

Accepted Manuscript

Stationarity for spatial point processes

Tonglin Zhang, Jorge Mateu

PII: S0047-259X(18)30175-1

DOI: <https://doi.org/10.1016/j.jmva.2018.11.001>

Reference: YJMVA 4435

To appear in: *Journal of Multivariate Analysis*

Received date: 31 March 2018

Please cite this article as: T. Zhang and J. Mateu, Stationarity for spatial point processes, *Journal of Multivariate Analysis* (2018), <https://doi.org/10.1016/j.jmva.2018.11.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Substationarity for spatial point processes

Tonglin Zhang^{a,*}, Jorge Mateu^b

^a*Department of Statistics, Purdue University, 250 North University Street, West Lafayette, IN 47907, USA*

^b*Departament de Matemàtiques, Universitat Jaume I, Campus Riu Sec, 12071 Castellón, Spain*

Abstract

This article aims to introduce the concept of substationarity for spatial point processes (SPPs). Substationarity is a new concept that has never been studied in the literature. Substationarity means that the distribution of an SPP can only be invariant under location shifts within a linear subspace of the domain. This notion lies theoretically between stationarity and nonstationarity. To formally propose the approach, the article provides the definition of substationarity and estimation of the first-order intensity function, including the subspace. As this may be unknown, we recommend using a parametric method to estimate the linear subspace and a nonparametric one to estimate the first-order intensity function given the linear subspace. It is thus a semiparametric approach. The simulation study shows that both the estimators of the linear subspace and the first-order intensity function are reliable. In an application to a Canadian forest wildfire data set, the article concludes that substationarity of wildfire occurrences may be assumed along the longitude, indicating that latitude is a more important factor than longitude in Canadian forest wildfire studies.

Keywords: Intensity functions, Kernel methods, Nonstationarity, Semiparametric estimation, Spatial point processes, Substationarity.

AMS subject classifications: 62M30, 62G05

1. Introduction

The goal of this article is to introduce and develop the concept of substationarity for spatial point processes (SPPs). Substationarity is a new concept that has not been studied in the literature of SPPs. It can theoretically bridge stationarity and nonstationarity, which are two well-known concepts in the literature of spatial statistics. Substationarity means that the distribution of an SPP is only invariant under any location shift within a linear subspace of the domain. Stationarity means that the distribution is invariant under any location shift within the entire domain. If a stationary process is also invariant under rotations, then it is isotropic. Nonstationarity is the complementary concept of stationarity. It means that the distribution of the SPP can be affected by at least one location shift in the domain. If an SPP is substationary, then its distribution may still be affected by a location shift outside the linear subspace. Therefore, the intersection of substationarity and nonstationarity is not empty.

The idea of the research is motivated by recent work on typical events in natural hazards. According to its scientific definition, a natural hazard is a naturally occurring event that might have a negative effect on human or the environment. Natural hazards include wildfires, tornados, and earthquakes. In our work on forest wildfires, we identified an inhomogeneous wildfire pattern in the forests of the province of Alberta, Canada [44]. The proportion of large wildfires in the north was higher than that in the south, but the frequency of wildfires in the south was higher than in the north. This is consistent with a previous finding that wildfire activities in Canadian boreal forests are less affected by their longitude [27], which means that substationarity might hold along the longitude.

Statistical approaches to SPPs are important in many scientific disciplines such as forestry [37], epidemiology [10], wildfires [46, 48], or earthquakes [32, 47]. In statistics, an SPP is treated as a pattern of random points developed in a Euclidean space. The number of points within a bounded subset of the Euclidean space is finite. Point distributions and dependence structures are modeled by intensity functions [9]. The simplifying assumptions of stationarity and isotropy

*Corresponding author. Email address: tlzhang@purdue.edu(T. Zhang).

have been developed to make the analysis convenient. Various well-known tools have been proposed. Examples include the K -function [35], the L -function [3], and the pair correlation function [38]. As stationarity is an important assumption, additional methods have been proposed to evaluate it [6, 18, 45].

Because of concerns around the stationarity assumption, recent research often models SPPs under nonstationarity [31]. An important concept called the second-order intensity-reweighted stationarity (SOIRS) has also been proposed [1]. This concept is powerful in the joint analysis of the first- and second-order intensity functions under nonstationarity. With the aid of SOIRS, various methods for nonstationary contexts have been proposed [11, 19, 20, 22, 40]. SOIRS only specifies the relationship between the first- and second-order intensity functions. It makes no assumption related to substationarity, implying that statistical approaches to substationarity can be combined with SOIRS.

The purpose of this article is to develop a formal statistical approach to substationarity in SPPs, including the concept of substationarity and corresponding estimation methods. Since the linear subspace may be unknown, estimation of the subspace must also be involved. We propose to estimate the subspace via a parametric method and intensity functions given the subspace by a nonparametric method. Therefore, we classify our estimation as a semiparametric approach. The nonparametric component provides the intensity function given the subspace and the parametric component supplies the linear subspace. We evaluate properties of our approach by simulations and applications. In simulations, we evaluate the performance of the estimator of the linear subspace and the first-order intensity function by studying their mean square error (MSE) and mean integrated square error (MISE) values, respectively. In applications, we implement our approach to a boreal forest wildfire data set. We conclude that estimation under substationarity can provide more precise and reliable results than under nonstationarity.

To the best of our knowledge, this article is the first one to discuss formally the concept of substationarity for SPPs. The term substationarity has appeared previously in the literature [12] but it referred to a completely different concept. Previous concepts of spatial or temporal stationarity in the context of spatio-temporal data can be treated as examples of substationarity but the subspace in these examples is completely known; see p. 308 in [7]. Therefore, one does not need to estimate the subspace under spatial or temporal stationarity. Besides, the impact of the second-order properties is important in our work, but it is not taken into account in these examples. Our interest is to propose the concept not only for the first-order but also for higher-order intensity functions. As this scenario has not been previously studied, it is important to have a formal statistical definition of substationarity at the beginning. Although many research problems can be specified, we only focus on estimation of the first-order intensity functions under substationarity, which includes estimation of the subspace. Many nonparametric or semiparametric methods can be adopted, but we only study the kernel method since it is convenient.

The article is organized as follows. In Section 2, we review the method of SPPs. In Section 3, we provide the definition of substationarity, including evaluations of its theoretical properties. In Section 4, we propose a semiparametric approach to estimate the first-order intensity function including the subspace. In Section 5, we evaluate the performance of our estimator by simulations. In Section 6, we apply our approach to the Alberta forest wildfire data. The paper ends with some discussion in Section 7.

2. Spatial point processes

Theoretically, an SPP $\mathcal{N}(\mathcal{S})$ on \mathcal{S} is composed of random points in a measurable subset $\mathcal{S} \subseteq \mathbb{R}^d$. It is treated as the restriction of \mathcal{N} , the SPP on the entire \mathbb{R}^d , with points in \mathcal{S} only. Thus, points of \mathcal{N} in \mathcal{S}^c (the complementary set of \mathcal{S}) are unobserved. Let \mathcal{B} and $\mathcal{A} \subseteq \mathcal{B}(A)$ be the collections of Borel sets of \mathbb{R}^d and a measurable $A \subseteq \mathbb{R}^d$, respectively. Let $N(A)$ and N be the numbers of points in A and \mathbb{R}^d , respectively. Then, $N(A)$ is finite if A is bounded and $\Pr\{N(A) = 0\} = 1$ for any $A \in \mathcal{B}(\mathbb{R}^d)$ satisfying $|A| = 0$, where $|\cdot|$ is the Lebesgue measure on \mathbb{R}^d .

Let $\lambda_k(\mathbf{s}_1, \dots, \mathbf{s}_k)$ be the k th-order intensity function of \mathcal{N} . An SPP \mathcal{N} is k th-order stationary if

$$\lambda_\ell(\mathbf{s}_1, \dots, \mathbf{s}_\ell) = \lambda(\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_\ell + \mathbf{h}) \quad (1)$$

for any $\mathbf{h} \in \mathbb{R}^d$ and $\ell \leq k$. It is strongly stationary if (1) holds for any $\ell \in \mathbb{N} = \{1, 2, \dots\}$. We say that $\mathcal{N}(\mathcal{S})$ is k th-order stationary and strongly stationary, respectively, if it can be derived by restricting a k th-order stationary or a strongly stationary \mathcal{N} on \mathcal{S} . The mean structure of \mathcal{N} is $\mu(A) = E\{N(A)\} = \int_A \lambda(\mathbf{s}) d\mathbf{s}$, where $\lambda(\mathbf{s}) = \lambda_1(\mathbf{s})$ is its first-order intensity function. Let the covariance function of \mathcal{N} be $\Gamma(\mathbf{s}_1, \mathbf{s}_2) = \{g(\mathbf{s}_1, \mathbf{s}_2) - 1\}\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2) + \lambda(\mathbf{s}_1)\delta_{\mathbf{s}_1, \mathbf{s}_2}(\mathbf{s}_2, \mathbf{s}_2)$, where

$g(\mathbf{s}_1, \mathbf{s}_2) = \lambda_2(\mathbf{s}_1, \mathbf{s}_2) / \{\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)\}$ is the pair correlation function and $\delta_{\mathbf{s},\mathbf{s}}$ is the point measure at $(\mathbf{s}, \mathbf{s}) \in \mathbb{R}^d \times \mathbb{R}^d$. Then, the covariance structure of N is

$$\text{cov}\{N(A_1), N(A_2)\} = \int_{A_1} \int_{A_2} \Gamma(\mathbf{s}_1, \mathbf{s}_2) d\mathbf{s}_2 d\mathbf{s}_1 = \int_{A_1} \int_{A_2} \{g(\mathbf{s}_1, \mathbf{s}_2) - 1\} \lambda(\mathbf{s}_1) \lambda(\mathbf{s}_2) d\mathbf{s}_2 d\mathbf{s}_1 + \mu(A_1 \cap A_2).$$

If $g(\mathbf{s}_1, \mathbf{s}_2)$ only depends on $\mathbf{s}_1 - \mathbf{s}_2$ or $\|\mathbf{s}_1 - \mathbf{s}_2\|$ such that it can be expressed as $g(\mathbf{s}_1 - \mathbf{s}_2)$ or $g(\|\mathbf{s}_1 - \mathbf{s}_2\|)$, then N is called a second-order intensity-reweighted stationary (SOIRS) or a second-order intensity-reweighted isotropic (SOIRI) SPP. SOIRS and SOIRI are important concepts for nonstationary SPPs, as they can model the first- and second-order intensity functions together [1].

If N is first-order stationary, then $\lambda(\mathbf{s}) = c$ and $\mu(A) = c|A|$ for some $c > 0$. If N is second-order stationary, then $\lambda(\mathbf{s}) = c$, $\mu(A) = c|A|$, $g(\mathbf{s}_1, \mathbf{s}_2) = g(\mathbf{s}_1 - \mathbf{s}_2)$,

$$\text{cov}\{N(A_1), N(A_2)\} = c^2 \int_{A_1} \int_{A_2} \{g(\mathbf{s}_1 - \mathbf{s}_2) - 1\} d\mathbf{s}_2 d\mathbf{s}_1 + c|A_1 \cap A_2|,$$

and

$$\text{var}\{N(A)\} = c^2 \int_A \int_A \{g(\mathbf{s}_1 - \mathbf{s}_2) - 1\} d\mathbf{s}_2 d\mathbf{s}_1 + c|A|.$$

If N is Poisson, then $g(\mathbf{s}_1, \mathbf{s}_2) = 1$, $\text{cov}\{N(A_1), N(A_2)\} = \mu(A_1 \cap A_2)$, and $\text{var}\{N(A)\} = E\{N(A)\}$ for any bounded $A, A_1, A_2, \in \mathcal{B}(\mathbb{R}^d)$. Thus, only the mean structure is important for Poisson SPPs.

3. Substationarity

The main purpose of this section is to provide a formal definition of substationarity and its properties. As substationarity is a new concept which has not been studied in the literature of SPPs before, it is also important to provide an asymptotic theory under substationarity. The theory is needed in the evaluation of theoretical properties of estimation in Section 4.

Definition 1. We say that N is k th-order substationary in a linear subspace $\mathcal{L} \subseteq \mathbb{R}^d$ if (1) holds for any $\mathbf{h} \in \mathcal{L}$ and $\ell \leq k$. We say that N is strongly substationary in \mathcal{L} if N is k th-order substationary in \mathcal{L} for any $k \in \mathbb{N}$. For any $S \subseteq \mathbb{R}^d$, we say that $N(S)$ is k th-order substationary or strongly substationary in \mathcal{L} , or $\mathcal{L} \cap S$ equivalently, if $N(S)$ can be restricted by a k th-order substationary or strongly substationary N in \mathcal{L} on S .

Obviously, if N is k th-order substationary and its k th-order intensity function almost surely exists, then (1) holds almost surely with respect to the Lebesgue measure of \mathbb{R}^d for any $\mathbf{h} \in \mathcal{L}$ and distinct $\mathbf{s}_1, \dots, \mathbf{s}_\ell \in \mathbb{R}^d$ with $\ell \leq k$. If N is k th-order substationary in \mathcal{L} , then it is also k th-order substationary in any linear subspace $\mathcal{L}' \subseteq \mathcal{L}$. Therefore, the linear subspace \mathcal{L} in Definition 1 is generally not unique.

