

Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates

M.I. Borrajo^{a,*}, W. González-Manteiga^a, M.D. Martínez-Miranda^b

^a*Department of Statistics, Mathematical Analysis and Optimisation, Universidade de Santiago de Compostela*

^b*Department of Statistics and Operations Research, Universidad de Granada*

Abstract

The bias-variance trade-off for inhomogeneous point processes with covariates is theoretically and empirically addressed. A consistent kernel estimator for the first-order intensity function based on covariates is constructed, which uses a convenient relationship between the intensity and the density of events location. The asymptotic bias and variance of the estimator are derived and hence the expression of its infeasible optimal bandwidth. Three data-driven bandwidth selectors are proposed to estimate the optimal bandwidth. One of them is based on a new smooth bootstrap proposal which is proved to be consistent under a Poisson assumption. The other two are a rule-of-thumb method based on assuming normality, and a simple non-model-based approach. An extensive simulation study is accomplished considering Poisson and non-Poisson scenarios, and including a comparison with other competitors. The practicality of the new proposals is shown through an application to real data about wildfires in Canada, using meteorological covariates.

Keywords: Spatial point processes, First-order intensity, Kernel estimation, Bandwidth selection, Covariates

2010 MSC: 62G10, 62G20, 60G55, 60-02

*Department of Statistics. Faculty of Mathematics. Campus Vida - E15782, Santiago de Compostela. Email: mariaisabel.borrajo@usc.es

The authors provide the data set and R code to replicate the real data analysis performed and this appears as annexes in the electronic version of the manuscript.

1. Introduction

Spatial point processes are mathematical models that describe the geometrical structure of patterns formed by events, which are distributed randomly in number and space. In the last decades we have seen an explosion in the literature devoted to point processes. New methodology has been developed based on a profound theoretical work and mainly driven by applications. Point patterns are observed in many different fields such as ecology, Illian et al. (2009) and Law et al. (2009); epidemiology, Diggle (1990) and Gatrell et al. (1996); astronomy, Babu and Feigelson (1996) and Kerscher (2000); forestry, Stoyan and Penttinen (2000); seismology, Ogata (1988), Ogata (1998), Ogata and Zhuang (2006) and Schoenberg (2011). An overview of general theory on point processes and some classical applications can be found in Daley and Vere-Jones (1988), Moller and Waagepetersen (2003), Illian et al. (2008), Diggle (2013) and Baddeley et al. (2015).

Understanding the first-order intensity function is one of the main goals in point process theory. The intensity uniquely determines the expected number of events in any measurable subset of its domain. Hence accurate estimation of the intensity is crucial to describe the mean behaviour of a point process. Parametric estimation of the intensity has been considered using for instance a likelihood score, such as the Akaike information criteria, see Van Lieshout (2000), Moller and Waagepetersen (2003) and Diggle (2013), or pseudolikelihood procedures, see for example Waagepetersen (2007). In the Bayesian framework, Yue and Loh (2007) proposed a semiparametric intensity estimator and Illian et al. (2012) considered intensity modelling based on log-Gaussian Cox processes. Parametric estimation might be simpler and it is appropriate when the context suggests a specific form, see Schoenberg (2005), but in general, nonparametric estimation methods based on kernels or splines are called to better describe the intensity, allowing more flexibility.

Diggle (1985) proposed the first kernel intensity estimator, based on the common kernel density estimator of Parzen (1962) and Rosenblatt (1956). The main drawback of this estimator is its lack of consistency, which has almost limited its use to exploratory analysis. To overcome this problem, Cucala (2006) developed asymptotic theory for Diggle's estimator, introducing the concept of "density of events locations", and using that the intensity and the density functions differ only by a constant.

A key challenge in kernel estimation is the choice of the bandwidth parameter. Small bandwidths give very detailed estimates and hence small

bias and large variance, while large bandwidths reduce the variance at the expense of increasing the bias. Hence an appropriate choice of the bandwidth, controlling the bias-variance trade-off, is required to use any kernel estimator in practice. There has been great theoretical developments on this problem in areas of statistics such as regression, density and hazard estimation (see for example Marron (1988), Scott (1992) and Silverman (1986) for an earlier full description of the problem), meanwhile in the context of point processes it has received less attention. Diggle (1985) proposed a bandwidth selector based on the minimisation of the mean squared error of his estimator. Later, Diggle and Marron (1988) showed the equivalence, for one-dimensional Cox processes, between Diggle (1985)'s bandwidth and the standard least-squares cross-validation method used in kernel density estimation. This paper also described and used the convenient relationship between density and intensity estimation mentioned above. Brooks and Marron (1991) proved the optimality of the least-squares cross-validation bandwidth for one-dimensional inhomogeneous Poisson point processes. Later Fuentes-Santos et al. (2015) developed an extension of Cucala's theory to the two-dimensional case, and proposed a plug-in bandwidth selection method based on bootstrapping. Recently Cronie and van Lieshout (2018) have addressed the bandwidth selection problem for Diggle (1985)'s kernel estimator suggesting a simple bandwidth estimate which does not use any model assumptions about the underlying process. Unfortunately this bandwidth choice is not designed to optimally balance the bias-variance trade-off of the intensity kernel estimator and then one cannot expect accurate intensity estimates in practice.

Kernel estimation of the intensity has been also consistently performed including additional information. Point processes may have extra information in two ways: marks and covariates, see (Illian et al., 2008, Chap. 5). The main difference between them is that marks are necessarily linked to the events, while covariates provide information about the whole observation region. Guan (2008) developed a kernel intensity estimator assuming that the intensity depends on some observed continuous covariates through an unknown continuous function. Baddeley et al. (2012) proposed a slightly different estimator based on the same idea. Regarding the bandwidth parameters, Guan (2008) suggested a simple but computationally intense cross-validation method to choose the bandwidth for his estimator. And Baddeley et al. (2012) did not specifically address the bandwidth selection problem for his estimator but used in practice the common rule-of-thumb borrowed from

density estimation (see Silverman (1986)).

In this paper we revisit theoretically and empirically the traditional bias-variance trade-off for the case of inhomogeneous point process with covariates. First we construct a theoretical framework to guarantee the consistency of a new kernel intensity estimator based on covariates. The estimator is defined by using the relationship between density and intensity functions in a similar way to Cucala (2006). Second we address the problem of bandwidth selection for the new kernel estimator. We suggest three new data-driven bandwidth selectors: a simple rule-of-thumb based on assuming normality and inspired on Silverman (1986)'s; a bootstrap estimator derived from a consistent resampling procedure; and an extension of the bandwidth selector proposed by Cronie and van Lieshout (2018) to the case of intensity depending on covariates. For simplicity we define the resampling procedure assuming that the underlying process is Poisson but other approaches are possible if specific information is available. Similarly we prove the consistency of both the kernel intensity estimator and the resampling procedure in the Poisson case. This Poisson assumption nevertheless seems not to be restrictive in practice as our empirical studies will show.

The paper accomplishes both theoretical and practical aspects but to make it easier to follow we have deferred most of the technical details to the appendices. Our theoretical developments are to the best of our knowledge new in the field. Although several methods already exist in the context of this paper, the theory is sparse in many ways. Particularly no theoretical bias-variance expressions, and hence no ad-hoc bandwidth selectors, have been previously developed for point processes with covariates. The practicality of our proposals is supported by simulations and a data analysis. Our simulations show a good performance of our proposals, which are competitive with other existing procedures. Simulations also show that our proposals perform well even when simulating from non-Poisson processes such as log-gaussian Cox and Neymann-Scott processes, and considerably better than the currently existing methods in these contexts.

The theoretical framework we have set up in the paper is defined for one single one-dimensional covariate. In practice it is of interest to consider several covariates. The relationship between the intensity and the density function can be also considered to define a multivariate extension of our intensity estimator. Moreover this link opens a promising way of solving problems of practical interest in spatial statistics, just by bringing into spatial statistics what has been done for decades in the density context (see for

example Chacón and Duong (2018), Scott (2015), as well as Klemelä (2009), for a recent revision on this topic). This paper fully describes this link in the one-dimensional case and details briefly the multivariate extension.

The rest of the paper is organised as follows. In Section 2, we briefly review kernel intensity estimation for point process with and without covariates. Section 3 is devoted to set up the theoretical framework for kernel intensity estimation with covariates and to develop asymptotic theory. In Section 4 a new resampling procedure is proposed based on smooth bootstrap and its consistency is shown. Section 5 describes the three new bandwidth selection methods. An extensive simulation study is carried out in Section 6 to analyse the finite-sample properties of the new proposals for Poisson and non-Poisson point processes, including a comparison with other existing competitors. Section 7 provides an illustration of the new proposals with a real data set of wildfires in Canada. Appealing spatio-temporal and multivariate extensions are described in Section 8, and finally some conclusions are drawn in Section 9. All computations have been performed in R. The wildfire data set and the R code with original functions to reproduce the application are provided as supplementary material.

2. Kernel intensity estimation

Let X be a point process defined in a region $W \subset \mathbb{R}^2$ with finite positive area. Let X_1, \dots, X_N be a realisation of the process with N being the random variable counting the number of events. The first-order intensity, from now on referred as intensity, is defined as:

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \frac{E[N(dx)]}{|dx|},$$

where $|dx|$ denotes the area of an infinitesimal region containing the point $x \in \mathbb{R}^2$ (see Diggle (2013)).

Diggle (1985) proposed a kernel intensity estimator for one-dimensional point processes, which has been extended to the plane as:

$$\hat{\lambda}_H^D(x) = \frac{\sum_{i=1}^N \mathbf{K}_H(x - X_i)}{p_H(x)}, \quad x \in W \subset \mathbb{R}^2.$$

Here H is a matrix of bandwidth parameters, $\mathbf{K}_H(x) = |H|^{-1/2} \mathbf{K}(H^{-1/2}x)$, where \mathbf{K} is a two-dimensional kernel function, and $p_H(x) = \int_W \mathbf{K}_H(x - y) dy$ is an edge correction term.

This kernel estimator has been widely used during decades and mostly limited to exploratory analysis due to its lack of consistency. To overcome this problem, Cucala (2006) introduced the concept of “density of events locations” for one-dimensional point processes. He defined such density as $\lambda_0(x) = \lambda(x)/m$, where $m = \int_W \lambda(x)dx$ is the expected number of events lying on W . And he proposed the following kernel estimator for λ_0 :

$$\hat{\lambda}_{0,h}(x) = \frac{1}{N} \sum_{i=1}^N K_h(x - X_i) 1_{\{N \neq 0\}}, x \in \mathbb{R},$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, with K being a one-dimensional kernel function and h a scalar bandwidth parameter. Here 1_{Ω} denotes the indicator function. Cucala (2006) proved the consistency of his estimator under an infill structure asymptotic framework. Recently Fuentes-Santos et al. (2015) have extended Cucala’s estimator to the two-dimensional case using a two-dimensional kernel density estimator that involves a matrix of bandwidths.

Moving on to the framework of point processes with covariates, let $Z : W \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ be a spatial continuous covariate that is exactly known in every point of the region of interest W , and let Z_1, \dots, Z_N be the realisation of the transformed process, i.e, $Z_i = Z(X_i)$. In practice this covariate will commonly be known in an enough amount of points spread over the region, so the values for the rest of the points can be interpolated and it can be assumed that these values are indeed the real ones (see Baddeley et al. (2012)).

Following Baddeley et al. (2012) let us assume that the intensity can be described from the known covariate through the model:

$$\lambda(u) = \rho(Z(u)), u \in W \subset \mathbb{R}^2, \tag{1}$$

where ρ is an unknown function. As Z is known, the only target for intensity estimation is the function ρ , for which Baddeley et al. (2012) proposed, among others, the following kernel estimator:

$$\hat{\rho}_W(z) = \sum_{i=1}^N \frac{1}{g^*(Z_i)} K_h(z - Z_i),$$

where g^* is the derivative of the spatial cumulative distribution function defined in the next section. To estimate the bandwidth parameter h in

practice the authors suggested the common Silverman's rule-of-thumb, see Silverman (1986).

