likelihood of each part is introduced multiplicatively in only one equation.

3.3. Statistical inference.

3.3.1. *SPDE approach.* The SPDE approach allows to represent a Gaussian Field with the Matérn covariance function defined in (3.7) as a discretely indexed spatial random process which produces significant computational advantages ([25]). Gaussian Fields are defined directly by their first and second order moments and their implementation is highly time consuming and provokes the so-called “big n problem”. This is due to the computational costs of $O(n^3)$ to perform a matrix àlgebra operation with $n \times n$dense covariance matrices, which is notably bigger 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 given by (3.8), which is a link between the GF and the GMRF. This link 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 ([25]).

In particular 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,

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

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 funcions are not random, but rather are chosen to be piecewise linear on each triangle

$$\varphi_l(s) = \begin{cases} 1 \ at \ vertix \ 1 \\ 0 \ elsewhere \end{cases}$$

The key is to calculate the weights $\{\omega_l\}$, 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 ([41]).

Thus, expression (3.4) defines an explicit link between the Gaussian field $X(s)$ and the Gaussian Markov random field, and it is 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) are assumed smoothed functions, in particular RW1 ([33]). Thus, RW1 for the Gaussian vector $x = (x_1, \dots, x_n)$ is constructed assuming independent increments

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

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

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

where$Q = \tau R$ and R is the structure matrix reflecting the neighbourhood structure of the model ([33]).

Considering a spatio-temporal geostatistical data we need to specify a valid spatio-temporal covariance function defined by $Cov\left(y_{it}, y_{jq}\right) = \sigma_C^2 M(s_i, s_j | t, q)$where $\sigma_C^2 > 0$ is the variance component and $M(s_i, s_j | t, q)$ is the Matérn spatio-temporal covariance function. Depending on our assumptions the spatio-temporal covariance function can be adapted to each situation. In the case of stationarity in space and time, the spatio-temporal covariance function can be specified as a function of the spatial Euclidean distance Δ_{ij}, and of the temporal lag $\Delta_{tq} = |t - q|$so it is defined by $Cov(y_{it}, y_{jq}) = \sigma_C^2 M(\Delta_{ij}; \Delta_{tq})$. If we assume separability, the spatio-temporal covariance function is given by $Cov(y_{it}, y_{jq}) = \sigma_C^2 M_1(\Delta_{ij}) M_2(\Delta_{tq})$, with M_1 and M_2 being the spatial and temporal correlation functions, respectively. Alternatively it is possible to consider a purely spatial covariance function given by $Cov(y_{it}, y_{jq}) = \sigma_C^2 M(\Delta_{ij})$ when t=q and 0 otherwise. In this last case, the temporal evolution could be introduced assuming that the spatial process evolves in time following an autoregressive dynamics ([20]).

Assuming separability we need to define the Matérn spatial covariance function which controls the spatial correlation at distance $\|h\| = \|s_i - s_j\|$ and this covariance is given by

(3.7) $$M(h|\nu, k) = \frac{2^{1-\nu}}{\Gamma(\nu)}(k\|h\|)^{\nu} K_{\nu}(k\|h\|)$$

whereK_ν is a modified Bessel function of the second kind and k>0 is a spatial scale parameter whose inverse, $1/k$, is sometimes referred to as a correlation length. The smoothness parameter ν>0 defines the Hausdorff dimension and the differentiability of the sample paths ([18]). Specifically, we tried ν=1,2,3 ([31]). Using the expression defined in (3.7), 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) ([41])

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

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

In the general spatial point process context, intensity stands for the number of events (fires in our case) per unit area. When considering the 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, Λjt is difficult to compute, and so we use instead the approximation, Λjt ≈ |sj | exp(ηjt (sj)), where ηjt (sj) is a 'representative value' (i.e., it represents the intensity or number of fires in a particular cell given by a linear predictor of covariates and other terms) ([41]), within the cell and |sj | is the area of the cell sj. To treat this kind of problems, Cox processes are widely used. In particular, Log Gaussian Cox processes (LGCP), which define a class of flexible models are particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field ([22]; [41]) and they are characterised by their intensity surface being modeled as

(3.9) $$\log(\lambda(s)) = Z(s)$$

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

3.3.2. *LGCP*. Conditional on a realization of $Z(s)$, a log-Gaussian Cox process is an inhomogeneous Poisson process. Considering a bounded region $\Omega \subset \mathbb{R}^2$ and given the intensity surface and a point pattern Y, the likelihood for a LGCP is of the form

$$\pi(Y|\lambda) = exp\left(|\Omega| - \int_{\Omega} \lambda(s)ds \prod_{s_i \in Y} \lambda(s_i)\right) \tag{3.10}$$

where the integral is complicated by the stochastic nature of $\lambda(s)$. 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 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 ([23]).

The basic idea is that, as we have explained in previous paragraphs, from a Gaussian Field (GF) with a Matérn covariance function, we 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.

Because our data is potentially zero inflated, as not all our events will become big fires, in this paper we present a spatial Poisson hurdle model to address these particular aspects of the data.

3.3.3. *Bayesian computation*. In a statistical analysis, to estimate a general model it is useful to model the mean for the i-th unit by means of an additive linear predictor, defined on a suitable scale

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m z_{mi} + \sum_{l=1}^{L} f_l(\upsilon_{li}) \tag{3.11}$$

where α is a scalar which represents the intercept, $\beta = (\beta_1, \dots, \beta_M)$ are the coefficients which quantify the effect of some covariates $z = (z_1, \dots, z_M)$ on the response, and $f = \{f_1(.), \dots, f_L(.)\}$ is a collection of functions defined in terms of a set of covariates $\upsilon = (\upsilon_1, \dots, \upsilon_L)$. From this definition, varying the form of the functions $f_l(.)$ we can estimate different kind of models, from standard and hierarchical regression, to spatial and spatio-temporal models ([36])

Given the specification in (3.8), the vector of parameters is represented by θ ={ α, β, f}.

In our case, assuming that the subscript i denotes the wildfire, the subscript j the municipal district and the subscript t (t=1994... 2011) the year of occurrence of the wildfire, for each cause, we specify the log-intensity of the Poisson process by a linear predictor ([23]) of the form

(3.12) $$\eta_{ijt}(s_j) = \alpha_{0j} + \beta_1 G_{ijt} + \beta_2 Z_{jt} + \beta_3 W_j + S_j + \tau_t + \upsilon_{jt}$$

where α_{0j} represents the heterogeneity accounting for variation in relative risk across different municipals districts,G_{ijt} represents those covariates which depend on the wildfire, the municipal district and the time, Z_{jt} represents those covariates which depend onthe municipal district and the time, W_j corresponds to those covariates which only depend on the municipal district,S_j is the spatial dependence, τ_t is the temporal dependence, and υ_{jt} is the spatio-temporal interaction.

Note that, we assume separability between spatial and temporal patterns and allow interaction between the two components.

Following the Bayesian paradigm we can obtain the marginal posterior distributions for each of the elements of the parameters vector

(3.13) $$p(\theta_i|y) = \int p(\psi|y)p(\theta_i|\psi, y)d\psi$$

and (possibly) for each element of the hyper-parameters vector

(3.14) $$p(\psi_k|y) = \int p(\psi|y)pd\psi_{-k}$$

Thus, we need to compute: (i) $p(\psi|y)$, from which all the relevant marginals $p(\psi_k|y)$ can be obtained, and (ii) $p(\theta_i|\psi, y)$, which is needed to compute the marginal posterior for the parameters. The INLA approach exploits the assumptions of the model to produce a numerical approximation to the posteriors of interest, based on the Laplace approximation ([43]).

