Nonparametric Method for Modeling Clustering Phenomena in Emergency Calls Under Spatial-Temporal Self-Exciting Point Processes

CHENLONG LI [1,2], ZHANJIE SONG[1,3], (Member, IEEE), AND XU WANG[2]

[1]School of Mathematics, Tianjin University, Tianjin 300350, China
[2]Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada
[3]Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China

Corresponding author: Zhanjie Song (zhanjiesong@tju.edu.cn)

ABSTRACT In this paper, a nonparametric spatial-temporal self-exciting point process is proposed to model clustering features in emergency calls. Gaussian kernel density functions are considered. The expectation-maximization algorithm is adopted for estimating the model. A simulation study is designed to carefully examine the performance of the proposed nonparametric method. The spatial-temporal patterns of the emergency calls in Montgomery County of Pennsylvania are studied using the proposed nonparametric model. The results demonstrate that the proposed nonparametric model captures the clustering phenomena present in the emergency calls from Montgomery County very well. Further, the proposed parameter estimation method results in robust and precise estimates.

INDEX TERMS Spatial-temporal point processes, emergency calls, nonparametric model, maximum likelihood estimation, expectation-maximization algorithm.

I. INTRODUCTION

With the growth of major cities in hazard prone areas, many researches have focused on improving the system of early hazard warnings and emergency responses [7]. Many different types of crime data have been extensively studied for analyzing hazard prone areas in the framework of random point processes [8], [11], [14], [17], [18], [22]–[24], [32]. One important characteristic of crime data is the clustering feature in the dimensions of space and time [17], [18], [22], [24].

Self-exciting point processes were developed and discussed in the context of seismology with the Neymann-Scott models and other cluster process models [37]. The explicit form of a self-exciting point process was formally defined by [15]. The Epidemic Type Aftershock Sequence (ETAS) model developed in seismology is an early important application of self-exciting point processes [27]. Self-exciting point processes perform particularly well in modeling

The associate editor coordinating the review of this manuscript and approving it for publication was Jihad Aljaam.

earthquakes due to the dependence between the main shocks and aftershocks [20]. The dependence between the major and subsequence events has also been noted in criminology [17], [18], [24]. Many factors play roles in the connection between one main crime and subsequent ones [24]. Lewis [18] pointed out that crimes may be correlated due to exogenous factors such as the state of economy, the month of the year, the change in military operations, etc., rather than "caused" by endogenous factors such as repeated offending behaviors. For example, burglars may repeatedly attack clusters of nearby targets because the local vulnerabilities are known to them [4]. A gang shooting may incite waves of retaliatory violence in a local region of the rival gang [33]. The contagious spread of crimes leads to the presence of clustering features in both space and time [24]. To account for the clustering features observed in crime data, self-exciting point processes were used to describe random collections of crimes where the occurrence of one crime increases the likelihood of other crimes occurring shortly thereafter [24].

However, there is limited research on modeling emergency calls with stochastic models. Wang *et al*. [38] performed a simple descriptive statistical analysis for emergency calls of a metropolitan city in China. The application of spatial-temporal self-exciting point processes on modeling and characterizing emergency calls is a relatively new research field. Emergency call data provide the information on when and where people encounter emergent events. One emergency call may be associated with many other emergency calls within a short time period. This dependency among emergency calls is because sometimes emergency calls are the quick actions of the insiders who were involved in the emergencies. The insiders can be the real triggers, the witnesses or the indirect insiders who are told about the already happened emergencies. Studying emergency calls allows us to understand how human beings react to emergent events, and how the news of accidents is spread in order to identify and characterize hazard prone areas.

In this paper, we focus on analyzing and modeling an interesting emergency call dataset from Montgomery county of Pennsylvania. We explore the data with a nonparametric method, which allows us to construct a flexible spatial-temporal self-exciting point process. The proposed nonparametric model introduces some scaling parameters in both the background and triggering intensity functions. The intensity functions can be expressed as the product of the scaling parameters and densities, and the densities are formatted using Gaussian kernel density functions. Viewing the estimation of the proposed self-exciting point process as an incomplete data problem (see Section III-C), the Expectation and Maximization algorithm (EM) is employed to obtain the maximum likelihood estimates (MLEs) of the scaling parameters. With the proposed nonparametric spatial-temporal self-exciting point process along with the EM algorithm, we seek to tackle the following three questions:

(a) Does the proposed nonparametric model work flexibly and effectively for spatial-temporal self-exciting point processes?

(b) Are the clustering phenomena present in emergency calls? If they do exist, what are the characteristics of the clustering phenomena?

(c) Can the proposed nonparametric spatial-temporal self-exciting point process model emergency calls very well?

For the first question, we use a parametric spatial-temporal self-exciting point process to simulate a toy catalog [24], and then examine the flexibility and effectiveness of the proposed nonparametric version. The simulation algorithm proposed in [43] is adopted to generate the simulated data. For the second question, the clustering features are studied using the estimated triggering intensity function. For the third question, we use a typical residual analysis, which is a simple and efficient model diagnostic method [31], to examine the performance of the proposed nonparametric spatial-temporal self-exciting point process in modeling emergency calls. The contributions of this paper are: (1) The MLEs of the scaling parameters have explicit formulae; (2) A novel explanation

of the proposed nonparametric spatial-temporal self-exciting point process is given based on the varied bandwidth kernel density estimation (KDE) method; (3) A detailed analysis of point processes provides insights about spreading patterns of the information in emergencies by human beings and the characteristics of hazard prone areas.

The rest of this paper is organized as follows. Section II describes the framework of spatial-temporal self-exciting point processes. Section III proposes a nonparametric spatial-temporal self-exciting point process. In this section, the MLEs of the nonparametric spatial-temporal self-exciting point process are derived based on the EM, and the overall Goodness-of-fit as well as a simulation algorithm for the estimated spatial-temporal self-exciting point process model are also introduced. Section IV provides the performance of the proposed method using a simulated dataset. The proposed model is implemented on an interesting emergency call dataset from Montgomery county, Montana, in Section V. Finally, Section VI summarizes the results and future research directions.

II. SPATIAL-TEMPORAL SELF-EXCITING POINT PROCESSES

A spatial-temporal point process X is a random collection of points with each point falling in an observed metric space $S \times T \subseteq \mathbb{R}^2 \times \mathbb{R}$. A spatial-temporal point process is uniquely determined by specifying its intensity process [19]. In a more general case the distributions of intensity functions of random point processes are conditional, not only on the time since the last event, but also on any additional information regarding the past history that may affect the distribution of the remaining times.

Let N be a simple counting process and $\mathscr{H}_t$ be the collection of all events observed within the time interval $(-\infty, t), t \in T$. The conditional intensity process $\lambda(\mathbf{s}, t)$ of a spatial-temporal point process is the expected rate describing the frequency of points occurring around space location $\mathbf{s}$ and time t, conditional on the history $\mathscr{H}_t$, $t \in T$, consisting of the set of location and time of all events that occur prior to time t. In other words, $\mathscr{H}_t$ is a family of sigma-algebras generated by the events occurring at times up to, but not including t. The definition of conditional intensity process is given in Equation (1), if the limits in Equation (1) exist.

$$\begin{cases} \lim_{\substack{\triangle \mathbf{s} \to 0 \\ \triangle t \to 0}} \dfrac{P(N([\mathbf{s}, \mathbf{s} + \triangle \mathbf{s}) \times [t, t + \triangle t)) = 1 | \mathscr{H}_t)}{\triangle \mathbf{s} \triangle t} = \lambda(\mathbf{s}, t | \mathscr{H}_t), \\ \lim_{\substack{\triangle \mathbf{s} \to 0 \\ \triangle t \to 0}} \dfrac{P(N([\mathbf{s}, \mathbf{s} + \triangle \mathbf{s}) \times [t, t + \triangle t)) > 1 | \mathscr{H}_t)}{\triangle \mathbf{s} \triangle t} = 0, \end{cases}$$

$$(1)$$

where $\mathbf{s} := (x, y) \in S$ represents a space location.

