Computational Statistics and Data Analysis 194 (2024) 107929

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Goodness-of-fit test for point processes first-order intensity

M.I. Borrajo [a,*], W. González-Manteiga [a], M.D. Martínez-Miranda [b]

[a] CITMAga, Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain
[b] Department of Statistics and Operations Research, Universidad de Granada, Spain

ARTICLE INFO

Keywords:
Point processes
First-order intensity
Goodness-of-fit
Covariates

ABSTRACT

Modelling the first-order intensity function is one of the main aims in point process theory. An appropriate model describes the first-order intensity as a nonparametric function of spatial covariates. A formal testing procedure is presented to assess the goodness-of-fit of this model, assuming an inhomogeneous Poisson point process. The test is based on a quadratic distance between two kernel intensity estimators. The asymptotic normality of the test statistic is proved and a bootstrap procedure to approximate its distribution is suggested. The proposal is illustrated with two applications to real data sets, and an extensive simulation study to evaluate its finite-sample performance.

1. Introduction

The understanding of spatial point processes is crucial in many different fields: ecology (Illian et al., 2009 and Law et al., 2009); epidemiology (Lawson, 2013); seismology (Ogata and Zhuang, 2006 and Schoenberg, 2011); forestry (Stoyan and Penttinen, 2000) and geology (Foxall and Baddeley, 2002). Methodological statistics have also approached this field from different perspectives, see for example Daley and Vere-Jones (1988) focused on a measure theory approach, and, Diggle (2013) and Cressie (2015) advocating for a more exploratory perspective.

A spatial point process is a stochastic process governing the location of a random number of events, $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, irregularly placed in a planar region $W \subset \mathbb{R}^2$. Throughout this paper, point processes and patterns are denoted in bold capitals, and events are denoted in bold. The spatial distribution of events in Poisson point processes, i.e., those with independent events, is determined by the first-order intensity function, see Illian et al. (2009) and Diggle (2013). This function measures the expected number of events per unit area, and is defined as follows:

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

where N is the counting measure, i.e., N is the random variable that counts the number of points lying on a given set, $\mathbb{E}$ denotes the expectation and $|\cdot|$ is the Lebesgue measure of the corresponding space, area in the planar case. Poisson point processes have been widely studied. A Poisson point process verifies, by definition, that the number of events in any bounded region of the process's underlying space follow a Poisson distribution. Moreover, considering a collection of disjoint and bounded subregions of the underlying space, the number of events of a Poisson point process in each bounded subregion is independent of all the others. This property is known as complete randomness, complete independence or independent scattering. In other words, there is a lack

* Corresponding author.
E-mail address: mariaisabel.borrajo@usc.es (M.I. Borrajo).

https://doi.org/10.1016/j.csda.2024.107929
Received 2 May 2023; Received in revised form 24 January 2024; Accepted 25 January 2024

Available online 1 February 2024

of interaction among the events, which motivates the Poisson process being sometimes called a completely random process and fully characterised by its first-order intensity function.

However, events in a generic spatial point process may not occur independently, and interactions between them are characterized through second-order properties such as the K-function, see Ripley (1977), and the second-order intensity function, $\lambda_2(x,y) = \lim_{|dx|\to 0,|dy|\to 0} \frac{\mathbb{E}[N(dx)N(dy)]}{|dx||dy|}$. The conditional intensity function

$$\lambda_x(x) = \lambda_c(x|\mathbf{y}) = \frac{\lambda_2(x,\mathbf{y})}{\lambda(\mathbf{y})}$$

determines the intensity at a point x conditional on having an event at $\mathbf{y}$, see Diggle (2013), and characterizes uniquely the distribution in any spatial point process. For spatial Poisson point processes, $\lambda_c(x,\mathbf{y}) = \lambda(x)$.

To estimate the first-order intensity function the classical approach consists of fitting a parametric function by maximum likelihood, see Waagepetersen (2007), Moller and Waagepetersen (2003) and Diggle (2013), or pseudo-likelihood procedures, see Van Lieshout (2000). It is well-known that parametric solutions may generate unreliable estimates which do not represent the underlying true model. Hence nonparametric approaches are valuable and powerful alternatives. Diggle (1985) proposed the first kernel intensity estimator which has been widely used in exploratory analysis. The main disadvantage of this estimator is its lack of consistency. To overcome this issue (Cucala, 2006, Sect. 5.3) proposed an improvement based on the relationship between intensity and density functions: both are non-negative functions and the difference relies on the intensity function not necessarily integrating the unit.

Recently, real application requirements have motivated a new scenario based on the inclusion of covariates. Considering spatially varying covariates has been a big step forward on point process theory, and it has been mostly addressed so far from a parametric perspective. See for example Waagepetersen (2007) performing inference on Neymann-Scott processes depending on a linear combination of spatial covariates; Guan and Loh (2007) estimating coefficients of covariates using the Poisson likelihood estimator, previously proposed in Schoenberg (2005); and Guan and Shen (2010) proposing a method to estimate the coefficients of the covariates improving the Poisson likelihood estimator. In the nonparametric context the inclusion of covariates has been less prolific. Guan (2008) proposes a kernel intensity estimator, assuming that the intensity function depends on some observed spatially varying covariates through an unknown continuous function. Later Baddeley et al. (2012) postulates the intensity model

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

where $Z : W \subset \mathbb{R}^2 \to \mathbb{R}$ is a spatial continuous covariate that is exactly known in every point of the region of interest W, and ρ is an unknown real function. This proposal provides intensity estimators based on local likelihood and kernels. Borrajo et al. (2020) have detailed the theoretical properties of a related kernel intensity estimator under this intensity model, including extensions to more than one spatial covariates and appropriate data-driven bandwidth selectors, which was absolutely missing so far.

Little attention has been paid to test the goodness-of-fit of intensity models with covariates. Some proposals on model selection and testing covariate significance have been made. Under a parametric approach, Yue and Loh (2015) and Thurman et al. (2015) focused on Neymann-Scott and cluster processes, respectively. A general formulation proposed by Díaz-Avalos et al. (2014) assumed that the intensity depends on a linear combination of several covariates, and they proposed a test to determine whether any of the coefficients associated to the covariates may be null. While there might exist exploratory techniques to check the adequacy of the covariate representation of the intensity (see for example Baddeley et al., 2015), to the extend of our knowledge, no formal goodness-of-fit test has been proposed in the literature. In this work we define such a test based on the comparison through a quadratic distance of the intensity model including the covariates, and the "classical" nonparametric model that only takes into account the spatial coordinates. Under the Poisson assumption, our procedure is well defined, and we are able to derive its asymptotic properties, as well as a convenient bootstrap calibration. The extension of the test to non-Poisson scenarios is also possible by specifying some more characteristics for the underlying process.

This paper is organised as follows. In Section 2 we introduce additional notation and briefly describe the nonparametric intensity estimators that we use to define the test statistic. Section 3 presents our goodness-of-fit test, providing the asymptotic distribution of the test statistic, and a bootstrap procedure to approximate its critical values in practice. The proposed methodology is applied to real data sets in Section 4. The finite-sample performance of the test is analysed in Section 5 through an extensive simulation study, where the simulated models have been defined based on the real data sets previously presented. Extensions to non-Poisson processes are discussed in Section 6, and some conclusions are drawn in Section 7.

2. First-order intensity modelling and estimation

Let $\mathbf{X}$ be a Poisson point process defined in a region $W \subset \mathbb{R}^2$, where W is assumed to have finite positive area. Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be a realisation of the process, where N is the random variable counting the number of events. Diggle (1985) proposed the first kernel intensity estimator for one-dimensional point processes, which has been easily extended to the plane as:

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

where H is a two-dimensional bandwidth matrix, $\mathbf{K}_H(x) = |H|^{-1/2}\mathbf{K}\left(H^{-1/2}x\right)$, with $\mathbf{K}$ being a bivariate kernel function, and $p_H = \int_W |H|^{-1/2}\mathbf{K}(H^{-1/2}(x-y))dy$ is the edge correction term. Here $|H|$ denotes the determinant, and $H^{-1/2}$ a power of matrix H.

This intensity estimator has been widely used during decades for exploratory analysis, but its lack of consistency has limited the inference performed with it. One approach to overcome this problem involves estimating the "density of events locations" or relative density. In this scenario, the Poisson assumption is needed to develop the asymptotic theory, and guarantee the consistency of the kernel estimator of the density of event locations. Notice that we cannot distinguish between hetereogeneity and interaction in an observed point pattern unless we have some additional information, such as a parametric model, Diggle (2013). As it has been pointed out by Fuentes-Santos et al. (2015), the common practice in the analysis of spatial point patterns is assuming that the point process is Poisson, estimating the first-order intensity, and then estimating the second-order properties to test the Poisson assumption. Therefore, the Poisson assumption is not very restrictive in practice.

