Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires

August 11, 2008

Jesper Møller, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark.
Email: jm@math.aau.dk.

Carlos Diaz-Avalos, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Apartado Postal 20-726 Del. Alvaro Obregón, México D.F., C.P. 0100.
Email: carlos@sigma.imas.unam.mx.

Abstract: Spatio-temporal Cox point process models with a multiplicative structure for the driving random intensity, incorporating covariate information into temporal and spatial components, and with a residual term modelled by a shot-noise process, are considered. Such models are flexible and tractable for statistical analysis, using spatio-temporal versions of intensity and inhomogeneous K-functions, quick estimation procedures based on composite likelihoods and minimum contrast estimation, and easy simulation techniques. These advantages are demonstrated in connection to the analysis of a relatively large dataset consisting of 2796 days and 5834 spatial locations of fires. The model is compared with a spatio-temporal log-Gaussian Cox point process model, and likelihood-based methods are discussed to some extent.

Keywords: composite likelihood; Cox process; forest fires; inhomogeneous K-function; intensity; log-Gaussian process; minimum contrast estimation; multiplicative model; pair correlation function; Poisson process; simulation; shot-noise process; spatio-temporal point process.

1 Introduction

Cox processes (Cox, 1955) are very useful for modelling aggregated point patterns, in particular in the spatial case where the two main classes of models are log-Gaussian Cox processes and shot-noise Cox processes (Møller et al., 1998; Wolpert and Ickstadt, 1998; Brix, 1999; Møller, 2003; Diggle, 2003; Møller and Waagepetersen, 2004, 2007; Møller and Torrisi,

2005, Hellmund et al., 2008). For spatio-temporal point pattern data, spatio-temporal log-Gaussian Cox process models have recently found different applications (Brix and Møller, 2001; Brix and Diggle, 2001, 2003; Brix and Chadoeuf, 2002; Diggle et al., 2005). In this paper, we study instead spatio-temporal shot-noise Cox point process models, and demonstrate that such models are flexible and tractable for statistical inference, not least when compared to spatio-temporal log-Gaussian Cox processes. As an application example, we consider how to model forest fires, where details of the dataset are given in Section 2.3. Forest fires represent a problem of considerable social importance, see e.g. Brillinger et al. (2006) and Section 2.1.

Consider a spatio-temporal Cox process $\mathbf{X} = \{\mathbf{X}_t : t \in \mathbb{Z}\}$, with discrete time $t \in \mathbb{Z}$ (the set of integers) and $\mathbf{X}_t$ a planar point process. Underlying this is a stochastic process $\mathbf{\Lambda} = \{\mathbf{\Lambda}_t : t \in \mathbb{Z}\}$, where each $\mathbf{\Lambda}_t = \{\lambda(u,t) : u \in \mathbb{R}^2\}$ is a locally integrable non-negative stochastic process, so that conditional on $\mathbf{\Lambda}$, the $\mathbf{X}_t$ are mutually independent Poisson processes with intensity functions $\mathbf{\Lambda}_t$. Local integrability means that for any bounded $B \subset \mathbb{R}^2$ and $t \in \mathbb{Z}$, $\int_B \lambda(u,t)\,\mathrm{d}u < \infty$, and hence $\mathbf{X}_t$ can be viewed as a locally bounded subset of $\mathbb{R}^2$. We assume a multiplicative decomposition of the random intensity,

$$\lambda(u,t) = \lambda_1(u)\lambda_2(t)S(u,t), \quad \mathrm{E}S(u,t) = 1, \quad (u,t) \in \mathbb{R}^2 \times \mathbb{Z} \tag{1}$$

where $\lambda_1(u)$ and $\lambda_2(t)$ are non-negative deterministic functions, while $S(u,t)$ is a spatio-temporal process with unit mean. A spatio-temporal log-Gaussian Cox process is obtained if $S(u,t)$ is a log-Gaussian process. Diggle et al. (2005) refer to $\lambda_1(u)\lambda_2(t)$ as the 'normal pattern', and consider a non-parametric model for the 'pattern of spatial variation (λ_1)', a parametric model for the 'pattern of temporal variation (λ_2)', and a parametric log-Gaussian model for the 'residual (S)'. Our model is different in that it incorporates important covariate information into $\lambda_1(u)$ and $\lambda_2(t)$ (Section 5.1) so that the model becomes inhomogeneous in both space and time, and it uses a shot-noise model for the residual process (see next paragraph) which is considered to account for unobserved random effects (including unobserved covariates). For the purpose of identifiability, λ_1 is assumed to be a density over a spatial observation window W, so that in our application example, λ_2 becomes the mean number of forest fires in W per day.

We consider the particular case where

$$S(u,t) = \delta \sum_{s=-\infty}^{\infty} \sum_{y \in \mathbf{\Phi}_s} \varphi(u-y, t-s) \tag{2}$$

where δ is a positive parameter, the second sum is over the points of a stationary Poisson process $\mathbf{\Phi}_s$ with intensity $\omega > 0$ (not depending on $s \in \mathbb{Z}$), and φ is a joint density on $\mathbb{R}^2 \times \mathbb{Z}$ with respect to the product measure of Lebesgue measure on $\mathbb{R}^2$ and counting measure on $\mathbb{Z}$. Hence the bound $\mathrm{E}S(u,t) = 1$ in (1) means that $\delta = 1/\omega$. Further, we assume that the point processes $\mathbf{\Phi}_t$, $t \in \mathbb{Z}$, are mutually independent. Note that the point processes $\mathbf{X}_t$ are dependent, unless $\varphi(u,t) = 0$ whenever $t \neq 0$. Many formulae and calculations reduce when the kernel φ is separable,

$$\varphi(u,t) = \phi(u)\chi(t), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{Z} \tag{3}$$

where ϕ is a density function on $\mathbb{R}^2$ and χ is a probability density function on $\mathbb{Z}$.

For the forest fire dataset, we have daily data together with covariate information about vegetation, elevation, slope, exposure, and temperature. We aim at fitting a relatively simple parametric model providing a good descriptive fit to the data and accounting for the dependence of covariates. Since we consider a rather large dataset consisting of 2796 days and 5834 spatial locations of fires, and the likelihood is intractable (Section 2.2), we focus on developing quick estimation procedures based on composite likelihoods and minimum contrast estimation procedures, extending ideas used in the spatial case (Møller and Waagepetersen, 2004, 2007) to the spatio-temporal case. More complicated likelihood-based methods are briefly discussed at the end of the paper. Much of the methodology presented apply as well on other problems than forest fires, including spatial epidemiology, where e.g. it could be interesting to apply our approach for the spatio-temporal incidence of non-specific gastroenteric symptoms considered in Diggle et al. (2003, 2005) and Diggle (2007).

The remainder of this paper is organized as follows. Section 2 introduces some notation, presents the dataset, and discusses questions of scientific interests. Sections 3-4 study various useful properties of the spatio-temporal shot-noise Cox process $\mathbf{X}$ relying only on the general structure (1)-(2) or (1)-(3). This includes the interpretation and simulation of the process as a Poisson cluster process, and how to define and estimate by non-parametric methods useful summary statistics such as the intensity, pair correlation, and inhomogeneous K-functions. Section 5 specifies our parametric model for λ_1, λ_2, and φ, under which further useful results can be derived. Section 6 fits the model, using the intensity and inhomogeneous K-functions in connection to estimation equations based on partial likelihoods and minimum contrast estimation procedures, and we use various summary statistics and simulations to check the fitted model. Section 7 concludes with a comparison to spatio-temporal log-Gaussian Cox processes and a discussion on prediction and likelihood based analysis.

2 Preliminaries

2.1 Forest fires

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. Fire is also important for the biological and ecological interrelations between many animal and plant species. It has the potential to change the species composition and hence the landscape. In many regards, fire can be thought as a grazing animal that removes plant material and debris, thereby giving many seeds that remained dormant in the forest soil a chance to germinate.

The analysis of forest fire occurrences is a research area that has been active for many years. Fire-history studies commenced in the early 20th century, mainly in the United States and Australia. Such studies are used to analyze the extent of fires and the timing of their occurrences. However, because the variables ignitions and lightning strikes do not correlate well, fire history data do not necessarily reflect the actual fire pattern (McKenzie et al.,

2000). Statistical modelling of forest fires appeared in the late 1970's with the works of Wilkins (1977) and Dayananda (1977). More recent statistical works include Peng et al. (2005), Brillinger et al. (2006), and the references therein.

The forest fire dataset considered in Section 2.3 and further on in this paper is from Northwestern America. In the forest ecosystems of Northwestern America, fire is the most important disturbance in a wide range of geographic scales, and it is difficult to visualize the existence of wide forest areas without the presence of an intense fire pattern. Yet however, fire is considered a hazard to human life and property. In the first half of the past century, this lead to a campaign to suppress fires from many American forests. Nowadays, ecologists and government agencies have realized the damage that such suppression programs were doing and now management strategies are different. Good management practices require understanding the role that biological and physical factors play in the pattern of fire occurrences in space and in time, and to assess the potential risk posed by such fire pattern to human property, it is necessary to develop statistical models. Thus, high quality information about space-time dynamics of fire is needed for long term resource management of forest ecosystems (De Long, 1998; McKenzie et al., 2000). In a forest stand, the risk of fire is usually related to variables such as air temperature and humidity, vegetation type, elevation and rainfall (Besie and Johnson, 1995). It is unlikely that proper conditions for fire ignition be present at the same time in a broad area, so fire occurrence may be considered as a local phenomena. The local random nature of fire ignitions as well as its dynamics in time permit to idealize the occurrence of fires as a space-time point process. Part of the spatial variation in fire occurrences is expected to be explained by covariate information available, but the rest of the spatial variation remains unexplained and may be modelled by some spatial random process. In this paper it will be the residual process S in (1) given by the shot-noise process (2).

2.2 Notation and likelihood

At this place it is appropriate to introduce some further notation and briefly discuss the likelihood function for the spatio-temporal shot-noise Cox process.

We consider a bounded spatial observation window $W \subset \mathbb{R}^2$ and a finite temporal observation window $T \subset \mathbb{Z}$, so that for each time $t \in T$, a finite point pattern $\mathbf{x}_t \subset W$ is observed. Thus $\mathbf{x} = \{\mathbf{x}_t : t \in T\}$ specifies the data, apart from any covariate information (the specific notation for the covariates are given in Section 2.3). We consider $\mathbf{x}_t$ and $\mathbf{x}$ to be realizations of $\mathbf{X}_t \cap W$ and $\mathbf{X}_{W\times T} = \{\mathbf{X}_t \cap W : t \in T\}$, respectively. The corresponding unobserved random intensity is denoted $\mathbf{\Lambda}_{W\times T} = \{\mathbf{\Lambda}_t \cap W : t \in T\}$. Due to the multiplicative structure (1), the 'spatial margin' $\mathbf{x}_W = \cup_{t\in T}\mathbf{x}_t$, i.e. all observed points in W, and the 'temporal margin' $\mathbf{n}_T = \{n_t : t \in T\}$, where $n_t = n(\mathbf{x}_t)$ is the observed number of points at time t, will naturally play a particular important role. The corresponding point processes to these margins are denoted $\mathbf{X}_W = \mathbf{X}_{\cup T} \cap W$ and $\mathbf{N}_T = \{N_t : t \in T\}$, where $\mathbf{X}_{\cup T} = \cup_{t\in T}\mathbf{X}_t$ is almost surely a disjoint union, and where $N_t = n(\mathbf{X}_t \cap W)$.

