Gradient angle-based analysis for spatiotemporal point processes

Tonglin Zhang

Department of Statistics, Purdue University
250 North University Street, West Lafayette, IN 47907-2066 e-mail: tlzhang@purdue.edu

Yen-Ning Huang

Department of Statistics, Indiana University, 1020 E Kirkwood Ave, Bloomington, IN 47405

Abstract: Spatiotemporal point processes (STPPs) are important in modeling randomly appeared events developed in space and time. Statistical methods of STPPs have been widely used in applications. In all of these methods, evaluations and inferences of intensity functions are the primary issues. The present article proposes a new method, which attempts to evaluate angles of gradient vectors of intensity functions rather than the intensity functions themselves. According to the nature of many natural and human phenomena, the evaluation of angle patterns of the gradient vectors is more important than the evaluation of their magnitude patterns because changes of angle patterns often indicate global changes of these phenomena. This issue is investigated by simulation studies, where significant variations of gradient angle patterns are identified only when modes of intensity functions are changed. To study these phenomena, the article proposes an analysis method for gradient angles of the first-order intensity function of STPPs. The proposed method is used to analyze aftershock earthquake activities caused by great mainshock earthquakes occurred in Japan 2011 and Indian Ocean 2004, respectively, where a significant global change in the second case is identified.

1. Introduction

The goal of the research is to develop a gradient-based approach to spatiotemporal point processes (STPPs) which can be used to describe the global trend of point occurrences. To develop the approach, it is important to have a statistical way which can characterize the tendency of gradient vectors of the intensity functions. It will point out that the usage of directions of gradient vectors is more important than the usage of their magnitudes. Based on this point, we propose a new statistical approach to gradient angle patterns of the first-order intensity function of STPPs. An important assumption is the angle invariant condition,

which means that the angle (or the direction) of gradient vectors of the first-order intensity function in the whole study area is time invariant. The angle invariant condition is useful in modeling many natural or human phenomena. For instance, the frequently used ETAS (epidemic type aftershock sequences) model for aftershock earthquakes [26, 40, 27] assumes that the occurrences of aftershock earthquakes are symmetric about their mainshock earthquake center, which satisfies the angle invariant condition. A similar issue has also been pointed out in infectious disease studies since the spread of an infectious disease is roughly symmetric about the original occurrence center [24]. In these examples, the direction of the tendency is more important than the magnitude since the change of directions represents a global change while the change of magnitudes only represents a local change. This issue has been justified in our simulation studies. We find that gradient angle patterns have significant changes only when the mode of the intensity functions has been changed. A mode change is often caused by the change of the intensity functions in the entire study region. Therefore, the statistical inference on angle patterns of gradient vectors of intensity functions is more important than the inference on the whole gradient vectors themselves. This motivates us to study gradient angle-based analysis methods in our research.

An STPP is a collection of random locations of points appeared in their spatiotemporal domain. Specific applications in the development of statistical approaches to spatial point processes (SPPs) or STPPs have appeared in many disciplines, including forestry [17], forest wildfires [28, 31, 38], earthquakes [19], and spatial epidemiology [11]. Many concepts and methods for SPPs have been previously proposed. Examples include the K-function [29], the L-function [9], the pair correlation function [34], stationarity tests [15, 39], nonstationary analyses [1, 4], likelihood and composite likelihood analyses [8, 32, 36], first-order analyses [17, 37], and second-order analyses [16]. However, none of them have studied gradient-based properties of spatial or spatiotemporal point patterns. The present article is the first one considering such an issue.

Although we only focus on STPPs, we believe that gradient-based methods are also important in the analysis SPPs. Gradients belong to the first-order properties. They describe changes of the first-order intensity function locally. However, gradient angles can also provide the global trend. According to insights from our simulations for STTP to be displayed in Section 5, we conclude that nonparametric estimation of angle patterns is more reliable than that of magnitude patterns in terms of the bandwidth selection. Understanding of angle patterns is more critical than understanding of magnitude or entire intensity patterns. We investigate possible impacts of our methods on invasive species problems under the framework of SPPs. In ecology, the invasive species problem describes the spread of a certain plant or animal that is not native to a specific region, which is believed to cause damage to the local environment or human society [2]. If locations of invasive species are provided by a spatial point pattern, then its magnitude pattern only provides the local trend of the invasion but the angle pattern can also provide the source center. Thus, the analysis of angle pattern is more important than the analysis of magnitude pattern.

We have encountered many new challenges in the development of methods for gradient patterns of STPPs. Our proposed method is different from those in the literature in a variety of ways. First, in most articles of STPPs in the literature, the interest is to investigate the mean and covariance structures, which are often described by the first-order intensity and pair correlation functions. However in our method, the interest is to investigate the gradient patterns of STPPs, which can be further specified by the magnitudes and angles of gradient vectors. Second, unlike the scenarios considered in the classical methods, the direction of a gradient vector may not be available in a stationary process since the gradient of the first-order intensity function is zero everywhere. Thus, our method is developed for a nonstationary process only, implying that it is not recommended analyzing stationary spatiotemporal point patterns. Third, for nonstationary spatiotemporal point patterns, the change of angle patterns of gradient vectors of intensity functions is more interesting than the only change of magnitude patterns since the previous one represents a global change while the latter one represents a local change. Our simulations support the conclusion that slight local changes of intensity functions do not affect the gradient angle patterns significantly. Therefore, we can use our method to study whether there is a global change. It is possible for us to conclude the existence of a global change if gradient angle patterns are significantly changed. Since the approach relies on the derivation of angles of gradient vectors, we propose a kernel-based estimator to compute them. We provide asymptotic properties of our estimator. We carry out simulations to study its performance. Our method is applied to study the gradient patterns of aftershock earthquakes in the 2011 Japan *Great Tohoku Earthquake* data and the 2004 Indian Ocean *Great Sumatra-Andaman Earthquake* data. We conclude that the angle patterns are almost invariant in the *Great Tohoku Earthquake* data but not in *Great Sumatra-Andaman Earthquake* data.

The statistical method proposed in this article is motivated from previous gradient-based techniques which have been extensively used in image processing and pattern recognition. Overall, estimation of the gradient pattern of an image is a fundamental task in any field of object recognition, where the earliest work appeared about forty years ago [12]. After that, gradient-based methods have been extensively used in pattern recognition [25]. Basic gradient-based criteria, which are detection, localization, and uniqueness of recognition, have been proposed [5]. It is well acknowledged that images can be easily visualized if a gradient-based method is used. For example, images can be segmented into visually sensible regions by finding watershed regions in a gradient magnitude image [23]. Multiscale analysis of intensity extrema in the gradient magnitude image can be used to define a hierarchy on the corresponding watershed regions [13]. In addition to pattern recognition, gradient estimation has been investigated in nonparametric statistics [6], where the main interest is estimation of partial derivatives (i.e., gradients) of a smoothness function but not the smoothness function itself. Although gradient-based approaches have been used in other fields, this kind of approaches has never been considered in the literature of spatial or spatiotemporal point processes.

The rest of the article is organized as follows. In Section 2, we propose a few important concepts and properties of the first-order intensity function of STPPs based on its gradient angles. In Section 3, we propose nonparametric kernel estimation of angles of gradient vectors of the first-order intensity function. In Section 4, we provide the asymptotic properties of our estimator, which includes the consistency, the asymptotic normality, and the bandwidth selection. In Section 5, we carry out a simulation study to evaluate the properties of our estimator. In Section 6, we apply our method to study the angle patterns of the *Great Tohoku Earthquake* data and the *Great Sumatra-Andaman Earthquake* data. In Section 7, we provide a discussion. The proofs of our asymptotic properties are displayed in the Appendix.

2. Gradient angles of the first-order intensity function

Let N be an STPP on $\mathcal{D} \times \mathcal{T}$, where $\mathcal{D} \in \mathscr{B}(\mathbb{R}^2)$ is the spatial domain, $\mathcal{T} \in \mathscr{B}(\mathbb{R})$ is the temporal domain, and $\mathscr{B}(A)$ represents the collection of Borel subsets of A. Let $N(A \times B)$ be the number of points in $A \times B$, where $A \in \mathscr{B}(\mathcal{D})$ and $B \in \mathscr{B}(\mathcal{T})$. Then, $N(A \times B)$ is finite if A and B are bounded. For a fixed $B \in \mathscr{B}(\mathcal{T})$, we can define a *collapsed* SPP denoted by $N_B(\cdot)$ with $N_B(A) = N(A \times B)$ for any $A \in \mathscr{B}(\mathcal{D})$. Throughout this article, we assume that $\mathcal{D}$ is a bounded two-dimensional manifold with $|\partial \mathcal{D}| = 0$ and $\mathcal{T}$ is a bounded interval, where ∂ is the boundary operation and $|A|$ means the Lebesgue measure of A in $\mathbb{R}^2$. For convenience, we assume $\mathcal{T} = [0, T]$ with $T \in \mathbb{R}^+$ and write $\mathbf{s} = (x, y)$ and $\mathbf{s}_i = (x_i, y_i)$ for any $\mathbf{s}, \mathbf{s}_i \in \mathcal{D}$.

Let the kth-order intensity function of N be denoted by $\lambda_k((\mathbf{s}_1, t_1), \cdots, (\mathbf{s}_k, t_k))$, $k \in \mathbb{N}^+$. Write the first-order intensity function of N as $\lambda(\mathbf{s}, t) = \lambda_1(\mathbf{s}, t)$. For any $A, A_1, A_2 \in \mathscr{B}(\mathcal{D})$ and $B, B_1, B_2 \in \mathscr{B}(\mathcal{T})$, there are

$$\mathrm{E}[N(A \times B)] = \int_B \int_A \lambda(\mathbf{s}, t) d\mathbf{s} dt,$$

and

$$\begin{aligned}
&\mathrm{Cov}[N(A_1 \times B_1), N(A_2 \times B_2)] \\
&= \int_{B_2} \int_{A_2} \int_{B_1} \int_{A_1} [g((\mathbf{s}, t), (\mathbf{s}', t')) - 1] \lambda(\mathbf{s}, t) \lambda(\mathbf{s}', t') d\mathbf{s} dt d\mathbf{s}' dt' \\
&\quad + \int_{B_1 \cap B_2} \int_{A_1 \cap A_2} \lambda(\mathbf{s}, t) d\mathbf{s} dt,
\end{aligned}$$

where $g((\mathbf{s}, t), (\mathbf{s}', t')) = \lambda_2((\mathbf{s}, t), (\mathbf{s}', t'))/[\lambda(\mathbf{s}, t)\lambda(\mathbf{s}', t')]$ if $\lambda(\mathbf{s}, t)\lambda(\mathbf{s}', t') > 0$ or $g((\mathbf{s}, t), (\mathbf{s}', t')) = 1$ if $\lambda(\mathbf{s}, t)\lambda(\mathbf{s}', t') = 0$ is the pair correlation function.

Assume $\lambda(\mathbf{s}, t)$ is continuous differentiable at any interior point $(\mathbf{s}, t)$ in $\mathcal{D} \times \mathcal{T}$. Let

$$\nabla_s \lambda(\mathbf{s}, t) = (\nabla_1 \lambda(\mathbf{s}, t), \nabla_2 \lambda(\mathbf{s}, t)) = \left(\frac{\partial \lambda(\mathbf{s}, t)}{\partial x}, \frac{\partial \lambda(\mathbf{s}, t)}{\partial y} \right) \qquad (2.1)$$

be the gradient of the first-order intensity function projected to the spatial domain. If $\|\nabla_s \lambda(\mathbf{s}, t)\| > 0$, then there exists a unique function $\theta(\mathbf{s}, t) \in [-\pi, \pi)$ such that

$$\cos \theta(\mathbf{s}, t) = \frac{\nabla_1 \lambda(\mathbf{s}, t)}{\|\nabla_s \lambda(\mathbf{s}, t)\|} \tag{2.2}$$

and

$$\sin \theta(\mathbf{s}, t) = \frac{\nabla_2 \lambda(\mathbf{s}, t)}{\|\nabla_s \lambda(\mathbf{s}, t)\|}. \tag{2.3}$$

The function $\theta(\mathbf{s}, t)$ represents the spatial angle patterns of $\lambda(\mathbf{s}, t)$, which can be used to describe the directions of the gradient vectors of $\lambda(\mathbf{s}, t)$ in the spatial domain.

