Journal de la Société Française de Statistique

# Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences

**Titre:** Détection et modélisation de structures spatio-temporelles multi-échelles : le cas des feux de forêt

Edith Gabriel[1,2] Thomas Opitz[2] Florent Bonneu[1]

**Abstract:** Wildfires due to natural origin or arson come along with important economic and ecological risks. Typically, the structure of relative risk of fire events is highly complex, shows strong variation over space and time and is driven by numerous covariates like land use, climate and weather. We here adopt a point process approach to study space-time large-scale variations and local interaction behavior in a data set reporting geolocalized fire events and burnt surfaces on a daily basis in the Bouches-du-Rhône county in Southern France, marked by a Mediterranean climate and high touristic activity. After a review of the existing literature, we explore the influence of land use and climatic covariates like temperature and precipitation on the probability of event occurrence. Statistical challenges arise from the multi-scale structure of data defined over various supports like fine grids for land use covariates, coarse grids for fire position leading to positional uncertainty, and meteorogical series observed at irregularly spaced measurement sites. We fit a log-Gaussian Cox process including covariate information and nonparametric spatial and temporal effects to our data through the technique of Integrated Nested Laplace Approximation, and we analyze the residual interaction structure through the space-time $K$-function adapted to the setting of second-order intensity reweighted stationarity. Specifically, we also study inhibitive effects that arise locally in time and space after fire events with relatively large burnt surfaces.

**Résumé :** Les feux de forêt, qu'ils soient d'origine naturelle ou malveillante, représentent d'importants risques économiques et écologiques. La structure relative aux éclosions de feux est complexe, avec de fortes variations spatiales et temporelles, souvent dirigées par de nombreuses covariables telles que l'occupation des sols, le climat et la météo. Dans cet article, nous nous intéressons à l'éclosion de feux recensés quotidiennement depuis 1981 dans les Bouches-du-Rhône, région très touristique marquée par un climat méditérannéen, et pour lesquels nous connaissons la surface brûlée associée. Nous adoptons une approche basée sur les processus ponctuels afin d'étudier les variations spatio-temporelles à grande échelle et les interactions à petite échelle. Après un aperçu sur la litterature existante, nous explorons l'influence des covariables climatiques (comme la température, les précipitations) et de l'occupation des sols sur la probabilité d'occurrence d'un feu. Les questions statistiques sont soulevées par la structure multi-échelles des données, qui sont de plus définies sur des supports variés, comme des grilles très fines pour l'occupation des sols, des grilles grossières pour la localisation des feux qui engendrent de l'incertitude de positionnement, et des séries météorologiques observées en des sites irrégulièrement répartis. Nous ajustons à ces données un modèle de Cox log-gaussien incluant l'information des covariables et des effets spatio-temporels via la méthode INLA (Integrated Nested Laplace Approximation) et nous analysons la structure d'interaction résiduelle via la fonction $K$ spatio-temporelle inhomogène. Nous mettons en évidence et modélisons en particulier les effets inhibiteurs localement présents dans l'espace et dans le temps induits par de très grandes surfaces brûlées.

***Keywords:*** INLA, multi-scale data, K-function, point process, space-time dependence, wildfire
***Mots-clés :*** dépendance spatio-temporelle, données multi-échelles, feux de forêt, processus ponctuel
***AMS 2000 subject classifications:*** 62-01, 62-07, 62F99, 62M40

[1] LMA EA2151, Avignon University, F-84000 Avignon, France.
E-mail: edith.gabriel@univ-avignon.fr ; E-mail: thomas.opitz@inra.fr
[2] BioSP, INRA, F-84914 Avignon, France.
E-mail: florent.bonneu@univ-avignon.fr


 ISSN: 2102-6238

## 1. Introduction

Mediterranean climate is characterized by hot, dry summers and cool, wet winters. Such a climate associated with vegetation types and human activity makes Mediterranean countries very exposed to wildfires, which have become the primary cause of forest destruction in this region. Wildfire prevention begins with understanding the stochastic mechanisms governing the intensity of occurrences and the severity of fires. The scientific reflection on risk prevention needs to consider realistic modeling of wildfires and their consequences in order to propose and improve decision tools that should remain simple and easy to use. Point processes offer a suitable setting to tackle realistic spatio-temporal modeling of wildfire occurrences and their spatial extent based on climate and environmental data, allowing us to take into account their complex interaction structures.

Since 1973, the French Government maintains the continuously updated Prométhée database [1] of wildfire occurrences in the Mediterranean region to develop statistical tools for spatial and temporal comparisons and a better knowledge of wildfire causes. Locations and dates of wildfires are recorded with associated characteristics such as the burnt area. In this paper, we analyze and model wildfires in Bouches-du-Rhône (Southern France) since 1981. Wildfire occurrence may depend on the presence and concomitance of several factors: climatic (temperature, humidity, wind speed), environmental (vegetation types, urbanization, road network) and human activity. Whilst human activity is the main cause of wildfires, climatic and environmental conditions are a prior condition to their outbreak and their propagation. Thus, before modeling wildfire occurrences, it is necessary to better understand the correlations between these phenomena to allow a good estimation and prediction of risks.

The analysis of point patterns allows us to highlight the factors driving trends and interactions in the spatial distribution and the temporal structure of points and determine the observation scales of these relationships. The methods for analyzing and modeling point processes have been used in the context of wildfires (Genton et al., 2006; Serra et al., 2014; Turner, 2009; Xu and Schoenberg, 2011), but the spatial and temporal data have been treated separately or strongly aggregated (per year, per spatial area) which is unsatisfactory when one wants to understand and model the stochastic mechanisms of spatio-temporal interaction. However, some statistical methods exist for studying the spatio-temporal structures of such data (Bonneu, 2007; Gabriel and Diggle, 2009); they yield a mechanistic or empirical modeling approach Gabriel (2016).

The spatio-temporal structure of the distribution of wildfires is very complex since, in practice, the dependence cannot be separated in space and time. The spatial heterogeneity of wildfires depends on the spatial distribution of current land use like vegetation, urban zones or wetlands (Juan et al., 2012; Møller and Diaz-Avalos, 2010; Pereira et al., 2013; Serra et al., 2014). However, as it will be shown in Section 3, it also depends on the past, because changes in vegetation due to fires affect the probability of wildfire occurrence during a regeneration period. In the literature, the hypothesis of separability in space and time is often assumed without any test because it allows decomposing the problem into two modeling steps, one in space and one in time, or to consider separable covariance matrices. This simplifying assumption led thus to progress on other scientific barriers.

We aim to detect and model multi-scale spatio-temporal structures in wildfire occurrences. Log-Gaussian Cox processes, along with the INLA method for inference and prediction, are

[1] http://www.promethee.com/


 ISSN: 2102-6238

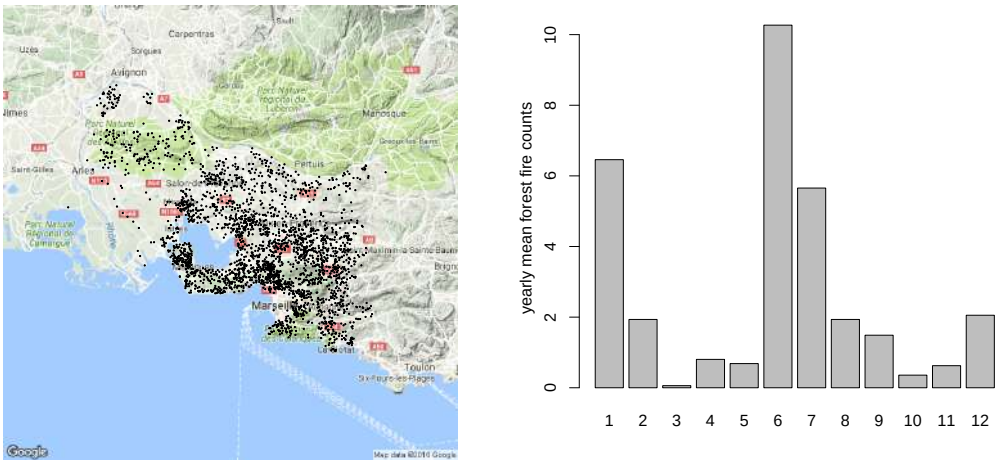


FIGURE 1. *Left: Map of forest fire locations recorded in the Bouches-du-Rhône department in Southern France. Right: Mean yearly number of forest fires with burnt surface superior to* 1*ha with respect to months.*

particularly useful to model clustered events (see, *e.g.*, Pereira et al., 2013 and Serra et al., 2014 in the context of wildfires). In this paper we show that they can also deal with more complex structures, allowing further temporal inhibition at small spatial scales and thus providing more accurate predictions.