Definition 2. We say that N is k th-order intrinsically substationary or intrinsically strongly substationary in \mathcal{L} if it is substationary or strongly substationary in \mathcal{L} but not in any linear subspace \mathcal{L}' of \mathbb{R}^d satisfying $\mathcal{L} \subseteq \mathcal{L}'$ but $\mathcal{L} \neq \mathcal{L}'$. We say that $N(S)$ is k th-order intrinsically substationary or intrinsically strongly substationary in \mathcal{L} , or $\mathcal{L} \cap S$ equivalently, if it can be restricted by a k th-order intrinsically substationary or intrinsically strongly substationary in \mathcal{L} on S .

If N is substationary in both \mathcal{L}_1 and \mathcal{L}_2 , then N is also substationary in $\text{span}(\mathcal{L}_1, \mathcal{L}_2)$. Thus, the linear subspace \mathcal{L} in Definition 2 is unique. A k th-order intrinsically substationary N in \mathcal{L} is k th-order stationary if and only if $\mathcal{L} = \mathbb{R}^d$. If N is intrinsically substationary in \mathcal{L} , then it is substationary in any linear subspace \mathcal{L}' of \mathcal{L} but not in any linear subspace \mathcal{L}' of \mathbb{R}^d strictly covering \mathcal{L} .

The concept can be extended to the second-order intensity-reweighted substationarity models, which can be treated as an extension of SOIRS models. Note that SOIRS satisfies $g(\mathbf{s}_1, \mathbf{s}_2) = g(\mathbf{s}_1 + \mathbf{h}, \mathbf{s}_2 + \mathbf{h})$ for any $\mathbf{h} \in \mathbb{R}^d$. Based on the idea of Definition 1 and 2, we can define the second-order intensity-reweighted substationarity models by assuming that the condition only holds for any $\mathbf{h} \in \mathcal{L}$. Therefore, it is weaker than SOIRS.

The first-order property under substationarity can be easily understood. Let N be first-order substationary in \mathcal{L} and r be the dimension of \mathcal{L} . Then $\mu(A) = \mu(A + \mathbf{h})$ for any $\mathbf{h} \in \mathcal{L}$. The statement holds in a more general case.

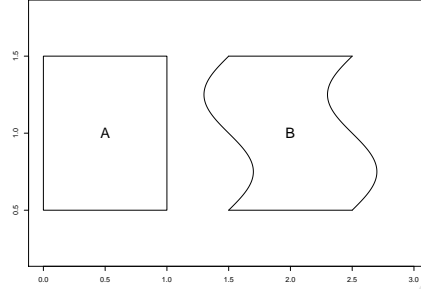


Figure 1: Equality of expected counts in two subsets under substationarity along the horizontal axis.

Suppose that \mathcal{N} is substationary in the horizontal axis of \mathbb{R}^2 (i.e., $d = 2$, such that $\mathcal{L} = \{(x, 0) : x \in \mathbb{R}\}$). Then, the first-order intensity of \mathcal{L} only depends on the vertical value of the point, indicating that we can express $\lambda(\mathbf{s}) = \lambda(y)$ for any $\mathbf{s} = (x, y) \in \mathbb{R}^2$. Let ν_r be the Lebesgue measure on \mathbb{R}^r . For any $A \subseteq \mathbb{R}^2$, we have $\mu(A) = \int_{-\infty}^{\infty} \lambda(y) \nu_1(A_y) dy$, where $A_y = \{x : \mathbf{s} = (x, y) \in A\}$. For any measurable bounded $A, B \subseteq \mathbb{R}^2$, we may still have $\mu(A) = \mu(B)$ even if $B \neq A + \mathbf{h}$ for any $\mathbf{h} \in \mathcal{L}$ (e.g., the case displayed in Figure 1). However, this method cannot be used to study the second-order property. The second-order property is important in asymptotic properties of our estimation method if \mathcal{N} is not Poisson. This is related to the general conclusion given by Theorem 3 and Corollary 1 below.

Theorem 1. (First-order property). *Let \mathcal{N} be substationary in $\mathcal{L} \subseteq \mathbb{R}^d$. For any measurable bounded sets $A, B \in \mathbb{R}^d$, if $\nu_r(A_{\mathbf{v}}) = \nu_r(B_{\mathbf{v}})$ almost surely for any $\mathbf{v} \in \mathbb{R}^d$, where $A_{\mathbf{v}} = \{\mathbf{s} \in A : \mathbf{s} - \mathbf{v} \in \mathcal{L}\}$, then $E\{N(A)\} = E\{N(B)\}$.*

Proof. Let $\mathbf{u}_1, \dots, \mathbf{u}_d$ be a set of orthogonal bases of \mathbb{R}^d with the previous r vectors forming the orthogonal bases of \mathcal{L} . Let $\mathcal{L}^\perp = \{\mathbf{v} \in \mathbb{R}^d : \mathbf{v} = x_{r+1}\mathbf{u}_{r+1} + \dots + x_d\mathbf{u}_d \text{ for some } x_{r+1}, \dots, x_d \in \mathbb{R}\}$ be the orthogonal space of \mathcal{L} . Let $\mathbf{s}_{\mathcal{L}}$ and $\mathbf{s}_{\mathcal{L}^\perp}$ be the orthogonal projections of \mathbf{s} on \mathcal{L} and \mathcal{L}^\perp , respectively. The first-order intensity function of \mathcal{N} is $\lambda(\mathbf{s}) = \lambda(\mathbf{s}_{\mathcal{L}})$ for any $\mathbf{s} \in A$. Then, $\mu(A) = \int_{\mathbf{s} \in A} \lambda(\mathbf{s}) d\mathbf{s} = \int_{\mathcal{L}^\perp} \lambda(\mathbf{s}_{\mathcal{L}}) \nu_r(A_{\mathbf{s}_{\mathcal{L}}}) d\mathbf{s}_{\mathcal{L}} = \int_{\mathbf{s} \in B} \lambda(\mathbf{s}) d\mathbf{s} = \mu(B)$. \square

Theorem 1 can be used to study the relationship between the expected numbers of counts between two regions. As higher-order intensity functions are involved, it is convenient for us to study asymptotics of the joint distribution. In the literature, the first step is to derive the asymptotics for Poisson SPPs [26, 44]. The asymptotics for other SPPs is evaluated under weak dependence given by strong mixing condition [36]. Following this idea, we study asymptotic properties of Poisson SPPs first.

Let $A_{z,\mathcal{L}} = \{\mathbf{v} + z\mathbf{u} : \mathbf{v} + \mathbf{u} \in A, \mathbf{v} \in \mathcal{L}^\perp, \mathbf{u} \in \mathcal{L}\}$ and $A_{\mathbf{v},z,\mathcal{L}} = \{\mathbf{s} \in A_{z,\mathcal{L}} : \mathbf{s} - \mathbf{v} \in \mathcal{L}\}$ for any $A \in \mathcal{B}(\mathbb{R}^d)$, where \mathcal{L} is a linear subspace of \mathbb{R}^d . Then, $A_{\mathbf{v},z,\mathcal{L}} = \{\mathbf{v} + z\mathbf{u} : \mathbf{v} + \mathbf{u} \in A, \mathbf{u} \in \mathcal{L}\}$ for given $\mathbf{v} \in \mathcal{L}^\perp$ and $z \in \mathbb{R}$, implying that $\nu_r(A_{\mathbf{v},z,\mathcal{L}}) = z^r \nu_r(A_{\mathbf{v},1,\mathcal{L}})$. If \mathcal{N} is substationary in \mathcal{L} and A is bounded, then $\mu(A_{z,\mathcal{L}}) = \int_{\mathbf{s} \in A_{z,\mathcal{L}}} \lambda(\mathbf{s}) d\mathbf{s} = z^r \int_{\mathcal{L}^\perp} \lambda(\mathbf{s}_{\mathcal{L}}) \nu_r(A_{\mathbf{s}_{\mathcal{L}},1,\mathcal{L}}) d\mathbf{s}_{\mathcal{L}} = z^r \mu(A)$.

Assume that \mathcal{N} is Poisson. Then, $\text{var}\{N(A_{z,\mathcal{L}})\} = \mu(A_{z,\mathcal{L}}) = z^r \mu(A)$ and $M_{z,\mathcal{L}}(A) = z^{-r/2}\{N(A_{z,\mathcal{L}}) - \mu(A_{z,\mathcal{L}})\} \rightsquigarrow \mathcal{N}\{0, \mu(A)\}$ as $z \rightarrow \infty$. Let \mathcal{A} be a collection of Borel sets of \mathbb{R}^d . Let $\mathcal{A}_{z,\mathcal{L}}, N(\mathcal{A}_{z,\mathcal{L}})$, and $\mu(\mathcal{A}_{z,\mathcal{L}})$ be vectors composed of $A_{z,\mathcal{L}}, N(A_{z,\mathcal{L}})$, and $\mu(A_{z,\mathcal{L}})$ for all $A \in \mathcal{A}$, respectively. If \mathcal{A} is a finite collection of disjoint subsets such that it can be expressed as $\mathcal{A} = \{A_1, \dots, A_m\}$ with disjoint A_1, \dots, A_m , then

$$M_{z,\mathcal{L}}(\mathcal{A}) \rightsquigarrow \mathcal{N}[0, \text{diag}\{\mu(\mathcal{A})\}], \quad (2)$$

where $M_{z,\mathcal{L}}(\mathcal{A})$ is the vector composed of $M_{z,\mathcal{L}}(A)$ for all $A \in \mathcal{A}$.

For any $V \in \mathcal{B}(\mathcal{L}^\perp)$, let $A_{\mathbf{t},V} = (0, t_1\mathbf{u}_1] \times \dots \times (0, t_r\mathbf{u}_r] \times V$, where $t_i > 0$, $(0, t_i\mathbf{u}_i] = \{\mathbf{s} = x\mathbf{u}_i : 0 < x \leq t_i\}$, and $\mathbf{u}_1, \dots, \mathbf{u}_r$ form a set of orthogonal bases of \mathcal{L} . If \mathcal{N} is Poisson, then

$$\begin{pmatrix} M_{z,\mathcal{L}}(A_{\mathbf{t},V}) \\ M_{z,\mathcal{L}}(A_{\mathbf{t}',V}) \end{pmatrix} \rightsquigarrow \mathcal{N} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mu(A_{\mathbf{t},V}) & \mu(A_{\mathbf{t} \wedge \mathbf{t}',V}) \\ \mu(A_{\mathbf{t} \wedge \mathbf{t}',V}) & \mu(A_{\mathbf{t}',V}) \end{pmatrix} \right\}, \quad (3)$$

as $z \rightarrow \infty$. The finite-dimensional Central Limit Theorem of $N(\mathcal{A}_{z,\mathcal{L}})$ can be derived by (2) and (3), but it is not enough for us to study properties of the estimator of the first-order intensity proposed in the next section. To study its properties, we need the functional Central Limit Theorem of $M_{z,\mathcal{L}}(\mathcal{A})$ when \mathcal{A} contains an infinitely number of measurable subsets of \mathbb{R}^d . A typical method to show the functional Central Limit Theorem is to combine the finite-dimensional asymptotics with the tightness [42]. A typical method to prove the tightness is the evaluation of the bracketing entropy number, which is used in the following theorem.

Theorem 2. *Let \mathcal{N} be a Poisson substationary SPP in \mathcal{L} . If $\mathcal{A}_V = \{A_{\mathbf{t},V} : \mathbf{t} = (t_1, \dots, t_r) \in [0, \infty)^r\}$ for some $V \in \mathcal{B}(\mathcal{L}^\perp)$, then $M_{z,\mathcal{L}}(\mathcal{A}_V)$ weakly converges to a mean 0 Gaussian random field on $[0, \infty)^r$ with the covariance structure given by the right-hand side of (3).*

Proof. We use the standard empirical process approach. Let $\mathcal{A}_{V,\mathbf{a}} = \{A_{\mathbf{t},V} : \mathbf{t} = (t_1, \dots, t_r)^\top \in [0, a_1] \times \dots \times [0, a_r]\}$ for any $\mathbf{a} = (a_1, \dots, a_r)^\top \in (0, \infty)^r$. Let $F(\mathbf{t}) = \mu(A_{\mathbf{t},V})/\mu(A_{\mathbf{a},V})$ for any $\mathbf{t} \leq \mathbf{a}$. Then F is an r -dimensional marginal uniformly distributed CDF on the σ -field generated by $\mathcal{A}_{V,\mathbf{a}}$. Let F_i be the i th CDF of F . For any $\epsilon \in (0, 1)$, there exists an integer J such that $r/\epsilon^2 \leq J \leq r/\epsilon^2 + 1$. Let $x_{ij} = ja^{-1}/(J+1)$ for all $j \in \{0, 1, \dots, J+1\}$. Then, $\epsilon^2/(\epsilon^2 + r) \leq F_i(x_{i(j+1)}) - F_i(x_{ij}) \leq \epsilon^2/r$.