It is clear that $\theta(\mathbf{s}, t)$ is not well-defined if $\|\nabla_s \lambda(\mathbf{s}, t)\| = 0$, which is attained if $(\mathbf{s}, t)$ is a spatial stationary point of $\lambda(\mathbf{s}, t)$ (i.e. point $(\mathbf{s}, t)$ satisfying $\nabla_s \lambda(\mathbf{s}, t) = \mathbf{0}$). We say spatial stationary points of $\lambda(\mathbf{s}, t)$ are isolated if there exists a $\delta > 0$ such that $\|\mathbf{s}_1 - \mathbf{s}_2\| > \delta$ for any distinct spatial stationary points $(\mathbf{s}_1, t)$ and $(\mathbf{s}_2, t)$ of $\lambda(\mathbf{s}, t)$. If spatial stationary points of $\lambda(\mathbf{s}, t)$ are isolated, then for a given interior point $\mathbf{s}_0 \in \mathcal{D}$ we can define $\theta(\mathbf{s}, t)$ for any $\mathbf{s} \neq \mathbf{s}_0$ in a sufficiently small neighborhood of $\mathbf{s}_0$, indicating that gradient angles of an STPP are almost surely available. We do not expect that our method can be used if $\lambda(\mathbf{s}, t)$ is only a function of t in the entire spatial domain. As $\nabla_s \lambda(\mathbf{s}, t) = \mathbf{0}$ everywhere, we cannot define $\theta(\mathbf{s}, t)$ at any $\mathbf{s} \in \mathcal{D}$. Therefore, our method is developed for a spatially nonstationary STPP.

Definition 1. *The STPP N is called angle invariant at* $\mathbf{s} \in \mathcal{D}$ *with* $\|\nabla_s \lambda(\mathbf{s}, t)\| > 0$ *if* $\theta(\mathbf{s}, t)$ *is invariant in t for any* $t \in \mathcal{T}$. *The STPP N is called angle invariant if* $\theta(\mathbf{s}, t)$ *is angle invariant at every interior point* $\mathbf{s} \in \mathcal{S}$ *with nonzero* $\|\nabla_s \lambda(\mathbf{s}, t)\|$ *values. We write* $\theta(\mathbf{s}) = \theta(\mathbf{s}, t)$ *for any interior point* $\mathbf{s} \in \mathcal{D}$ *with nonzero* $\|\nabla_s \lambda(\mathbf{s}, t)\|$ *values. We call* $\lambda(\mathbf{s}, t)$ *angle invariant if N is angle invariant.*

Let $\tilde{N}$ be an SPP on $\mathcal{D}$ with its first-order intensity function denoted by $\tilde{\lambda}(\mathbf{s})$. Assume $\tilde{\lambda}(\mathbf{s})$ is continuous differentiable at any interior point $\mathbf{s} \in \mathcal{D}$. Let

$$\nabla \tilde{\lambda}(\mathbf{s}) = (\nabla_1 \tilde{\lambda}(\mathbf{s}), \nabla_2 \tilde{\lambda}(\mathbf{s})) = \left(\frac{\partial \tilde{\lambda}(\mathbf{s})}{\partial x}, \frac{\partial \tilde{\lambda}(\mathbf{s})}{\partial y} \right)$$

be the gradient vector of $\tilde{\lambda}(\mathbf{s})$. Then, $\nabla \tilde{\lambda}(\mathbf{s})$ exists if $\mathbf{s}$ is an interior point of $\mathcal{D}$. The angle function $\tilde{\theta}(\mathbf{s})$ of $\nabla \tilde{\lambda}(\mathbf{s})$ can be defined similarly as those in (2.2) and (2.3). If $\nabla \tilde{\lambda}(\mathbf{s}) \neq \mathbf{0}$ (i.e., not a stationary point), then it is uniquely determined by

$$\cos \tilde{\theta}(\mathbf{s}) = \frac{\nabla_1 \tilde{\lambda}(\mathbf{s})}{\|\nabla \tilde{\lambda}(\mathbf{s})\|}$$

and

$$\sin \tilde{\theta}(\mathbf{s}) = \frac{\nabla_2 \tilde{\lambda}(\mathbf{s})}{\|\nabla \tilde{\lambda}(\mathbf{s})\|}.$$

Let $\tilde{\lambda}(\mathbf{s})$ and $\tilde{\lambda}'(\mathbf{s})$ be the first-order intensity functions of two SPPs. Let $\tilde{\theta}(\mathbf{s})$ and $\tilde{\theta}'(\mathbf{s})$ be the angle functions of $\nabla\tilde{\lambda}(\mathbf{s})$ and $\nabla\tilde{\lambda}'(\mathbf{s})$, respectively. We say the two functions $\tilde{\theta}$ and $\tilde{\theta}'$ are equal if $\tilde{\theta}(\mathbf{s}) = \tilde{\theta}'(\mathbf{s})$ at any interior point $\mathbf{s} \in \mathcal{D}$ when none of $\|\nabla\tilde{\lambda}(\mathbf{s})\|$ and $\|\nabla\tilde{\lambda}'(\mathbf{s})\|$ is zero. We define that the angle functions of $\nabla\tilde{\lambda}(\mathbf{s})$ and $\nabla\tilde{\lambda}'(\mathbf{s})$ are equal at $\mathbf{s}$ if one of $\|\nabla\tilde{\lambda}(\mathbf{s})\|$ and $\|\nabla\tilde{\lambda}'(\mathbf{s})\|$ is zero. Then, we provide the following definition.

Definition 2. *Let $\tilde{N}$ and $\tilde{N}'$ be two distinct SPPs on $\mathcal{D}$ with their first-order intensity functions denoted by $\tilde{\lambda}(\mathbf{s})$ and $\tilde{\lambda}'(\mathbf{s})$, respectively. Assume $\tilde{\lambda}(\mathbf{s})$ and $\tilde{\lambda}'(\mathbf{s})$ are continuously differentiable. We say that $\tilde{N}$ and $\tilde{N}'$ have similar (first-order) patterns if the angle functions of $\nabla\tilde{\lambda}(\mathbf{s})$ and $\nabla\tilde{\lambda}'(\mathbf{s})$ are equal at every interior point $\mathbf{s} \in \mathcal{D}$.*

If stationary points of $\tilde{\lambda}(\mathbf{s})$ and $\nabla\tilde{\lambda}'(\mathbf{s})$ are isolated, then they do not affect the concept of similarity given by Definition 2. This provides the following theorem.

Theorem 1. *Let N be an STPP on $\mathcal{D} \times \mathcal{T}$ with a continuous differentiable $\lambda(\mathbf{s}, t)$, whose stationary points are isolated. Then, N is angle invariant if and only if N_{B_1} and N_{B_2} have similar patterns for any subintervals $B_1, B_2 \subseteq \mathcal{T}$.*

Theorem 1 provides a way to study the angle invariant property of the first-order intensity function of an STPP. It is useful if an STPP only has one or a few stationary points. An example is the case when $\lambda(\mathbf{s}, t)$ is derived from an ETAS model (to be discussed in Example 3 below). Since $\lambda(\mathbf{s}, t)$ only has one stationary point, which is its mode, the angle functions can be studied at any points other than the mode. It is expected to see similar patterns of the first-order intensity function if the STPP is compared among different time periods. We provide a few examples to illustrate the issues discussed in Definition 2 and Theorem 1.

Example 1 (First-order separable model)**.** If the STPP N has a separable first-order intensity function such that $\lambda(\mathbf{s}, t) = \lambda_s(\mathbf{s})\lambda_t(t)$ for any $(\mathbf{s}, t) \in \mathcal{D} \times \mathcal{T}$, then $\nabla_s\lambda(\mathbf{s}, t) = \lambda_t(t)\nabla\lambda_s(\mathbf{s})$. Therefore, $\theta(\mathbf{s}, t)$ does not depend on t implying that N is angle invariant.

Example 2 (Condensed model)**.** Assume the first-order intensity function of N can be written as $\lambda(\mathbf{s}, t) = F[f(\mathbf{s}), t] = F[f(x, y), t]$, where f is a bivariate function and F is a positive bivariate function. Then,

$$\nabla_s\lambda(\mathbf{s}, t) = F_1[f(\mathbf{s}), t]\left(\frac{\partial f(\mathbf{s})}{\partial x}, \frac{\partial f(\mathbf{s})}{\partial y}\right),$$

where F_1 is the first component of the partial derivative vector of F. Therefore, $\theta(\mathbf{s}, t)$ does not depend on t implying that N is angle invariant. A few interesting cases can be derived. If $f(x, y) = x^2 + y^2$ and $F(u, v)$ is decreasing in u, then an isotropic model is derived. If $f(x, y) = ax^2 + 2bxy + cy^2$ with $b^2 - ac < 0$ and $a, c > 0$, then an elliptical model is derived. Both models are angle invariant.

Example 3 (ETAS model)**.** The ETAS model is a spatiotemporal marked point process model representing the activities of earthquakes of a certain magnitude

and larger in a region during a period of time [26, 27, 40]. The model includes background activities and aftershock clusters. Events in each aftershock cluster are independently produced by the mainshock earthquake. The size of the aftershock cluster is related to the magnitude of the mainshock earthquake. The ETAS model is composed of the background rates and the intensity functions derived from all of the mainshock earthquakes and their aftershock clusters. If there is only one extremely large mainshock earthquake with magnitude M^* at space-time location $(\mathbf{s}^*, t^*)$, then the first-order intensity function of the locations of aftershock earthquakes can be approximately expressed as

$$\lambda(\mathbf{s}, t) = \kappa(M^*) u(t - t^*) v(\mathbf{s} - \mathbf{s}^* | M^*). \tag{2.4}$$

It is clear that $\lambda(\mathbf{s}, t)$ given by (2.4) is angle invariant. Model (2.4) is useful in the study of aftershock earthquake activities produced by an extremely large earthquake. The reason is because the impact of other mainshock earthquakes and the impact of background earthquakes can be roughly ignored within a few months after its occurrence. The ETAS model will be used with more details in Section 6.

3. Estimation

We propose a nonparametric kernel method to estimate $\theta(\mathbf{s})$ for an angle invariant STPP. Suppose that observations of N are collected in $\mathcal{D} \times \mathcal{T}$ and denoted by $(\mathbf{s}_1, t_1), \cdots, (\mathbf{s}_n, t_n)$, where $n = N(\mathcal{D} \times \mathcal{T})$ is the total number of events. To estimate $\lambda(\mathbf{s}, t)$, we can use the classical Berman-Diggle estimator [4, 10] as

$$\hat{\lambda}_{\mathbf{h}}(\mathbf{s}, t) = \frac{1}{C_{\mathbf{h}}(\mathbf{s}, t)} \sum_{i=1}^{n} K_{\mathbf{h}}(\mathbf{s} - \mathbf{s}_i, t - t_i), \tag{3.1}$$

where $\mathbf{h} = (h_1, h_2)$ is the bandwidth (h_1 for space and h_2 for time), $K_{\mathbf{h}}(\mathbf{s}, t) = K(\mathbf{s}/h_1, t/h_2)/(h_1^2 h_2)$ with $K(\mathbf{s}, t) = K_{(1,1)}(\mathbf{s}, t)$ being a kernel function and $C_{\mathbf{h}}(\mathbf{s}, t) = \int_{\mathcal{D}} \int_{\mathcal{T}} K_{\mathbf{h}}(\mathbf{s} - \mathbf{u}, t - v) dv d\mathbf{u}$ being the edge-correction.

