

Spatio-temporal modelling of wildfires in Catalonia, Spain, 1994–2008, through log Gaussian Cox processes

L. Serra^{1,2}, M. Saez^{1,2}, D. Varga^{1,3}, A. Tobías^{1,4},
P. Juan⁵ & J. Mateu⁵

¹*Research Group on Statistics, Applied Economics and Health (GRECS),
University of Girona, Spain*

²*CIBER of Epidemiology and Public Health (CIBERESP), Spain*

³*Geographic Information Technologies and
Environmental Research Group, University of Girona, Spain*

⁴*Institute of Environmental Assessment and Water Research (IDAEA),
Spanish Council for Scientific Research (CSIC), Barcelona, Spain*

⁵*Department of Mathematics, Campus Riu Sec,
University Jaume I of Castellon, Spain*

Abstract

Forest fire management is not only an emergency task, the preventive task could be even more important, being better to avoid the risk of a forest fire ignition before it starts or minimize its hazard, rather than later trying to extinguish it. If we associate wildfires with their spatial coordinates, along with other variables, it is possible to identify them by means of a spatio-temporal stochastic process. Spatio-temporal clustering of wildfires could indicate the presence of risk factors. In fact, what is usually of interest is to assess their dependence on covariates. Two were the objectives in this paper. Firstly, to evaluate how the extent of clustering in wildfires differs across marks. Secondly, to analyze the influence of covariates on trends in the intensity of wildfire locations. We analyzed the spatio-temporal patterns produced by wildfire incidences in Catalonia, located in the north-east of the Iberian Peninsula. The total number of fires recorded in the studied area, during the period 1994–2008, was 10,783. In addition to the locations of the fire centroids, several marks and spatial covariates were considered. We specified spatio-temporal log-Gaussian Cox



process models. Models were estimated using Bayesian inference for Gaussian Markov Random Field (GMRF) through the Integrated Nested Laplace Approximation (INLA) algorithm. The results allow us to quantify and assess possible spatial relationships between the distribution of risk of ignition and possible explanatory factors. We believe the methods shown in the paper may contribute to the prevention and management of wildfires, which are not random in space or time.

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

1 Introduction

Forest fires are now the main cause of forest destruction in the countries of the Mediterranean basin causing enormous economic and ecological damage as well as loss of human life. Fire risk is seriously important in the Mediterranean region due to its marked seasonality, with high temperatures and low humidity in prolonged summers (extending from June to October and sometimes even longer). These extreme conditions allow even a small heat source (lightning, a spark, match, a cigarette butt) to set off a violent and dangerous fire. Together with the heat and lack of humidity of this season, the inland summer winds, characterized by its great speed and strong desiccant power, such as Tramuntana in Catalonia, help spread the fires, transporting them quickly from one place to another. Also dry and cold winds of winter increase the danger of fire [1].

Weather conditions determine the composition of Mediterranean forests. For this reason, due to the prolonged summer droughts, most of these forests have stabilized based on species that need fire during their reproductive cycle. Pine forests are large masses on both sides, north and south of the Mediterranean. In particular, the Aleppo pine (*Pinus halepensis*) is the most widespread on the coasts of Spain, France, Italy, Greece, Turkey, Morocco, Algeria and Tunisia. These species are characterized by physiological mechanisms that connect natural reproduction with fire, for example the opening of the cones by the intense heat. These species may also have a high content of resin and essential oils, extremely flammable [1].

In the particular case of Catalonia, a region located in the northeast of the Iberian Peninsula which represents 6.4% of the Spanish national territory, it is necessary to take into account that there is also an important process of afforestation of the different agricultural areas as well as an increase of the abandonment of rural activities. For everything, Catalonia is in a situation of extreme vulnerability to the risk of fires.

Forest fires management is not only an emergency task, the preventive task could be even more important, being better avoid risk of a forest fire ignition before it start or minimize its hazard, rather than later try to extinguish it.