Operationally, INLA proceeds by first exploring the marginal joint posterior for the hyper-parameters $\hat{p}(\psi|y)$ in order to locate the mode; a grid search is then performed and produces a set G of “relevant” points $\{\psi^*\}$ together with a corresponding set of weights, $\{w_{\psi^*}\}$to give the approximation to this distribution. Each marginal posterior $\hat{p}(\psi^*|y)$ can be obtained using interpolation based on the computed values and correcting for (probable) skewness, e.g. by using log-splines. For each ψ^*, the conditional posteriors $\hat{p}(\theta_i|\psi^*, y)$ are then evaluated on a grid of selected values for θ_i and the marginal posteriors $\hat{p}(\theta_i|y)$ are obtained by numerical integration ([6])

(3.15) $$\hat{p}(\theta_i|y) \approx \sum_{\psi^* \in G} \hat{p}(\theta_i|\psi^*, y)\hat{p}(\psi^*|y)w_{\psi^*}$$

Given the specification in (3.12), 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 (3.7). 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.

Note that in both parts of the model we control for heterogeneity, spatial dependence and spatio-temporal extra variability. Models are estimated using Bayesian inference for Gaussian Markov Random Field (GMRF) through the Integrated Nested Laplace Approximation (INLA).

The use of INLA and the SPDE algorithms produce massive savings in computational times and allow the user to work with relatively complex models in an efficient way. All analyses are carried out using the R freeware statistical package (version 2.15.2) ([32]) and the R-INLA package ([33]).

4. RESULTS

We note that, in general, wildfires caused by natural causes are not larger than 50ha. The same happens for those fires caused by unknown causes or for those rekindled. For this reason, even if we have analysed the forth causes we focus our results only on big wildfires caused by negligence and accidents and on those caused intentionally or arson.

4.1. **First stage results**.

We first consider a logistic regression to model the probability of a wildfire becoming larger than a particular area. Table 1 shows the significant factors of the logistic model distinguishing by the three sizes (50ha, 100ha and 150ha) and considering wildfires occurred by negligence and accidents (cause 2) and those caused by intention or arson (cause 3). The main factors that have an influence in the presence of wildfires (larger than a given extension of hectares) are the orientation and the land use. Taking into account the rest of the covariates considered we can see that the hill shade, the distance to anthropic areas and the maximum temperature have no influence in the probability of a fire to become larger than a specific area. Table 2 shows the means of the posterior distributions for the hyper-parameters of the first stage considering the three sizes of area analysed. The heterogeneity, the time and the interaction have a small impact and moreover, their values decrease when the extension of the wildfires increases. We can also appreciate that there are not big differences between the two causes. On the other hand, the values of the spatial component show that there is an important spatial dependence, especially for wildfires occurred by negligence and accidents.

In Figures 4 and 5, we show the marginal distribution of hyper-parameters κ, τ, ρ, heterogeneity, time and interaction for Causes 2 and 3. In all of them, the distribution is Gamma, the distributions are similar for both causes. Finally, Figure 6 shows the prediction of the probability of a fire to become larger than 50ha as well as the standard deviation of this prediction. Looking at the wildfires occurred by negligence and accidents we can see that higher probabilities are concentrated around the main urban areas of Catalonia: Girona (in the north-east), Barcelona (in the middle of the coast), Tarragona (in the south along the coast) and Lleida (in the centre west). There are also high probabilities in the north-west, corresponding to a large forest area. With respect to intentional and arson wildfires the probabilities are less concentrated than in wildfires occurred by negligence and accidents but are also higher in the same areas. Regarding the standard deviation we do not appreciate alarming values. On the second cause higher values are found where the probabilities are also higher. The third cause presents lower values of deviation than wildfires occurred by negligence and accidents meaning that the model works better with wildfires occurred by intention or arson.

	Cause 2			Cause 3		
	50	100	150	50	100	150
(Intercept)	X	X	X	X	X	X
factor(Aspect)2						
factor(Aspect)3		X				
factor(Aspect)4				X	X	X
factor(Slope)2						
factor(Slope)4						
factor(Slope)5			X			
factor(Altitude)3	X					
factor(Land use)1	X					
factor(Land use)3		X	X			
factor(Land use)4					X	X
factor(Land use)6						X
ftmin 3					X	
ftmin 5		X				

Table 1. Significative factors for the logistic model in the first stage of the analysis.

	50ha		**100ha**		**150ha**	
	Cause 2	**Cause 3**	**Cause 2**	**Cause 3**	**Cause 2**	**Cause 3**
Heterogeneity	0.000054	0.000054	5.212E-09	5.192E-09	3.959E-09	5.247E-09
Space	0.246900	0.148810	0.3908300	0.0520790	0.0884000	0.0131780
Interaction	0.000043	0.000043	3.885E-09	3.827E-09	3.408E-09	3.762E-09
Time (year)	0.000053	0.000049	5.187E-09	5.135E-09	4.444E-09	4.759E-09

Table 2. Means of the posterior distributions for the hyper-parameters of the first stage.

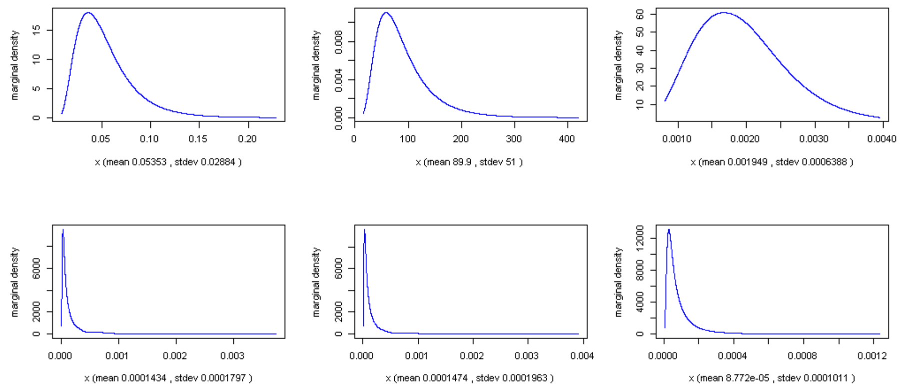


Figure 4. From Top-Left to Bottom-Right: Marginal posterior distribution for κ, τ , ρ, heterogeneity, time and interaction, respectively, for Cause 2.