The critical problem of modeling such point processes is to determine how the conditional intensity process depends on the past history [36]. Typically, it is dealt by specifying special structures for the conditional intensity process. The self-exciting point process models are one important kind of

conditional intensity process models [19]. Given observed events with location $\mathbf{s}_i$ and time t_i up to time t ($i = 1, 2, 3 \cdots$), we have the following definition for the conditional intensity process of a spatial-temporal self-exciting point process.

Definition 1: Given an observed event sequence $(\mathbf{s}_i, t_i)$, $i = 1, 2, \ldots$, and $t_i < t$, a spatial-temporal self-exciting point process is simply a point process N such that N has a conditional intensity process written in Equation 2.

$$\lambda(\mathbf{s}, t|\mathscr{H}_t) = \mu(\mathbf{s}, t) + \int_{S \times (-\infty, t)} g(\mathbf{s} - \boldsymbol{\xi}, t - \varsigma)N(d\boldsymbol{\xi}, d\varsigma)$$
$$:= \mu(\mathbf{s}, t) + \sum_{i: t_i < t} g(\mathbf{s} - \mathbf{s}_i, t - t_i), \quad (2)$$

for $(\mathbf{s}, t) \in S \times T$, where $N(d\boldsymbol{\xi}, d\varsigma) = 1$ if an infinitesimal element $(d\boldsymbol{\xi}, d\varsigma)$ includes an event $(\mathbf{s}_i, t_i)$ for some i, otherwise $N(d\boldsymbol{\xi}, d\varsigma) = 0$.

Equation (2) is closely related to a branching process, i.e., each point of a self-exciting process is either an **immigrant (background)** or a **descendant (offspring or triggering)** [34]. A point of a self-exciting process occurs at location $\mathbf{s}_i \in S$ and time $t_i \in \mathbb{R}$ is called an immigrant if it is viewed as a point generated from an inhomogeneous Poisson process with the intensity function $\mu(\mathbf{s}_i, t_i)$. Then, it generates offsprings at future location $\mathbf{s}$ and time $t > t_i$ from an inhomogeneous Poisson process with the intensity function $g(\mathbf{s} - \mathbf{s}_i, t - t_i)$. All offsprings related to this immigrant are called the descendants of this immigrant $(\mathbf{s}_i, t_i)$. Each immigrant generates either zero or more descendants independently. The immigration intensity function $\mu(\mathbf{s}, t)$ governs the frequency at which new immigrants arrive. Whenever a point event occurs, it is either an immigrant or a descendant, and the conditional intensity process is increased temporarily, i.e., events arrive at a higher frequency for certain time windows. The increase in intensity causes secondary point events, which in turn can spawn descendants of their own. How fast this effect decays in time is governed by the triggering intensity function $g(x, y, t)$.

In most applications of self-exciting point processes, the background intensity function $\mu(x, y, t)$ is assumed to be stationary [15], [20], [27], [40], [42]. Existing works often restrict the triggering intensity function as an exponential function in time [1], [2], [13]. The distribution in space of the triggering intensity function, according to the research fields of seismicity, crime and security, is often considered as an isotropic Gaussian kernel [24], [28], [42]. A more flexible approach was proposed using a set of basic functions to explore the triggering intensity function [35], [39], [41], where the coefficients and even the basic functions [41] are iteratively updated and refined. An alternative way to estimate the triggering intensity function non-parametrically was proposed in [3] by solving a set of p Wiener-Hopf systems in p^2 dimensions. Zhuang *et al.* [42] considered using a varied bandwidth kernel density function method to estimate the background intensity function. Marsan and Lengline [20] proposed a complete nonparametric method, namely the model-independent declustering algorithm (MISD),

for estimating both the background and triggering intensity functions. Fox *et al.* [12] extended this method to the cases in which the background intensity function is inhomogeneous.

Given the conditional intensity process defined in Equation (2) over the observation period $D := [t_*, t^*)$, the parameter estimates can be obtained by maximizing the log-likelihood function [10] written in Equation (3)

$$\log L = \int \int \int_{S \times D} \log \lambda(\mathbf{s}, t) N(d\mathbf{s}, dt)$$
$$- \int \int \int_{S \times D} \lambda(\mathbf{s}, t) dx dy dt$$
$$:= \sum_{i=1}^{n} \log \lambda(\mathbf{s}_i, t_i) - \int \int \int_{S \times D} \lambda(\mathbf{s}, t) dx dy dt. \quad (3)$$

Note that the log-likelihood function depends on the choice of the observation period D. The history $\mathscr{H}_t$ defined in Equation (2) is adjusted as a collection of all events observed during the time interval $[t_*, t), t \leq t^*$.

III. NONPARAMETRIC MODEL AND MODEL DIAGNOSTICS

In this section, we consider spatial-temporal self-exciting point processes with inhomogeneous background intensity functions. We first present the proposed nonparametric model and then explore the estimation method based on the EM algorithm. To examine the Goodness-of-fit of the proposed nonparametric model, the residual analysis with the thinning method proposed in [31] is adopted.

A. NONPARAMETRIC MODEL

In an exploratory context for modeling point events, parametric models are not always flexible enough to cover all the possible structures of the intensity processes, therefore nonparametric models are very useful. In this paper the spatial-temporal self-exciting point process model considered has an inhomogeneous background intensity function $\mu(x, y, t)$ and a nonparametric triggering intensity function $g(x, y, t)$. One can assume that the spatial-temporal background intensity function $\mu(x, y, t)$ can be separated as

$$\mu(x, y, t) = \alpha u(x, y)v(t), \quad (4)$$

where α is a positive scaling factor controlling the overall background intensity function [24], [35], [42]; $\int_{\mathbb{R}^2} u(x, y) dx dy = 1$ and $\int_{\mathbb{R}} v(t) dt = 1$. Then the varied bandwidth KDE can be used to estimate $u(x, y)$ and $v(t)$ as

$$u(x, y) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_{1,i}}(x - x_i^b, y - y_i^b), \quad (5)$$

$$v(t) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_{2,i}}(t - t_i^b), \quad (6)$$

where $\{(x_i^b, y_i^b, t_i^b)\}_{i=1}^{n_b}$ represents the background events; n_b represents the number of background events; $d_{1,i}$ and $d_{2,i}$ are the varied bandwidths of space and time respectively calculated for each background event i; $K_{d_{1,i}}(x - x_i^b, y - y_i^b)$

and $K_{d_{2,i}}(t - t_i^b)$ denote Gaussian kernel functions:

$$K_{d_{1,i}}(x - x_i^b, y - y_i^b) = \frac{1}{2\pi \sigma_x^b \sigma_y^b (d_{1,i})^2} \exp\{-\frac{(x - x_i^b)^2}{2(\sigma_x^b d_{1,i})^2} - \frac{(y - y_i^b)^2}{2(\sigma_y^b d_{1,i})^2}\},$$

$$K_{d_{2,i}}(t - t_i^b) = \frac{1}{\sqrt{2\pi} \sigma_t^b d_{2,i}} \exp\{-\frac{(t - t_i^b)^2}{2(\sigma_t^b d_{2,i})^2}\},$$

where $\sigma_x^b, \sigma_y^b, \sigma_t^b$ are the standard deviations of the coordinates of the background event points $\{(x_i^b, y_i^b, t_i^b)\}$. The bandwidths $d_{1,i}$ and $d_{2,i}$ are computed by finding the radius of the smallest disk centered at the scaling data of $\{(x_i^b, y_i^b)\}$ and $\{t_i^b\}$, which have a unit variance in each coordinate, and the smallest disk contains at least n_1 and n_2 extra events, respectively.