Given that the number of events of an inhomogeneous Poisson point process, N, has distribution $Pois\left(\int_W \lambda(x)dx\right) \equiv Pois(m)$, Cucala (2006) defined the density of event locations, $\lambda_0(x) = \lambda(x)/m$. Using kernel estimators, this idea has been followed by Fuentes-Santos et al. (2015), and the kernel estimator of the relative density is defined:

$$\hat{\lambda}_{0,H}(x) = \frac{1}{p_H(x)N}|H|^{-1/2}\sum_{i=1}^{N}\mathbf{K}\left(H^{-1/2}(x-X_i)\right)1_{\{N\neq 0\}}. \tag{2}$$

Closely related to the above estimator, Borrajo et al. (2020) have proposed a consistent kernel intensity estimator using covariates for the multivariate model

$$\lambda(x) = \boldsymbol{\rho}(\mathbf{Z}(x)),\ x \in W \subset \mathbb{R}^2, \tag{3}$$

where $\boldsymbol{\rho}$ is an unknown p-dimensional function and $\mathbf{Z} = (Z^1, \ldots, Z^p)$ is a p-dimensional covariate. Each component, $Z^j : W \to \mathbb{R}$, is a spatial continuous covariate that is exactly known in every point of the region of interest W.

Under model (3), the grounds of Borrajo et al. (2020)'s proposal to estimate the intensity function, λ, rely on a result stating that, if X is a point process in $W \subset \mathbb{R}^2$ fulfilling equation (3) and $\mathbf{Z}$ is defined as above, then the transformed point process $\mathbf{Z}(X)$ is a point process in $\mathbb{R}^p$ with intensity $\boldsymbol{\rho}\mathbf{g}^\star$. In this context, $\mathbf{g}^\star$ is defined as $\mathbf{g}^\star(\cdot) = |W|\mathbf{g}(\cdot)$, where $\mathbf{g}$ is the first derivative of the spatial cumulative distribution function of $\mathbf{Z}$ and $\mathbf{G}(\mathbf{z}) = \frac{1}{|W|}\int_W 1_{\{\mathbf{Z}(u)\leq \mathbf{z}\}}du$. Hence, the density of events locations associated to the transformed point process is $\mathbf{f}(\cdot) = \frac{\rho(\cdot)\mathbf{g}^\star(\cdot)}{m}$ with $m = \int_{\mathbb{R}^p}\boldsymbol{\rho}(\mathbf{z})\mathbf{g}^\star(\mathbf{z})d\mathbf{z} = \int_W \lambda(x)dx$. With these definitions a kernel estimator for the density of events locations $\mathbf{f}$ is defined as follows:

$$\hat{\mathbf{f}}_B(\mathbf{z}) = \mathbf{g}^\star(\mathbf{z})\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\mathbf{g}^\star(\mathbf{Z}_i)}\mathbf{L}_B\left(\mathbf{z}-\mathbf{Z}_i\right)1_{\{N\neq 0\}}, \tag{4}$$

where $\mathbf{L}$ is a multivariate radially symmetric kernel function, and B is a p-dimensional bandwidth matrix. From (4) the kernel estimator for the intensity function λ in model (3) is given by:

$$\hat{\lambda}_B(x) = \sum_{i=1}^{N}\frac{1}{\mathbf{g}^\star(\mathbf{Z}_i)}\mathbf{L}_B(\mathbf{Z}(x)-\mathbf{Z}_i). \tag{5}$$

3. The proposed test

The first-order intensity assumption detailed in (1) has been on the grounds of several proposals, see for example Baddeley et al. (2012), Baddeley et al. (2013) or Borrajo et al. (2020). No formal testing procedure has been proposed, and only graphical diagnostics for validating the covariate effect term in the model are available so far.

In this scenario, we want to test the null hypothesis

$$H_0 : \lambda(x) = \boldsymbol{\rho}(\mathbf{Z}(x)),\ x \in W,$$

versus a general alternative in which the intensity function is not explained completely by the covariates, i.e.,

$$H_1 : \lambda(x) \neq \boldsymbol{\rho}(\mathbf{Z}(x)),\ \text{for some } x \in W.$$

For simplicity in the process' notation, and to be able to apply consistency results in Fuentes-Santos et al. (2015), our proposal is developed in $\mathbb{R}^2$. However, the ideas and formulas can be directly extended to a generic Euclidean space, $\mathbb{R}^d$, taking into account that an extra effort on proving the consistency of kernel methods in $\mathbb{R}^d$ needs to be done.

As it has been introduced in Section 1, with Poisson assumption, the process under the null hypothesis is well-defined. However, in Section 6 we detail some ideas on a more general version of our proposal in terms of the nature of the process.

3.1. Test statistic

For convenience we express the problem in terms of the relative density of event location instead of the corresponding intensity function. Notice that the null hypothesis can be equivalently rewritten as

$$H_0 : \lambda_0(x) = \boldsymbol{\rho}(\mathbf{Z}(x))/m,$$

where recall that $\lambda_0(x) = \lambda(x)/m$ is the relative density of event locations, with $m = \int_W \lambda(x)dx$, denoting the expected number of events in W.

Let us now define the quadratic distance between the two theoretical functions of interest:

$$\int_W (\lambda(x)/m - \boldsymbol{\rho}(\mathbf{Z}(x))/m)^2\, dx = \int_W \left(\lambda_0(x) - \boldsymbol{\rho}(\mathbf{Z}(x))/m\right)^2 dx.$$

From this theoretical distance we construct our test statistic by plugging-in kernel estimates of the relative densities. For $\lambda_0(x)$ we use the two-dimensional kernel estimator given in (2), while for $\boldsymbol{\rho}(\mathbf{Z}(x))$ we use the p-dimensional kernel estimator given in (5), and replace the unknown m by N. Hence, our test statistic is formally defined as follows:

$$T_n = \int_W \left(\hat{\lambda}_{0,H}(x) - \hat{\boldsymbol{\rho}}_{0,B}(\mathbf{Z}(x))\right)^2 dx, \tag{6}$$

where $\hat{\lambda}_{0,H}(x)$ has been defined in (2) and $\hat{\rho}_{0,B}(x)$ is defined from (5) as:

$$\hat{\rho}_{0,B}(\mathbf{Z}(x)) = \frac{\hat{\lambda}_B(x)}{N} 1_{\{N\neq 0\}} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\mathbf{g}^{\star}(\mathbf{Z}_i)} \mathbf{L}_B\left(\mathbf{z} - \mathbf{Z}_i\right) 1_{\{N\neq 0\}}.$$

Notice that T_n measures the discrepancy between the null and the alternative hypotheses. For big values of T_n we should reject the null hypothesis in favour of the alternative, i.e., the intensity of the Poisson process cannot be explained by the covariates information through model (3). On the other hand small values of T_n indicate that the intensity of the Poisson process may be explained by the covariates through such model.

Remark 1. An important fact in the definition of this testing procedure is that it is based on a Poisson assumption. Hence, testing this condition is essential to obtain an accurate interpretation of the result of the test, i.e., when rejecting the null hypothesis, it is crucial to determine whether the rejection is due to the failure of the Poisson assumption or the falseness of H_0 itself.

Hence, in practice, we use the tools available in the R-package *spatstat* to perform the test described in Loosmore and Ford (2006) (`dclf.test`) and Ripley (1977, 1981) (`mad.test`) for the K and L functions to verify the Poisson assumption before conducting the goodness-of-fit test.

3.2. Asymptotic properties

For spatial and spatio-temporal point processes, three asymptotic regimes can be formulated: increasing-domain, infill asymptotics, Ripley (1988), and nearly infill sampling, as it is known the combination of both of them. As the ideas in this paper are related to density estimation, the asymptotic regime considered is infill structure asymptotic framework (Diggle and Marron, 1988), which states that the expected number of events tends to infinity for a fixed bounded observation domain. For a better understanding, take into account that the mathematical background behind this asymptotic framework is a sequence of counting processes $N \equiv N_m$, where $m = \mathbb{E}[N_m]$, and it is this expected number of events the quantity that increases to infinity.

For simplicity here we develop the asymptotic theory for just one covariate ($p = 1$). The result can be directly transferred, with similar arguments, to $p > 1$, but at the expense of more complexity and cumbersome notation.

Some notation and regularity conditions need to be introduced. We use the notation $\circ$ for the convolution between two functions, $tr(\cdot)$ for the trace of a matrix, D^2 for the Hessian operator and $R(h) = \int h^2(x)dx$, for any function h. $W = \mathbb{R}^2$ is required for the consistency of the estimator $\hat{\lambda}_{0,H}$, see Fuentes-Santos et al. (2015). Additionally we assume:

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

(A.2) The bandwidth value $b \equiv b(m)$, denoted in the general scenario $p > 1$ as B, verifies the following $\lim_{m\to\infty} b = 0$ 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.3) The bandwidth matrix H is symmetric and positive-definite, such that all entries of H tend to zero and $m^{-1}|H|^{-1/2} \to 0$ as m increases.

(A.4) $\boldsymbol{K}$ is a continuous, symmetric, square integrable bivariate density function such that $\int_{\mathbb{R}^2} uu^T \boldsymbol{K}(u)du = \boldsymbol{\mu_2(K)} I_2$ with $\boldsymbol{\mu_2(K)} < \infty$ and I_2 denoting the two-dimensional identity matrix.

(A.5) $Z(x)$ is a continuity point of ρ for all $x \in W$.

