# CHARACTERISATION OF POINT PROCESSES WITH COVARIATES THROUGH THE FIRST-ORDER INTENSITY FUNCTION

M.I. Borrajo[1], W. González-Manteiga [1] e M.D. Martínez-Miranda[2]

[1]Departamento de Estatística, Análise Matemática e Optimización. Universidade de Santiago de Compostela.
[2] Departamento de Estadística e Investigación Operativa. Universidad de Granada.

## ABSTRACT

Point processes are a branch of spatial statistics, whose main aim is to study the geometrical structure of patterns formed by objects that are distributed randomly in number and space. This type of data arise in many different fields such as ecology, J. B. Illian, Møller, and Waagepetersen (2009); epidemiology, P. J. Diggle (1990); astronomy, Babu and Feigelson (1996); forestry, Stoyan and Penttinen (2000) and seismology, Ogata and Zhuang (2006) and Schoenberg (2011). The first-order intensity function is one of the characteristic functions of a point processes and has generated a great interest since the eighties. Covariates are a way of gathering extra information, allowing more precise models for point processes without assuming parametric restrictions.

This work provides two main contributions: first a new kernel intensity estimator is proposed and a theoretical framework to guarantee its consistency is developed. A new bootstrap resampling procedure as well as two new data-driven bandwidth selection methods are suggested. The second contribution is a test o check whether the assumed model with covariates is or not appropriate, for which a $L^2-$distance based statistic is proposed and is proved to be asymptotically normal. The good behaviour of all the proposed techniques is shown through different extensive simulation studies and they are also applied to a real data set consisting of locations of gold deposits in the Murchison area of Western Australia.

**Keywords**: Point processes; Covariates; First-order intensity; Kernel estimation; Testing significance.

## 1. INTRODUCTION TO POINT PROCESSES

Point process theory has its roots in the introduction of the Poisson distribution in 1837 by Siméon-Denis Poisson, which derived it as a passage to the limit from the binomial distribution, It allowed to predict the pattern in which random events of low probability occurs in the course of a large number of trials. The main advances in the field of point processes have been done during the last decades of the 20th century, see Daley and Vere-Jones (1988), Moller and Waagepetersen (2003), J. Illian, Penttinen, Stoyan, and Stoyan (2008), P. J. Diggle (2013) and Baddeley, Rubak, and Turner (2015) for an overview of the theoretical background of point processes.

In this work we have focused on spatial point processes, i.e, point processes in $\mathbb{R}^2$. First of all we need to introduce some basic concepts that are fundamental in the understanding of the theory.

**Definition 1** *Let $X$ be a point process in $W \subset \mathbb{R}^2$, let $\mathcal{P}(W)$ denote the parts of $W$, i e., the family of all possible subsets of $W$; then $N : \mathcal{P}(W) \to \mathbb{Z}^+$, where $N(A)$ denote the number of occurrences of the process in the set $A \in \mathcal{P}(W)$, is defined as the* ***counting measure*** *associated with the process $X$.*

**Definition 2** *Let $X$ be a point process in $W \subset \mathbb{R}^d$ and $N$ its associated counting measure, then a realisation (or a sample) of the process, $X_1, \ldots, X_N$, is called a* ***point pattern****.*

Figure 1 provides examples of point patterns. These graphics gather the existing possibilities for a pattern in terms of "clusterisation": the one on the left is the most clustered, i.e., we can clearly distinguish groups of points that lie together; the one on the right is regular, the points tend to avoid each other; and the one in the middle, shows points randomly placed and do not following any specific pattern.

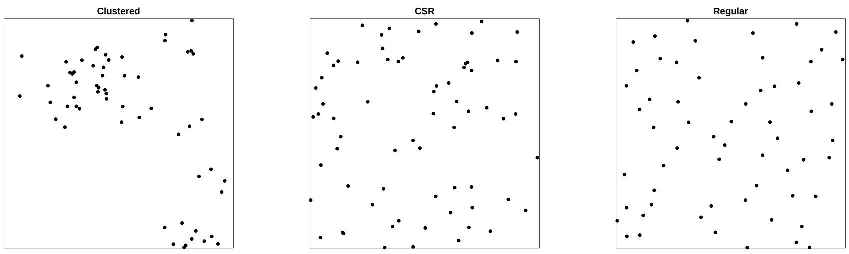



Figure 1: Example of clustered (left), random (middle) and regular (right) point patterns.

Let focus on the random case. In this situation the points have no preference for any spatial location (homogeneity) and the information about the outcome in one region of space has no influence on the outcome in other regions (independence). This situation is known as complete spatial randomness (CSR) and is formally characterise by the following properties:

CSR1 The number of events in a planar region $A \subset W$ follows a Poisson distribution with mean $\lambda|A|$, with $\lambda \in \mathbb{R}^+$.

CSR2 Conditional to $N = n$ events in $W$, the points $X_1, \ldots, X_n$ are an independent identically distributed (i.i.d.) sample from the uniform distribution in $W$.

The interest on CSR lies on the basis that it represents an idealised standard which should be tested as a first step in any practical situation. In case we accept this hypothesis the modelisation of the process became almost trivial, see P. J. Diggle (2013) for more explanation on this.