Furthermore, one can split the spatial-temporal triggering intensity function $g(x, y, t)$ into two parts:

$$g(x, y, t) = \beta h(x, y, t), \tag{7}$$

where $\beta < 1$ is a positive scaling factor and $\int \int \int_{\mathbb{R}^2 \times \mathbb{R}} h(x, y, t) dx dy dt = 1$. Function $h(x, y, t)$ can be viewed as the underlying joint probability density function for both space and time between a background event and its triggering events. Then the varied bandwidth KDE of $h(x, y, t)$ can be given by

$$h(x, y, t) = \frac{1}{n_o} \sum_{i=1}^{n_o} K_{d_{3,i}}(x - x_i^o, y - y_i^o, t - t_i^o), \tag{8}$$

where $\{(x_i^o, y_i^o, t_i^o)\}_{i=1}^{n_o}$ represent the inter-point distances between the triggering events and their parent events; n_o represents the number of the triggering events; $K_{d_{3,i}}(x - x_i^o, y - y_i^o, t - t_i^o)$ denotes Gaussian kernel functions written as

$$K_{d_{3,i}}(x - x_i^o, y - y_i^o, t - t_i^o)$$
$$= \frac{1}{\sigma_x^o \sigma_y^o \sigma_t^o (\sqrt{2\pi} d_{3,i})^3} \exp\{-\frac{(x - x_i^o)^2}{2(\sigma_x^o d_{3,i})^2} - \frac{(y - y_i^o)^2}{2(\sigma_y^o d_{3,i})^2} - \frac{(t - t_i^o)^2}{2(\sigma_t^o d_{3,i})^2}\},$$

where $\sigma_x^o, \sigma_y^o, \sigma_t^o$ are the sample standard deviations of the coordinates of inter-point distances $\{(x_i^o, y_i^o, t_i^o)\}$; the bandwidth $d_{3,i}$ is computed by finding the radius of the smallest disk centered at the scaling data of $\{(x_i^o, y_i^o, t_i^o)\}$, which has a unit variance in each coordinate, and the smallest disk contains at least n_3 extra events. Combining Equations (4)-(8), we propose a general form of nonparametric models as

$$\lambda(x, y, t | \mathcal{H}_t) = \alpha u(x, y) v(t)$$
$$+ \sum_{i: t_i < t} \beta h(x - x_i, y - y_i, t - t_i). \tag{9}$$

B. MAXIMUM LIKELIHOOD ESTIMATION
Since the branching structure is unknown, we can view the estimation of self-exciting point processes as incomplete data problems. Then we use the EM algorithm to attain the maximum likelihood estimates (MLEs) of the scaling

parameters (α, β) and update the kernel densities. The EM algorithm has been widely used in estimating the background and triggering intensity functions for both parametric and nonparametric models [12], [13], [16], [18], [20], [24], [34]. A good overview of the EM algorithm and its extensions is provided by [21]. In the rest of this section, we derive the MLEs of the proposed nonparametric model based on the EM algorithm.

Suppose we have observed a realization of a spatial-temporal self-exciting point process, with event locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and times $\{t_1, \ldots, t_n\}$ over a spatial region S and a temporal window D. Define random variables

$$\zeta_i = \begin{cases} i & \text{if event } i \text{ is a background event,} \\ j & \text{if event } i \text{ is triggered by event } j, i \neq j. \end{cases} \tag{10}$$

If a branching structure is incorporated, the complete data log-likelihood can be decomposed into the likelihood functions for the background and triggering events separately as

$$\log L^c(\Theta)$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \mathbf{1}_{\{\zeta_i = j, j = i\}} \log(\mu(x_i, y_i, t_i))$$
$$- \int \int \int_{S \times D} \mu(x, y, t) dx dy dt$$
$$+ \sum_{j=1}^{n} \left[\sum_{i: t_j < t_i} \mathbf{1}_{\{\zeta_i = j\}} \log(g(x_i - x_j, y_i - y_j, t_i - t_j)) \right.$$
$$\left. - \int \int \int_{S \times [t_j, t^*)} g(x - x_j, y - y_j, t - t_j) dx dy dt \right]$$
$$= \sum_{i=1}^{n} \mathbf{1}_{\{\zeta_i = i\}} \log(\mu(\mathbf{s}_i, t_i))$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}_{\{\zeta_i = j\}} \log(g(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j))$$
$$- \int \int \int_{S \times D} \lambda(\mathbf{s}, t) dx dy dt, \tag{11}$$

where $\Theta = \{\zeta_i, i = 1, \ldots, n; \Psi\}$ and $\Psi = \{\alpha, u(x, y), v(t), \beta, h(x, y, t)\}$; $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Let $\Psi^{(k)} := \{\alpha^{(k)}, u^{(k)}(x, y), v^{(k)}(t), \beta^{(k)}, h^{(k)}(x, y, t)\}$ be the set of values of Ψ at the kth iteration. Then the E-step and the M-step can be calculated as:

E-Step Calculating $Q(\Psi; \Psi^{(k)})$

$$Q(\Psi; \Psi^{(k)})$$
$$= E_{\Psi^{(k)}}\{\log L^c(\Theta) | (\mathbf{x}, \mathbf{y}, \mathbf{t})\}$$
$$= \sum_{i=1}^{n} p_{ii}^{(k)} \log(\mu(x_i, y_i, t_i))$$
$$- \int \int \int_{S \times D} \mu(x, y, t) dx dy dt$$
$$+ \sum_{j=1}^{n} \left[\sum_{i: t_j < t_i} p_{ij}^{(k)} \log(g(x_i - x_j, y_i - y_j, t_i - t_j)) \right.$$
$$\left. - \int \int \int_{S \times [t_j, t^*)} g(x - x_j, y - y_j, t - t_j) dx dy dt \right], \tag{12}$$

where

$$p_{ij}^{(k)} := E(\mathbf{1}_{\{\zeta_i = j\}})$$

$$= \begin{cases} \dfrac{\mu^{(k)}(x_i, y_i, t_i)}{\lambda^{(k)}(x_i, y_i, t_i)}, & j = i, \\[2ex] \dfrac{g^{(k)}(x_i - x_j, y_i - y_j, t_i - t_j)}{\lambda^{(k)}(x_i, y_i, t_i)}, & t_i > t_j, j > 0, \\[2ex] 0, & t_i < t_j, j > 0; \end{cases} \tag{13}$$

and $\mu^{(k)}(x, y, t) = \alpha^{(k)} u^{(k)}(x, y) v^{(k)}(t)$; $g^{(k)}(x, y, t) = \beta^{(k)} h^{(k)}(x, y, t)$; $\lambda^{(k)}(x, y, t) = \mu^{(k)}(x, y, t) + \sum_{i:t_i < t} g^{(k)}(x - x_i, y - y_i, t - t_i)$;