If we associate wildfires with their spatial coordinates, along with other variables, it is possible to identify them by means of a spatio-temporal stochastic process. Spatio-temporal clustering of wildfires could indicate the presence of risk factors. In fact, what is usually of interest is to assess their dependence on

covariates. The problem here is that our data set did not have the structure of the classical point pattern, i.e. a regular lattice [2], but it was very complicated to deal with using standard methods. Log Gaussian Cox processes (LGCP) are a class of flexible models that are particularly useful in the context of modelling aggregation relative to some underlying unobserved environmental field [2–4]. Recently [4], a flexible framework for fitting complicated LGCPs using integrated nested Laplace approximation (INLA) [5] was developed. This approach, however, is still based on a regular lattice. Although, this leads to consistent estimates, if the lattice is fine enough and appropriately discretised [6], could be very 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 [2]. In this paper we tried another more computationally efficient approach based on the stochastic partial differential equation (SPDE) models [7]. 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 using Bayesian inference for GMRFs, it was possible to adopt the INLA algorithm that gives significant computational advantages [5]. On the other, following that approach [2], the specification of the Gaussian random field was completely separated from the approximation of the LGCP likelihood, leading, again, to far greater flexibility.

2 Methods

2.1 Data setting

The total number of fires recorded in the study area during the period 1994–2008 was 10,783. 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 1994 to 2008) and the buffer to the wildfire belonged. Spatial covariates were also considered. In particular, eight continuous covariates (topographic variables – slope, aspect and hill shade – proximity to human populations – urban area – proximity to concomitants of human activity – roads and railroads – meteorological variables – maximum and minimum temperatures); and one categorical variable (land use).

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 TIN surface. In this article, 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 was 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.

The distance from the location of the wildfire to urban areas and to roads and railroads were constructed considering a geographical layer in each case. On the layer analyzed we used the function Euclidian distance included in the application Spatial Analyst of ArcGis10 in order to get two new layers called distance-area and distance-roads depending on the case. These layers were continuous and they were defined as a raster layer.

We also used land use maps (1:250 000) of Catalonia (32 000 Km²) using classification techniques applied on existing LANDSAT MSS images on four different years (1992, 1997, 2002 and 2010). In particular, we assigned the land use map just before the date of each wildfire. In this paper we reclassified land use into twenty-two categories: outside the scope of Catalonia; continental water; seawater; place exposed to blizzards; road infrastructures; suburbs; downtowns; industrial and commercial areas; herbaceous dry-farmed; herbaceous irrigated crops; dry fruit; irrigated fruit; vineyard; alpine meadows; transitional wood-land shrubs; sclerophyllus vegetation; deciduous forest; coniferous forest; wetlands vegetation; soil with little or no vegetation; burned areas; and sandy areas and beaches.

We also included the temperatures (maximum and minimum) that occurred in the location of the wildfire, up to seven days before the occurrence of the same. The estimation of the temperature at the point and from the previous day to a week before, of the occurrence of the wildfire was done by a two-step Bayesian model. Further details can be found elsewhere [8].

2.2 Buffer construction

Instead of constructing a fine regular lattice, we constructed an irregular grid using buffers. In particular, for each wildfire we constructed a buffer considering a distance of 1,500 meters from the centroid. Initially we had a lot of buffers, fortunately many of them overlapping. Therefore, we were overlaying 'layers' of buffers, from the first layer where no buffer was overlapped to the last with the maximum number of overlaps, so that, ultimately, 'cells' (the final group of buffers) did not overlap and a wildfire belonged to only one of the 'cells'. Finally, we took this last, as our layer, with a total of 3,782 'cells' (6.65 wildfires per cell on average, standard deviation: 61.17; median: 1; first quartile: 1; third quartile: 4). Note that we did not have a regular lattice, but a very irregular one.

2.3 Statistical model

Let N_{jt} denote the observed number of wildfires in specific cell $s_j, j=1, \dots, 3,782$, and year t ($t=1994, \dots, 2008$). As a consequence of the definition of the LGCP, N_{jt} may be considered as an independent Poisson random variable [2],

$$N_{jt} \sim \text{Poisson}(\Lambda_{jt}) \quad (1)$$



The problem is that the total intensity in each cell, Λ_{jt} , was impossible to compute and we used instead the approximation, $\Lambda_{jt} \approx |s_j| \exp(\eta_{jt}(s_j))$; where $\eta_{jt}(s_j)$ was a 'representative value' [2] within the cell and $|s_j|$ the area of the cell s_j .

This approximation, allowed us to describe the log-intensity of the Poisson processes by a linear predictor [4]:

$$\eta_{ijt}(s_j) = \beta_j + \log(Esp_{jt}) + \sum_{\alpha} \beta_{\alpha} z_{\alpha, it} + S_j + \tau_t + \nu_{jt} \quad (2)$$

Note that we specified our LGCP with some special features. First, we specified a spatio-temporal mixed model with two levels, the wildfire, with subscript i ($i=1, \dots, 10,783$); and the cell to which the wildfire belonged, with subscript j ($j=1, \dots, 3,782$). In addition, the subscript t ($t=1994, \dots, 2008$) denoted the year of occurrence of the wildfire.