The Poisson point process is one of the most used and studied due to its particularly convenient properties. There are two different types: the homogeneous and the nonhomogeneous. The first one is the simplest existing model and it corresponds exactly to CSR, it is usually denoted by $Pois(\lambda)$ where $\lambda$ is known as rate or intensity. It satisfies some interesting properties, such as stationarity (the distribution of the point process is invariant under translations) and isotropy (the distribution is invariant under rotations around the origin). The nonhomogeneous Poisson point process is an extension of this one, where the rate value $\lambda$ is now a function of the location, so $\lambda(x)$ with $x \in W$, and it is used in many practical situations. Formally, the nonhomogeneous Poisson point process is defined as follows:

**Definition 3** *Let $X$ be a point process in $W \subset \mathbb{R}^2$ with $N$ its associated counting measure and a spatially varying function $\lambda(x)$. Then we say that $X$ is an nonhomogeneous Poisson point process with intensity $\lambda(x)$ if it verifies:*

*NHP1 $N(A)$ has a Poisson distribution with mean $\int_A \lambda(x)dx$ for any $A \subset W$.*

*NHP2 Conditional to $N = n$, the $n$ events in $W$ form an independent random sample from the distribution in $W$ with density proportional to $\lambda(x)$.*

There is another relevant characteristic that usually arise in point process theory, the ***edge effect***. As we have seen, point process are stochastic mechanism that are observed in a bounded region, even though they may be defined in a bigger one. Hence, when studying them we have to take care about it. We will see later in this manuscript some assumptions that allow us to directly avoid this effect, and when this does not happen, how edge correction terms are introduced.

Marks and covariates are two different ways of including some extra information in a point process model, (J. Illian et al., 2008, Chap. 5). The main difference among them is that marks are directly linked to the events, while covariates include information about the whole observation region. This second scenario is where we will work on.

The rest of this paper is structured as follows: in Section 2 we make a brief overview on the existing methods in kernel intensity estimation. In Section 3 we set up a new framework for kernel intensity estimation with covariates and develop asymptotic theory for it; we propose a smooth bootstrap procedure and two new data-driven bandwidth selection methods; a simulation study over all these new procedures and the existing competitors is carried out. Section 4 is devoted to test the goodness-of-fit of the intensity model previously assumed through an $L^2$ test statistic, which asymptotic normality is proved and it behaviour is analysed in a simulation study. An application to a real data set is detailed in Section 5 and finally we draw some conclusions in Section 6.

## 2. THE FIRST-ORDER INTENSITY FUNCTION

Modelling the first-order intensity function is one of the main aims in point process theory. This function computes the mean number of events per (length, area or volume) unit, and it is one of the functions that can completely characterizes a point process.

Let $X$ be a point process defined in a region $W \subset \mathbb{R}^2$, where $W$ is assumed to have finite positive area and let $X_1, \ldots, X_N$ be a realisation of the process. The first-order intensity, from now on referred as intensity, is defined following P. J. Diggle (2013) as:

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

where $|dx|$ denotes the area of an infinitesimal region containing the point $x \in W$.

There is extensive literature on parametric point processes models and intensity estimation under this assumption, see Schoenberg (2005). Assuming a parametric model for the intensity function may be a way of estimating it, using for instance a likelihood score such as the Akaike Information Criteria (AIC), see Van Lieshout (2000), Moller and Waagepetersen (2003) and P. J. Diggle (2013) or pseudolikelihood procedures, see Waagepetersen (2007). In the Bayesian context J. B. Illian, Sørbye, and Rue (2012) proposed some models based on log-gaussian Cox processes. However it is well known that we can obtain unreliable estimates when the assumed parametrization deviates from the true intensity. This is the main reason that supports the use of nonparametric techniques, in particular kernel methodology. P. J. Diggle (1985) proposed the first kernel intensity estimator for one dimensional point process, which have been easily extended to the plane $\mathbb{R}^2$:

$$\hat{\lambda}_H^D(x) = \frac{\sum_{i=1}^N K_H(x - X_i)}{p_H(x)}, \quad x \in W, \tag{1}$$

where $H$ is a bandwidth matrix, $K_H(x) = |H|^{-1/2} K\left(H^{-1/2}x\right)$ and $p_H = \int_W |H|^{-1/2} K(H^{-1/2}(x-y))dy$ is an edge correction term where $K$ denotes a kernel function.

This estimator has been widely used during decades for exploratory analysis, but the inference performed with it has been limited due to its lack of consistency. To overcome this problem, Cucala (2006) defined the "density of events locations" as $\lambda_0(x) = \lambda(x)/m$, where $m = \int_W \lambda(x)dx$ is the expected number of events lying on W. He proposes a kernel estimator:

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

with $1_{\{\}}$ denoting the indicator function and $h$ a one-dimensional bandwidth parameter. He proves its consistency under an increasing domain asymptotic framework. Fuentes-Santos, González-Manteiga, and Mateu (2015) extended these ideas to the two-dimensional situation using bandwidth matrices, as it has been done in the context of multivariate density estimation.

Now, let $Z : W \subset \mathbb{R}^2 \to \mathbb{R}$ be a spatial continuous covariate that is exactly known in every point of the region of interest $W$ and $Z_1, \ldots, Z_N$ the realisation of the transformed process, i.e, $Z_i = Z(X_i)$. In practice this covariate will commonly be know 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.