The paper is structured as follows. Section 2 introduces the data and analyzes the influence of covariates through exploratory statistics and plots. Multi-scale spatio-temporal dependence and interaction are tested and described in Section 3. The log-Gaussian Cox process models and the related predictions are presented in Section 4.

## 2. Wildfire-related data for Southern France

### *2.1. The Prométhée dataset*

We consider a record of fire starting points for the years 1981 to 2015 for the Bouches-du-Rhône department in Southern France whose surface area amounts to around $5100km^2$. The spatial resolution is given by the DFCI coordinates spanning a grid in the Lambert 93 projection with quadratic grid cells covering approximately $4km^2$ each. The point coordinates of fires correspond to the center of the grid cell where the fire started. The value of burnt surface is also available for each event. The left display of Figure 1 shows a map of the Bouches-du-Rhône area, at the border of the Mediterranean sea, with black dots for the positions of DFCI cell centers where wildfires having burnt more than $1ha$ of surface have taken place. The right display of Figure 1 reports yearly means of the number of fire occurrences with respect to the twelve months. Clearly, the event intensity peaks during the hot summer season, but we also observe a high value for January, perhaps owing to accidental fires from New Year's Eve festivities and fireworks.

For weather data, we use freely available observation series from the Global Historical Climate Network (GHCN) hosted by the National Climatic Data Center[2] (NCDC). We here work with

[2] www.ncdc.noaa.gov/cdo-web/


 ISSN: 2102-6238

daily observation series of average temperatures, cumulated precipitation and maximum sustained wind speed for one measurement station (Marignane Airport, close to Marseille) in the Bouches-du-Rhône department; the observation record covers the full Prométhée period with only relatively few missing data. Since spatial interpolation of climate data is a very complex task in itself, we will not take into account the spatial variation of climate variables, but rather consider the Marignane data as a good proxy for the whole study region Bouches-du-Rhône, which is relatively small.

### *2.2. Detecting the influence of dynamic and static covariates*

The Bouches-du-Rhône region and Southern France in general welcome large numbers of tourists and local holiday-makers especially during the hot summer season. Therefore, it would be interesting to take their impact on forest fire risk into consideration. For lack of related data in our analysis, we cannot directly study this effect. Nevertheless, climatic covariates can be considered as a very rough proxy for touristic activity, which avoids the difficult tasks of quantifying tourist activity of and disentangling the highly correlated climate and touristic effects. In unreported preliminary work, we have further investigated the effect of weekdays (*e.g.*, week-end vs. other days of the week), but we could not detect a significant effect. An isolated peak in the event counts also appears around New Year's Eve, which we conjecture to be attributable to accidents with fireworks.

#### *Land use*

We can expect land use to be highly informative on the risk of wildfires. The presence of vegetation and its type (conifer forests, deciduous forests, heathland, ...) is crucial, and we can trivially exclude fire occurrence over expanses of water. The effects of infrastructure are more difficult to capture (road networks, tourism, urbanization, ...). Moreover, almost one-dimensional land elements like roads or rivers have little surface and would require a spatial resolution that is computationally prohibitive, or they have to be transformed into more appropriate, heuristically motivated representations like local densities or distances to the closest element.

The influence of the land use covariate information available in our data is not straightforward to estimate since it requires a spatially explicit model with high resolution in order to disentangle the effect of land use and other unknown spatially varying effects; moreover, the resolution of fire positions is considerably lower in our data set. Land use is a dynamic process evolving over time, but our data are static and report the current state; they therefore do not take into account changes that may have taken place between the start and the end of the study period. Concerning the estimation of regression coefficients to describe the influence of land use covariates, we refer to the modeling Section 4 where all effects are jointly estimated in a log-Gaussian Cox process model.

#### *Weather and climate*

Weather conditions at the day under consideration and during the period preceding this day are known to have a strong impact on the risk of wildfire occurrences. Indeed, it is natural to assume


 ISSN: 2102-6238

that heat waves without significant precipitation characterize high-risk periods. The effect of wind may be less obvious. In general, winds may fan the flames and can lead to a rapid spread of fire and sparks across large areas. However, during days with the strong, so-called Mistral winds blowing through the Bouches-du-Rhône area touristic outdoor activities may be rather low; therefore, certain risks are reduced during such days, and the effect of wind could be highly nonlinear and difficult to detect. Notice that, in practice, the effect of high-risk weather may be further mitigated by administrative measures, as for instance the limitation of public access to natural areas with vegetation.

We start with an exploratory analysis. The scatterplot in Figure 2 illustrates daily temperature/precipitation data with additional information on wildfire occurrence in the Bouches-du-Rhône, and the associated wind speeds. Since durations of high-temperature periods and the cumulation of precipitation over several days are more important than isolated instantaneous measures of such variables, we here propose to consider the average temperature and the cumulated precipitation over the 7 or 28 preceding days including the day of study. We expect that such "aggregated" variables will lead to more significant effects. We see that most fire events are associated to rather high temperatures and to rather low cumulated precipitation amounts for both aggregation periods, with mostly 0 values in the case of 7 day aggregation. A clear influence of wind speed is difficult to discern. To investigate further if climatic variables are different from seasonal mean behavior when fires occur, we look at Figure 3. It shows monthly boxplots for the above-mentioned 7 day aggregated weather variables by distinguishing days with fire from the totality of days in the study period. Average temperatures preceding fire events (Figure 3, left display) may be slightly higher during summer months, but the effect of temperature over the whole year is mixed. For precipitation (Figure 3, middle display) we observe a general tendency towards lower values before fires occur. Finally, we cannot make out noteworthy effects of wind speed (Figure 3, right display) on the wildfire risk. We now turn to Figure 4, which reports spline regression curves that explain the effect of the different segments of values of the precedingly defined weather covariates on the presence of daily fire occurrences in the Bouches-du-Rhône. Bernoulli regression was conducted through `R`'s `glm`-function. Effects are clearly nonlinear, and the results also confirm that the co-occurrence patterns of temperature and precipitation have a strong impact on fire risk.

Finally, we try to shed some more light on the influence of wind speeds through the exploratory diagrams in Figure 5. The left plot shows the counts of daily fire events against observed wind speeds, while the right one opposes burnt surfaces of fire events to the wind speed observed at the event date. Again, drawing conclusions from a visual inspection of these plots is not straightforward. In particular, we cannot detect any monotonous effect of wind speed on the burnt surfaces. There may even be a contrary effect since very large fires never co-occur with very large wind speeds in our data.

## 3. Interaction of wildfire occurrences over space and time

Our data set is of the form $\{(s_i,t_i), i = 1,\dots,n = 2831\}$, where $(s_i,t_i)$ corresponds to a wildfire spatially indexed by a DFCI cell center $s_i = (x_i,y_i)$ and year $t_i$. These data can be viewed as a realization of a spatio-temporal process observed in a bounded spatial window $W$ (which corresponds to the Bouches-du-Rhône department) and a finite temporal window $T = [1981,2015]$.