According to Theorem 1, it is not necessary to consider a full spatiotemporal kernel function in estimation of $\theta(\mathbf{s})$. Here, we recommend considering estimation of the first-order intensity function of $N_{\mathcal{T}}$, which is denoted by $\lambda_{\mathcal{T}}(\mathbf{s})$. We use a purely spatial full symmetric kernel in (3.1) such that $K(\mathbf{s}, t) = K(\mathbf{s}) = K(\|\mathbf{s}\|)$. Then, (3.1) becomes

$$\hat{\lambda}_{h, \mathcal{T}}(\mathbf{s}) = \frac{1}{C_h(\mathbf{s})} \sum_{i=1}^{n} K_h(\mathbf{s} - \mathbf{s}_i), \tag{3.2}$$

where $C_h(\mathbf{s}) = \int_{\mathcal{D}} K_h(\mathbf{s} - \mathbf{u}) d\mathbf{u}$ is the edge correction. Using notation of Stochastic integrals, (3.2) can be equivalently expressed as

$$\hat{\lambda}_{h, \mathcal{T}}(\mathbf{s}) = \frac{1}{C_h(\mathbf{s})} \int_{\mathcal{D}} \int_{\mathcal{T}} K_h(\mathbf{s} - \mathbf{u}) N(d\mathbf{u} \times dv). \tag{3.3}$$

In the following, we always use $\mathbf{u}$ to represent the integration variable for space and v to represent the integration variable for time. We treat (3.3) as an estimator of the first-order intensity function of $N_{\mathcal{T}}$, which will be used to derive an estimator of $\theta(\mathbf{s})$. To get the estimator, we consider the gradient of $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$, denoted by $\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$, which is composed of

$$
\begin{aligned}
&\nabla_j\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}) \\
&= \int_{\mathcal{D}}\int_{\mathcal{T}}[\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u})}{\partial z} - \frac{1}{C_h^2(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial z}K_h(\mathbf{s}-\mathbf{u})]N(d\mathbf{u}\times dv),
\end{aligned}
\tag{3.4}
$$

where $z = x$ if $j = 1$ and $z = y$ if $j = 2$. The kernel estimator of $\nabla\lambda_{\mathcal{T}}(\mathbf{s})$ is

$$
\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}) = (\nabla_1\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}),\nabla_2\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})) = \left(\frac{\partial\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\partial x},\frac{\partial\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\partial y}\right).
\tag{3.5}
$$

If $\|\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})\| > 0$, then we can uniquely solve a function $\hat{\theta}_h(\mathbf{s})$ on $[-\pi,\pi)$ by

$$
\begin{aligned}
\cos\hat{\theta}_h(\mathbf{s}) &= \frac{\nabla_1\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\|\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})\|}, \\
\sin\hat{\theta}_h(\mathbf{s}) &= \frac{\nabla_2\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\|\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})\|}.
\end{aligned}
\tag{3.6}
$$

The expected vector and the covariance matrix of $\nabla\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ are important in the evaluation of the properties of $\hat{\theta}_h(\mathbf{s})$, where the method of Stochastic Integrals in Campell's theorem is useful. Taking the expected values of (3.4), we have

$$
\begin{aligned}
&\mathrm{E}(\nabla_j\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})) \\
&= \int_{\mathcal{D}}[\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u})}{\partial u} - \frac{1}{C_h^2(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial x}K_h(\mathbf{s}-\mathbf{u})]\lambda_{\mathcal{T}}(\mathbf{u})d\mathbf{u},
\end{aligned}
\tag{3.7}
$$

where $u = x$ if $j = 1$ and $u = y$ if $j = 2$. Taking the variances and covariance of (3.4), we have

$$
\begin{aligned}
&\mathrm{Cov}(\nabla_{j_1}\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}),\nabla_{j_2}\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})) \\
&=\mathrm{Cov}(\frac{\partial\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\partial z_1},\frac{\partial\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\partial z_2}) \\
&= \int_{\mathcal{D}}\int_{\mathcal{D}}\int_{\mathcal{T}}\int_{\mathcal{T}}[\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u})}{\partial z_1} - \frac{1}{C_h(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial z_1}K_h(\mathbf{s}-\mathbf{u})] \\
&\quad \times [\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u}')}{\partial z_2} - \frac{1}{C_h^2(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial z_2}K_h(\mathbf{s}-\mathbf{u}')] \\
&\quad \times [g((\mathbf{u},v),(\mathbf{u}',v'))-1]\lambda(\mathbf{u},v)\lambda(\mathbf{u}',v')d\mathbf{u}d\mathbf{u}'dvdv' \\
&\quad + \int_{\mathcal{D}}[\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u})}{\partial z_1} - \frac{1}{C_h^2(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial z_1}K_h(\mathbf{s}-\mathbf{u})] \\
&\quad \times [\frac{1}{C_h(\mathbf{s})}\frac{\partial K_h(\mathbf{s}-\mathbf{u})}{\partial z_2} - \frac{1}{C_h^2(\mathbf{s})}\frac{\partial C_h(\mathbf{s})}{\partial z_2}K_h(\mathbf{s}-\mathbf{u})]\lambda_{\mathcal{T}}(\mathbf{u})d\mathbf{u},
\end{aligned}
\tag{3.8}
$$

where $z_1 = x$ if $j_1 = 1$, $z_1 = y$ if $j_1 = 2$, $z_2 = x$ if $j_2 = 1$, and $z_2 = y$ if $j_2 = 2$. Equation (3.8) can be used to compute the whole variance-covariance matrix of $\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ since it represents a formula of the variances if $j_1 = j_2$ or covariance if $j_1 \neq j_2$. If N is Poisson, then (3.8) becomes

$$
\begin{aligned}
&\mathrm{Cov}(\nabla_{j_1} \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}), \nabla_{j_2} \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})) \\
&= \int_{\mathcal{D}} \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_1} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_1} K_h(\mathbf{s} - \mathbf{u}) \right] \\
&\quad \times \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_2} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_2} K_h(\mathbf{s} - \mathbf{u}) \right] \lambda_{\mathcal{T}}(\mathbf{u}) d\mathbf{u}.
\end{aligned}
\tag{3.9}
$$

The function $\hat{\theta}_h(\mathbf{s})$ given by (3.6) is treated as an estimator of $\theta(\mathbf{s})$, which is well-defined if $\|\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})\| > 0$. It is clear that $\hat{\theta}_h(\mathbf{s})$ depends on the bandwidth h and the kernel function $K(\mathbf{s})$, which does not rely on $\mathcal{T}$. However, it does not mean that the temporal structure of the spatiotemporal point pattern is not important in $\hat{\theta}_h(\mathbf{s})$. According to (3.8), the temporal structure affects the properties of $\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ and therefore it also affects the properties of $\hat{\theta}_h(\mathbf{s})$. Based on our asymptotic results to be displayed in Section 4, we find that the temporal dependence structure has significant impacts on the bias, the variance, and the optimal bandwidth selection in $\hat{\theta}_h(\mathbf{s})$.

4. Asymptotics

We provide the asymptotic properties of $\hat{\theta}_h(\mathbf{s})$ under $T \to \infty$ with a bounded connected $\mathcal{D}$ in this section. The asymptotic properties include the consistency, the asymptotic normality, and the optimal strategy for the selection of h in (3.6). The derivation is based on the properties of $\hat{\theta}_h(\mathbf{s})$, where the asymptotic properties of $\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ are important. We use $\xrightarrow{P}$ to represent convergence in probability and $\xrightarrow{D}$ to represent convergence in distribution.

To derive the asymptotic properties, the most important regularity condition is the strong mixing condition, which was first proposed by [30] and then followed by many other authors [18, 20]. The strong mixing condition has later been used for asymptotics in spatial point processes [21]. Since $\mathcal{D}$ is bounded, we modify the definition of the strong mixing condition such that it can be interpreted by $T \to \infty$.

Let P be the distribution of N. Denote $\mathscr{F}_t$ as the σ-field generated by $\{N(A \times B) : A \subseteq \mathcal{D}, B \subseteq [0,t]\}$ and $\mathscr{G}_t$ as the σ-field generated by $\{N(A \times B) : A \subseteq \mathcal{D}, B \subset (t,T]\}$. Define

$$
\alpha_T(v) = \sup\{|P(E_1 \cap E_2) - P(E_1)P(E_2)| : E_1 \in \mathscr{F}_t, E_2 \in \mathscr{G}_{t+v}\}, \tag{4.1}
$$

where $0 < v < T$. The coefficient of strong mixing condition introduced by [30] then can be modified as

$$
\alpha(v) = \sup_{T \in \mathbb{R}^+} \alpha_T(v). \tag{4.2}
$$

To derive the asymptotic properties, we need the following regularity conditions.

(C1) N is simple and angle invariant with uniformly bounded first-order, second-order, third-order, and fourth-order intensity functions.

(C2) The function $\lambda_s(\mathbf{s}) = \lim_{T \to \infty} T^{-1} \int_0^T \lambda(\mathbf{s}, t) dt$ positively exists and is third-order continuously differentiable at every interior point $\mathbf{s} \in \mathcal{D}$.

(C3) There exists a $u > 0$ such that

$$\lim_{d_t(A_1 \times B_1, A_2 \times B_2) \to \infty} t^{-(1+u)} \mathrm{Corr}[N(A_1 \times B_1), N(A_2 \times B_2)] = 0,$$

for any $A_1 \times B_1, A_2 \times B_2 \subseteq \mathcal{D} \times \mathcal{T}$, where $d_t(A_1 \times B_1, A_2 \times B_2) = \min_{t_1 \in B_1, t_2 \in B_2} |t_1 - t_2|$.

(C4) There exists a uniformly bounded continuously differentiable bivariate function $\varphi(\mathbf{u}, \mathbf{u}')$ on $\mathcal{D}$ such that

$$\varphi(\mathbf{u}, \mathbf{u}') = \lim_{T \to \infty} \frac{1}{T} \int_0^T \int_0^T [g((\mathbf{u}, v), (\mathbf{u}', v)) - 1] \lambda(\mathbf{u}, v) \lambda(\mathbf{u}', v) dv dv'.$$

(C5) There exists a $\beta > 2$ and $\gamma = 2/\beta$ such that $\limsup_{a \in \mathbb{R}^+} \|N(\mathcal{D} \times [a, a + u])\|_\beta < \infty$ and $\int_0^\infty \alpha^{1-\gamma}(v) dv < \infty$ for any $u > 0$.

(C6) The kernel function K is full symmetric and satisfies $\lim_{u \to \infty} u^2 K(u) = 0$.

Condition (C1) concerns the properties of the first-order and second-order intensity functions, which can be easily understood. To derive the asymptotics of $\hat{\theta}_h(\mathbf{s})$, we need the functional central limit theorem theory for spatial point process. It can be established by Corollary 1 of [18]. The condition that λ_4 exists and is bounded upto $k = 4$ is proposed to apply the corresponding conclusion. If Condition (C2) also holds, then based on Theorem 1 there exists a function $\lambda'(\mathbf{s}, t)$ such that for any interior point $\mathbf{s} \in \mathcal{D}$ there is

$$\nabla \lambda_s(\mathbf{s}) = (\cos \theta(\mathbf{s}), \sin \theta(\mathbf{s})) \lim_{T \to \infty} T^{-1} \int_0^T \lambda'(\mathbf{s}, t) dt. \qquad (4.3)$$

Condition (C3) states that N is weakly dependent, which is strengthened by Condition (C4). The connection between (C3) and (C4) is given in Lemma 1 below. Condition (C5) is the usual mixing condition for functional central limit theorems with a weakly dependent structure [18]. For the kernel function, if Condition (C6) holds, then the term related to $\partial C_h(\mathbf{s})/\partial x$ and $\partial C_h(\mathbf{s})/\partial y$ in (3.7) and (3.8) can be ignored in asymptotic studies.