In some cases it can be assumed that spatial point process intensity depends on the covariate, see for instance Baddeley, Chang, Song, and Turner (2012), so it can be written

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

where $\rho$ is an unknown function. As $Z$ is known, only $\rho$ needs to be estimated in order to obtain an estimation of $\lambda$, which is the target.

To this purpose it is necessary to deal with the transformed univariate point process, $Z(X)$, and establish the theoretical relationship between this one and the original spatial point process $X$. If $X$ is a Poisson point process in $W \subset \mathbb{R}^2$ with intensity function (2), then $Z(X)$ is a Poisson point process in $\mathbb{R}$ with intensity $\rho g^\star$ and with the same expected number of events, where $g^\star$ is the unnormalized version of the derivative of the spatial cumulative distribution function.

The proposals in Guan (2008) and Baddeley et al. (2012), following the assumption (2), are similar kernel intensity estimators. Guan (2008) develop a kernel estimator based on the definition of the distance between two points by the distance trough their covariates values:

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

with $q_h(u) = \int_W K_h(||Z(u) - Z(s)||)ds$ the edge correction term.

Considering the increasing domain asymptotic framework and adding also some suitable assumptions, the consistency of his proposal is proved. A bandwidth selection criterion using cross-validation techniques is defined, as well as a dimension reduction tool that allows to handle with high-dimensional covariates.

Baddeley et al. (2012) propose two types of nonparametric intensity estimators based on two nonparametric density estimators, one relies on local likelihood and the other on kernels. We will focus on the last one, in particular on a kernel intensity estimator for the $\rho$ function with a one-dimensional covariate:

$$\hat{\rho}_W(z) = \sum_{i=1}^N \frac{1}{g^\star(Z_i)} K_h(z - Z_i). \tag{3}$$

Note that to obtain the bandwidth parameter $h$ in practice, Baddeley et al. (2012) use the common Silverman's rule-of-thumb for density estimation applied to the transformed data and they do not have any ad-hoc procedure.

## 3. A CONSISTENT THEORETICAL FRAMEWORK FOR KERNEL INTENSITY ESTIMATION

### 3.1. New intensity estimation procedure and its asymptotics

In this section we work under the transformed space assuming (2), and the point process obtained from the original one, $X$, through the covariate, $Z(X)$, defined in the previous section.

Now, following the idea of Cucala (2006) we use the relationship between the intensity and the density function, and we define the following "artificial" density function:

$$f(\cdot) = \frac{\rho(\cdot) g^\star(\cdot)}{m}. \tag{4}$$

What we propose is to take profit of this relationship: firstly estimating the density, $f$, and then going back to our target problem, that is the estimation of the intensity $\lambda(u) = \rho(Z(u))$.

Following the above notation, we define the estimator of the relative density as follows:

$$\hat{f}_h(z) = g^\star(z) \frac{1}{N} \sum_{i=1}^N \frac{1}{g^\star(Z_i)} K_h(z - Z_i)\, 1_{\{N \neq 0\}}, \tag{5}$$

where $K$ is a univariate kernel function and $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$. This is an estimator of $f$, and once we have it, we can go back to the intensity function by plug-in and letting $\hat{\lambda}(u) = \hat{\rho}_h(Z(u))$, where $\hat{\rho}_h$ can be replaced by (3).

In the following statement we obtain the value of the pointwise mean and variance of $\hat{f}_h$ with the corresponding error rates, as well as its mean squared error (MSE), which is defined as $MSE(h,z) = E\left[\left(\hat{f}_h(z) - f(z)\right)^2\right]$.

Hereafter we will establish that our point process $X$ is an nonhomogeneous Poisson point process in $W \subset \mathbb{R}^2$. Although the intensity estimator we propose, as well as the bandwidth selectors, can be applied to non-Poisson processes, the previous assumption is required to prove the consistency of the estimator. We also need to introduce some regularity conditions:

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

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

(A.3) $G$ is three times differentiable.

(A.4) $z$ is a continuity point of $\rho$.

(A.5) $\rho$ is three times differentiable.

Notice that we use an infill structure asymptotic framework, which means that the observation region remains fixed while the sample size increases. In this scenario the bandwidth $h$ is considered as a function of the sample size, this is, $h \equiv h(m)$ and hence it is a sequence of values when $m \to \infty$.

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

$$E\left[\hat{f}_h(z)\right] = \frac{g^\star(z)(K_h \circ \rho)(z)}{m}\left(1 - e^{-m}\right) \qquad \text{and}$$
$$Var\left[\hat{f}_h(z)\right] = A(m)\frac{(g^\star(z))^2}{n}\left(K_h^2 \circ \frac{\rho}{g^\star}\right)(z) - (A(m) + e^{-2m} - e^{-m})(g^\star(z))^2(K_h \circ \rho)^2(z),$$

*where $\circ$ denotes the convolution between two functions. Moreover, adding condition (A.5) we have:*

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

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

Now, defining the mean integrated square error (MISE) as $MISE(h) = E\int\left(\hat{f}_h(z) - f(z)\right)^2 dz$, and denoting by $AMISE$ its asymptotic version, the next result is a consequence of Theorem 1:

**Corolary 1** *Under conditions (A.1) to (A.3) and (A.5),*

$$MISE(h) = e^{-2m}R(f) + (1 - e^{-m})^2\frac{h^4}{4}R\left(\frac{\rho''g^\star}{m}\right)\mu_2^2(K) -$$
$$- e^{-m}(1 - e^{-m})h^2\mu_2(K)\int_{\mathbb{R}}\frac{g^\star(z)\rho''(z)f(z)}{m}dz + \frac{A(m)}{h}R(K) +$$
$$+ o(h^2(1 - e^{-m})e^{-m}) + o(h^4(1 - e^{-m})^2) + o\left(\frac{A(m)}{mh}\right) \qquad \text{and}$$

$$AMISE(h) = (1 - e^{-m})^2\frac{h^4}{4}R\left(\frac{\rho''g^\star}{m}\right)\mu_2^2(K) + \frac{A(m)}{h}R(K).$$

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

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

The proofs of the results included in this Section 3 can be seen in Borrajo, González-Manteiga, and Martínez-Miranda (2017b).

### 3.2. Resampling bootstrap method

Another important point in statistical inference are the resampling procedures, which have been widely used in different contexts to perform inference and calibrate the distribution of statistics in goodness-of-fit tests. The smooth bootstrap procedure for point processes with covariates we propose in this section is based on the work of Cao (1993) for kernel density estimation, and Cowling, Hall, and Phillips (1996), for the intensity estimation of a Poisson point process.

Let $X_1, \dots, X_n$ be a realisation of the spatial point process $X$, construct $Z_1, \dots, Z_n$ the associated realisation of the transformed univariate process, let $\hat{f}_g$ be the density estimator in (5) and $\hat{\rho}_g$ the estimator defined in the previous section, where $g$ is a pilot bandwidth.

Now, conditional on $Z_1, \dots, Z_n$, let $N^* \sim Poiss(\hat{m})$ with $\hat{m} := \int_{\mathbb{R}} \hat{\rho}_g(z) g^\star(z) dz$, generate $n^*$ a realisation of this random variable $N^*$ and then draw $Z_1^*, \dots, Z_{n^*}^*$ by sampling randomly with replacement $n^*$ times from the distribution with density proportional to $g^\star \hat{\rho}_g$, i.e. $\tilde{f}_g = \frac{\hat{\rho}_g g^\star}{\hat{m}}$.

Denote by $Z^*$ the random variable generated by the bootstrap method presented above. From the bootstrap sample define the density estimator as:

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

hence, using equation (4) we get the associated estimator of $\rho$:

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

and then we plug-in it in (2) to obtain an estimator of $\lambda$.

The following result provides the expression of the mean, variance and mean squared error of $\hat{f}_h^*$ under the bootstrap distribution; hereafter we use $E^*$, $Var^*$ and $MSE^*$ to refer to them.

**Theorem 2** *Under hypothesis (A.1) to (A.4) we get:*

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

*where* $A(\hat{m}) := E^*\left[\frac{1}{N^*} 1_{\{N^* \neq 0\}}\right]$. *Moreover, adding condition (A.5) we have*

$$MSE^*(h,z) = e^{-2\hat{m}} (\tilde{f}_g(z))^2 + \frac{h^4}{4} (\hat{\rho}_h''(z))^2 \frac{(g^\star(z))^2}{\hat{m}^2} \mu_2^2(L)(1 - e^{-\hat{m}})^2 -$$
$$- e^{-\hat{m}}(1 - e^{-\hat{m}}) h^2 \tilde{f}_g(z) \frac{\hat{\rho}_g''(z) g^\star(z)}{\hat{m}} \mu_2(L) + \frac{A(\hat{m})}{h} R(L) + o_P(h^4(1 - e^{-\hat{m}})^2) +$$
$$+ o_P(h^2(1 - e^{-\hat{m}}) e^{-\hat{m}}) + o_P\left(\frac{A(\hat{m})}{\hat{m} h}\right); \quad (8)$$

*remind that we have defined* $\tilde{f}_g(\cdot) = \frac{\hat{\rho}_g(\cdot) g^\star(\cdot)}{\hat{m}}$.

In the same way as we have done for Corollary 1, the integrated and asymptotic version of the $MSE^*$ can be easily deduced from the previous result:

**Corolary 2** *Under conditions (A.1) to (A.3) and (A.5),*

$$MISE^*(h) = e^{-2\hat{m}} R(\tilde{f}_g) + \frac{h^4}{4} R\left(\frac{\hat{\rho}_g^{''} g^\star}{\hat{m}}\right) \mu_2^2(L)(1-e^{-\hat{m}})^2 -$$

$$- e^{-\hat{m}}(1-e^{-\hat{m}})h^2\mu_2(L)\int \frac{\tilde{f}_g(z)\hat{\rho}_g^{''}(z)g^\star(z)}{\hat{m}}dz + \frac{A(\hat{m})}{h}R(L)+$$

$$+ o(h^4(1-e^{-\hat{m}})^2) + o_P(h^2(1-e^{-\hat{m}})e^{-\hat{m}}) + o_P\left(\frac{A(\hat{m})}{\hat{m}h}\right) \qquad and$$

$$AMISE^*(h) = \frac{h^4}{4} R\left(\frac{\hat{\rho}_g^{''} g^\star}{\hat{m}}\right) \mu_2^2(L)(1-e^{-\hat{m}})^2 + \frac{A(\hat{m})}{h}R(L).$$

*Therefore the asymptotic expression of the optimal bootstrap bandwidth is:*

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

*which is a plug-in version of* (6)*.*

All the results above lead to the following Corollary.