Let $X_\epsilon = \{\mathbf{x} = (x_1, \dots, x_r) : x_i = x_{ij} \text{ for some } j \in \{0, 1, \dots, J+1\}\}$. Then, $\#X_\epsilon = (J+2)^r \leq \{(r+3)/\epsilon^2\}^r$. For any $g_{\mathbf{x}} \in \mathcal{G} = \{I_{\mathbf{x}} : \mathbf{x} \in [0, a_1] \times \dots \times [0, a_r]\}$, we can find $\mathbf{x}', \mathbf{x}'' \in X_\epsilon$ such that $\mathbf{x}' \leq \mathbf{y} \leq \mathbf{x}''$ but there is no $\mathbf{x}^* \in X_\epsilon$ satisfying $x'_i < x_i^* < x''_i$ for some $i \in \{1, \dots, r\}$, where x_i, x_i^* , and x''_i are the i th components of \mathbf{x}, \mathbf{x}^* , and \mathbf{x}'' , respectively. Then, $g_{\mathbf{x}'} \leq g_{\mathbf{x}} \leq g_{\mathbf{x}''}$ and

$$\|g_{\mathbf{x}''} - g_{\mathbf{x}'}\|_F^2 \leq \sum_{i=1}^r \{F_i(x''_i) - F_i(x'_i)\} \leq \epsilon^2.$$

Since

$$\int_0^1 \ln^{1/2}(\#X_\epsilon) d\epsilon \leq \int_0^1 [r \ln(r+3) + 2 \ln \epsilon]^{1/2} d\epsilon < \infty,$$

\mathcal{G} is F -Donsker; see p. 270 in [39], implying that the conclusion holds in $[0, a_1] \times \dots \times [0, a_r]$ for any $\mathbf{a} \in (0, \infty)^r$. The final conclusion is drawn by letting $a_i \rightarrow \infty$ for all $i \in \{1, \dots, r\}$. \square

We derive the functional Central Limit Theorem of $M_{z,\mathcal{L}}(\mathcal{A})$ for Poisson SPPs by Theorem 2. We now turn our attention to the functional Central Limit Theorem for other types of SPPs. A critical issue is to account for dependence. It requires us to study second-order properties.

Let A and B be bounded measurable subsets of \mathbb{R}^d . If $\|\mathbf{h}\|$ is sufficiently large such that $(A + \mathbf{h}) \cap B = \emptyset$, then

$$\text{cov}\{N(A + \mathbf{h}), N(B)\} = \int_A \int_B \{g(\mathbf{s}_1 - \mathbf{s}_2 - \mathbf{h}) - 1\} \lambda(\mathbf{s}_1) \lambda(\mathbf{s}_2) d\mathbf{s}_2 d\mathbf{s}_1.$$

If $g(\mathbf{s}_1 - \mathbf{s}_2 - \mathbf{h}) \rightarrow 1$ as $\|\mathbf{h}\| \rightarrow \infty$, then $\text{cov}\{N(A + \mathbf{h}), N(B)\} \rightarrow 0$, indicating that $N(A + \mathbf{h})$ and $N(B)$ are almost independent. To theoretically address this issue, we need to assume that \mathcal{N} satisfies the strong mixing condition. This approach was first introduced for dependent random variables in [36] and later extended to stationary SPPs in [26]. Here we want to modify it to substationary SPPs.

Suppose that \mathcal{N} is substationary in \mathcal{L} . Let $\mathcal{B}(A)$ be the collection of Borel sets generated by A . Denote the diameter of A by $\rho(A)$ and $\rho(A_1, A_2)$ as the minimum distance between A_1 and A_2 , where $\rho(A) = \sup_{\mathbf{s}, \mathbf{s}' \in A} \|\mathbf{s} - \mathbf{s}'\|$ and $\rho(A_1, A_2) = \min_{\mathbf{s} \in A_1, \mathbf{s}' \in A_2} \|\mathbf{s} - \mathbf{s}'\|$. Let

$$\alpha(u, v) = \sup\{|\Pr(U_1 \cap U_2) - \Pr(U_1)\Pr(U_2)| : U_1 \in \mathcal{B}(A_1), U_2 \in \mathcal{B}(A_2), \\ \rho(A_1, A_2) \geq u, \rho(A_1) \leq v, \rho(A_2) \leq v, A_1, A_2 \in \mathcal{B}(\mathbb{R}^d)\}$$

be the mixing coefficients, where $\Pr(U)$ is the probability of U formulated by the distribution of \mathcal{N} . We say \mathcal{N} is strongly mixing if $\alpha(zu, zv) \rightarrow 0$ as $z \rightarrow \infty$.

We show the functional Central Limit Theorem of $M_{z,\mathcal{L}}(\mathcal{A}_{\mathbf{t},V})$ for $\mathbf{t} \in [0, \infty)^r$ and $V \in \mathcal{B}(\mathcal{L}^\perp)$ by the traditional method. It was initially proposed by [25] and later modified by [23]. The main idea is to split $A_{z,\mathcal{L}}$ for $A \in \mathcal{A}_{\mathbf{t},V}$ into two components \mathcal{B} and \mathcal{C} . Both \mathcal{B} and \mathcal{C} can be expressed as sums of disjoint blocks, where counts in blocks of \mathcal{B}

are almost independent and counts in blocks of C can be ignored. This method is popular for proving the asymptotic normality of stationary time series and weakly dependent SPPs. Since the proof of our functional Central Limit Theorem is a simple use of the method, we here introduce it only briefly.

Theorem 3. Assume that N is strongly mixing and substationary in \mathcal{L} . If the fourth intensity function of N is uniformly bounded and

$$\int_0^\infty z^{d-1/2} \alpha(zu, zv) dz < \infty \quad (4)$$

for any $u, v > 0$, then $M_{z,\mathcal{L}}(\mathcal{A}_V)$ weakly converges to a Gaussian process with independent increments.

Proof. Let $A_i = U_i \times V$ for any disjoint $U_1, \dots, U_m \in \mathcal{B}(\mathcal{L})$ and $V \in \mathcal{B}(\mathcal{L}^\perp)$. Let $\mathcal{A} = \{A_1, \dots, A_m\}$. Using the method of Theorem 1.3 in [25], we can partition \mathcal{A} into many small blocks, denoted by $\mathcal{B} = \{\mathcal{B}_1, \dots, \mathcal{B}_{k_1}\}$ and $C = \{C_1, \dots, C_{k_2}\}$, where $k_1, k_2 \rightarrow \infty$ as $z \rightarrow \infty$, such that

$$\min_{B \in \mathcal{B}_{j,n}, B' \in \mathcal{B}_{j',n}, j \neq j'} \rho(B, B') \geq u \quad \text{and} \quad N(\mathcal{A}_{z,\mathcal{L}}) = N(\mathcal{B}_{z,\mathcal{L}}) + N(C_{z,\mathcal{L}}).$$

By the method of Theorem 1.4 in [25], we can choose k_1 such that it is bounded by $z^{(1+u)/(2d)}$ for any positive u , if z is sufficiently large. Then,

$$\left| \mathbb{E} \left\{ e^{it \sum_{j=1}^m M_{z,\mathcal{L}}(A_j)} \right\} - \prod_{j=1}^{k_1} \mathbb{E} \left\{ e^{it M_{z,\mathcal{L}}(\mathcal{B}_j)} \right\} \right| \leq 4\kappa_1 \alpha(zu, zv),$$

where $v = \max\{\rho(U_i)\}$. If (4) holds, then the right-hand side of the above goes to 0 as $z \rightarrow \infty$. Since λ_4 is uniformly bounded, we conclude that Lyapounov's Condition holds, implying that the asymptotic normality holds. Then, the Central Limit Theorem of $M_{z,\mathcal{L}}(\mathcal{A})$ holds for finite \mathcal{A} . By the same method in the proof of the tightness that we have shown in Theorem 2, we can show the tightness of the distribution of $M_{z,\mathcal{L}}(\mathcal{A}_V)$ for a sufficiently large z . It provides the functional Central Limit Theorem of $M_{z,\mathcal{L}}(\mathcal{A}_V)$ and the conclusion of the theorem. \square

Corollary 1. If all conditions of Theorem 3 hold, then there exists $C > 0$ such that for any $A \in \mathcal{B}(\mathbb{R}^d)$, one has $M_{z,\mathcal{L}}(A) \rightsquigarrow N[0, C^2 \mu(A)]$.

Proof. We assume that there exists $\mathbf{t} \in \mathbb{R}^r$ and $v \in \mathcal{B}(\mathcal{L}^\perp)$ such that $A = A_{\mathbf{t},V}$. If we partition $(0, t_1 \mathbf{u}_1] \times \dots \times (0, t_r \mathbf{u}_r]$ into countable small rectangles, denoted by $\mathcal{A} = \{U_i : i \in \mathbb{N}\}$, then $A_{\mathbf{t},V} = \bigcup_{i=1}^\infty U_i \times V$. By Theorem 3, we have $M_{z,\mathcal{L}}(\mathcal{A}) \rightsquigarrow N(0, D_{\mathcal{A}})$, where $D_{\mathcal{A}}$ is a diagonal matrix determined by the property of \mathcal{A} . There exists a σ -finite measure $\tilde{\mu}$ on \mathbb{R}^d such that $M_{z,\mathcal{L}}(A_{\mathbf{t},V}) \rightsquigarrow N(0, \tilde{\mu}(A_{\mathbf{t},V}))$. Since \mathcal{A}_V is a π -system, we conclude that $\tilde{\mu}$ is uniquely determined. Thus, $\tilde{\mu}(A)$ is proportional to $\mu(A)$, implying the conclusion. \square

In practice, an important issue is to estimate the first-order intensity function $\lambda(\mathbf{s})$ under substationarity with an unknown \mathcal{L} . As $\lambda(\mathbf{s})$ only varies in \mathcal{L} , it is equivalent to estimate $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ and \mathcal{L} together. Since it is generally inappropriate to model $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ parametrically, we propose a nonparametric method for the intensity function. As \mathcal{L} can be formulated by a rotation of a linear subspace spanned by coordinates, we propose a parametric method to estimate it. Therefore, we classify our estimation as a semiparametric approach. The functional Central Limit Theorems given by Theorems 2 and 3 provide the theoretical basis of the estimation method.

4. Estimation

Let N be substationary in $\mathcal{L} \subseteq \mathbb{R}^d$. Assume that points of N are only collected in bounded $S \in \mathcal{B}(\mathbb{R}^d)$ such that they can be represented by $N(S)$. Denote n as the observed value of $N(S)$. Our main interest is to estimate $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ and \mathcal{L} simultaneously. We propose a two-step method. In the first step, we estimate $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ for a given \mathcal{L} , where a nonparametric method is adopted. In the second step, we estimate \mathcal{L} , where a parametric method is adopted. The implementation of the second step needs results of the first step.

We propose a kernel-based method to estimate $\lambda(\mathbf{s})$ for a given \mathcal{L} . We investigate the usual kernel-based method without using substationarity [8]. This provides an estimator of $\lambda(\mathbf{s})$ as

$$\hat{\lambda}_h(\mathbf{s}) = C_h^{-1}(\mathbf{s}) \int_S K_h(\mathbf{s}' - \mathbf{s}) N(d\mathbf{s}') \quad (5)$$

when $N(\mathcal{S}) > 0$, where $K_h(\mathbf{s}) = K(\mathbf{s}/h)/h^d$ with bandwidth $h \in (0, \infty)$ is a kernel density function on \mathbb{R}^d and $C_h(\mathbf{s}) = \int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}'$ is the Berman–Diggle boundary correction [2]. For convenience, one usually defines $\hat{\lambda}_h(\mathbf{s}) = 0$ if $N(\mathcal{S}) = 0$. By Campbell's theorem, we have

$$E\{\hat{\lambda}_h(\mathbf{s})\} = C_h^{-1}(\mathbf{s}) \int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) \lambda(\mathbf{s}') d\mathbf{s}'$$

and

$$\text{var}\{\hat{\lambda}_h(\mathbf{s})\} = C_h^{-2}(\mathbf{s}) \int_{\mathcal{S}} \int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) K_h(\mathbf{s}'' - \mathbf{s}) \{g(\mathbf{s}', \mathbf{s}'') - 1\} \lambda(\mathbf{s}') \lambda(\mathbf{s}'') d\mathbf{s}' d\mathbf{s}'' + C_h^{-2}(\mathbf{s}) \int_{\mathcal{S}} K_h^2(\mathbf{s}' - \mathbf{s}) \lambda(\mathbf{s}') d\mathbf{s}'.$$

We modify (5) for a substationary \mathcal{N} in \mathcal{L} and obtain an estimator of $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ (or $\lambda(\mathbf{s})$, equivalently) as

$$\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp}) = C_{h,\mathcal{L}^\perp}^{-1}(\mathbf{s}_{\mathcal{L}^\perp}) \sum_{i=1}^n K_{h,\mathcal{L}^\perp}(\mathbf{s}_{i,\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) = C_{h,\mathcal{L}^\perp}^{-1}(\mathbf{s}_{\mathcal{L}^\perp}) \int_{\mathcal{S}} K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) \mathcal{N}(d\mathbf{s}') \quad (6)$$

if $N(\mathcal{S}) > 0$ and $\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp}) = 0$ if $N(\mathcal{S}) = 0$, where $\mathbf{s}_{i,\mathcal{L}^\perp}$ is the projection of \mathbf{s}_i on \mathcal{L}^\perp , $K_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp}) = K(\mathbf{s}_{\mathcal{L}^\perp}/h)/h^r$ is a kernel density function on \mathcal{L}^\perp , and $C_{h,\mathcal{L}^\perp}(\mathbf{s}) = \int_{\mathcal{S}} K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) d\mathbf{s}'$ is still the boundary correction. Still by Campbell's theorem, we have

$$E\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} = C_{h,\mathcal{L}^\perp}^{-1}(\mathbf{s}_{\mathcal{L}^\perp}) \int_{\mathcal{S}} K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) \lambda(\mathbf{s}'_{\mathcal{L}^\perp}) d\mathbf{s}',$$

and

$$\begin{aligned} \text{var}\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} &= C_{h,\mathcal{L}^\perp}^{-2}(\mathbf{s}_{\mathcal{L}^\perp}) \int_{\mathcal{S}} \int_{\mathcal{S}} K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) K_{h,\mathcal{L}^\perp}(\mathbf{s}''_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) \{g(\mathbf{s}', \mathbf{s}'') - 1\} \lambda(\mathbf{s}'_{\mathcal{L}^\perp}) \lambda(\mathbf{s}''_{\mathcal{L}^\perp}) d\mathbf{s}' d\mathbf{s}'' \\ &\quad + C_{h,\mathcal{L}^\perp}^{-2}(\mathbf{s}) \int_{\mathcal{S}} K_{h,\mathcal{L}^\perp}^2(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) \lambda(\mathbf{s}'_{\mathcal{L}^\perp}) d\mathbf{s}'. \end{aligned} \quad (7)$$

If $r = d$, then $\mathcal{L} = \mathbb{R}^d$ and (6) becomes

$$\hat{\lambda} = n/|\mathcal{S}|. \quad (8)$$

Since \mathcal{N} is stationary in this case, the first-order intensity function is a constant, indicating that the estimator must be a constant.