Lemma 1. *Assume (C1)–(C3) hold. If $f((\mathbf{u}, v), (\mathbf{u}', v'))$ is a uniformly bounded continuous bivariate function on $\mathcal{D} \times \mathcal{T}$, then*

$$\frac{1}{T} \int_0^T \int_0^T f((\mathbf{u}, v), (\mathbf{u}', v'))[g((\mathbf{u}, v), \mathbf{u}', v') - 1] \lambda(\mathbf{u}, v) \lambda(\mathbf{u}', v') dv dv' \qquad (4.4)$$

is uniformly bounded.

Lemma 1 states that the right-hand side of the equation given by Condition (C4) is bounded, implying that $\varphi(\mathbf{u}, \mathbf{u}')$ is well-defined if the limit exists. Let

$$\mu_{h,z}(\mathbf{s}) = \lim_{T \to \infty} T^{-1} \mathrm{E}(\nabla_j \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})),$$

where $z = x$ if $j = 1$ and $z = y$ if $j = 2$, and

$$\sigma_{h,z_1 z_2}(\mathbf{s}) = \lim_{T \to \infty} T^{-1} \mathrm{Cov}(\nabla_{j_1} \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}), \nabla_{j_2} \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})),$$

where $z_1 = x$ if $j_1 = 1$, $z_1 = y$ if $j_1 = 2$, $z_2 = x$ if $j_2 = 1$, and $z_2 = y$ if $j_2 = 2$. Using (C2) and (C3), there are

$$\mu_{h,z}(\mathbf{s}) = \int_{\mathcal{D}} \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z} K_h(\mathbf{s} - \mathbf{u}) \right] \lambda_s(\mathbf{u}) d\mathbf{u} \qquad (4.5)$$

and

$$
\begin{aligned}
& \sigma_{h,z_1 z_2}(\mathbf{s}) \\
&= \int_{\mathcal{D}} \int_{\mathcal{D}} \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_1} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_1} K_h(\mathbf{s} - \mathbf{u}) \right] \\
&\quad \times \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u}')}{\partial z_2} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_2} K_h(\mathbf{s} - \mathbf{u}') \right] \varphi(\mathbf{u}, \mathbf{u}') d\mathbf{u} d\mathbf{u}' \\
&\quad + \int_{\mathcal{D}} \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_1} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_1} K_h(\mathbf{s} - \mathbf{u}) \right] \\
&\quad \times \left[\frac{1}{C_h(\mathbf{s})} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_2} - \frac{1}{C_h^2(\mathbf{s})} \frac{\partial C_h(\mathbf{s})}{\partial z_2} K_h(\mathbf{s} - \mathbf{u}) \right] \lambda_s(\mathbf{u}) d\mathbf{u}.
\end{aligned}
\qquad (4.6)
$$

Equations (4.5) and (4.6) are important in our asymptotic studies.

Lemma 2. *Assume (C1)–(C5) hold. Let $\mu_h(\mathbf{s}) = (\mu_{h,x}(\mathbf{s}), \mu_{h,y}(\mathbf{s}))^T$ and*

$$\Sigma_h(\mathbf{s}) = \begin{pmatrix} \sigma_{h,xx}(\mathbf{s}) & \sigma_{h,xy}(\mathbf{s}) \\ \sigma_{h,yx}(\mathbf{s}) & \sigma_{h,zz}(\mathbf{s}) \end{pmatrix}.$$

If $\mathbf{s}$ is an interior point of $\mathcal{D}$, then for any fixed h there is

$$T^{-\frac{1}{2}} [\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}) - \mu_h(\mathbf{s})] \xrightarrow{D} N(0, \Sigma_h(\mathbf{s})) \qquad (4.7)$$

as $T \to \infty$.

Theorem 2. *Assume (C1)–(C5) hold and $\mu_h(\mathbf{s}) \neq \mathbf{0}$. Let*

$$
\begin{aligned}
\theta_h(\mathbf{s}) &= \arccos(\mu_{h,y}(\mathbf{s})/\|\mu_h(\mathbf{s})\|), \ \ if \ \mu_{h,y}(\mathbf{s}) \geq 0, \\
&= - \arccos(\mu_{h,y}(\mathbf{s})/\|\mu_h(\mathbf{s})\|), \ \ if \ \mu_{h,y}(\mathbf{s}) < 0,
\end{aligned}
\qquad (4.8)
$$

and

$$
\begin{aligned}
\tau_h^2(\mathbf{s}) &= \|\mu_h(\mathbf{s})\|^{-2} [\mu_{h,y}^2(\mathbf{s}) \sigma_{h,xx}(\mathbf{s}) - 2\mu_{h,x}(\mathbf{x}) \mu_{h,y}(\mathbf{x}) \sigma_{h,xy}(\mathbf{s}) \\
&\quad + \mu_{h,x}^2(\mathbf{s}) \sigma_{h,yy}(\mathbf{s})].
\end{aligned}
\qquad (4.9)
$$

Then for any fixed h, there is

$$\sqrt{T}(\hat{\theta}_h(\mathbf{s}) - \theta_h(\mathbf{s})) \xrightarrow{D} N(0, \tau_h^2(\mathbf{s})) \qquad (4.10)$$

as $T \to \infty$.

Theorem 3. *(Consistency) Assume (C1)–(C5) hold and $\|\nabla_s \lambda_{\mathcal{T}}(\mathbf{s})\| \neq 0$. If $h \to 0$ and $h^4 T \to \infty$ as $T \to \infty$, then $\hat{\theta}_h(\mathbf{s}) \xrightarrow{P} \theta(\mathbf{s})$ at any interior point $\mathbf{s} \in \mathcal{D}$.*

Theorem 4. *(Asymptotic Normality) Assume (C1)–(C6) hold and $\|\nabla_s \lambda_{\mathcal{T}}(\mathbf{s})\| \neq 0$. Let $\lambda_{s,x}(\mathbf{s}) = \partial \lambda_s(\mathbf{s})/\partial x$ and $\lambda_{s,y}(\mathbf{s}) = \partial \lambda_s(\mathbf{s})/\partial y$. Denote $\dot{f}(\mathbf{s})$ as the vector of gradient and $\ddot{f}(\mathbf{s})$ as the Hessian matrix of $f(\mathbf{s})$. Denote*

$$\begin{aligned}\nu(\mathbf{s}) =& \frac{1}{2(\lambda_{s,x}^2(\mathbf{s}) + \lambda_{s,y}^2(\mathbf{s}))} \\ & \times \int_{\mathbb{R}^2} K(\mathbf{u})\mathbf{u}^T [\lambda_{s,x}(\mathbf{s})\ddot{\lambda}_{s,y}(\mathbf{s}) - \lambda_{s,y}(\mathbf{s})\ddot{\lambda}_{s,y}(\mathbf{s})]\mathbf{u}d\mathbf{u}\end{aligned} \qquad (4.11)$$

and

$$\tau^2(\mathbf{s}) = \lambda_s(\mathbf{s})\|\nabla\lambda_s(\mathbf{s})\|^2 (\xi_x \lambda_{s,y}^2(\mathbf{s}) + \xi_y \lambda_{s,y}^2(\mathbf{s})), \qquad (4.12)$$

where $\xi_z = \int_{\mathcal{D}} [\frac{\partial K_h(\mathbf{u})}{\partial z}]^2 d\mathbf{u}$ with $z = x$ and $z = y$, respectively. If $h = cT^{-\frac{1}{8}}$ as $T \to \infty$, then

$$T^{\frac{1}{4}}(\hat{\theta}_h(\mathbf{s}) - \theta(\mathbf{s})) \xrightarrow{D} N(c^2 \nu(\mathbf{s}), \tau^2(\mathbf{s})/c^4). \qquad (4.13)$$

It is important to provide the best strategy for the selection of the bandwidth h used in the estimator. Here we consider the usual mean integrated squared error (MISE) of $\hat{\theta}_h(\mathbf{s})$ criterion which is defined as

$$\text{MISE}(\hat{\theta}_h(\mathbf{s})) = \int_{\mathcal{D}} \text{E}[\hat{\theta}_h(\mathbf{s}) - \theta(\mathbf{s})]^2 d\mathbf{s}. \qquad (4.14)$$

The best strategy for the selection of h is derived by minimizing $\text{MISE}(\hat{\theta}_h(\mathbf{s}))$.

Theorem 5. *Assume (C1)–(C6) hold. The $MISE(\hat{\theta}_h(\mathbf{s}))$ is asymptotically minimized by taking $h = cT^{-\frac{1}{8}}$ with*

$$c = \left\{ \frac{\int_{\mathcal{D}} \tau^2(\mathbf{s})d\mathbf{s}}{\int_{\mathcal{D}} \nu^2(\mathbf{s})d\mathbf{s}} \right\}^{\frac{1}{8}} \qquad (4.15)$$

as $T \to \infty$.

5. Simulation

We carried out simulation studies to evaluate the advantage of our gradient angle-based analysis method comparing with traditional intensity-based analysis methods. We wanted to address three basic issues. In the first, we wanted to study the sensitivity of $\hat{\theta}_h(\mathbf{s})$ to the bandwidth h. As the choice of the bandwidth is critical in estimating intensity functions, it is important to study the impact of the values of h on estimates of gradient angles. This was evaluated within

individual realizations. In the second, we wanted to study the reliability of $\hat{\theta}_h(\mathbf{s})$ in the entire study region. We wanted to know whether the pattern of $\hat{\theta}_h(\mathbf{s})$ was reliable to the change of intensity functions. We designed a mode change to represent the global change of intensity functions. This was evaluated between individual realizations. In the third, we evaluated the precision of $\hat{\theta}_h(\mathbf{s})$. We compared $\hat{\theta}_h(\mathbf{s})$ with its true values. This was conducted via a large number of simulation repetitions.

We evaluated the performance of $\nabla\hat{\lambda}_{h,\tau}(\mathbf{s})$ displayed by (3.6) via simulation studies. We simulated realizations from Poisson STPPs and Poisson cluster STPPs on $\mathbb{R}^2 \times [0,T]$ with $T = 1, 2, 5, 10$. We considered these processes due to their popularity in modeling geological and ecological data. In both processes, we chose $\lambda(\mathbf{s},t) = \kappa\phi[(\mathbf{s}-\mathbf{s}_0)/\sigma_0(t)]/\sigma_0^2(t)$ with $\kappa = 1000T$, where $\phi(\mathbf{s})$ was the standard bivariate normal density, $\sigma_0(t) = 0.5 + 0.2(t/T)$ was a time variant scale parameter, and $\mathbf{s}_0$ was either fixed or varied to represent the gradient center of $\nabla_s\lambda(\mathbf{s},t)$ or the spatial mode of $\lambda(\mathbf{s},t)$ equivalently. If $\mathbf{s}_0$ was fixed, then $\lambda(\mathbf{s},t)$ depended on time but angles of its gradient vectors projected to its spatial domain did not, implying that $\lambda(\mathbf{s},t)$ was angle invariant; otherwise, $\lambda(\mathbf{s},t)$ was not angle invariant. To study the impact of varied $\mathbf{s}_0$ on our approach, we divided the time domain into two periods. We fixed $\mathbf{s}_0$ at distinct locations in each of the two periods. This reflected the case when a sudden event occurred at the beginning of the second period (or the end of the first period) which significantly changed the global pattern of the point occurrences. It was similar to the issue to be discussed in *Great Sumatra-Andama earthquake* data set in Section 6. We chose $\kappa = 1000, 2000, 5000, 10000$ such that the total expected number of points were equal to 1000, 2000, 5000, and 10000, respectively.

In order to generate a Poisson STPP, we first generated the total number of points n from a Poisson random variable with expected values equal to κ. We then generated temporal locations $t_1, \cdots, t_n$ uniformly on $[0,T]$. For every t_i, we generated the spatial locations $\mathbf{s}_i$ independently from $\phi[(\mathbf{s}-\mathbf{s}_0)/\sigma_0(t_i)]/\sigma_0^2(t_i)$. A Poisson STPP was then derived. In order to generate a Poisson cluster STPP, we first generated the total number of parent points n_0 from a Poisson random variable with expected values equal to κ/k. We then generated parent temporal locations $t_1, \cdots, t_{n_0}$ uniformly on $[0,1]$ as well as their spatial locations $\mathbf{s}_1, \cdots, \mathbf{s}_{n_0}$ independently from $\phi[(\mathbf{s}-\mathbf{s}_0)/\sigma_0(t_i)]/\sigma_0^2(t_i)$. After parent points were derived, we independently generated offspring points around parent points, where each parent point had the $Poisson(k)$ number of offsprings and the spatiotemporal locations of offspring points to their parent points were determined by a Gaussian distribution with the standard deviation equal to σ. We chose $k = 4$ and $\sigma = 0.01$ in our simulation.