 ISSN: 2102-6238

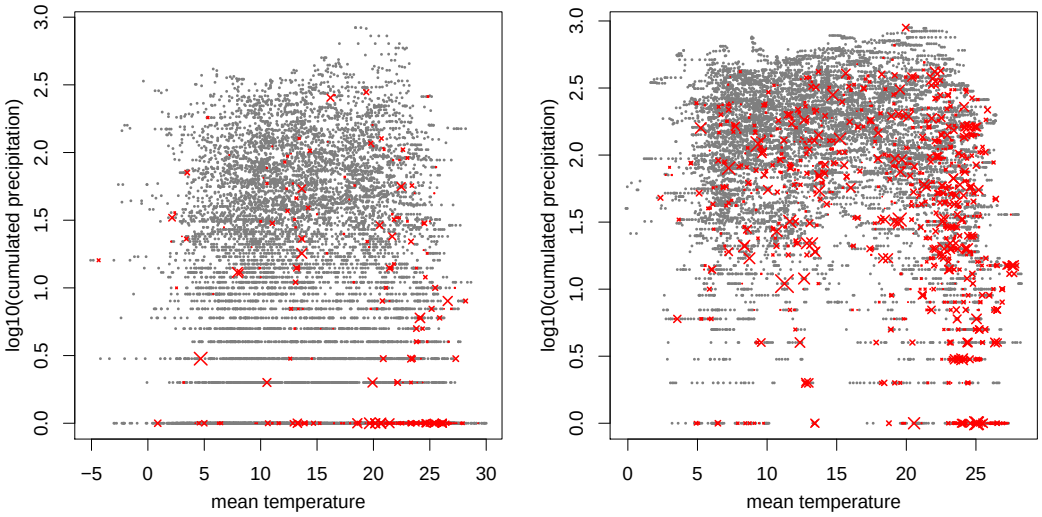


FIGURE 2. *Climate covariates and fire events. Climate covariates (left: average* 7 *day temperature preceding the reported day, base-*10 *logarithm of* 1+*cumulated precipitation; right: with average* 28 *day values) are shown for all days of the study period. Days with fire events are marked red, and the size of the* × *symbol is proportional to the wind speed observed at the event date.*

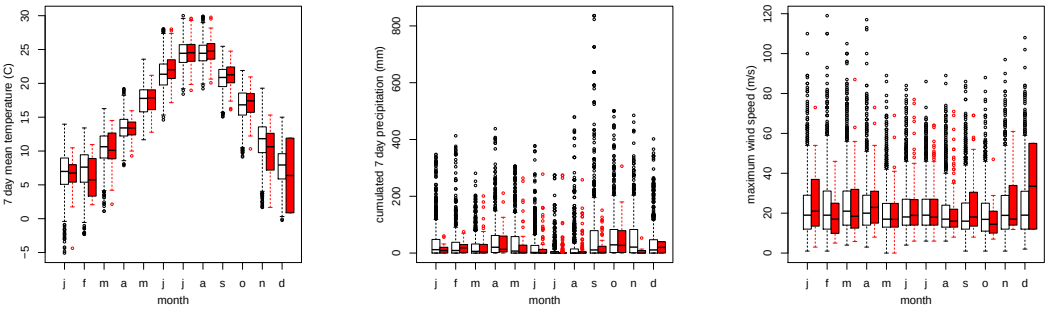


FIGURE 3. *Boxplots of weather data by month. Red boxplots represent the subsample of days with observed fire events. Left:* 7 *day average temperature preceding the reported day. Middle:* 7 *day cumulated precipitation preceding the reported day. Right: Wind speed for the reported day.*


 ISSN: 2102-6238

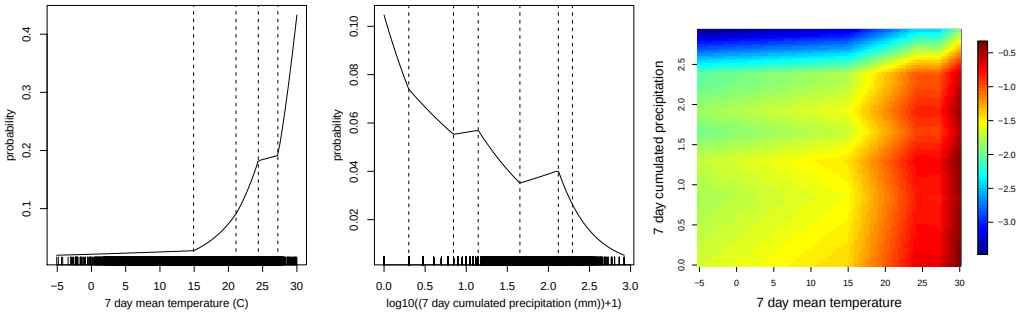


FIGURE 4. *Nonlinear Bernoulli regression of daily fire occurrence indicator variables in the Bouches-du-Rhône on climate covariates. Regression curves are based on linear splines with interior knot locations indicated by dashed vertical lines. The rugs in the left and middle display show the observed covariates. Left: Occurrence probabilities predicted from 7 day average temperatures. Middle: Occurrence probabilities predicted from 7 day cumulated precipitation. Right: Occurrence probabilities predicted from average temperatures and cumulated precipitation (regression with two spline curves for the two covariates).*

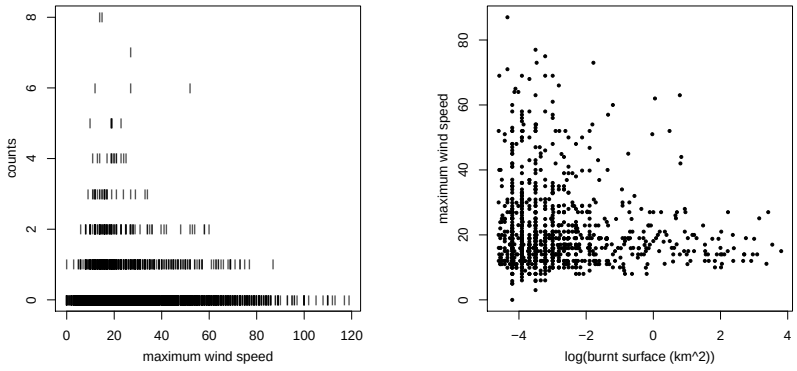


FIGURE 5. *Fire events in the Bouches-du-Rhône with respect to wind speeds. Left: Daily event count plotted against observed wind speeds. Right: Burnt surface against observed wind speed.*


 ISSN: 2102-6238

We refer to the recent review of Gonzalez-Monsalve et al. (2016) which describes statistical models and methods that can be used to analyze the spatio-temporal behaviour of point patterns. Spatio-temporal patterns of wildfires are essentially characterized through two components: a trend and some interaction between events (see e.g. Møller and Diaz-Avalos, 2010; Juan et al., 2012). As it will be described in Section 4, the trend incorporates covariate information. In this section, we analyze the interaction of wildfires over space and time.