M-Step Maximizing $Q(\Psi; \Psi^{(k)})$

The probabilities p_{ij}, $i, j \in \{1, 2, \ldots, n\}$, allow estimating $u^{(k+1)}(x, y)$, $v^{(k+1)}(t)$ and $h^{(k+1)}(x, y, t)$ using kernel density functions as

$$u^{(k+1)}(x, y) = \frac{\sum_{i=1}^n p_{ii}^{(k)} K_{d_{1,i}}(x - x_i, y - y_i)}{\sum_{i=1}^n p_{ii}^{(k)}},$$

$$v^{(k+1)}(t) = \frac{\sum_{i=1}^n p_{ii}^{(k)} K_{d_{2,i}}(t - t_i)}{\sum_{i=1}^n p_{ii}^{(k)}},$$

$$h^{(k+1)}(x, y, t)$$
$$= \frac{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)} K_{d_{ij}}(x - x_{ij}, y - y_{ij}, t - t_{ij})}{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)}},$$

where

$$\widehat{\sigma}_x^b = \frac{1}{\sum_{i=1}^n p_{ii}^{(k)}} \sum_{i=1}^n p_{ii}^{(k)} \left(x_i - \frac{\sum_{i=1}^n p_{ii}^{(k)} x_i}{\sum_{i=1}^n p_{ii}^{(k)}} \right)^2,$$

then $\widehat{\sigma}_y^b$ and $\widehat{\sigma}_t^b$ are estimated similarly;

$$\widehat{\sigma}_x^o = \frac{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)} \left(x_{ij} - \frac{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)} x_{ij}}{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)}} \right)^2}{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)}},$$

then $\widehat{\sigma}_y^o$ and $\widehat{\sigma}_t^o$ are estimated similarly; $x_{ij} := x_i - x_j$, $y_{ij} := y_i - y_j$ and $t_{ij} := t_i - t_j$.

Alternatively, a Monte Carlo-based method proposed in [24] can be used to gain efficiency of using KDEs when the data size is large. One can estimate $u^{(k+1)}(x, y)$, $v^{(k+1)}(t)$ and $h^{(k+1)}(x, y, t)$ as

$$u^{(k+1)}(x, y) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_{1,i}}(x - x_i^b, y - y_i^b), \tag{14}$$

$$v^{(k+1)}(t) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_{2,i}}(t - t_i^b), \tag{15}$$

$$h^{(k+1)}(x, y, t) = \frac{1}{n_o} \sum_{i=1}^{n_o} K_{d_{3,i}}(x - x_i^o, y - y_i^o, t - t_i^o), \tag{16}$$

where the background events $\{x_i^b, y_i^b, t_i^b\}_{i=1}^{n_b}$ and the triggering-parent inter-point distances $\{x_i^o, y_i^o, t_i^o\}_{i=1}^{n_o}$ are sampled from $p_{ij}^{(k)}$, $i, j \in \{1, \ldots, n\}$; n_b is the number of the

background events, while n_o is the number of the triggering events.

One can calculate the scaling parameters by equaling the corresponding partial derivatives to zero, i.e.,

$$\beta^{(k+1)} = \frac{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)}}{\sum_{j=1}^n \int \int \int_{S \times [t_j, t^*]} h^{(k)}(x - x_j, y - y_j, t - t_j) dx dy dt}, \tag{17}$$

$$\alpha^{(k+1)} = \frac{\sum_{i=1}^n p_{ii}^{(k)}}{\int \int \int_{S \times D} u^{(k)}(x, y) v^{(k)}(t) dx dy dt}. \tag{18}$$

To simplify the above computation, we adopt the following approximation Equations (19) and (20). The integral terms of Equation (17) can be approximated by 1 using the fact that the spatial-temporal distances between the triggering events and their parent events are usually much smaller than the study region $S \times D$. This approximation was also considered in [34]. For the similar reasoning, the integral term of Equation (18) can also be approximated by 1. Using these approximations, we have

$$\beta^{(k+1)} \approx \frac{\sum_{j=1}^n \sum_{i=j+1}^n p_{ij}^{(k)}}{n} \approx \frac{n_o}{n}, \tag{19}$$

$$\alpha^{(k+1)} \approx \sum_{i=1}^n p_{ii}^{(k)} \approx n_b. \tag{20}$$

Algorithm 1 summarizes our proposed Monte Carlo-based EM algorithm for estimating the parameters of the proposed nonparametric model.

Algorithm 1 Estimation

Input: Event locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and times $\{t_1, \ldots, t_n\}$. Initialize $P^{(0)} := (p_{ij}^{(0)})$ by

$$P^{(0)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1/2 & 1/2 & 0 & \cdots & 0 & 0 \\ 1/4 & 1/4 & 1/2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1/(2(n-1)) & 1/(2(n-1)) & 1/(2(n-1)) & \cdots & 1/(2(n-1)) & 1/2 \end{bmatrix}_{n \times n}$$

Output: The parameters $\Psi^{(k)}$
−① Sample background events $\{x_j^b, y_j^b, t_j^b\}_{j=1}^{n_b}$ and triggering-parent inter-point distances $\{x_i^o, y_i^o, t_i^o\}_{i=1}^{n_o}$ from $P^{(k-1)}$;
−② Estimate $u^{(k)}(x, y)$, $v^{(k)}(t)$ and $\alpha^{(k)}$ from the sampled data using Equations (14), (15) and (20);
−③ Estimate $h^{(k)}(x, y, t)$ and $\beta^{(k)}$ from the sampled data using Equations (16) and (19);
−④ Update the probabilities in matrix $P^{(k)} = (p_{ij}^{(k)})$ using Equation (13)
−⑤ If the errors $L_2 := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (p_{ij}^{(k+1)} - p_{ij}^{(k)})^2 < \varepsilon$, then the algorithm is converged (in practice we take $\varepsilon = 10^{-4}$); Otherwise, repeat Steps ①-⑤ until convergence.

Remark 1: In [42], zhuang et al. suggested choosing a value between 15 and 100 for n_1. As well as in [24], both n_2 and n_3

take values between 15-100. To overcome the over-fitting problem, the bandwidths should be greater than some small values due to measurement errors (e.g., the lower bound of space radiuses may be set as 0.02 km and the lower bound of time radiuses may be set as 0.05 h). In addition, $P^{(0)}$ is one possible initialization. One can use other initial matrices (see the examples in [12]).

Remark 2: Other approaches used to estimate the background and triggering intensity functions have been discussed in [42], [20], [24], and [12]. In [42], Zhuang *et al.* used semi-parametric models and estimated the parameters using the iterative numerical optimization methods [28]. In [24], a nonparametric estimate of $g(x, y, t)$, which was consistent with our approximate estimate, was adopted directly without explanation. In [12] and [20], Fox *et al.* and Marsan and Lengline considered using piecewise constant functions to model the background and triggering intensity functions, and then using histogram estimators for parameter estimation.

C. MODEL DIAGNOSTICS AND SIMULATION

In this section, we examine the Goodness-of-fit of the proposed nonparametric spatial-temporal self-exciting point process for emergency calls. The methods for evaluating the Goodness-of-fit are presented using residual analysis methods, such as thinning, re-scaling, and superposition, which involve transforming point processes with conditional intensity functions and then inspecting the uniformity of the results [9]. Residual analysis methods can help identify defects in spatial-temporal models and suggest ways in which the models may be improved [31].