Denote θ the collection of all unknown model parameters, i.e. regression parameters for $\lambda_1(u)$ and $\lambda_2(t)$ together with parameters for the residual process $S(u,t)$. These parameters are specified in detail in Section 5. Until Section 5, for ease of presentation, we suppress in

the notation that $\lambda_1(u)$, $\lambda_2(t)$, and $S(u,t)$ depend on θ. The likelihood function $l(\theta;\mathbf{x})$ is proportional to the density of the spatio-temporal Cox process $\mathbf{X}_{W\times T}$ with respect to the distribution defined by i.i.d. unit rate Poisson processes on W and indexed by T. We have

$$l(\theta;\mathbf{x}) = \left[\prod_{u\in\mathbf{x}_W} \lambda_1(u)\right]\left[\prod_{t\in T}\lambda_2(t)^{n_t}\right] \times \mathrm{E}_\theta\left[\exp\left(-\sum_{t\in T}\int_W \lambda_1(u)\lambda_2(t)S(u,t)\,\mathrm{d}u\right)\prod_{t\in T}\prod_{u\in\mathbf{x}_t} S(u,t)\right] \tag{4}$$

where E_θ denotes expectation under the model with parameter θ, see e.g. Møller and Waagepetersen (2007). In general the expectation in (4) has no closed form expression, and the likelihood is intractable. In most of this paper, we avoid this problem and use simple and quick inference procedures, while Section 7 discusses ways of doing more complicated likelihood-based inference.

2.3 Data

For our application example, W is the area known as the Blue Mountains, a 71,351 km^2 large region included mainly in Eastern Oregon, US, see Figure 1. This area is characterized by the perennial presence of smoldering fires. Due to the increased pressure from human activities and to the suppression of fires since early 1930's causing an accumulation of plant material, there is a potential for the outburst of wild fires. The main ignition sources are lightning strikes, camp fires, and machinery (Agee, 1993). We consider only lightning-caused fires, which comprise over 90% of the total fires observed in the Blue Mountains.

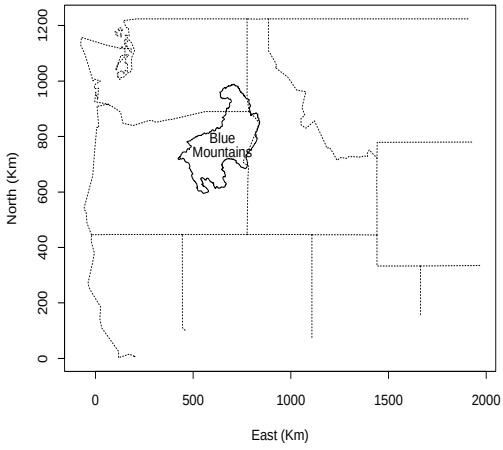

Figure 1: Geographic location of the Blue Mountains.

The fires were reported on a daily basis from April 1, 1986 to November 25, 1993, where the total number of fires in the Blue Mountains area was 5834. Accordingly, $T = \{1,\ldots,m\}$

with $m = 2796$ days, and $\mathbf{x}_t$ is the observed point pattern of forest fires on day t, with $t = 1$ corresponding to 4/1/1986 and $t = m$ to 11/25/1993. Also spatial covariate information is available (see Figure 2), where for each spatial location $u \in W$ and time $t \in T$, $V(u)$ denotes the vegetation type classified into 9 categories (left panel), $C(u)$ the slope-exposure type classified into 16 categories (central panel), and $E(u)$ the elevation (right panel) centered around 1750 m, the average elevation in the study area. Here a subdivision consisting of rectangular cells of size 3.51 km in the East-West direction and 2.54 km in the North-South direction is used so that $V(u), C(u), E(u)$ can be considered to be (approximately) constant within each cell. Furthermore, temporal covariate information is provided by the average temperature $T(t)$ during each day t (see Figure 4). The average temperature is computed from temperature records reported for three meteorological stations inside the study area.

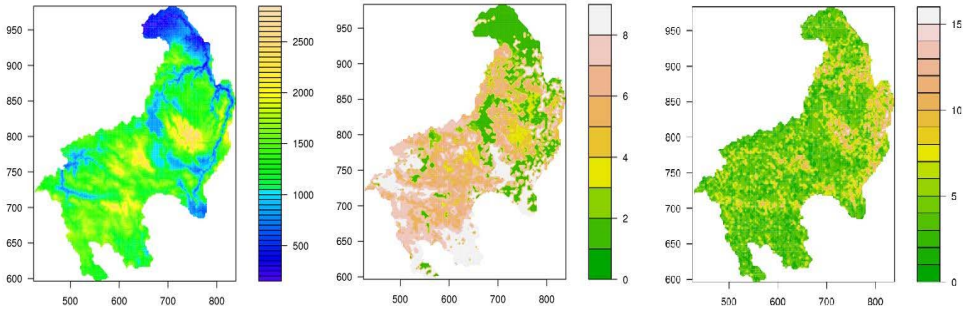

Figure 2: Elevation map in meters above sea level (left panel), spatial distribution of the 9 vegetation categories (central panel) and the 15 slope-exposure categories (right panel) in the Blue Mountains area.

Figure 3 shows the pattern $\mathbf{x}_W$ of all fires within W together with a non-parametric estimate of λ_1 (Diggle, 1985). Clearly, fire presence is more intense in certain areas of the Blue Mountains and seems related to the spatial covariates in Figure 2. For example, fires are very rare at elevation below 1000 m above sea level, are common at elevations ranging from 1500 to 2200 m as well as for slope-exposure categories 3-6 (which correspond to moderate slopes facing southwards), and frequently occur in areas covered by vegetation types 5 and 6 (which correspond to Ponderosa pine and Douglas Fir). Figure 4 shows the number of fires n_t at the different days, two estimates of λ_2, a non-parametric estimate (Silverman, 1986) and a parametric estimate, where the latter is discussed in Section 6.1, and the temperature $T(t)$. There is an apparent seasonal pattern of fires over time, with most of the fires occurring in the period from late spring to early fall. Seemingly there is also some relationship between the patterns of fires and temperatures. Figure 5 shows the spatial patterns of fire occurrences in four consecutive weeks during the summer of 1987. These plots (and further similar plots which are omitted here) illustrate that at a weekly time scale, fires tend to occur in small clusters.

The space-time point pattern dataset considered in this paper was in Diaz-Avalos et al. (2001) aggregated into a pixel-time array, considering for each pixel-time location a

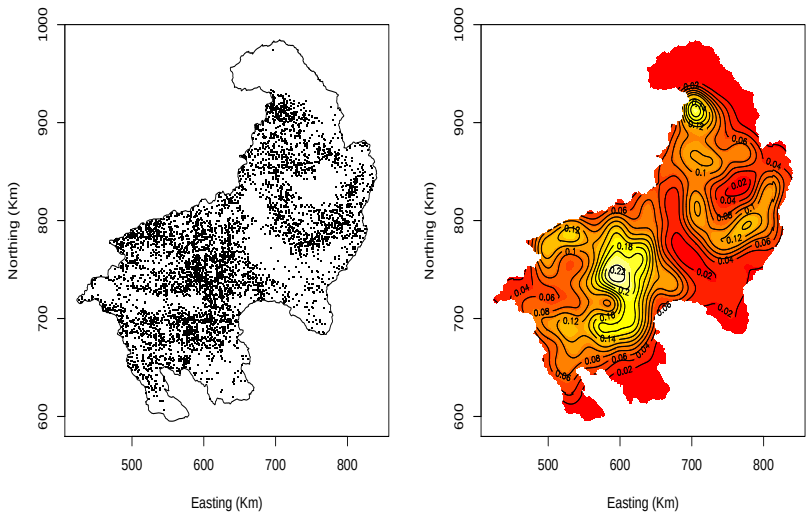

Figure 3: Spatial pattern of fire occurrences in the Blue Mountains from 04/01/1986 to 11/25/1993 (left panel) and non-parametric estimate of the spatial intensity function $\lambda_1(u)$ (right panel).

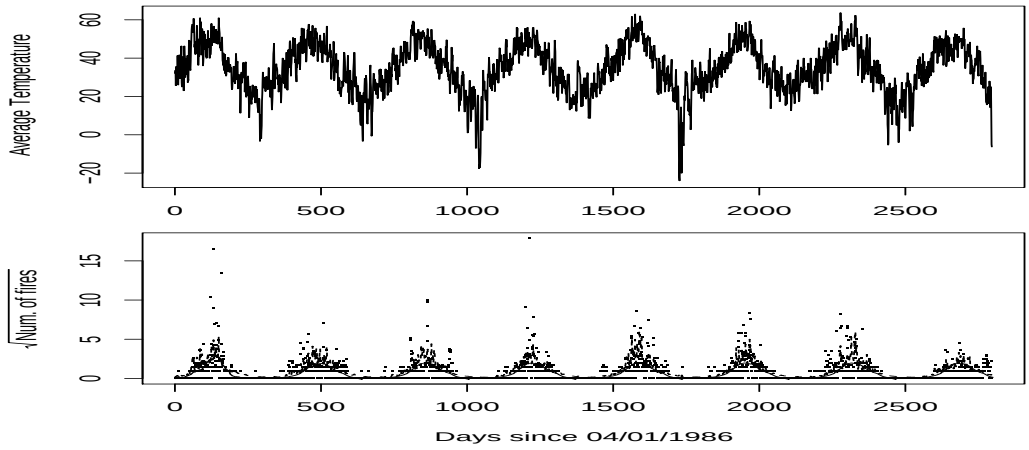

Figure 4: Upper panel: average daily temperature (Fahrenheit) in the Blue Mountains area. Lower panel: square root of the daily number of fire occurrences in the Blue Mountains (solid dots), a non-parametric estimate (solid line), and a parametric estimate (dashed line) of the temporal intensity function $\lambda_2(t)$.

binary variable which is one if a least one fire occurred and zero otherwise, where the pixels correspond to the rectangular cells of the subdivision described above. Diaz-Avalos et al. (2001) fitted then a generalized linear mixed model in a Bayesian framework.

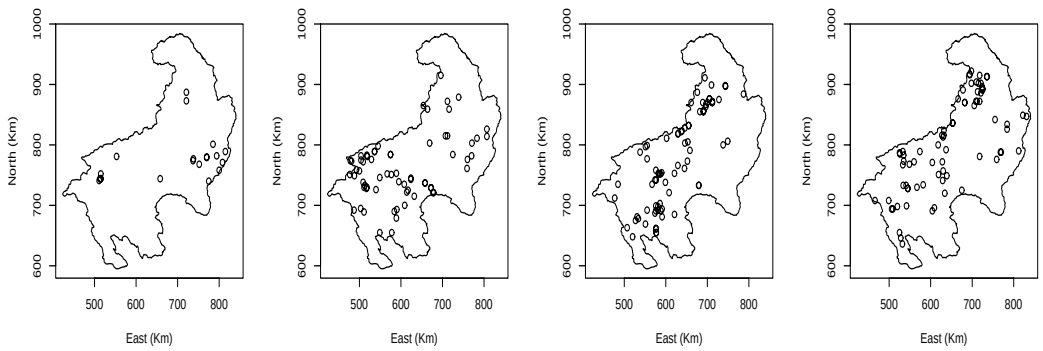

Figure 5: Spatial patterns of fire occurrences in four consecutive weeks.