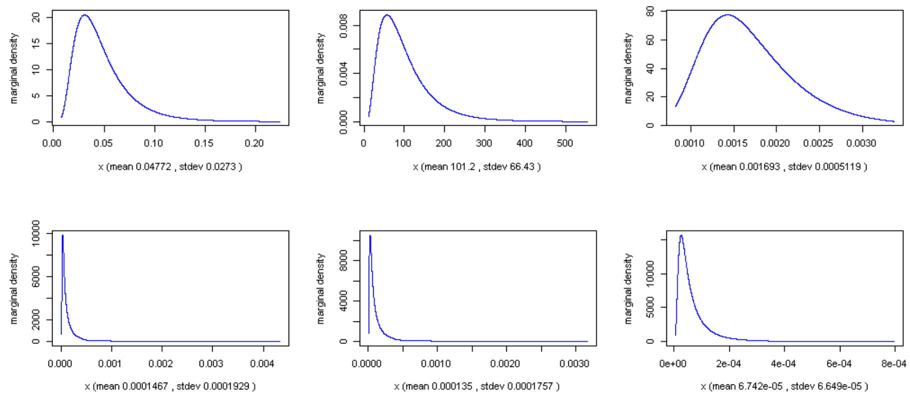


Figure 5. From Top-Left to Bottom-Right: Marginal posterior distribution for κ, τ , ρ, heterogeneity, time and interaction, respectively for Cause 3.

4.2. **Second stage results**. In the second stage we model the frequencies of wildfires (larger than a specific area) per spatial unit. Table 3 shows the values of the hyper-parameters. It is important to note that in this second stage the spatial values are not included. The reason is because there is a too high correlation between the spatial dependence component, S_i, and the spatio-temporal interaction, v_{jt}, that prevents the model from working properly. Therefore, we introduce the spatial random effect through the interaction. The heterogeneity is quite much significant than in the first stage, especially for intentional wildfires and arson. Something similar happens with the interaction. It is much larger than in the first stage and it is also more representative for wildfires occurred by intention and arson. Finally, with respect to the temporal dependence, this is also larger than in the first stage but it has almost no variation between the two causes. In addition there are not relevant differences between the three extensions of hectares in any of the three hyper-parameters analysed. In Figure 7, we show the marginal posterior distribution of hyper-parameters for heterogeneity, time and interaction for Causes 2 and 3. In all of them, the distribution is Gamma. Finally, Figure 8 shows the predicted number of wildfires larger than 50ha per spatial unit. Wildfires occurred by negligence and accidents and those caused by intention or arson present the same pattern of distribution according to the probabilities obtained in the first stage of the model. In general, big wildfires are concentrated along the coast being denser around the metropolitan area of Barcelona. Looking at the standard deviations we point out that intention wildfires and arson have very low values so, again, we note that the model correctly fits wildfires occurred intentionally or arson.

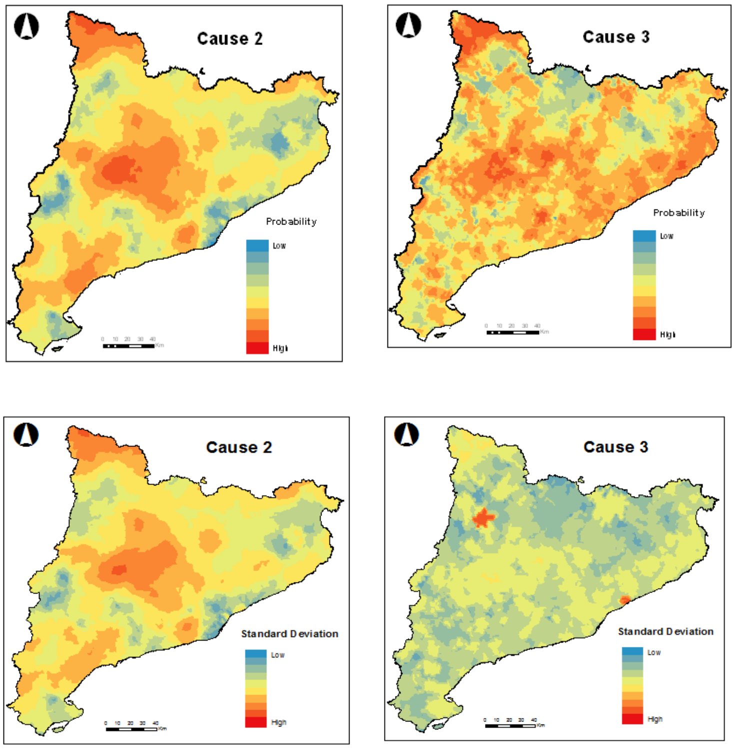


Figure 6. Top: Prediction maps for Cause 2 and Cause 3. Bottom: Standard Deviation for the prediction under Cause 2 and Cause 3.

	50ha		**100ha**		**150ha**	
	Cause 2	**Cause 3**	**Cause 2**	**Cause 3**	**Cause 2**	**Cause 3**
Heterogeneity	0.116645	1.083424	0.116918	1.088495	0.116836	1.089681
Interaction	0.000181	0.010143	0.000177	0.010101	0.000180	0.009634
Time (year)	0.000048	0.000048	0.000047	0.000048	0.000048	0.000040

Table 3. Hyper-parameters for the model in the second stage.

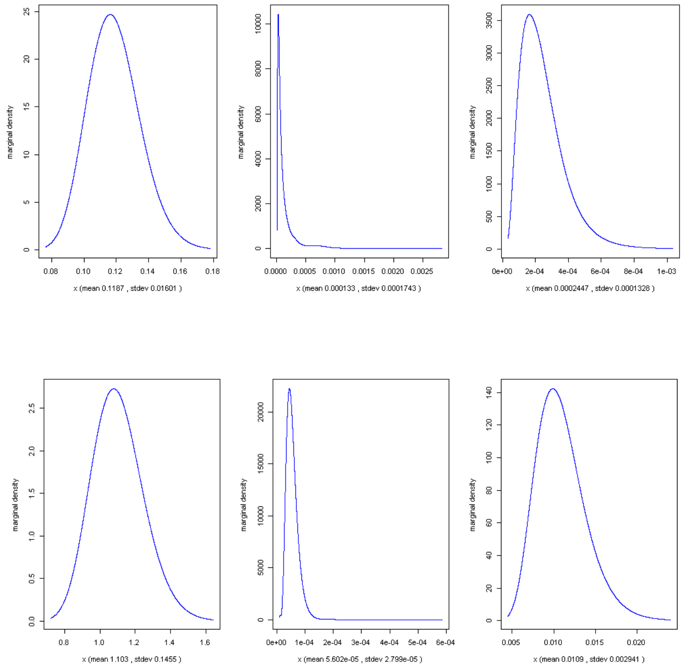


Figure 7.Posterior distribution of the hyper-parameters for the second stage. Left: heterogeneity, Middle: time and Right: interaction. First line: Cause 2, second line: Cause 3.