1) OVERALL GOODNESS-OF-FIT

There are many model diagnostic methods for self-exciting point processes [5], [6], [29], [31]. We decided to use the residual analysis methods with the thinning idea proposed in [31] since it is simple and efficient. The thinning method uses the property that any process characterized by its conditional intensity process may be thinned to obtain a homogeneous Poisson process [26]. The following procedure is a standard thinning algorithm [31]:

(1) Denote $k = \inf_{(x_i, y_i, t_i)} \widehat{\lambda}(x_i, y_i, t_i)$, i.e., the largest lower bound (infimum) of $\lambda(x_i, y_i, t_i)$;
(2) For event i, calculate the quantity $p_i = \frac{k}{\widehat{\lambda}(x_i, y_i, t_i)}$;
(3) Retain event i with probability p_i.

For the thinning process, one can use Ripley's K-function [30], which calculates the proportion of events per unit area within a given distance, to perform model diagnostics. This method detects whether the thinning process still has clusters not accounted by the model [31]. The most commonly used K-function with edge-corrected estimators is given in Equation (21) [30]:

$$\widehat{K}(d) = |S| n^{-2} \sum_i \sum_{j \neq i} w(\mathbf{s}_i, \mathbf{s}_j)^{-1} \mathbf{1}_{d_{ij} < d}, \qquad (21)$$

where $|S|$ is the volume of the observation region, d_{ij} is the distance between the ith and jth points, and the weight

function, $w(\mathbf{s}_i, \mathbf{s}_j)$ is the proportion of the circumference of a circle, which has center $\mathbf{s}_i$ and radius $||\mathbf{s}_j - \mathbf{s}_i||_2$ that lies in the study area. Further, the L−function, estimated by $L(d) = \sqrt{\frac{\widehat{K}(d)}{\pi}}$ and based on $\widehat{K}(d)$, has a more stable variance than that of Ripley's K−function. Thus we use L−function to perform model diagnostics. Algorithm 2 gives the diagnostic procedure based on the Monte Carlo simulation and the L−function.

Algorithm 2 Diagnosis

Input: Initial tuning parameter k, the estimated intensity function $\widehat{\lambda}(x, y, t)$, a number of simulations M and a discrete value d

Output: The empirical 95% confidence bounds of the L-function

−① Using the thinning method to obtain a thinning process;

−② Calculate the value of the L-function;

−③ Repeat Steps ① and ② M times;

−④ Calculate the empirical mean of the L-function;

−⑤ Simulate a homogeneous Poisson process with intensity k over the observed region S and repeat Step ② M times;

−⑥ Calculate the empirical 95% confidence bounds of the L-function corresponding to M realizations of the homogeneous Poisson process.

---

2) SIMULATION

Here we introduce a simple and efficient simulation method proposed by [43]. For a thorough review of simulation methods and the rationality of the thinning method, please see [10], [26], and [25].

IV. NUMERICAL EXPERIMENTS
A. SIMULATED DATA

In this section, a simulated data generated from the conditional intensity function as defined in Equation (22) with the size around 6000 is used for verifying the performance of the proposed method. Number 6000 is a common data size used in most applications of earthquakes and crimes [24], [42]. The background and triggering intensity functions for this simulation are written as:

$$\begin{cases} \mu(x, y, t) = \dfrac{\overline{\mu}}{2\pi(\sigma_\mu)^2} \exp(-\dfrac{(x-c)^2}{2\sigma_\mu^2} - \dfrac{(y-d)^2}{2\sigma_\mu^2}), \\ g(x, y, t) = \dfrac{\theta\omega}{2\pi\sigma_x\sigma_y} \exp(-\omega t - \dfrac{x^2}{2\sigma_x^2} - \dfrac{y^2}{2\sigma_y^2}), \end{cases} \quad (22)$$

where $\overline{\mu} = 5.71$, $\sigma_\mu = 4.5$, $c = 10$, $d = 10$, $\theta = 0.2$, $\omega^{-1} = 10$, $\sigma_x = 0.01$ and $\sigma_y = 0.1$. The simulation is carried out using **Algorithm 3** in a 20 by 20 region of the space. In order to have a realization of the point process at a steady state, the first and last 2000 points are discarded in each simulation. Mohler *et al.* [24] used a similar conditional intensity process for their simulation study. However their triggering

Algorithm 3 Simulation

Input: The estimated intensity functions $\widehat{\mu}(x, y, t)$ and $\widehat{g}(x, y, t)$

Output: Data simulation

−① Calculate $\widehat{m} = \int \int_{\mathbb{R}^2} \int_{\mathbb{R}} \widehat{g}(x, y, t)dtdxdy$;

−② Generate events from the background with the intensity function $\widehat{\mu}(x, y, t)$ over region $S \times D$, using the thinning method. Name this catalog of events $G^{(0)}$;

−③ Let $l = 0$;

−④ For event i in $G^{(l)}$, simulate its $N^{(i)}$ triggering events where $N^{(i)} \sim \text{Poisson}(\widehat{m})$. The triggering events' location and time are generated using the probability density function $\widehat{h}$, normalized from the triggering intensity function $\widehat{g}$. Name these triggering events $O_i^{(l)}$;

−⑤ Let $G^{(l+1)} = \bigcup_{i \in G^{(l)}} O_i^{(l)}$;

−⑥ If $G^{(l)}$ is not empty, set $l = l + 1$ and return to Step ④. Otherwise, return $G = \bigcup_{j=0}^{l} G^{(j)}$;

−⑦ Discard the points in G falling in $\mathbb{R}^2 \times \mathbb{R} \setminus S \times [t_*, t^*)$, the rest points are the final set of simulated events.

intensity function $g(x, y, t)$ has $\int \int \int_{\mathbb{R}^3} g(x, y, t)dxdydt > 1$, which contradicts the definition of conditional intensity function [19]. We correct this contradiction and proposed the triggering intensity function as Equation (22).

B. FITTING SIMULATED DATASET

We estimate the conditional intensity process Equation (9) based on the simulated datasets generated from Equation (22) using **Algorithm 1**. The tuning parameters of the varied bandwidth KDEs, i.e., the bandwidths $d_{1,i}$, $d_{2,i}$ and $d_{3,i}$, are set as (15, 100, 15) based on the previous research in [24] and [42].

In Fig. 1, the errors L_2, the log-likelihood value and the branching coefficient β are plotted against the number of iterations. We observe that the errors converge quickly within the first 10 iterations and then stabilize. It can be seen that the log-likelihood value converges to some maximum value. The branching coefficient converges to the true value.

In Fig. 2, we plot the estimated marginal densities of the background and triggering intensity functions against the actual distributions at the 50th iteration of one implementation of the Monte Carlo sampling. We observe that the estimated marginal densities of the background and triggering events are close to the true densities except that the edge effect of KDEs makes the estimated marginal densities deviate from the true densities at the boundary.