Previously Guan (2008) proposed a closely-related kernel intensity estimator for the case of a multidimensional covariate, $\mathbf{Z} = (Z_1, \dots, Z_p) : W \subset \mathbb{R}^2 \rightarrow \mathbb{R}^p$:

$$\hat{\lambda}_h^G(u) = \frac{\sum_{i=1}^N K_h(\|\mathbf{Z}(u) - \mathbf{Z}(X_i)\|)}{q_h(u)}.$$

Here $q_h(u) = \int_W K_h(\|\mathbf{Z}(u) - \mathbf{Z}(s)\|) ds$ is the edge correction term. Considering the increasing domain asymptotic framework and adding also some suitable assumptions, the author proved the consistency of the estimator. In practice he suggested a simple, but computationally-intensive, cross-validation method to estimate the bandwidth h .

3. A consistent kernel intensity estimator based on covariates

In this section we define our kernel intensity estimator based on a one-dimensional covariate Z and show its consistency. In our derivations we work under the transformed space assuming model (1), and with the point process obtained from the original one, X , through the covariate, $Z(X)$ (more details about it are given in Appendix A).

We need to introduce some definitions and additional notation. The spatial cumulative distribution function of Z is defined as

$$G(z) = \frac{1}{|W|} \int_W 1_{\{Z(u) \leq z\}} du,$$

where $|W|$ denotes the area of the region $W \subset \mathbb{R}^2$. Let assume that G has a first derivative g , for which we need Z to be differentiable with non-zero gradient, and let denote by $g^*(\cdot) = |W|g(\cdot)$ and $G^*(\cdot) = |W|G(\cdot)$ the unnormalised versions. The results detailed in Appendix A show that $Z(X)$ is indeed a point process with intensity ρg^* .

To derive our consistent kernel estimator we follow Cucala (2006) and use the relationship between the intensity and the density function. We define the density function for this problem as the relative density of the transformed point process $Z(X)$:

$$f(z) = \frac{\rho(z)g^*(z)}{m}. \tag{2}$$

Our idea is to construct a kernel estimate of f and then plug-in it in expression (2), jointly with an appropriate estimate of m , and thus to derive an

estimator of ρ . This gives the estimator of the intensity λ through equation (1).

Following the pre-established notation, we define our estimator of the relative density f :

$$\hat{f}_h(z) = g^*(z) \frac{1}{N} \sum_{i=1}^N \frac{1}{g^*(Z_i)} K_h(z - Z_i) 1_{\{N \neq 0\}}. \quad (3)$$

Now we need to estimate m that is the expected number of events. For simplicity we suggest the sample size N as an estimator. From these estimators and using equations (2) and (1) we derive our kernel intensity estimator as:

$$\hat{\lambda}_h(u) = \hat{\rho}_h(Z(u)) = \frac{N \hat{f}_h(Z(u))}{g^*(Z(u))} = \sum_{i=1}^N \frac{1}{g^*(Z_i)} K_h(Z(u) - Z_i). \quad (4)$$

Remark that, for the particular estimates of the relative density and m we propose, our final intensity estimator shares the same expression as Baddeley et al. (2012)'s estimator detailed in the previous section. However our proposal benefits for being conveniently constructed to guarantee the consistency, to facilitate the theoretical developments and to allow consistent bootstrap methods. This construction also allows intuitive multivariate extensions, including the time dimension, as we discuss later in Section 8.

In the analysis of spatial point patterns two asymptotic frameworks can be assumed: the increasing-domain and the infill-structure. The former, proposed by Cressie (1993) and used also in Guan (2008), dictates that the expected number of events tends to infinity with the size of the observation region. Under this framework, which increases the number of events keeping the distance between them, we only add information in the boundary, instead of elsewhere, and when the bandwidth h tends to zero, the estimated intensity at each point depends on an expected number of events tending to zero. The infill-structure asymptotics, proposed in Diggle and Marron (1988), assume that the observation region remains fixed while the sample size increases. In this case the bandwidth is considered as a function of the expected sample size, this is, $h \equiv h(m)$, and hence as a sequence of values when $m \rightarrow \infty$, which is closer to the classical density framework. We assume here the infill-structure asymptotic framework.

The asymptotic properties of our intensity estimator are then driven by those of the relative density estimator, \hat{f}_h , given in (3). Next we obtain its

mean and variance, and hence the asymptotic expression of mean squared error, $MSE(h, z) = E[\{\hat{f}_h(z) - f(z)\}^2]$. All proofs are provided in the appendices. To simplify the theoretical calculations we assume that the point process X is inhomogeneous Poisson in $W \subset \mathbb{R}^2$, so that we can obtain closed-form expressions. Additionally we introduce some regularity conditions:

$$(A.1) \quad \int_{\mathbb{R}} K(z)dz = 1; \int_{\mathbb{R}} zK(z)dz = 0 \text{ and } \mu_2(K) := \int_{\mathbb{R}} z^2K(z)dz < \infty.$$

$$(A.2) \quad \lim_{m \rightarrow \infty} h = 0 \text{ and } \lim_{m \rightarrow \infty} \frac{A(m)}{h} = 0, \text{ where } A(m) := E \left[\frac{1}{N} 1_{\{N \neq 0\}} \right].$$

(A.3) G is three times differentiable.

(A.4) z is a continuity point of ρ .

(A.4') ρ is three times differentiable.

Theorem 3.1. *Under conditions (A.1) to (A.4) we have that:*

$$\begin{aligned} E \left[\hat{f}_h(z) \right] &= \frac{g^*(z)(K_h \circ \rho)(z)}{m} (1 - e^{-m}), \\ \text{Var} \left[\hat{f}_h(z) \right] &= A(m) \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) - \\ &\quad - (A(m) + e^{-2m} - e^{-m})(g^*(z))^2 (K_h \circ \rho)^2(z), \end{aligned}$$

where \circ denotes the convolution between two functions. Moreover, adding condition (A.4') we have:

$$\begin{aligned} MSE(h, z) &= e^{-2m} f^2(z) + (1 - e^{-m})^2 \frac{h^4}{4} \left(\frac{\rho''(z)g^*(z)}{m} \right)^2 \mu_2^2(K) \\ &\quad - e^{-m}(1 - e^{-m})h^2 \mu_2(K) \frac{(g^*(z))^2 \rho(z) \rho''(z)}{m^2} + \frac{A(m)}{h} f(z) R(K) \\ &\quad + o(h^2(1 - e^{-m})e^{-m}) + o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right), \end{aligned}$$

where $R(K) = \int_{\mathbb{R}} K^2(z)dz$.

The mean integrated square error is defined as:

$$MISE(h) = E \int \left\{ \hat{f}_h(z) - f(z) \right\}^2 dz, \quad (5)$$

and its asymptotic expression is denoted by $AMISE$. As a consequence of Theorem 3.1, the following result provides the asymptotic expression of the optimal bandwidth defined as the minimiser of the $AMISE$.

Corollary 3.2. *Under conditions (A.1) to (A.3) and (A.4'),*

$$\begin{aligned} MISE(h) &= e^{-2m}R(f) + (1 - e^{-m})^2 \frac{h^4}{4} R\left(\frac{\rho'' g^*}{m}\right) \mu_2^2(K) \\ &\quad - e^{-m}(1 - e^{-m})h^2 \mu_2(K) \int_{\mathbb{R}} \frac{g^*(z)\rho''(z)f(z)}{m} dz + \frac{A(m)}{h} R(K) \\ &\quad + o(h^2(1 - e^{-m})e^{-m}) + o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right), \\ AMISE(h) &= (1 - e^{-m})^2 \frac{h^4}{4} R\left(\frac{\rho'' g^*}{m}\right) \mu_2^2(K) + \frac{A(m)}{h} R(K). \end{aligned}$$

As a consequence the optimal bandwidth value which minimises $AMISE$ is:

$$h_{AMISE} = \left(\frac{A(m)R(K)}{\mu_2^2(K)(1 - e^{-m})^2 R\left(\frac{\rho'' g^*}{m}\right)} \right)^{1/5}. \quad (6)$$

4. A consistent bootstrap method under the Poisson assumption

Here we propose a new bootstrap procedure for point processes with covariates assuming that the underlying process is inhomogeneous Poisson. It consists of an extension of the smooth bootstrap method of Cao (1993), defined for kernel density estimation, and it is based on the resampling algorithms described in Cowling et al. (1996) for Poisson intensities.

We use the bootstrap procedure to estimate the $MISE$ of the density estimator given in (3), and later to define a bootstrap bandwidth selector for our kernel intensity estimator given in (4). This is only possible because of the Poisson assumption, which allows us to get closed-form expressions of the error measures and of the optimal bandwidth value that we can then use to propose a bandwidth selector. Other bootstrap procedures have been used in spatial processes such as the block bootstrap. It has the advantage of not requiring the Poisson assumption. However such procedure does not lead to a closed-form expression for the bandwidth selector and an intensive resampling algorithm has to be used in practice.

Recall that X_1, \dots, X_N is a realisation of the spatial point process X , and Z_1, \dots, Z_N is the corresponding realisation of the transformed process. From these we construct bootstrapped Z_1^*, \dots, Z_n^* as follows:

1. Conditional on Z_1, \dots, Z_N , we define $N^* \sim Poiss(\hat{m})$, with $\hat{m} := \int_{\mathbb{R}} \hat{\rho}_b(z) g^*(z) dz$.
2. We generate one value n^* from N^* .
3. We draw $Z_1^*, \dots, Z_{n^*}^*$ by sampling randomly with replacement from the distribution with density $\frac{\hat{\rho}_b g^*}{\hat{m}}$, where b is a pilot bandwidth whose practical choice is detailed in Section 5.2.

Now, from the bootstrapped sample $Z_1^*, \dots, Z_{n^*}^*$ and N^* , we define the bootstrap estimator of the density f as:

$$\hat{f}_h^*(z) = g^*(z) \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{1}{g^*(Z_i^*)} K_h(z - Z_i^*) 1_{\{N^* \neq 0\}}, \quad (7)$$

and hence the corresponding bootstrap estimator of λ :

$$\hat{\lambda}_h^*(u) = \sum_{i=1}^{N^*} \frac{1}{g^*(Z_i^*)} K_h(Z(u) - Z_i^*).$$

The following result provides the expression of the mean, the variance and MSE of $\hat{f}_h^*(z)$ under the bootstrap distribution. Hereafter E^* , Var^* and MSE^* denote the mean, the variance and the MSE respectively, under the bootstrap distribution.

Theorem 4.1. *Under conditions (A.1) to (A.4),*

$$\begin{aligned} E^* \left[\hat{f}_h^*(z) \right] &= \frac{g^*(z)}{\hat{m}} (K_h \circ \hat{\rho}_b)(z) (1 - e^{-\hat{m}}), \\ Var^* \left[\hat{f}_h^*(z) \right] &= \frac{(g^*(z))^2}{\hat{m}} \left(K_h^2 \circ \frac{\hat{\rho}_b}{g^*} \right) (z) A(\hat{m}) - \frac{(g^*(z))^2}{\hat{m}^2} (K_h \circ \hat{\rho}_b)^2(z) (A(\hat{m}) \\ &\quad + e^{-2\hat{m}} - e^{-\hat{m}}), \end{aligned}$$

where $A(\hat{m}) := E^* \left[\frac{1}{N^*} 1_{\{N^* \neq 0\}} \right]$. Moreover, adding condition (A.4'), we have

that:

$$\begin{aligned}
MSE^*(h, z) &= e^{-2\hat{m}} \left(\frac{\hat{\rho}_b(z)g^*(z)}{\hat{m}} \right)^2 + \frac{h^4}{4} (\hat{\rho}_h''(z))^2 \frac{(g^*(z))^2}{\hat{m}^2} \mu_2^2(K) (1 - e^{-\hat{m}})^2 \\
&- e^{-\hat{m}} (1 - e^{-\hat{m}}) h^2 \frac{\hat{\rho}_b(z)g^*(z)}{\hat{m}} \frac{\hat{\rho}_b''(z)g^*(z)}{\hat{m}} \mu_2(K) + \frac{A(\hat{m})}{h} R(K) \\
&+ o_P(h^4(1 - e^{-\hat{m}})^2) + o_P(h^2(1 - e^{-\hat{m}})e^{-\hat{m}}) + o_P\left(\frac{A(\hat{m})}{\hat{m}h}\right). \tag{8}
\end{aligned}$$

The next corollary provides expressions of the mean integrated squared error and its asymptotic approximation under the bootstrap distribution.