The following theorem provides the normal limiting distribution of the statistic T_n. The result is based on the central limit theorem for the integrated squared error of multivariate kernel density estimators given by Hall (1984), see Appendix for the proof. In the development of this proof it is important to note that by using linearization techniques proposed in Collomb (1976), condition (A.3) implies that $A(m)|H|^{-1/2} \to 0$ as m increases:

$$\mathbb{E}\left[\frac{1}{N} 1_{\{N\neq 0\}}\right] = \mathbb{E}\left[\frac{1_{\{N\neq 0\}}}{N}\right] = \mathbb{E}\left[\frac{\phi}{\xi}\right] = \frac{\mathbb{E}[\phi]}{\mathbb{E}[\xi]} + c + c^{(\nu)} + \frac{(-1)^{\nu}\sigma^{1,\nu}}{\mathbb{E}[\xi]^{\nu}},$$

where the second and subsequent addends are smaller than the ratio of expectations. Hence, we can write that, in the limit,

$$\mathbb{E}\left[\frac{1}{N}1_{\{N\neq 0\}}\right] \simeq \frac{\mathbb{E}[\phi]}{\mathbb{E}[\xi]} = \frac{\mathbb{E}\left[1_{\{N\neq 0\}}\right]}{\mathbb{E}[N]} = \frac{\mathbb{P}(N\neq 0)}{m} = \frac{1-e^{-m}}{m} \simeq \frac{1}{m}.$$

And from this, the implication of (A.3) is straightforward. Also the Poisson assumption, implying second order independence, which is essential in the proof to compute the conditional expectations.

Theorem 1. *Under conditions (A.1) to (A.5), and under the null hypothesis, $H_0 : \lambda(x)/m = \rho(Z(x))/m$ for all $x \in W$,*

$$\frac{T_n - \mu_{T_n}}{\sigma_{T_n}} \xrightarrow{m\to\infty} N(0,1),$$

with

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

$$\sigma^2_{T_n} = A(m)|H|^{-1/2}\iint \lambda_0^2(x)\lambda_0(y)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))dxdy + 2A(m)|H|^{-1/2}R(\lambda_0)R(\boldsymbol{K}).$$

Remark 2. To extend this result to general covariate dimension $p > 1$ a modification of condition (A.2) is needed:

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

This would derive into modifying the details along the proof (Appendix A) by changing the real value bandwidths of the covariate-based estimator into matrices and determinants. However, this has no direct impact on the final result of the asymptotic distribution, because the higher order terms are those of the nonparametric estimator (2) which does not include covariates.

In practice, the asymptotic distribution given in Theorem 1 could be approximated estimating m by the observed sample size n, and $A(m)$ by $1/n$, see Cucala (2006). The slow convergence rate of the asymptotic distribution, which is detailed in the standardize factor, i.e. $A(m)|H|^{-1/2}$, and which is slower than parametric rate, indicates that this normal distribution may not be the best way to calibrate the test in practice, specially for small patterns. For small expected sample sizes, the real distance of the statistic to normality may not be small enough to provide with accurate results. Our proposal is to use bootstrap methods to approximate this asymptotic distribution. The bootstrap procedure, detailed in the following section, needs the result stated on Theorem 1 to formally prove its consistency, see González-Manteiga and Crujeiras (2013) for a extent discussion of different goodness-of-fit tests using bootstrap methods for calibration. The idea behind the proof is to use similar arguments to those of Theorem 1 using the plug-in estimators, and taking into account that the second extra step is basically averaging the resulting values from a consistent smooth bootstrap.

3.3. Bootstrap approximation to the null distribution of the test statistic

Here we describe smooth bootstrap procedures inspired in Cao (1993) and Cowling et al. (1996) to resample under the null hypothesis, using the Poisson assumption. In Section 6 we discuss possible extensions.

We consider the general multivariate case with $p \geq 1$. Under the null hypothesis the intensity can be written as a function of the p covariates, $\lambda(x) = \boldsymbol{\rho}(\mathbf{Z}(x))$. So let us consider a pilot kernel intensity estimate $\hat{\lambda}_R(x)$ as in (5), with a pilot bandwidth matrix R. With this pilot estimate and conditional on the observed pattern $\mathbf{X}_1, \ldots, \mathbf{X}_N$, we suggest first a simpler bootstrap algorithm for the calibration of the test based on simulation from the null hypothesis, and second a two-stage bootstrap algorithm motivated by the insights from Baddeley et al. (2017).

Algorithm 1 Simple smooth bootstrap procedure to compute the critical value of the test statistic.

1: Draw the number of events, n^*, from a random variable N^* having a Poisson distribution with intensity $\int_W \hat{\lambda}_R(x)dx$.
2: Draw a bootstrap sample, $\mathcal{X}^* = \{\mathbf{X}_1^*, \ldots, \mathbf{X}_{n^*}^*\}$, by sampling randomly n^* times from the distribution with density proportional to $\hat{\lambda}_R$ in (5).
3: Compute the bootstrapped test statistic, $T_{n^*}^*$, using the expression in (6), but applied to the bootstrap sample.
4: Repeat steps 1 to 3 M times leading to bootstrapped test statistics, $T_{n^*}^{*1}$, …, $T_{n^*}^{*M}$, and approximate the critical value of the test at level α by the $[(1-\alpha)M]$-th order statistic of these M values. Compute the p-value as $p = \frac{1+\sum_{i=1}^{M} 1_{\{T_n \leq T_{n^*}^{*i}\}}}{M+1}$, where T_n is the value of the test statistics from the original sample defined in (6).

Baddeley et al. (2017) discussed that classical Monte Carlo tests, as the derived from our algorithm above, are biased when the null hypothesis is composite. This can be overcome using a two-stage or double bootstrap. The motivation comes from the observation that the p-value derived from the algorithm is non-uniformly distributed when the null has to be estimated. A two-stage bootstrap will provide an adjustment of this p-value using second stage p-values (see also Davison and Hinkley, 1997). The full procedure is described below.

Algorithm 2 Two-stage smooth bootstrap procedure to compute the critical value of the test statistic.

1: Run Algorithm 1 to derive the first stage p-value p based on M bootstrap samples under the null hypothesis estimated with a pilot kernel intensity estimate $\hat{\lambda}_R(x)$. Keep the M bootstrap samples: $\mathcal{X}^{*,1}, \ldots, \mathcal{X}^{*,M}$.
2: **for** $i = 1, \ldots, M$ with each sample $\mathcal{X}^{*,i}$ **do**
3: Estimate the intensity using the same kernel estimator as in the first stage, denote it by $\hat{\lambda}_R^{*,i}$.
4: Run Algorithm 1 using $\hat{\lambda}_R^{*,i}$ as the pilot intensity and derive the corresponding p-value p_i.
5: Compute the adjusted p-value as $p_{adj} = \frac{1+\sum_{i=1}^{M} 1_{[p_i \leq p]}}{M+1}$.

The algorithm above admits some variations. First the bandwidth for the pilot intensity estimator could be chosen differently in the two stages. Second, the number of bootstrap samples simulated in the first and second stages could be also different, see discussion in Baddeley et al. (2017).

Algorithm 2 is computational much more demanding than Algorithm 1. For this reason in our simulations we restricted to the latter but computed both in the real data applications. Our simulations in Section 5 will show that the bootstrap approximation of the null distribution derived from Algorithm 1 performs quite well, providing a test with the correct empirical size, and reasonable power levels.

In practice the two algorithms require the choice of several bandwidth parameters, including the pilot bandwidth mentioned before. Optimal bandwidth selection for tests based on kernel smoothing is still an open problem. Nevertheless in practical situations is often suitable to use good automatic bandwidth selectors designed for estimation, even though they might not be theoretically optimal for testing purposes. In point processes this has been considered for example in the former paper by Fuentes-Santos et al. (2015); in the density field for testing different characteristics, see for example Silverman (1981) and Mammen et al. (1992) for multimodality tests, Hyndman and Yao (2002) for symmetry tests and in the regression context see Eubank and Hart (1992) for goodness-of-fit. To derive our test we need two bandwidth matrices H, for the intensity estimator of Fuentes-Santos et al. (2015), and B for the covariate-based intensity estimator of Borrajo et al. (2020). Following the precedent literature we use the automatic data-driven bandwidth selectors proposed for the estimators in Fuentes-Santos et al. (2015) and Borrajo et al. (2020), respectively. In the numerical analyses showed later we will see that these choices seem to be appropriate, while for the pilot bandwidth R we have evaluated the test in a suitable range of values to validate our results.

4. Data analyses

In this section we illustrate our proposal using two real data sets. The first one is the Murchison data set that has been used by A. Baddeley in his papers, books and software, see for example Baddeley et al. (2012) and Baddeley et al. (2015). The second data set consists of wildfires in Canada and it has been obtained from the Canadian Wildland Fire Information System http://cwfis.cfs.nrcan.gc.ca/home.

4.1. Murchison gold deposits

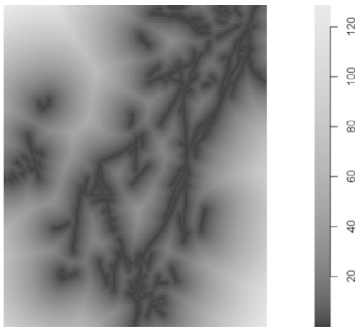
Murchison geological survey data shown in Fig. 1 (left) record the spatial locations of 255 gold deposits and the surrounding geological faults. These data came from a 330×394 km region in the Murchison area of Western Australia and they have been obtained by Watkins and Hickman (1990). As it has been remarked in Baddeley et al. (2012), at scale 1:500000, the gold deposits spatial extension is negligible and they can be considered as points without losing generality. Note also that the real gold deposits and faults are three-dimensional, while here we use a two-dimensional projection. Moreover, some geological faults may have been missed because they are not recorded by direct observation but in magnetic field surveys or geologically inferred from discontinuities in the rock sequences.