Fig. 3 shows the estimated values of the centered $L-$function, $L(d)-d$, and the 95% confidence intervals of the homogeneous Poisson process. For a homogeneous Poisson process, $L(d) - d = 0$, so the departure from 0 indicates inhomogeneity. It can be seen that the thinning residuals are evidently homogeneous as the estimated values of the
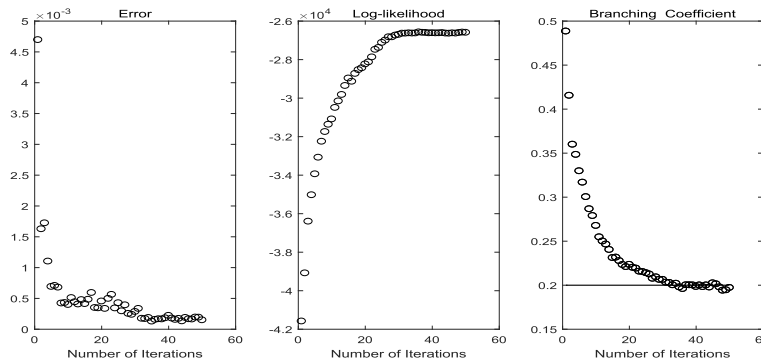

FIGURE 1. The estimated (red lines) and actual (black lines) marginal densities of the background (top) and triggering (bottom) intensity functions.
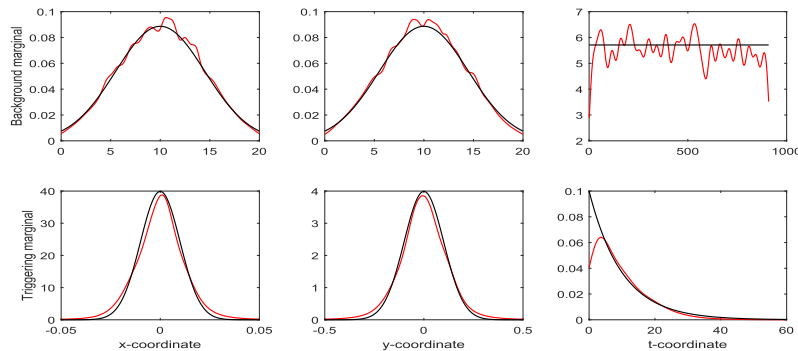

FIGURE 2. The error L_2 (left), the log-likelihood (middle) and the branching coefficient β (right) are plotted against the number of iterations for both background and triggering marginals.
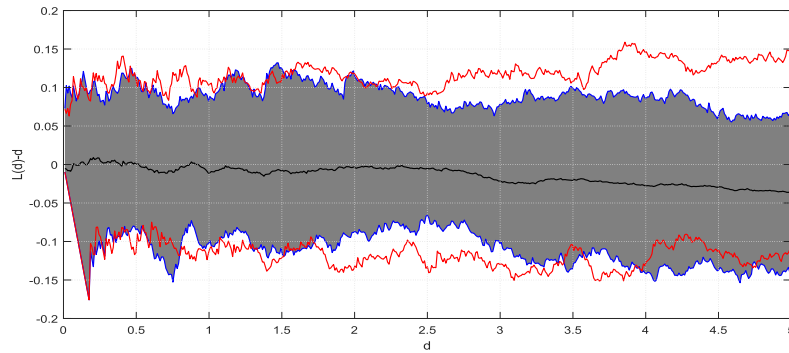
FIGURE 3. Residual analysis with the thinning method. Lower red: the 5% confidence bound of the estimated values of the centered $L-$function, $L(d) - d$, of the homogeneous Poisson process; Upper red: the 95% confidence bound of the estimated values of the centered $L-$function of the homogeneous Poisson process; Lower green: the 5% confidence bound of the estimated values of the centered $L-$function of the thinning residuals; Upper green: the 95% confidence bound of the estimated values of the centered $L-$function of the thinning residuals; Middle black: the empirical mean of the estimated values of the centered $L-$function of the thinning residuals; Gray region: the confidence region of the estimated values of the centered $L-$function of the thinning residuals.
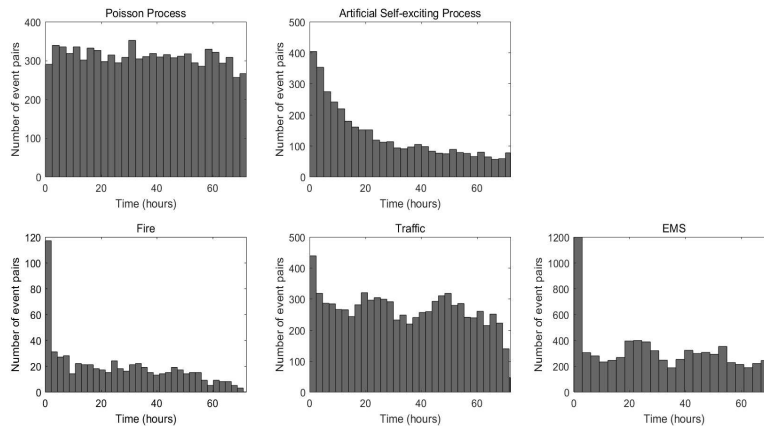

FIGURE 4. Histograms of time (less than 72 hours) of the homogeneous Poisson process (top left), the simulated self-exciting point process (top right) and emergency calls between events separated by 200 meters or less (bottom).

centered $L-$function, are entirely within the 95% confidence bounds.

V. EMERGENCY CALL DATA ANALYSIS

A. DATA DESCRIPTION

Montgomery County is a county located in Pennsylvania. The data collected by the Montgomery County government contain 140545 reported emergency calls occurring in a rectangular area between longitudes $-76°$ and $-74.8°$ and latitudes $39.9°$ and $40.5°$ during September 20, 2016, and September 20, 2017. The data include the information of occurrence time stamps, longitudes, latitudes, emergency descriptions, addresses, etc., and consist of 3 groups: Emergency Medical Services (EMS), Fire and Traffic. The data are publicly available at the link https://www.kaggle.com/mchirico/montcoalert. The data have been transferred from the system of longitude and latitude to the Universal Transverse Mercator (UTM).

The clustering features of the emergency calls are clearly present in Fig. 4. The top left of Fig. 4 displays the distribution of the time by hours between pairs of events of a homogeneous Poisson process separated in space by 200 meters or less. It is approximately uniformly distributed under this situation. Meanwhile, the top right of Fig. 4 shows the distribution of the time between pairs of events from the simulated self-exciting point process data. It can be seen that the histogram shows a spike at shorter time, indicating the clustering features of the simulated data. The bottom plots of Fig. 4 present the histograms of the time between nearby emergency call pairs of all three types of emergency call events separated in space by 200 meters or less, for all recorded emergency calls. Again we observe the clustering features of emergency calls indicated by a spike during a short time period.

B. MODELING EMERGENCY CALL DATA

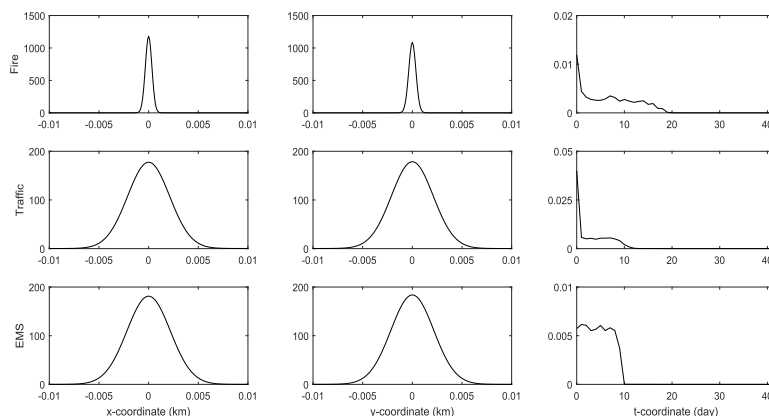Now we use the proposed approach to model the Montgomery County emergency call data. We first study the triggering