Corollary 4.2. *Under conditions (A.1) to (A.3) and (A.4'),*

$$\begin{aligned}
MISE^*(h) &= e^{-2\hat{m}} R\left(\frac{\hat{\rho}_b g^*}{\hat{m}}\right) + \frac{h^4}{4} R\left(\frac{\hat{\rho}_b'' g^*}{\hat{m}}\right) \mu_2^2(K) (1 - e^{-\hat{m}})^2 \\
&- e^{-\hat{m}} (1 - e^{-\hat{m}}) h^2 \mu_2(K) \int \frac{\hat{\rho}_b(z) \hat{\rho}_b''(z) (g^*)^2(z)}{\hat{m}^2} dz + \frac{A(\hat{m})}{h} R(K) \\
&+ o(h^4(1 - e^{-\hat{m}})^2) + o_P(h^2(1 - e^{-\hat{m}})e^{-\hat{m}}) + o_P\left(\frac{A(\hat{m})}{\hat{m}h}\right), \\
AMISE^*(h) &= \frac{h^4}{4} R\left(\frac{\hat{\rho}_b'' g^*}{\hat{m}}\right) \mu_2^2(K) (1 - e^{-\hat{m}})^2 + \frac{A(\hat{m})}{h} R(K).
\end{aligned}$$

As a consequence, the asymptotic expression of the optimal bootstrap bandwidth for \hat{f}_h^* is:

$$h_{AMISE^*} = \left(\frac{A(\hat{m})R(K)}{\mu_2^2(K)(1 - e^{-\hat{m}})^2 R\left(\frac{\hat{\rho}_b'' g^*}{\hat{m}}\right)} \right)^{1/5}, \tag{9}$$

which is a plug-in estimate of the MISE-optimal bandwidth given in (6).

All these results above lead to the following corollary.

Corollary 4.3. *Under assumptions (A.1) to (A.4') MISE* and AMISE* are consistent estimators of MISE and AMISE, respectively, and therefore h_{AMISE^*} is a consistent estimator of \hat{h}_{AMISE} .*

5. Data-driven bandwidth selection

We propose three new bandwidth selectors for our kernel intensity estimator. The two first ones are inspired by common selectors in density estimation. And the last one is an adaptation of the selector proposed by Cronie and van Lieshout (2018) to the context of point processes with covariates assumed in this paper.

5.1. A rule-of-thumb bandwidth selector

Following the original idea in Silverman (1986) for common kernel density estimation, we aim to derive a plug-in estimator of the infeasible optimal bandwidth given in (6), assuming normality. In our context we assume that the underlying relative density given in (2) is Normal with mean μ and variance σ^2 . Thus we can estimate the parameters μ and σ^2 from the data and derive estimators of the unknown values in (6).

Remark that in common kernel density estimation the unknown values are immediately derived from the normality assumption, however in our case it is slightly more complicated. There are three unknowns to be estimated in (6): the value $R\left(\frac{\rho''g^*}{m}\right)$, the expected number of points, m , and the expectation $A(m)$ defined in condition (A.2). We suggest to estimate m by the observed sample size n , and $A(m)$ by $1/n$. And finally, using the normality of the density f , which is related to ρ as $f(z) = \frac{\rho(z)g^*(z)}{m}$, we derive the expression:

$$\rho''(z) = m \left(\frac{f''(z)}{g^*(z)} - \frac{2f'(z)(g^*(z))'}{(g^*(z))^2} - \frac{f(z)(g^*(z))''}{(g^*(z))^2} + \frac{2f(z)((g^*(z))')^2}{(g^*(z))^3} \right),$$

and then compute $R\left(\frac{\rho''g^*}{m}\right)$ using numerical integration. Replacing all those estimates in (6) we have the rule-of-thumb bandwidth that we denote hereafter by \hat{h}_{RT} .

5.2. A bootstrap bandwidth selector

Our bootstrap bandwidth selector is given by the expression:

$$\left(\frac{A(\hat{m})R(K)}{\mu_2^2(K)(1 - e^{-\hat{m}})^2 R\left(\frac{\rho_b''g^*}{\hat{m}}\right)} \right)^{1/5},$$

derived in Corollary 4.2. This is a consistent estimator of the MISE-optimal bandwidth for the estimator of the relative density in (3), which defines our kernel intensity estimator.

Notice that, given a pilot bandwidth b , the above expression is completely known from the data, but numerical integration is required to compute the values \hat{m} , $A(\hat{m})$ and $R(\frac{\hat{\rho}_b'' g^*}{\hat{m}})$. This can be done using for instance the Simpson's rule.

Following Cao (1993) and Borrajo et al. (2017) we derive that the order of the pilot bandwidth b in our context is $m^{-1/7}$. Hence we propose a data-driven pilot bandwidth choice just by rescaling the rule-of-thumb bandwidth as:

$$\hat{b} = \frac{\hat{m}^{-1/5}}{\hat{m}^{-1/7}} \hat{h}_{RT}, \quad (10)$$

The bootstrap bandwidth selector with the data-driven pilot bandwidth (10) is denoted hereafter as \hat{h}_{Boot} .

5.3. A non-model-based bandwidth selector

The non-model-based bandwidth selector proposed by Cronie and van Lieshout (2018) relies on the fact that $\int_W \frac{1}{\lambda(x)} \lambda(x) dx = |W|$, which let them build a discrepancy measure between the inverse of the kernel intensity estimator by Diggle (1985) and the area of the observation region, $|W|$. Minimising this discrepancy measure they obtain a data-driven bandwidth selection procedure. Remark that this criterion does not aim to optimise the bias-variance trade-off of the kernel intensity estimator and therefore does not guarantee to lead to good intensity estimates. On the other hand such a simple procedure does not rely on any model assumption that makes it very appealing for applications.

Here we transfer Cronie and van Lieshout (2018)'s approach to the context of point processes with covariates under model (1). It involves replacing Diggle's estimator by ours, detailed in Section 3. Hence, our non-model-based bandwidth selector, denoted by \hat{h}_{NM} , is defined as:

$$\hat{h}_{NM} = \arg \min_{h>0} (T(h) - |W|),$$

where $T(h) = \sum_{i=1}^N \hat{\rho}_h(Z_i)^{-1}$ inside W , and $|W|$ elsewhere.

6. Finite sample study

Here we analyse the performance of our intensity estimator with the three new bandwidth selection proposals through simulations. We have considered two scenarios: first the case where the underlying process is Poisson and second log-Gaussian Cox and Neymann-Scott processes inspired in the models simulated by Waagepetersen (2007), Guan and Loh (2007), Guan and Shen (2010) and Mrkvicka et al. (2014).

6.1. The Poisson scenario

To start with, we analyse the performance of our proposals with models fulfilling the assumptions required in the theory, which include that the intensity can be expressed as a function of one single covariate in the form of (1), and the underlying process is inhomogeneous Poisson. We define three models with intensity function in the unit square given by:

$$\lambda(u) = \exp(\beta_0 + \beta_1 Z(u)), \quad u \in W = [0, 1] \times [0, 1],$$

where β_0 and β_1 are known parameters, and Z denotes a covariate. The models are constructed using two different covariates that are shown in Fig. 1. The first covariate on the left, Z_1 , is a realisation of a Gaussian random field, with zero mean and exponential covariance structure $C(r) = \sigma^2 \exp(-r/\gamma)$, with parameters $\sigma = 0.1$ and $\gamma = 0.1$. The second covariate on the centre, dR , is a rescaled version into the unit square of the “distance to letter R” defined in Baddeley et al. (2012).

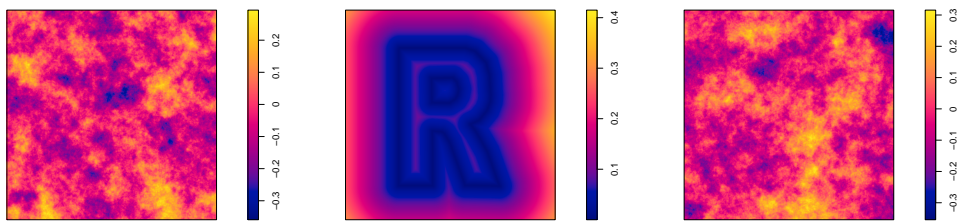


Figure 1: Plot of the covariates used in the simulated models, Z_1 (left), dR (centre) and the error term used to perturb the information given to the estimator by the real covariate (right).

Under the Poisson assumption, the three models defined are characterised

by the following three intensity functions:

$$\begin{aligned}\lambda_1(u) &= \exp(6 + 4Z1(u)), \\ \lambda_2(u) &= \exp(6 + 4(Z1 + \epsilon)(u)), \\ \lambda_3(u) &= \exp(5 - 3dR(u)).\end{aligned}\tag{11}$$

The second model includes an error term, ϵ , to perturb the covariate, in order to see how the estimators perform using only partial information on the real covariate generating the process. We have generated this error term considering a different realisation of the same Gaussian random field we have considered for $Z1$. A plot of this error term is shown in Fig. 1, and the three intensity functions in (11) are represented in Fig. 2.

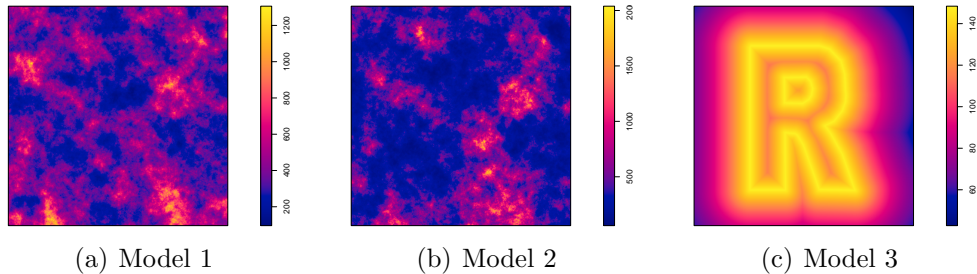


Figure 2: Intensity functions characterising the three simulated models.

We have simulated 500 realisations for each model and different expected sample sizes covering a wide range of possibilities, $m = 50, 100, 250$ and 500 . Remark that to guarantee that the mean number of events in the unit square is m , we have conveniently rescaled the intensity functions above.

Under these models we have evaluated the performance of our intensity estimator, given in (4), with the three new bandwidth proposals: the rule-of-thumb bandwidth, \hat{h}_{RT} , the bootstrap bandwidth, \hat{h}_{Boot} , and the non-model-based bandwidth, \hat{h}_{NM} . From each simulated sample we evaluate the performance of the resulting intensity estimates, $\hat{\lambda}$, through the relative integrated squared error:

$$ISE_{rel} = \int_W \left(\frac{\hat{\lambda}(u) - \lambda(u)}{\lambda(u)} \right)^2 du.\tag{12}$$

Besides, from the 500 simulated samples for each case, we compute three error criteria for each estimator. The two first are the average and the standard

deviation of the relative integrated errors, denoted respectively by e_1 and e_2 . The third one is defined as the relative bias of the bandwidth estimates. In this case we measure how close are our bandwidth estimates to the infeasible optimal bandwidth minimising the *MISE* in (5). We compute a Monte Carlo approximation of such optimal bandwidth, which we denote by \hat{h}_{MISE} , and consider it as a benchmark in our simulations. Thus, the relative bias of a bandwidth estimate \hat{h} is defined as $e_3 = \text{mean} \left((\hat{h} - \hat{h}_{\text{MISE}}) / \hat{h}_{\text{MISE}} \right)$.

In the comparison among our three proposals we include four competitors in the literature. On the one hand, under the covariate intensity model (1), we consider Baddeley et al. (2012)'s approach that uses the common Silverman's rule-of-thumb bandwidth for density estimation, \hat{h}_{Badd} ; and a least-squares cross-validated bandwidth defined in the same way as for common density estimation, \hat{h}_{CV} . On the other hand we consider estimation without covariates through Diggle's intensity estimator. In this case we compute two bandwidth estimates: the one presented in Diggle (1985), \tilde{h}_{Dig} , and the non-model-based bandwidth estimate originally proposed by Cronie and van Lieshout (2018), \tilde{h}_{CVL} .