We compare the MSEs (mean square errors) of $\hat{\lambda}_h(\mathbf{s})$ and $\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})$ as $h \rightarrow 0$ and $hz \rightarrow \infty$ in the case where $\mathcal{S} = A_{z,\mathcal{L}}$ for a bounded $A \in \mathcal{B}(\mathbb{R}^d)$. We find that the bias of $\hat{\lambda}_h(\mathbf{s})$, which is given by $\text{Bias}\{\hat{\lambda}_h(\mathbf{s})\} = E\{\hat{\lambda}_h(\mathbf{s})\} - \lambda(\mathbf{s})$, can go to 0 as $h \rightarrow 0$, but it can simultaneously cause $\text{var}\{\hat{\lambda}_h(\mathbf{s})\} \rightarrow \infty$. To make $\text{var}\{\hat{\lambda}_h(\mathbf{s})\}$ small, we need to choose a large h , but it increases $\text{Bias}\{\hat{\lambda}_h(\mathbf{s})\}$. Thus, $\text{MSE}\{\hat{\lambda}_h(\mathbf{s})\} = [E\{\hat{\lambda}_h(\mathbf{s})\} - \lambda(\mathbf{s})]^2 + \text{var}\{\hat{\lambda}_h(\mathbf{s})\}$ cannot go to 0 as $z \rightarrow \infty$. However, by selecting an appropriate h , we can make $\text{MSE}\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} \rightarrow 0$.

Theorem 4. Let \mathcal{N} be substationary in \mathcal{L} and $\mathcal{S} = A_{z,\mathcal{L}}$ for a bounded $A \in \mathcal{B}(\mathbb{R}^d)$ with $r > 0$ and $|\partial A| = 0$. Suppose that all conditions of Theorem 3 hold. Assume that $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$ is positive and continuous in the interior of \mathcal{S} and $v_r(A_{\mathbf{v}})$ is almost surely continuous in any $\mathbf{v} \in A^\perp$. For an interior point \mathbf{s} of A , if $h \rightarrow 0$ and $hz \rightarrow \infty$, then $\text{MSE}\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} \rightarrow 0$ as $z \rightarrow \infty$.

Proof. For an interior point of $\mathbf{s} \in A$, one has that

$$E\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} = \left\{ \int_{\mathcal{L}^\perp} v_r(A_{\mathbf{s}_{\mathcal{L}^\perp} + h\mathbf{v}}) K(\mathbf{v}) d\mathbf{v} \right\}^{-1} \int_{\mathcal{L}^\perp} v_r(A_{\mathbf{s}_{\mathcal{L}^\perp} + h\mathbf{v}}) K(\mathbf{v}) \lambda(\mathbf{s}_{\mathcal{L}^\perp} + h\mathbf{v}) d\mathbf{v}.$$

If $h \rightarrow 0$ as $z \rightarrow \infty$, then $\lim_{z \rightarrow \infty} E\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} = \lambda(\mathbf{s}_{\mathcal{L}^\perp})$ by the continuity of $v_r(A_{\mathbf{v}})$ and $\lambda(\mathbf{s}_{\mathcal{L}^\perp})$. By (7), we have

$$\begin{aligned} \text{var}\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})\} &= \left\{ \int_A K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) d\mathbf{s}' \right\}^{-2} \int_A \int_A K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) K_{h,\mathcal{L}^\perp}(\mathbf{s}''_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) [g(\mathbf{s}', \mathbf{s}'' + z(\mathbf{s}''_{\mathcal{L}^\perp} - \mathbf{s}'_{\mathcal{L}^\perp})) - 1] \\ &\quad \times \lambda(\mathbf{s}'_{\mathcal{L}^\perp}) \lambda(\mathbf{s}''_{\mathcal{L}^\perp}) d\mathbf{s}' d\mathbf{s}'' + z^{-r} \left\{ \int_A K_{h,\mathcal{L}^\perp}(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) d\mathbf{s}' \right\}^{-2} \int_A K_{h,\mathcal{L}^\perp}^2(\mathbf{s}'_{\mathcal{L}^\perp} - \mathbf{s}_{\mathcal{L}^\perp}) \lambda(\mathbf{s}'_{\mathcal{L}^\perp}) d\mathbf{s}'. \end{aligned}$$

By Theorem 3, we conclude that the first term of the above goes to 0 as $z \rightarrow \infty$. Therefore, we only need to study the second term, which satisfies

$$(hz)^{-r} \left\{ \int_{\mathcal{L}^\perp} \nu_r(A_{\mathbf{s}_{\mathcal{L}^\perp} + h\mathbf{v}}) K(\mathbf{v}) d\mathbf{v} \right\}^{-2} \int_{\mathcal{L}^\perp} K^2(\mathbf{v}) \lambda(\mathbf{s}_{\mathcal{L}^\perp} + h\mathbf{v}) d\mathbf{v} \rightarrow 0$$

as $hz \rightarrow \infty$. This concludes the argument. \square

Example 1. We interpret Theorem 4 of \mathcal{L} in a special case. Assume that \mathcal{N} is substationary in \mathbb{R}^2 and $\mathcal{L} = \{(x, 0) : x \in \mathbb{R}\}$ such that $d = 2$, $r = 1$, and the first-order intensity function can be expressed as $\lambda(\mathbf{s}) = \lambda(y)$, where $\mathbf{s} = (x, y)$. Suppose that $S = [0, z] \times [0, \omega]$ such that observations of \mathcal{N} can be expressed by points within $[0, z] \times [0, \omega]$, denoted by $\mathbf{s}_1, \dots, \mathbf{s}_n$, where $n = N(S)$ is the total number of observed points. If we choose $K(\mathbf{s}) = (2\pi)^{-1} e^{-(x^2+y^2)/2}$ for the case where substationarity is not accounted for, then $K_h(\mathbf{s}) = \phi(x/h)\phi(y/h)/h = (2\pi h)^{-1} e^{-(x^2+y^2)/(2h^2)}$, where ϕ is the PDF of $\mathcal{N}(0, 1)$. By (5), we have

$$E\{\hat{\lambda}_h(\mathbf{s})\} = \left\{ \int_0^z \int_0^\omega \frac{1}{2\pi h^2} e^{-\frac{(x'-x)^2+(y'-y)^2}{2h^2}} dy' dx' \right\}^{-1} \int_0^z \int_0^\omega \frac{1}{2\pi h^2} e^{-\frac{(x'-x)^2+(y'-y)^2}{2h^2}} \lambda(\mathbf{s}') ds'.$$

Then, $\lim_{h \rightarrow 0} E\{\hat{\lambda}_h(\mathbf{s})\} = \lambda(\mathbf{s})$, implying that the bias of $\hat{\lambda}_h(\mathbf{s})$ can only disappear as $h \rightarrow 0$ but this can make $\text{var}\{\hat{\lambda}_h(\mathbf{s})\}$ large. If we choose $K(y) = (2\pi)^{-1/2} e^{-y^2/2}$ for the case where substationarity is accounted for, then $K_{h,\mathcal{L}^\perp}(y) = \phi(y/h)/h = (\sqrt{2\pi}h)^{-1} e^{-y^2/(2h^2)}$. By (6), we have

$$E\{\hat{\lambda}_{h,\mathcal{L}^\perp}(y)\} = \left\{ \int_0^\omega \frac{1}{\sqrt{2\pi}h} e^{-\frac{(y'-y)^2}{2h^2}} dy' \right\}^{-1} \int_0^\omega \frac{1}{\sqrt{2\pi}h} e^{-\frac{(y'-y)^2}{2h^2}} \lambda(y) dy.$$

Then, $\lim_{h \rightarrow 0} E\{\hat{\lambda}_{h,\mathcal{L}^\perp}(y)\} = \lambda(y)$, implying that the bias of $\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$ also disappears as $h \rightarrow 0$. By (7), we have

$$\begin{aligned} \text{var}\{\hat{\lambda}_{h,\mathcal{L}^\perp}(y)\} &= \left\{ \int_0^\omega \frac{1}{\sqrt{2\pi}h} e^{-\frac{(y'-y)^2}{2h^2}} dy' \right\}^{-2} \int_0^\omega \int_0^\omega \frac{1}{2\pi h^2} e^{-\frac{(y'-y)^2+(y''-y)^2}{2h^2}} \lambda(y') \lambda(y'') dy' dy'' \\ &\quad \times \frac{1}{z} \int_0^z [g\{(0, y'), (x'', y'')\} - 1] dx'' dy'' + \frac{1}{z} \left\{ \int_0^\omega \frac{1}{\sqrt{2\pi}h} e^{-\frac{(y'-y)^2}{2h^2}} dy' \right\}^{-2} \int_0^\omega \frac{1}{\sqrt{2\pi}h} e^{-\frac{(y'-y)^2}{2h^2}} \lambda(y') dy'. \end{aligned}$$

If all conditions of Theorem 3 hold, then $\lim_{x'' \rightarrow \infty} g\{(0, y'), (x'', y'')\} - 1 = 0$. Thus, the first term of the above goes to 0 as $z \rightarrow \infty$ and the second term goes to 0 if $zh \rightarrow \infty$, satisfying the conclusion of Theorem 4.

As \mathcal{L} is also unknown, we need to estimate it, too. We want to modify the composite likelihood approach. Let \mathcal{L}_0 be the true linear subspace with dimension r_0 . If $r_0 = 0$, then \mathcal{N} is not substationary in any linear subspace of \mathbb{R}^d . If $r_0 = d$, then \mathcal{N} is stationary in the entire \mathbb{R}^d . Otherwise, \mathcal{N} is substationary in \mathcal{L} but nonstationary in \mathbb{R}^d . Without loss of generality, we assume $\mathcal{L}_0 = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{r_0})$, where $\mathbf{e}_1, \dots, \mathbf{e}_d$ is the orthonormal basis of \mathbb{R}^d . Let $\mathbf{u}_1, \dots, \mathbf{u}_d$ be another orthonormal basis of \mathbb{R}^d such that $\mathcal{L} = \text{span}(\mathbf{u}_1, \dots, \mathbf{u}_r)$. Then, we can find an orthogonal transformation (or orthogonal rotation) \mathbf{Q} such that $\mathcal{Q} = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_r) = \mathcal{L}$. Let \mathcal{Q} be the collection of orthogonal transformation in \mathbb{R}^d and $\hat{\mathcal{L}}$ be an estimator of \mathcal{L} . Then, $\text{span}(\mathbf{e}_1, \dots, \mathbf{e}_r) = \mathbf{Q}^{-1} \hat{\mathcal{L}}$, which provides an estimator of \mathbf{Q} , implying that estimating \mathcal{L} is equivalent to estimating $\mathbf{Q} \in \mathcal{Q}$. By the inverse of \mathbf{Q} , we can study the relationship between $\hat{\mathcal{L}}$ and \mathcal{L}_0 based on $\mathbf{u}_1, \dots, \mathbf{u}_d$ instead of $\mathbf{e}_1, \dots, \mathbf{e}_d$. Then, we have a fixed \mathcal{L} in $K_{h,\mathcal{L}}$ but a varied \mathcal{L}_0 , indicating that the impact of h can be removed and the problem becomes purely parametric. Therefore, consistency and optimal rate of convergence in the parametric case are expected. We provide a proof of this in the following.