3 General properties

This section discusses some useful general properties of the spatio-temporal shot-noise Cox process with the structure (1)-(2) or (1)-(3).

3.1 Poisson cluster process interpretation

We can view the spatio-temporal shot-noise Cox process $\mathbf{X}$ as a spatio-temporal Poisson cluster process constructed as follows. Let $\mathbf{\Phi} = \{\mathbf{\Phi}_t : t \in \mathbb{Z}\}$ be the spatio-temporal Poisson process underlying (2). Considering all points $y \in \mathbf{\Phi}_s$ for all times $s \in \mathbb{Z}$, we have that $\mathbf{X}_t$ conditional on $\mathbf{\Phi}$ is distributed as the superposition $\cup_{y,s}\mathbf{X}_t^{(y,s)}$ of mutually independent Poisson processes $\mathbf{X}_t^{(y,s)}$ on $\mathbb{R}^2$, where $\mathbf{X}_t^{(y,s)}$ has intensity function

$$\lambda_t^{(y,s)}(u) = \lambda_1(u)\lambda_2(t)\delta\varphi(u-y, t-s), \quad u \in \mathbb{R}^2.$$

Note that each 'cluster' $\mathbf{X}_t^{(y,s)}$ is finite, with the number of points (the 'offspring') being Poisson distributed with mean

$$n_t^{(y,s)} = \lambda_2(t)\delta \int \lambda_1(u)\varphi(u-y, t-s)\,\mathrm{d}u \tag{5}$$

and the offspring density is proportional to $\lambda_1(u)\varphi(u-y, t-s)$ for $u \in \mathbb{R}^2$. We refer to y as a 'mother point'. This simplifies in the separable case (3) where

$$n_t^{(y,s)} = \lambda_2(s)\chi(t-s)\delta \int \lambda_1(u)\phi(u-y)\,\mathrm{d}u$$

and the offspring density is proportional to $\lambda_1(u)\phi(u-y)$. Furthermore, still conditional on $\mathbf{\Phi}$, for all $y \in \mathbf{\Phi}_s$ and all $s, t \in \mathbb{Z}$, the clusters $\mathbf{X}_t^{(y,s)}$ are mutually independent.

The Poisson cluster process interpretation is due to (2), and it becomes useful when simulating $\mathbf{X}$ and making predictions as discussed in Sections 3.2 and 7. On the other hand, rather than interpreting the clustering as a mechanism causing forest fires, we consider (2) as a flexible and tractable way of modelling a random intensity function.

3.2 Simulation

In principle we can make a simulation of $\mathbf{X}_{W\times T}$ using a dominating spatio-temporal shot-noise Cox process constructed as follows. Assume that $\lambda_1^{\max} = \sup_{u\in W}\lambda_1(u)$ is finite. Since the distribution of $\mathbf{X}_{W\times T}$ does not depend on how $\lambda_1(u)$ is specified outside W, we can take $\lambda_1(u) = \lambda_1^{\max}$ for $u \notin W$. Suppose we have simulated $\mathbf{\Phi}$ (of course this will be impossible in practice, since $\mathbf{\Phi}$ is infinite). Consider a 'dominating' cluster $\mathbf{D}_t^{(y,s)}$, which is a Poisson process with intensity function

$$\lambda_1^{\max}\lambda_2(t)\delta\varphi(u-y,t-s) \geq \lambda_t^{(y,s)}(u), \quad u \in \mathbb{R}^2.$$

We assume that these dominating clusters are mutually independent for all $y \in \mathbf{\Phi}_s$ and $s,t \in \mathbb{Z}$. Note that $\mathbf{D}_t^{(y,s)}$ is usually easy to generate, particularly in the separable case (3) where the number of points in $\mathbf{D}_t^{(y,s)}$ is Poisson distributed with parameter $\lambda_2(t)\delta\lambda_1^{\max}\chi(t-s)$, and where the offspring density is $\phi(u-y)$, $u \in \mathbb{R}^2$. Then we obtain a simulation of each $\mathbf{X}_t \cap W$ by an independent thinning of $\mathbf{D}_t^{(y,s)}$, i.e. each point $u \in \mathbf{D}_t^{(y,s)}$ is included in $\mathbf{X}_t \cap W$ with probability $(\lambda_1(u)/\lambda_1^{\max})\mathbf{1}[u \in W]$, where $\mathbf{1}[\cdot]$ is the indicator function, and whether such points are included or not are mutually independent events.

In practice we suggest to make an approximate simulation, using a finite version of $\mathbf{\Phi}$, where we evaluate the error done by the approximation as explained below. Alternatively, perfect (or exact) simulation algorithms can be developed along similar lines as in Brix and Kendall (2002) and Møller (2003), however, for our application example and probably also most other applications, we find it much easier and sufficient to use the approximate simulation algorithm.

The finite version of $\mathbf{\Phi}$ is obtained by restricting it to a bounded region $\tilde{W}\times\tilde{T} \supseteq W\times T$. Let $\tilde{\mathbf{\Phi}}_t = \mathbf{\Phi}_t \cap \tilde{W}$, which is simply a homogeneous Poisson process on $\tilde{W}$ with intensity ω, and the processes $\tilde{\mathbf{\Phi}}_t$, $t \in \tilde{T}$, are mutually independent. We approximate the residual process (2) by

$$\tilde{S}(u,t) = \delta\sum_{s\in\tilde{T}}\sum_{y\in\tilde{\mathbf{\Phi}}_s}\varphi(u-y,t-s), \quad (u,t) \in \mathbb{R}^2\times\mathbb{Z} \tag{6}$$

and consider a spatio-temporal shot-noise Cox process $\{\tilde{\mathbf{X}}_t : t \in \mathbb{Z}\}$ driven by the random intensity

$$\tilde{\lambda}(u,t) = \lambda_1(u)\lambda_2(t)\tilde{S}(u,t), \quad (u,t) \in \mathbb{R}^2\times\mathbb{Z}. \tag{7}$$

A realization of $\tilde{\mathbf{X}}_{W\times T} = \{\tilde{\mathbf{X}}_t \cap W : t \in T\}$ is then an approximate simulation of $\mathbf{X}_{W\times T}$, and clearly, $\tilde{\mathbf{X}}_{W\times T}$ is almost surely a finite spatial-temporal Cox point process. The simulation of $\tilde{\mathbf{X}}_{W\times T}$ may easily be done along similar lines as above. For example, in the separable case (3), the steps are as follows, assuming $\lambda_1^{\max} = \sup_{u\in W}\lambda_1(u) < \infty$ and letting $\nu(s) = \sum_{t\in T}\lambda_2(t)\chi(t-s)$ for $s \in \tilde{T}$.

- Generate the mother processes $\tilde{\mathbf{\Phi}}_s$, $s \in \tilde{T}$.
- For each $s \in \tilde{T}$ and $y \in \tilde{\mathbf{\Phi}}_s$,
  - (i) generate a realization $n(y,s)$ from a Poisson distribution with parameter $\lambda_1^{\max}\nu(s)\delta$;

(ii) generate $n(y,s)$ i.i.d. points with density $\phi(u-y)$, $u \in \mathbb{R}^2$;

(iii) make an independent thinning, where we retain each point u from (ii) with probability $(\lambda_1(u)/\lambda_1^{\max})\mathbf{1}[u \in W]$;

(iv) to each retained point u from (iii) associate a time t_u generated from the density $p_s(t) = \lambda_2(t)\chi(t-s)/\nu(s)$, $t \in T$.

- For each $t \in T$, return all retained points u with $t_u = t$ (no matter which $s \in \tilde{T}$ and $y \in \tilde{\mathbf{\Phi}}_s$ are associated to u). These points constitute the approximate simulation of $\mathbf{X}_t \cap W$.

Since $\tilde{\mathbf{X}}_{W\times T}$ is a subprocess of $\mathbf{X}_{W\times T}$ some points may be missing, however, if $\tilde{W} \times \tilde{T}$ is chosen sufficiently large, we expect the two processes to be close. To evaluate this error, consider the mean number of missing points

$$M_T = \mathrm{E}\sum_{t\in T}(n(\mathbf{X}_t \cap W) - n(\tilde{\mathbf{X}}_t \cap W)).$$

As in Møller (2003), by conditioning on $\mathbf{\Phi}$, using Campbell's theorem (see e.g. Møller and Waagepetersen, 2004) and the fact that $\delta = 1/\omega$, we obtain

$$M_T = \int_W \lambda_1(u) \sum_{t\in T} \lambda_2(t) \left[1 - \sum_{s\in\tilde{T}} \int_{\tilde{W}} \varphi(u-y, t-s)\,\mathrm{d}y\right] \mathrm{d}u. \tag{8}$$

This expression simplifies in the separable case as shown in Section 5.4.

For example, suppose that $\varphi(u,t) = 0$ whenever $t < 0$ or $t > 1-k$, where $k \leq 1$ is a fix integer. Then mother points born at time s can only create offspring at times $s, s+1, \ldots, s+1-k$, and so we naturally take $\tilde{T} = \{k, \ldots, m\}$. As exemplified in Section 5.4, the choice of $\tilde{W}$ then depends on how small we want M_T.

3.3 Spatio-temporal margins

The spatio-temporal shot-noise Cox process possesses the appealing property that the superposition $\mathbf{X}_{\cup T} = \cup_{t\in T}\mathbf{X}_t$ is a spatial inhomogeneous shot-noise Cox process on $\mathbb{R}^2$, where the process is driven by the random intensity

$$\lambda_{\cup T}(u) = \delta\lambda_1(u) \sum_{s=-\infty}^{\infty} \sum_{t\in T} \lambda_2(t) \sum_{y\in\mathbf{\Phi}_s} \varphi(u-y, t-s), \quad u \in \mathbb{R}^2. \tag{9}$$

Similarly, $\mathbf{N}$ is a shot-noise Cox process on $\mathbb{Z}$ driven by

$$\lambda_{\int_W}(t) = \delta\lambda_2(t) \sum_{s=-\infty}^{\infty} \sum_{y\in\mathbf{\Phi}_s} \int_W \lambda_1(u)\varphi(u-y, t-s)\,\mathrm{d}u, \quad t \in \mathbb{Z}. \tag{10}$$

Thus techniques for analyzing and simulating spatial respective temporal shot-noise Cox processes apply for the spatio-temporal margins.

4 Summary statistics

The structures (1)-(2) or (1)-(3) imply simple moment expressions and second order intensity reweighted stationarity (Section 4.1). Thereby it becomes possible to define inhomogeneous K-functions (Sections 4.2-4.3), which are estimated by non-parametric methods and used for exploratory analysis. Moreover, the results become useful in connection to estimation and model checking for parametric models as discussed in Section 6.

4.1 Intensity and pair correlation functions

For $t, t_1, t_2 \in \mathbf{Z}$ with $t_1 \neq t_2$, let $\rho(\cdot, t)$ denote the intensity function of the spatial point process $\mathbf{X}_t$, $g((\cdot, t), (\cdot, t))$ the pair correlation function of $\mathbf{X}_t$, and $g((\cdot, t_1), (\cdot, t_2))$ the cross pair correlation function of $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$, see e.g. Møller and Waagepetersen (2004). Since $\mathbf{X}$ is a spatio-temporal Cox process, for any $t, t_1, t_2 \in \mathbf{Z}$ and $u, u_1, u_2 \in \mathbb{R}^2$,

$$\rho(u, t) = \mathrm{E}\lambda(u, t), \quad g((u_1, t_1), (u_2, t_2)) = \mathrm{E}[\lambda(u_1, t_1)\lambda(u_2, t_2)]/[\rho(u_1, t_1)\rho(u_2, t_2)]$$

and we refer to $\rho(\cdot, \cdot)$ and $g(\cdot, \cdot, \cdot, \cdot)$ as the intensity and pair correlation functions of $\mathbf{X}$. Note that g describes the 'normalized' second order moment properties, where $g((\cdot, t), (\cdot, t)) = 1$ if $\mathbf{X}_t$ is a Poisson process, and $g((\cdot, t_1), (\cdot, t_2)) = 1$ if $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$ are independent.