Second, we included in the model (2), as an offset, the expected number of wildfires in cell j (and year t), Esp_{jt} . We constructed this variable as a draw (one per cell) from a Poisson distribution with mean equal to the average of wildfires 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 but, in the relative risk (RR) of wildfires per cell and year and in the effect of covariates on such relative risk. That is to say, if the risk of occurrence of a wildfire was higher ($RR > 1$), equal to ($RR = 1$) or less ($RR < 1$) than expected.

Third, note that we included only spatial covariates, $z_{\alpha, it}$, as explanatory variables of the relative risk of a wildfire. That is to say, all covariates were included at the level of the wildfire, not the cell (notice that the subscript was i). β_{α} denoted (unknown) parameters associated to covariates. With the exception of temperatures (both, maximum and minimum), we introduced all covariates in quintiles, allowing for a non-linear effect of them on relative risk.

Fourth, we introduced in (2) four random effects: i) heterogeneity: i.e. β_{jt} , accounting for variation in relative risk across different cells; ii) spatial dependence: S_j ; iii) temporal dependence: τ_t and; iv) spatio-temporal interaction: ν_{jt} . Note that we assume separability between spatial and temporal patterns and, at best, allow interaction between the two components.

The heterogeneity was specified as a vector of independent and Gaussian distributed random variable on j , with constant precision [9].

For the spatial covariance structure we used the Matérn family of covariance functions and a nugget term, over j .

$$\text{cov}(\varepsilon_j, \varepsilon_{j'}) = M(|j - j'|, r^2 \sigma^2, \rho, \nu) + (1 - r^2) \sigma^2 I(j = j') \quad (3)$$

M was the Matérn function [10]; $\sigma^2 I$ denoted the sill (the total variance of the innovation process); $r^2 \sigma^2$ was the variance of the spatially correlation portion of the process; $(1 - r^2) \sigma^2$ corresponded to the nugget (the variability unique to a given station); ρ was the range of the process (the size of the region