For an intensity function $\lambda(\mathbf{s}) = \lambda(\mathbf{s})$, let

$$\ell(\lambda) = \ell\{\lambda(\mathbf{s})\} = \sum_{i=1}^n \ln \lambda(\mathbf{s}_i) - \int_S \lambda(\mathbf{s}) d\mathbf{s} = \sum_{i=1}^n \ln \lambda(\mathbf{s}_i) - \kappa \quad (9)$$

be the log-likelihood function of $\mathcal{N}(S)$ if \mathcal{N} is Poisson, where $\kappa = E\{N(S)\} = \int_S \lambda(\mathbf{s}) d\mathbf{s}$. Then, $\ell\{\lambda(\mathbf{s})\}$ can be treated as the composite log-likelihood of $\mathcal{N}(S)$ if \mathcal{N} is non-Poisson [20]. Therefore, we can estimate \mathcal{L} by

$$\hat{\mathbf{Q}}_h = \underset{\mathbf{Q} \in \mathcal{Q}}{\text{argmax}} \ell(\hat{\lambda}_{h,\mathcal{L}^\perp}), \quad (10)$$

which provides an estimator of $\hat{\mathcal{L}}_h$. We need to determine the best choice of h in $\hat{\mathcal{L}}_h$. Various methods related to this problem have been proposed. One method is to use the generalized cross validation (GCV) approach [16]. Another method relies on its asymptotic performance based on mean integrated square errors (MISEs); see Section 24.2 in [39]. We study this problem in Section 5.

We evaluate asymptotic properties, including consistency and asymptotic normality, of $\hat{\mathcal{L}}_h$. We still assume that $\mathcal{S} = A_{z, \mathcal{L}_0}$ for a bounded $A \in \mathcal{B}(\mathbb{R}^d)$. We use the traditional method based on the Kullback–Leibler divergence (or information). For any $\tilde{\lambda} = \tilde{\lambda}(\mathbf{s})$ and $\lambda = \lambda(\mathbf{s})$, two options of intensity functions of \mathcal{N} , the Kullback–Leibler divergence is defined as $D(\tilde{\lambda}, \lambda) = \ell(\tilde{\lambda}) - \ell(\lambda)$. Let $\lambda_0(\mathbf{s})$ be the true intensity function. By the Shannon–Kolmogorov Information Inequality (see p. 113 in [14]), we have $E\{D(\lambda, \lambda_0)\} < 0$ for any distinct λ and λ_0 . To show consistency of $\hat{\mathcal{L}}_h$, we need to study the asymptotic performance of $D(\hat{\lambda}_{h, \mathcal{L}^\perp}, \hat{\lambda}_{h, \mathcal{L}_0^\perp})$. By $\int_{\mathcal{S}} \hat{\lambda}_{h, \mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp}) d\mathbf{s} = n$ for any \mathcal{L} and K_{h, \mathcal{L}^\perp} , we have

$$D_h(\mathcal{L}, \mathcal{L}_0) = \sum_{i=1}^n \ln \frac{\hat{\lambda}_{h, \mathcal{L}^\perp}(\mathbf{s}_i)}{\hat{\lambda}_{h, \mathcal{L}_0^\perp}(\mathbf{s}_i)} = \int_{\mathcal{S}} \ln \frac{\hat{\lambda}_{h, \mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})}{\hat{\lambda}_{h, \mathcal{L}_0^\perp}(\mathbf{s}_{\mathcal{L}^\perp})} f(d\mathbf{s})$$

when $n > 0$ and $D_h(\mathcal{L}, \mathcal{L}_0) = 0$ when $n = 0$. Since asymptotic properties of $D_h(\mathcal{L}, \mathcal{L}_0)$ cannot be directly implied by the Shannon–Kolmogorov Information Inequality, we have to provide a proof. We show it for Poisson SPPs first and then extend it to other SPPs. The role of the following lemma is equivalent to the role of the optimality condition in maximum likelihood estimation based on the Kullback–Leibler information; see p. 112 in [14].

Lemma 1. Assume that \mathcal{N} is Poisson with $\lambda(\mathbf{s}) > 0$ for all $\mathbf{s} \in \mathcal{S}$. Let $\tilde{D}_h(\mathcal{L}, \mathcal{L}_0) = \sum_{i=1}^n \tilde{D}_{i,h}(\mathcal{L}, \mathcal{L}_0)$, where

$$\tilde{D}_{i,h}(\mathcal{L}, \mathcal{L}_0) = \ln \frac{C_{h, \mathcal{L}^\perp}^{-1}(\mathbf{s}_i, \mathcal{L}^\perp) \sum_{j=1, j \neq i}^n K_{h, \mathcal{L}^\perp}(\mathbf{s}_j, \mathcal{L}^\perp - \mathbf{s}_i, \mathcal{L}^\perp)}{C_{h, \mathcal{L}_0^\perp}^{-1}(\mathbf{s}_i, \mathcal{L}^\perp) \sum_{j=1, j \neq i}^n K_{h, \mathcal{L}^\perp}(\mathbf{s}_j, \mathcal{L}_0^\perp - \mathbf{s}_i, \mathcal{L}_0^\perp)}.$$

If $r_0 \geq 1$, then $E\{\tilde{D}_h(\mathcal{L}, \mathcal{L}_0)\} \leq 0$ and the equality holds if and only if $\mathcal{L} \subseteq \mathcal{L}_0$.

Proof. Note that $f(\mathbf{s}) = \lambda(\mathbf{s})/\kappa$ is a PDF on \mathcal{S} and $\mathcal{N}(\mathcal{S}) \sim \mathcal{P}(\kappa)$. Let $\mathcal{N}_{-i}(\mathcal{S}) = \{\mathbf{s}_j : j \in \{1, \dots, n\}, j \neq i\}$ when $n \geq 2$ and $\mathcal{N}_{-i}(\mathcal{S}) = \emptyset$ when $n \in \{0, 1\}$. Let $\xi_{-i,0,\mathcal{L}}(A) = 0$ and

$$\xi_{-i,n}(A) = \sum_{j=1, j \neq i}^n \int_A C_{h, \mathcal{L}^\perp}^{-1}(\mathbf{s}_{\mathcal{L}^\perp}) K_{h, \mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp} - \mathbf{s}_j, \mathcal{L}^\perp) f(\mathbf{s}) d\mathbf{s}$$

when $n \geq 1$ for any $A \in \mathcal{B}(\mathcal{S})$. Then

$$\xi_{-i,n,\mathcal{L}}(d\mathbf{s}) = \{C_{h, \mathcal{L}^\perp}^{-1}(\mathbf{s}_i, \mathcal{L}^\perp) \sum_{j=1, j \neq i}^n K_{h, \mathcal{L}^\perp}(\mathbf{s}_j, \mathcal{L}^\perp - \mathbf{s}_i, \mathcal{L}^\perp)\} f(d\mathbf{s})$$

given $\mathcal{N}_{-i}(\mathcal{S})$, implying that $\xi_{-i,n,\mathcal{L}}(d\mathbf{s})/\xi_{-i,n,\mathcal{L}}(\mathcal{S})$ is a probability measure on \mathcal{S} . When $n \geq 1$, by Jensen's inequality, we have

$$\int_{\mathcal{S}} \tilde{D}_{i,h}(\mathcal{L}, \mathcal{L}_0) \frac{\xi_{-i,n,\mathcal{L}_0}(d\mathbf{s})}{\xi_{-i,n,\mathcal{L}_0}(\mathcal{S})} \leq \ln \int_{\mathcal{S}} \frac{\xi_{-i,n,\mathcal{L}}(d\mathbf{s})}{\xi_{-i,n,\mathcal{L}}(\mathcal{S})} = 0,$$

where the equality holds if and only if $\mathcal{L} \subseteq \mathcal{L}_0$. □

Theorem 5. Assume that \mathcal{N} is Poisson and intrinsically substationary in \mathcal{L}_0 with $r_0 \geq 1$ and $\lambda(\mathbf{s}) > 0$ for all $\mathbf{s} \in \mathcal{S}$. Suppose that there exists a function $\Psi(\mathbf{s})$ such that $E\|\Psi(\mathbf{s})\| < \infty$ and $n^{-1}D_h(\mathcal{L}, \mathcal{L}_0) \leq \Psi(\mathbf{s})$ almost surely. Then $\hat{\mathcal{L}}_h \subseteq \mathcal{L}_0$ almost surely.

Proof. We immediately have $n^{-1}D_h(\mathcal{L}, \mathcal{L}_0) - E\{n^{-1}D_h(\mathcal{L}, \mathcal{L}_0)|n\} \xrightarrow{a.s.} 0$ by the Ergodic Theorem; see p. 315 in [4]. Since \mathcal{Q} is compact by the method of Theorem 16(a) in [14], we can further show that

$$\sup_{\mathcal{L}} |n^{-1}D_h(\mathcal{L}, \mathcal{L}_0) - E\{n^{-1}D_h(\mathcal{L}, \mathcal{L}_0)|n\}| \xrightarrow{a.s.} 1.$$

Since $n^{-1}\tilde{D}_h(\mathcal{L}, \mathcal{L}_0) = n^{-1}D_h(\mathcal{L}, \mathcal{L}_0) + O(1/n)$, we can show the conclusion by Lemma 1 and the standard method in the proof of consistency of the MLE; see Theorem 17 in [14]. □

Theorem 6. Let $\dot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}})$ and $\ddot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}})$ be the gradient and Hessian matrix of $\ell_i(\hat{\lambda}_{h,\mathcal{L}}) = \ln \hat{\lambda}_{h,\mathcal{L}}(\mathbf{s}_i) - 1$ with respect to \mathcal{L} . Assume that $|\partial A| = 0$, $|A| > 0$, and \mathcal{N} is Poisson and intrinsically substationary in \mathcal{L} with $r_0 \geq 1$, and $\lambda(\mathbf{s})$ is positive, second-order continuous, and uniformly bounded in \mathcal{S} . If $r = r_0$, then $\sqrt{n}(\hat{\mathcal{L}}_h - \mathcal{L}_0) \rightsquigarrow \mathcal{N}[\mathbf{0}, \mathbf{E}^{-1}\{\ddot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}})\}]$.

Proof. By (9) and the fact that $\int_{\mathcal{S}} \hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}) d\mathbf{s} = n$, we have $\ell(\hat{\lambda}_{h,\mathcal{L}^\perp}) = \sum_{i=1}^n \ell_i(\hat{\lambda}_{h,\mathcal{L}})$. After ignoring the small order terms in the Taylor expansion, we approximately have

$$\mathbf{0} = \dot{\ell}(\hat{\lambda}_{h,\mathcal{L}^\perp}) \approx \dot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp}) + \ddot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp})(\hat{\mathcal{L}} - \mathcal{L}_0),$$

implying that

$$\sqrt{n}(\hat{\mathcal{L}} - \mathcal{L}_0) = \{n^{-1}\ddot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp})\}^{-1}\{(1/\sqrt{n})\dot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp})\} + o_p(1).$$

Since $\mathbf{s}_1, \dots, \mathbf{s}_n$ are iid with PDF $f(\mathbf{s})$ for a given n , we conclude that $\dot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp})/\sqrt{n}$ is asymptotically normal by the functional Central Limiting Theorem in the iid case; see Chapter 9 in [39]. Then,

$$(1/\sqrt{n})\dot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp}) \rightsquigarrow \mathcal{N}[\mathbf{0}, \mathbf{E}\{\dot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}_0})\dot{\ell}_i^\top(\hat{\lambda}_{h,\mathcal{L}_0})\}].$$

By the Ergodic Theorem, we have $n^{-1}\ddot{\ell}(\hat{\lambda}_{h,\mathcal{L}_0^\perp}) \xrightarrow{a.s.} \mathbf{E}\{\ddot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}_0})\}$. We draw the final conclusion by $\mathbf{E}\{\dot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}_0})\dot{\ell}_i^\top(\hat{\lambda}_{h,\mathcal{L}_0})\} = \mathbf{E}\{\ddot{\ell}_i(\hat{\lambda}_{h,\mathcal{L}_0})\}$. \square

Theorem 7. Suppose that all conditions of Theorem 3 hold. Assume that $|\partial A| = 0$, $|A| > 0$, and \mathcal{N} is intrinsically substationary in \mathcal{L}_0 with $r_0 \geq 1$, and $\lambda(\mathbf{s})$ is positive, second-order continuous, and uniformly bounded in \mathcal{S} . If $r = r_0$, then $\sqrt{n}(\hat{\mathcal{L}}_h - \mathcal{L}_0)$ weakly converges to a multivariate normal distribution.

Proof. It can be proven by the method in the proof of Theorem 3 with the conclusion of Theorem 6. \square

The Central Limit Theorem of $\sqrt{n}(\hat{\mathcal{L}}_h - \mathcal{L}_0)$ for the Poisson case is extended to a non-Poisson case by Theorem 7. Note that the asymptotic holds only when $r = r_0$. We need to address the case when $r \neq r_0$. If $r > r_0$, then \mathcal{N} cannot be substationary in \mathcal{L} . We cannot have consistency of $\hat{\lambda}_{h,\mathcal{L}}(\mathbf{s})$ for any \mathcal{L} . If $r < r_0$, by Theorem 5, we have $\hat{\mathcal{L}}_h \subseteq \mathcal{L}_0$ almost surely. Therefore, it is important to evaluate $\hat{\mathcal{L}}_h$ which is studied in the next section.