FIGURE 5. Triggering marginals: $g(x)$ (left), $g(y)$ (middle) and $g(t)$ (right) estimated using the proposed nonparametric self-exciting method for three types of emergency calls, i.e., fire, traffic and EMS respectively.
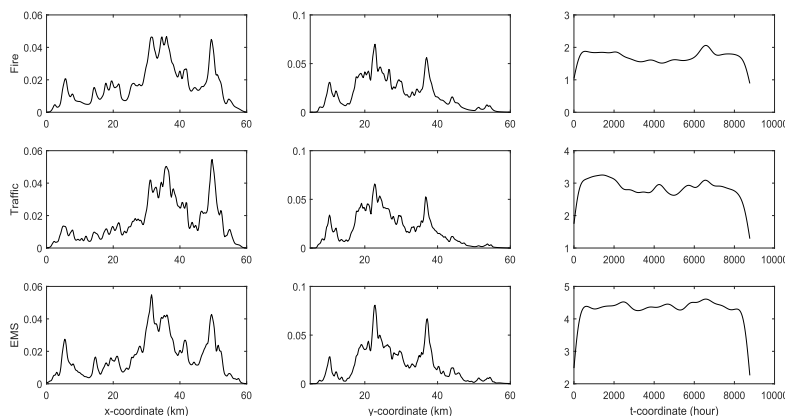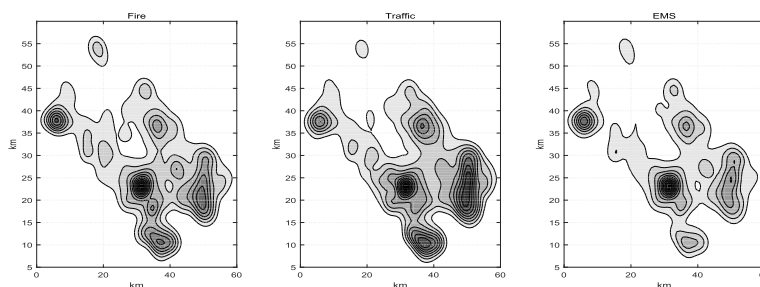

FIGURE 6. Background marginals: $g(x)$ (left), $g(y)$ (middle) and $g(t)$ (right) estimated using the proposed nonparametric self-exciting method for three types of emergency calls, i.e., fire, traffic and EMS respectively.

FIGURE 7. Background space marginal $\mu(x, y)$ estimated using the proposed nonparametric self-exciting method for three types of emergency calls, i.e., fire, traffic and EMS respectively.

intensity function, which represents the clustering features of the emergency calls. The estimated marginal densities of the triggering intensity function are shown in Fig. 5. The presence of clustering features can be seen clearly, as these estimated marginal densities of space appear to approximate Gaussian distributions with small variances. Fig. 5 shows that an emergency call may trigger other emergency calls in a very close distance and time. This is because these emergency calls may be triggered by the same emergency event and reported by nearby insiders within a time interval from a few minutes to several hours. For different types of

emergency calls, the estimated marginal densities of spatial-temporal are different. The emergency calls of fire were made by the insiders within a 30 meters × 30 meters area and a time interval from zero to several days because fire is easy to spread and may last for long time. The emergency calls of traffic were made by the insiders within a bigger area than that of fire because the traffic accidents occur within a 50 meters × 50 meters area, but the time interval is less than one hour because the traffic accidents are reported immediately by the insiders and resolved in a shorter time period comparing to fire emergencies. The emergency calls of the EMS show a
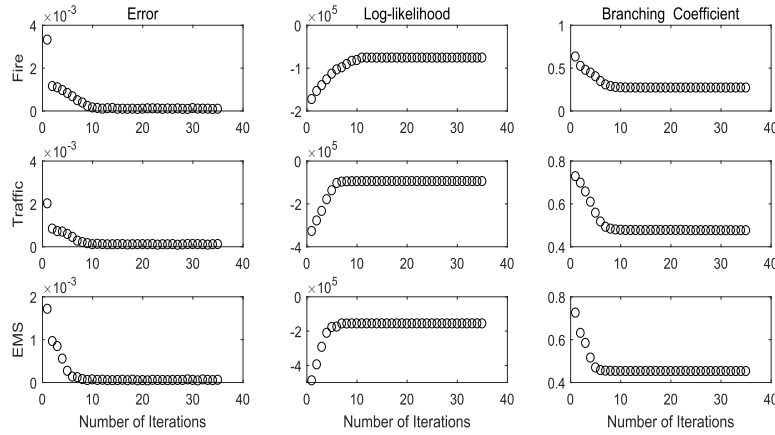
FIGURE 8. Error L_2 (left), log-likelihood (middle) and branching coefficient (right) against iterations for the varied bandwidth KDE method.
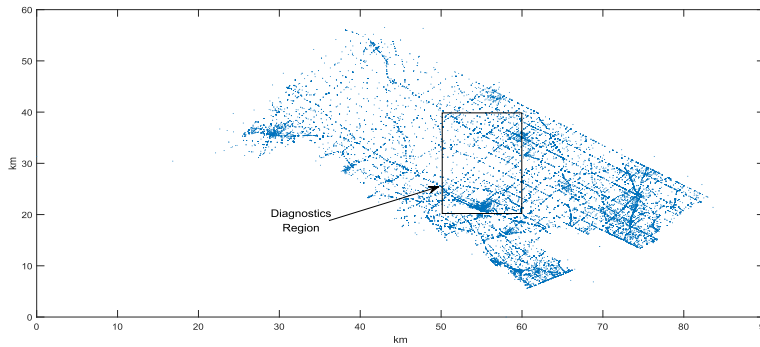

FIGURE 9. The selected examining region of size $[50; 60] \times [20; 40]$ km$\times$km.
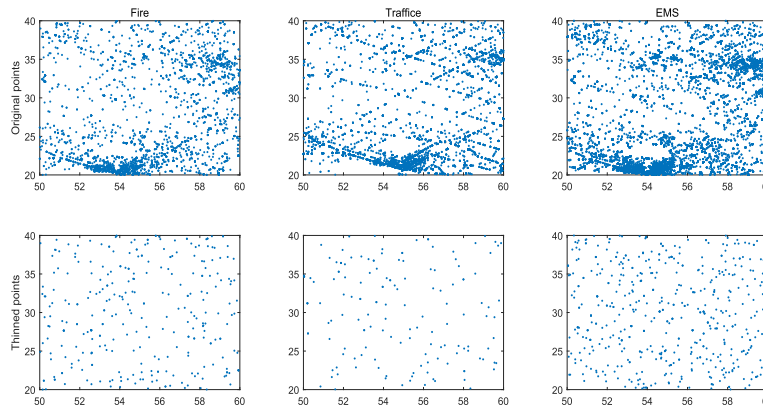

FIGURE 10. The original location of emergency calls and the generated location with the thinning process.

very different clustering pattern, i.e., these calls were made by the insiders within a uniform time interval from 0 to 10 days, which is consistent with the discovery that the crimes with high probabilities happen within 10 days after a "main" crime happened [24].

Figs. 6 and 7 display the marginal densities of the estimated background intensity function, which represent the occurrence rate of spontaneous, untriggered emergency calls. Here we observe that the estimated background intensity function does not exhibit fluctuations on a time scale of

months/days/hours, which are different from the fluctuations due to self-excitation. The background intensity function is spatially varied, and may be caused by the environmental heterogeneity in Montgomery County, such as the population density.