### 3.1. *Temporal inhibition at close spatial distances*

An important aspect of modeling is the study of the geometric small-scale patterns that can arise at close distances in space and time around a given fire event. Physically, we expect that vegetation and other combustible material has been burnt throughout the surface covered by a wildfire and will progressively regenerate afterwards. Therefore, it is natural to assume that no wildfires can take place over the same surface immediately after an event, and ecological knowledge about regeneration cycles suggests that this inhibitive pattern between neighboring events can span several years. To provide empirical evidence of this claim, we here propose a simple and robust measure through the *normalized empirical intensity ratio index* $I_t$ for years $t = 1981, 1982, \ldots$ defined as the ratio of two counts: $(i)$ number of events observed in $T_{follow} = [t+1, t+\Delta t]$ that are spatially close to wildfires observed in $T_{ref} = [t-2, t]$ and $(ii)$ number of events observed outside the period $T_{follow} \cup T_{ref} = [t-2, t+\Delta t]$ that are spatially close to wildfires observed in $T_{ref}$. In addition, we normalize these counts with the respective global intensities for the periods $T_{ref}$ and $(T_{follow} \cup T_{ref})^C$. Therefore, index values below 1 indicate temporal inhibition between fire events at close distances, while values above 1 suggest temporal clustering. More precisely, the following notion of proximity is used in the calculation of the index. For wildfires $(s_i, a_i)$ observed around $s_i$ with surface $a_i$ taking place during $T_{ref}$, we consider the union of discs centered at $s_i$ with area $\alpha a_i$, where $\alpha > 0$ is a scaling parameter, which we try to tune for minimizing the positional uncertainty. Ideally, we would like to work with the number of follow-up events taking place on the precise position of burnt surfaces, but due to the lack of precise locations for fire sites and burnt surfaces in our data this is not possible. Instead, after having checked several values for $\alpha$ in preliminary analyses, we here choose $\alpha = 0.75^2$ and consider only relatively large conditioning events from $T_{ref}$. Discs are centered at the grid cell centers of the DFCI coordinates of the events under consideration. Figure 6 shows indexes calculated from the full data in the Prométhée data base for pivotal years $t = 1981$ to 2012 and various minimum surface parameters for events entering into the calculation. The results indicate that it is difficult to detect inhibitive patterns when keeping relatively small events exceeding $10ha$ (Figure 6, left display) owing to the noisy positional data, but inhibition becomes more clear when using conditioning events in $T_{ref}$ above $50ha$ (middle display) or above $100ha$ (right display) where many follow-up periods $T_{follow}$ contain no event at all spatially close to the ones in $T_{ref}$. Although it would be rather complicated to calculate suitable confidence intervals for these index values, we still underline that the event counts for the various indicator sets are sufficiently high to draw reliable conclusions on the presence of temporal inhibition from Figure 6. We will thus consider temporal inhibition at close distances as a fact. A good explicative and predictive generative model should therefore integrate a component that allows quantifying its extent. Moreover, the space-time exploratory statistics related to the $K$-function reported in the following section may reveal further details


 ISSN: 2102-6238

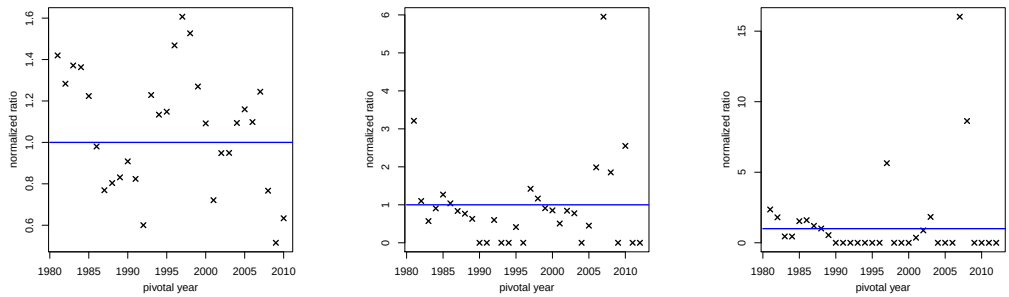


FIGURE 6. *Normalized empirical intensity ratio index $I_t$ for various parameters. Left: Results for events with at least* 10*ha of burnt surface and a* 5 *year follow-up period. Middle: Results for conditioning events with at least* 50*ha of burnt surface and other events with* 20*ha minimum surface, and* 3 *year follow-up period. Right: Results for conditioning events with at least* 100*ha of burnt surface and other events with* 25*ha minimum surface, and a* 3 *year follow-up period.*

about small-scale and intermediate-scale interaction patterns.

### 3.2. *Spatio-temporal inhomogenous K-function*

By making second-order stationarity and isotropy assumptions, one assumes that the intensity of the process is constant over the spatio-temporal domain and that the second-order moment intensity reduces to a function of spatial and temporal distances. Whilst widely used in practice, the assumption of second-order stationarity is not plausible in most of environmental applications and can lead to the detection of spurious interaction patterns. Gabriel and Diggle (2009) extended to the spatio-temporal setting the definition of a weaker assumption: the second-order intensity reweighted stationarity (SOIRS), see Baddeley et al. (2000). For SOIRS and isotropic spatio-temporal point processes $\Phi$, Gabriel and Diggle (2009) define the spatio-temporal inhomogenous $K$-function as:

$$K(h_s,h_t) = \frac{1}{|A|}\mathbb{E}\left[\sum_{(s_i,t_i)\in\Phi\cap A}\sum_{(s_j,t_j)\in\Phi\setminus(s_i,t_i)}\frac{\mathbb{I}\{\|s_i-s_j\|\le h_s;|t_i-t_j|\le h_t\}}{\lambda(s_i,t_i)\lambda(s_j,t_j)}\right],$$

where $h_s,h_t>0$, $A$ is a compact, and its related empirical estimator is

$$\widehat{K}(h_s,h_t) = \sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{w_{ij}}\frac{\mathbb{I}\{\|s_i-s_j\|\le h_s\,;\,|t_i-t_j|\le h_t\}}{\lambda(s_i,t_i)\lambda(s_j,t_j)},$$

where $w_{ij}$ is a spatio-temporal edge correction factor, for instance one of those defined in Gabriel (2014).

As illustrated in Gabriel et al. (2010); Tamayo-Uria et al. (2014), the spatio-temporal inhomogeneous $K$-function can be used to test for spatio-temporal clustering and interaction. Under the null hypothesis $H_0^c$ that the pattern is a realization of a Poisson process with intensity $\lambda(s,t)$, we have $K(h_s,h_t)=2\pi h_s^2 h_t$. We can construct confidence envelopes from simulations of the process with distribution $\mathscr{P}ois(\lambda(s,t))$. Under the null hypothesis, $H_0^i$, that the pattern is a realization of a pair of independent spatial and temporal, second-order intensity reweighted stationary point


 ISSN: 2102-6238

TABLE 1. *Summary of intensity models used for estimating the spatio-temporal inhomogeneous K-function.*

| | **Model** | | | |
|---|---|---|---|---|
| | A<br>Purely spatial | B<br>Space-time separable with random walk | C<br>Space-time separable with time-replicated Gaussian random field | D<br>Kernel smoothing |
| **Spatial intensity** | | | | |
| land use covariates | ✓ | ✓ | ✓ | |
| $W(s)$: spatial marginal effects | ✓ | ✓ | | |
| $W_t(s)$: spatial effects i.i.d in time | | | ✓ | |
| Gaussian kernel | | | | ✓ |
| **Temporal intensity** | | | | |
| uniform over the years | ✓ | | | |
| climate covariates | | ✓ | ✓ | |
| $W(t)$: temporal marginal effects | | ✓ | | |
| Gaussian kernel | | | | ✓ |

processes, we have $K(h_s, h_t) \propto K_S(h_s)K_T(h_t)$; i.e., the process is second-order separable (Gabriel and Diggle, 2009; Møller and Ghorbani, 2012). We can build confidence envelopes from estimates calculated from independent permutations of event times, holding their locations fixed Diggle et al. (1995). For simulations and estimation, we here used the package `stpp` (Gabriel et al., 2013) of the `R` (R Core Team, 2016) software library.