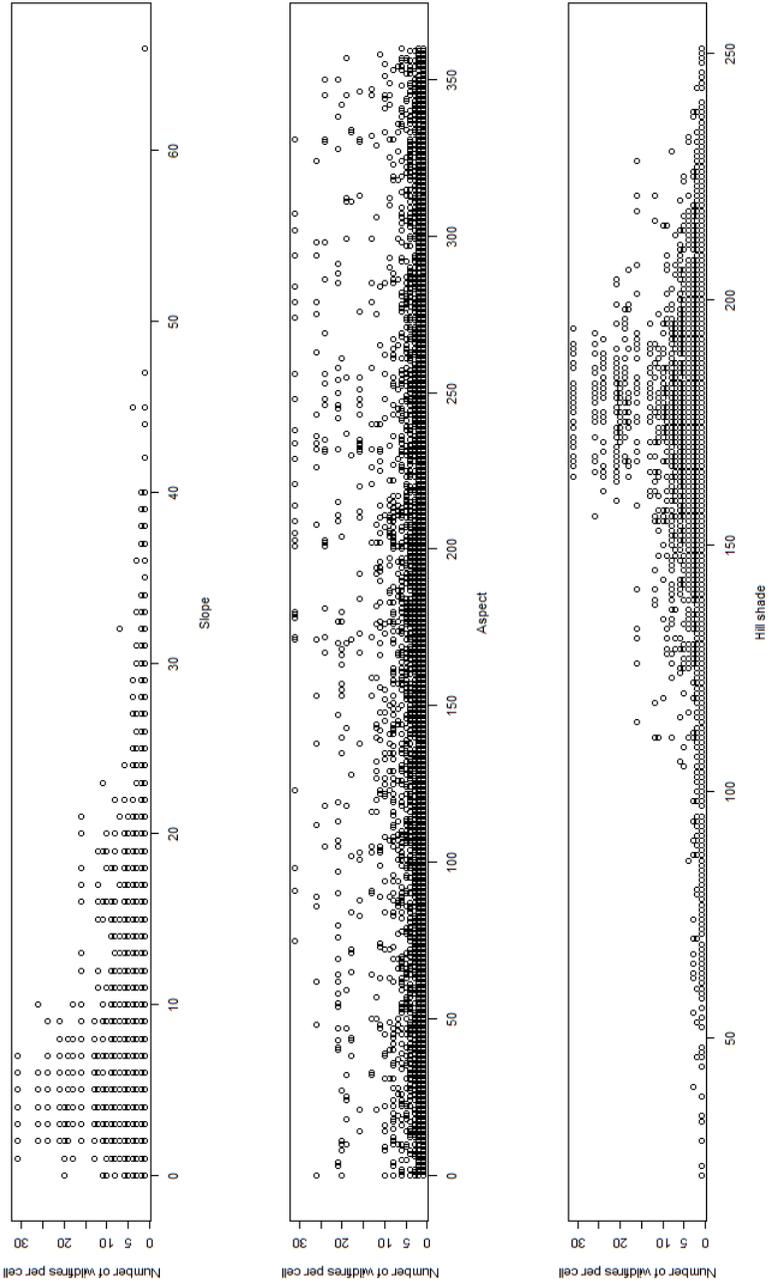


Figure 1: Wildfires per cell on topographic covariates.



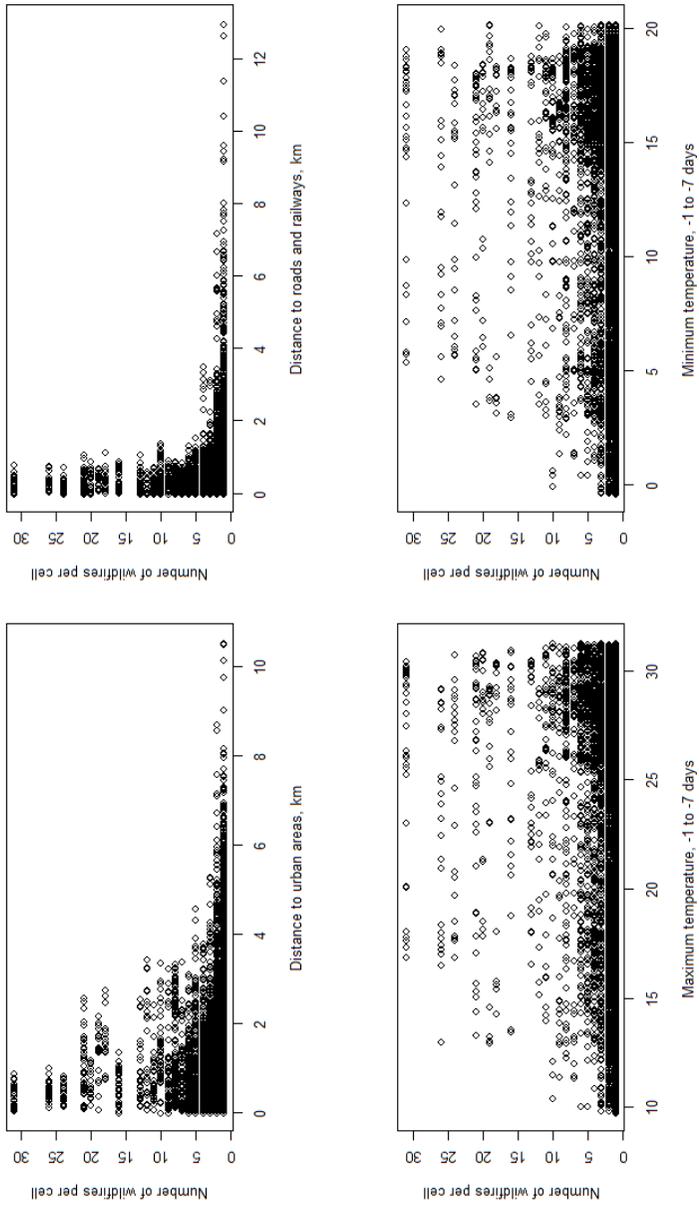


Figure 2: Wildfires per cell on the rest of continuous covariates.



where the process was significantly correlated); and ν was the smoothness of the process (specifically, we tried $\nu=1,2,3$ the only available values for the model for the time being, R-INLA project [11]).

Both the temporal dependence (on t) and the spatio-temporal interaction (on j and t) we assumed smoothed functions, in particular random walks of order 1 [9].

All the analyses were carried out with the R freeware statistical package (version 2.14.1) [12] and the R-INLA package [11].