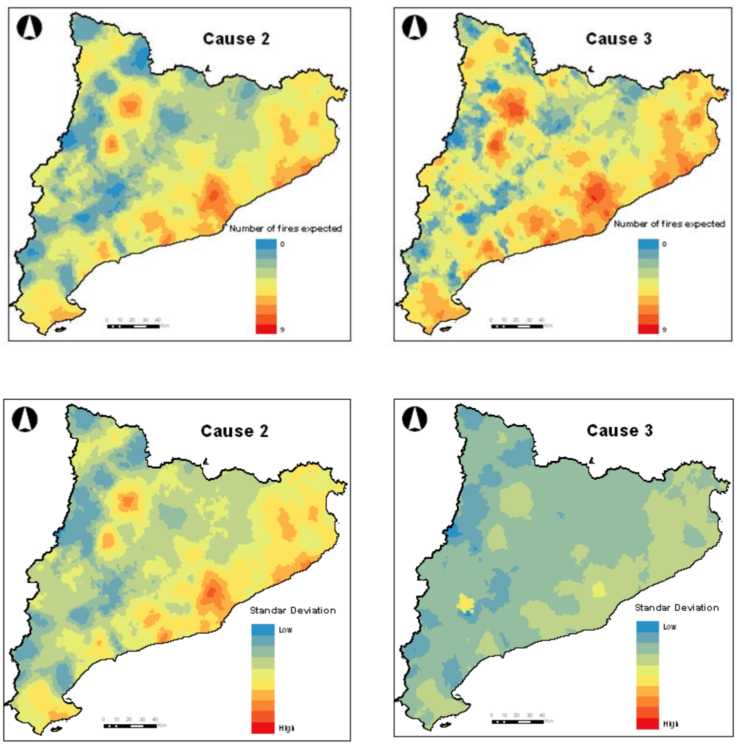


Figure 8. Number of fires expected Maps: On the Top: Cause 2 and Cause 3 and on the Bottom: Cause 2-sd and Cause 3-sd.

5. DISCUSSION

The main finding of this study is that big wildfires are mostly caused by human actions either by negligence and accidents or by intention or arson. These results make sense with what the bibliography shows and what we have commented in the introduction; over 95% of the fires in Europe are due to human causes.

Normally a natural wildfire does not spread as much as an intentional wildfire and so, the number of wildfires which are larger than a big extension, is not enough to obtain results. Analyzing the four causes separately we noticed no significant results for wildfires caused by natural causes and for those caused by unknown causes or rekindled. In fact separating wildfires by cause and by its extension we almost did not have wildfires caused by natural causes nor unknown causes or rekindled. In particular in our data there are only 15 wildfires bigger than 50ha occurred by natural causes compared to 180 caused by negligence or accidents. Our model does not work properly with such a limited small number of data so, even if we have studied the four causes, we have restricted the study to the second and the third causes. To analyse and estimate the number of zeros in a dataset there are different statistical alternatives. On one hand we have the ZIP model, which is 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 [27] 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 the positive values is governed by a zero-truncated count model. In the ZIP models, unlike the hurdle model, there are thought to be two kinds of zeros, ”true zeros” and ”excess zeros”. Although the practical results are very similar in both approaches, hurdle models are most appropriate in our case, since every wildfire can turn into a big wildfire and therefore, every point is susceptible to become larger than a specific number of hectares.

ACKNOWLEDGEMENTS

Work partially funded by grant MTM2010-14961 from the Spanish Ministry of Science and Education.

REFERENCES

[1] Amaral-Turkman, M.A., Turkman, K.F., Le Page. Y., Pereira, J.M.C. (2011). Hierarchical space-time models for fire ignition and percentage of land burned by wildfires. *Environmental and Ecological Statistics*, **18**, 601-617.

[2] Alan A. Ager, Nicole M. Vaillant, Mark A. Finney, Haiganoush K. Preisler (2012). Analyzing wildfire exposure and source-ink relationships on a fire prone forest landscape *Forest Ecology and Management*, **267**, 271-283.

[3] Bachmann A. and Allgower, B. (2001). A consistent wildland fire risk terminology is needed. *Fire Management Today*, **61** (4), 28-33

[4] Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, 1-42.

[5] Berman, M. and Turner, T.R. (1992). Approximanting point process likelihoods with GLM. *Journal of Applied Statistics*, **41**, 31-38.

[6] Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013). Spatial and Spatio-temporal models with R-INLA.*Spatial and Spatio-temporal Epidemiology*, **4**, 33-49.

[7] Chuvieco, E. (2009). *Earth Observation of Wildland Fires in Mediterranean Ecosystems*.Springer-Verlag Berlin Heidelberg.

[8] Comas, C., Palahi, M., Pukkala, T. and Mateu, J. (2009). Characterising forest spatial structure through inhomogeneous second order characteristics.*Stochastic Environmental Research and Risk Assessment*, **23**, 387-397

[9] Comas, C. and Mateu, J. (2011). Statistical inference for Gibbs point processes based on field observations. *Stochastic Environmental Research and Risk Assessment*, **25**, 287-300.

[10] Cressie, N.A.C. (1993). *Statistics for Spatial Data(Revised ed.).* New York: Wiley.

[11] Díaz-Avalos, C., Peterson, D. L., Alvarado C. E., Ferguson, Sue A. and Besag, J.E. (2001). Space-time modelling of lightning-caused forest fires in the Blue Mountains, Oregon. *Canadian Journal of Forest Research*, **31**(9), 1579-1593.

[12] Daley, D. and Vere-Jones, D. (2003). An Introduction to the Theory of Point Processes, 2nd edition. Volume I. Springer, New York.

[13] Deb, D.P. and Trivedi, P.K. (2002). The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics*, **21**, 601-625.

[14] Diggle, PJ. (2003). *Statistical Analysis of Spatial Point Patterns*. (2nd ed). Arnold, London UK.

[15] Dillon, G.K., Holden, Z.A., Morgan, P., Crimmins, M.A., Heyerdahl, E.K. and Luce, C.H. (2011). Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. *Ecosphere*, **2** (12), 1-33.

[16] Flannigan, M. and Wotton, B. (2001). *Climate, weather, and area burned*. In Forest Fires: Behavior and Ecological Effects E. Johnson and K. Miyanishi (eds), 351-373. San Diego: Academic Press.

[17] 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.

[18] Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn Cross-Covariance functions for multivariate random fields. *Journal of the American Statistical Association*, **105** (491), 1167-1177.

[19] Gonzalez, J.R. and Pukkala, T. (2007). Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research*, **126**, Issue 3, 421-429.

[20] Harvill, J. L. (2010). Spatio-temporal processes. *WIREs Computational Statistics*, **2**, 375-382.

[21] Hirsch, R.E., Lewis, B.D., Spalding, E.P., and Sussman, M.R. (1998). A role for the AKT1 potassium channel in plant nutrition.*Science*, **8**, 918-921.

[22] Illian, J.B. and Hendrichsen, D.K. (2010). Gibbs point process models with mixed effects. *Environmetrics*, **21**, 341-353.

[23] Illian, J.B., Sorbye, S.H. and Rue, H. (2012). A toolbox for fitting complex spatial point processes models using integreted nested Laplace approximations (INLA). *The Annals of Applied Statistics*, **6**, Issue 4, 1499-1530.

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

[25] Lindgren, F., Rue, H. and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields the SPDE approach. *Journal of the Royal Statistical Society, Series B*, **73**, 423-498.

[26] 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.

[27] Mullahy, J. (1986). Specification and Testing of Some Modified Count Data Models.*Journal of Econometrics*, **33**, 341-365.

[28] Neelon, B., Ghosh, P. and LoebsMullahy, P.F. (2013). A spatial Poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of the Royal Statistical Society. Series A*, **176** (2), 389-413.

[29] Ordóñez, C., Saavedra, A., Rodríguez-Pérez, J.R., Castedo-Dorado, F. and Covin, E. (2012). Using model-based geostatistics to predict lightning-caused wildfires.*Environmental Modelling and Software*, **29** (1), 44-50.

[30] Piñol, J., Terradas, J. and Lloret, R. (1998). Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. *Climate Change*, **38**, 345-357.

[31] Plummer, M., Penalized, L. (2008). Functions for Bayesian model Comparation. *Biostatistics*, **9** (3), 523-539.