In Baddeley et al. (2012) the aim on studying these data is to "specify zones of high prospectivity to be explored for gold", so they have already assumed that the influence of the fault information is relevant, and it may actually explain the localization of gold deposits under the one-dimensional model (1). From the faults it is possible to construct a covariate by computing the distance from every point in the observation region to the nearest fault, see Fig. 1 (right). This covariate has been used to model the intensity of the process through (1). Our goal is to check whether this intensity model is indeed appropriate. As we have previously indicated, before applying our goodness-of-fit test we need to verify the inhomogeneous Poisson assumption. Hence, we have conducted the Loosmore and Ford (2006) and Ripley (1977, 1981) tests obtaining p-values of 0.12 and 0.09, respectively, and then assuming that this Murchison data set may come from an inhomogeneous Poisson point process.

To apply the goodness-of-fit test described in the previous section we need to choose several bandwidth parameters. As previously discussed we propose to use automatic data-driven bandwidth selectors for the two bandwidths required to compute the test statistic in (6): the bandwidths B, which is a scalar in this case, and a two-dimensional matrix H. We use the bandwidth selectors proposed in Borrajo et al. (2020) based on a bootstrap procedure, and Fuentes-Santos et al. (2015), to estimate B and H, respectively. Additionally for the bootstrap method we need one more scalar bandwidth, this is, the pilot bandwidth R. In this case we use different values within a suitable range to thoughtfully validate our results. The range has been chosen around 0.52 (which is the value of the bandwidth selector of Borrajo et al., 2020) taking into account the scale of the data.

Table 1 shows the obtained p-values, considering $M = 500$ replications in the bootstrap algorithm. From these results we cannot reject the null hypothesis not having enough evidence against. All the bandwidths in the considered range provide high p-values. Our conclusion is therefore that, once it is assumed that the process fulfils a Poisson condition, there is statistical evidence supporting that

(a) Murchison geological survey data, geological faults.



(b) Covariate information for the Murchison data set: distance to the nearest geological fault (in meters).

Fig. 1. Geological faults (left) and covariate constructed from them (right) used in the Murchison data set.

the geological faults are enough to explain the location of the gold deposits in the form shown in model (1). In this particular data set the Poisson assumption seems to be appropriate since the correlation in the data may be negligible, i.e., having a gold deposit at a certain location does not favour or prevent from having another gold deposit nearby.

Table 1
P-values of the test for different pilot bandwidth values R in the bootstrap algorithm (Murchison data set).

R	0.30	0.40	0.50	0.60	0.70
p-value Algorithm 1	0.94	0.89	0.80	0.70	0.60
p-value Algorithm 2	0.80	0.87	0.72	0.49	0.38

4.2. Wildfires in Canada

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 a long tradition on recording information about their wildfires; and also studies from many different perspectives have been carried out: Walter et al. (2014), Rogers et al. (2013), Di Iorio et al. (2013), Flannigan and Harrington (1988). It is well known that fire activity in Canada mostly relies on meteorological elements such as long periods without rain and high temperatures. In this context we are interested in studying the influence of these meteorological variables on the spatial distribution of wildfires.

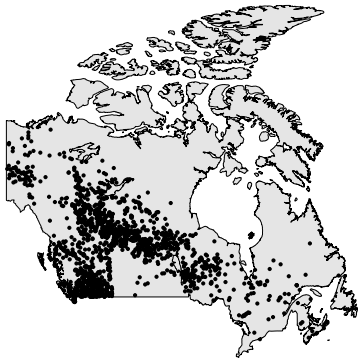


Fig. 2. Wildfires in Canada during June 2015.

The wildfire data set and also a complete meteorological information from the last decades is available at the Canadian Wildland Fire Information System website (http://cwfis.cfs.nrcan.gc.ca/home). The fire season in Canada lasts from late April until August, with a peak of activity in June and July, hence we analyse the influence of meteorological covariates on wildfires during June 2015, a total number of 1841, see Fig. 2.

In this study we focus on two covariates: temperature and precipitation, see Fig. 3 left and right panels, respectively. The reason is that the above mentioned bibliography has determined them as the main necessary factor in the ignition of a wildfire. It is important

to note that for inferential purposes we have removed two regions (Northwest Territories and Nunavut, mostly covered by ice layers) from the whole observation window (Canada). There are no fires registered on those regions and we cannot perform inference with such lack of information. We apply our proposed goodness-of-fit test to check the intensity dependence on the covariates in the form of model (3). We aim to conclude whether the temperature or/and the precipitation can explain the spatial distribution of the wildfires in Canada during June 2015. To this aim we check the goodness-of-fit of two intensity models. First the one-dimensional that uses the temperature as the only covariate, and then precipitation, in both cases $p = 1$. Then we tried a two-dimensional model using both temperature and precipitation together ($p = 2$). The inhomogeneous Poisson assumption needs to be verified before conducting the goodness-of-fit test. We use once again the proposals in Loosmore and Ford (2006) and Ripley (1977, 1981) using in the one-dimensional and the two-dimensional the corresponding intensity functions, with the effect of one or two covariates, respectively. In all the models, the tests do not reject the null hypothesis and then we can assume that the Poisson condition is fulfilled.

(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).

Fig. 3. Covariates used in the intensity estimation of the wildfires in Canada during June 2015.

Regarding to the first model, with just the temperature as covariate, we compute the test statistic using the same kind of bandwidth choices as for the Murchison data described before. Again we consider different bandwidth values in a suitable range for the pilot bandwidth in the bootstrap calibration. For these data it seems that a suitable range would be around 0.35, which is the value of the automatic bandwidth selector. In this case we get the same results for all the bandwidths in the range and both algorithms, rejecting the null hypothesis based on p-values that are always close to zero. Similar results are obtained when using precipitation as covariate. Our conclusion is that the temperature or precipitation cannot explain on its own the wildfires spatial distribution.

We move then to the second model and include the two covariates, temperature and precipitation. In this two-dimensional context the bandwidth selectors of Borrajo et al. (2020) do not apply so we have used the plug-in rule for multivariate densities developed in Chacón and Duong (2010), which has been implemented in the function `Hpi` of the R-package *ks*. The result from our test is similar, and the null hypothesis is again rejected based on a p-value close to zero with both algorithms.

This leads us to conclude that, even the covariates are likely to have an effect on the distribution of the wildfires in Canada, this is not enough to explain the first-order intensity by themselves in the way shown in (3). In this case the idea of being able to explain wildfires occurrences only with temperature and precipitation information may be too ambitious, and other meteorological covariates (or some indexes gathering several variables) should be used to analyse this process.

Following this idea, we have used the Canadian Forest Fire Weather Index (FWI), which consists of six components that account for the effects of fuel moisture and weather conditions on fire behaviour. Calculation of the components is based on consecutive daily observations of temperature, relative humidity, wind speed, and 24-hour precipitation, for more details and information on this see https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi. Using only this index as covariate is not neither enough to explain wildfire behaviour, possibly more complex models should be accounted for.

5. Simulation study

This section is devoted to analyse the performance of our proposal through Monte Carlo simulations. Under a Poisson assumption we define two intensity models based on the two real data sets previously presented and analysed. To reduce the burden of computation we only consider a one-dimensional covariate ($p = 1$). The first model is based on the wildfires data set and the theoretical intensity, represented in Fig. 4, has been constructed by computing the kernel intensity estimator of Borrajo et al. (2020) with the temperature as the covariate. The second model has been constructed in a similar way but using the Murchison data set, with the covariate being the distance to the nearest geological fault. The intensity function in this case is represented in Fig. 5.

Fig. 4. Theoretical intensity function for the first model analysed in the simulation study, obtained by applying a kernel intensity estimator to the Canada wildfire data set.

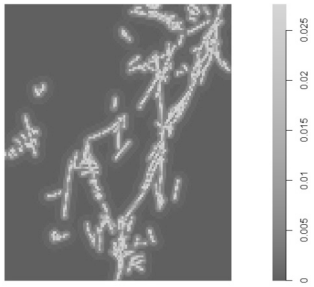


Fig. 5. Theoretical intensity function for the second model analysed in the simulation study, obtained by applying a kernel intensity estimator to the Murchison data set.

To evaluate the empirical level of our test we generate samples under the null hypothesis. To evaluate the power of the test we simulate samples under the alternative. We define alternatives by including a multiplicative deviation from the null models as follows:

$$\lambda(x) = \lambda_{\text{null}}(x)t(x).$$

Here λ_{null} denotes the intensity under the null and t is a perturbation function that depends on a parameter that controls the discrepancy from the null. We have considered this function as a diagonal band that crosses the observation region nullifying the extension out of it, with a smooth change. The band is wider or thinner depending on a one-dimensional parameter: when its wide enough it covers the whole observation region, so we approach the null hypothesis, and as it becomes thinner the model gets away from it.