5. Simulation

We carried out simulation studies to evaluate the performance of $\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}) = \hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s}_{\mathcal{L}^\perp})$ given by (6). This also contained estimation of \mathcal{L} . We simulated realizations from Poisson and Poisson cluster SPPs in a rectangular region $\mathcal{S} = [0, z] \times [0, \omega]$, the region used in Example 1. We chose $\omega = 1$ in our simulations. We selected these processes because they are popular in modeling ecological, environmental, geographical data. In both processes, we chose the first-order intensity function as $\lambda(\mathbf{s}) = 100\{\Gamma^{-1}(a)/\Gamma(2a)\}y^{a-1}(1-y)^{a-1}$ for a selected $a \geq 1$ such that we always had $\kappa = \mathbf{E}\{N(\mathcal{S})\} = 100z$. Note that $\lambda(y)/100$ is the PDF of $\mathcal{B}(a, a)$ distribution. We chose a equal to 1.0, 1.5, 2.0, 2.5, and 3.0 in our simulations. If $a = 1$, then \mathcal{N} was stationary in the entire \mathbb{R}^2 , where we had $r_0 = 2$; otherwise, it was only substationary in $\mathcal{L} = \{(x, y) : x \in \mathbb{R}\}$, where we had $r_0 = 1$. Since \mathcal{L} might be unknown, we also evaluated the performance of $\hat{\mathcal{L}}_h$, the estimator of \mathcal{L} given by (10).

To obtain a Poisson SPP, we first generated the number of points from the $\mathcal{P}(\kappa)$ distribution and then identically and independently generated the locations of these points. The horizontal values of these points were generated from the uniform distribution on $[0, 1]$. The vertical values of these points were generated from the $\mathcal{B}(a, a)$ distribution. To obtain a Poisson cluster SPP, we first generated their parent points from a Poisson SPP with its first-order intensity function equal to $\lambda(\mathbf{s})$ by the same method for the Poisson SPP. After parent points were derived, we generated offspring points. Each parent point generated $\mathcal{P}(\gamma)$ offspring points independently. The position of each offspring point relative to its parent point was defined as a radially symmetric Gaussian random variable with a standard deviation σ . We chose $\gamma = 5$ and $\sigma = 0.02$ in all the cases of Poisson cluster SPPs that we studied.

We studied two cases in the implementation of $\hat{\lambda}_{h,\mathcal{L}}(\mathbf{s})$. In the first case, we assumed that \mathcal{L} was known such that we could directly use (6). We chose $K_{h,\mathcal{L}^\perp}(y) = \phi(y/h)/h$ as the density of $\mathcal{N}(0, h^2)$. Then, we had $C_{h,\mathcal{L}^\perp}(y) = z[\Phi\{(\omega - y)/h\} - \Phi\{-y/h\}]$, where Φ is the CDF of $\mathcal{N}(0, 1)$. This provided

$$\hat{\lambda}_{h,\mathcal{L}^\perp}(y) = \frac{1}{z} \left\{ \Phi\left(\frac{\omega - y}{h}\right) - \Phi\left(-\frac{y}{h}\right) \right\}^{-1} \sum_{i=1}^n \frac{1}{h} \phi\left(\frac{y_i - y}{h}\right)$$

for $y \in (0, \omega)$. In the second case, we assumed that \mathcal{L} was unknown. We also estimated \mathcal{L} . Note that any one-dimensional linear subspace of \mathbb{R}^2 can be expressed as

$$\mathcal{L}_\theta = \{(u \cos \theta, u \sin \theta) : u \in \mathbb{R}\}, \quad (11)$$

for some $\theta \in [-\pi/2, \pi/2]$, leading to $\mathcal{L}_\theta^\perp = \{(-v \sin \theta, v \cos \theta) : v \in \mathbb{R}\}$. We chose $K_{h, \mathcal{L}_\theta^\perp}(v) = \exp(-(v/h)/h)$ on \mathcal{L}_θ^\perp .

To apply (6), we computed the analytic expression of $C_{h, \mathcal{L}_\theta^\perp}(v)$ for all $\theta \in [-\pi/2, \pi/2]$. If $\theta = 0$, then $C_{h, \mathcal{L}_0^\perp}(v) = z[\Phi\{(\omega - v)/h\} - \Phi(-v/h)]$. If $\theta = -\pi/2$, then $C_{h, \mathcal{L}_{-\pi/2}^\perp}(v) = \omega[\Phi\{(z - v)/h\} - \Phi(-v/h)]$. If $0 < \theta < \pi/2$, then $-z \sin \theta \leq v \leq \cos \theta$ and

$$\begin{aligned} C_{h, \mathcal{L}_\theta^\perp}(v) = & \frac{z \sin \theta + v}{\sin \theta \cos \theta} \left[\Phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \wedge 0 - v}{h} \right\} - \Phi \left\{ \frac{-z \sin \theta - v}{h} \right\} \right] \\ & + \left(\frac{z}{\cos \theta} \wedge \frac{\omega}{\sin \theta} \right) \left[\Phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \vee 0 - v}{h} \right\} - \Phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \wedge 0 - v}{h} \right\} \right] \\ & + \frac{\omega \cos \theta - v}{\sin \theta \cos \theta} \left[\Phi \left\{ \frac{\omega \cos \theta - v}{h} \right\} - \Phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \vee 0 - v}{h} \right\} \right] + \frac{h}{\sin \theta \cos \theta} \left[\phi \left(\frac{-z \sin \theta - v}{h} \right) \right. \\ & \left. + \phi \left(\frac{\omega \cos \theta - v}{h} \right) - \phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \wedge 0 - v}{h} \right\} - \phi \left\{ \frac{(\omega \cos \theta - z \sin \theta) \vee 0 - v}{h} \right\} \right]. \end{aligned}$$

If $-\pi/2 < \theta < 0$, then $0 \leq v \leq -z \sin \theta + \cos \theta$ and

$$\begin{aligned} C_{h, \mathcal{L}_\theta^\perp}(v) = & \left\{ \frac{z}{\cos \theta} \wedge \left(-\frac{\omega}{\sin \theta} \right) \right\} \left[\Phi \left\{ \frac{(-z \sin \theta) \vee (\omega \cos \theta - v)}{h} \right\} - \Phi \left\{ \frac{(-z \sin \theta) \wedge (\omega \cos \theta - v)}{h} \right\} \right] \\ & + \frac{z \sin \theta - \omega \cos \theta + v}{\sin \theta \cos \theta} \left[\Phi \left\{ \frac{-z \sin \theta + \omega \cos \theta - v}{h} \right\} - \Phi \left\{ \frac{(-z \sin \theta) \vee (\omega \cos \theta - v)}{h} \right\} \right] \\ & - \frac{v}{\sin \theta \cos \theta} \left[\Phi \left\{ \frac{(-z \sin \theta) \wedge (\omega \cos \theta - v)}{h} \right\} - \Phi \left\{ \frac{-v}{h} \right\} \right] + \frac{h}{\sin \theta \cos \theta} \left[\phi \left\{ \frac{(-z \sin \theta) \vee (\omega \cos \theta - v)}{h} \right\} \right. \\ & \left. + \phi \left\{ \frac{(-z \sin \theta) \wedge (\omega \cos \theta - v)}{h} \right\} - \phi \left(\frac{-z \sin \theta + \omega \cos \theta - v}{h} \right) - \phi \left(\frac{v}{h} \right) \right]. \end{aligned}$$

For a given $\theta \in [-\pi/2, \pi/2]$, we calculated $\hat{\lambda}_{h, \mathcal{L}_\theta^\perp}(\mathbf{s}_i)$ by (6) as

$$\hat{\lambda}_{h, \mathcal{L}_\theta^\perp}(v) = C_{h, \mathcal{L}_\theta^\perp}^{-1}(v) \sum_{i=1}^n \frac{1}{h} \phi \left(\frac{y_i \cos \theta - x_i \sin \theta - v}{h} \right) \quad (12)$$

for all $v \in [(-z \sin \theta) \wedge 0, \cos \theta + (-z \sin \theta) \vee 0]$, where x_i and y_i are horizontal and vertical coordinates of \mathbf{s}_i , respectively. We defined $\mathcal{Q} = \{\theta : \mathbf{Q}_\theta\}$ in the implementation of (10), where $\mathbf{Q}_\theta \mathbf{s} = y \cos \theta - x \sin \theta$ was an orthogonal project from \mathbb{R}^2 to \mathcal{L}_θ . The estimator $\hat{\theta}_h$ was the value of θ corresponding to $\hat{\mathbf{Q}}_h$ given by (10). We calculated the value of $\hat{\lambda}_{h, \hat{\mathcal{L}}^\perp}(v)$ with $\hat{\mathcal{L}} = \mathcal{L}_{\hat{\theta}_h}$. It was treated as the estimator of $\lambda(\mathbf{s})$ under substationarity with an unknown \mathcal{L} . It was compared with $\hat{\lambda}_{h, \mathcal{L}^\perp}(v)$, the estimator of $\lambda(\mathbf{s})$ with a known \mathcal{L} . In the end, we calculated $\hat{\theta}_h$ by (10) and (12).

We evaluated the performance of the MSE of $\hat{\theta}_h$ and the MISE of $\hat{\lambda}_{h, \mathcal{L}_\theta^\perp}(v)$ for selected a , z , and h . The performance of $\hat{\lambda}_{h, \mathcal{L}_\theta^\perp}(v)$ was compared with that of $\hat{\lambda}_h(\mathbf{s})$ given by (5) and $\hat{\lambda}$ under stationarity given by (8), where we chose $K(\mathbf{s})$ as the density of the standard bivariate normal distribution in the computation of $\hat{\lambda}_h(\mathbf{s})$.

We simulated 1000 realizations for each selected cases. To evaluate the performance of $\hat{\theta}_h$, we computed its MSE value by

$$\frac{1}{1000} \sum_{i=1}^{1000} \hat{\theta}_{hi}^2,$$

where $\hat{\theta}_{hi}$ was the value of $\hat{\theta}_h$ in the i th realization; see Table 1. The results of other cases showed that the root MSE of $\hat{\theta}_h$ were all close to 0, indicating that the estimator was accurate. The MSE of $\hat{\theta}_h$ decreased as z increased. This

Table 1: Simulations (with 1000 replications) for root MSE of $\hat{\theta}_h$ (given by degrees) with respect to selected a , z , r_0 and h for Poisson and Poisson cluster processes.

a	z	h for Poisson Process				h for Poisson Cluster Process			
		0.01	0.02	0.05	0.1	0.01	0.02	0.05	0.1
1.5	1	4.16	4.31	4.70	4.98	4.31	4.53	5.00	5.31
	2	4.00	3.75	3.70	3.73	4.17	4.20	4.55	4.77
	5	3.32	2.13	1.14	0.94	4.34	4.17	3.78	3.24
	10	1.63	0.43	0.29	0.32	4.84	4.22	2.71	1.62
2.0	1	3.88	4.16	4.34	4.41	4.20	4.33	4.84	5.13
	2	3.34	3.10	2.58	2.43	3.92	4.02	4.21	4.32
	5	1.12	0.76	0.51	0.54	3.32	2.96	2.20	1.60
	10	0.29	0.21	0.17	0.18	2.31	1.39	0.61	0.51
2.5	1	3.83	3.96	4.19	4.13	4.14	4.34	4.77	5.07
	2	3.04	2.59	1.96	1.84	3.78	3.89	3.90	3.81
	5	0.76	0.59	0.43	0.41	2.33	2.05	1.77	1.18
	10	0.23	0.17	0.13	0.15	0.63	0.56	0.39	0.38
3.0	1	3.72	3.94	3.83	3.77	4.11	4.29	4.73	4.93
	2	2.60	2.32	1.64	1.56	3.62	3.67	3.68	3.63
	5	0.64	0.50	0.36	0.37	1.75	1.50	1.13	0.99
	10	0.21	0.16	0.12	0.13	0.53	0.43	0.32	0.32

was interpreted by Theorem 4. The MSE decreased as a increased since the strength of nonstationarity increased as a became large. For the same a and z values, the MSE of $\hat{\theta}_h$ was also affected by the bandwidth h in the kernel approach is always an important issue to be investigated. In all the cases that we studied, the MSE of $\hat{\theta}_h$ in the Poisson SPPs was lower than those in the Poisson cluster SPPs. This was expected as for the same κ value, the number of independent clusters in the Poisson cluster SPPs was lower than the number of independent points in the Poisson SPPs.

We did not include the case when $a = 1$ in Table 1 as θ_0 was not well defined. As it induced $r_0 = 2$ and $\mathcal{L}_0 = \mathbb{R}^2$, we always had $\hat{\mathcal{L}}_h \subseteq \mathcal{L}_0$. To evaluate the impact of r_0 , we examined the performance of $\ell(\hat{\lambda}_{h,\mathcal{L}^\perp})$ as θ varied; see Figure 2. A strange pattern was discovered when $a = 1$ as it had $r = 1 < r_0 = 2$. All of the remaining cases had a unique maximum as we had $r = r_0 = 1$. Findings of this sort provided us a method to diagnose whether $r = r_0$.