[32] R Development Core Team. (2011). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, http://www.r-project.org/.

[33] R-INLA project [available in: http://www.r-inla.org/home, accessed on August 13, 2012].

[34] Reus Dolz, M.L. and Irastorza, F. (2003). Estado del Conocimiento de causas sobre los incendios forestales en España. *APAS and IDEM Estudio sociologico sobre la percepcion de la población española hacia los incendios forestales*. www.idem21.com/descargas/pdfs/IncediosForestales.pdf.

[35] Røder, 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 actors. A case study in the Ayora region (eastern Spain).*Remote Sensing of Environment*, **112** (1), 259-273.

[36] Rue H., Martino S. and Chopin N. (2009). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society*, Series B, **71**, 319-392.

[37] Saez, M., Barcelo, M.A., Tobias, A., Varga, D., Ocaña-Riola, R., Juan, P. and Mateu, J. (2012). Space-time interpolation of daily air temperatures.*Journal of Environmental Statistics*, **3** (5).

[38] San-Miguel-Ayanza, J., Rodrigues, M., Santos de Oliveira, S., Kemper Pacheco, C., Moreira, F., Duguy, B. and Camia, A. (2012). Land cover change and fire regime in the European Mediterranean region. *Post-Fire Mangement and Restoration of Southern European Forests Managing Forest Ecosystems*, F. Moreira, M. Arianoustsou, P. Corona, J. de las Heras (Eds.), Springer, Berlin, Heidelberg. 21-43.

[39] San-Miguel-Ayanza, J., Moreno, J. M. and Camia, A. (2013). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*, **294**, 11-22.

[40] Silva, J.S., Vaz, P., Moreira, F., Catry, F. and Rego, F.C. (2011). Wildfires as a major driver of landscape dynamics in three fire-prone areas of Portugal.*Landscape and Urban Planning*, **101**, Issue 4, 349-358.

[41] Simpson, D., Illian, J., Lindgren, F., Sorbye, S.H. and Rue, H. (2011). Going off grid: computationally efficient inference for log-Gaussian Cox processes.Technical Report, Trondheim University.

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

[43] Tiernery, L. and Kadane, J.B. (1986). Accurate Approximations for posterior moments and marginal densities.*Journal of the American Statistical Association*, **81**, 82-86.

[44] Turner, R. (2009). Point patterns of forest fire locations. *Environmental and Ecological Statistics*, **16**, 197-223.

[45] Wang, Z., Ma, R. and Li, S. (2012). Assessing area-specific relative risks from large forest fire in Canada. *Environmental and Ecological Statistics*, To appear.

[46] Wisdom M. Dlamini, A. (2010). Bayesian belief network analysis of factor influencing wildfire occurrence in Swaziland.*Environmental Modelling and Software*, **25**, Issue 2, 199-208.

[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.

Chapter 4. Discussion

The first part of this Thesis is restricted to inhomogeneous spatial models, where the temporal scale is fixed, and only the spatial component is modeled. After the first results, a second analysis includes time into the model, thus considering spatio-temporal point process models. One good thing about this approach is that we were able to model and evaluate the corresponding spatio-temporal interaction. We are aware of some approaches which had already considered this modelling, but they consider independent spatial replication in time which is not realistic. In the context of spatio-temporal modelling, one useful approach is to model the spatio-temporal intensity function as an additive or multiplicative form of the spatial and temporal intensities, and then adding a spatio-temporal residual component (for example Diggle et al. 2005).

During this work we have started using some covariates (slope, aspect, altitude, hill shade and land use) and we have completed the analysis adding more covariates such as proximity to anthropic areas and climatic variables (maximum and minimum temperatures). Climatic variables could explain the spatial structure but we are not sure on what drives the temporal variation of wildfires occurrences over time. However, we can note that land use varies over time and it has an effect on the temporal variation of the wildfire counts.

Knowing that models for forest fire occurrence have been studied using different approacheswe have chosen 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 is 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 have used in our analysis has allowed us to find the class of models that best fits the occurrence of wildfires distinguishing by cause.

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 have also considered the time component and the interaction between space and time, and we have not considered any element that follows a CAR model. On contrary, we have modeled the spatial correlation by the Matérn spatial covariance function using a regular lattice through stochastic partial differential equation (SPDE)to transform the initial Gaussian Field to a Gaussian Markov Random Field (GMRF).

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 previous than 2011 and they do not take advantage of the current SPDE approach (Krainski, 2013) as we have done.

Finally, it is worth noting that efforts to suppress wildfires have become an important problem in recent 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 changes in wildfires behaviour could cause catastrophic damages as human lives, economics and environmental losses.

The analysis of wildfire incidence in Catalonia presented in this Thesis provides important clues as to which risk factors are associated with which different causes. The results of our analysis have provided a deeper insight into factors associated with wildfire incidence in Catalonia.

A future work might incorporate more covariates related to the occurrence of wildfires such as humity or wind. Moreover, a further research should focus on considering space and time separable instead of approaching no separability by means of the interaction between them.

Chapter 5.Conclusions

Throughout this work we have achieved several findings. First, the extent of clustering in wildfires differs through the years and through the considered causes of wildfire ignitions. Second, covariates such as land use, slope, aspect and hill shade influence the trends in the intensity of wildfire locations. Third, spatio-temporal point process is the most sensible approach to model forest fires occurrence. Fourth, Hurdle model is the best to model the occurrence of big wildfires (wildfires greater than a given extension of hectares: 50ha, 100ha or 150ha). Finally, maps of wildfire risks built from the estimated models, by year and cause of ignition, are useful tools in preventing and managing vulnerability levels.

All these settlements are specyfied below following the objectives described at the beginning of this Thesis:

Specific objective 1.–(1) To evaluate how the extend of clustering in wildfires differs across the years they occurred.

From the first article we have concluded that wildfires are not random in space or time and that, despite the variability found among marks, especially over time, the model that best fits the spatial distribution ofwildfires is the area-interaction point process model. In addition, the analysis of wildfire incidence in Catalonia has provided important clues as to which risk factors are associated with which different causes.

Specific objective 2, 3 and 5.–(2) To analyse the influence of covariates on trends in the intensity of wildfire locations.(3) To analyse the spatio-temporal patterns produced by those wildfire incidences by considering the influence of covariates on trends in the intensity of wildfire locations. (5) To build maps of wildfire risks, by year and cause of ignition, in order to provide a tool for preventing and managing vulnerability levels.

From the second articlewe have foundthat covariates affect differently depending on the cause of the wildfire. On the one hand, wildfires started either through negligence and accidents or intentionally are associated with low elevation locations, which are easily accessible to most people, particularly arsonists. In addition, the relative risk of wildfires caused by negligence or accident is lower than 1 for locations far from urban areas, roads and railways due to the lower human presence and activities in such locations. On the other hand, for wildfires caused by nature we have conclude that 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 wildfire. This, added to the lower human presence in such locations, facilitates the spreading of wildfire without control. An increased gradient in the relative risk is 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. 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 those wildfires caused by unknown causes or rekindled, elevation is the only covariate which does not have any significant influenceon trends in the intensity of wildfire locations. However, it must be said, that elevation and distance from urban areas should be correlated, which may make it difficult to attribute single factors to wildfire occurrence. This complex model structure is most likely due to the fact that here we have a mix of wildfires from all of the different causes.