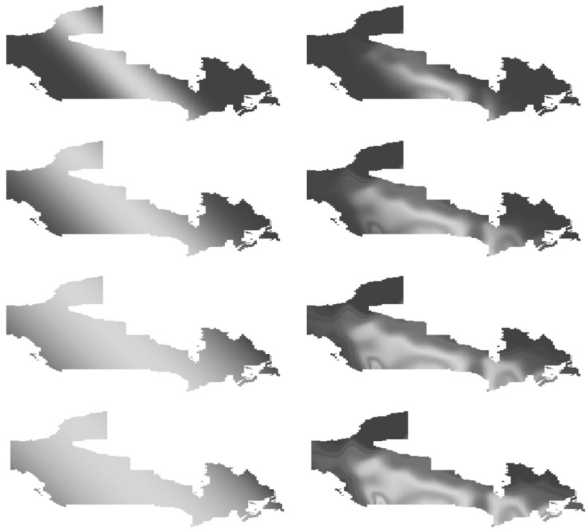


Fig. 6. Representation of the t_C functions (first column) and the resulting intensity (second column) for the four values of the parameter $d_C = 6, 12, 20, 30$ for the Canada model.

Since the observation regions of each model are different we have defined two different functions but following the same idea. In both cases the diagonal band is based on a univariate normal density, $\phi(\cdot, \mu, \sigma)$, depending on a parameter: d_C for the Canada data model and d_M for the Murchison data model. This gives alternatives for the first case with function $t_C(u, v) = \phi(u, 15 - v - v_{0C}, d_C)$, where $v_{0C} = 60.40$ is the middle point of the y-axis in the observation region, and d_C taking values $6, 12, 20$ and 30. In Fig. 6 we represent, for each value of the parameter d_C, both the t_C function (first column) and the final intensity function (second column).

For the Murchison model, the function is defined as $t_M(u, v) = \phi(u - u_{0M}, u - v_{0M}, d_M)$, where $u_{0M} = 517.69$ and $v_{0M} = 6900.61$ are the middle point of the x and y-axis respectively, and d_M is the parameter taking values $10, 20, 40$ and 60. We have decided to rotate the diagonal band due to the distribution of the data and the covariate information, but it could have been done in the other way. Take also into account that due to this rotation, even a thin band collects a lot of information from the covariate which may cause smaller rejection proportions. Indeed, we have performed the same simulations with a rotated band and we obtained higher rejection proportions. In Fig. 7 we represent the t_M functions (first row) as well as the final intensities (second row). Remark that we have not included the scales in Fig. 6 and 7 because the values of the intensity depend on the expected sample size, so they will change in each of the situations considered in the study.

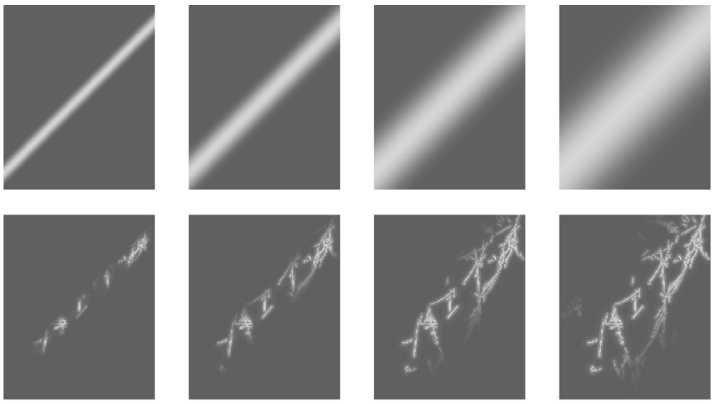


Fig. 7. Representation of the t_M functions (first row) and the resulting intensity (second row) for the four values of the parameter $d_M = 10, 20, 40, 60$ for the Murchison model.

We have computed our test in the above described scenarios, considering 5000 Monte Carlo replications. To choose the required bandwidths we have followed the same considerations as for the real data in the previous section. To approximate the test statistic distribution under the null we have used 200 bootstrap samples, this has shown to be enough for the considered situations.

Table 2 and Table 3 show the results of the test for the Canada model and the Murchinson model, respectively. These tables provide the observed rejection proportions for different situations that go from the null hypothesis, to evaluate the level of the test (case of $d_\bullet = \infty$), to the furthest away situation from it to evaluate the power of the test ($d_C = 6$ and $d_M = 10$, respectively). From these results we can see that the empirical level is accurate. Under the null hypothesis ($d_\bullet = \infty$) the observed rejections are about 5% that is the nominal level in this study. Under the alternative ($d_\bullet < \infty$) the rejection proportions give us the power of the test. The proportions seem to be higher for the first model where, even for $d_C = 20$ that is a situation near to the null, the power values are high for medium and large expected sample sizes. In the second model, the values do not reach those levels. As we have explained above, keeping in mind the spatial distribution of the gold deposits, even for thin bands we are gathering a lot of information from the covariate and hence we are not really as far from the null hypothesis as we might think.

Table 2
Rejection proportions for Canada model, with different values of the parameter controlling the discrepancy from the null hypothesis, d_C, and four expected sample sizes.

	$d_C = 6$	$d_C = 12$	$d_C = 20$	$d_C = 30$	$d_C = \infty$
$m = 50$	1	0.6852	0.1454	0.0722	0.0480
$m = 100$	1	0.9308	0.2232	0.0826	0.0502
$m = 200$	1	0.9986	0.4076	0.1060	0.0516
$m = 500$	1	1	0.8266	0.1744	0.0520

Table 3
Rejection proportions for Murchison model, with different values of the parameter controlling the discrepancy from the null hypothesis, d_M, and four expected sample sizes.

	$d_M = 10$	$d_M = 20$	$d_M = 40$	$d_M = 60$	$d_M = \infty$
$m = 50$	0.9584	0.3858	0.1048	0.0636	0.049
$m = 100$	0.9926	0.4774	0.1049	0.0680	0.0496
$m = 200$	0.9998	0.6404	0.1136	0.0610	0.0504
$m = 500$	1	0.9376	0.1727	0.0633	0.0492

We have also represented the rejection proportions shown in the tables on a graph depending on the distance to the null hypothesis and the sample size, see Fig. 8. From the graph we can quickly confirm that, under the null hypothesis, the proportion of rejections is around the nominal level in all scenarios. This proportion increases as we get away from the null and, as expected, it increases faster with the growth of the sample size. We can also visualize here the fact that for the Murchison model the rejection proportions are generally lower than for the Canada model, due to the relationship between the orientation of the band and the distribution of the data that we have previously remarked.

In summary our test shows good level and power results for the considered scenarios. Notice that we have not included any other method to compare with in this study because, to the extent of our knowledge, there is no existing formal statistical test addressing the same problem.

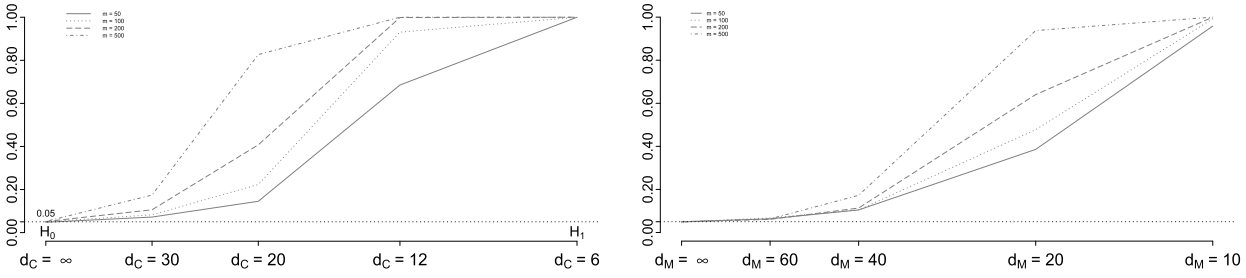


Fig. 8. Representation of the rejection proportions for the different simulated scenarios of the Canada model (left) and Murchison model (right).

6. Further extension

6.1. Non-Poisson point process

In the previous sections we rely on a Poisson assumption that may be too restrictive in some situations. There are practical situations in which Poisson models do not apply because of interactions between points. One example is the analysis of earthquakes where the point process cannot be characterised by its first-order intensity and the interaction, generally divided into attraction (producing clusters) and inhibition (originating regular patterns), needs to be described.

Our test can be extended to a more general class of point processes if more information is available. As it has been pointed out in Diggle et al. (2007) we can not simultaneously estimate the first and second order structure of an inhomogeneous point process from a single realization without any additional information, such as covariates or the specification of a parametric model. The reason behind this issue is that while the covariate-based intensity estimator chases the first order intensity, the one based only on locations chases the conditional first order intensity, and then we can not directly use their discrepancy to detect the presence of covariates in our model.

Our idea to solve this drawback, which is nor trivial nor straightforward, is to have some previous knowledge on the type of process and then use, for example, a parametric estimator replacing that in (2) in our statistic (which could still be defined based on an quadratic distance), and replicate our calibration process for this new version. The latter is actually the main barrier to overcome in this general scenario, which would require an adaptation of the bootstrap procedure.

One last remark regarding to this extension is about theoretical results. If the Poisson assumption does not hold theoretical developments as those derived in this paper are harder to achieve, or even intractable, depending on the process. As we have described above, this does not diminish the practical applicability and Monte Carlo techniques, see for example Díaz-Avalos et al. (2014), can be applied to simulate and approximate the test distribution.