The simulation results are shown in Tables 1, 2 and 3. Looking at the values of measure e_1 in these tables, we can see that the bootstrap bandwidth performs better than the others in most of the cases, and when this does not occur, it is still competitive. Our rule-of-thumb and Baddeley et al. (2012)'s proposal are not far away from it, even though the former shows a slightly better behaviour, specially for small sample sizes. The cross-validation method is only competitive in Model 2, where the covariate information provided to the estimator is perturbed with an error term. Regarding the non-model-based selector, in general it performs poorly, and specially the original approach for Diggle's estimator by Cronie and van Lieshout (2018). The results are even worse than those for the simple Diggle (1985)'s approach in Model 1 and 3. Remark that this non-model-based bandwidth estimate does not aim a good intensity estimate but a good estimate of the observation region area. Recall that Diggle's estimator does not use covariate information.

In terms of variability, measured by e_2 , our bootstrap bandwidth provides the best results specially for big samples. Our rule-of-thumb and the cross-validation method are also competitive, and all of them beat Diggle's estimates and our non-model-based bandwidth estimate.

We now look at the bias of bandwidth estimates through measure e_3 . For Model 1 the bootstrap bandwidth outperforms the others except for

the biggest sample size, where our non-model-based bandwidth gives in general the smaller values. However, as we saw in measure e_1 , it does not necessarily provide with a final good estimate for the intensity. The non-model-based bandwidth sometimes moves below the optimal (negative e_3), sometimes above (positive e_3) while showing quite often large relative ISE's (measure e_1). For Model 2 and Model 3, the non-model-based reaches the smaller values. Notice also that our rule-of-thumb bandwidth and Baddeley et al. (2012)'s approach generally show biases in the same direction (bandwidth smaller than the optimal one). Note that the cross-validation method is not designed to estimate the optimal bandwidth minimising the MISE but the integrated squared error (ISE), which might also explain the high values shown in this measure. Remark that for Diggle's estimator, and its two bandwidth choices, we do not compute the measure e_3 , the reason is that the optimal bandwidth minimising the MISE in (5) is not a benchmark in this case.

Table 1: Simulation results for Model 1 under a Poisson scenario. Our proposals \hat{h}_{RT} , \hat{h}_{Boot} and \hat{h}_{NM} , are compared with Baddeley et al. (2012)'s, \hat{h}_{Badd} , a cross-validation method, \hat{h}_{CV} , and two bandwidths for Diggle's estimator, \tilde{h}_{CVL} and \tilde{h}_{Dig} .

	\hat{h}_{MISE}	\hat{h}_{Badd}	\hat{h}_{RT}	\hat{h}_{Boot}	\hat{h}_{CV}	\hat{h}_{NM}	\tilde{h}_{CVL}	\tilde{h}_{Dig}
$m = 50$								
e_1	0.0662	0.1222	0.0881	0.0656	0.2688	0.1052	0.4262	0.3801
e_2	0.0735	0.1089	0.0848	0.0701	0.0699	0.1247	0.1799	0.1574
e_3	—	-0.6132	-0.4352	0.0533	6.2584	0.0666	—	—
$m = 100$								
e_1	0.0388	0.0706	0.0518	0.0391	0.2665	0.0721	0.3201	0.2709
e_2	0.0345	0.0494	0.0391	0.0342	0.0499	0.0701	0.1070	0.0902
e_3	—	-0.6238	-0.4577	0.0686	6.5941	0.0929	—	—
$m = 200$								
e_1	0.0248	0.0443	0.0332	0.0250	0.2731	0.0508	0.2433	0.2170
e_2	0.0220	0.0339	0.0268	0.0225	0.0347	0.0456	0.0644	0.0574
e_3	—	-0.6274	-0.4706	0.0859	8.1597	0.0875	—	—
$m = 500$								
e_1	0.0127	0.0231	0.0173	0.0131	0.2768	0.0334	0.1728	0.1858
e_2	0.0091	0.0159	0.0120	0.0093	0.0201	0.0298	0.0355	0.0342
e_3	—	-0.6254	-0.4690	0.1566	10.1974	0.0777	—	—

Table 2: Simulation results for Model 2 under a Poisson scenario. Our proposals \hat{h}_{RT} , \hat{h}_{Boot} and \hat{h}_{NM} , are compared with Baddeley et al. (2012)'s, \hat{h}_{Badd} , a cross-validation method, \hat{h}_{CV} , and two bandwidths for Diggle's estimator, \tilde{h}_{CVL} and \tilde{h}_{Dig} .

	\hat{h}_{MISE}	\hat{h}_{Badd}	\hat{h}_{RT}	\hat{h}_{Boot}	\hat{h}_{CV}	\hat{h}_{NM}	\tilde{h}_{CVL}	\tilde{h}_{Dig}
$m = 50$								
e_1	0.3070	0.3414	0.3115	0.3045	0.2708	0.3955	0.7010	0.6785
e_2	0.1451	0.1567	0.1436	0.1470	0.0501	0.2512	0.2957	0.2799
e_3	—	-0.4386	-0.2109	0.4183	0.2839	-0.0929	—	—
$m = 100$								
e_1	0.2574	0.2830	0.2708	0.2810	0.2660	0.3540	0.5273	0.5306
e_2	0.1270	0.1109	0.1061	0.1092	0.0329	0.1605	0.1698	0.1687
e_3	—	-0.9252	-0.8985	-0.8111	0.4187	-0.1031	—	—
$m = 200$								
e_1	0.2269	0.2514	0.2453	0.2617	0.2685	0.3280	0.4000	0.4569
e_2	0.0717	0.0727	0.0660	0.0665	0.0247	0.1022	0.0987	0.1095
e_3	—	-0.9334	-0.9099	-0.8290	0.5097	-0.1391	—	—
$m = 500$								
e_1	0.2065	0.2281	0.2264	0.2421	0.2710	0.3031	0.2860	0.4167
e_2	0.0313	0.0363	0.0355	0.0368	0.0157	0.0729	0.0484	0.0655
e_3	—	-0.9419	-0.9252	-0.8540	0.5688	-0.1992	—	—

Table 3: Simulation results for Model 3 under a Poisson scenario. Our proposals \hat{h}_{RT} , \hat{h}_{Boot} and \hat{h}_{NM} , are compared with Baddeley et al. (2012)'s, \hat{h}_{Badd} , a cross-validation method, \hat{h}_{CV} , and two bandwidths for Diggle's estimator, \tilde{h}_{CVL} and \tilde{h}_{Dig} .

	\hat{h}_{MISE}	\hat{h}_{Badd}	\hat{h}_{RT}	\hat{h}_{Boot}	\hat{h}_{CV}	\hat{h}_{NM}	\tilde{h}_{CVL}	\tilde{h}_{Dig}
$m = 50$								
e_1	0.0970	0.1173	0.1012	0.1001	0.1798	0.2060	0.2392	0.2055
e_2	0.0845	0.1013	0.0877	0.0855	0.0621	0.2313	0.1381	0.1034
e_3	—	-0.3810	-0.1910	0.2486	2.3346	-0.0313	—	—
$m = 100$								
e_1	0.0669	0.0735	0.0679	0.0704	0.1689	0.1652	0.1714	0.1139
e_2	0.0419	0.0507	0.0441	0.0396	0.0407	0.1461	0.0761	0.0530
e_3	—	-0.2854	-0.1001	0.3110	3.4686	-0.0192	—	—
$m = 200$								
e_1	0.0446	0.0460	0.0447	0.0487	0.1632	0.2632	0.1295	0.0669
e_2	0.0237	0.0251	0.0234	0.0227	0.0293	0.4099	0.0465	0.0271
e_3	—	-0.1730	0.0243	0.4022	4.8131	-0.0101	—	—
$m = 500$								
e_1	0.0244	0.0244	0.0251	0.0283	0.1587	0.2441	0.0927	0.0406
e_2	0.0094	0.0093	0.0088	0.0090	0.0183	0.1866	0.0292	0.0137
e_3	—	-0.0003	0.2245	0.5455	0.5229	-0.0039	—	—

To complete the analysis we compare our proposal with the intensity estimator by Guan (2008). For our intensity estimator we now consider only our bootstrap bandwidth, which has shown good results in all scenarios, and the MISE benchmark. For Guan’s intensity estimator we consider the cross-validation bandwidth suggested by the author, $\hat{h}_{CV_{Guan}}$, and as a benchmark, the (Monte Carlo approximated) optimal MISE bandwidth for the estimator, $\hat{h}_{MISE_{Guan}}$. Notice that the optimal bandwidths for our kernel intensity estimator and Guan’s estimator are then different. The simulation results are summarised in Table 4 and correspond to the same 500 simulated samples considered in previous tables. For simplicity we only report the measures e_1 and e_2 .

Table 4: Comparison with Guan (2008)’s estimator under a Poisson scenario.

	$\hat{h}_{MISE_{Guan}}$	$\hat{h}_{CV_{Guan}}$	\hat{h}_{MISE}	\hat{h}_{Boot}	$\hat{h}_{MISE_{Guan}}$	$\hat{h}_{CV_{Guan}}$	\hat{h}_{MISE}	\hat{h}_{Boot}
<i>Model 1</i>	$m = 50$				$m = 100$			
e_1	0.0815	0.1855	0.0662	0.0656	0.0481	0.1885	0.0388	0.0391
e_2	0.0786	0.1367	0.0735	0.0701	0.0412	0.1034	0.0345	0.0342
<i>Model 2</i>	$m = 50$				$m = 100$			
e_1	0.3316	0.4479	0.3070	0.3045	0.2777	0.4027	0.2574	0.2810
e_2	0.1804	0.2474	0.1451	0.1470	0.1092	0.1756	0.1270	0.1092
<i>Model 3</i>	$m = 50$				$m = 100$			
e_1	0.0518	0.0737	0.0970	0.1001	0.0309	0.0442	0.0669	0.0704
e_2	0.0530	0.0822	0.0845	0.0855	0.0305	0.0457	0.0419	0.0396
	$\hat{h}_{MISE_{Guan}}$	$\hat{h}_{CV_{Guan}}$	\hat{h}_{MISE}	\hat{h}_{Boot}	$\hat{h}_{MISE_{Guan}}$	$\hat{h}_{CV_{Guan}}$	\hat{h}_{MISE}	\hat{h}_{Boot}
<i>Model 1</i>	$m = 200$				$m = 500$			
e_1	0.0288	0.2358	0.0248	0.0250	0.0147	0.2407	0.0127	0.0131
e_2	0.0243	0.0892	0.0220	0.0225	0.0097	0.0640	0.0091	0.0093
<i>Model 2</i>	$m = 200$				$m = 500$			
e_1	0.2482	0.3804	0.2269	0.2617	0.2283	0.3593	0.2065	0.2421
e_2	0.0707	0.1303	0.0717	0.0665	0.0393	0.0825	0.0313	0.0368
<i>Model 3</i>	$m = 200$				$m = 500$			
e_1	0.0181	0.0356	0.0446	0.0487	0.0092	0.0227	0.0244	0.0238
e_2	0.0152	0.0274	0.0237	0.0227	0.0064	0.0155	0.0094	0.0090

First we have to point out the computational burden of performing cross-validation for Guan’s estimator, along with the numerical integration that this estimator requires. It seems to be a drawback of this approach that is well known for cross-validation methods. Another issue that we have observed about Guan’s cross-validation proposal is the non existence of a global minimum for the cross-validation score in a number of samples. In our simulations this issue has led to the selection of the bandwidth at the boundary

of the minimisation interval, which has occurred between 45 and 213 times out of 500, depending on the model and the sample sizes (it is more likely for smaller sizes). Ignoring these issues of the practical proposal by Guan, we can see from the numbers reported in Table 4 that our proposal performs considerably better for the two first models. In the best situation for the intensity estimators, that is using the infeasible benchmarks, \hat{h}_{MISE} and $\hat{h}_{\text{MISE}_{\text{Guan}}}$, our estimator achieves smaller relative errors in average with slightly lower variability. Considering the data-driven bandwidth choices for each estimator our bootstrap approach clearly beats the cross-validation method. On the other hand, Model 3 seems to be a good scenario for Guan’s estimator and his cross-validation method, but even in this case our bootstrap proposal is still competitive.