3 Results

In Figures 1 and 2 we show some scatter plots of the number of wildfires per cell on the continuous covariates. With respect to topographic variables (Figure 1), it seems that the higher slope greater number of wildfires. The number of fires was also higher when hill shade between 150 and 200, approximately. It seems that there was not relationship between aspect and the number of wildfires per cell. A shorter distance from both, urban areas and roads-railways, greater number of wildfires per cell (less distance for roads and railways).

At higher (maximum and minimum) temperatures, on average, in the preceding days (until a week before) greater number of wildfires per cell. In this later case, however, the dispersion was very high.

Table 1: Results of the estimation. Statistically significant fixed effects.

Fixed effects	Relative risk	95% credible interval
Hill shade		
Q1 <159	0.71497	0.57037 0.91228
Q3 173-180	0.77538	0.61917 0.98865
Q4 180-189	0.77712	0.62005 0.99151
Q5 >189	0.76306	0.60806 0.97435
Aspect		
Q4 201-266	1.04736	1.00857 1.08719
Q5 >266	1.05272	1.00305 1.10412
Slope		
Q3 5-8	0.93666	0.90460 0.96944
Q4 8-13	1.04221	1.00283 1.08264
Maximum temperature		
Sixth previous day	1.05061	1.01454 1.08752
Seventh previous day	1.04171	1.01262 1.07017
Minimum temperature		
Second previous day	1.04248	1.01613 1.06937
Fourth previous day	1.02908	1.00404 1.05465

Table 2: Results of the estimation random effects.

Random effects	Mean (standard deviation)
Heterogeneity (variance)	0.40370 (0.00977)
Temporal dependency (variance)	0.14505 (0.02932)
Spatial dependency	
Range	80.060 (115.706)
Variance	0.00685
Interaction (variance)	0.37342 (0.02853)

Results of the estimation of model (2) are shown in Tables 1 and 2. In Table 1 we show the estimated fixed effects that resulted statistically significant, that is to say, the credible interval for relative risks did not contain the unity.

Note that the greater the shadow the lower the relative risk of a wildfire in a cell (RR ranged from 22% to 31% lower than in the second quintile, which was not statistically significant). With respect to aspect, the relative risk of a wildfire was 4.7–5.2% higher than in the third first quintiles. Slopes from 5–8% led to 4% less of risk, whereas from 8–13% implied 4.2% more risk. With respect to the temperatures of the previous days, in both cases, maximum and minimum, greater temperature implied more risk, although the risk was higher and farthest for maximum temperature than for minimum temperature.

With respect to the random effects, note that it seems that the importance of the heterogeneity and the interaction random effects were more important than the spatial dependency random effect. However, note that the range (distance at which the spatial correlation becomes equal to 0.1) was estimated in 80 km. Temporal dependency was also important.

In Figure 3 we show the effect on relative risk of the temporal dependency. Note that, in general, the effect has been growing until 2005, decreasing thereafter.

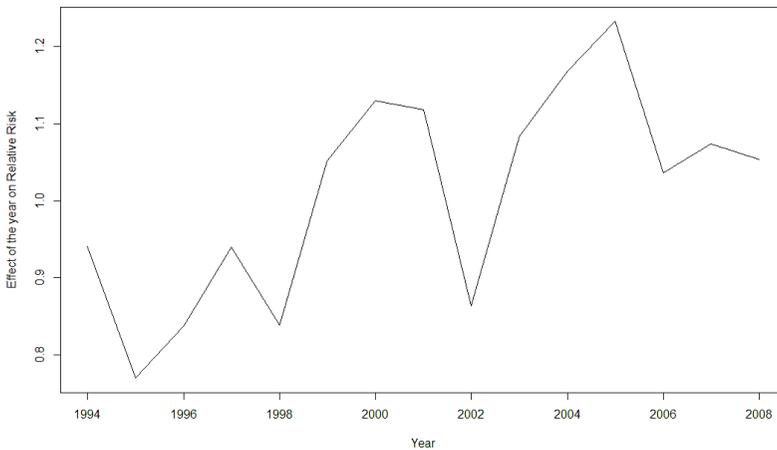


Figure 3: Effect of the temporal dependency on relative risk of wildfire.