For our multiplicative model (1),

$$\rho(u, t) = \lambda_1(u)\lambda_2(t) \tag{11}$$

and ρ does not depend on the specification of the residual process, i.e. whether we consider a spatio-temporal shot-noise or log-Gaussian or another kind of Cox process. Further,

$$g((u_1, t_1), (u_2, t_2)) = \mathrm{E}[S(u_1, t_1)S(u_2, t_2)] \tag{12}$$

specifies the second order moment properties of the residual process. Combining (2) and (12) with the Slivnyak-Mecke theorem (Mecke, 1967; Møller and Waagepetersen, 2004), we obtain

$$g((u_1, t_1), (u_2, t_2)) = 1 + \delta\varphi * \tilde{\varphi}(u_1 - u_2, t_1 - t_2) \tag{13}$$

where $*$ denotes convolution, and $\tilde{\varphi}(u, t) = \varphi(-u, -t)$. Thus $g \geq 1$, reflecting the fact that $\mathbf{X}$ exhibits aggregation in both space and time as made clear by the Poisson cluster interpretation (Section 3.1).

By (13), g is space-time stationary, i.e. $g((u_1, t_1), (u_2, t_2)) = g(u, t)$ depends only on $u_1, u_2 \in \mathbb{R}^2$ and $t_1, t_2 \in \mathbb{Z}$ through the spatial difference $u = u_1 - u_2$ and the numerical time difference $t = |t_1 - t_2|$. Consequently, the point processes $\mathbf{X}_t$ are second order intensity reweighted stationary (Baddeley et al., 2000) with identical pair correlation functions $g(u, 0)$, and pairs of point processes $(\mathbf{X}_{t_1}, \mathbf{X}_{t_2})$ with the same value of $|t_1 - t_2| > 0$ are cross second order intensity reweighted stationary (Møller and Waagepetersen, 2004) with identical cross pair correlation function $g(u, |t_1 - t_2|)$.

4.2 Inhomogeneous K-functions

Henceforth we assume that g is isotropic, i.e. for all $u_1, u_2 \in \mathbb{R}^2$ and all $t_1, t_2 \in \mathbb{Z}$, we have that

$$g((u_1, t_1), (u_2, t_2)) = g(r, t)$$

depends only on the spatial distance $r = \|u_1 - u_2\|$ and the numerical time difference $t = |t_1 - t_2|$. This assumption will be satisfied for the parametric model of $\mathbf{X}$ introduced later, it simplifies the exposition in the sequel, and it is convenient for non-parametric estimation of the pair correlation function based on kernel estimation (Stoyan and Stoyan, 2000; Diggle et al., 2005). However, the various expressions of K-functions and their estimates below can easily be modified along similar lines as in Møller and Waagepetersen (2004) to cover the general case of (13).

The inhomogeneous spatio-temporal K-function at times $t_1, t_2 \in \mathbb{Z}$ is defined by

$$K(r, t_1, t_2) = K(r, t) = 2\pi \int_0^r s g(s, t) \, \mathrm{d}s, \quad r \geq 0, \ t = |t_1 - t_2|. \tag{14}$$

This is an extension of the definition of the spatio-temporal K-function for the stationary case considered in Diggle et al. (1995), where $K(\cdot, 0)$ is the usual inhomogeneous K-function of each spatial point process $\mathbf{X}_t$ (Baddeley et al., 2000), while if $t_1 \neq t_2$, $K(\cdot, t_1, t_2)$ is the inhomogeneous cross-K-function of $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$ (Møller and Waagepetersen, 2004). Clearly, since g is isotropic, there is a one-to-one correspondence between g and K.

For $u_1, u_2 \in \mathbb{R}^2$, let w_{u_1,u_2} denote Ripley's edge correction factor given by $2\pi\|u_1 - u_2\|$ divided by the length of arcs obtained by the intersection of W with the circle with center u_1 and radius $\|u_1 - u_2\|$ (Ripley, 1976). The non-parametric estimate

$$\hat{K}(r, t_1, t_2) = \sum_{u_1 \in \mathbf{x}_{t_1}, \, u_2 \in \mathbf{x}_{t_2} : \, u_1 \neq u_2} \frac{\mathbf{1}[\|u_1 - u_2\| \leq r] w_{u_1,u_2}}{\rho(u_1, t_1)\rho(u_2, t_2)} \tag{15}$$

is unbiased and in contrast to non-parametric estimation of g avoids kernel estimation. Consequently, for $t = 0, 1, \ldots, m-1$, the average

$$\hat{K}(r, t) = \frac{1}{m - t} \sum_{t'=1}^{m-t} \hat{K}(r, t', t' + t), \tag{16}$$

is an unbiased non-parametric estimate of $K(r, t)$. The unbiased property may be less important in practice, since the intensity function ρ is usually unknown, and we have to plug in an estimate of ρ in (15), where we use the non-parametric estimates of λ_1 and λ_2 given in Figures 3-4.

We have $K(r, 0) = \pi r^2$ in the special case of a Poisson process $\mathbf{X}_t$, and $K(r, t_1 - t_2) = \pi r^2$ in the special case where $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$ are independent, cf. Proposition 4.4 in Møller and Waagepetersen (2004). We consider L-functions given by $L = \sqrt{K/\pi}$, whereby $L(r, 0) - r$ is zero if $\mathbf{X}_t$ is a Poisson process, and $L(r, t_1 - t_2) - r$ is zero if $\mathbf{X}_{t_1}$ and $\mathbf{X}_{t_2}$ are independent.

For the spatio-temporal shot-noise Cox process, (13) and (14) imply that $L(r, t) - r \geq 0$, with in general a strict inequality (unless the kernel $\varphi(u, t)$ is zero whenever $t \neq 0$).

Since the variation of $\hat{K}(r, t_1, t_2)$ can be huge when n_{t_1} or n_{t_2} is small, for integers $c \geq 0$, define a truncated version by $\hat{K}_c(r, t_1, t_2) = \hat{K}(r, t_1, t_2)$ if both $n_{t_1} \geq c$ and $n_{t_2} \geq c$, and $\hat{K}_c(r, t_1, t_2) = 0$ otherwise. Denote $\hat{K}_c(r, t)$ the average of such $\hat{K}_c(r, t_1, t_2)$-functions, and $\hat{L}_c(r, t)$ the corresponding L-function. For example, if $t = 0$, Figure 6 shows clearly how the variation of $\hat{L}_c(r, 0)$-functions is reduced as c increases, and the functions $\hat{L}_{15}(r, 0)$ and $\hat{L}_{20}(r, 0)$ look very similar. Plots of $\hat{L}_c(r, t)$-functions with $c = 15$ and simulated 95%-inter-quantile envelopes obtained under a fitted spatio-temporal Poisson model, using an intensity function $\lambda_1(u)\lambda_2(t)$ given by either the non-parametric estimates in Figures 3-4 or the parametric estimates obtained later in Section 6.1, and assuming independence between the $\mathbf{X}_t$-processes, show clearly the poor fit of such Poisson models. For instance, Figure 7 shows such plots when $c = 15$, $t = 0, 1, 2$, and $\lambda_1(u)\lambda_2(t)$ is the parametric estimate. For details on how the 95%-inter-quantile envelopes are obtained, see Møller and Waagepetersen (2004).

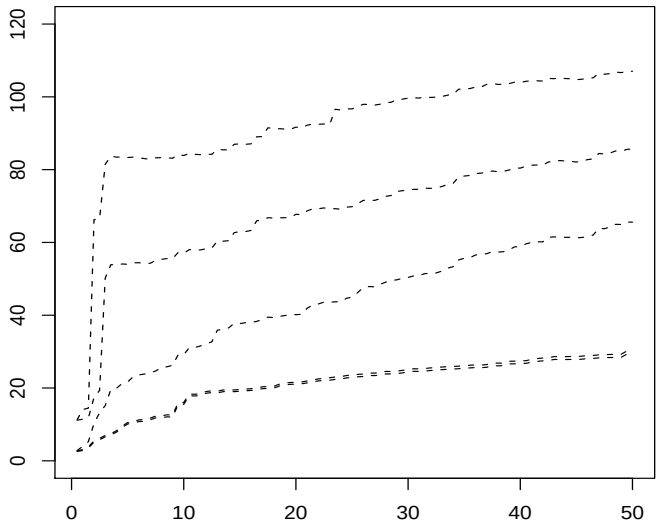

Figure 6: Non-parametric estimates $\hat{L}_c(r, 0) - r$ with $c = 0, 2, 5, 15, 20$ (from top to bottom).

4.3 Summaries for the spatio-temporal margins

The spatial shot-noise Cox process $\mathbf{X}_{\cup T}$ has intensity function

$$\rho_{\cup T}(u) = \lambda_1(u) \sum_{t \in T} \lambda_2(t), \quad u \in \mathbb{R}^2 \tag{17}$$

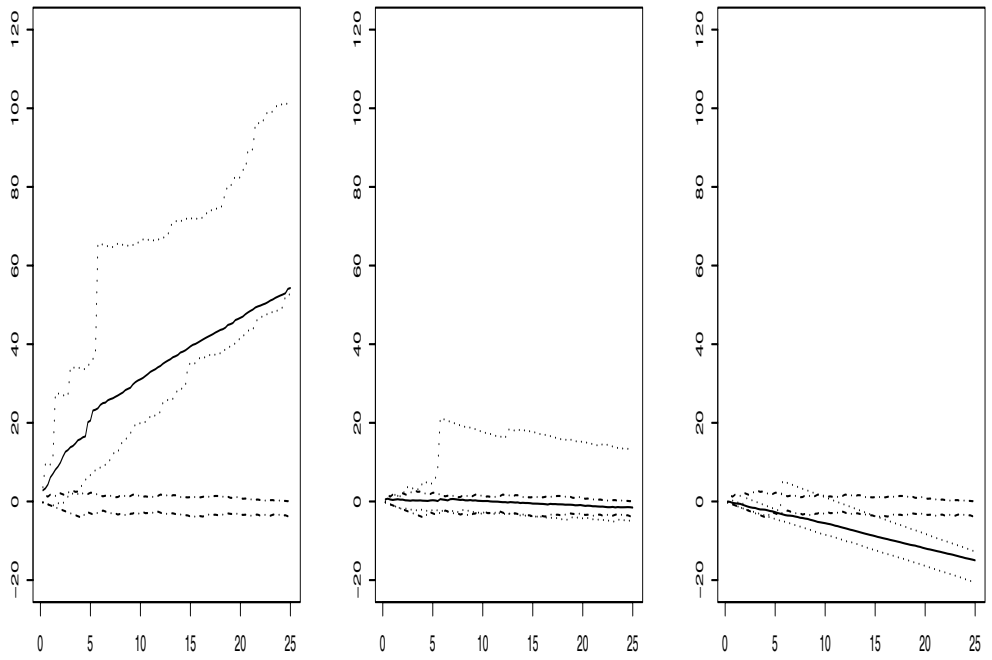

Figure 7: Non-parametric estimates $\hat{L}_c(0,r) - r$ (left panel), $\hat{L}_c(1,r) - r$ (central panel) and $\hat{L}_c(2,r) - r$ (right panel), together with 95%-inter-quantile envelopes obtained from 39 simulations of either the fitted spatio-temporal Poisson process (dot-dashed lines; $\lambda_1(u)\lambda_2(t)$ is given by the parametric estimates obtained in Section 6.1) or the fitted spatio-temporal shot-noise Cox process (dotted line; the model is fitted in Section 6). Here $c = 15$.