**Corolary 3** *Under assumptions (A.1) to (A.4) $MISE^*$ and $AMISE^*$ are consistent estimators of $MISE$ and $AMISE$, respectively.*

### 3.3. Data-driven bandwidth selection

We have proposed in this work a kernel intensity estimation based on (5), and as it is well known this implies to choose a bandwidth parameter which determines the degree of smoothness to be considered in the estimation. The choice of the bandwidth parameter is crucial and it has motivated several papers in the literature in the recent decades, see for example Marron (1988), Scott (1992) and Silverman (1986) for an earlier full description of the problem. There is a lot of theory developed on this issue in areas of statistics such as density estimation and regression, meanwhile in the context of point processes it has received less attention. P. J. Diggle (1985) proposed a bandwidth selector based on the minimisation of the mean squared error (MSE) of his estimator. Later, P. Diggle and Marron (1988) showed the equivalence, for Cox processes in $\mathbb{R}$, between that procedure and the standard least-squares cross-validation method used in kernel density estimation. This is an example of the strong connection between this two problems, density and intensity estimation. Brooks and Marron (1991) proved the optimality of the least-squares cross-validation bandwidth for one-dimensional nonhomogeneous Poisson point processes. Fuentes-Santos et al. (2015) develop an extension of Cucala's theory to the two-dimensional case, and propose a two-dimensional bandwidth selection procedure based on a bootstrap method.

We describe now two new bandwidth selection methods for the intensity estimation based on (5) that we propose. These methods consist of adaptations of common selectors in the field of density estimation that have not yet been defined nor implemented in the point process framework. We propose a Normal scale rule (rule-of-thumb) and a bootstrap selector derived from the consistent resampling bootstrap procedure detailed in the previous section. All these proposals are based on estimating the infeasible optimal expression (6) where $m$, $A(m)$ and $\rho^{''}$ are unknown elements.

*Rule-of-thumb for bandwidth selection*

The basis idea of this method is the same as in Silverman (1986): we assume that the underlying density (4) is Normal, $N(\mu, \sigma)$, with the parameters being estimated from the data, and in this way we replace the unknown values in (6).

In the point processes framework the computation is slightly different from the one used in the context of density estimation, because here the density is only a feature to get the intensity. To begin with, we have to remark that in our context, (6) has some other unknown elements apart from $f$, such as $m$ and $A(m)$. The first one is the expected number of points, that in practice can be estimated by the sample size $n$, and the second one by $1/n$.

The only unknown element left is $\rho^{''}$. However, assuming that $f = \frac{\rho g^\star}{m}$ is Normal, we can derive that

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

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

*Bootstrap for bandwidth selection*

The asymptotic expression of the optimal bootstrap bandwidth can be considered to derive a consistent bandwidth estimate. Cao (1993) suggested such approach for kernel density estimation with complete data and in Borrajo, González-Manteiga, and Martínez-Miranda (2017a) detail the result for length-biased data, as well as some remarks to extend it to general weighted data.

The expression (9) that we use to build this selector, has a few quantities that need to be computed: $\hat{m}$, $A(\hat{m})$ and $R(\frac{\hat{\rho}_g^{''}g^\star}{\hat{m}})$. The first two can be easily calculated through numerical integration methods such as Simpson's rule, while the last one requires some more development.

The main challenge in the estimation of $R(\frac{\hat{\rho}_g^{''}g^\star}{\hat{m}})$ is to obtain an appropriate value for the pilot bandwidth $g$. Regarding Cao (1993) and Borrajo et al. (2017a) we can assume that the order of that bandwidth in our context is $m^{-1/7}$, and that the constant has a slight influence on the final result. Hence we propose to use as pilot bandwidth a re-scaled version of the rule-of-thumb previously defined $\hat{g} = (m^{-1/5}/m^{-1/7})\hat{h}_{RT}$; obviously in practice we do not know the value of $m$, so we use the best approximation we can have which is the sample size of the corresponding realization of the point process.

Then, the bootstrap bandwidth we propose is:

$$\hat{h}_{\mathrm{Boot}} = \left(\frac{A(\hat{m})R(L)}{\mu_2^2(L)(1-e^{-\hat{m}})^2 R\left(\frac{\hat{\rho}_{\hat{g}}^{''}g^\star}{\hat{m}}\right)}\right)^{-1/5}.$$

We perform a simulation study to analyse the behaviour of the methods proposed in this work. Firstly we analyse the performance of the intensity estimator defined in (3), using our two bandwidths selection proposals: the rule of thumb and the bootstrap method. We compare these with the only bandwidth selector that has been previously proposed by Baddeley et al. (2012), which is the common Silverman's rule-of-thumb for density estimation.

An overview of the results of this study indicates that in general, the bootstrap bandwidth seems to perform better than the others in most of the cases, and when this does not occur, our procedure is still competitive. Any of the other methods are not far away from it, even though the rule-of-thumb specifically designed for spatial point processes shows a slightly better behaviour than the Silverman's rule-of-thumb, specially for small sample sizes. In terms of variability, the three compared methods are similar, even though the bootstrap estimates shows in general smaller values.