6.2. Non-Poisson scenarios

Our bootstrap bandwidth selector for our kernel intensity estimator has been developed assuming that the underlying process is Poisson. We now check whether this assumption is crucial for the performance of the estimator by simulating two non-Poisson models: a log-Gaussian Cox process and a Neymann-Scott process.

The log-Gaussian scenario is defined by the intensity model, $\log \lambda(x) = \beta_0 + \beta_1 Z(x) + G(x)$, where Z is any of the covariates in Fig. 1, and G is again a zero-mean Gaussian random field with exponential covariance model given by $C(r) = \sigma^2 \exp(-r/\gamma)$, where $\sigma = 0.1$ and $\gamma = 0.1$. We have analysed the effect of different covariates ($Z1$, $Z1 + \varepsilon$, dR) and γ values ($\gamma = 0.1$ and $\gamma = 0.2$) leading to similar results, so we present in Table 5 the values obtained for $Z = Z1$ and $\gamma = 0.1$.

The Neymann-Scott process is simulated as follows: we first simulate the parent generation as inhomogeneous Poisson point processes with intensities defined by $\kappa \frac{\lambda}{\int \lambda}$, where λ can be any of the functions in (11). Then, conditioned on each of those points, the number of “offsprings” associated with each parent is also inhomogeneous Poisson and their locations are driven again by the covariate-dependent intensities previously mentioned. The parameter $\kappa = 20, 50, 100, 200$, indicates the expected number of parents to be simulated, and then for a fixed m it has an impact on the rate of “offsprings” per parent (and then on the cluster shape), i.e., if $m = 100$ and $\kappa = 20$, the mean of the Poisson process driven the parents is 20, and to achieve a final expected number of 100, the expected number of “offsprings” per parent is 5. Again, the results obtained for different covariates and κ ’s values lead to

similar conclusions, so here we only include the results for $Z1$ and $\kappa = 50$ in Table 6.

From Tables 5 and 6 we can see the good performance of our kernel intensity estimator with the bootstrap bandwidth, even though the Poisson assumption is not fulfilled. This shows not only the theoretical interest of our results but also their practical value. Focusing on the results for the log-Gaussian case in Table 5, the bootstrap outperforms the other approaches in both e_1 and e_2 error criteria, except for the smaller sample size where our non-model approach, \hat{h}_{NM} , seems to have the lowest variance. Regarding the bias of the bandwidth estimates, measured by e_3 , the bootstrap is again performing better than the others. The bootstrap bandwidth tends to be smaller than the optimal one as it is shown by the negative bias. The same happens to the other bandwidth estimates with the exception of the non-model-based one which tends to oversmooth.

Table 5: Simulation results for log-Gaussian Cox processes with covariate $Z = Z1$ and $\gamma = 0.1$.

	\hat{h}_{MISE}	\hat{h}_{Badd}	\hat{h}_{RT}	\hat{h}_{Boot}	\hat{h}_{NM}	\tilde{h}_{CvL}	\tilde{h}_{Dig}
$m = 50$							
e_1	0.1120	0.1727	0.1355	0.1126	0.1239	0.4460	0.4061
e_2	0.2700	0.3286	0.2778	0.2685	0.1563	0.1891	0.1730
e_3	—	-0.1143	-0.0976	-0.0475	0.0528	—	—
$m = 100$							
e_1	0.0566	0.0889	0.0662	0.0571	0.0840	0.3320	0.2843
e_2	0.0671	0.0910	0.0683	0.0696	0.0763	0.1105	0.0985
e_3	—	-0.0849	-0.0676	-0.0186	0.0883	—	—
$m = 200$							
e_1	0.0395	0.0645	0.0519	0.0400	0.0641	0.2563	0.2307
e_2	0.0380	0.0557	0.0464	0.0385	0.0504	0.0644	0.0602
e_3	—	-0.0686	-0.0588	-0.0157	0.0866	—	—
$m = 500$							
e_1	0.0234	0.0346	0.0267	0.0241	0.0430	0.1779	0.1975
e_2	0.0102	0.0172	0.0122	0.0106	0.0310	0.0339	0.0345
e_3	—	-0.0464	-0.0315	0.0114	0.0750	—	—

For the Neymann-Scott processes we can see again in Table 6 a good performance of the bootstrap selector, as well as our rule-of-thumb bandwidth, both outperform the non-model-based approach. The non-model-based bandwidth of Cronie and van Lieshout (2018) for Diggle’s estimator

behaves poorly. Remark that this scenario corresponds to a clustered process where the clusters are driven by the covariate, which in this particular case is a realisation of a Gaussian random field.

From our simulations above we have shown that our bootstrap proposal seems to work well under the two considered non-Poisson processes. Further investigations are necessary to generalise our results to other more complicated scenarios.

Table 6: Simulation results for Neymann-Scott processes with covariate $Z1$ and $\kappa = 50$.

	\hat{h}_{MISE}	\hat{h}_{Badd}	\hat{h}_{RT}	\hat{h}_{Boot}	\hat{h}_{NM}	\tilde{h}_{CvL}	\tilde{h}_{Dig}
$m = 50$							
e_1	0.1023	0.1550	0.1224	0.1029	0.1303	0.4472	0.4046
e_2	0.2632	0.2776	0.2395	0.2484	0.1990	0.2333	0.2126
e_3	—	-0.1111	-0.0941	-0.0448	0.0624	—	—
$m = 100$							
e_1	0.0635	0.0955	0.0737	0.0639	0.0931	0.3343	0.2862
e_2	0.1003	0.1064	0.0952	0.0982	0.1074	0.1480	0.1398
e_3	—	-0.0975	-0.0803	-0.0307	0.0846	—	—
$m = 200$							
e_1	0.0477	0.0683	0.0570	0.0482	0.0736	0.2721	0.2474
e_2	0.0510	0.0573	0.0498	0.0511	0.0709	0.1217	0.1303
e_3	—	-0.0881	-0.0779	-0.0334	0.0692	—	—
$m = 500$							
e_1	0.0335	0.0435	0.0358	0.0340	0.0529	0.1986	0.2137
e_2	0.0354	0.0330	0.0313	0.0351	0.0518	0.0747	0.0995
e_3	—	-0.0620	-0.0471	-0.0039	0.0563	—	—

7. Canadian wildfires

Forest fires are one of the most important natural disturbances since the last Ice Age and they represent a huge social and economic problem. Canada has quite a long tradition on recording information about their wildfires; and also studies from many different perspectives have been carried out: Flannigan and Harrington (1988), Di Iorio et al. (2013), Rogers et al. (2013), Walter et al. (2014). It is quite well known that fire activity in Canada mostly relies on meteorological elements such as high temperatures and long periods without rain.

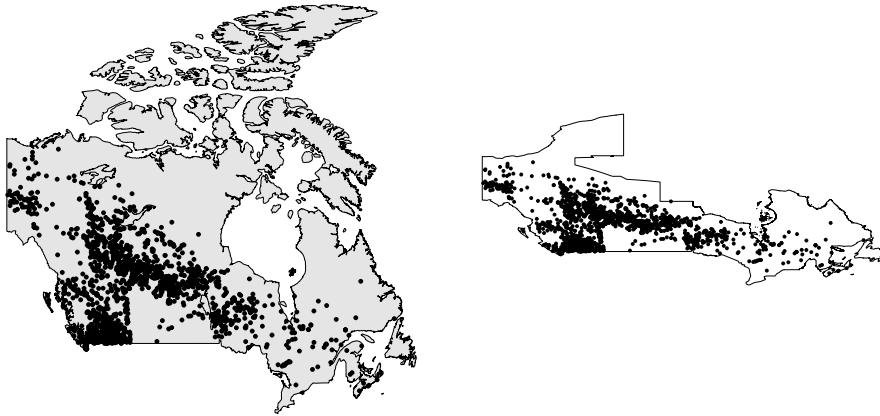
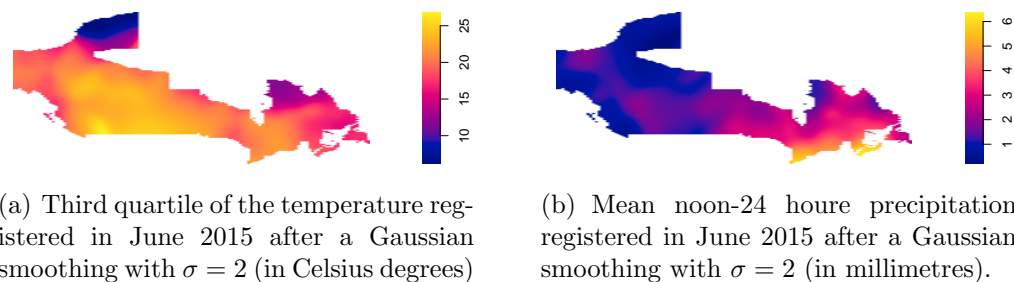


Figure 3: Wildfires in Canada during June 2015 over the whole country (left), and only on the reprojected observation region (right).

We are interested in studying the spatial influence on some of these meteorological variables on the distribution of wildfires. The wildfire data set and complete meteorological information from the last decades are available at the Canadian Wildland Fire Information System website (<http://cwfis.cfs.nrcan.gc.ca/home>). In the analysis we have removed two regions, Northwest Territories and Nunavut, as those are ice covered land and no forest grows. The fire season in Canada lasts from late April until August, with a peak of activity in June and July, hence we are interested in analysing wildfires during June 2015 (see Fig. 3), and we focus our attention on two covariates, temperature and precipitation (see Fig. 4).



(a) Third quartile of the temperature registered in June 2015 after a Gaussian smoothing with $\sigma = 2$ (in Celsius degrees)

(b) Mean noon-24 hour precipitation registered in June 2015 after a Gaussian smoothing with $\sigma = 2$ (in millimetres).

Figure 4: Covariates to be used in the intensity estimation of the wildfires in Canada during June 2015.

Under our intensity model in (1) we look for a single one-dimensional covariate driving the intensity of the wildfire events. As the temperature

seems to have a high influence on the distribution of the wildfires, we perform the estimation including it as the explanatory covariate. Remark that instead of considering the maximum value during June in every point of the region, we have computed the third quartile to avoid dealing with extreme values. In this real scenario we compare not only the results obtained using the different bandwidth selectors described in the previous sections but also, in order to detect the influence of the covariate, we have included Diggle’s kernel intensity estimator that uses only point pattern coordinates with two bandwidth selectors.

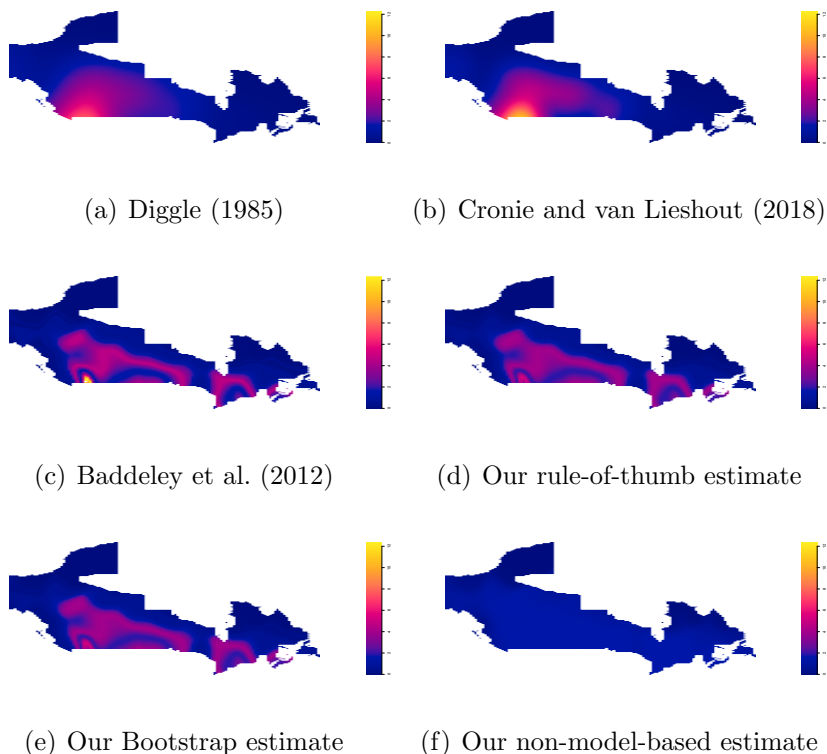


Figure 5: Comparison of intensity estimators using the temperature as the covariate.