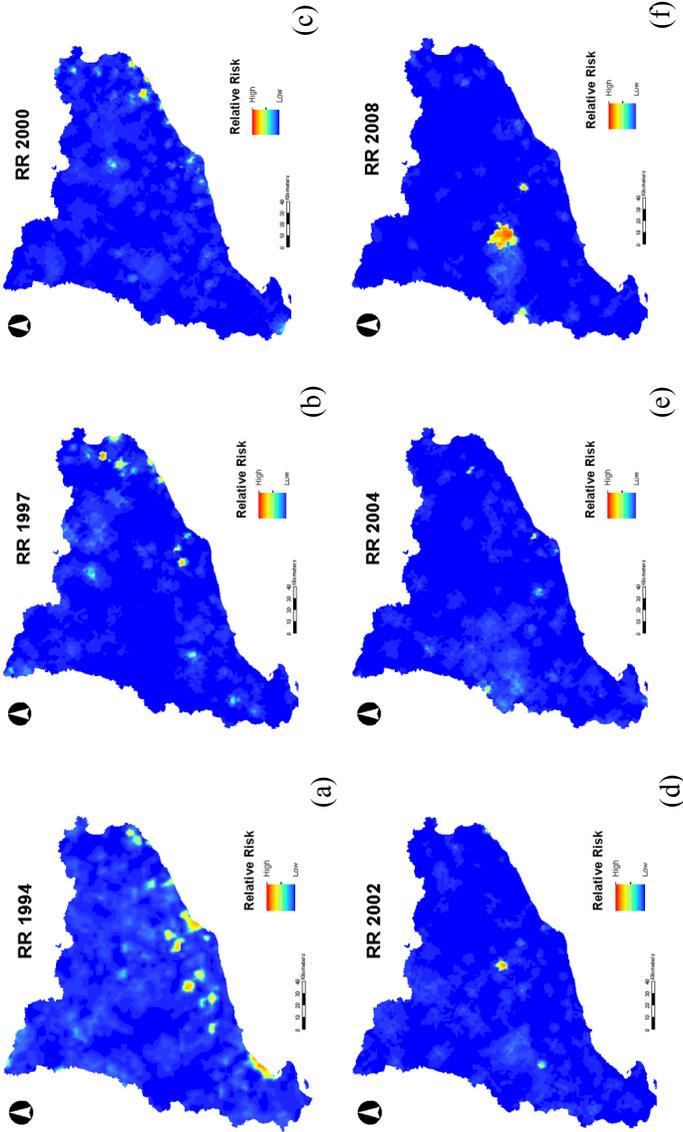


Figure 4: Relative risk of wildfires on selected years.

In Figure 4, finally, we show the relative risks of wildfires (on selected years). Note that the risk was higher than expected only in some areas, particularly, in 1994 and in the central part of Catalonia in 2008.

Acknowledgements

Part of the work that has led to this paper was done while Laura Serra was doing a pre-doctoral stage at the Department of Mathematical Sciences of the Norwegian University of Science and Technology, Trondheim, Norway. We would like to thank, in particular Håvard Rue and Finn Lindgren. We would also like to thank Irene Moreno and the Wessex Institute, for their kindness and patience with us.

References

- [1] Vélez R. Forest fires in the Mediterranean: Regional Perspective [in Spanish] <http://www.fao.org/docrep/t9500S/t9500s02.htm>
- [2] Simpson, D., Illian, J., Lindgren, F., Sørbye, S.H., Rue, H. Going off grid: Computationally efficient inference for log-Gaussian Cox processes, 2011, <http://www.math.ntnu.no/~danesi/S10-2011.pdf>
- [3] Møller, J., Syversveen, A.R., Waagepetersen, R.P. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, pp. 451–482, 1998.
- [4] Illian, J.B., Sørbye, S.H., Rue, H., Hendrichsen, D. Fitting a log Gaussian Cox process with temporally varying effects - a case study, 2010, <http://www.math.ntnu.no/inla/r-inla.org/papers/S17-2010.pdf>
- [5] Rue, H., Martino, S., Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, pp.319–392, 2009.
- [6] Waagepetersen, R. Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics & Probability Letters*, **66(3)**, pp. 229–235, 2004.
- [7] Lindgren, F., Rue, H., Lindstrom, J. 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)**, pp.423–498, 2011.
- [8] Saez, M., Barceló, M.A., Tobias, A., Varga, D., Ocaña-Riola, R., Juan, P., Mateu, J. Space-time interpolation of daily air temperatures. *Journal of Environmental Statistics* (in press), 2012.
- [9] Latent models. The R INLA project. <http://www.r-inla.org/models/latent-models>
- [10] Stein, M.L. *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer: New York, 1999.
- [11] R-INLA project (2011). URL: <http://www.r-inla.org/home>
- [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, 2011, URL <http://www.R-project.org/>