We also evaluated the performance of four different estimators of the first-order intensity functions; see Table 2. We used $\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$ to represent the case when θ was known. We used $\hat{\lambda}_{h,\hat{\mathcal{L}}^\perp}(y)$ to represent the case when θ was unknown. We used $\hat{\lambda}_h(s)$ to represent the case when substationarity was not taken into account. We used $\hat{\lambda}$ to represent the case when stationarity was assumed. All of the minimum MISEs (mean integrated square errors) were reached by $\hat{\lambda}$ when $a = 1$ as the SPPs were stationary in this case. The root MISEs of $\hat{\lambda}$ increased in a since the strength of nonstationarity became large as a increased. For the same a and h values, the root MISEs of $\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$ and $\hat{\lambda}_{h,\hat{\mathcal{L}}^\perp}(y)$ decreased in z . We interpreted this by Theorem 4. The root MISEs of $\hat{\lambda}_h(s)$ did not vary significantly as z changed since the size of the region was not a critical issue in its computation. For all of the cases with $a > 1$ that we studied, the root MISEs of $\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$ and $\hat{\lambda}_{h,\hat{\mathcal{L}}^\perp}(y)$ were lower than those of $\hat{\lambda}_h(s)$ and $\hat{\lambda}$, indicating that efficiency was gained by accounting for substationarity. The root MISEs of $\hat{\lambda}$ diverged as z increased because the SPPs was not stationary when $a \neq 1$.

6. Application

We applied our approach to the Alberta Forest Wildfire data. The Alberta Forest Wildfire data consisted of forest wildfire activities that took place in Alberta, Canada, from 1931 to 2012. The Canadian Alberta Forest Service initiated the modern era of wildfire record starting in 1931. Since 1996, paper-based wildfire information was no longer retained. The wildfire historical data were entered at the field level on the Fire Information Resource Evaluation System (FIREs), which can be freely downloaded from the internet. We collected the historical forest wildfire data from 1996 to 2010 within a rectangle spanned from 117 longitude West to 110 longitude West in the horizontal direction and from 54.7 latitude North to 58 latitude North in the vertical direction; see Figure 3(a). We treated the

Table 2: Simulations (with 1000 replications) of root MISE of $\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$, $\hat{\lambda}_{h,\mathcal{L}^\perp_{\theta_h}}(v)$, $\hat{\lambda}_h(s)$, and $\hat{\lambda}$ with respect to selected a , z , and h in the Poisson and Poisson cluster processes.

a	z	h	Poisson Process				Poisson Cluster Process			
			$\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$	$\hat{\lambda}_{h,\mathcal{L}^\perp_{\theta_h}}(v)$	$\hat{\lambda}_h(s)$	$\hat{\lambda}$	$\hat{\lambda}_{h,\mathcal{L}^\perp}(y)$	$\hat{\lambda}_{h,\mathcal{L}^\perp_{\theta_h}}(v)$	$\hat{\lambda}_h(s)$	$\hat{\lambda}$
1	1	0.01	42.69	47.20	158.63	7.89	81.99	88.57	179.06	17.68
		0.02	30.28	33.31	110.09	8.23	64.44	70.07	157.49	18.32
		0.05	19.57	20.87	48.12	7.97	43.17	45.92	98.38	18.15
		0.10	14.50	14.93	25.74	7.96	32.41	33.75	56.77	18.54
	2	0.01	30.23	34.73	158.14	5.85	57.54	67.03	179.68	12.39
		0.02	21.33	24.62	109.37	5.65	45.77	52.51	156.55	11.96
		0.05	13.89	15.48	47.08	5.73	30.91	34.92	97.86	12.54
		0.10	10.09	10.79	24.90	5.62	22.74	24.60	54.99	12.76
	5	0.01	19.04	23.81	158.07	3.71	36.75	47.05	179.69	8.00
		0.02	13.68	17.23	109.28	3.51	29.01	36.83	156.05	8.28
		0.05	8.70	10.93	46.76	3.49	19.12	24.30	96.51	7.79
		0.10	6.40	7.56	24.31	3.49	14.23	17.16	53.56	7.73
	10	0.01	13.46	19.00	157.94	2.47	26.34	37.49	179.55	5.76
		0.02	9.67	13.63	109.05	2.45	20.60	29.39	156.09	5.52
		0.05	6.16	8.87	46.59	2.49	13.73	19.91	96.34	5.60
		0.10	4.48	6.01	24.14	2.57	10.17	14.14	53.25	5.83
	2	0.01	40.85	45.03	153.49	37.09	77.57	83.74	176.81	41.48
		0.02	29.04	31.95	102.77	39.72	60.99	65.82	151.79	41.41
		0.05	19.12	20.57	45.67	38.95	41.15	43.98	92.52	41.58
		0.10	17.14	17.75	26.86	39.14	31.92	33.00	53.78	41.57
	2	0.01	28.76	33.75	122.42	38.74	56.00	63.74	176.78	40.11
		0.02	20.58	24.17	82.07	38.84	43.68	50.73	151.36	40.05
		0.05	14.11	15.85	45.01	38.72	29.24	33.76	91.24	39.98
		0.10	13.62	14.74	26.11	38.71	23.91	26.01	52.58	40.10
	5	0.01	18.20	21.53	153.05	38.60	35.61	45.99	176.88	39.10
		0.02	13.14	15.87	102.69	38.60	27.43	36.15	150.76	39.14
		0.05	9.35	9.96	44.51	38.59	18.93	23.49	90.86	39.01
		0.10	11.12	11.57	25.70	38.61	16.99	19.00	51.78	39.04
	10	0.01	12.90	14.38	153.20	38.52	25.30	34.22	176.78	38.77
		0.02	9.21	11.03	102.62	38.53	19.76	25.11	150.76	38.83
		0.05	7.07	7.41	44.28	38.52	14.12	16.13	90.52	38.76
		0.10	9.25	10.19	25.52	38.54	13.79	14.70	51.72	38.76
3	1	0.01	39.55	42.57	148.66	59.06	73.23	79.67	174.60	60.47
		0.02	27.74	30.07	97.35	59.06	57.74	62.95	147.18	60.44
		0.05	18.07	19.74	42.77	59.06	38.45	41.24	86.89	60.45
		0.10	18.44	19.15	26.90	59.05	32.26	33.30	52.23	60.67
	2	0.01	27.37	31.53	148.40	58.87	52.97	60.94	174.22	59.62
		0.02	17.15	22.28	96.93	58.89	41.09	47.90	146.42	59.64
		0.05	12.90	14.20	42.00	58.85	27.79	31.93	85.44	59.66
		0.10	15.28	16.19	26.10	58.89	24.25	26.55	50.55	59.59
	5	0.01	17.23	19.52	147.97	58.78	33.58	40.71	174.06	59.07
		0.02	12.27	13.56	96.55	58.78	25.99	31.32	145.19	59.06
		0.05	8.80	9.32	41.79	58.77	17.90	20.87	85.27	59.04
		0.10	13.21	13.68	25.62	58.76	18.25	19.58	49.77	59.06
	10	0.01	12.34	13.58	148.18	58.73	23.86	27.81	174.15	58.86
		0.02	8.65	9.28	96.40	58.74	18.46	21.18	145.86	58.88
		0.05	6.71	7.02	41.55	58.74	13.13	14.34	84.74	58.87
		0.10	12.52	12.73	25.41	58.73	15.13	15.80	49.18	58.88

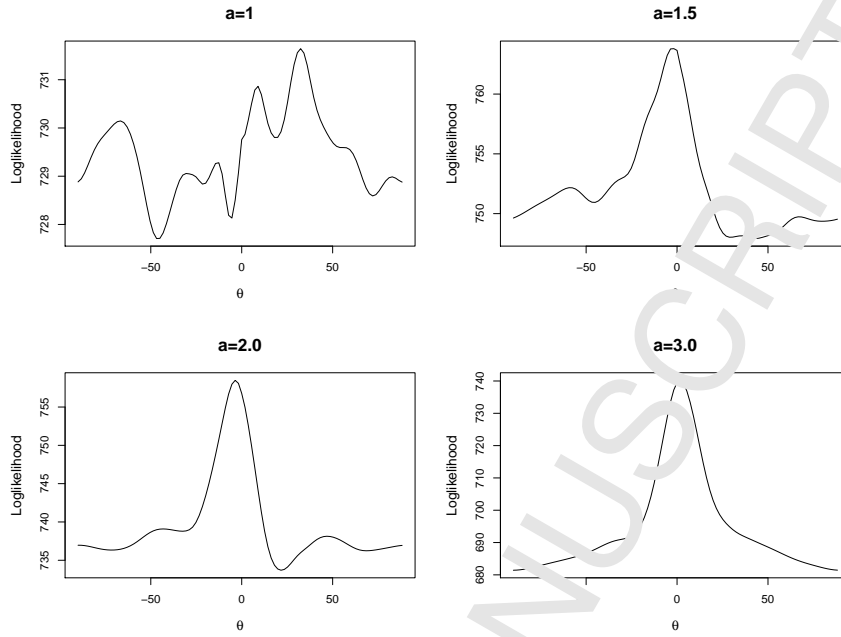


Figure 2: Plots of $\ell(\hat{\lambda}_{h,\mathcal{L}^\perp})$ as functions of θ varies when $h = 0.01$ based on realizations in the simulation.

rectangle as the study region in our approach. The region contained 8125 wildfire occurrences with all of the three greatest wildfires occurred in Alberta forests during the 15 year period. The greatest wildfire occurred in 2002 at 111.8 longitude West and 55.5 latitude North with an area burned of 2387 squared kilometers. The second greatest wildfire occurred in 1998 at 116.5 longitude West and 54.7 latitude North with an area burned of 1631 squared kilometers. The third greatest wildfire occurred in 1998 at 114.3 longitude West and 47.5 latitude West with an area burned of 1554 squared kilometers. The total burned area in the region was over 60% of the total burned area in the entire region.

The study region contained a large portion of boreal forests in Alberta, which was dominated in plain areas. The geographical distribution of boreal forest wildfires is considered as a major dominant disturbance in the high latitude area of the North Hemisphere [34]. A small portion of boreal forests of Alberta was in the mountain areas, located in the southwestern region of Alberta. We focused our study on the plain areas since tree densities and topographic conditions were significantly different between the mountain and plain areas.

We evaluated the scientific background about substationarity of wildfire occurrences in Alberta Forests. It was pointed out that wildfire activities in boreal forest were significantly affected by latitude but not by longitude [43]. It was expected to have low number of wildfire occurrences with high values of area burned in the north than those in the south [44], indicating that substationarity might be assumed along the longitude. To confirm this, we calculated the estimates of $\lambda(\mathbf{s})$ with the standard bivariate normal kernel via (5) under nonstationarity. We used a few bandwidth values and found that the results were not stable; see Figures 3(b), 3(c), and 3(d). However, all of our results showed that the estimates of the intensity were high in the south but low in the north.

We followed previous findings and assumed that fire occurrences were substationary in a linear space of \mathbb{R}^2 , where the linear space was $\mathcal{L} = \mathcal{L}_0$ given by (11). We computed $\hat{\theta}_h$ with a normal kernel in (10). We treated $\hat{\theta}_h$ as an estimator of θ for a given h . We compared values of $\hat{\theta}_h$ with various choices of h . We found that $\hat{\theta}_h$ was reliable. For instance, we got $\hat{\theta}_h = -0.002$ (given by arc degree) if $h = 0.01$, $\hat{\theta}_h = -0.001$ if $h = 0.02$, $\hat{\theta}_h = -0.003$ if $h = 0.05$, and $\hat{\theta}_h = -0.007$ if $h = 0.1$. Therefore, we had $\hat{\theta}_h \approx 0$, indicating that we might simply choose $\mathcal{L} = \mathcal{L}_0$ in our method. To investigate this, we compared $\ell\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s})\}$ and $\ell\{\hat{\lambda}_{h,\mathcal{L}_0^\perp}(\mathbf{s})\}$ with selected h in (9). We found that $\ell\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s})\} - \ell\{\hat{\lambda}_{h,\mathcal{L}_0^\perp}(\mathbf{s})\}$ were 1.66, 0.89, 3.38, and 7.05 when h was 0.01, 0.02, 0.05, and 0.1, respectively. Comparing these values with the differences of log-likelihood functions affected by h , which were often more than a few hundred, we concluded that the impact of $\ell\{\hat{\lambda}_{h,\mathcal{L}^\perp}(\mathbf{s})\} - \ell\{\hat{\lambda}_{h,\mathcal{L}_0^\perp}(\mathbf{s})\}$ could be ignored. Therefore, it was enough for us to use $\theta = 0$ in the computation of the

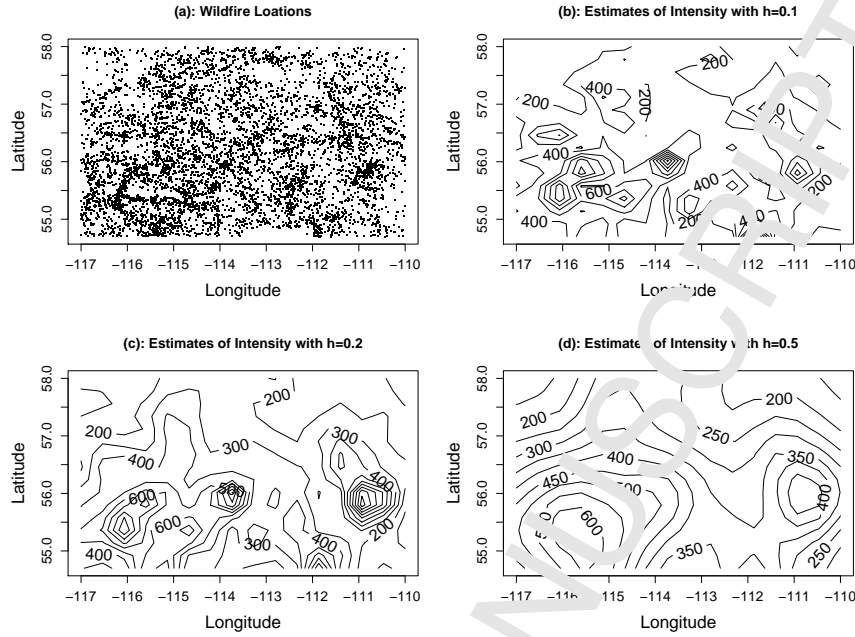


Figure 3: Wildfires locations and estimates of the intensity under nonstationarity in Alberta Forests from 1996 to 2010 in the selected region, where bandwidths were given by degrees.

estimator of the first-order intensity function.