To perform these analyses, we assumed that the intensity is separable in space and time. It has been estimated by incorporating different spatial and temporal effects as summarized in Table 1 and described in Section 4. For each of these models we computed the spatio-temporal inhomogeneous $K$-function both from the data and from 100 patterns simulated under each of the null hypotheses $H_0^c$ and $H_0^i$. Figure 7 highlights spatial and temporal distances at which spatio-temporal clustering (in yellow), interaction (in red), or both (in orange) has been detected. When modeling the inhomogeneity through a purely spatial intensity reflecting land use covariates (intensity model A) or by more complex non-parametric estimation of the spatial and temporal components (model D), we mainly observe that there is spatio-temporal clustering for different ranges of spatial and temporal distances (as the spatial intensity is simply an average over the years). Clustering is absent at small spatial distances, and we conjecture that this is at least partly due to the spatio-temporal inhibition highlighted in the previous section, which also leads to the existence of spatio-temporal interaction. Models B and C correct these effects. In particular, model C features a spatial random effect, i.i.d. in time, which can capture the spatial inhomogeneity separately for each year. It further shows some spatio-temporal interaction at spatial and temporal distances below 20 *km* and 8 years respectively. Model B appears to be the one leading to the weakest spatio-temporal residual structure.


 ISSN: 2102-6238

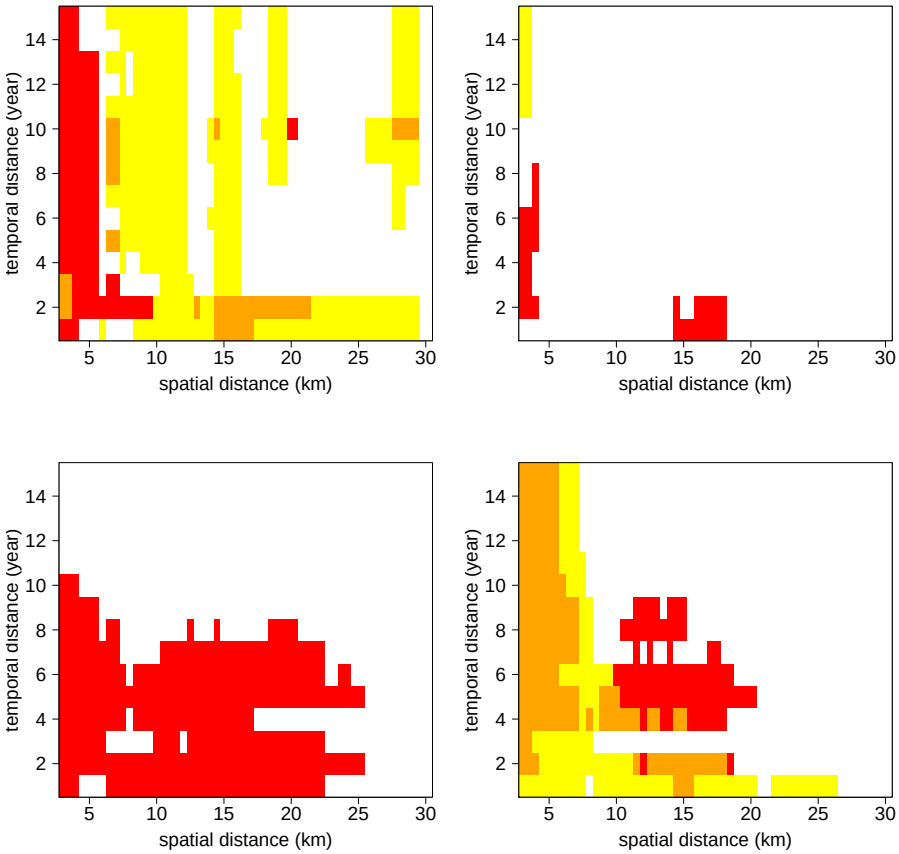


FIGURE 7. *Comparison between the spatio-temporal inhomogeneous K-function and tolerance envelopes indicating spatio-temporal clustering (yellow), interaction (red), both (orange), when the intensity is estimated according to model A (top left), model B (top right), model C (bottom left) and model D (bottom right).*


 ISSN: 2102-6238

## 4. Log-Gaussian Cox process models

### 4.1. Model building and inference

In the light of the previous analyses, we here propose to construct stochastic models that incorporate the available covariate information for the sake of predicting wildfire risk intensities. Notice that the predicted intensities of the models below have already been used in the estimation of the space-time inhomogeneous $K$ function in the preceding section. We concede that some if not all of our models will not capture the full extent of the stochasticity in the wildfire event process, but still provide highly valuable information. The challenging task of constructing and fitting a model that captures the full complexity of the space-time intensity variations and stochastic structure is left for future research. We consider models with a stochastic intensity of log-Gaussian type and a space-year resolution for incorporating covariate information. Models of different complexity are considered for a Gaussian space-time effect $W(s,t)$, whose spatial component is always based on the flexible yet computationally convenient Matérn-like spatial Gauss–Markov random fields arising as approximate solutions to certain stochastic partial differential equations (see Lindgren et al., 2011; Simpson et al., 2016 for details). The model is specified for years $t \in \{1981,\dots,2015\}$ in the following way:

$$\Lambda(s,t) = \exp\left(\beta_0 + \beta_{inhib} z_{inhib}(s,t) + \sum_{\text{land use}} \beta_{land,i} z_{land,i}(s) + \sum_{j=1}^{3} \beta_{clim,j} z_{clim,j}(t) + W(s,t)\right), \tag{1}$$

with covariates $z_{land,i}$ related to land use, $z_{clim,j}$ to climate and $z_{inhib}$ to fires in the same DFCI cell during the 5 years preceding $t$.

We now explain the components in more detail. We have studied the following three structures for $W(s,t)$:

$$W(s,t) = W(s) \qquad \text{spatial marginal effect,} \tag{2}$$

$$W(s,t) = W(s) + W(t) \qquad \text{spatial and temporal marginal effects,} \tag{3}$$

$$W(s,t) = W_t(s) \qquad \text{spatial effects i.i.d. in time for } t = 1981,\dots,2015. \tag{4}$$

Models (2) and (3) can be considered either as a relatively simple log-Gaussian Cox processes, or as a Bayesian model for a Poisson process where the prior for the intensity is purely spatial in (2) and is space-time separable in (3). The marginal temporal effect $W(t)$ is here chosen as a first-order random walk with a sum-to-zero constraint. Model (4) incorporates higher stochasticity into the model through its replicated spatial effects; it is therefore capable to model clustering of events at the yearly level. To correctly normalize the spatio-temporal intensity for each year of the full observation period in models (2) and (3), we add the offset term $-\log$(number of observation years) to $\log\Lambda(s,t)$.

In models (2) and (3), a purely temporal effect is given through $\sum_{j=1}^{3} \beta_{clim,j} z_{clim,j}(t) + W(t)$, while $\sum_{\text{land use}} \beta_{land,i} z_{land,i}(s) + W(s)$ is a purely spatial effect. The inhibition effect $\beta_{inhib} z_{inhib}(s,t)$ is a novelty in our model compared to the existing literature and breaks the space-time separation of our model. It artificially integrates a repulsive pattern into the process, which remains well-defined since we can simulate the process iteratively for each time step by conditioning on the realization of the preceding time steps.


 ISSN: 2102-6238

For land use, we include the covariate values averaged over each cell when we use a spatial discretization that is different from the initial very precise spatial resolution of the raw covariate data: proportions of land occupied by water (`water`), by forest (`forest`), specifically by coniferous forest (`conif`), specifically by deciduous forests (`decid`), by paved roads (`road`), by paths and dirt roads (`path`), by buildings (`build`), or by any kind of vegetation (`veget`); average values of altitude (`alti`) and of the norm of the gradient (`grad`).