From Fig. 3 and Fig. 5 we can assure that the information given by the temperature is relevant for the estimation. Diggle’s proposal in graph (a) can barely identify the area with more fires and, Cronie and van Lieshout (2018)’s proposal in graph (b) is doing it a bit better, but it can just get the

pattern in the centre of the region. On the other hand, using the information of the temperature, makes the intensity estimation more similar to the point process structure. Except for our non-model-based approach in graph (f) of Fig. 5, which clearly oversmooths, the other three bandwidth selectors provide similar intensity estimates.

Now we consider a more complicated model where both the temperature and the precipitation have an influence on the intensity, which seems to be natural in this example. To allow for several covariates we propose in Section 8 a multivariate extension of our model. See the model description in (13) and our multivariate intensity estimator in (14). Our multivariate extension involves a multivariate density estimate, which in the two-dimensional case requires to estimate a two-dimensional bandwidth matrix. In this example we have focused on estimating the bandwidth for the multivariate density estimator, and considered the multivariate plug-in rule proposed by Duong and Hazelton (2003) (here again we can see the advantage of our construction that allows to borrow methods from the density context to transfer them to point processes). This method is implemented in the function `Hpi` of the R-package `ks`, Duong (2017), and has the advantage of quick computation. Other choices are possible (see for example the recent book of Chacón and Duong (2018), for a revision on this topic) such as the popular cross-validation method. However cross-validation methods are quite computationally-intensive as we discussed in our simulations in the one-dimensional case, and become prohibited in a multivariate context.

Fig. 6 shows the result of the multivariate intensity estimation with the plug-in bandwidth. Comparing it with Fig. 5, we can see that the intensity estimate based on temperature and precipitation seems to better represent the wildfire point pattern in Fig. 3. This indicates (under model assumptions) that both covariates seem to contribute to the explanation of the resulting wildfire distribution.

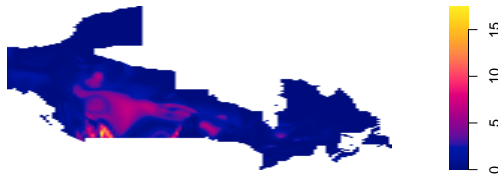


Figure 6: Kernel intensity estimation performed with two covariates, temperature and precipitation.

8. Further extensions

8.1. Spatio-temporal point processes

Including the time dimension is an appealing extension of the spatial processes considered in this paper. To the extent of our knowledge spatio-temporal point processes have not been addressed so far for the covariate model we have focused on in this paper. Here we provide a first approach to include the time in our covariate intensity model.

Let us define a spatio-temporal intensity depending on a covariate:

$$\lambda(u, t) = \rho(Z(u, t)),$$

where u and t are, respectively, the spatial and temporal coordinates, $Z : W \times T \subset \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$, is the covariate and ρ is an unknown real function. As it has been done before, we assume that Z is known so, we only need to estimate ρ in order to estimate λ .

As we have in this situation a covariate defined in the spatio-temporal domain, the theory previously developed in this paper immediately applies to this new situation, reproducing all the results just taking into account the change of dimension in the domain of the covariate.

Another possible framework in this spatio-temporal context can be the covariate having only a spatial or a temporal dependence, i.e, that Z is either a function of space or either a function of time but not both together. Hence, the intensity model is

$$\lambda(z, t) = \rho(Z(x), t) \quad \text{or} \quad \lambda(x, z) = \rho(x, Z(t)).$$

Remark that here ρ is still a real function but with a multivariate domain. In this case, following the literature about kernel intensity estimation in general spatio-temporal point processes, we propose the following estimators:

$$\hat{\rho}(z, t) = \sum_{i=1}^N \frac{1}{g^*(Z_i)} K_h(z - Z_i) K_s(t - t_i) \text{ and}$$

$$\hat{\rho}(x, z) = \sum_{i=1}^N \frac{1}{g^*(Z(t_i))} |H|^{-1/2} \mathbf{K}(H^{-1/2}(x - X_i)) K_s(Z(t) - Z(t_i)),$$

where $(X_1, t_1), \dots, (X_N, t_N)$ is the spatio-temporal pattern, K is a one-dimensional kernel function, s and h are one-dimensional bandwidth parameters, \mathbf{K} is a radially symmetric bivariate density function, H is a two-dimensional bandwidth matrix, and Z_i and g^* have been previously defined

in the manuscript. The theoretical developments provided in this paper may be extended to these situations with the appropriate regularity conditions.

8.2. Multivariate extension

Although the framework we have set up in this paper is defined to use one single one-dimensional covariate, it is of interest the extension to the multivariate case. Let us denote by $\mathbf{Z} = (Z_1, \dots, Z_p)$ a p -dimensional covariate providing possible significant information about the process, where $Z_i : W \subset \mathbb{R}^2 \rightarrow \mathbb{R} \forall i \in \{1, \dots, p\}$ are one-dimensional continuous covariates. We assume the following multivariate intensity model:

$$\lambda(u) = \boldsymbol{\rho}(\mathbf{Z}(u)). \quad (13)$$

Our proposal is to estimate λ in this context using again the convenient relationship between the intensity and the density functions described for a one-dimensional covariate. To formally define our multivariate estimator we first define the spatial cumulative distribution function in this multivariate context:

$$\mathbf{G}(z) = \frac{1}{|W|} \int_W 1_{\{\mathbf{Z}(u) \leq z\}} du,$$

where $\mathbf{Z}(u) \leq z$ refers to $(Z_1(u) \leq z_1) \cap \dots \cap (Z_p(u) \leq z_p)$, with $z \in \mathbb{R}^p$. We denote by \mathbf{g} the first derivative of \mathbf{G} , and by \mathbf{G}^* and \mathbf{g}^* we denote the unnormalised versions of \mathbf{G} and \mathbf{g} , respectively.

From these functions we define the corresponding multivariate relative density by $\mathbf{f}(\cdot) = \frac{\boldsymbol{\rho}(\cdot)\mathbf{g}^*(\cdot)}{m}$, and suggest the following standard kernel density estimator:

$$\hat{\mathbf{f}}_h(z) = \mathbf{g}^*(z) \frac{1}{N} \sum_{i=1}^N \frac{1}{\mathbf{g}^*(Z_i)} \mathbf{K}_H(z - Z_i) 1_{\{N \neq 0\}}, \quad (14)$$

where \mathbf{K} is a multivariate radially symmetric kernel function and H is a p -dimensional bandwidth matrix. Using derivations provided in Appendix D, our multivariate intensity estimator for model (13) can be written as:

$$\hat{\lambda}_H(x) = \sum_{i=1}^N \frac{1}{\mathbf{g}^*(Z_i)} \mathbf{K}_H(Z(u) - Z_i), \quad (15)$$

Analogous theoretical developments to those presented in Section 3 and Section 4 can be extended to the multivariate situation following the steps

we have already detailed in Appendix D, and applying some specifically tools designed for the multivariate analysis, which can be found in Scott (1992), Cucala (2006) and Fuentes-Santos et al. (2015). Full derivations are beyond the scope of this paper, as well as deriving specific asymptotic theory supporting bandwidth selection for our intensity estimator in the multivariate context. Practical suggestions for bandwidth selection have been given in Section 8.

9. Conclusions

We have considered kernel intensity estimation in the context of spatial point processes with covariates. We have set up an innovative theoretical framework in which we have been able to properly address the bias-variance trade-off of our kernel intensity estimator. We have derived the asymptotically optimal bandwidth for the estimator and proposed three data-driven bandwidth selectors: a rule-of-thumb, a bootstrap bandwidth and a non-model-based bandwidth. The two first are model-based and assume, respectively, normality and/or a Poisson underlying process. The third one does not rely on any assumption about the process and it is an extension (using covariates) of the selector proposed by Cronie and van Lieshout (2018). Our empirical studies have shown that our bootstrap bandwidth selector has the best performance, and it seems to happen even if the underlying process is not Poisson. The simple rule-of-thumb offers good results in some cases and the non-model-based approach behaves poorly in both Poisson and non-Poisson scenarios. This is not surprising taking into account that the non-model-based approach does not aim a good compromise between the bias and the variance of the kernel intensity estimator. Such bias-variance trade-off is not considered by the original non-model-based approach by Cronie and van Lieshout (2018) either. Having a reliable and completely model-free approach for bandwidth selection in the context of spatial point processes (with or without covariates) is still an open problem. In our theoretical developments in this paper, the Poisson assumption has been used to derive in detail the expressions of the mean squared error of our kernel intensity estimator. The assumption is convenient to simplify the calculations, however other generalizations should be possible but more complex. Similarly our bootstrap bandwidth is derived from a resampling procedure that makes use of the Poisson assumption. Even though it seems to work fine for some empirical situations where the assumption does not hold, it is possible to

adapt the resampling procedure to describe other processes if there is available information about them.

We have described an application with real data about wildfires in Canada. The data application has shown that our kernel intensity estimator based on covariates with the bootstrap bandwidth is a good choice in practice. This data application has also motivated a multivariate extension of our proposals to include more than one covariate. The extension comes naturally taking into account the structure of our intensity estimator, as a properly rescaled density estimator. The multivariate extension has been briefly described in this paper including not only more than one covariate but also the time dimension. We have used the extension in practice to allow two covariates to describe the wildfire intensity and it seems to work well. The theoretical developments in the multivariate framework requires further work. Dealing with a high number of covariates would also require further refinements of our proposal. The fully nonparametric density estimator involved in our multivariate proposal is limited by the curse of dimensionality. One solution to this problem consists of imposing an additive or multiplicative structure to the intensity function, or formulating a semiparametric model. We think this way forward is promising in the field of point process and this would generalise for example the appealing fully parametric model of Díaz-Avalos et al. (2014). The developments in this paper establish the basis for these further extensions.

10. Acknowledgments

The authors are grateful for constructive comments from two anonymous Reviewers and the Associate Editor. This work has been partially supported by the Spanish Ministry of Economy and Competitiveness, through grants number MTM2013-41383P and MTM2016-76969P, which includes support from the European Regional Development Fund (ERDF). Support from the IAP network StUDyS from Belgian Science Policy (P6/07), is also acknowledged. M.I. Borrajo has been supported by FPU grant (FPU2013/00473) from the Spanish Ministry of Education. The authors also acknowledge the Canadian Wildland Fire Information System for their activity in recording and freely providing all the real data used in this paper.

Appendix A. Useful tools for spatial point processes with covariates

Following Daley and Vere-Jones (1988) and Reiss (2012) we develop the following result that allows us to establish the model for the transformed point processes in the covariate space:

Theorem A. 1. *Let X be a spatial point process in $W \subset \mathbb{R}^2$ with intensity function of the form (1) for some real-valued function ρ . Then, $Z(X)$ is a one-dimensional point process in \mathbb{R} with intensity function ρg^* . Moreover, if the original point process is Poisson, then the transformed one preserves this property and it is also Poisson.*

The relationship between X and $Z(X)$ can be extended to the expected number of events through this result adapted from Federer (1969):

Theorem A. 2. *Let $W \subset \mathbb{R}^2$ be a bounded subset, $Z : W \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ a differentiable function with non-zero gradient in every point of W . Then, for $l : W \rightarrow \mathbb{R}$, defined as $l(u) = \lambda(u) \|\nabla Z(u)\|^{-1} : W \rightarrow \mathbb{R}^+$, it holds:*

$$\int_W \lambda(u) du = \int_W \lambda(u) \|\nabla Z(u)\|^{-1} \|\nabla Z(u)\| du = \int_{\mathbb{R}} \int_{Z^{-1}(y)} \lambda(u) \|\nabla Z(u)\|^{-1} dH(u) dy,$$

where $Z^{-1}(y) = \{u \in W / Z(u) = y\}$ and dH is the one-dimensional Hausdorff measure.