and pair correlation function

$$g_{\cup T}(u_1, u_2) = 1 + \delta \frac{\sum_{t_1,t_2 \in T} \lambda_2(t_1)\lambda_2(t_2)\varphi * \tilde{\varphi}(u_1 - u_2, t_1 - t_2)}{\sum_{t_1,t_2 \in T} \lambda_2(t_1)\lambda_2(t_2)}, \quad u_1, u_2 \in \mathbb{R}^2. \tag{18}$$

Equations (17) and (18) follow straightforwardly from (11) and (13), or alternatively by using (9) and the Slivnyak-Mecke Theorem. Note that $g_{\cup T}(u_1, u_2) = g_{\cup T}(u_1 - u_2)$, meaning that $\mathbf{X}_{\cup T}$ is second order intensity reweighted stationary. Since g is assumed to be isotropic, it follows that also $g_{\cup T}(u_1, u_2) = g_{\cup T}(\|u_1 - u_2\|)$ is isotropic. Consequently, $\mathbf{X}_{\cup T}$ has inhomogeneous K-function

$$K_{\cup T}(r) = 2\pi \int_0^r s g_{\cup T}(s)\, \mathrm{d}s, \quad r \geq 0. \tag{19}$$

Its corresponding L-function is denoted $L_{\cup T}$, and

$$\hat{K}_{\cup T}(r) = \sum_{u_1 \in \mathbf{x},\, u_2 \in \mathbf{x}:\, u_1 \neq u_2} \frac{\mathbf{1}[\|u_1 - u_2\| \leq r] w_{u_1,u_2}}{\rho_{\cup T}(u_1)\rho_{\cup T}(u_2)} \tag{20}$$

is an unbiased non-parametric estimate. As $\rho_{\cup T}(u)$ is unknown, we estimate it in (20) by using (17) and the non-parametric estimates of $\lambda_1(u)$ and $\lambda_2(t)$ from Figures 3-4. Figure 8

shows $\hat{L}_{\cup T}(r) - r$ together with 95%-inter-quantile envelopes obtained from 39 simulations of an inhomogeneous Poisson process on W with the intensity function given by the non-parametric estimate of $\rho_{\cup T}(u)$. Figure 8 clearly shows that the pattern of forest fires is more aggregated than the fitted inhomogeneous Poisson process.

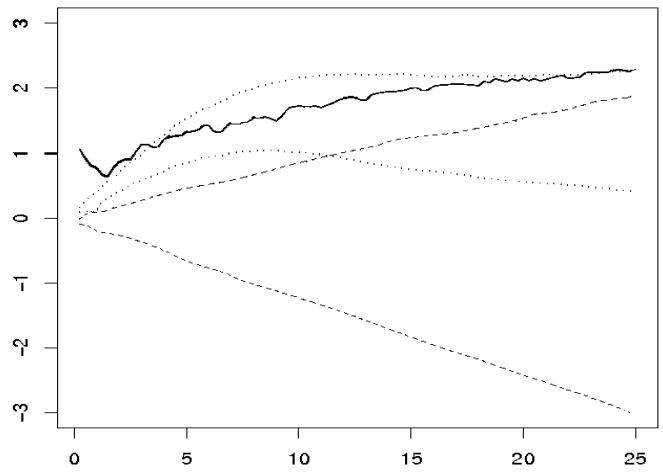

Figure 8: Non-parametric estimate of $\hat{L}_{\cup T}(r) - r$ (solid line) and 95%-inter-quantile envelopes obtained from 39 simulations of the fitted inhomogeneous Poisson process on W (dashed lines) or the spatio-temporal shot-noise Cox process in Section 6 (dotted lines).

The temporal point process $\mathbf{N}$ has intensity function $\lambda_2(t)$, $t \in \mathbb{Z}$, since we have imposed the identifiability condition that λ_1 integrates to one over W, cf. Section 1. For $t_1, t_2 \in \mathbb{Z}$, the covariance function $\mathrm{Cov}(t_1, t_2) = \mathrm{Cov}(N_{t_1}, N_{t_2})$ is given by

$$\mathrm{Cov}(t_1, t_2) = \delta \lambda_2(t_1) \lambda_2(t_2) \int_W \int_W \lambda_1(u_1) \lambda_1(u_2) \varphi * \tilde{\varphi}(u_1 - u_2, t_1 - t_2) \, \mathrm{d}u_1 \, \mathrm{d}u_2. \tag{21}$$

This follows from (11) and (13) or alternatively from the Slivnyak-Mecke Theorem in combination with (10). Equation (21) implies that the corresponding correlation function is stationary, i.e. $\mathrm{Corr}(t_1, t_2) = \mathrm{Corr}(|t_1 - t_2|)$, where

$$\mathrm{Corr}(t) = \frac{\int_W \int_W \lambda_1(u_1) \lambda_1(u_2) \varphi * \tilde{\varphi}(u_1 - u_2, t) \, \mathrm{d}u_1 \, \mathrm{d}u_2}{\int_W \int_W \lambda_1(u_1) \lambda_1(u_2) \varphi * \tilde{\varphi}(u_1 - u_2, 0) \, \mathrm{d}u_1 \, \mathrm{d}u_2}, \quad t = 0, 1, \ldots. \tag{22}$$

As expected, $\mathrm{Corr}(t) \geq 0$.

As exemplified in Section 5.3, (18)-(22) simplify when φ is a separable kernel as in (3). In particular, we then see that $N_t / \lambda_2(t)$ is a second-order stationary time series, with mean one and correlation function

$$\mathrm{Corr}(t) = \frac{\chi * \tilde{\chi}(t)}{\chi * \tilde{\chi}(0)}, \quad t = 0, 1, \ldots. \tag{23}$$

Hence, when $t \ll m$, the usual non-parametric estimate of the correlation function (e.g. Priestley, 1983) is given by

$$\widehat{\mathrm{Corr}}(t) = \frac{\sum_{s=1}^{m-t} [(n_s n_{s+t})/(\lambda_2(s)\lambda_2(s+t)) - 1]}{\sum_{s=1}^{m} [(n_s/\lambda_2(s))^2 - 1]}. \tag{24}$$

Figure 9 shows $\widehat{\mathrm{Corr}}(t)$ for $t = 1, \ldots, 19$, where λ_2 in (24) is replaced by the non-parametric estimate from Figure 4. For $t \geq 20$, $\widehat{\mathrm{Corr}}(t)$ is effectively zero.

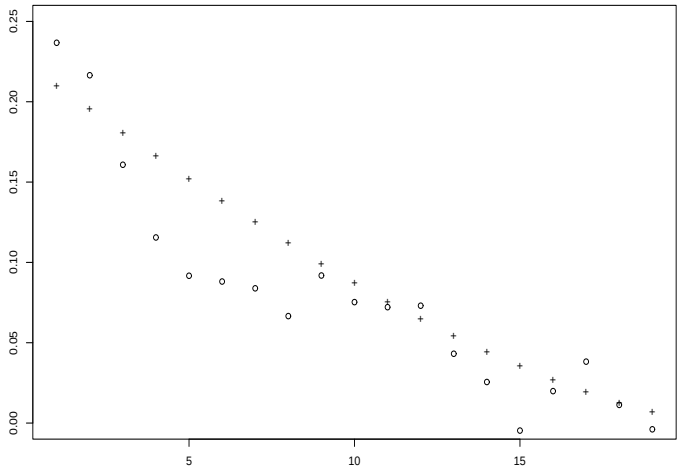

Figure 9: Correlations at lags $1, \ldots, 19$: the non-parametric estimate $\widehat{\mathrm{Corr}}(t)$ (open dots) and a parametric estimate (pluses) derived in Section 6.2.

5 Parametric model

Sections 5.1-(5.2) specify our parametric model for λ_1, λ_2, and φ. Thereby closed form expressions for theoretical correlation and K-functions can be derived (Section 5.3) and the error of the approximate simulation algorithm of the spatio-temporal shot-noise Cox process can be evaluated (Section 5.4).

5.1 Modelling the normal pattern

For the forest fire data, we assume that

$$\begin{aligned} \log \lambda_1(u) = & c(\beta^{V,C,E}) + \sum_{i=1}^{9} \beta_i^V \mathbf{1}[V(u) = i] + \sum_{j=1}^{16} \beta_j^C \mathbf{1}[C(u) = j] + \\ & \beta_1^E E(u) + \beta_2^E E(u)^2 \end{aligned} \tag{25}$$

and

$$\log \lambda_2(t) = \beta_0 + \beta_1^S \cos(\eta_t t) + \beta_2^S \sin(\eta_t t) + \beta_3^S \cos(2\eta_t t) + \beta_4^S \sin(2\eta_t t) + \beta_1^T T(t) + \beta_2^T T(t)^2 + \beta_3^T T(t)^3 + \beta_4^T T(t)^4 + \beta_5^T T(t)^5 \quad (26)$$

where the notation means the following. The β's are real regression parameters, where for the purpose of identifiability we impose the bounds that $\sum_{i=1}^{9} \beta_i^V = 0$ and $\sum_{i=1}^{16} \beta_i^C = 0$. Further, $\eta_t = 2\pi/365$ is the frequency for years with 365 days, and $\eta_t = 2\pi/366$ for leap years. Furthermore, $c(\beta^{V,C,E})$ is a normalizing constant depending on

$$\beta^{V,C,E} = (\beta_1^V, \ldots, \beta_9^V, \beta_1^C, \ldots, \beta_{16}^C, \beta_1^E, \beta_2^E)$$

so that λ_1 becomes a density over the spatial region W.

In (25), a log-linear form with respect to β_i^V and β_j^C is assumed for convenience, since we have no specific information on how to model the functional dependence of the covariates V and C. Furthermore, we consider a quadratic dependence of $E(u)$. Since a plot of the number of fires versus elevation showed a bell shaped form that resembles a normal distribution, with mean value about 1750 m (corresponding to $E(u) = 0$), β_2^E is expected to be negative and controlling how spread fire occurrences are around $-\beta_1^E/(2\beta_2^E)$, which is expected to be close to zero. Indeed this is confirmed by the parameter estimates $\hat{\beta}_2^E = -0.604$ and $-\hat{\beta}_1^E/(2\hat{\beta}_2^E) = -0.0243$ obtained by the method in Section 6.1.

In (26), β_0 is the general intercept, time-of-year effects are modelled by a sine-cosine wave plus its first harmonics, and the effect of the temperature is modelled by a fifth order polynomial. We have also investigated results based on including sixth and seven order terms which seemed unnecessarily, while including less terms provided a less good fit when we compared parametric estimates of λ_2 with the data n_t, $t \in T$.