Aggregating climate information to a yearly scale is far from being straightforward and many approaches could be used to encode interesting information and to obtain significant variables. To our knowledge, existing literature does not provide in-depth analysis of how exactly climate data should be aggregated for fire risk prediction in our study region. Therefore, we have conducted a preliminary descriptive analysis of correlation patterns between various aggregation approaches over one week to one month, whose results are not reported in detail. Based on this preliminary analysis of the correlation structure between various candidates of aggregated climatic covariates, we use three variables $z_{clim,j}(t)$, $j = 1,\dots,3$, $t = 1981,\dots,2015$, in our model. Their calculation relies on two aggregated daily variables computed for each day of the observation period, consisting of the mean temperature and the cumulated precipitation amount over the 4 weeks preceding the day under consideration. For each year, $z_{clim,1}(t)$ is then given as the proportion of days where this temperature variable lies above the $1-28/365.25$ quantile and this precipitation variable falls below the $28/365.25$ quantile, with quantiles calculated over all 35 years. Therefore, $z_{clim,1}(t)$ indicates the proportion of times that 4-week return levels are exceeded for extreme hot 4-week temperatures and are underrun for extreme low 4-week precipitation in year $t$. Next, $z_{clim,2}(t)$ reports the proportion of underruns of the median of 4-week precipitation during year $t$, and $z_{clim,3}(t)$ the proportion of underruns of the the 0.1-quantile of 4-week precipitation during year $t$. To make these 3 yearly climate covariates better comparable between years, we divide each by its mean over the whole observation period. For instance, if $z_{clim,2}(2001) = 2$, then we had twice as much exceedances of the median of 4-weekly cumulated precipitation in 2001 compared to the average year.

Finally, we have included a covariate $z_{inhib}(s,t)$ that is related to the proportion of a DFCI cell affected by fires in the 5 years preceding $t$. A fire in the cell $s$ during $t-\Delta t$ with $\Delta t \in \{1,\dots,5\}$ and burnt surface $a$ contributes $(6-\Delta t)/5 \times \min(a, \text{cell surface})/(\text{cell surface})$ to $z_{inhib}(s,t)$ (with cell surface equals to $4km^2$ for the DFCI cells).

For estimating the posterior means of covariate coefficients and of the spatio-temporal effect $W(s,t)$, we use the framework of Integrated Nested Laplace Approximation (Rue et al., 2009), implemented in the `INLA` package of `R` (Lindgren and Rue, 2015). Appropriate, only weakly informative priors were chosen for the estimated effects and hyperparameters like the effective range and the variance of the spatial Gaussian fields. The triangulation of space used to discretize the Gaussian random fields for obtaining a Gauss–Markov approximation according to the approach of Lindgren et al. (2011) is shown in Figure 8. The dense triangulation nodes inside the Bouches-du-Rhône department (and stretching slightly beyond) correspond to the DFCI cell centers. Notice that we here have to extend the triangulation to a larger domain to ensure approximate second-order stationarity of the Gaussian fields within the study region. To keep computing costs moderate, a less dense triangulation has been chosen for the nodes relatively far away from the study region; see Lindgren et al. (2011) and Lindgren and Rue (2015) for more technical details on these calculations and tuning steps. Overall, this mesh has 3054 nodes,


 ISSN: 2102-6238

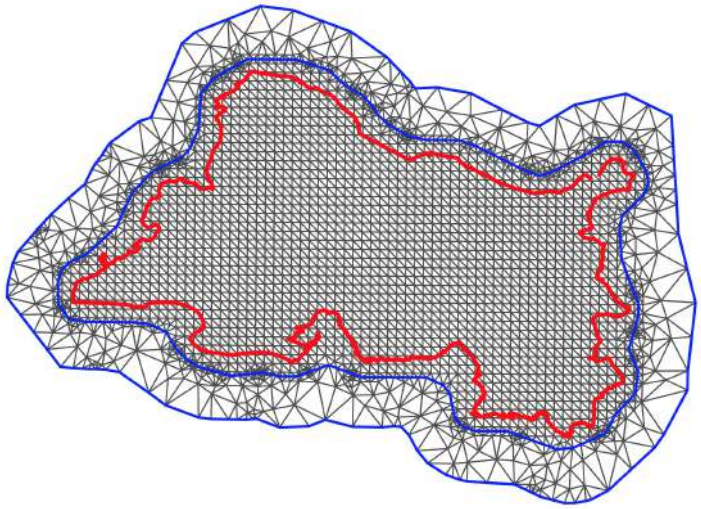

FIGURE 8. *Triangulation used for the discretization of the spatial effect $W(s)$. Red curve: boundary of the study region (Bouches-du-Rhône department). Inner blue curve: boundary of densely triangulated region. Outer blue curve: boundary of the extended domain.*

meaning that $W(s)$ and each of the replicates $W_t(s)$ for $t = 1981, 1982, \ldots$, are represented through 3054 Gaussian variables located at the nodes of the mesh. Owing to the high computational complexity when estimating hyperparameters in models with a complicated latent Gaussian structure like model (4), we fix the i.i.d. Gaussian fields' effective range to $15km$ and the variance to 1.

### *4.2. Estimation results*

Computing times on a standalone computer with 2 cores available for R-INLA were approximately 15 minutes for models (2) and (3), and 60 minutes for model (4). To start, we look at the spatial prediction of yearly fire event intensities in Figure 9 from the fit of the purely spatial intensity model (2) where time-varying covariates are set to 0; *i.e.*, $\beta_{inhib} = 0$ and $\beta_{clim,j} = 0$, $j = 1,2,3$. Notice that we have computed the predicted intensities over the $250 \times 250$ high-resolution image grid by injecting the recalculated average covariate values for the grid cells of the image grid into the fitted model. This corresponds to a change of spatial support with respect to the covariate information in the fitted model. The combination of spatial covariate information with a nonparametric spatial effect appears to have well reproduced the spatial variation in event intensities.

In the remainder of this section, we analyze the results in more detail and we focus on models which include all covariates such that time-dependence is captured. In model (3), we estimated several hyperparameters: the posterior mean of the effective spatial range in $W(s)$ where the correlation is approximately 0.1 is $15.4km$, with 95% credible interval $(10.9, 21.4)km$; for the variance of $W(s)$, we obtain the posterior mean 1.37 with credible interval $(1.06, 1.76)$; for the variance of the yearly random walk innovations, we get a posterior estimate of 0.14.


 ISSN: 2102-6238

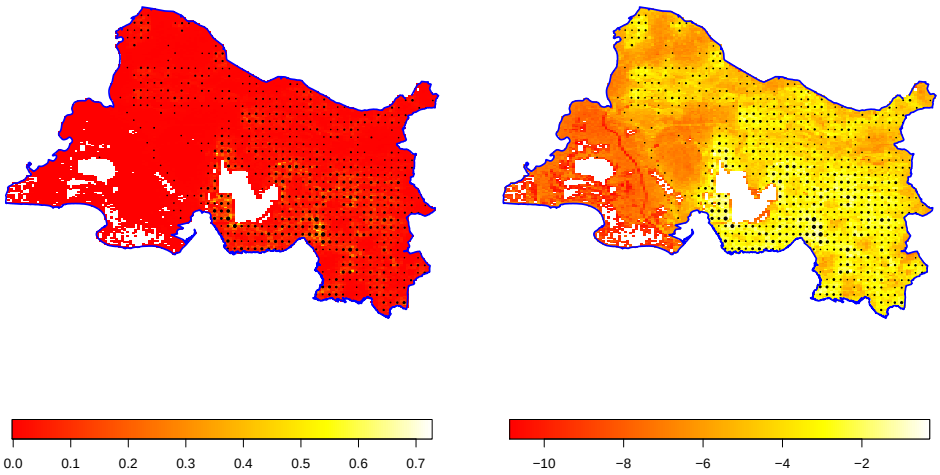


FIGURE 9. *Posterior means of spatial intensities of yearly fire events (per km$^2$). Left: posterior mean. Right: log of posterior mean. Points indicate fire events observed over the observation period; they are located at the DFCI cell centers where fires have been observed and the surface of points is proportional to the number of observed fire events. Image points with more than* 95% *of water surface, corresponding to approximate* 0 *values, are shown in white.*