Applying this result to the unnormalised spatial cumulative distribution function $G^*(z) = |W|G(z)$ we have:

$$\begin{aligned} G^*(z) &= \int_W 1_{\{Z(u) \leq z\}} du = \int_W 1_{\{Z(u) \leq z\}} \|\nabla Z(u)\|^{-1} \|\nabla Z(u)\| du \\ &= \int_{\mathbb{R}} \int_{Z^{-1}(y)} 1_{\{Z(u) \leq z\}} \|\nabla Z(u)\|^{-1} dH(u) dy = \int_{-\infty}^z \int_{Z^{-1}(y)} \|\nabla Z(u)\|^{-1} dH(u) dy. \end{aligned}$$

Deriving with respect to z , we get $g^*(z) := (G^*)'(z) = \int_{Z^{-1}(z)} \|\nabla Z(u)\|^{-1} dH(u)$.

We can now rewrite the relationship between the original spatial point process intensity and the transformed one through an integral, which also

implies that the expected number of events in the corresponding region is the same in both processes:

$$\begin{aligned}
m &:= \int_W \lambda(u) du = \int_{\mathbb{R}} \int_{Z^{-1}(y)} \lambda(u) \|\nabla Z(u)\|^{-1} dH(u) dy \\
&= \int_{\mathbb{R}} \int_{Z^{-1}(y)} \rho(Z(u)) \|\nabla Z(u)\|^{-1} dH(u) dy \\
&= \int_{\mathbb{R}} \rho(y) \int_{Z^{-1}(y)} \|\nabla Z(u)\|^{-1} dH(u) dy = \int_{\mathbb{R}} \rho(y) g^*(y) dy.
\end{aligned}$$

Hence, we can finally have a one-dimensional inhomogeneous Poisson point process.

Appendix B -. Proof of Theorem 3.1

Recall that N is not a constant but a random variable, so we have a double stochastic scenario, on the one hand the randomness provided by N and on the other the randomness of the point process. To deal with this we use the conditional mean and we consider also some tools related to real number series.

Consider first the mean:

$$\begin{aligned}
E \left[\hat{f}_h(z) \right] &= E \left[E \left[\hat{f}_h(z) | N = n > 0 \right] \right] = \sum_{n=1}^{\infty} E \left[\hat{f}_h(z) | N = n \right] \mathbb{P}(N = n) \\
&= \sum_{n=1}^{\infty} E \left[\hat{f}_h(z) | N = n \right] \frac{e^{-m} m^n}{n!} = \sum_{n=1}^{\infty} \frac{g^*(z)}{m} (K_h \circ \rho)(z) \frac{e^{-m} m^n}{n!} \\
&= \frac{g^*(z)(K_h \circ \rho)(z)}{m} e^{-m} \left(\sum_{n=1}^{\infty} \frac{m^n}{n!} - 1 \right) = \frac{g^*(z)(K_h \circ \rho)(z)}{m} (1 - e^{-m}),
\end{aligned}$$

where we have used that:

$$\begin{aligned}
E \left[\hat{f}_h(z) | N = n \right] &= E \left[g^*(z) \frac{1}{n} \sum_{i=1}^n \frac{1}{g^*(Z_i)} K_h(z - Z_i) \right] = g^*(z) E \left[\frac{1}{g^*(Z_1)} K_h(z - Z_1) \right] \\
&= g^*(z) \int_{\mathbb{R}} \frac{1}{g^*(s)} K_h(z - s) \frac{\rho(s) g^*(s)}{m} ds = \frac{g^*(z)}{m} (K_h \circ \rho)(z).
\end{aligned}$$

Next, the variance:

$$\begin{aligned}
\text{Var} \left[\hat{f}_h(z) \right] &= E \left[\hat{f}_h^2(z) \right] - E \left[\hat{f}_h(z) \right]^2 = A(m) \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) \\
&\quad - (A(m) + e^{-m} - 1) \frac{(g^*(z))^2}{m^2} (K_h \circ \rho)^2(z) - \frac{(g^*(z))^2 (K_h \circ \rho)^2(z)}{m^2} (1 - e^{-m})^2 \\
&= A(m) \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) - (A(m) + e^{-2m} - e^{-m}) (g^*(z))^2 (K_h \circ \rho)^2(z),
\end{aligned}$$

where we have used the notation $A(m) := \left[\frac{1}{N} \mathbf{1}_{\{N \neq 0\}} \right]$ and,

$$\begin{aligned}
E \left[\hat{f}_h^2(z) \right] &= E \left[E \left[\hat{f}_h^2(z) | N = n > 0 \right] \right] = \sum_{n=1}^{\infty} E \left[\hat{f}_h^2(z) | N = n \right] \mathbb{P}(N = n) \\
&= \sum_{n=1}^{\infty} E \left[\hat{f}_h^2(z) | N = n \right] \frac{e^{-m} m^n}{n!} = \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) \sum_{n=1}^{\infty} \frac{e^{-m} m^n}{nn!} \\
&\quad + \frac{(g^*(z))^2}{m^2} (K_h \circ \rho)^2(z) \sum_{n=1}^{\infty} \frac{n-1}{n} \frac{e^{-m} m^n}{n!} = \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) E \left[\frac{1}{N} \mathbf{1}_{\{N \neq 0\}} \right] \\
&\quad + \frac{(g^*(z))^2}{m^2} (K_h \circ \rho)^2(z) E \left[\frac{N-1}{N} \mathbf{1}_{\{N \neq 0\}} \right] = A(m) \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) \\
&\quad - (A(m) + e^{-m} - 1) \frac{(g^*(z))^2}{m^2} (K_h \circ \rho)^2(z),
\end{aligned}$$

as well as:

$$\begin{aligned}
E \left[\hat{f}_h^2(z) | N = n \right] &= E \left[(g^*(z))^2 \frac{1}{n^2} \left(\sum_{i=1}^n \frac{1}{g^*(Z_i)} K_h(z - Z_i) \right)^2 \right] \\
&= \frac{(g^*(z))^2}{n^2} E \left[\left(\sum_{i=1}^n \frac{1}{g^*(Z_i)} K_h(z - Z_i) \right)^2 \right] \\
&= \frac{(g^*(z))^2}{n^2} E \left[\sum_{i=1}^n \frac{1}{(g^*(Z_i))^2} K_h^2(z - Z_i) + \sum_{i \neq j} \frac{1}{g^*(Z_i)} \frac{1}{g^*(Z_j)} K_h(z - Z_i) K_h(z - Z_j) \right] \\
&= \frac{(g^*(z))^2}{n} E \left[\frac{1}{(g^*(Z_1))^2} K_h^2(z - Z_1) \right] + \frac{(g^*(z))^2}{n^2} n(n-1) \left(E \left[\frac{1}{g^*(Z_1)} K_h(z - Z_1) \right] \right)^2 \\
&= \frac{(g^*(z))^2}{n} \int_{\mathbb{R}} \frac{1}{(g^*(s))^2} K_h^2(z - s) \frac{\rho(s) g^*(s)}{m} ds \\
&\quad + \frac{(g^*(z))^2 (n-1)}{n} \left(\int_{\mathbb{R}} \frac{1}{g^*(s)} K_h(z - s) \frac{\rho(s) g^*(s)}{m} ds \right)^2 \\
&= \frac{(g^*(z))^2}{mn} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) + \frac{(g^*(z))^2 (n-1)}{nm^2} (K_h \circ \rho)^2 (z).
\end{aligned}$$

Lastly compute the MSE in terms of bias and variance. We apply a Taylor expansion of second order to the mean and of first-order to the variance, getting:

$$\begin{aligned}
E \left[\hat{f}_h(z) \right] &= \frac{g^*(z)}{m} (K_h \circ \rho) (z) (1 - e^{-m}) \\
&= \frac{g^*(z)}{m} \left(\rho(z) + \frac{1}{2} h^2 \mu_2(K) \rho''(z) + o(h^2) \right) (1 - e^{-m}), \\
Bias \left[\hat{f}_h(z) \right] &= \frac{g^*(z)}{m} \left(-e^{-m} \rho(z) + (1 - e^{-m}) \frac{1}{2} h^2 \rho''(z) \mu_2(K) + o(h^2 (1 - e^{-m})) \right), \\
Var \left[\hat{f}_h(z) \right] &= A(m) \frac{(g^*(z))^2}{m} \left(K_h^2 \circ \frac{\rho}{g^*} \right) (z) \\
&\quad - (A(m) + e^{-2m} - e^{-m}) (g^*(z))^2 (K_h \circ \rho)^2 (z) \\
&= A(m) \frac{(g^*(z))^2}{m} \frac{1}{h} \frac{\rho}{g^*} (z) R(K) + o \left(\frac{A(m)}{mh} \right).
\end{aligned}$$

where $\mu_2(K) = \int_{\mathbb{R}} z^2 K(z) dz$ and $R(K) = \int_{\mathbb{R}} K^2(u) du$. Hence,

$$\begin{aligned} MSE \left[\hat{f}_h(z) \right] &= e^{-2m} f^2(z) + (1 - e^{-m})^2 \frac{h^4}{4} \left(\frac{\rho''(z) g^*(z)}{m} \right)^2 \mu_2^2(K) - \\ &\quad - e^{-m} (1 - e^{-m}) h^2 \mu_2(K) \frac{(g^*(z))^2 \rho(z) \rho''(z)}{m^2} + \frac{A(m)}{h} f(z) R(K) \\ &\quad + o(h^2(1 - e^{-m})e^{-m}) + o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right); \end{aligned}$$

from which we can easily derive its asymptotic expression:

$$\begin{aligned} AMSE \left[\hat{f}_h(z) \right] &= e^{-2m} f^2(z) + (1 - e^{-m})^2 \frac{h^4}{4} \left(\frac{\rho''(z) g^*(z)}{m} \right)^2 \mu_2^2(K) - \\ &\quad - e^{-m} (1 - e^{-m}) h^2 \mu_2(K) \frac{(g^*(z))^2 \rho(z) \rho''(z)}{m^2} + \frac{A(m)}{h} f(z) R(K). \end{aligned}$$

Appendix C -. Proof of Theorem 4.1

To compute the MSE^* of (7) we follow the same structure we have established for Theorem 3.1, with the difference that we are now under the bootstrap distribution.

The pointwise mean is:

$$\begin{aligned} E^* \left[\hat{f}_h^*(z) \right] &= E^* \left[E^* \left[\hat{f}_h^*(z) | N^* = n^* > 0 \right] \right] = \sum_{n^*=1}^{\infty} E^* \left[\hat{f}_h^*(z) | n^* = n^* \right] \mathbb{P}^*(N^* = n^*) \\ &= \sum_{n^*=1}^{\infty} E^* \left[\hat{f}_h^*(z) | N^* = n^* \right] \frac{e^{-\hat{m}} \hat{m}^{n^*}}{n^*!} = \sum_{n^*=1}^{\infty} \frac{g^*(z)}{\hat{m}} (K_h \circ \hat{\rho}_b)(z) \frac{e^{-\hat{m}} \hat{m}^{n^*}}{n^*!} \\ &= \frac{g^*(z)}{\hat{m}} (K_h \circ \hat{\rho}_b)(z) (1 - e^{-\hat{m}}), \end{aligned}$$

where we have used that:

$$\begin{aligned} E^* \left[\hat{f}_h^*(z) | N^* = n^* > 0 \right] &= E^* \left[g^*(z) \frac{1}{n^*} \sum_{i=1}^{n^*} \frac{1}{g^*(Z_i^*)} K_h(z - Z_i^*) \right] \\ &= g^*(z) E^* \left[\frac{1}{g^*(Z_1^*)} K_h(z - Z_1^*) \right] \\ &= \frac{g^*(z)}{\hat{m}} \int K_h(z - s) \hat{\rho}_b(s) ds = \frac{g^*(z) (K_h \circ \hat{\rho}_b)(z)}{\hat{m}}. \end{aligned}$$

We now apply a second order Taylor expansion to the mean in order to obtain the bias:

$$Bias^* \left[\hat{f}_h^*(z) \right] = -e^{-\hat{m}} \frac{\hat{\rho}_b(z)g^*(z)}{\hat{m}} + \frac{h^2}{2} \frac{\hat{\rho}_b''(z)g^*(z)}{\hat{m}} \mu_2(K)(1 - e^{-\hat{m}}) + o(h^2(1 - e^{-\hat{m}})).$$