5.2 Modelling the kernel

For the forest fire data, we also assume a separable kernel

$$\varphi(u,t) = \phi_{\sigma^2}^{(2)}(u)\chi_\zeta(t), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{Z}. \quad (27)$$

Further,

$$\phi_{\sigma^2}^{(2)}(u) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right)$$

is the density of a radially symmetric bivariate normal distribution with standard deviation $\sigma > 0$ (the spatial band-width). Furthermore, as suggested by Figure 9, $\chi_\zeta(t)$ is concentrated on $0, \ldots, t^* - 1$ with $t^* = 20$, and

$$\chi_\zeta(t) = \zeta(t^* - t), \quad t = 1, \ldots, t^* - 1, \quad (28)$$

is a decreasing linear function so that χ_ζ becomes a probability density function. Thus $\chi_\zeta(0) = 1 - \zeta t^*(t^* - 1)/2$ and $0 \le \zeta \le 2/[t^*(t^* - 1)]$. The parameters σ and ζ are correlation parameters, where the positive association between points (u_1, t_1) and (u_2, t_2) of $\mathbf{X}$ increases as σ or ζ increases.

5.3 Second order properties

Our parametric model assumptions imply the following closed form expressions for the second order characteristics.

From (27) we obtain the joint density

$$\varphi * \tilde{\varphi}(u,t) = \phi^{(2)}_{2\sigma^2}(u)\chi_\zeta * \tilde{\chi}_\zeta(t). \tag{29}$$

Note that $\chi_\zeta * \tilde{\chi}_\zeta$ is a symmetric probability density on $1-t^*,\ldots,0,\ldots,t^*-1$, which for $t = 1,\ldots,t^*-1$ is given by

$$\begin{aligned}\chi_\zeta * \tilde{\chi}_\zeta(t) = {} & \zeta(t^*-t)(1-\zeta t^*(t^*-1)/2) + \zeta^2/6\times \\ & \left[(t^*-1)t^*(2t^*-1) - t(t+1)(2t+1) - 3t^*(t^*+t+1)(t^*-t-1)(t+1)\right].\end{aligned} \tag{30}$$

Combining (13) and (29), we obtain the pair correlation function of $\mathbf{X}$, which is seen to be isotropic, where $g((u_1,t_1),(u_2,t_2)) = g(r,t)$ is a decreasing function of both $r = \|u_1-u_2\|$ and $t = |t_1-t_2|$. It also follows that the positive association between points (u_1,t_1) and (u_2,t_2) of $\mathbf{X}$ increases as σ or ζ increases. Note that $g(r,0)$ is of the same form as the pair correlation function of an inhomogeneous modified Thomas process (Møller and Waagepetersen, 2007). Combining (13)-(14) and (29), we obtain the inhomogeneous K-function of $\mathbf{X}$,

$$K(r,t) = \pi r^2 + \delta\left[1-\exp\left(-r^2/(4\sigma^2)\right)\right]\chi_\zeta * \tilde{\chi}_\zeta(t). \tag{31}$$

By (18)-(19) and (27), the inhomogeneous K-function of $\mathbf{X}_{\cup T}$ becomes

$$\begin{aligned}K_{\cup T}(r) = {} & \pi r^2 + \delta\left[1-\exp\left(-r^2/(4\sigma^2)\right)\right]\times \\ & \frac{\sum_{t_1,t_2\in T}\lambda_2(t_1)\lambda_2(t_2)\chi_\zeta * \tilde{\chi}_\zeta(|t_1-t_2|)}{\sum_{t_1,t_2\in T}\lambda_2(t_1)\lambda_2(t_2)}.\end{aligned} \tag{32}$$

Finally, the correlation function of $\mathbf{N}$ is obtained by combining (23) and (30).

5.4 The error of the approximate simulation algorithm

For the approximate simulation algorithm in Section 3.2, we obtain the following details under our parametric model assumptions. The mean cluster size given by (5) becomes $n_t^{(y,s)} = \lambda_1(y)\lambda_2(s)\delta\chi_\zeta(t-s)$, and the offspring distribution is simply a bivariate normal distribution with mean y and independent coordinates with variance σ^2. Since we have observations at times $t = 1,\ldots,m$, and the clusters of offspring $\mathbf{X}_t^{(y,s)}$ are empty whenever $t-s \geq t^*$, we let $\tilde{T} = \{2-t^*,\ldots,m\}$. Hence, in terms of (8) the error of the approximate simulation algorithm becomes

$$M_T = \int_W \lambda_1(u)\sum_{t\in T}\lambda_2(t)\left[1-\omega e(u;\sigma)\right]\,\mathrm{d}u \tag{33}$$

where

$$e(u;\sigma) = \int_{\tilde{W}} \phi^{(2)}_{\sigma^2}(u-y)\,\mathrm{d}y.$$

We let the extended spatial window be a rectangle $\tilde{W} = [a_1, a_2] \times [b_1, b_2]$ so that the smallest distance from its boundary to W is 3σ. Then for $u = (v, w) \in \tilde{W}$,

$$e(u;\sigma) = [F((b_1 - v)/\sigma) - F((a_1 - v)/\sigma)][F((b_2 - w)/\sigma) - F((a_2 - w)/\sigma)]$$

with F denoting the cumulative standard normal distribution function. When σ is equal to its estimate obtained in Section 6.2, we obtain $M_T = 0.008$, indicating that effectively no points will be missing in a simulation.

6 Quick non-likelihood estimation methods

In general the likelihood function for θ is intractable, cf. (4). Møller and Waagepetersen (2007) provide a general discussion of quick non-likelihood estimation procedures based on the intensity and pair correlation functions of a parametric spatial point process model. These procedures divide into estimation equations motivated heuristically as limits of composite likelihood functions and estimation equations obtained by minimum contrast methods. This section adapts such estimation equations to the parametric spatio-temporal shot-noise Cox process considered in this paper. We let θ_1 specify the unknown parameters of the normal pattern and θ_2 the unknown parameters of the residual process, where θ_1 and θ_2 are variation independent, cf. Section 5. Sections 6.1-6.2 consider the estimation of θ_1 and θ_2, respectively.

6.1 Estimation of the regression parameters

The composite likelihood (Lindsay, 1988) based on the intensity function can be obtained in various ways. See Møller and Waagepetersen (2007) for the case of a single spatial point process, which easily extend to our spatio-temporal case. Asymptotic properties of maximum composite likelihood estimates are studied in Schoenberg (2004) and Waagepetersen (2007).

One way of obtaining a composite likelihood is to consider the limit of composite likelihood functions for Bernoulli trials concerning absence or presence of points of the point processes $\mathbf{X}_t \cap W$, $t \in T$, within infinitesimally small cells partitioning the spatial observation window W. Due to the multiplicative form of the intensity function (11), we consider a simpler procedure, where we separate into composite likelihoods for respective the spatial margin $\mathbf{x}_W$ and the temporal margin $\mathbf{n}_T$. Let $\theta_1 = (\theta_{1,1}, \theta_{1,2})$, $\theta_{1,1} = (\alpha, \beta^{V,C,E})$, and $\theta_{1,2} = (\beta_0, \beta_1^S \ldots, \beta_4^S, \beta_1^T, \ldots, \beta_5^T)$, where $\alpha = c(\beta^{V,C,E}) + \log \sum_{t \in T} \lambda_2(t; \theta_{1,2})$. By (17) and (25)-(26),

$$\log \rho_{\cup T}(u;\theta_1) = \alpha + \sum_{i=1}^{9} \beta_i^V \mathbf{1}[V(u) = i] + \sum_{j=1}^{16} \beta_j^C \mathbf{1}[C(u) = j] + \beta_1^E E(u) + \beta_2^E E(u)^2$$

is the log intensity function of $\mathbf{X}_W$, and $\lambda_2(t; \theta_{1,2})$ is the intensity function of $\mathbf{N}_T$. The log composite likelihoods become

$$L_W(\theta_1; \mathbf{x}_W) = \sum_{u \in \mathbf{x}_W} \log \rho_{\cup T}(u;\theta_1) - \int_W \rho_{\cup T}(u;\theta_1)\, \mathrm{d}u \tag{34}$$

corresponding to the log likelihood for a Poisson process with intensity function $\rho_{\cup T}(u;\theta_1)$, and

$$L_T(\theta_{1,2};\mathbf{n}_T) = \sum_{t\in T} n_t \log \lambda_2(t;\theta_{1,2}) - \sum_{t\in T} \lambda_2(t;\theta_{1,2}) \tag{35}$$

corresponding to the log likelihood for a Poisson model with mean $\lambda_2(t;\theta_{1,2})$. Note that (34)-(35) do not depend on the specification of the residual process, i.e. whether we consider a spatio-temporal shot-noise or log-Gaussian or another kind of Cox process.

The maximum composite likelihood estimate $\hat{\theta}_{1,2}$ based on (35) is easily found using e.g. the software package `R`. The corresponding parametric estimate of λ_2 is shown in Figure 4, and it shows some discrepancies to the non-parametric estimate of λ_2, though overall they both follow the general pattern of the daily number of fires. The parametric estimate tends to be higher, particularly during the winter months, since (26) considers the trend in the number of fires and not the 'local' number of fires, while the non-parametric kernel estimate and the scarce number of fires occurring during the winter months cause low values of the non-parametric estimate at the 'local' level.

Suppose we plug in the estimate $\hat{\theta}_{1,2}$ into (34). Maximization of (34) with respect to $\theta_{1,2}$ is complicated by the fact that α depends on the normalizing constant $c(\beta^{V,C,E})$. For computational convenience, we ignore this dependence and treat first α as a real parameter which is variation independent of $\beta^{V,C,E}$. We propose then to obtain the estimate $\hat{\theta}_{1,1}$ which maximizes (34) with respect to $\theta_{1,1}$. This corresponds to the maximum composite likelihood estimate based on the intensity function of $\mathbf{X}_W$ which is proportional to λ_1, when λ_1 is unnormalized. This estimate is easily found using the software package `spatstat` (Baddeley and Turner, 2005, 2006). Second we normalize $\lambda_1(u;\hat{\theta}_{1,1})$ so that it becomes a density function. Note that the estimate of $\beta^{V,C,E}$ is unaffected by this normalization.

The left panel in Figure 10 shows this normalized estimate $\lambda_1(u;\hat{\theta}_{1,1})$, which recognizes the presence of areas with low intensity of fires, mainly in the North-East part of W. Such areas correspond mostly to agricultural land. Although fires are not impossible in such regions, the absence of a good amount of dry debris in the ground makes a lightning-caused ignition a rare event. However, $\lambda(u,\hat{\theta}_{1,1})$ overestimates the intensity in the area between $750-800$ Km East and $750-820$ Km North, as well as in the two southern tips of W. Since the actual values of the estimates $\hat{\theta}_{1,1}$ and $\hat{\theta}_{1,2}$ may perhaps appear to be less interesting for the reader, we omit them here (the estimated values of β_1^E and β_2^E were given in Section 5.1). We just notice here that, as expected, positive large values $\hat{\beta}_i^V$ correspond to regions with more fires, negative and small values $\hat{\beta}_i^V$ correspond to regions with few (or none) fires, positive and large values $\hat{\beta}_j^C$ correspond to regions with exposure slopes facing South or with modest or low slopes, and negative and small values $\hat{\beta}_j^C$ correspond to regions with exposure slopes facing North or with steep slopes.

The right panel in Figure 10 shows the residuals obtained by subtracting the non-parametric estimate of λ_1 in Figure 3 (normalized so that it integrates to one) from $\lambda_1(u;\hat{\theta}_{1,1})$. The residual image shows zero values in the areas where the presence of fires is scarce. In such areas both the parametric and non-parametric estimates of λ_1 agree. For the areas where fire presence was more intense, the residual image does not show a systematic pattern of positive and negative values. This indicates that both estimators follow with an acceptable degree of approximation the fire pattern observed in W.