In addition, throughout this second article we have proved that there is a spatio-temporal interaction and that clear different characteristics exist between the distributions of wildfires, depending on each cause.

Specific objective 4, 6 and 7.–(4) To model the occurrence of big wildfires (greater than a given extension of hectares) using an adapted two-part econometric model, specially a Hurdle model. (6) To analyse which factors have more influence in generating wildfires bigger than a given extension (50ha, 100ha or 150ha).(7) To evaluate two different statistical alternatives (ZIP models and Hurdle models) to analyse and estimate the excess of zeros of a stochastic process.

From the third article we have justified that big wildfires are mostly caused by human actions either through negligence and accidents or intentionally but not by natural causes. Analysing the four causes separately we noticed no significant results for wildfires caused by natural causes and for those caused by unknown causes or rekindled.

Furthermore we have concluded that among different statistical alternativesto analyse and estimate the number of zeros in a dataset, as ZIP models, Hurdle models are most appropriate in analysing big wildfires occurrence, since every wildfire can turn into a big wildfire and therefore, every point is susceptible to become larger than a specific number of hectares.

Summarysing, the main conclusions of this Thesis are:

1. Wildfires are not random in space or time and so we are able to model them.

2. The traditional methodology of spatial statistics, which the model that best fits the pattern of wildfires is the area-interaction, has shown that there are variability in space and time. This conclusion has made possible to apply a spatio-temporal methodology using mixed models.

3. The spatio-temporal mixed model used to analyse the occurrence of wildfires in Catalonia is a new approach which allow quantifying and assessing possible spatial relationships between the distribution of risk of ignition and causes.

4. Big wildfires are not attributed to natural causes and the best model to analyse them is the Hurdle model.

5. The methodology used through this Thesis may be useful in fire management decision-making and planning.and so may contribute to the prevention and management of wildfires

Chapter 6.References

Agee, JK. *Fire Ecology of Pacific Northwest Forests*. Washington DC: Island Press, 1993.

Ager AA, Vaillant NM, Finney MA, Preisler HK. Analyzing wildfire exposure and source-ink relationships on a fire prone forest landscape *Forest Ecology and Management*2012; 267:271-283.

Amaral-Turkman MA, Turkman KF, Le Page Y, Pereira JMC. Hierarchical spatio-temporal models for fire ignition and percentage of land burned by wildfires.*Environmental and Ecological Statistics*2011; 18:601-617.

Anderson HE.Forest fuel ignitability.*Fire Technology* 1970; 6:312-319.

Arunachalam V.Point processes, product densities and cellular biology [in Spanish]. *Revista Colombiana de Estadística* 2005; 28(1):1-16.

Bachmann A, Allgower B.The need for a consistent wildfire risk terminology. Neuenschwander L, Ryan K, Golberg G. (Eds*.).Crossing the Millenium:Integrating SpatialTechnologies and Ecological Principles for a New Age in Fire Management*,TheUniversity of Idaho and the International Association of Wildland Fire, Moscow, ID 2000; 67-77.

Bachmann A, Allgower B. A consistent wildland fire risk terminology is needed. *Fire Management Today*2001; 61 (4):28-33.

Baddeley AJ, van Lieshout MNM. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* 1995; 47:601-619.

Baddeley AJ, Moller J, Waagepetersen R. Non and semi- parametric estimation of interactions in inhomogeneous point patterns. *Statistica Neerlandica* 2000; 54: 329-350.

Baddeley AJ, Turner R. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 2005; 12:1-42.

Badia A, Saurí D, Cerdan R, Llurdés KC. Causality and management of forest fires in Mediterranean environments: an example from Catalonia. *Global Environmental Change Part B: Environmental Hazards* 2002; 4(1): 23-32.

Basset JM. Flashover: development and control [in Spanish] [Available at: http://www.kume.cl/KFiles/File/Fuego_y_flashover.pdf, accessed on January 23, 2011].

Beringer J. Community fire safety at the urban/rural interface: The bushfire risk. *Fire Safety Journal* 2000; 35:1-23.

Beringer J, Hutley LB, Tapper NJ, Coutts A, Kerley A, O'Grady A.P. Fire impacts on surface heat, moisture and carbon fluxes from a tropical savanna in northern Australia. *International Journal of Wildland Fire*2003; 12(4):333–340.

Berman M, Turner TR. Approximating point process likelihoods with GLM. *Journal of Applied Statistics* 1992; 41:31-38.

Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion).*Journal of the Royal Statistical Society, Series B* 1974; 36:192-236.

Blanco J, García D, Castellnou M, Molina D, Grillo F, Pous E. *Basic Course on Wildfires* [in Spanish]. Universitat de Lleida, Cabildo de Gran Canaria, Generalitat de Catalunya, Spain, 2008[Available at: http://www.slideshare.net/gtfsaltominho/curso-basico-de-incendios-forestales, accessed on May 13, 2011].

Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and Spatio-temporal models with R-INLA.*Spatial and Spatio-temporal Epidemiology* 2013; 4:33-49.

Boada M. Catalunya, land of forests without knowing [in Catalan]. *Món Sostenible* July 2012 [Available at: http://www.monsostenible.net/catala/opinions/catalunya-pais-de-boscos-sense-saber-ho/, accessed on November 23, 2010].

Boulandier J, Esparza F, Garayoa J, Orta C, Anitua P. Fire fighting Manual [in Spanish]. Pamplona: Gobierno de Navarra, 2001 [Available at: http://www.bomberosdenavarra.com/documentos/ficheros documentos/indices.pdf, last accessed August 1, 2013].

Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*. New Jersey: Holden-Day, 1976.

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

Breslow NE. Extra-Poisson variation in log-linear models.*Applied Statistics* 1984; 33(1):38-44.

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

Butry DT, Prestemon JP. Spatio-temporal wildland arson crime functions. *American Agricultural Economics Association* 2005 [Available at: http://ageconsearch.umn.edu/handle/19197, accessed on March 7, 2011].

Cardille JA, Ventura SJ, Turner MG. Environmental and social factors influencing wildfires in the Upper Midwest, United States. *Ecological Applications* 2001; 11(1):111–127.

Carey H, Schumann M. *Modifying wildfire behaviour: The effectiveness of fuel treatments: the status of our knowledge*. Santa Fe, Southwest Region: National Community Forestry Centre, 2003.

Carmo M, Moreira F, Casimiro P, Vaz P. Land use and topography influence on wildfire occurrence in northern Portugal. *Landscape and Urban Planning* 2011; 100:169–176.

Castellnou M, Arilla E, López M. Fire management in Catalonia: first steps. *International conference on prevention strategies of fires in southern Europe* Barcelona 2005; 1-9.

Castellnou M, Nebot E, Miralles M. Therole of fire inlandscape management.*IV International Wildfire Fire Conference* Sevilla 2007 [Available at: http://www.fire.uni-freiburg.de/sevilla-2007/contributions/doc/SESIONES_TEMATICAS/ST1/Castellnou_et_al__2_SPAIN_Cat.pdf , last accessed August 1, 2013].

Castellou M, Miralles M. The changing face of wildfires.*Crisis Response Journal* 2009; 5(4):56-57.