At the beginning of our simulations, we evaluated the performance of the classical estimator of intensity functions based on simulated realizations. We ignored the time of occurrences and attmpted to estimate $\lambda_{\mathcal{T}}(\mathbf{s})$ by $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ using (3.2) for each individual realization. We chose $K_h(\mathbf{s})$ as the density of the independent bivariate normal distribution with both means equal to 0 and both variances equal to h^2. We used various options of h to study the impact
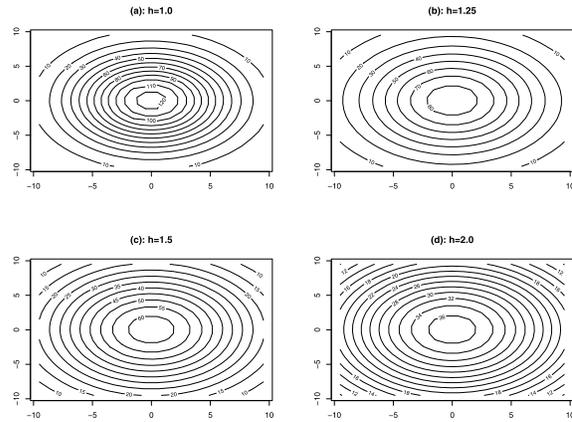
FIG 1. *Contour plots of estimated intensity derived within a realization in the simulation for a Poisson STPP with respect to selected h when $\kappa = 1000$ (or $T = 1$) and $\mathbf{s}_0$ is fixed at $(0,0)$.*

of the bandwidth on $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$. For a given h, we studied variations of $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ by comparing the values between different realizations. We found that for each individual realization values of $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ highly depended on the choices of h (Figure 1). For distinct realizations, values of $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ significantly changed even if values of h were not changed (not shown). Although their values varied greatly, the circular shapes in the contour plot of $\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})$ were not significantly affected by h. The entire pattern was extremely reliable among distinct realizations. Based on the findings, we guessed that the gradient angles of intensity functions might not be sensitive to the bandwidth h although it could significantly affect estimates of intensity functions. This motivated us to study the performance of estimated gradient angles rather than estimated values of intensity functions.

The performance of $\hat{\theta}_h(\mathbf{s})$ given by (3.6) were evaluated in two situations. In the first situation, we assumed $\mathbf{s}_0$ was always fixed at $(0,0)$, implying that $\lambda(\mathbf{s},t)$ was angle invariant in the entire time domain. In the second situation, we partitioned the time domain into two periods. The first period was formed by time within $[0, T/2]$, where $\mathbf{s}_0$ was fixed at $(0,0)$. The second period was formed by time within $(T/2, T]$, where $\mathbf{s}_0$ was fixed at $(0.5,0)$. This situation reflected the case when a sudden event occurred at $t = 0.5T$, which moved the gradient center of $\nabla_s \lambda(\mathbf{s},t)$ or the spatial mode of $\lambda(\mathbf{s},t)$ from $(0,0)$ to $(0.5,0)$.

We used all of the point occurrences in the computation of $\hat{\theta}_h(\mathbf{s})$ when $\mathbf{s}_0$ was always fixed at $(0,0)$ (Figure 2). The values of $\hat{\theta}_h(\mathbf{s})$ contained the length and the direction information of the gradient vectors. It showed that the lengths of the gradient vector were significantly affected by the choices of h but the angles were not. This was consistent with our previous findings obtained from Figure 1. To study the reliability of $\hat{\theta}_h(\mathbf{s})$, we compared the gradient angle patterns among different realizations in our simulations. We did not find any significant changes visually.
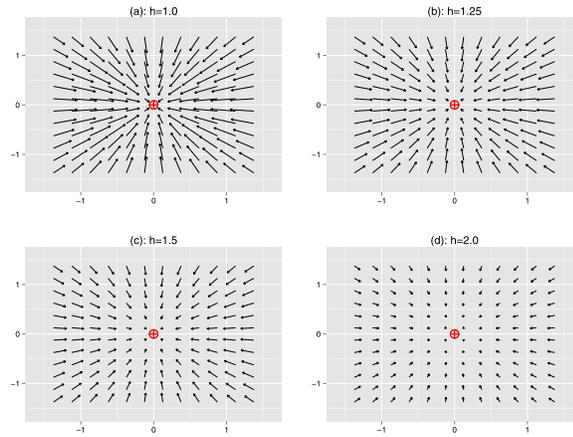
FIG 2. *Estimated gradient patterns derived within a realization in the simulation for a Poisson STPP with respect to selected h when* $\kappa = 1000$ *(or* $T = 1$*) and* $\mathbf{s}_0$ *is fixed at* $(0,0)$.
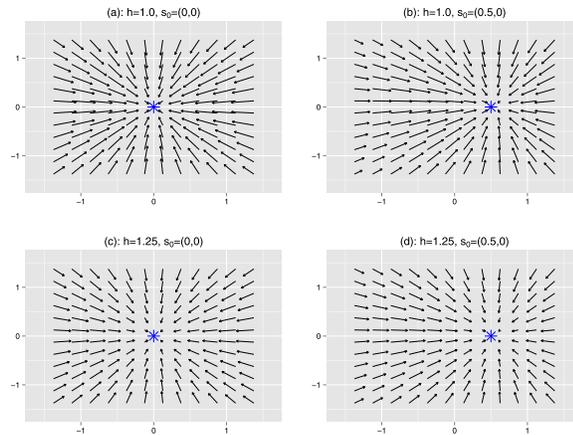

FIG 3. *Estimated gradient patterns derived from realizations in the simulation for a Poisson STPP with respect to selected h when* $\kappa = 1000$ *(or* $T = 1$*) and* $\mathbf{s}_0$ *moves from* $(0,0)$ *in the first period to* $(0.5,0)$ *in the second period.*

We studied the performance of $\hat{\theta}_h(\mathbf{s})$ according to the two time periods when $\mathbf{s}_0$ varied (Figure 3). In the first time period, we computed the values of $\hat{\theta}_h(\mathbf{s})$ using points with $t_i \leq 0.5T$. In the second time period, we computed the values of $\hat{\theta}_h(\mathbf{s})$ using points with $t_i > 0.5T$. We had identified significant changes in the comparison of gradient angle patterns between the two periods. The gradient center was almost fixed at $(0,0)$ in the first period or at $(0.5,0)$ in the second period, indicating that a global change of point occurrences was identified in the data. Gradient angles were not sensitive to the choices of h in both periods, which were also reliable in the comparison of patterns between distinct realizations.

Simulation results for the root MISE of the differences between the true and estimated spatial angles of gradient vectors of the first-order intensity functions in the Poisson STPPs and Poisson cluster STPPs, respectively, when $\mathbf{s}_0 = (0,0)$, $\kappa = 1000T$, and the number of replications is 10,000.

T	Poisson for different h				Poisson cluster for different h			
	2.2	2.0	1.8	1.6	2.2	2.0	1.8	1.6
1	0.0305	0.0304	0.0310	0.0318	0.0334	0.0332	0.0329	0.0345
2	0.0213	0.0212	0.0215	0.0216	0.0229	0.0247	0.0239	0.0246
5	0.0135	0.0134	0.0135	0.0136	0.0150	0.0148	0.0153	0.0158
10	0.0096	0.0095	0.0096	0.0100	0.0107	0.0105	0.0106	0.0109

Based on our findings in the two situations, we further studied the performance of $\hat{\theta}_h(\mathbf{s})$ if points were partitioned into more than two time periods (not shown). When $\mathbf{s}_0$ was fixed at $(0,0)$, as the case considered in the first situation, we found that the angle patterns were almost identical among all of these periods. If $\mathbf{s}_0$ varied from $(0,0)$ to $(0.5,0)$, as the case considered in the second period, the angle patterns had significant changes as time approached to $0.5T$. After that, no significant changes were founded. Therefore, we conclude that the usage of gradient angle patterns can be more efficient in the study of the change of global patterns of spatiotemporal point data.

Although we have only displayed the results for Poisson STPPs, we also studied similar issues for Poisson cluster STPPs. We did not found any significant differences in our studies for Poisson cluster STPPs. Therefore, we decided not to show the results. In addition, we used a data driven approach to study the robustness of our findings. We carried out the cross-validation approach for individual realizations. We did not find any significant changes of the patterns of $\hat{\theta}_h(\mathbf{s})$ in the approach. Based on our studies, we conclude that the findings reflected by Figures 2 and 3 should be general in all types of STPPs.

In the end, we evaluated the accuracy of $\hat{\theta}_h(\mathbf{s})$ when $\mathbf{s}_0$ was fixed at $(0,0)$, indicating that the issue was considered only for an angle invariant $\lambda(\mathbf{s},t)$ in the entire temporal domain. We computed the angle difference between the true and the estimated angles at each point in the study area using

$$d_{\theta,h}(\mathbf{s}) = 2\arcsin\left(\frac{\|\theta(\mathbf{s}) - \hat{\theta}_h(\mathbf{s})\|}{2}\right),$$

where $\theta(\mathbf{s})$ always pointed to the spatial origin based on our simulation patterns. After values of $d_{\theta,h}(\mathbf{s})$ were derived, we computed the mean integrated square error (MISE) of angle differences by integrating the squares of their values over the whole study area. We then compared the root MISE values for selected bandwidth h based on 10,000 simulation replications (Table 1). We found that the MISEs of the Poisson cluster STPPs were a little bit larger than the MISEs of the Poisson SPPs. The MISE decreased as κ increased. The change of MISEs was not large in all of the cases that we had studied.

In summary, we found that angles of gradient vectors of the first-order intensity function were not sensitive to the choice of the bandwidth. Gradient angle

patterns were extremely stable among different realizations. We did not find any siginificant changes if the mode of the first-order intensity function did not vary. Estimates of gradient angles were more precise than estimates of intensity functions. Therefore, it is important to pay more attention to the change of angle patterns in practice.

6. Applications

We expected that our method would have wide applications in natural hazard studies. Earthquakes are considered as the most important natural hazard events. They often result in numerous deaths and damages. This motivated us to apply our method to earthquake data. Many sources of earthquake data have been established and available via internet and can be downloaded for free. Examples include the websites of the United States National Geophysical (USGS) data center, the Northern California Earthquake Data Center (NCEDC), the GeoCommunity data center, and many others. The database contains dates, depths, longitude, latitude, and magnitude of earthquakes at the regional or global level from thousands of years ago to recent years.

An important issue in the analysis of earthquake data is the problem of earthquake clusters caused by aftershocks. The presence of earthquake clusters often makes it hard to understand earthquake activities. A number of statistical methods have been proposed for earthquake clusters. One of the most important model is the epidemic-type aftershock sequences (ETAS) model [26, 27, 40]. The ETAS model, which is defined by a conditional intensity function only affected by ancestors but not offsprings, has been widely used for earthquakes [3, 7, 35]. The ancestors may be understood as mainshock earthquakes, which are described by their spatiotemporal locations and corresponding magnitudes. If an extremely large mainshock earthquake occurs, then within a short time period the ETAS model is primarily dominated by its aftershock patterns, which implies the impact of the other mainshock earthquakes can be ignored. Let the magnitude and the spatiotemporal location of the extremely large mainshock earthquake be denoted by M^* and $(\mathbf{s}^*, t^*) = (x^*, y^*, t^*)$, respectively. Then, the conditional intensity function [27, 40] can be approximately expressed as

$$\lambda^*(\mathbf{s}, t, M) = j(M)[\mu(x,y) + \kappa(M^*)u(t - t^*)v(x - x^*, y - y^*|M^*)], \qquad (6.1)$$

where $\mu(x,y)$ represents the background earthquakes, $j(M)$ satisfies $\int_0^\infty j(m)dm = 1$, M represents the magnitude, and $(\mathbf{s}, t) = (x, y, t)$ represents the spatiotemporal location of aftershock earthquakes. By integrating out the magnitude term given by $j(M)$, if the point pattern is only considered within a short time period after the extremely large mainshock earthquake, then $\mu(x,y)$ can be ignored, which implies that the intensity function of the point pattern can be approximately expressed by (2.4) in Example 3 of Section 2, which is angle invariant.