TABLE 2. *Summary of posterior estimates for the coefficients of the fixed effects (intercept, preceding fire events, climatic and land use covariates). Significant effects according to* 95% *credible intervals are marked with* $^\star$.

| **Covariate** | **Posterior mean** | 95% **credible interval** |
|---|---|---|
| intercept$^\star$ | -7.3 | [-8.06,-6.56] |
| inhib | -0.13 | [-0.64,0.35] |
| clim1$^\star$ | 0.07 | [0.05,0.1] |
| clim2$^\star$ | 0.57 | [0.39,0.75] |
| clim3$^\star$ | 0.25 | [0.2,0.31] |
| water$^\star$ | -3.64 | [-5.03,-2.36] |
| forest | 0.59 | [-0.01,1.19] |
| conif$^\star$ | 0.79 | [0.19,1.4] |
| decid | -0.3 | [-1.12,0.51] |
| alti$^\star$ | -0.54 | [-0.79,-0.28] |
| grad$^\star$ | 0.25 | [0.11,0.38] |
| road$^\star$ | 0.49 | [0.33,0.65] |
| path$^\star$ | 0.29 | [0.12,0.47] |
| build | -0.04 | [-0.1,0.01] |
| veget$^\star$ | 0.67 | [0.13,1.22] |


 ISSN: 2102-6238

Table 2 sums up the estimates of the linear covariate coefficients. Although our exploratory analysis on a larger region provided empirical evidence for an inhibitive effect when fire events took place in the preceding years at close distance, this effect is not significant for the Bouches-du-Rhône data since `inhib`'s credible interval covers 0. Here, this is probably due to the smaller data set and also to the inclusion of smaller events whose burnt surface is negligible compared to the DFCI cell size, leading to high positional uncertainty. Nevertheless, the posterior mean of the coefficient is slightly negative, hinting at a tendency towards inhibition. On the other hand, this result also indicates that the occurrence of a wildfire does not considerably reduce the risk of new occurrences shortly after and nearby in a radius of around one to two kilometers corresponding to the size of a DFCI cell. The three climate covariates are clearly significant: hot and dry weather increases fire risk. It is striking that the strongest effects are due only to the scarcity of precipitation over longer periods, whereas temperature seems to play a (relatively) minor role. Nevertheless, such results should be interpreted with caution, since there also is a strong negative correlation between temperature and precipitation during the summer months in the study region. With respect to land use, we can state that the presence of coniferous forests strongly favors wildfires, whereas the credible interval of forest presence stretches slightly beyond 0. The risk for deciduous forest appears to be relatively lower compared to forest in general, although the effect is not significant in our model. Other significant effects are the presence of vegetation and of dense road or path networks. Fire risk tends to significantly decrease with altitude in our study region, whereas strong gradients (hillsides, ...) come along with higher relative risk. The presence of buildings seems to be associated to relatively smaller fire risk. This seems logical since there usually is little combustible vegetation material in highly urbanized places, and even in the contrary case where buildings are embedded into natural or semi-natural spaces and therefore close to potential high-risk sites we can assume that surveillance is high.

We now investigate in more detail if our intensity model is able to capture the temporal fluctuation in the number of yearly fire events. Figure 10 shows the time effects when averaging the posterior mean over space for each year. The left hand plot corresponds to model (2) which does not include a nonparametric time effect through the random walk $W(t)$; therefore, temporal variability is captured solely through climate covariates. The right hand plot of Figure 10 reports the estimated intensity when including the random walk effect $W(t)$ as in model (3). We see that our climate covariates does indeed explain part of the temporal variability, but adding the nonparametric effect to the model improves the fit considerably. More specifically, 17 out of the 35 random walk innovations are significantly different from 0 at the 95% level, *i.e.*, they have credible intervals that do not overlap 0.

### *4.3. Fire risk prediction*

Our statistical models allow prediction of fire risk into the future if observed or artificial covariates are available for these years. When they have not (yet) been observed, one may use a predictive model for such covariates to study fire risk with respect to the predicted mean or quantiles of covariates; otherwise, it may be instructive to compare the fire risk predictions for a small number of manually defined scenarios. Based on the preceding results, we illustrate this approach by making prediction of fire intensities for the year 2016 by fitting model (3) with marginal spatial and temporal effects using fire data until 2015 and climatic covariates observed for 2016. By


 ISSN: 2102-6238

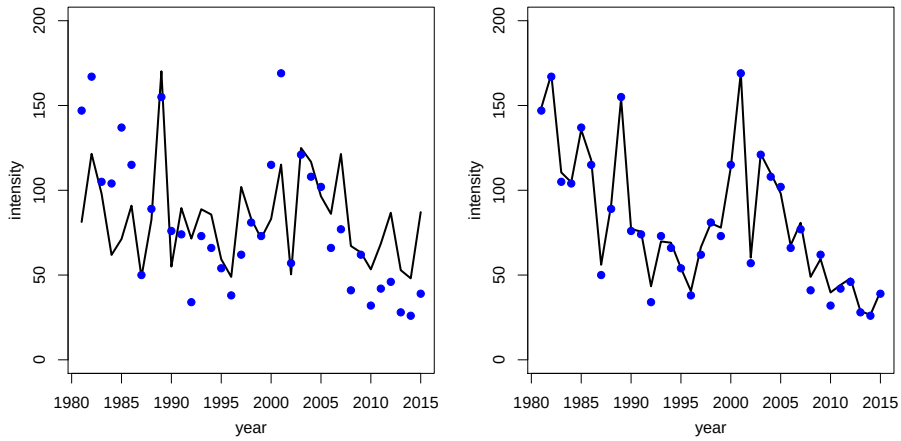


FIGURE 10. *Spatially averaged posterior means of temporal intensities of yearly fire events. Left: temporal variability captured through climate covariates only. Right: temporal variability captured through climate covariates and random walk. The blue points indicate the observed number of fire events for each year.*

comparing observed fire events in 2016 with predictions, we can check the practical utility of our model. In 2016, we only have climatic covariates and fire event data from January to October such that our predictions and comparisons can only be approximate, but they already cover the summer period where most fire events take place. Predictions are obtained by refitting model (3) with INLA, where we include the year 2016 with observed covariates but with missing fire event data; notice that making predictions in the Bayesian framework of INLA corresponds to using `NA` data for the values to predict. We then take the resulting posterior means of intensity estimates for 2016 as our prediction. Figure 11 shows a rather good match in terms of spatial variation between the predicted log-intensities and the observed fire events from 2016. It is noteworthy that there has been a jump in the number of observed events (98) compared to preceding years (28 in 2013, 26 in 2014, 39 for 2015); thanks to climatic covariate information, this jump is appropriately captured by our model whose prediction for the overall number of events in 2016 amounts to 90.

## 5. Discussion

After reviewing statistical methods from the literature to explore and model wildfire risk, we have adopted a point process approach from stochastic geometry to provide a detailed analysis of data collected for the Bouches-du-Rhône department in Southern France. The high complexity of such data is due to the strong nonstationarities over space and time, often explicable through various effects of covariates, and to multiple observations scales of event data and covariates in terms of spatial and temporal discretization and aggregation.