6.2. Weighted L^2 distance

Although the framework we have set up in this paper is based on measuring the discrepancy between two estimators by using a classical L^2 distance, a weighted version of it could be implemented:

$$T_{\omega,n} = \int_W \left(\hat{\lambda}_{0,H}(x) - \hat{\rho}_{0,B}(\mathbf{Z}(x))\right)^2 \omega(x)dx.$$

As far as this weight function is deterministic, all the theoretical developments may be replicated for the weighted version without increasing their complexity. This weight function may allow us to deal with scenarios where all regions are not of the same importance, and may also be used to avoid possible edge effects by limiting the importance of those areas on the test statistic. However, there is a drawback including this weight function, which is a possible increase of the computational cost. The reason is that the region where this weight function is defined, i.e. its domain, might be very irregular, so its definition and the corresponding grid might require developing efficient algorithms.

7. Conclusions

In this paper we have defined a goodness-of-fit test to check the adequacy of an existing model in the field of point processes, using nonparametric techniques. Under such model the first-order intensity of a process can be expressed as a function of spatial covariates. We have completely described the test for inhomogeneous Poisson point processes, using the theoretical framework developed in Borrajo et al. (2020). It has allowed us to detail the asymptotic distribution of the test statistic, as well as a bootstrap method used to accomplish its calibration in practice. Our proposal has been illustrated using two real data examples, where one or two covariates have been used to describe the intensity of the underlying point process. We have defined theoretical intensity models based on these two real situations and used them to carry out a simulation study. This has helped to better understand the performance of our test in finite samples while keeping close to the real practice. The simulation results are quite satisfactory and show good values in terms of level and power in several scenarios. To the extent of our knowledge our goodness-of-fit test does not have any competitors to be compared with.

For real situations where the Poisson assumption does not hold we have briefly described an extension of our test. Owning some extra information on the underlying process, would let us appropriately estimate its first-order intensity without covariates. Then an adaptation of our test statistic could be implemented.

The methods proposed in this paper can be extended to address appealing research questions such as testing the significance of some covariates, like in the paper of Díaz-Avalos et al. (2014), with the advantage of being formulated in a completely nonparametric way. Moreover the goodness-of-fit of the log-linear covariate representation of the first-order intensity assumed in Díaz-Avalos et al. (2014) could be checked using the ideas presented in our paper. These are left for future research.

Acknowledgements

The authors are grateful for the very constructive comments from the three anonymous reviewers and Associate Editor which helped to improve this manuscript. The authors acknowledge the support through Grant PID2020-116587GB-I00 funded by

MCIN/AEI/10.13039/501100011033. The authors also acknowledge the Canadian Wildland Fire Information System for their activity in recording and freely providing part of the real data used in this paper.

Appendix A. Proof of Theorem 1

Along this proof we derive the mean and variance of the statistic T_n, as well as assuring its asymptotic normality. We first rewrite the statistic as:

$$\begin{aligned} T_n &= \int_W \left(\hat{\lambda}_{0,H}(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx \\ &= \int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx + \int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx + 2\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)\left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right) dx. \end{aligned} \tag{A.1}$$

Applying mean and variance operators to (A.1), we obtain

$$\mathbb{E}\left[T_n\right] = E\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx\right] + \mathbb{E}\left[\int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx\right] + 2\mathbb{E}\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)\left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right) dx\right] \tag{A.2}$$

and

$$\begin{aligned} \mathbb{V}ar\left[T_n\right] &= \mathbb{V}ar\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx\right] + \mathbb{V}ar\left[\int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx\right] \\ &+ 4\mathbb{V}ar\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)\left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right) dx\right] \\ &+ 2\mathbb{C}ov\left(\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx, \int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx\right) \\ &+ 4\mathbb{C}ov\left(\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx, \int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)\left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right) dx\right) \\ &+ 4\mathbb{C}ov\left(\int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx, \int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)\left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right) dx\right). \end{aligned} \tag{A.3}$$

The idea is to compute mean and variance of each of the addends, and then deal with the covariances. Let us start with the first addend in (A.1):

$$\begin{aligned} \mathbb{E}\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx\right] &= \int \mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\right] dx - 2\int \lambda_0(x)\mathbb{E}\left[\hat{\lambda}_{0,H}(x)\right] dx + R(\lambda_0) \\ &= A(m)|H|^{-1/2}R(K) + \frac{1}{4}\mu_2^2(K)\int tr^2(H D^2\lambda_0(x))dx + o\left(A(m)|H|^{-1/2}\right) + o(tr(H)) \end{aligned} \tag{A.4}$$

and

$$\begin{aligned} \mathbb{V}ar\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx\right] &= \iint \Big(\mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\hat{\lambda}^2_{0,H}(y)\right] + 2\lambda_0^2(y)\mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\right] - 4\lambda_0(y)\mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\hat{\lambda}_{0,H}(y)\right] \\ &\quad -4\lambda_0^2(x)\lambda_0(y)\mathbb{E}\left[\hat{\lambda}_{0,H}(y)\right] + 4\lambda_0(x)\lambda_0(y)\mathbb{E}\left[\hat{\lambda}_{0,H}(x)\hat{\lambda}_{0,H}(y)\right] + \lambda_0^2(x)\lambda_0^2(y)\Big) dxdy - (A.4)^2 \\ &= R(K)o(A(m)|H|^{-1/2}) - 2R(K)R(\lambda_0)o(A(m)|H|^{-1/2}) - 6R(\lambda_0)o(tr(H)) \end{aligned} \tag{A.5}$$

For the second addend we also need the relationship established in Theorem A.1 and Theorem A.2 in Borrajo et al. (2020). We finally obtain that both, mean and variance, are negligible in comparison with the terms obtained for the first addend, in particular we have found that

$$\mathbb{E}\left[\int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx\right] = o(A(m)) \qquad \text{and} \qquad \mathbb{V}ar\left[\int_W \left(\lambda_0(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx\right] = O(A(m)),$$

which are both smaller than $o(A(m)|H|^{-1/2} + tr(H))$ corresponding to the first addend.

The mean and variance of the third addend are also simpler because several terms are negligible with respect to the main term in the first addend's variance. We obtain:

$$2\mathbb{E}\left[\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)\left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)dx\right]=\mu_2(K)\int \lambda_0(x)tr(H\,D^2\lambda_0(x))dx+o(tr(H))$$

and

$$\begin{aligned}&\mathbb{V}ar\left[2\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)\left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)dx\right]\\&=4A(m)|H|^{-1/2}\iint \lambda_0^2(x)\lambda_0(y)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))dxdy+4\iint \lambda_0(x)\lambda_0(y)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(A(m)|H|^{-1/2})dxdy\\&+o(A(m)).\end{aligned}$$

Focusing now on the covariances terms, we obtain that

$$\mathbb{C}ov\left[\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)^2dx,\int_W \left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)^2dx\right]=A(m)|H|^{-1/2}R(\lambda_0)R(K)+R(\lambda_0)R(K)o(A(m)|H|^{-1/2})+o(A(m)),\tag{A.6}$$

and the other two terms,

$$\mathbb{C}ov\left[\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)^2dx,\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)\left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)dx\right]\tag{A.7}$$

and

$$\mathbb{C}ov\left[\int_W \left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)^2dx,\int_W \left(\hat{\lambda}_{0,H}(x)-\lambda_0(x)\right)\left(\lambda_0(x)-\hat{\rho}_{0,b}(Z(x))\right)dx\right],\tag{A.8}$$

are smaller than the main term of the first addend's variance.

In all the results regarding expectations, variances and covariances, the following computations are needed:

$$\mathbb{E}\left[K_H(x-X_1)\right]=\int K_H(x-u)\lambda_0(u)du=\lambda_0(x)+\frac{1}{2}\mu_2(K)tr(H\,D^2\lambda_0(x))+o(tr(H)),$$

$$\mathbb{E}\left[K_H^2(x-X_1)\right]=\int K_H^2(x-u)\lambda_0(u)du=|H|^{-1/2}\lambda_0(x)R(K)+o(|H|^{-1/2}),$$

$$\mathbb{E}\left[K_H(x-X_1)K_H(y-X_1)\right]=\int K_H(x-u)K_H(y-u)\lambda_0(u)du=|H|^{-1/2}\lambda_0(x)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))+o(|H|^{-1/2}),$$

$$\mathbb{E}\left[K_H^2(x-X_1)K_H(y-X_1)\right]=\int K_H^2(x-u)K_H(y-u)\lambda_0(u)du=|H|^{-1}\lambda_0(x)(\boldsymbol{K}^2\circ\boldsymbol{K})(H^{-1/2}(x-y))+o(|H|^{-1}),$$

$$\mathbb{E}\left[K_H^2(x-X_1)K_H^2(y-X_1)\right]=\int K_H^2(x-u)K_H^2(y-u)\lambda_0(u)du=|H|^{-3/2}\lambda_0(x)(\boldsymbol{K}^2\circ\boldsymbol{K}^2)(H^{-1/2}(x-y))+o(|H|^{-3/2}),$$

$$\mathbb{E}\left[\hat{\lambda}_{0,H}(x)\right]=(1-e^{-m})E\left[K_H(x-X_1)\right]=\lambda_0(x)+\frac{1}{2}\mu_2(K)tr(H\,D^2\lambda_0(x))+o(tr(H)),$$