To investigate whether a spatiotemporal point pattern of aftershocks of extremely large earthquake satisfies Model (2.4), we collected the historical earthquake data from the NCEDC website. The website contained global earthquake

activities since 1900. Since our interests were earthquake patterns in this century, we collected earthquake data with magnitude greater than or equal to 4.0 from 2000 to 2014 from the website. The data contained 197 major earthquakes (≥ 7.0) and 16 great earthquakes (≥ 8.0). Since Model (2.4) was approximately derived for extremely large earthquakes based on the ETAS, we focused on great earthquakes in the data. We had two interesting and important cases. The first case was the *Great Tohoku Earthquake* which occurred in March 11, 2011 with magnitude 9.1 at 38.30 latitude North and 142.37 longitude East. The Great Tohoku earthquake was the most powerful earthquake ever recorded to have hit Japan [33], which caused over twenty thousand life lost. Importantly, the *Great Tohoku Earthquake* also produced a severe tsunami which caused serious nuclear accidents in Fukushima Nuclear Power Plants affecting hundreds of thousands of residents in a few thousands square kilometers area. Besides the *Great Tohoku Earthquake*, the second important case was the *Great Sumatra-Andama Earthquake* which occurred in December 26, 2004 with magnitude 9.0 at 3.30 latitude North and 95.98 longitude East. The *Great Sumatra-Andama earthquake* was the most powerful earthquake in the area of Indian Ocean near the coast of Indonesia in the recorded history [22]. The earthquake caused over two hundred thousand life lost in over fourteen countries. Most of those were killed by the tsunami triggered by the earthquake.

To investigate the earthquake pattern caused by the *Great Tohoku Earthquake*, we focused on aftershock earthquakes that occurred within an area from 30 latitude north to 45 latitude north and from 130 longitude east to 150 longitude east from March 11, 2001 to September 10, 2001, which contained a six-month time period of aftershock earthquake occurrences of the *Great Tohoku Earthquake*. The dataset contained 4543 aftershock earthquakes with magnitude greater than or equal to 4.0. There were 656 earthquakes with magnitude between five and six, 72 earthquakes between six and seven, and three earthquakes between seven and eight. The largest aftershock earthquake with magnitude 7.9 and the second largest aftershock earthquake with magnitude 7.7 occurred in the same day of the mainshock earthquake. After the first day, the largest aftershock earthquake with magnitude 7.1 occurred about 27 days later. Therefore, there were no extremely large (i.e. great) aftershock earthquakes in the *Great Tohoku Earthquake* dataset.

We partitioned the sixth-month time period into six 30-day or 31-day time intervals. Within each time interval, we estimated the gradient pattern of aftershock occurrences using $\nabla\hat{\lambda}_{h,\tau}(\mathbf{s})$ given by (3.5), where the normal kernel with bandwidth equal to 100km was used. The bandwidth value was obtained by a cross-validation procedure with a number of candidate values. We also considered other bandwidth values and found that the patterns were similar. We compared the patterns of $\nabla\hat{\lambda}_{h,\tau}(\mathbf{s})/n$ (Figure 4), where n was the number of aftershock earthquakes in the time period. We had n equal to 3087 the first, 563 in the second, 308 in the third, 217 in the fourth, 195 in the fifth, and 173 in the sixth time periods, respectively. We had several interesting findings in the comparison between these patterns: (a) the gradient centers of all of the six patterns were almost fixed at 37.5 latitude north and 142.5 longitude east,
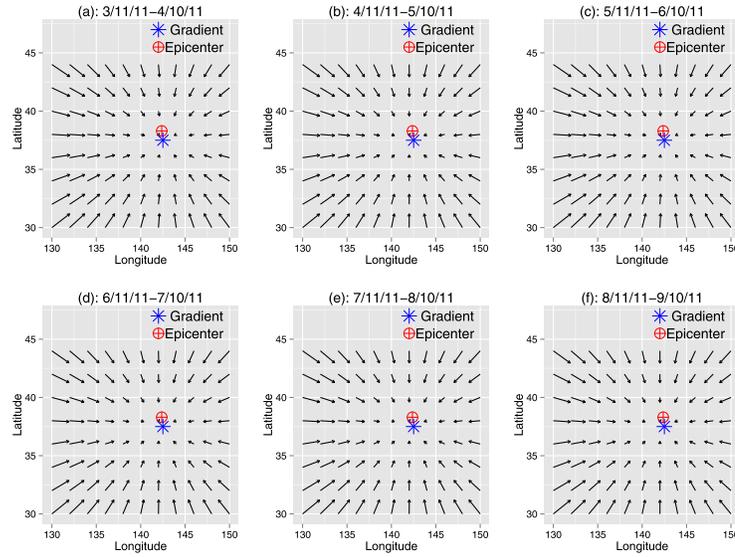
FIG 4. *Gradient patterns of aftershock earthquakes for the 2011 Great Tohoku Earthquake, where the mainshock (magnitude* 9.1*) occurred on March 11, 2011.*

which was about 90km away from the epicenter; and (b) the angles of all of gradient vectors considered in the study area almost kept constants but their lengths varied significantly. Based on our findings, we could roughly conclude that the patterns displayed by Figure 4 were consistent with the assumptions of the ETAS model.

To investigate the earthquake pattern caused by *Great Sumatra-Andama earthquake*, we focused on aftershock earthquakes that occurred within an area from 10 latitude south to 15 latitude north and from 90 longitude east to 105 longitude east from December 26, 2004 to June 26, 2005, which contained a six-month time period of aftershock earthquake occurrences of the *Great Sumatra-Andama earthquake*. The dataset contained 4381 aftershock earthquakes with magnitude greater than or equal to 4.0. There were 687 earthquakes with magnitude between five and six, 43 earthquakes with magnitude between six and seven, 3 earthquakes with magnitude between seven and eight, and one extremely large aftershock earthquake with magnitude 8.6 at 2.09 latitude north and 97.11 longitude east on March 28, 2005. We also partitioned the sixth-month time period into six 30-day or 31-day time intervals. Within each time interval, we also computed $\nabla\hat{\lambda}_{h,\tau}(\mathbf{s})$ using (3.5), where the normal kernel with bandwidth equal to 100km were used. This was also obtained by a cross-validation procedure with a number of candidate values. We also considered other bandwidth values and found that the patterns were similar. Similarly, we compared the patterns of $\nabla\hat{\lambda}_{h,\tau}(\mathbf{s})/n$ (Figure 5). We had n equal to 1288 in the first, 964 in the second, 201 in the third, 1348 in the fourth, 343 in the first, and 237 in the sixth time periods, respectively. Based on Figure 5, we had: (a) the gradient center of the
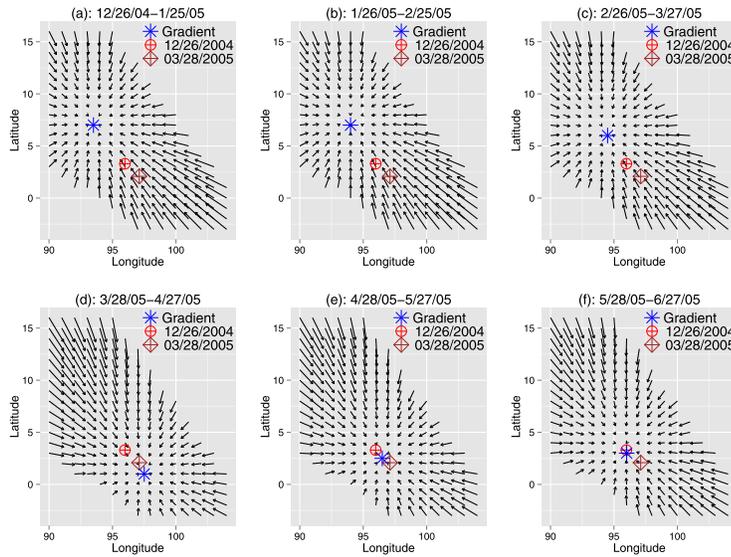
FIG 5. *Gradient patterns of aftershock earthquakes for the 2004 Great Sumatra-Andaman Earthquake, where the mainshock (magnitude* 9.0*) and the largest aftershock (magnitude* 8.6*) occurred on December 26, 2004 and March 28, 2005, respectively.*

first three months was almost fixed at latitude 7.0 north and longitude 93.5 east, which was about 500km northwest of the mainshock earthquake epicenter; (b) the largest aftershock earthquake with magnitude 8.6 occurred three month later at the opposite direction of the first three month gradient center from the location of the mainshock earthquake, which was at about 184km southeast of the mainshock epicenter; and (c) the gradient center gradually moved back to the mainshock epicenter after the largest aftershock earthquake occurred and the two centers were almost identical a few months later. It was clear that the gradient patterns displayed by Figure 5 was inconsistent with the assumption of the ETAS model. The ETAS model assumes aftershock earthquakes occur symmetrically in space about the epicenter of the mainshock earthquake, but this was not true in the pattern displayed by Figure 5.

In summary, we concluded that the angle patterns of gradient vectors of the first-order intensity function was almost time invariant in the *2011 Great Tohoku Earthquake* data but not in the *2004 Great Sumatra-Andaman Earthquake* data. The possible reason was the existence of extremely large aftershock earthquakes a few months later, which appeared in the *2004 Great Sumatra-Andaman Earthquake* data but not in the *2011 Great Tohoku Earthquake* data. The extremely large aftershock earthquake reflected the unbalanced release of energy which significantly affected the earthquake aftershock pattern. The ETAS model would be appropriate in modeling the aftershock earthquake patterns if there were not extremely large aftershock earthquakes. Otherwise, other models should be considered. If the gradient center of aftershock earthquakes was far away from the

epicenter, then it was possible that there existed an extremely large aftershock earthquake in the opposite direction of the gradient center. This is an important future research topic.

7. Discussion

We have proposed a gradient angle-based analysis method for spatiotemporal point processes (STPPs). The method is developed based on a gradient-based assumption, called the angle invariant assumption, for the first-order intensity function of STPPs. If the angle invariant assumption holds, then the angle of gradient vectors of the first-order intensity in the spatial domain does not vary in time but the length may be time varied. This concept is important in the study of many human and natural phenomena when clusters of point patterns are present. Because angle patterns of gradient vectors within clusters are more reliable than their length patterns as time varies, it is possible for us to capture the main feature of STPPs by studying patterns of gradient angles of its first-order intensity function. The method based on gradient angles suggests that one should pay more attention to the angles rather than the lengths of gradient vectors for dynamics of spatiotemporal point patterns. If gradient angle patterns have significant changes, then a global change may appear.

Since only the first-order property is assumed, the choice of the second-order property is flexible. Both the first-order and the second-order intensity functions are important in applications. It is interesting to combine our method with other methods for analysis of the second-order properties. For instance, if the second-order intensity is also modeled by the second-order intensity-reweighted stationary or isotropic process [1], then both the first-order and the second-order intensity functions are specified. The joint use of our gradient angle-based method and other statistical methods for second-order properties may provide efficient and effective approaches to understand the dynamics of spatiotemporal point patterns.