Looking at the bias of bandwidth estimates, i.e., the relative difference between the selected bandwidth and the benchmark $\hat{h}_{\mathrm{MISE}}$ (the minimiser of the Monte Carlo approximation based on 500 samples of the MISE criterion), the bootstrap bandwidth selector outperforms far more better than the others in most of the studied models. It is also shown that the rule-of-thumb and Silverman's procedures show the bias in the same direction, to be more specific all of them tend to choose smaller bandwidths than the optimal one, while in general the bootstrap selector does the opposite.

To complete our analysis, we have carried out a parallel simulation study to compare our proposals to the competitor described in Guan (2008). For Guan's intensity estimator we have considered two bandwidth choices: the (Monte Carlo approximated) optimal MISE bandwidth, which is considered as a benchmark for this estimator, and the practical least-squares cross-validation bandwidth proposed in his paper. For our proposal we have considered the intensity estimator with the benchmark $\hat{h}_{\text{MISE}}$ defined above, and our bootstrap bandwidth selector. Our proposal performs considerably better than Guan's for all the simulated models except one. In the best case for the intensity estimators, this is calculating the intensity estimators with the infeasible benchmarks, our estimator achieves smaller relative errors with slightly lower variability. Considering the practical bandwidth choices for each estimator our bootstrap approach clearly beats the cross-validation method. On the other hand, there is one model that seems to be a good scenario for Guan's estimator and his practical cross-validation method, but even in this case our bootstrap proposal is still competitive. The detailed results of this simulation study can be seen in Borrajo et al. (2017b).

## 4. TESTING THE COVARIATE DEPENDENCE OF THE INTENSITY

Until now we have assumed model (2) being true and we have developed all the methodology under this assumption. Hence, it would be relevant to check whether this model is appropriate; so we will now focus on testing the significance of the covariate.

We formally want to test a null hypothesis $H_0 : \lambda(x) = \rho(Z(x))$, $x \in W$ versus a general alternative in which the intensity function is not explained completely through the covariate. The idea is to define a test statistic based on a $L^2$-distance between the classical kernel intensity estimator (1) and our proposal in Borrajo et al. (2017b). Due to the lack of consistency of Diggle's proposal we have decide to do a equivalent comparison using the concept of "density of events location" of Cucala (2006) instead of using the intensities; i.e., the null hypothesis can be equivalently rewritten as $H_0 : \lambda_0(x) = \rho(Z(x))/m$, with $\lambda_0(x) = \lambda(x)/m$ and $m = \int_W \lambda(x)dx$. Hence, the test statistic is defined as:

$$T = \int_W \left(\hat{\lambda}_{0,H}(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx, \tag{10}$$

where $\hat{\lambda}_{0,H}(x) = \frac{1}{N p_H(x)} \sum_{i=1}^N K_H(x - X_i) 1_{\{N \neq 0\}}$ is the bivariate estimation for the density of events location proposed by Fuentes-Santos et al. (2015), $H$ is a bandwidth matrix, $p_H(x)$ is the edge correction term, $\hat{\rho}_{0,b}(Z(x)) = \frac{\hat{\rho_b(x)}}{N} 1_{\{N\neq 0\}}$ with $\hat{\rho}_b(x) = \sum_{i=1}^N \frac{1}{g^\star(Z(X_i))} L_b(Z(x) - Z(X_i))$, $b$ is a bandwidth parameter, $K$ and $L$ are kernel functions, $K_H(u) = |H|^{-1/2} K(H^{-1/2}u)$ and $L_b(u) = \frac{1}{b} L\left(\frac{u}{b}\right)$.

*Asymptotic properties and calibration*

Hall (1984) proposed a central limit theorem for the integrated square error of multivariate kernel density estimators, which have a similar structure to our test statistic.

Hereafter we will assume that $W = \mathbb{R}^2$ to avoid the edge effects in the theoretical developments, and we introduce some regularity conditions:

(A.6) $\int_{\mathbb{R}} L(z)dz = 1$; $\int_{\mathbb{R}} zL(z)dz = 0$ and $\mu_2(L) := \int_{\mathbb{R}} z^2 L(z)dz < \infty$.

(A.7) $\lim_{m\to\infty} b = 0$ and $\lim_{m\to\infty} \frac{A(m)}{b} = 0$, where $A(m) := \mathbb{E}\left[\frac{1}{N} 1_{\{N\neq 0\}}\right]$.

(A.8) The bandwidth matrix $H$ is symmetric and positive-definite, and such that all entries of $H$ tends to zero, and $m^{-1}|H|^{-1/2} \to 0$ as $m$ increases.

(A.9) $K$ is a continuous, symmetric, square integrable bivariate density function such that $\int_{\mathbb{R}^2} uu^T K(u)du = \mu_2(K) Id_2$ with $\mu_2(K) < \infty$.

(A.10) $Z(x)$ is a continuity point of $\rho$.