$$\begin{aligned}\mathbb{E}\left[\hat{\lambda}_{0,H}^2(x)\right]&=A(m)\mathbb{E}\left[K_H^2(x-X_1)\right]+(1-e^{-m}-A(m))\mathbb{E}^2\left[K_H(x-X_1)\right]\\&=A(m)|H|^{-1/2}\lambda_0(x)R(K)+\lambda_0^2(x)+\mu_2(K)\lambda_0(x)tr(H\,D^2\lambda_0(x))\\&+\frac{1}{4}\mu_2^2(K)tr^2(H\,D^2\lambda_0(x))+A(m)\lambda_0^2(x)+A(m)\mu_2(K)\lambda_0(x)tr(H\,D^2\lambda_0(x))\\&+\frac{1}{4}A(m)\mu_2^2(K)tr^2(H\,D^2\lambda_0(x))+2\lambda_0(x)o(tr(H))+R(K)o(A(m)|H|^{-1/2})\\&+\mu_2(K)tr(H\,D^2\lambda_0(x))o(tr(H))+o(tr^2(H))+o(A(m))\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\hat{\lambda}_{0,H}(x)\hat{\lambda}_{0,H}(y)\right]&=A(m)\mathbb{E}\left[K_H(x-X_1)K_H(y-X_1)\right]+(1-e^{-m}-A(m))\mathbb{E}\left[K_H(x-X_1)\right]\mathbb{E}\left[K_H(y-X_1)\right]\\&=A(m)|H|^{-1/2}\lambda_0(x)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))+(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(A(m)|H|^{-1/2})+\lambda_0(x)\lambda_0(y)\\&+\frac{1}{2}\lambda_0(x)\mu_2(K)tr(H\,D^2\lambda_0(y))\lambda_0(x)o(tr(H))+\frac{1}{2}\lambda_0(y)\mu_2(K)tr(H\,D^2\lambda_0(x))\\&+\frac{1}{4}\mu_2^2(K)tr(H\,D^2\lambda_0(x))tr(H\,D^2\lambda_0(y))\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\mu_2(K)tr(H\,D^2\lambda_0(x))o(tr(H))+\lambda_0(y)o(tr(H))+\frac{1}{2}\mu_2(K)tr(H\,D^2\lambda_0(y))o(tr(H))+A(m)\lambda_0(x)\lambda_0(y)\\
&+\frac{1}{2}A(m)\lambda_0(x)\mu_2(K)tr(H\,D^2\lambda_0(y))+A(m)\lambda_0(x)o(tr(H))+\frac{1}{2}\lambda_0(y)\mu_2(K)tr(H\,D^2\lambda_0(x))\\
&+\frac{1}{4}A(m)\mu_2^2(K)tr(H\,D^2\lambda_0(x))tr(H\,D^2\lambda_0(y))+o(A(m)tr(H)),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\hat{\lambda}_{0,H}(y)\right] &= B(m)\mathbb{E}\left[K_H^2(x-X_1)K_H(y-X_1)\right]+A(m)\mathbb{E}\left[K_H^2(x-X_1)\right]\mathbb{E}\left[K_H(y-X_1)\right]\\
&+2A(m)\mathbb{E}\left[K_H(x-X_1)K_H(y-X_1)\right]\mathbb{E}\left[K_H(x-X_1)\right]+(1-e^{-m})\mathbb{E}^2\left[K_H(x-X_1)\right]\mathbb{E}\left[K_H(y-X_1)\right]\\
&=A(m)|H|^{-1/2}\lambda_0(x)\lambda_0(y)R(K)+\frac{1}{2}A(m)|H|^{-1/2}\lambda_0(x)\mu_2(K)R(K)tr(H\,D^2\lambda_0(x))\\
&+\lambda_0(x)R(K)o(A(m)|H|^{-1/2}tr(H))+\lambda_0(y)R(K)o(A(m)|H|^{-1/2})\\
&+\frac{1}{2}R(K)\mu_2(K)tr(H\,D^2\lambda_0(y))o(A(m)|H|^{-1/2})\\
&+R(K)o(A(m)|H|^{-1/2}tr(H))+2A(m)|H|^{-1/2}\lambda_0^2(x)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))\\
&+A(m)|H|^{-1/2}\mu_2(K)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))\lambda_0(x)tr(H\,D^2\lambda_0(x))\\
&+2(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(A(m)|H|^{-1/2}tr(H))+2\lambda_0(x)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(A(m)|H|^{-1/2})\\
&+\mu_2(K)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))tr(H\,D^2\lambda_0(x))o(A(m)|H|^{-1/2})+o(A(m)|H|^{-1/2}tr(H))\\
&+\lambda_0^2(x)\lambda_0(y)+\frac{1}{2}\mu_2(K)\lambda_0^2(x)tr(H\,D^2\lambda_0(y))+\lambda_0^2(x)o(tr(H))+\mu_2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(x))\\
&+\frac{1}{2}\mu_2^2(K)\lambda_0(x)tr(H\,D^2\lambda_0(x))tr(H\,D^2\lambda_0(y))+\mu_2(K)\lambda_0(x)tr(H\,D^2\lambda_0(x))o(tr(H))+2\lambda_0(x)\lambda_0(y)p(tr(H))\\
&+\mu_2(K)\lambda_0(x)tr(H\,D^2\lambda_0(y))o(tr(H))+2\lambda_0(x)o(tr^2(H))+\lambda_0(y)o(tr^2(H))x+\frac{1}{2}\mu_2(K)tr(H\,D^2\lambda_0(y))o(tr^2(H))\\
&+o(tr^3(H)),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\hat{\lambda}^2_{0,H}(x)\hat{\lambda}^2_{0,H}(y)\right] &= C(m)\mathbb{E}\left[K_H^2(x-X_1)K_H^2(y-X_1)\right]+2B(m)\mathbb{E}\left[K_H^2(x-X_1)K_H(y-X_1)\right]\mathbb{E}\left[K_H(y-X_1)\right]\\
&+2B(m)\mathbb{E}\left[K_H(x-X_1)K_H^2(y-X_1)\right]\mathbb{E}\left[K_H(x-X_1)\right]+B(m)\mathbb{E}\left[K_H^2(x-X_1)\right]\mathbb{E}\left[K_H^2(y-X_1)\right]\\
&+2B(m)\mathbb{E}^2\left[K_H(x-X_1)K_H(y-X_1)\right]+A(m)\mathbb{E}\left[K_H^2(x-X_1)\right]\mathbb{E}^2\left[K_H(y-X_1)\right]\\
&+4A(m)\mathbb{E}\left[K_H(x-X_1)K_H(y-X_1)\right]\mathbb{E}\left[K_H(x-X_1)\right]\mathbb{E}\left[K_H(y-X_1)\right]\\
&+A(m)\mathbb{E}\left[K_H^2(y-X_1)\right]\mathbb{E}^2\left[K_H(x-X_1)\right]\\
&+(1-e^{-m})\mathbb{E}^2\left[K_H(x-X_1)\right]\mathbb{E}^2\left[K_H(y-X_1)\right]\\
&=A(m)|H|^{-1/2}\lambda_0(x)\lambda_0^2(y)R(K)+A(m)|H|^{-1/2}R(K)\mu_2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(y))\\
&+R(K)\lambda_0^2(y)o(A(m)|H|^{-1/2})+R(K)\mu_2(K)\lambda_0(y)tr(H\,D^2\lambda_0(y))o(A(m)|H|^{-1/2})\\
&+4A(m)|H|^{-1/2}\lambda_0^2(x)\lambda_0(y)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))\\
&+2A(m)|H|^{-1/2}\lambda_0^2(x)\mu_2(K)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))tr(H\,D^2\lambda_0(y))\\
&+4A(m)\lambda_0^2(x)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(|H|^{-1/2}tr(H))\\
&+2A(m)|H|^{-1/2}\lambda_0(x)\lambda_0(y)\mu_2(K)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))tr(H\,D^2\lambda_0(x))\\
&+4\lambda_0(x)\lambda_0(y)(\boldsymbol{K}\circ\boldsymbol{K})(H^{-1/2}(x-y))o(A(m)|H|^{-1/2})+o(A(m)|H|^{-1/2}tr(H))+A(m)|H|^{-1/2}\lambda_0^2(x)\lambda_0(y)R(K)\\
&+A(m)|H|^{-1/2}R(K)\mu_2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(x))+R(K)\lambda_0^2(x)o(A(m)|H|^{-1/2})\\
&+R(K)\mu_2(K)\lambda_0(x)tr(H\,D^2\lambda_0(x))o(A(m)|H|^{-1/2})+\lambda_0^2(x)\lambda_0^2(y)\\
&+\lambda_0^2(x)\lambda_0(y)\mu_2(K)tr(H\,D^2\lambda_0(y))2\lambda_0^2(x)\lambda_0(y)o(tr(H))+\frac{1}{4}\lambda_0^2(x)\mu_2^2(K)tr^2(H\,D^2\lambda_0(y))\\
&+\lambda_0^2(x)\mu_2(K)tr(H\,D^2\lambda_0(y))o(tr(H))+\lambda_0^2(y)\lambda_0(x)\mu_2(K)tr(H\,D^2\lambda_0(x))\\
&+\mu_2^2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(x))tr(H\,D^2\lambda_0(y))+2\mu_2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(x))o(tr(H))\\
&+2\lambda_0(x)\lambda_0^2(y)o(tr(H))+2\mu_2(K)\lambda_0(x)\lambda_0(y)tr(H\,D^2\lambda_0(y))o(tr(H))+\frac{1}{4}\mu_2^2(K)\lambda_0^2(y)tr^2(H\,D^2\lambda_0(x))\\
&+\lambda_0^2(y)\mu_2(K)tr(H\,D^2\lambda_0(x))o(tr(H))+o(A(m))+o(tr^2(H)),
\end{aligned}
$$

with $B(m)=\mathbb{E}\left[\frac{1}{N^2}1_{\{N\neq 0\}}\right]$ and $C(m)=\mathbb{E}\left[\frac{1}{N^3}1_{\{N\neq 0\}}\right]$.