We expect that our gradient angle-based method will have wide applications in natural hazard and infectious disease studies. Many natural hazard phenomena can be described by spatiotemporal point processes with marks. The angle pattern of gradient vectors describes the tendency of changes of occurrences, which may depend on the magnitude (i.e. marks) of events. The usage of angle patterns is likely to provide solutions to visualize the linkage between different time periods of the occurrences of natural hazard events. For infectious diseases, the angle pattern is even more important since it can be easily interpreted as the way of the spread of a disease. We suspect that the way of infection is changed if a significant change of the gradient angle pattern has been found.

There are a few possible extensions to our approach. First, the gradient angle-based analysis method only needs that the angle of gradient vector is invariant in time. It does not provide any further methods to model the angle patterns. To make the model more precise, one can use a parametric method to model the angle pattern such that a parametric approach is derived. Second, although we have provided a kernel-based nonparametric approach to estimate the angle

pattern, we have not considered other approaches such as the smoothing spline approach [14]. Therefore, many nonparametric approaches may be considered. Third, explanatory variables are usually important in statistics but the problem has not been considered in our approach. It is of interest to consider a way to incorporate explanatory variables in our model and this should be an important problem in future research.

Appendix A: Proofs

Proof of Theorem 1: Let $\lambda(\mathbf{s}, t)$, $\lambda_B(\mathbf{s})$ and $\lambda_{\tilde{B}}(\mathbf{s})$ be the first-order intensity functions of N, N_B, and $N_{\tilde{B}}$, respectively. If N is angle invariant, then there exists a function $\theta(\mathbf{s})$ defined on $\mathcal{D}$ taking values on $[-\pi, \pi)$ such that

$$\frac{\partial \lambda(\mathbf{s}, t)}{\partial x} = \|\nabla_s \lambda(\mathbf{s}, t)\| \cos \theta(\mathbf{s})$$

and

$$\frac{\partial \lambda(\mathbf{s}, t)}{\partial y} = \|\nabla_s \lambda(\mathbf{s}, t)\| \sin \theta(\mathbf{s}).$$

Then, the gradient of $\lambda_B(\mathbf{s})$ is

$$\nabla \lambda_B(\mathbf{s}) = (\int_B \frac{\partial \lambda(\mathbf{s}, t)}{\partial x} dt, \int_B \frac{\partial \lambda(\mathbf{s}, t)}{\partial y} dt) = (\cos \theta(\mathbf{s}), \sin \theta(\mathbf{s})) \int_B \|\nabla_s \lambda(\mathbf{s}, t)\| dt,$$

which implies that the angle of $\nabla \lambda_B(\mathbf{s})$ is still $\theta(\mathbf{s})$. Therefore, the angle of $\nabla \lambda_B(\mathbf{s})$ does not depend on B, which implies that N_B and $N_{\tilde{B}}$ have similar patterns for any $B, \tilde{B} \subseteq \mathcal{T}$. In addition, if the angle of $\nabla \lambda_B(\mathbf{s})$ is invariant in B, then there exists a function $\theta(\mathbf{s})$ defined on $\mathcal{D}$ taking values on $[-\pi, \pi)$ such that

$$\nabla \lambda_B(\mathbf{s}) = (\cos \theta(\mathbf{s}), \sin \theta(\mathbf{s})) \|\nabla \lambda_B(\mathbf{s})\| = (\int_B \frac{\partial \lambda(\mathbf{s}, t)}{\partial x} dt, \int_B \frac{\partial \lambda(\mathbf{s}, t)}{\partial y} dt). \quad (A.1)$$

Let $B_{t_0}(a) = [t_0 - a, t_0 + a]$ for any $t_0 \in \mathbb{R}$ and $a \in \mathbb{R}^+$. Then, $\|\nabla_s \lambda(\mathbf{s}, t_0)\| = \lim_{a \to 0} \|\nabla \lambda_{B_{t_0}(a)}\|/(2a)$ for any $t \in \mathcal{T}$. Since $\lambda(\mathbf{s}, t)$ is continuously differentiable in $\mathbf{s}$ and t, $\|\nabla \lambda_B(\mathbf{s})\|$ is continuous in $\mathbf{s}$. Therefore, $\theta(\mathbf{s})$ is continuous in $\mathbf{s}$ if $\sup_{t \in \mathcal{T}} \|\nabla_s \lambda(\mathbf{s}, t)\| > 0$, which implies that $\theta(\mathbf{s})$ is uniquely determined by (A.1). If $\|\nabla_s(\mathbf{s}, t)\| = 0$ for any $t \in \mathcal{T}$, then $\theta(\mathbf{s})$ is not uniquely defined in (A.1). In this case, we have $\|\nabla \lambda_B(\mathbf{s})\| = 0$ for any $B \subseteq \mathcal{T}$, which is not included in Definition 1. Therefore, N is angle invariant. $\square$

Proof of Lemma 1: Let $a_i = (i/m)T$ for any positive T with $i = 0, \cdots, m$. Define $U(\mathbf{u}; \epsilon)$ as the small disc centered at $\mathbf{u}$ with diameter ϵ in $\mathcal{D}$. Let

$$S_m = \frac{1}{T} \sum_{i=1}^{m} \sum_{j=1, j\neq i}^{m} f((\mathbf{u}, v), (\mathbf{u}', v))$$
$$\times \operatorname{Corr}\{N[U(\mathbf{u}; \epsilon) \times [a_{i-1}, a_i]], N[U(\mathbf{u}'; \epsilon) \times [a_{j-1}, a_j]]\},$$

where $\mathbf{u}$ and $\mathbf{u}'$ are interior points of $\mathcal{D}$. If $|f((\mathbf{u}, v), (\mathbf{u}', v'))| < M$ and $\lambda(\mathbf{s}, t) \leq C$, then there exists a $c > 0$ such that

$$
\begin{aligned}
|S_m| \leq & \frac{1}{T} \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} |f((\mathbf{u}, v), (\mathbf{u}', v)) \\
& \times \operatorname{Corr}\{N[U(\mathbf{u}; \epsilon) \times [a_{i-1}, a_i]], N[U(\mathbf{u}'; \epsilon) \times [a_{j-1}, a_j]]\}| \\
\leq & \frac{1}{T} \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} cM\lambda(\mathbf{u}, a_i)\lambda(\mathbf{u}', a_j) \\
& \times [\max(1, \frac{T}{m}(j - i))]^{-(1+\alpha)} |U(\mathbf{u}, \epsilon)||U(\mathbf{u}', \epsilon)| \\
\leq & \frac{1}{T} \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} cMC^2 [\max(1, \frac{T}{m}(j - i))]^{-(1+\alpha)} |U(\mathbf{u}, \epsilon)||U(\mathbf{u}', \epsilon)| \\
\rightarrow & 2cMC^2 (1 + \int_1^T t^{-(1+\alpha)} dt) |U(\mathbf{u}, \epsilon)||U(\mathbf{u}', \epsilon)|
\end{aligned}
$$

as $m \rightarrow \infty$. Therefore, S_m absolutely converges as $T \rightarrow \infty$. The conclusion is drawn since with Condition (C2) the quantity defined by (4.4) is the limit of $S_m/(|U(\mathbf{u}, \epsilon)||U(\mathbf{u}', \epsilon)|)$ under $m \rightarrow \infty$ and $\epsilon \rightarrow 0$. □

Proof Lemma 2: For any measurable $A \subseteq \mathcal{D}$ and $m \in \mathbb{R}^+$, let $X_i = N(\mathcal{A} \times [(i-1)T/m, iT/m)) - \int_A \int_{(i-1)T/m}^{iT/m} \lambda(\mathbf{s}, t) dt d\mathbf{s}$. Then, $N(A \times \mathcal{T}) = \sum_{i=1}^{m} X_i$. Using Condition (C1), there are $E(X_i) = 0$ and $V(X_i) < \infty$. Using Conditions (C3) and (C4), there is

$$
b_h^2(A) = \lim_{T \rightarrow \infty} \frac{V[N(A \times \mathcal{T})]}{T} = \int_A \int_A \varphi(\mathbf{u}, \mathbf{u}') d\mathbf{u} d\mathbf{u}' - \int_D \lambda_s(\mathbf{u}) d\mathbf{u}.
$$

Since the conclusion of Corollary 1 in [18] can be applied if Condition (C5) holds, there is $N(A \times [0, aT])/(b_h(A)\sqrt{T})$ weakly converges to the standard Brownian motion on $[0, 1]$. Note that this conclusion holds for any fixed measurable $A \subseteq \mathcal{D}$. Using the standard method of simple functions for Lebesgue Integral, we can show the left-hand side of (4.7) is normally distributed. Using (4.5) and (4.6), we derive the conclusion of the lemma. □

Proof of Theorem 2: Using the polar transformation of $x = r\cos\omega$ and $y = r\sin\omega$, where $0 \leq r < \infty$ and $\omega \in [-\pi, \pi)$, there are $r = \sqrt{x^2 + y^2}$ and $\omega = \arccos(x/\sqrt{x^2 + y^2})$ if $y > 0$ and $\omega = -\arccos(x/\sqrt{x^2 + y^2})$ if $y < 0$. The gradient of ω is $\nabla\omega = (-y/(x^2 + y^2), x/(x^2 + y^2))$. Putting $x = \mu_{h,x}(\mathbf{s})$ and $y = \mu_{h,y}(\mathbf{s})$ in ω and $\nabla\omega$, we have $\cos\theta_h(\mathbf{s}) = \mu_{h,y}(\mathbf{s})/\|\mu_h(\mathbf{s})\|$ and $\sin\theta_h(\mathbf{s}) = \mu_{h,x}(\mathbf{s})/\|\mu_h(\mathbf{s})\|$. Then, the conclusion of the theorem is drawn using the Delta Theorem. □

Proof of Theorem 3: For any interior point $\mathbf{s} \in \mathcal{D}$, there are $\lim_{h \rightarrow 0} C_h(\mathbf{s}) = 1$ and $\lim_{h \rightarrow 0} \|\nabla C_h(\mathbf{s})\| = 0$. Using the integral by part in (4.5), there is

$$\lim_{h \to 0} \mu_{h,z}(\mathbf{s}) = \lim_{h \to 0} \int_{\mathcal{D}} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z} \lambda_s(\mathbf{u}) d\mathbf{u}$$

$$= \lim_{h \to 0} \int_{\mathcal{D}} \frac{\partial \lambda_s(\mathbf{u})}{\partial z} K_h(\mathbf{s} - \mathbf{u}) d\mathbf{u}$$

$$= \frac{\partial \lambda_s(\mathbf{s})}{\partial z}.$$

Therefore, $\lim_{h \to 0} T^{-1} \mu_h(\mathbf{s}) = \nabla \lambda_s(\mathbf{s})$. For $\sigma_{h,z_1 z_2}$ given by (4.6), we study $\nabla K_h(\mathbf{s}) = h^3 \nabla K(\mathbf{s})$ and find $\lim_{h \to 0} h^{3+v} \sup_{\mathbf{s} \in \mathcal{D}} \|\nabla K_h(\mathbf{s})\| = 0$. Then

$$\lim_{h \to 0} h^4 \sigma_{h,z_1 z_2} \leq \lim_{h \to 0} h^3 \sup_{\mathbf{s} \in \mathcal{D}} \|\nabla K_h(\mathbf{s})\| \int_{\mathcal{D}} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z_2} \lambda_s(\mathbf{u}) d\mathbf{u}$$

$$= \lim_{h \to 0} h^3 \sup_{\mathbf{s} \in \mathcal{D}} \|\nabla K_h(\mathbf{s})\| \frac{\partial \lambda_s(\mathbf{s})}{\partial z_2}$$

$$< \infty.$$

Therefore,

$$\lim_{T \to \infty} \|\mathrm{V}(\frac{\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{T})\| = \lim_{T \to \infty} \frac{1}{T} \|\mathrm{V}(\frac{\hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\sqrt{T}})\| \leq \lim_{T \to \infty} \frac{1}{h^3 T} h^3 \|\Sigma_h(\mathbf{s})\| = 0,$$

where $\|\mathbf{M}\|$ represents the L_2-norm of the matrix $\mathbf{M}$. Therefore, using the Chebychev inequality, we have $T^{-1} \nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s}) \xrightarrow{P} \nabla \lambda_s(\mathbf{s})$. Using the Continuous Mapping theorem, there is

$$\frac{\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})}{\|\nabla \hat{\lambda}_{h,\mathcal{T}}(\mathbf{s})\|} \xrightarrow{P} (\cos \theta(\mathbf{s}), \sin \theta(\mathbf{s})).$$