**Theorem 3** *Under conditions (A.6) to (A.10) and assuming the null hypothesis $H_0 : \lambda(x)/m = \rho(Z(x))/m \, \forall x \in W$ holds*

$$\frac{T - \mu_T}{\sigma_T} \longrightarrow N(0,1), \ \textit{where}$$

$$\mu_T = A(m)|H|^{-1/2} R(K) + \frac{1}{2}\mu_2(K) \int \lambda_0(x) tr(HD^2\lambda_0(x))dx + \frac{1}{4}\mu_2^2(K) \int tr^2(HD^2\lambda_0(x))dx,$$

$$\sigma_T^2 = A(m)|H|^{-1/2} \int\int \lambda_0^2(x)\lambda_0(y)(K \circ K)(H^{-1/2}(x-y))dxdy + 2A(m)|H|^{-1/2} R(\lambda_0)R(K),$$

*with $tr(\cdot)$ denoting the trace of a matrix.*

The proof of this result has been detailed in Borrajo, González-Manteiga, and Martínez-Miranda (2017c).

However, this asymptotic distribution may not be the best way to calibrate our test because it requires some extra estimations, and as the convergence rate may be slow it is not suitable for small patterns. Our proposal to deal with this inaccuracy is to use our bootstrap procedure previously presented in Section 3.

We have used the smooth bootstrap procedure inspired in Cao (1993) and Cowling et al. (1996) to resample under the null hypothesis, hence, let $\hat{\lambda} = \hat{\rho}_t(Z(x))$ with $t$ a pilot bandwidth. Now, conditional on $X_1, \ldots, X_N$ let $N^* \sim Pois\left(\int_W \hat{\rho}_t(Z(x))dx\right)$, generate $n^*$ a realisation of this random variable $N^*$ and then draw $X_1^*, \ldots, X_{n^*}$ by sampling randomly $n^*$ times from the distribution with density proportional to $\hat{\lambda} = \hat{\rho}_t(Z(x))$. Note that Cowling et al. (1996) remarked that the kernel and the bandwidth matrix used in the smooth bootstrap do not need to be the same as those to perform the statistic.

We have conducted a simulation study to analyse the performance of the nonparametric test introduced above. We have checked the normality of $T$ under the null hypothesis and compared the probabilities of rejecting $H_0$ provided by the asymptotic and bootstrap calibrations. We have also analysed the power of the test for different ways and degrees of departure from the null hypothesis.

We have chosen two different models constructed using real data sets to study the test under the null hypothesis; and the results show that, as expected, the asymptotic normality is not performing well for small sample sizes. Meanwhile, the bootstrap calibration, with the appropriate bandwidth (a reescaled version of the bootstrap bandwidth selector previously presented), shows a good behaviour for the two models and the different sample sizes.

The alternative hypothesis are generated based on the two models under the null, adding a parameter that allow us to control the distance to the null hypothesis and study the behaviour of the test in different situations. Here the values obtained for the power are also high. All the numbers and details of this simulations can be seen in Borrajo et al. (2017c).

## 5. APPLICATION TO A REAL DATA SET

Here we describe a data analysis to illustrate our proposals in this work. We use the same data set as Baddeley et al. (2012), which is available in Baddeley, Turner, et al. (2005). In this data set the points are the locations of gold deposits observed in a geological survey of the Murchison region of Western Australia, and the covariate is the distance from any point of the observation region to the nearest fault; the observation region is contained in a 330×400 km rectangle. In Figure 2 we can see the 255 locations (left) and the covariate information (right).

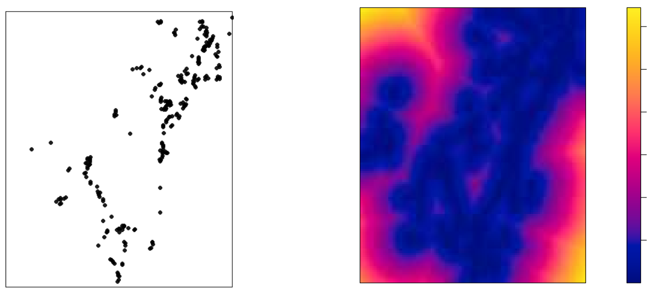
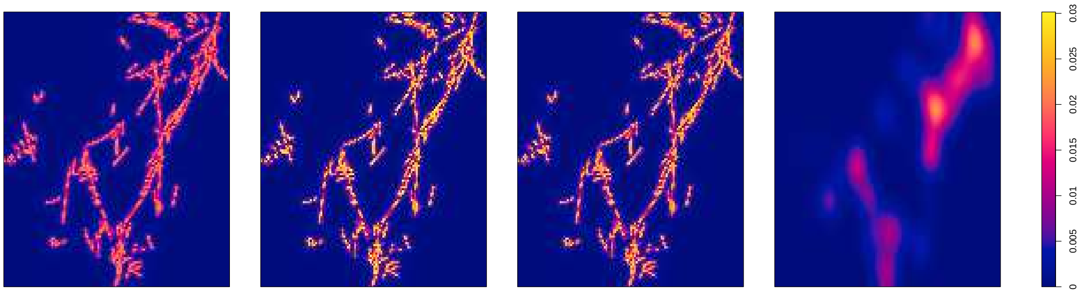



Figure 2: Location of the gold deposits (left) and distance to the faults (right).

First of all we need to compute our test statistic in order to know whether the model (2) is appropriate to describe the intensity. Actually this data set and the covariate has been already used without checking the assumption using a formal statistical test. The reason was that there were no specific methodology for it; this gap is now filled with our proposal. After computing the test we obtain a value of the statistic that allow us to accept the null hypothesis and then all the techniques presented above are applicable (p-values between 0.602 and 0.804 for a suitable range of bandwidth values).