To derive the variance we work as follows:

$$\begin{aligned} Var^* \left[\hat{f}_h^*(z) \right] &= E^* \left[\hat{f}_h^{*2}(z) \right] - E^* \left[\hat{f}_h^*(z) \right]^2 \\ &= \frac{(g^*(z))^2}{\hat{m}} \left(K_h \circ \frac{\hat{\rho}_b}{g^*} \right) (z) A(\hat{m}) - \frac{(g^*(z))^2}{\hat{m}^2} (K_h \circ \hat{\rho}_b)^2(z) (A(\hat{m}) + e^{-2\hat{m}} - e^{-\hat{m}}). \end{aligned}$$

Applying a first-order Taylor expansion we get:

$$Var^* \left[\hat{f}_h^*(z) \right] = \frac{\hat{\rho}_b(z)g^*(z)}{\hat{m}} R(K) \frac{A(\hat{m})}{h} + o\left(\frac{A(\hat{m})}{\hat{m}h} \right),$$

where we have used:

$$\begin{aligned} E^* \left[(\hat{f}_h^*(z))^2 \right] &= E^* \left[E^* \left[\hat{f}_h^{*2}(z) | N^* = n^* \right] \right] = \sum_{n^*=1}^{\infty} E^* \left[\hat{f}_h^{*2}(z) | n^* = n^* \right] \mathbb{P}^* (N^* = n^*) \\ &= \sum_{n^*=1}^{\infty} E^* \left[\hat{f}_h^{*2}(z) | N^* = n^* \right] \frac{e^{-\hat{m}} \hat{m}^{n^*}}{n^*!} = \frac{(g^*(z))^2}{\hat{m}} \left(K_h^2 \circ \frac{\hat{\rho}_b}{g^*} \right) (z) \sum_{n^*=1}^{\infty} \frac{e^{-\hat{m}} \hat{m}^{n^*}}{n^* n^*!} \\ &+ \frac{(g^*(z))^2}{\hat{m}^2} (K_h \circ \hat{\rho}_b)^2(z) \sum_{n^*=1}^{\infty} \frac{n^* - 1}{n^*} \frac{e^{-\hat{m}} \hat{m}^{n^*}}{n^*!} = \frac{(g^*(z))^2}{\hat{m}} \left(K_h^2 \circ \frac{\hat{\rho}_b}{g^*} \right) (z) A(\hat{m}) \\ &- \frac{(g^*(z))^2}{\hat{m}^2} (K_h \circ \hat{\rho}_b)^2(z) (A(\hat{m}) + e^{-\hat{m}} - 1), \text{ where:} \end{aligned}$$

$$\begin{aligned} E^* \left[\hat{f}_h^{*2}(z) | N^* = n^* \right] &= E^* \left[\left(g^*(z) \frac{1}{n^*} \sum_{i=1}^{n^*} \frac{1}{g^*(Z_i^*)} K_h(z - Z_i^*) \right)^2 \right] \\ &= \frac{(g^*(z))^2}{n^{*2}} E^* \left[\sum_{i=1}^{n^*} \frac{1}{(g^*(Z_i^*))^2} K_h^2(z - Z_i^*) + \sum_{i \neq j} \frac{1}{g^*(Z_i^*)} \frac{1}{g^*(Z_j^*)} K_h(z - Z_i^*) K_h(z - Z_j^*) \right] \\ &= \frac{(g^*(z))^2}{n^{*2}} \left(n^* E^* \left[\frac{1}{(g^*(Z_1^*))^2} K_h^2(z - Z_1^*) \right] + n^*(n^* - 1) E^* \left[\frac{1}{g^*(Z_1^*)} K_h(z - Z_1^*) \right]^2 \right) \\ &= \frac{(g^*(z))^2}{n^* \hat{m}} \left(K_h^2 \circ \frac{\hat{\rho}_b}{g^*} \right) (z) + \frac{(g^*(z))^2 (n^* - 1)}{n^* \hat{m}^2} (K_h \circ \hat{\rho}_b)^2(z). \end{aligned}$$

Finally, gathering square bias and variance together we obtain the MSE^* in (8).

Appendix D -. Technical details of the multivariate extension

The extension of Theorem A.1 and Theorem A.2 in Appendix A to the multivariate case can follow applying results in Daley and Vere-Jones (1988) and Reiss (2012) .

Theorem D. 1. *Let X be a spatial point process in $W \subset \mathbb{R}^2$ with intensity function of the form (13) for some real-valued function ρ , and $\mathbf{Z} : W \subset \mathbb{R}^2 \rightarrow \mathbb{R}^p$, a continuous function. Then $\mathbf{Z}(X)$ is a p -dimensional point process in \mathbb{R}^p with intensity function $\rho \mathbf{g}^*$. Moreover, if the original point process is Poisson, then the transformed one is also Poisson.*

Theorem D. 2. *Let $W \subset \mathbb{R}^2$ be a bounded subset, $\mathbf{Z} : W \subset \mathbb{R}^2 \rightarrow \mathbb{R}^p$ a measurable, Lipschitz and differentiable function with non-zero Jacobian in every point of W , $J\mathbf{Z}(u) \neq 0$. Then, for $l : W \rightarrow \mathbb{R}$, defined as $l(u) = \lambda(u)(J\mathbf{Z}(u))^{-1} : W \rightarrow \mathbb{R}^+$, it holds:*

$$\int_W \lambda(u) du = \int_W \lambda(u)(J\mathbf{Z}(u))^{-1} J\mathbf{Z}(u) du = \int_{\mathbb{R}^p} \int_{\mathbf{Z}^{-1}(y)} \lambda(u)(J\mathbf{Z}(u))^{-1} dH^p(u) dy,$$

where $\mathbf{Z}^{-1}(y) = \{u \in W : \mathbf{Z}(u) = y\}$ and dH^p is the p -dimensional Hausdorff measure.

Applying this result to the unnormalised version of the spatial cumulative distribution function defined above, we have:

$$\mathbf{G}^*(z) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_p} \int_{\mathbf{Z}^{-1}(y)} (J\mathbf{Z}(u))^{-1} dH^p(u) dy_1, \dots, dy_p,$$

and deriving with respect to z , we get $\mathbf{g}^*(z) = \int_{\mathbf{Z}^{-1}(z)} (J\mathbf{Z}(u))^{-1} dH^p(u)$.

We can now rewrite the relationship between the original spatial point process intensity and the transformed one through an integral, in a similar way we have done it in the Appendix A for the one-dimensional case, implying that the expected number of events in the corresponding region is the same

in both processes:

$$\begin{aligned}
m &:= \int_W \lambda(u) du = \int_{\mathbb{R}^p} \int_{\mathbf{Z}^{-1}(y)} \lambda(u) (J\mathbf{Z}(u))^{-1} dH^p(u) dy \\
&= \int_{\mathbb{R}^p} \int_{\mathbf{Z}^{-1}(y)} \boldsymbol{\rho}(\mathbf{Z}(u)) (J\mathbf{Z}(u))^{-1} dH^p(u) dy = \int_{\mathbb{R}^p} \rho(y) \int_{\mathbf{Z}^{-1}(y)} (J\mathbf{Z}(u))^{-1} dH^p(u) dy \\
&= \int_{\mathbb{R}^p} \boldsymbol{\rho}(y) \mathbf{g}^*(y) dy.
\end{aligned}$$

References

- Babu, G. J. and Feigelson, E. D. (1996). *Astrostatistics*, volume 3. CRC Press.
- Baddeley, A., Chang, Y. M., Song, Y., and Turner, R. (2012). Nonparametric estimation of the dependence of a spatial point process on spatial covariates. *Statistics and Its Interface*, 5:221–236.
- Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC Press.
- Borrajo, M. I., González-Manteiga, W., and Martínez-Miranda, M. D. (2017). Bandwidth selection for kernel density estimation for length-biased data. *Journal of Nonparametric Statistics*, 29(3):636–668.
- Brooks, M. M. and Marron, J. S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1):157–165.
- Cao, R. (1993). Bootstrapping the mean integrated squared error. *Journal of Multivariate Analysis*, 45(1):137–160.
- Chacón, J. E., and Duong, T. (2018). *Multivariate kernel smoothing and its applications*. Chapman and Hall/CRC.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley-Interscience, revised edition edition.
- Cowling, A., Hall, P., and Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a poisson point process. *Journal of the American Statistical Association*, 91(436):1516–1524.

- Cronie, O. and van Lieshout, M. N. M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika*, 105(2):455–462.
- Cucala, L. (2006). *Espacements bidimensionnels et données entachés d’erreurs dans l’analyse des processus ponctuels spatiaux*. PhD thesis, Université des Sciences de Toulouse I.
- Daley, D. J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes*. Springer Verlag.
- Di Iorio, T., Anello, F., Bommarito, C., Cacciani, M., Denjean, C., De Silvestri, L., Di Biagio, C., di Sarra, A., Ellul, R., Formenti, P., et al. (2013). Long range transport of smoke particles from canadian forest fires to the mediterranean basin during june 2013. In *AGU Fall Meeting Abstracts*.
- Díaz-Avalos, C., Juan, P., and Mateu, J. (2014) Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. *Stochastic environmental research and risk assessment*, 28(3):593–609.
- Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34(2):138–147.
- Diggle, P. and Marron, J. S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association*, 83:793–800.
- Diggle, P. J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 153(3):349–362.
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press.
- Duong, T. (2017) ks: Kernel Smoothing. R package version 1.10.5. <https://CRAN.R-project.org/package=ks>
- Duong, T. and Hazelton, M. (2003) Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*, 15(1):17–30.

- Federer, H. (1969). Geometric measure theory. In *Die Grundlehren der mathematischen Wissenschaften*, volume 153. Springer-Verlag New York Inc.
- Flannigan, M. and Harrington, J. (1988). A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). *Journal of Applied Meteorology*, 27(4):441–452.
- Fuentes-Santos, I., González-Manteiga, W., and Mateu, J. (2015). Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial poisson point processes. *Scandinavian Journal of Statistics*, 43(2):416–435.
- Gatrell, A. C., Bailey, T. C., Diggle, P. J., and Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21(1):256–274.
- Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association*, 103(483):1238–1247.
- Guan, Y., and Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102(480):1377–1386.
- Guan, Y., and Shen, Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, 97(4):867–880.
- Illian, J. B., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons.
- Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16(3):389–405.
- Illian, J. B., Sørbye, S. H., and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, 6(4):1499–1530.
- Kerscher, M. (2000). Statistical analysis of large-scale structure in the universe. In *Statistical physics and spatial statistics*, volume 554, pages 36–71. Springer.

- Klemelä, J. S. (2009). *Smoothing of multivariate data: density estimation and visualization*. (Vol. 737). John Wiley and Sons.
- Mrkvicka, T., Muska, M. and Kubecka, J. (2014). Two step estimation for Neyman-Scott point process with inhomogeneous cluster centers. *Statistics and Computing*, 24(1): 91–100.
- Law, R., Illian, J., Burslem, D. F., Gratzner, G., Gunatilleke, C., and Gunatilleke, I. (2009). Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, 97(4):616–628.
- Marron, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics*, 13(3-4):187–208.
- Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.
- Ogata, Y. and Zhuang, J. (2006). Space–time etas models and an improved extension. *Tectonophysics*, 413(1):13–23.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.
- Reiss, R. D. (2012). *A course on point processes*. Springer Science & Business Media.
- Rogers, B., Randerson, J., and Bonan, G. (2013). High-latitude cooling associated with landscape changes from north american boreal forest fires. *Biogeosciences*, 10(2):699–718.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

- Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial–temporal point process. *Journal of Statistical Planning and Inference*, 128(1):79–93.
- Schoenberg, F. P. (2011). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98:789–795.
- Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization..* John Wiley and Sons.
- Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualisation.* Wiley.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* CRC press.
- Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, 15(1):61–78.
- Van Lieshout, M. (2000). *Markov point processes and their applications.* World Scientific.
- Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman–scott processes. *Biometrics*, 63(1):252–258.
- Walter, C., Freitas, S., Kraut, I., Rieger, D., Vogel, H., and Vogel, B. (2014). Influence of 2010 canadian forest fires on cloud formation on the regional scale. In *AGU Fall Meeting Abstracts*.
- Yue, Y. R., and Loh, J. M. (2011). Bayesian semiparametric intensity estimation for inhomogeneous spatial point processes. *Biometrics*, 67(3):937–946.