Castellnou M, Pagés J, Larrañaga A, Piqué M. *Standard fire risk map of Catalonia* [in Catalan]. GRAF-Bombers. Departament d'Interior. Generalitat de Catalunya i Centre Tecnològic Forestal de Catalunya, 2010.

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

Chuvieco E. *Earth Observation of Wildland Fires in Mediterranean Ecosystems*.Berlin Heidelberg: Springer-Verlag, 2009.

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

Comas C, Palahi M, Pukkala T, Mateu J. Characterising forest spatial structure through inhomogeneous second order characteristics. *Stochastic Environmental Research and Risk Assessment* 2009; 23:387-397.

Comas C, Mateu J. Statistical inference for Gibbs point processes based on field observations. *Stochastic Environmental Research and Risk Assessment* 2011; 25:287-300.

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

Cressie NAC *Statistics for Spatial Data (Revised ed.)*. New York: Wiley, 1993.

Daley D, Vere-Jones D.*An Introduction to the Theory of Point Processes*.New York: Springer Volume 2,2003.

Davis IB.The wildland/urban interface.Paradise or battleground?*Journal of Forestry*1990; 26-31.

De la Cruz Rot M. Introduction to mapped data analysis or some of the (many) things that I can do if I have coordinates [in Spanish].*Ecosistemas* 2006; 15(3):19-39.

Deb DP, Trivedi PK. The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* 2002; 21:601-625.

Díaz-Avalos C, Peterson DL, Alvarado E, Ferguson SA, Besag JE. Space-time modelling of lightning-caused forest fires in the Blue Mountains, Oregon. *Canadian Journal of Forest Research* 2001; 31(9):1579-1593.

Díaz-Delgado R, Pons X. 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* 2001; 147(1):67-74.

Diggle PJ. *Statistical Analysis of Spatial Point Patterns*.London: Academic Press, 1983.

Diggle PJ. *Statistical Analysis of Spatial Point Patterns (2nd Edition).*London UK: Arnold, 2003.

Diggle P, Zheng P, Durr P. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Journal of the Royal Statistical Society, Series C, Applied Statistics* 2005; 54(3):645-658.

Diggle P, Rowlingson B, Su TL. Point process methodology for on-line spatio-temporal disease surveillance.*Environmetrics* 2005; 16:423-434.

Dillon GK, Holden ZA, Morgan P, Crimmins MA, Heyerdahl EK, Luce CH. Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. *Ecosphere*2011; 2(12):1-33.

Dixon PM. Ripley’s K function.*Encyclopedia of Environmetrics* 2002; 3:1796-1803.

DMAH, 2002.*Map Basic Wildfire Danger* [in Catalan]. Departament de Medi Ambient Habitatge, Generalitat de Catalunya.[Available at http://www.gencat.cat/medinatural/incendis/mapaperill/, last accessed August 1, 2013].

DMAH, 2005b.*Fire Risk Map On-line* [in Catalan]. Departament de Medi Ambient i Habitatge, Generalitat de Catalunya [Available at http://www.gencat.cat/medinatural/incendis/mapaperill/, last accessed August 1, 2013].

Drever CR, Drever MC, Messier C, Bergeron Y, Flannigan M. Fire and the relative roles of weather, climate and landscape characteristics in the Great Lakes-St. Lawrence forest of Canada, *Journal of Vegetation Science*. 2008; 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, accessed on April 8, 2012].

European Environment Agency.*CLC2006 technical guidelines*. Luxembourg: Office for Official. Publications of the European Communities, 2007.

Flannigan MD, Wotton B. Climate, weather, and area burned. *Forest Fires: Behavior and Ecological Effects* 2001; 10:351-373.

Flannigan MD, Amiro BD, Logan KA, Stocks BJ, Wotton BM. Forest fires and climate change in the 21st century. *Mitigation and Adaptation Strategies for Global Change* 2006; 11(4):847-859.

Fuentes Santos I. *Statistical inferenceforspatialpoint processes. Application to the analysisof forest fires inGalicia* [in Spanish].Master’s degree dissertation. University of Coruña, Spain: 2008-2009.

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

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

Geisser S. *Predictive Inference: An Introduction*. London: Chapman & Hall, 1993.

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

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

Gonzalez JR, Pukkala T. Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research* 2007; 126:421-429.

Graham RT, McCaffrey S, Jain TB. *Science basis for changing forest structure to modify wildfire behavior and severity*. US Department of Agriculture Forest Service, 2004 [Available at: http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1161&context=barkbeetles, accessed on February 7, 2011].

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

Hardy CC. Wildland fire hazard and risk: Problems, definitions, and context. *Forest Ecology and Management* 2005; 211(1-2):73-82.

Harvill JL. Spatio-temporal processes. *WIREs Computational Statistics*2010;2:375-382.

Held L, Schrödle B, Rue H. Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA.*Statistical Modelling and Regression Structures*2009; 91-110 [Available at http://www.r-inla.org/papers, accessed on July 18, 2012].

Heymann Y, Steenmans CH, Croissille G, Bossard M. *CORINE Land Cover. Technical Guide*. Luxembourg: Office for Official Publications of the European Communities, 1994.

Hirsch RE, Lewis BD, Spalding EP, Sussman MR. A role for the AKT1 potassium channel in plant nutrition. *Science* 1998; 8: 918-921.

IDESCAT.*Basic Statistics of Catalonia. Demography and Quality of Life* [in Catalan]- Institut d'Estadística de Catalunya (IDESCAT) 2011 [Available at http://www.idescat.cat/dequavi/Dequavi?TC=444&V0=1&V1=1, accessed on October 23, 2011].

Illian JB, Hendrichsen DK. Gibbs point process models with mixed effects. *Environmetrics*2010; 21:341-353.

Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical analysis and modelling of spatial point patterns*. New York: Wiley, 2008.

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

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

Jensen JL, Møller J. Pseudo-likelihood for exponential family models of spatial point processes. *Annals of Applied Probability* 1991; 3:445-461.

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

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

Krainski ET. The R-INLA tutorial: SPDE models 2013 [Available at: http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf,accessed on May 13, 2013].

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

Latent models.The R INLA project.[Available at http://www.r-inla.org/models/latentmodels, accessed on August 13, 2012].

Lewis PAW, Shedler GS. Simulation of nonhomogenous Poisson processes by thinning. *Naval Research Logistics Quarterly* 1979; 26(3):403-413.

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

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

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

Manualsurvivalin the laboratory [in Spanish].Universidad de Alicante. Alicante, 2007.

Martino S, Rue H. *Case Studies in Bayesian Computation using INLA*in Complex data modeling and computationally intensive statistical methods(R-code) 2010 [Available at http://www.r-inla.org/papers, accessed on April 18, 2011].

Mateu J, Montes F.Modelling thespatial distributionof cysts in the porpoise stomachby Gibbs process [in Spanish].*Qüestiió* 1998; 22(1):175-194.

McRae RHD. Prediction of areas prone to lightning ignition.*International Journal of Wildland Fire* 1992; 2:123-130.

Millington JDA. Wildfire risk mapping: considering environmental change in space and time. *Journal of Mediterranean Ecology* 2005; 6(1):33-44.

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

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

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

Møller J, Díaz-Avalos C. Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics* 2010; 37:2-15.

Møller J, Waagepetersen RP. *Statistical inference and simulation for spatial point processes*. New York: CRC Press, 2003.