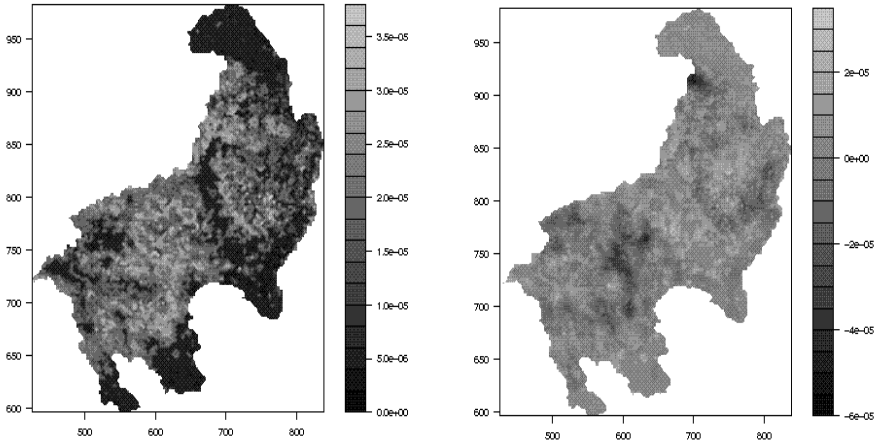

Figure 10: Left panel: parametric estimate of λ_1. Right panel: corresponding residuals when the non-parametric estimate in Figure 3 is subtracted.

6.2 Estimation of the residual process parameters

Let $\theta_2 = (\sigma, \delta, \zeta)$, where $\sigma > 0$, $\delta > 0$, $\zeta \in (0, 1)$ are variation independent. As in Section 6.1, one possibility is to separate into composite likelihoods for respective $\mathbf{x}_W$ and $\mathbf{n}_T$, but now using the second order product density

$$\rho^{(2)}((u_1, t_1), (u_2, t_2)) = \rho(u_1, t_1)\rho(u_1, t_1)g((u_1, t_1), (u_2, t_2)).$$

However, as in Møller and Waagepetersen (2007) we find the numerical computation of the corresponding score functions and their derivatives to be quite time consuming, and suggest instead minimizing a 'contrast' between the theoretical expressions of the second order characteristics $K_{\cup T}(u)$ and $\mathrm{Corr}(t)$ and their non-parametric estimates. Asymptotic properties of minimum contrast estimates are studied in the stationary case of spatial point processes in Heinrich (1992) and Guan and Sherman (2007).

The theoretical expression of the correlation function of $\mathbf{N}$,

$$\mathrm{Corr}(t; \zeta) = \chi_\zeta * \tilde{\chi}_\zeta(t)/\chi_\zeta * \tilde{\chi}_\zeta(0)$$

is given by (30), and it depends only on the parameter ζ. The minimum contrast estimate $\hat{\zeta}$ is the least square estimate obtained by minimizing

$$\sum_{t=1}^{t^*-1} \left[\mathrm{Corr}(t; \zeta) - \widehat{\mathrm{Corr}}(t)\right]^2$$

where $\widehat{\mathrm{Corr}}(t)$ is given by (24) when λ_2 is replaced by its parametric estimate from Section 6.1. We obtain $\hat{\zeta} = 0.00316$. The estimated function $\mathrm{Corr}(t; \hat{\zeta})$ is shown in Figure 9, and it is of a similar form as $\widehat{\mathrm{Corr}}(t)$.

The theoretical expression (32) of the inhomogeneous K-function of $\mathbf{X}_{\cup T}$ when we replace ζ by $\hat{\zeta}$ and $\lambda_2(t)$ by its parametric estimate $\lambda_2(t;\hat{\theta}_{1,2})$ is

$$K_{\cup T}(r;\sigma,\delta) = \pi r^2 + \delta' \left[1 - \exp\left(-r^2/(4\sigma^2)\right)\right]$$

where

$$\delta' = \delta \times \frac{\sum_{t_1,t_2 \in T} \lambda_2(t_1;\hat{\theta}_{1,2})\lambda_2(t_2;\hat{\theta}_{1,2})\chi_{\hat{\zeta}} * \tilde{\chi}_{\hat{\zeta}}(|t_1 - t_2|)}{\sum_{t_1,t_2 \in T} \lambda_2(t_1;\hat{\theta}_{1,2})\lambda_2(t_2;\hat{\theta}_{1,2})}.$$

Since δ and δ' are proportional, $\sigma > 0$ and $\delta' > 0$ are variation independent. The minimum contrast estimate $(\hat{\delta}', \hat{\sigma})$ is obtained by minimizing

$$\int_0^a \left(K_{\cup T}(r;\sigma,\delta)^b - \hat{K}_{\cup T}(r)^b\right)^2 \mathrm{d}r \tag{36}$$

where $\hat{K}_{\cup T}(r)$ is given by (17) and (20) when λ_1 and λ_2 are replaced by their parametric estimates from Section 6.1. Further, a and b are user-specified parameter. Partly following the recommendations in Diggle (2003), taking $a = 25$ and $b = 1/4$, we obtain $\hat{\sigma} = 6.363$ and $\hat{\delta} = 2672.725$, corresponding to an intensity $\hat{\omega} = 1/\hat{\delta} = 0.000374$ for each mother Poisson process $\mathbf{\Phi}_s$. For larger values $a \geq 25$, approximately the same estimates are obtained.

As a model check, Figures 7-8 shows the 95%-inter-quantile envelopes obtained from 39 simulations of the fitted space-time shot-noise Cox process. Figure 7 does not indicate any misfit, and both figures show a much better fit than for the Poisson model. In Figure 8, for distances beyond 2.5 Km, the fitted Cox model seems to capture well the second order characteristics for the point pattern of all fires. Given the size of W, and the size 3.51×2.54 Km of regions over which we have information about the spatial covariates (Section 2.3), the misfit at distances shorter than 2.5 Km is not surprising. This misfit is also not too important, since modelling (and prediction) of fire risk at local scale in most cases is of less interest.

7 Concluding remarks

Forest fires have an important influence on the environment, human health (often leading to fatalities) and property. As human population grows, it is becoming more important to design fire management plans. Overall our fitted spatio-temporal shot-noise Cox process describes well the daily fire patterns in the Blue Mountains from 04/01/1986 to 11/25/1993 by accounting for spatial and temporal covariates as well as for unobserved random effects (including unobserved covariates). It provides fire ecologists with a useful tool for fire management plans.

We have demonstrated how various summary statistics, particularly different kinds of inhomogeneous K-functions, can be constructed and used for exploratory analysis and model checking. Composite likelihoods and summary statistics have furthermore been used for obtaining fast estimation procedures. In fact the methodology in the previous sections may apply easily if the spatial-temporal shot-noise residual term S is replaced by a spatial-temporal log-Gaussian process, since first and second order moment expressions and simulation of spatial-temporal log-Gaussian processes are well-known. In the sequel we briefly

discuss how one could further proceed with a (much more time consuming) simulation based likelihood analysis, including how to make predictions. This will clarify some computational advantages of using the spatio-temporal shot-noise Cox process over the spatio-temporal log-Gaussian Cox process.

7.1 Prediction

For a specified (or estimated) spatial-temporal shot-noise Cox process $\mathbf{X}$, it may be of interest first to predict the unobserved random intensity function $\mathbf{\Lambda}$ given the data $\mathbf{X}_{W\times T} = \mathbf{x}$, and second to make daily predictions. Since the relevant conditional distributions are complicated, both predictions will be based on Markov chain Monte Carlos (MCMC) simulations, which also will be an important ingredient in connection to MCMC based likelihood inference in Section 7.2.

7.1.1 Predicting the random intensity

For our parametric model, the conditional distribution of $\mathbf{X}_{W\times T}$ given $\mathbf{\Lambda}$ is specified by the distribution of $\mathbf{X}_{W\times T}$ given the mother points $\{\mathbf{\Phi}_t : t \in \tilde{T}\}$, where $\tilde{T} = \{k, \dots, m\}$ and $k = -18$, cf. Section 5. Since this involves infinite many mother points, we consider the approximation $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}$ from Section 3.2. $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}} = \{\tilde{\mathbf{\Phi}}_t : t \in \tilde{T}\}$ consists of i.i.d. homogeneous Poisson processes $\tilde{\mathbf{\Phi}}_t$ on $\tilde{W}$, with intensity ω, and for $(u,t) \in W \times T$, we approximate $\lambda(u,t)$ by $\tilde{\lambda}(u,t)$ given in (7). Let $\mathbf{y}_t \subset \tilde{W}$ denote a finite point pattern (corresponding to a realization of $\tilde{\mathbf{\Phi}}_t$), and set $\mathbf{y} = \{\mathbf{y}_t : t \in \tilde{T}\}$. Conditional on the data $\mathbf{x}$, with respect to the distribution defined by i.i.d. unit rate Poisson processes on $\tilde{W}$ and indexed by $\tilde{T}$, the conditional density of $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}$ is given by

$$p(\mathbf{y}|\mathbf{x}) \propto \left[\prod_{t\in T} \exp\left(-\int_W \tilde{\lambda}(u,t)\,\mathrm{d}u\right) \prod_{u\in\mathbf{x}_t} \tilde{\lambda}(u,t)\right] \left[\prod_{t\in\tilde{T}} \omega^{n(\mathbf{y}_t)}\right] \tag{37}$$

where $\tilde{\lambda}(u,t)$ is obtained by replacing each $\tilde{\mathbf{\Phi}}_s$ in (6) by $\mathbf{y}_s$. The predictive distribution of the unobserved intensity function $\mathbf{\Lambda}_{W\times T}$ given the data is then approximated by (37). Note that in (37), for at least one $t \in \tilde{T}$, $\mathbf{y}_t$ has to be non-empty ($\mathbf{y}_t \neq \emptyset$), unless we had the unusual situation that the data $\mathbf{x}$ was empty. Note also that for each $t \in \tilde{T}$, the 'full conditional' of (37), i.e. the conditional distribution of $\tilde{\mathbf{\Phi}}_t$ given both all $\tilde{\mathbf{\Phi}}_s = \mathbf{y}_s$ with $s \in T \setminus \{t\}$ and $\mathbf{X}_{W\times T} = \mathbf{x}$, has density

$$p(\mathbf{y}_t|\mathbf{y}_s,\, s \neq t,\, \mathbf{x}) \propto \omega^{n(\mathbf{y}_t)} \prod_{t'\in T} \exp\left(-\int_W \tilde{\lambda}(u,t')\,\mathrm{d}u\right) \prod_{u\in\mathbf{x}_{t'}} \tilde{\lambda}(u,t'). \tag{38}$$