The results on the exploratory analysis with the space time intensity-reweighted $K$-function indicate that it is important to further investigate space-time interaction structures between events at the local level and to extend the class of models to true space-time dependence structures beyond the simple inclusion of a local inhibitive effect after fire events. It would further be useful to take into account the size of individual fires by considering a point process marked through


 ISSN: 2102-6238

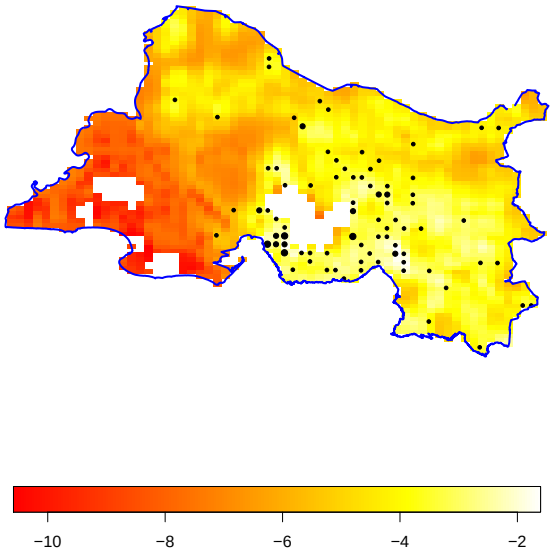


FIGURE 11. *Predicted spatial log-intensity for* 2016 *(per km*$^2$*) based on the model* (3) *fitted with fire event data until 2015 and climatic covariates for 2016. Points indicate fire events observed in 2016 from January to October; they are located at the DFCI cell centers where fires have been observed and the surface of points is proportional to the number of observed fire events. Image points with more than* 95% *of water surface, corresponding to approximate* 0 *values, are shown in white.*

burnt surfaces. Intensities and the strength of inhibitive patterns could then be estimated for each class of fire sizes, although the relatively small number of very large fires and the positional uncertainty in the fire database may render such modeling intricate.

Fires propagation can be related to wind direction leading to anisotropic shapes of burnt surfaces. Here, for sake of simplicity we have considered the isotropic case, both in the shape of burnt surfaces and in the point pattern. However further investigation could be done to check for anisotropic behavior in the point pattern, for instance using Comas et al. (2017)'s approach. If it were, our log-Gaussian Cox process model could include anisotropy as in Siino et al. (2017). Then, to handle anisotropy in INLA, we should estimate the anisotropy, model it and manually build precision matrices.

Concerning the approximations conducted when fitting the log-Gaussian Cox models in Section 4, several improvements are possible. Currently, our triangulation of the domain allows triangle edges to cross boundaries between two DFCI cells such that the random effects spill over between neighboring cells. A finer triangulation could remedy this problem. Explicit modeling of intra-year dynamics in wildfire risk would be an important progress and would enable the integration of climate covariate data with higher temporal and spatial resolution than in our model. However, such improvements make the model more complex, and consequently increase the estimation uncertainty, require more substantial computing resources and carry potential for problematic


 ISSN: 2102-6238

numerical and computational instabilities. Finally, using separate effects depending on the size of burnt surfaces similar to the approach of Serra et al. (2014) could be expedient to distinguish risk factors for small and large surface wildfires. In future work, we therefore project proposing a more complex space-time log-Gaussian Cox process model that adequately captures temporal intensity and dependence effects at small scales. It would further be interesting to investigate through statistical approaches the impact of forest fires on eco-systems and human society; for instance, an important challenge would be to propose joint modeling of forest fires and particulate matter concentrations like PM10 values, which are known to peak after important fire events. We estimate that the work on particulate matter modeling over mainland France proposed in RESSTE Network (2016) will be highly valuable for these follow-up research projects. The flexible modeling and fast estimation through INLA (Rue et al., 2009), see also the review oriented towards space-time modeling in Opitz (2016), give us justified hope that we can succeed in proposing complex and practically useful models of such multivariate multi-scale structures over space and time. We further think that functional regression approaches could improve the choice of indices and aggregated values that we use to characterize the contribution of recent weather conditions to instantaneous fire risk. This choice is currently conducted in a rather simplistic, informal and heuristic manner and it is yet based on observations from a single measurement site; but again, we are confident to propose a tractable semi-parametric functional regression approach with spatial dependence within the INLA framework, which may integrate improved physical knowledge about climate-related wildfire causes and may allow a more objective choice of aggregated climate variables with good predictive properties through formal model selection tools.

Finally, we plan to extend our statistical analysis to the data collected in the Prométhée base over a larger region covering most of Southern France, including the island of Corsica; we already used part of this data set to substantiate the local inhibitive effect after big wildfires. On the one hand, statistical modeling based on this larger data set could improve statistical power for inferring effects that are homogeneous over larger regions, while it also allows identifying regional differences at smaller scales in the relative risk of wildfires.

## 6. Acknowledgments

## References

Baddeley, A., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat Neerl*, 54(3):329–350.

Bonneu, F. (2007). Exploring and modeling fire department emergencies with a spatio-temporal marked point process. *Case Studies in Business, Industry and Governement Statistics*, 1:139–152.

Comas, C., Conde, J., and J, M. (2017). A second-order test to detect spatio-temporal anisotropic effects in point patterns. *Statistics*, page To appear.

Diggle, P., Chetwynd, A., Häggkvist, R., and Gooding, S. (1995). Second-order analysis of space-time clustering. *Stat Methods Med Res*, 4:124–136.

Gabriel, E. (2014). Estimating second-order characteristics of inhomogeneous spatio-temporal point processes: influence of edge correction methods and intensity estimates. *Methodoloy and Computing in Applied Probability*, 16(2):411:431.

Gabriel, E. (2016). Spatio-temporal point pattern analysis and modelling. In *Encyclopedia of GIS, 2nd Edition*.

Gabriel, E. and Diggle, P. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63:43–51.


 ISSN: 2102-6238

Gabriel, E., Rowlingson, B., and Diggle, P. (2013). `stpp`: An R package for plotting, simulating and analysing spatio-temporal point patterns. *J Stat Softw*, 53(2):1:29.

Gabriel, E., Wilson, D., Leatherbarrow, H., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, A., and Diggle, P. (2010). Spatio-temporal epidemiology of campylobacter jejuni enteritis, in an area of northwest england, 2000-2002. *Epidemiol Infect*, 138:1384–1390.

Genton, M., Butry, D., Gumpertz, M., and Prestemon, J. (2006). Spatio-temporal analysis of wildfire ignitions in the St Johns River water management district, Florida. *International Journal of Wildland Fire*, 15:87–97.

Gonzalez-Monsalve, J., Rodriguez-Cortes, F., Cronie, O., and Mateu, J. (2016). Spatio-temporal point process statistics: a review. *Spatial Statistics*.

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.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19).

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Møller, J. and Diaz-Avalos, C. (2010). Structured spatio-temporal shot-noise cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, 37(1):2–25.

Møller, J. and Ghorbani, M. (2012). Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Statistica Neerlandica*, 66:472–491.

Opitz, T. (2016). Latent Gaussian modeling and INLA: A review with focus on space-time applications. Technical report. Submitted to Journal de la Société Française de Statistique.

Pereira, P., Turkman, K., Amaral-Turkman, M., Sa, A., and Pereira, J. (2013). Quantification of annual wildfire risk; a spatio-temporal point process approach. *Statistica*, 73(1):55–68.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RESSTE Network (2016). Analyzing spatio-temporal data with R: Everything you always wanted to know – but were afraid to ask. Submitted to Journal de la Société Française de Statistique.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Diaz-Avalos, C., and Rue, H. (2014). Spatio-temporal log-gaussian cox processes for modelling wildfire occurrence: the case of catalonia, 1994–2008. *Environmental and Ecological Statistics*, 21(3):531–563.

Siino, M., Adelfio, G., Gabriel, E., and Mateu, J. (2017). Classes of anisotropic spatio-temporal log-gaussian cox processes to model earthquake events. Technical report.

Simpson, D., Illian, J., Lindgren, F., Sørbye, S., and Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*. In press.

Tamayo-Uria, I., Mateu, J., and Diggle, P. (2014). Modelling of the spatio-temporal distribution of rat sightings in an urban environment. *Spatial Statistics*.

Turner, R. (2009). Point patterns of forest fire locations. *Environmental and Ecological Statistics*, 16:197–223.

Xu, H. and Schoenberg, F. (2011). Point process modelling of wildfire hazard in Los Angeles county, California. *The Annals of Applied Statistics*, 5:684–704.


 ISSN: 2102-6238