Therefore, $\hat{\theta}_h(\mathbf{s}) \xrightarrow{P} \theta(\mathbf{s})$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 4: For any small sufficient h, using the Taylor expansion, there is

$$\lambda_{s,z}(\mathbf{s} - h\mathbf{u}) = \lambda_{s,z}(\mathbf{s}) - h\dot{\lambda}_{s,z}^T(\mathbf{s})\mathbf{u} + \frac{h^2}{2}\mathbf{u}^T \ddot{\lambda}_{s,z}(\mathbf{s})\mathbf{u} + o(h^2),$$

where $z = x$ or $z = y$. Using the integral by part, there is

$$\lim_{h \to 0} h^{-2} \nabla \mu_h(\mathbf{s})$$

$$= (\lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial x} \lambda_{\mathbf{s}}(\mathbf{u}) d\mathbf{u}, \lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial y} \lambda_{\mathbf{s}}(\mathbf{u}) d\mathbf{u})$$

$$= (\lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} K_h(\mathbf{s} - \mathbf{u}) \lambda_{\mathbf{s},x}(\mathbf{u}) d\mathbf{u}, \lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} K_h(\mathbf{s} - \mathbf{u}) \lambda_{\mathbf{s},y}(\mathbf{u}) d\mathbf{u})$$

$$= (\lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} K(\mathbf{u}) \lambda_{\mathbf{s},x}(\mathbf{s} - h\mathbf{u}) d\mathbf{u}, \lim_{T \to \infty} h^{-2} \int_{\mathcal{D}} K(\mathbf{u}) \lambda_{\mathbf{s},y}(\mathbf{s} - h\mathbf{u}) d\mathbf{u}).$$

Then,

$$\lim_{T \to \infty} h^{-2} [\nabla \mu_h(\mathbf{s}) - \nabla \mu(\mathbf{s})]$$

$$= (\frac{1}{2} \int_{\mathbb{R}^2} K(\mathbf{u})\mathbf{u}^T \ddot{\lambda}_{s,x}(\mathbf{s})\mathbf{u}d\mathbf{u}, \frac{1}{2} \int_{\mathbb{R}^2} K(\mathbf{u})\mathbf{u}^T \ddot{\lambda}_{s,y}(\mathbf{s})\mathbf{u}d\mathbf{u}).$$

Using $\theta_h(\mathbf{s})$ in (4.8), we derive the expression of the bias as

$$\nu(\mathbf{s}) = \lim_{h \to 0} h^{-2}(\theta_h(\mathbf{s}) - \theta(\mathbf{s}))$$

$$= \frac{1}{2(\lambda_{s,x}^2(\mathbf{s}) + \lambda_{s,y}^2(\mathbf{s}))} \int_{\mathbb{R}^2} K(\mathbf{u})\mathbf{u}^T [\lambda_{s,x}(\mathbf{s})\ddot{\lambda}_{s,y}(\mathbf{s}) - \lambda_{s,y}(\mathbf{s})\ddot{\lambda}_{s,y}(\mathbf{s})]\mathbf{u}d\mathbf{u}.$$

Next, we study the behavior of $\sigma_{h,z_1 z_2}$ as $h \to 0$. Since the first term in (4.6) can be ignored, for $z = x$ or $z = y$ there is

$$\lim_{h \to 0} h^4 \sigma_{h,zz} = \lim_{h \to 0} h^3 \int_{\mathcal{D}} [\frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial z}]^2 \lambda_s(\mathbf{u})d\mathbf{u}$$

$$= \lim_{h \to 0} \int_{\mathbb{R}^2} [\frac{\partial K(\mathbf{u})}{\partial z}]^2 \lambda_s(\mathbf{s} - h\mathbf{u})d\mathbf{u} = \lambda_s(\mathbf{s})\xi_z$$

and

$$\lim_{h \to 0} h^4 \sigma_{h,xy} = \lim_{h \to 0} h^4 \int_{\mathcal{D}} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial x} \frac{\partial K_h(\mathbf{s} - \mathbf{u})}{\partial y} \lambda_s(\mathbf{u})d\mathbf{u}$$

$$= \lim_{h \to 0} h^4 \int_{\mathbb{R}^2} K_h^2(\mathbf{s} - \mathbf{u}) \frac{\partial^2 \lambda_s(\mathbf{u})}{\partial x \partial y} d\mathbf{u}$$

$$= 0.$$

Then,

$$\lim_{h \to 0} h^4 \tau_h^2(\mathbf{s}) = \lambda_s(\mathbf{s}) \|\nabla \lambda_s(\mathbf{s})\|^2 (\xi_x \lambda_{s,y}^2(\mathbf{s}) + \xi_y \lambda_{s,y}^2(\mathbf{s})).$$

If $h = cT^{-\frac{1}{8}}$, then

$$T^{\frac{1}{4}}(\hat{\theta}_h(\mathbf{s}) - \theta(\mathbf{s})) = T^{\frac{1}{4}}[(\hat{\theta}_h(\mathbf{s}) - \theta_h(\mathbf{s})) + (\theta_h(\mathbf{s}) - \theta(\mathbf{s}))]$$

$$= \frac{h^2}{c^2}[\sqrt{T}(\hat{\theta}_h(\mathbf{s}) - \theta_h(\mathbf{s}))] + \frac{c^2}{h^2}(\theta_h(\mathbf{s}) - \theta(\mathbf{s}))$$

$$\xrightarrow{D} N(c^2\nu(\mathbf{s}), \tau^2(\mathbf{s})/c^4),$$

as $T \to \infty$. $\qquad\square$

Proof of Theorem 5: Straightforwardly, there is

$$\text{MISE}(\hat{\theta}_h(\mathbf{s})) = \int_{\mathcal{D}} [Bias^2(\hat{\theta}_h(\mathbf{s}) + \text{V}(\hat{\theta}_h(\mathbf{s}))]d\mathbf{s}$$

$$= \frac{1}{T^{1/2}} \int_{\mathcal{D}} [c^4\nu^2(\mathbf{s}) + \tau^2(\mathbf{s})/c^4]d\mathbf{s} + o_p(1/T^{1/2})$$

$$= \frac{c^4}{T^{1/2}} \int_{\mathcal{D}} \nu^2(\mathbf{s})d\mathbf{s} + \frac{1}{c^4 T^{1/2}} \int_{\mathcal{D}} \tau^2(\mathbf{s})d\mathbf{s} + o_p(1/T^{1/2}),$$

where is asymptotically minimized at $h = cT^{-\frac{1}{8}}$ with c given by (4.15). $\qquad\square$

References

[1] Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, **54**, 329–350. MR1804002

[2] Balderama, E., Schoenberg, F. P., Murray, E., and Rundel, P. W. (2012). Application of branching models in the study of invasive species. *Journal of the American Statistical Association*, **107**, 467–476. MR2980058

[3] Bansal, A. R., Dimri, V. P., Babu, K. K. (2013). Epidemic type aftershock sequence (ETAS) modeling of northeastern Himalayan seismicity. *Journal of Seismology*, **17**, 255–264.

[4] Berman, M. and Diggle, P. J. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of Royal Statistical Society Series B*,**51**, 81–92. MR0984995

[5] Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transaction of Pattern Analysis Machine Intelligence*, **8**, 679–698.

[6] Charnigo, R., Hall, B., Srinivasan, C. (2011). A generalized C_p Criterion for derivative estimation. *Technometrics*, **53**, 238–253. MR2857702

[7] Console, R., Murru, M., and Falcone, G. (2010). Probability gains of an epidemic-type aftershock sequence model in retrospective foresting of $M \geq$ 5 earthquakes in Italy. *Journal of Seismology*, **14**, 9–26.

[8] Cox, D. R. (1972). The statistical analysis of dependencies in point process. In *Stochastic Point Processes*, P. A. W. Lewis (ed), 55–66. New York: Wiley. MR0375705

[9] Diggle, P. J. (1983). *Statistical Analysis of Point Processes*. Academic Press, London. MR0743593

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

[11] Diggle, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Statistical Methods in Medical Research*, **16**, 325–336. MR2242245

[12] Fukunaga, K. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions of information theory*, **21**, 32–40. MR0388638

[13] Gauch, J. M. (1999). Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transcations on Image Processing*, **8**, 69–79.

[14] Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York. MR1876599

[15] Guan, Y. (2008). A KPSS test for stationarity for spatial point processes. *Biometrics*, **64**, 800–806. MR2526630

[16] Guan, Y. (2009). On nonparametric variance estimation for second-order statistics of inhomogeneous spatial point processes with known parametric intensity form. *Journal of the American Statistical Association*, **104**, 1482–1491. MR2750573

[17] Guan, Y. and Shen, Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, **97**, 867–880. MR2746157

[18] Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability*, **12**, 141–153. MR0723735

[19] Holden, L., Sannan, S. and Bungum, H. (2003). A stochastic marked point process model for earthquakes. *Natural Hazards and Earth System Sciences*, **3**, 95–101.

[20] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its applications*, **7**, 349–382. MR0148125

[21] Ivanoff, G. (1982). Central limit theorems for point processes. *Stochastic Processes and Their Applications*, **12**, 171–186. MR0651902

[22] Lay, T., et. al (2005). The Great Sumatra-Andaman earthquake of 26 December 2004. *Science*, **308**, 1127–1133.

[23] Meyer, F. and Beucher, S. (1990). Morphological segementation. *Journal of Visual Communication and Image Representation*, **1**, 21–46.

[24] Miller, J. C. (2008). Spread of infectious disease through cluster populations. *Journal of the Royal Society Interfact*, **12**, 1–14.

[25] Nevatie, R. (1986). Image Segmentation. *Handbook of pattern recognition and image processing*. Young, T. Y. and Fu, K. S. ed., Academic Press, chapter 9. MR0937927

[26] Ogata, Y. (1998). Space-time point process models for earthquake occurrences. *Annals of the Institution Statistical Mathematics*, **50**, 379–402.

[27] Ogata, Y. and Zhuang, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics*, **413**, 13–23.

[28] Peng, R. D., Schoenberg, F. P, and Woods, J. A. (2005). A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association*, **100**, 26–35. MR2166067

[29] Ripley, B. D. (1976). The second-order analysis of spatial point processes. *Journal of Applied Probability*, **23**, 255–266. MR0402918

[30] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, **42**, 43–47. MR0074711

[31] Schoenberg, F. P. (2004). Testing separability in spatial-temporal marked point processes. *Biometrics*, **60**, 471–481. MR2066282

[32] Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, **128**, 79–93. MR2110179

[33] Stein, S., Geller, R. J., and Liu, M. (2012). Why earthquake hazard maps often fail and what to do about it. *Tectonophysics*, **562–563**, 1–25.

[34] Stoyan, D. and Stoyan, H. (1996). Estimating pair correlation functions of planar cluster processes. *Biometrical Journal*, **38**, 259–271.

[35] Vere-Jones, D. and Zhuang, J. (2010). Distribution of the largest event in the critical epidemic-type aftershock-sequence model. *Physical Review E*, **78**, 047102. MR2748665

[36] Waagetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott process. *Biometrics*, **63**, 252–258. MR2345595

[37] Waagetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, **92**, 351–363. MR2422695

[38] Zhang, T. and Zhuang, Q. (2014). On the local odds ratio between points and marks in marked point processes. *Spatial Statistics*, **9**, 20–37. MR3326829

[39] Zhang, T. and Zhou, B. (2014). Test for stationarity for spatial point processes in an arbitrary region. *Journal of Agricultural, Biological and Environmental Statistics*, **19**, 387–404. MR3292563

[40] Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, **97**, 369–380. MR1941459