To simulate from (37), we may extend the Metropolis-Hastings algorithm for spatial shot-noise Cox processes studied in Møller (2003) to spatio-temporal shot-noise Cox processes. The extension is a Metropolis-within-Gibbs algorithm, using e.g. a systematic updating scheme by running through $t \in T$ and updating from each full conditional (38) using a birth-death Metropolis-Hastings algorithm (Geyer and Møller, 1994; Geyer, 1999; Møller

and Waagepetersen, 2003). For example, we may let a birth proposal happen with probability 0.5, where the new born mother point in $\tilde{W}$ has a density $f(y)$ which is proportional to an extension of $\lambda_1(y)$ to $\tilde{W}$, while a death proposal consists of omitting a uniformly selected point from the existing mother points. The acceptance probability of the proposal is then given as in the above-mentioned references. It can easily be calculated, since we can easily calculate the conditional density 37, as we can easily successively calculate $\tilde{\lambda}(u',t') = \lambda_1(u')\lambda_2(t')\tilde{S}(u',t')$ because

$$\tilde{S}(u',t') = \delta \sum_{s\in\tilde{T}} \sum_{y\in\mathbf{y}_s} \phi^{(2)}_{\sigma^2}(u'-y)\chi_\zeta(t'-s) \tag{39}$$

is a sum. For instance, in case of a birth proposal $\mathbf{y}'_t = \mathbf{y}_t \cup \{u\}$, we update $\tilde{\lambda}$ by

$$\tilde{\lambda}_{\text{new}}(u',t') = \tilde{\lambda}_{\text{old}}(u',t') + \lambda_1(u')\lambda_2(t')\delta\phi^{(2)}_{\sigma^2}(u'-u)\chi_\zeta(t'-t).$$

Stability properties of the Metropolis-within-Gibbs algorithm (irreducibility, aperiodicity, and various convergence properties) follow by similar arguments as in Møller (2003). Compared to the case of a spatio-temporal log-Gaussian Cox process, where a spatio-temporal extension of the Langevin-Hastings (or Metropolis adjusted Langevin) algorithm from Møller et al. (1998) has been used in Brix and Møller (2001), Brix and Diggle (2001, 2003), and Diggle et al. (2005), the Metropolis-within-Gibbs algorithm for the spatio-temporal shot-noise Cox process is much simpler and, as previous experience for the spatial case (Møller and Waagepetersen, 2004, 2007) indicates, it may be expected to be much faster.

7.1.2 Predicting the future

Assume that $\varphi(u,t) = 0$ whenever $t < 0$. This is satisfied for our parametric spatial-temporal shot-noise Cox process.

Consider first the case of making a prediction for 'tomorrow' (i.e. at time $m+1$) given the data $\mathbf{x}$. Assume that $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}$ has already been predicted, cf. Section 7.1.1. First we simulate a realization of $\tilde{\mathbf{\Phi}}_{m+1}$. Since $\tilde{\mathbf{\Phi}}_{m+1}$ is a Poisson process which is independent of $(\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}, \mathbf{X}_{W\times T})$, this simulation step is very much easier than for a spatio-temporal log-Gaussian Cox process model which involves the Metropolis-Langevin algorithm. Second, conditional on realizations of both $(\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}, \mathbf{X}_{W\times T})$ and $\tilde{\mathbf{\Phi}}_{m+1}$, we can easily simulate $\mathbf{X}_{m+1}\cap W$, since it is approximately a Poisson process with intensity function $\tilde{\lambda}(u, m+1)$, $u \in W$. If M_{m+1} is the mean number of missing points of this approximation, then M_{m+1} is given by (8) or (33) when we replace T in these equations by $m+1$.

Similarly, at time $m+2$, when we have predicted $(\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}, \tilde{\mathbf{\Phi}}_{m+1}, \mathbf{X}_{m+1}\cap W)$, we can easily make predictions of first $\tilde{\mathbf{\Phi}}_{m+2}$ and second $\mathbf{X}_{m+2}\cap W$. And so on at times $m+3, m+4, \ldots$, where of course the error done by the approximations accumulates.

7.2 Simulation based likelihood inference

For a spatio-temporal log-Gaussian Cox process, Diggle et al. (2005) and Diggle (2007) briefly discuss approximate maximum likelihood estimation based on MCMC procedures as treated

in Møller and Waagepetersen (2004). A missing data approach is used, where $\mathbf{\Lambda}_{W\times T}$ is the missing data, and so MCMC simulations are based on the Langevin-Hastings algorithm for the conditional distribution of $\mathbf{\Lambda}_{W\times T}$ given $\mathbf{X}_{W\times T} = \mathbf{x}$. For our spatio-temporal shot-noise Cox process, the missing data would instead be the mother points which may generate offspring equal to the data $\mathbf{x}$, or in practice we consider instead $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}$ as an approximation of these mother points. Then MCMC simulations are expected to be easier and faster, cf. Section 7.1.1. However, for both types of Cox models, due to the very complicated likelihood, it may not be straightforward to find the approximate maximum likelihood estimate, which may involve a combination of Newton-Raphson and estimation procedures for ratios of unknown normalizing constants, such as importance, bridge and path sampling, cf. Møller and Waagepetersen (2004).

As pointed out in Møller and Waagepetersen (2007), at least from a computational view point, Bayesian MCMC inference for Cox processes, and particularly shot-noise Cox processes, is often easier than maximum likelihood inference. For our spatio-temporal shot-noise Cox process, we would then need to impose a prior on the unknown parameter θ consisting of regression parameters and the parameters of the shot-noise residual process. The missing data is incorporated into the posterior distribution, which will then be the given by the conditional distribution of jointly θ and $\tilde{W}\times\tilde{T}$ given the data $\mathbf{x}$. We suggest to use a Metropolis-within-Gibbs algorithm where we alternate between updating $\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}}|(\theta,\mathbf{x})$ and $\theta|(\tilde{\mathbf{\Phi}}_{\tilde{W}\times\tilde{T}},\mathbf{x})$. The first type of update is again as described in Section 7.1.1, while we imagine that the second type of update would be a kind of Metropolis random walk update.

7.3 Summary statistics and residuals

Since our spatial-temporal shot-noise Cox process is inhomogeneous in both time and space, we have considered various inhomogeneous K-functions, cf. Section 4, and used these for model fitting and checking, cf. Section 6. Other kind of summary statistics, e.g. summary statistics based on inter-point distances such as F, G, and J-functions, have only been defined in the homogeneous case, and it seems not possible to extend them in a natural way to the inhomogeneous case, see Møller and Waagepetersen (2004) and the references therein. Alternatively, residuals may be used, as exemplified in Figure 10 and further discussed in Baddeley et al. (2005), Waagepetersen (2005), and Møller (2008).

Acknowledgements:

Supported by the Danish Natural Science Research Council, grant 272-06-0442, "Point process modelling and statistical inference".

References

[1] J.K. Agee. *Fire Ecology of Pacific Northwest Forests*. Island Press, Washington D.C., 1993.

[2] A. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350, 2000.

[3] A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

[4] A. Baddeley and R. Turner. Modelling spatial point patterns in R. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Case Studies in Spatial Point Process Modeling*, pages 23–74. Springer Lecture Notes in Statistics 185, Springer-Verlag, New York, 2006.

[5] A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of Royal Statistical Society Series B*, 67:617–666, 2005.

[6] W.C. Bessie and E.A. Johnson. The relative importance of fuels and weather on fire behavior in subalpine forest. *Ecology*, 76:747–762, 1995.

[7] D.R. Brillinger, H.K. Preisler, and J. W. Benoit. Probabilistic risk assessment for wildfires. *Environmetrics*, 17:623–633, 2006.

[8] A. Brix. Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31:929–953, 1999.

[9] A. Brix and J. Chadoeuf. Spatio-temporal modeling of weeds and shot-noise G Cox processes. *Biometrical Journal*, 44:83–99, 2002.

[10] A. Brix and P. J. Diggle. Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B*, 63:823–841, 2001.

[11] A. Brix and P. J. Diggle. Corrigendum: Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B*, 65:946, 2003.

[12] A. Brix and J. Møller. Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scandinavian Journal of Statistics*, 28:471–488, 2001.

[13] D. R. Cox. Some statistical models related with series of events. *Journal of the Royal Statistical Society Series B*, 17:129–164, 1955.

[14] P.W.A. Dayananda. Stochastic models for forest fires. *Ecological Modelling*, 3:309–313, 1977.

[15] S.C. De Long. Natural disturbance rate and patch size distribution of forest in Northern British Columbia: implications for forest management. *Northwest Science*, 72:35–48, 1998.

[16] C. Diaz-Avalos, D.L. Peterson, E. Alvarado, S.A. Ferguson, and J.E. Besag. Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canadian Journal of Forest Research*, 31:1579–1593, 2001.

[17] P. J. Diggle. A kernel method for smoothing point process data. *Applied Statistics*, 34:138–147, 1985.

[18] P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Arnold, London, second edition, 2003.

[19] P.J. Diggle. Spatio-temporal point processes: methods and applications. In B. Finkenstädt, L. Held, and V. Isham, editors, *Statistical Methods for Spatio-Temporal Systems*, pages 1–45. Chapman & Hall/CRC, Boca Raton, 2007.

[20] P.J. Diggle, A.G. Chetwynd, R. Häggkvist, and S. Morris. Second order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136, 1995.

[21] P.J. Diggle, L. Knorr-Held, B. Rowlingson, T. Su, P. Hawtin, and T. Bryant. Towards on-line spatial surveillance. In R. Brookmeyer and D. Stoup, editors, *Monitoring the Health of Populations: Statistical Methods for Public Health Surveillance*, pages 233–266. Oxford University Press, Oxford, 2003.

[22] P.J. Diggle, B. Rowlingson, and T. Su. Point process methodoly for on-line spatio-temporal disease surveillance. *Environmetrics*, 16:423–434, 2005.

[23] Y. Guan and M. Sherman. On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society Series B*, 69:31–49, 2007.

[24] L. Heinrich. Minimum contrast estimates for parameters of spatial ergodic point processes. In *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, pages 479–492, Prague, 1992. Academic Publishing House.

[25] G. Hellmund, M. Prokešová, and E.B.V. Jensen. Lévy based cox point processes. *Advances in Applied Probability*, 40, 2008. To appear.

[26] B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.

[27] D. McKenzie, D.L. Peterson, and J.K. Agee. Fire frequency in the interior Columbia River basin: Building regional models from fire history data. *Ecological Applications*, 10:1497–1516, 2000.

[28] J. Mecke. Stationäre zufällige Maße auf lokalkompakten Abelschen Gruppen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 9:36–58, 1967.

[29] J. Møller. Shot noise Cox processes. *Advances in Applied Probability*, 35:4–26, 2003.

[30] J. Møller. Inference. In W.S. Kendall and I. Molchanov, editors, *Stochastic Geometry: Highlights, Interactions and New Perspectives*. Clarendon-Press, Oxford, 2008. To appear.

[31] J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.

[32] J. Møller and G. L. Torrisi. Generalised shot noise Cox processes. *Advances in Applied Probability*, 37:48–74, 2005.

[33] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2004.

[34] J. Møller and R. P. Waagepetersen. Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics*, 34:643–711, 2007.

[35] R.D. Peng, F.P. Schoenberg, and J. Woods. A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of American Statistical Association*, 100:26–35, 2005.

[36] M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, London, 1981.

[37] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266, 1976.

[38] F. P. Schoenberg. Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128:79–93, 2004.

[39] B.W. Silverman. *Density Estimation*. Chapman and Hall, London, 1986.

[40] D. Stoyan and H. Stoyan. Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, 27:641–656, 2000.

[41] R. Waagepetersen. Discussion of the paper by Baddeley, Turner, Møller & Hazelton (2005). *Journal of the Royal Statistical Society Series B*, 67:662, 2005.

[42] R. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258, 2007.

[43] C.W. Wilkins. A stochastic analysis of the effect of fire on remote vegetation. Ph.D. thesis, University of Adelaide, 1977.

[44] R. L. Wolpert and K. Ickstadt. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85:251–267, 1998.