The last step in this proof is the asymptotic normality. Our test statistic can be expanded and written as:

$$T_n = \int_W \left(\hat{\lambda}_{0,H}(x) - \hat{\rho}_{0,b}(Z(x))\right)^2 dx = \frac{1}{N^2}\sum_{i=1}^{N}\int K_H(x-X_i)dx + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\int K_H(x-X_i)K_H(x-X_j)dx$$
$$+ \frac{1}{N^2}\sum_{i=1}^{N}\int \frac{1}{g^\star(Z(X_i))}L_b(Z(x)-Z(X_i))dx + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\int \frac{1}{g^\star(Z(X_i))}\frac{1}{g^\star(Z(X_j))}L_b(x-X_i)L_b(x-X_j)dx$$
$$- \frac{2}{N^2}\sum_{i=1}^{N}\int \frac{1}{g^\star(Z(X_i))}K_H(x-X_i)L_b(Z(x)-Z(X_i))dx - \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\int \frac{1}{g^\star(Z(X_j))}K_H(x-X_i)L_b(Z(x)-Z(X_j))dx \tag{A.9}$$

where each of the addends is a U-statistic on a Poisson point process, remarking that the sums do not allow duplicated points in the same expression. Moreover, every of the addends is absolutely convergent in the sense defined by Reitzner and Schulte (2013), hence following their Theorem 4.7 we can assure the normality of each term. Then the normality of our statistic with the mean and variance detailed in the main body of Theorem 1 is proved.

References

Baddeley, A., Chang, Y.M., Song, Y., Turner, R., 2012. Nonparametric estimation of the dependence of a spatial point process on spatial covariates. Stat. Interface 5, 221–236.

Baddeley, A., Chang, Y.M., Song, Y., Turner, R., 2013. Residual diagnostics for covariate effects in spatial point process models. J. Comput. Graph. Stat. 22 (4), 886–905.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. CRC Press.

Baddeley, A., Hardegen, A., Lawrence, T., Milne, R.K., Nair, G., Rakshit, S., 2017. On two-stage Monte Carlo tests of composite hypotheses. Comput. Stat. Data Anal. 114, 75–87.

Borrajo, M.I., González-Manteiga, W., Martínez-Miranda, M.D., 2020. Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates. Comput. Stat. Data Anal. 144.

Cao, R., 1993. Bootstrapping the mean integrated squared error. J. Multivar. Anal. 45 (1), 137–160.

Chacón, J.E., Duong, T., 2010. Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. Test 19 (2), 375–398.

Cowling, A., Hall, P., Phillips, M.J., 1996. Bootstrap confidence regions for the intensity of a Poisson point process. J. Am. Stat. Assoc. 91 (436), 1516–1524.

Cressie, N., 2015. Statistics for Spatial Data, revised edition. John Wiley & Sons.

Cucala, L., 2006. Espacements bidimensionnels et données entachés d'erreurs dans l'analyse des procesus ponctuels spatiaux. PhD thesis. Université des Sciences de Toulouse I.

Collomb, G., 1976. Estimation Non Parametrique de la Regression par la méthode du noyau. Ph.D. dissertation. Universit Paul Sabatier de Toulouse.

Daley, D.J., Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer Verlag.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK.

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., Mateu, J., 2014. Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. Stoch. Environ. Res. Risk Assess. 28 (3), 593–609.

Diggle, P., 1985. A kernel method for smoothing point process data. J. R. Stat. Soc., Ser. C, Appl. Stat. 34 (2), 138–147.

Diggle, P.J., Gómez-Rubio, V., Brown, P.E., et al., 2007. Second-order analysis of inhomogeneous spatial point processes using case-control data. Biometrics 63 (2), 550–557.

Diggle, P., Marron, J.S., 1988. Equivalence of smoothing parameter selectors in density and intensity estimation. J. Am. Stat. Assoc. 83, 793–800.

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press.

Eubank, R.L., Hart, J.D., 1992. Testing goodness-of-fit in regression via order selection criteria. Ann. Stat. 20 (3), 1412–1425.

Flannigan, M.D., Harrington, J.B., 1988. A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). J. Appl. Meteorol. 27 (4), 441–452.

Foxall, R., Baddeley, A., 2002. Nonparametric measures of association between a spatial point process and a random set, with geological applications. J. R. Stat. Soc., Ser. C, Appl. Stat. 51 (2), 165–182.

Fuentes-Santos, I., González-Manteiga, W., Mateu, J., 2015. Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial Poisson point processes. Scand. J. Stat. 43 (2), 416–435.

González-Manteiga, W., Crujeiras, R.M., 2013. An updated review of Goodness-of-Fit tests for regression models. Test 22, 361–411.

Guan, Y., 2008. On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. J. Am. Stat. Assoc. 103 (483), 1238–1247.

Guan, Y., Loh, J.M., 2007. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. J. Am. Stat. Assoc. 102 (480), 1377–1386.

Guan, Y., Shen, Y., 2010. A weighted estimating equation approach for inhomogeneous spatial point processes. Biometrika 97 (4), 867–880.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. J. Multivar. Anal. 14 (1), 1–16.

Hyndman, R.J., Yao, Q., 2002. Nonparametric estimation and symmetry tests for conditional density functions. J. Nonparametr. Stat. 14 (3), 259–278.

Illian, J.B., Møller, J., Waagepetersen, R.P., 2009. Hierarchical spatial point process analysis for a plant community with high biodiversity. Environ. Ecol. Stat. 16 (3), 389–405.

Law, R., Illian, J., Burslem, D.F., Gratzer, G., Gunatilleke, C., Gunatilleke, I., 2009. Ecological information from spatial patterns of plants: insights from point process theory. J. Ecol. 97 (4), 616–628.

Lawson, A.B., 2013. Statistical Methods in Spatial Epidemiology. John Wiley & Sons.

Loosmore, N.B., Ford, E.D., 2006. Statistical inference using the G or K point pattern spatial statistics. Ecology 87, 1925–1931.

Mammen, E., Marron, J.S., Fisher, N.I., 1992. Some asymptotics for multimodality tests based on kernel density estimates. Probab. Theory Relat. Fields 91 (1), 115–132.

Moller, J., Waagepetersen, R.P., 2003. Statistical Inference and Simulation for Spatial Point Processes. CRC Press.

Ogata, Y., Zhuang, J., 2006. Space–time ETAS models and an improved extension. Tectonophysics 413 (1), 13–23.
Reitzner, M., Schulte, M., 2013. Central limit theorems for u-statistics of Poisson point processes. Ann. Probab. 41 (6), 3879–3909.
Ripley, B.D., 1977. Modelling spatial patterns (with discussion). J. R. Stat. Soc., Ser. B 39, 172–212.
Ripley, B.D., 1981. Spatial Statistics. John Wiley and Sons.
Ripley, B.D., 1988. Statistical Inference for Spatial Processes. Cambridge University Press.
Rogers, B.M., Randerson, J.T., Bonan, G.B., 2013. High-latitude cooling associated with landscape changes from North American boreal forest fires. Biogeosciences 10 (2), 699–718.
Schoenberg, F.P., 2005. Consistent parametric estimation of the intensity of a spatial–temporal point process. J. Stat. Plan. Inference 128 (1), 79–93.
Schoenberg, F.P., 2011. Multidimensional residual analysis of point process models for earthquake occurrences. J. Am. Stat. Assoc. 98, 789–795.
Silverman, B.W., 1981. Using kernel density estimates to investigate multimodality. J. R. Stat. Soc., Ser. B, Methodol. 43 (1), 97–99.
Stoyan, D., Penttinen, A., 2000. Recent applications of point process methods in forestry statistics. Stat. Sci. 15 (1), 61–78.
Thurman, A.L., Fu, R., Guan, Y., Zhu, J., 2015. Regularized estimating equations for model selection of clustered spatial point processes. Stat. Sin. 25 (1), 173–188.
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., Vogel, B., 2014. Influence of 2010 Canadian forest fires on cloud formation on the regional scale. In: AGU Fall Meeting Abstracts.
Watkins, K.P., Hickman, A.H., 1990. Geological Evolution and Mineralization of the Murchison Province, Western Australia, vol. 1. Department of Mines, Western Australia.
Yue, Y.R., Loh, J.M., 2015. Variable selection for inhomogeneous spatial point process models. Can. J. Stat. 43 (2), 288–305.