In Figure 3 we perform intensity estimation using the covariate information and the three bandwidth selectors presented in Section 3, as well as Diggle's estimator that ignores the covariate information and uses only the information coming from the locations. We can see that the covariate is indeed useful and with it, we produce better estimations. Moreover if we compare the three bandwidth selectors, our proposals give higher intensity values than Silverman's rule used in Baddeley et al. (2012).




(a) Silverman (b) Rule-of-thumb (c) Bootstrap (d) Diggle

Figure 3: Kernel intensity estimators for the Murchison data set.

We have briefly shown here the possible applications of our methodology and the interest that it might generate in different fields, such as in forestry, where they have several meteorological and geological covariates that may be of interest in analysing point patterns of for example wildfires or species distribution. Another application to wildifires in Canada can be seen in Borrajo et al. (2017b).

## 6. CONCLUSIONS

In this manuscript we propose new methodology on the field of non-parametric inference for point processes with covariates. We have considered kernel intensity estimation and we have set up a theoretical framework that has allowed us to develop in detail the asymptotic expansions of the MSE, MISE and AMISE of our intensity estimator. Furthermore we have proposed a consistent smooth bootstrap procedure, and two new data-driven bandwidth selection methods. We also propose a new procedure to test if a given covariate is or not of interest for a point process, using the $L^2-$distance and the estimator previously developed. We detail the asymptotic normality of the test statistic and we use our bootstrap procedure to improve its calibration in practice. All the methodology is supported by extensive simulation studies that show the good behaviour of our proposals in many different situations. Finally we include an application to real data set showing the potential and the strength of the methods presented above.

## REFERENCES

Babu, G. J., & Feigelson, E. D. (1996). *Astrostatistics* (Vol. 3). CRC Press.

Baddeley, A., Chang, Y. M., Song, Y., & Turner, R. (2012). Nonparametric estimation of the dependence of a spatial point process on spatial covariates. *Statistics and Its Interface*, *5*, 221-236. doi: 10.4310/SII.2012.v5.n2.a7

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: methodology and applications with r.* CRC Press.

Baddeley, A., Turner, R., et al. (2005). Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, *12*(6), 1–42.

Borrajo, M. I., González-Manteiga, W., & Martínez-Miranda, M. D. (2017a). Bandwidth selection for kernel density estimation for length-biased data. *Journal of Nonparametric Statistics (To appear)*.

Borrajo, M. I., González-Manteiga, W., & Martínez-Miranda, M. D. (2017b). Bootrapping kernel intensity estimation for nonhomogeneous point processes depending on spatial covariates. *TEST (2nd review)*. Retrieved from `https://arxiv.org/abs/1703.03213`

Borrajo, M. I., González-Manteiga, W., & Martínez-Miranda, M. D. (2017c). *Testing covariate significante in spatial point processes first-order intensity (unpublished).*

Brooks, M. M., & 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. doi: 10.1016/0304-4149(91)90076-O

Cao, R. (1993). Bootstrapping the mean integrated squared error. *Journal of Multivariate Analysis*, *45*(1), 137–160. doi: 10.1006/jmva.1993.1030

Cowling, A., Hall, P., & 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.

Cucala, L. (2006). *Espacements bidimensionnels et données entachés d'erreurs dans l'analyse des procesus ponctuels spatiaux* (Doctoral dissertation, Université des Sciences de Toulouse I). Retrieved from `https://tel.archives-ouvertes.fr/tel-00135890/document`

Daley, D. J., & Vere-Jones, D. (1988). *An introduction to the theory of point processes.* New York: Springer Verlag.

Diggle, P., & 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. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *34*(2), 138–147. doi: 10.2307/2347366

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. doi: 10.2307/2982977

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns.* CRC Press.

Fuentes-Santos, I., González-Manteiga, W., & Mateu, J. (2015). Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial poisson point processes. *Scandinavian Journal of Statistics*, *43*(2), 416-435. doi: 10.1111/sjos.12183

Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association*, *103*(483), 1238–1247. doi: 10.1198/016214508000000526

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, *14*(1), 1–16. doi: 10.1016/0047-259X(84)90044-7

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns.* John Wiley & Sons. doi: 10.1002/9780470725160

Illian, J. B., Møller, J., & Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, *16*(3), 389–405. doi: 10.1007/s10651-007-0070-8

Illian, J. B., Sørbye, S. H., & Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested laplace approximation (inla). *The Annals of Applied Statistics*, 1499–1530.

Marron, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics*, *13*(3-4), 187–208.

Moller, J., & Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes.* CRC Press. doi: 10.1002/sim.1896

Ogata, Y., & Zhuang, J. (2006). Space–time etas models and an improved extension. *Tectonophysics*, *413*(1), 13–23. doi: 10.1016/j.tecto.2005.10.016

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. doi: 10.1016/j.jspi.2003.09.027

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. (1992). *Multivariate density estimation: Theory, practice and visualisation.* Wiley.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* CRC press.

Stoyan, D., & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, *15*(1), 61–78. doi: 10.1214/ss/1009212674

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. doi: 10.1111/j.1541-0420.2006.00667.x