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

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

National Interagency Fire Centre.*The Science of Windland* Fire [Available at: http://www.nifc.gov/, accessed on February 14, 2011].

National Wildfire Coordinating Group.*Wildfire Prevention Strategies* March 1998 [Available at: http://www.nwcg.gov/pms/docs/wfprevnttrat.pdf, accessed on February 14, 2011].

Neelon B, Ghosh P, LoebsMullahy PF. A spatial Poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of the Royal Statistical Society. Series A* 2013;176(2):389-413.

O'Donnell AJ. *Spatial and temporal patterns of wildfires in semi-arid south-western Australia*.PhD dissertation.University of Western, Australia: 2011.

Ohser J. On estimators for the reduced second moment measure of point processes.*Mathematische Operationsforschung und Statistik, series Statistics* 1983; 14.

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

Papangelou F. The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeitsth* 1974; 28:207-226.

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

Piñol J, Terradas J, Lloret F. Climate warming wildfire hazard, and wildfire occurrence in coastal Eastern Spain.*Climate Change* 1998; 38:345-357.

Piqué M, Castellnou M, Valor T, Pagés J, Larrañaga A, Miralles M, Cervera T. *Integrationof the riskof large firesForestals(GIF) in forest management: Firetype andvulnerability of structuresheadersforestfires*. Series:Sustainable Forest ManagementGuidelinesfor Catalonia (ORGEST) [in Catalan].Centre de la Propietat Forestal. Departament d'Agricultura, Ramaderia, Pesca, Alimentació i Medi Natural. Generalitat de Catalunya 2011.

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

Pyne SJ, Andrews PL, Laven RD. *Introduction to Wildland fire.* New York: Wiley, 1996.

R Development Core Team.*R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.ISBN 3-900051-07-0.[Available at http://www.r-project.org/, 2010, accessed on September 24, 2011].

R-INLA project. [Available at: 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. 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* 2012; 11:4.

Reus Dolz ML, Irastorza F. Stateof knowledgeof causesofforest fires inSpain. Sociologicalstudyon theperception oftheSpanishpopulationtoforest fires. APAS, 2003 [in Spanish] [Available at: http://www.idem21.com/descargas/pdfs/CAUSAS_IF.pdf, accessed on June 8, 2010].

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

Rifà A, Castellnou M. *The Catalan model of forest fires extinction* [in Spanish]. IV International Wildfire Fire Conference, Sevilla (Spain), 2007.

Ripley BD. The second-order analysis of stationary point processes. *Journal of Applied Probability* 1976; 13:256-266.

Ripley BD. Modelling spatial patterns.*Journal of the Royal Statistical Society, Series B* 1977; 39:172-212.

Ripley BD. *Statistical inference for spatial processes*.Cambridge CB2 1RP: Cambridge University Press, 1988.

Ripley BD, Kelly FP. Markov point processes. *Journal of the London Mathematical Society* 1977; 15:188-192.

Röder A, Hill J, Duguy B, Alloza JA, Vallejo R. 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* 2008; 112(1):259-273.

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

RothermelRC. *A mathematical model for predicting fire spread in wildland fuels*. Research paper INT-115. Ogden, UT: U.D. Department of Agriculture, Intermountain Forest and Range Experiment Station, 1972.

Rowlingson B, Diggle P. Splancs: spatial point pattern analysis code in S-Plus. *Computers and Geosciences* 1993; 19:627-655; (the original sources can be accessed at: http://www.maths.lancs.ac.uk/~rowlings/Splancs/).

Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 2009; 71(2):319-392[Available at http://www.math.ntnu.no/~hrue/r-inla.org/papers/inla-rss.pdf, accessed on May 21, 2011].

Saez M. The problem of repeated measures.Longitudinal epidemiologycal analysis [in Spanish].*Gaceta Sanitaria* 2001; 15(4):347-352.

Saez M, Saurina C. *Spatial Statistics and Epidemiology* [in Spanish]. Girona: Documenta Universitaria, 2007.

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

San-Miguel-Ayanza J, Rodrigues M, Santos de Oliveira S, Kemper Pacheco C, Moreira F, Duguy B, Camia A. Land cover change and fire regime in the European Mediterranean region.In Moreira F, Arianoustsou M, Corona P, de las Heras J, editors.Post-Fire Mangement and Restoration of Southern European Forests.Berlin, Heidelberg: Springer; 2012; 4: 21-43.

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

Scott A. The Pre-Quaternary history of fire.*Palaeogeography, Palaeoclimatology, Palaeoecology* 2000; 164:297-345.

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

Shen G, Mingjian Y, Xin-Sheng H, Xiangcheng M, Haibao R, I-Fang S, Ma K. Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology* 2009; 90:3033-3041.

Silva JS, Vaz P, Moreira F, Catry F, Rego FC. Wildfires as a major driver of landscape dynamics in three fire-prone areas of Portugal.*Landscape and Urban Planning* 2011; 101(4):349-358.

Simpson D, Illian J, Lindgren F, Sørbye SH, Rue H. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Technical Report Trondheim University* 2011. [Available at: http://www.math.ntnu.no/~daniesi/S10-2011.pdf, accessed on August 7, 2012].

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

Taylor BM, Diggle PJ.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* 2013.

The population in Catalonia, 2011 [in Catalan]: [Available at: http://blocs.xtec.cat/legosocials/2009/01/29/la-poblacio-a-catalunya/, accessed on March 15, 2012).

Thompson WA, Vertinsky I, Schreier H, Blackwell BA. Using forest fire hazard modelling in multiple use forest management planning.*Forest Ecology and Management* 2000; 134:163-176.

Tiernery L, Kadane JB. Accurate Approximations for posterior moments and marginal densities.*Journal of the American Statistical Association*1986;81:82-86.

Turner R. Point patterns of forest fire locations. *Environmental and Ecological Statistics*2009; 16:197-223.

Van Lieshout MNM.*Markov Point Processes and Their Applications*.London: Imperial College Press, 2000.

Varga D. *Landscape and Agricultural Abandonment in the Mediterranean Mountains.An approximation to the Hortmoier and Sant Aniol Valleys ('Alta Garrotxa') from the Landscape Ecology.*PhD dissertation. University of Girona, Spain:2007.

Vélez R. *Thefire management.Foundationsand experiences* [in Spanish]. Madrid: M^{c}Graw Hill. Interamericana de España S.L., 2000.

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

Wang Z, Ma R, Li S. Assessing area-specific relative risks from large forest fire in Canada.*Environmental and Ecological Statistics*2012, To appear.

Widom B, Rowlinson JS. New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics* 1970; 52:1670–1684.

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, accessed on November 4, 2011].

Winter G, Fried JS. Homeowner perspectives on fire hazard, responsibility, and management strategies at the wildland-urban interface.*Society and Natural Resources* 2000; 13(1):33-49.

Wisdom M. Dlamini, A. Bayesian belief network analysis of factor influencing wildfire occurrence in Swaziland. *Environmental Modelling and Software* 2010; 25(2):199-208.

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

Zheng P, Diggle P, Gómez-Rubio V. *Package 'spatialkernel'*, version 0.4-10, February 15, 2012 [Available at: http://cran.r-project.org/web/packages/spatialkernel/spatialkernel.pdf, accessed on March 19, 2012].