Simply using $\theta = 0$, we obtained $\mathcal{L}_0 = \{(x, 0) : x \in \mathbb{R}\}$. We computed values of $\hat{\lambda}_{h, \mathcal{L}_0^\perp}(y)$ with a few choices of h . We found that all of them were close, e.g., as those displayed by Figure 4. Therefore, we concluded that our approach was reliable. The intensity of wildfire occurrences was almost maximized at 55.8 latitude North. It decreased fast to the north but slowly to the south. The north part was consistent with our previous conclusion but the south part was a concern. We studied the reason by looking at the terrestrial ecozones. We found that ecozones in the south of the study region was dominated by grassland, which might affect the occurrences of forest wildfires [33, 43].

7. Discussion

In this article, we propose the concept of substationarity and provide a semiparametric method to estimate the first-order intensity function of a spatial point process, including the linear subspace. Although this has not been previously studied in the literature in SPPs, a related concept in geostatistics called spheric symmetry that are stationary along the longitudes but nonstationary along latitudes has been previously discussed [24, 28, 29]. A significant difference between the two concepts is that we assume that the linear subspace is unknown. Therefore, we also need to estimate the linear subspace.

The estimation method proposed in this article is a modified version of the traditional kernel density estimation for random variables. Traditional kernel density estimation is formulated under the assumption that sampling data are identically and independently collected from a continuous distribution. This is often violated because of dependency in spatial point data. A common way to account for dependency in SPPs is to use the second-order intensity functions. As specific relationships between the first-order and the second-order intensity functions can be formulated under the concept of SOIRs, it is possible to account for both of them simultaneously.

Although we have only discussed the kernel-based approach, two other nonparametric or semiparametric approaches may be used. The local polynomial approach is a modified version of the kernel approach [5, 13]. It is based on the idea of weighted localized polynomial regression, where the weights are determined by kernel functions of explanatory variables. The smoothing spline approach estimates a smooth function by minimizing a penalized likelihood function [17, 30, 41]. The penalized likelihood function has two terms. The negative log-likelihood term

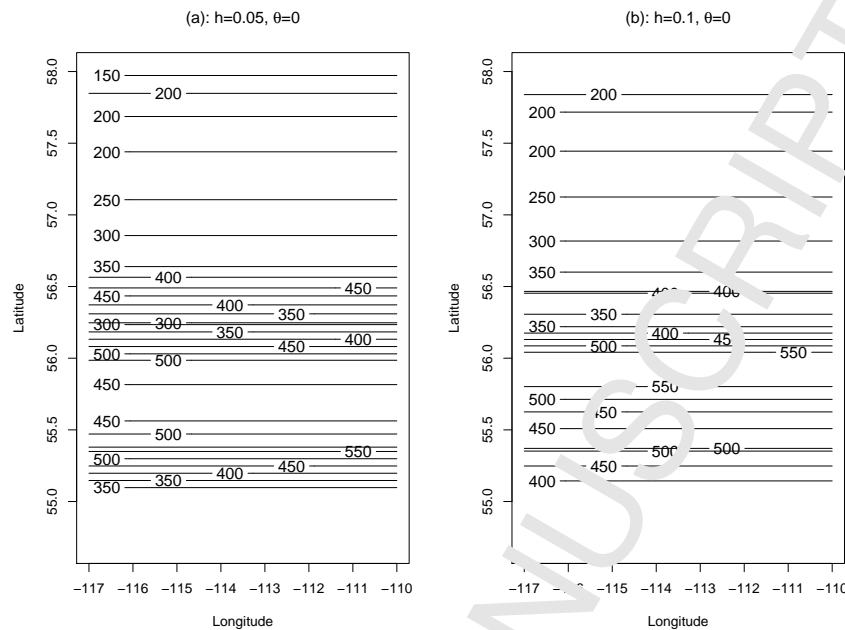


Figure 4: Estimates of the first-order intensity in the Alberta forest fire data under substationarity along the longitude.

controls the goodness-of-fit. The penalty term controls the smoothness. Both the local polynomial and the smooth spline approaches can be used to estimate intensity functions of SPPs under substationarity.

As a relative concept, nonsubstationarity is also important in spatial point data. A nonsubstationarity approach must be adopted if assumptions of substationarity fail. Based on the concept of substationarity, a few possible methods for nonsubstationarity may be proposed. An easy way is to borrow the idea of additive models in nonparametric statistics [15, 21]. Assume that the intensity functions of a nonstationary SPP can be expressed by the sum of intensity functions of a few substationary SPPs. If the intersection of the linear subspaces given by these substationary SPPs only contains the origin, then the nonstationary SPP is not substationary in any linear subspace of the domain. The structure of additive models for nonsubstationarity in SPPs is essentially different from the structure of additive models in nonparametric statistics. Additive models in SPPs attempt to model additivity by intensity functions. Additive models in nonparametric statistics attempt to model additivity by mean structures. Additive models in SPPs contain dependence structures but additive models in nonparametric statistics do not. This is an interesting research question to be studied in the future.

Acknowledgments

The authors appreciate great comments from the Editor-in-Chief, Prof. Christian Genest, an Associate Editor, and two anonymous referees that significantly improved the quality of the article.

References

- [1] A.J. Baddeley, J. Møller, R. Waagepetersen, Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, *Statist. Neerlandica* 54 (2000) 329–350.
- [2] M. Berman, J. Diggle, Estimating weighted integrals of the second-order intensity of a spatial point process, *J. R. Statist. Soc. B* 51 (1989) 81–92.
- [3] J. Besag, Contribution to the discussion of Dr. Ripley's paper, *J. R. Statist. Soc. B* 39 (1977) 193–195.
- [4] P. Billingsley, *Probability and Measure*, Wiley, New York, 1995.
- [5] W.S. Cleveland, S. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, *J. Amer. Statist. Assoc.* 83 (1988) 596–610.

- [6] S.N. Chiu, K.I. Liu, Stationarity tests for spatial point processes using discrepancies, *Biometrics* 69 (2013) 497–507.
- [7] N. Cressie, C.K. Wikle, *Statistics for Spatio-temporal Data*, Wiley, Hoboken, NJ, 2011.
- [8] P.J. Diggle, A kernel method for smoothing point process data, *Appl. Statist.* 34 (1985) 138–147.
- [9] P.J. Diggle, *Statistical Analysis of Spatial Point Patterns*, 2nd Edition, Wiley, New York, 2003.
- [10] P.J. Diggle, Spatio-temporal point processes, partial likelihood, foot and mouth disease, *Stat. Methods Med. Res.* 16 (2006) 325–336.
- [11] P.J. Diggle, G. Rubio, P.E. Brown, A.G. Chetwynd, S. Gooding, Second-order analysis of inhomogeneous spatial point processes using case-control data, *Biometrics* 63 (2007) 550–557.
- [12] R.M. Dudley, Fourier analysis of substationary processes with a finite moment, *Trans. Amer. Math. Soc.* 118 (1965) 360–375.
- [13] J. Fan, Local linear regression smoothers and their minimax efficiency, *Ann. Statist.* 21 (1993) 196–215.
- [14] T.S. Ferguson, *A Course in Large Sample Theory*, Chapman & Hall/CRC Press, Boca Raton, FL, 1996.
- [15] J.H. Friedman, W. Stuetzle, Projection pursuit regression, *J. Amer. Statist. Assoc.* 76 (1981) 811–823.
- [16] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics* 21 (1979) 215–223.
- [17] C. Gu, *Smoothing Spline ANOVA Models* 2nd Edition, Springer, New York, 2013.
- [18] Y. Guan, A KPSS test for stationarity for spatial point processes, *Biometrics* 64 (2008) 806–806.
- [19] Y. Guan, On nonparametric variance estimation for second-order statistics of inhomogeneous spatial point processes with known parametric intensity form, *J. Amer. Statist. Assoc.* 104 (2009) 1482–1491.
- [20] Y. Guan, Y. Shen, A weighted estimating equation approach for inhomogeneous spatial point processes, *Biometrika* 97 (2010) 867–880.
- [21] T.J. Hastie, R.J. Tibshirani, *Generalized Additive Models*, Chapman & Hall/CRC Press, Washington, DC, 1990.
- [22] P.A. Henrys, P.E. Brown, Inference for cluster inhomogeneous spatial point processes, *Biometrics* 65 (2009) 423–430.
- [23] N. Herrndorf, A functional central limit theorem for weakly dependent sequence of random variables, *Ann. Probab.* 12 (1994) 141–153.
- [24] M.T. Horrell, M.L. Stein, A covariance parameter estimation method for polar-orbiting satellite data, *Statistica Sinica* 25 (2015) 41–59.
- [25] I.A. Ibragimov, Some limit theorems for stationary processes, *Theory Probab. Appl.* 7 (1962) 171–186.
- [26] G. Ivanoff, Central limit theorems for point processes, *Stochastic Process. Appl.* 12 (1982) 171–186.
- [27] Y. Jiang, Q. Zhuang, Extreme value analysis of wildfires in Canadian boreal forest ecosystems, *Can. J. For. Res.* 41 (2011) 1836–1852.
- [28] R.H. Jones, Stochastic processes on a sphere, *Ann. Math. Statist.* 34 (1963) 213–218.
- [29] M. Jun, M.L. Stein, Nonstationary covariance models for global data, *Ann. Appl. Statist.* 2 (2008) 1271–1289.
- [30] G. Kimeldorf, G. Wahba, Some results on tchebycheffian spline functions, *J. Math. Anal. Appl.* 33 (1971) 82–94.
- [31] J. Møller, R.P. Waagepetersen, Modern statistics for spatial point processes (with discussion), *Scand. J. Statist.* 34 (2007) 685–711.
- [32] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.* 83 (1988) 9–27.
- [33] A.J. Pitman, G.T. Narisma, J. McAneney, The impact of climate change on the risk of forest and grassland fires in Australia, *Clim. Change* 84 (2007) 383–401.
- [34] J. Podur, D.L. Martell, K. Knight, Statistical quality control analysis of forest fire activity in Canada, *Can. J. For. Res.* 32 (2002) 195–205.
- [35] B.D. Ripley, The second-order analysis of spatial point processes, *J. Appl. Probab.* 23 (1976) 255–266.
- [36] M. Rosenblatt, A central limit theorem and a strong mixing condition, *Proc. Natl. Acad. Sci. USA.* 42 (1956) 43–47.
- [37] D. Stoyan, H. Stoyan, *Fractals, Random Shapes and Point Fields*, Wiley, New York, 1994.
- [38] D. Stoyan, H. Stoyan, Estimating pair correlation functions of planar cluster processes, *Biom. J.* 38 (1996) 259–271.
- [39] A.W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK, 1998.
- [40] R. Waagepetersen, An estimating function approach to inference for inhomogeneous Neyman-Scott process, *Biometrics* 63 (2007) 252–258.
- [41] G. Wahba, *Spline models for observational data*, CMS-NSF Regional Conference Series in Applied Mathematics, SIAM, 1990.
- [42] W. Whitt, Proofs of the martingale FCLT, *Probab. Surveys* 4 (2007) 268–302.
- [43] J. Xiao, Q. Zhuang, Drought effects on large fire activities in Canadian and Alaskan forests, *Environ. Res. Lett.* 2 (2007) 044003.
- [44] T. Zhang, Q. Zhuang, On the local odds ratio between points and marks in marked point processes, *Spatial Statist.* 9 (2014) 20–37.
- [45] T. Zhang, B. Zhou, Test for stationarity for spatial point processes in an arbitrary region, *J. Agric. Biol. Environ. Statis.* 19 (2014) 387–404.
- [46] T. Zhang, A Kolmogorov-Smirnov type test for independence between marks and points of marked point processes, *Electron. J. Statist.* 8 (2014) 2557–2584.
- [47] T. Zhang, On independence and separability between points and marks of marked point processes, *Statistica Sinica* 27 (2017) 207–227.
- [48] T. Zhang, R. Zhuang, Testing proportionality between the first-order intensity functions of spatial point processes, *J. Multivariate Anal.* 155 (2017) 72–82.