In Fig. 8, the errors L_2, the log-likelihood and the branching coefficient are plotted against the number of iterations. We observe that, from the left plot of Fig. 8, the errors converge quickly within the first 5 iterations and then stabilize. The middle plot of Fig. 8 shows the log-likelihood
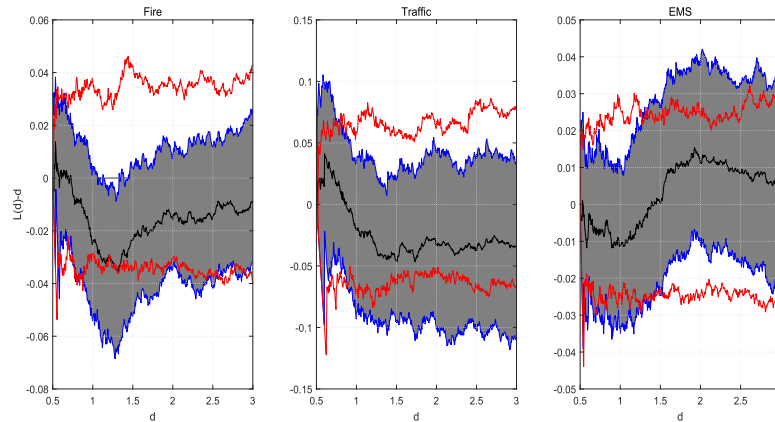
FIGURE 11. Thinning residuals. Lower red: the 5% confidence bound of the estimated values of the centered *L*−function, *L*(*d*) − *d*, of the homogeneous Poisson process; Upper red: the 95% confidence bound of the estimated values of the centered *L*−function of the homogeneous Poisson process; Lower green: the 5% confidence bound of the estimated values of the centered *L*−function of the thinning residuals; Upper green: the 95% confidence bound of the estimated values of the centered *L*−function of the thinning residuals; Middle black: the empirical mean of the estimated values of the centered *L*−function of the thinning residuals; Gray region: the confidence region of the estimated values of the centered *L*−function of the thinning residuals.

converges to some maximum value. We also see that, from the right plot of Fig. 8, the branching coefficient converges to some fixed value. All the convergence indicates that **Algorithm 1** performs well for these three types of emergency calls. We examine the Goodness-of-fit of the proposed nonparametric spatial-temporal self-exciting point process in a selected region of size $[50, 60] \times [20, 40]$. Fig. 10 shows the implementation of the thinning process. We observe that the thinning process behaves similarly to Poisson processes. Fig. 11 shows the estimated values of the centered L−function, $L(d) - d$, and the 95% confidence bounds of homogeneous Poisson processes of the selected region. Some spatial-temporal points occur at exactly the same location because of the location error. We therefore diagnose the estimated model using $L(d)$ function where d is greater than 0.5. It shows that the estimated values of the centered L−function are almost within the 95% confidence bounds. This result demonstrates that the proposed spatial-temporal self-exciting point process can well model the three types of emergency calls.

VI. CONCLUSION

We propose a nonparametric spatial-temporal self-exciting point process model for modeling emergency calls. Using the residual analysis method, we examine the performance of the proposed model. The clustering features are studied based on the triggering intensity function. Using the simulated data, we show that the proposed nonparametric structure performs flexibly and effectively for spatial-temporal self-exciting point processes. Furthermore, we show how our spatial-temporal self-exciting point process model can be used for modeling emergency calls. It can be seen that the proposed nonparametric model is able to capture the clustering features in emergency calls. The estimated background and triggering intensity functions can be used to distinguish the

areas with intrinsically high emergency calls and those with temporarily high emergency calls, as the former arises due to some specific geographical and environmental conditions and the latter may be caused by insiders' behaviors. These are also the reasons why different clustering patterns for different type of emergency calls are observed. The branching coefficients of the fire, traffic and EMS emergency calls are 26.90%, 47.55% and 45.2%, respectively. The clustering phenomena strongly present in the emergency call data in the researched region.

For future research, the branch coefficient β in Equation (9) can be extended as a function of space and time, such as the ETAS model where the branch coefficient is influenced by the magnitude. To model emergency calls, a varied branch coefficient is better because of the specific geographical and environmental conditions, such as the population density. Later, we will focus on studying a general form of branch coefficient spatial-temporal self-exciting point processes and their applications.

REFERENCES

[1] E. Bacry, K. Dayri, and J. F. Muzy, "Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data," *Eur. Phys. J. B*, vol. 85, no. 5, p. 157, 2012.

[2] E. Bacry, M. Bompaire, S. Gaïffas, and J. F. Muzy. (2015). "A generalization error bound for sparse and low-rank multivariate Hawkes processes." [Online]. Available: https://arxiv.org/abs/1501.00725

[3] E. Bacry and J. F. Muzy, "First- and second-order statistics characterization of Hawkes processes and non-parametric estimation," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2184–2202, Apr. 2016.

[4] W. Bernasco and P. Nieuwbeerta, "How do residential burglars select target areas?: *A new approach to the analysis of criminal location choice*," *Brit. J. Criminol.*, vol. 45, no. 3, pp. 296–315, 2004.

[5] A. Bray and F. P. Schoenberg, "Assessment of point process models for earthquake forecasting," *Stat. Sci.*, vol. 28, no. 4, pp. 510–520, 2013.

[6] A. Bray, K. Wong, C. D. Barr, and F. P. Schoenberg, "Voronoi residual analysis of spatial point process models with applications to California earthquake forecasts," *Ann. Appl. Statist.*, vol. 8, no. 4, pp. 2247–2267, 2014.

[7] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Secur. J.*, vol. 21, nos. 1–2, pp. 4–28, 2008.

[8] A. Clauset and R. Woodard, "Estimating the historical and future probabilities of large terrorist events," *Ann. Appl. Statist.*, vol. 7, no. 4, pp. 1838–1865, 2013.

[9] R. A. Clements, F. P. Schoenberg, and A. Veen, "Evaluation of space–time point process models using super-thinning," *Environmetrics*, vol. 23, no. 7, pp. 606–616, 2012.

[10] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*, vol. 1. New York, NY, USA: Springer-Verlag, 2003.

[11] M. Egesdal, C. Fathauer, K. Louie, and J. Neuman, "Statistical modeling of gang violence in Los Angeles," SIAM Undergraduate Res. Online, Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep., 2010, pp. 72–94. [Online]. Available: http://www.SIAM.org/students/siuro/vol3/S01045.pdf

[12] E. W. Fox, F. P. Schoenberg, and J. S. Gordon, "Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences," *Ann. Appl. Statist.*, vol. 10, no. 3, pp. 1725–1756, 2016.

[13] E. W. Fox, M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi, "Modeling e-mail networks and inferring leadership using self-exciting point processes," *J. Amer. Stat. Assoc.*, vol. 111, no. 514, pp. 564–584, 2016.

[14] W. Gorr, Y. Lee, D. Weisburd, and J. Eck, "Chronic and temporary crime hot spots," *Unraveling the Crime-Place Connection: New Directions in Theory and Policy-Advances in Criminological Theory*. New Brunswick, NJ, USA: Transaction Publishers, 2017.

[15] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, pp. 83–90, Apr. 1971.

[16] E. Lewis and G. Mohler, "A nonparametric EM algorithm for multiscale Hawkes processes," *J. Nonparametric Statist.*, vol. 1, no. 1, pp. 1–20, 2011.

[17] E. Lewis, G. Mohler, P. J. Brantingham, and A. Bertozzi, "Self-exciting point process models of insurgency in Iraq," *Secur. J.*, vol. 25, no. 3, pp. 244–264, 2010.

[18] E. A. Lewis, "Estimation techniques for self-exciting point processes with applications to criminal behavior," Ph.D. dissertation, Dept. Math., Univ. California, Los Angeles, Los Angeles, CA, USA, 2012.

[19] T. J. Liniger, "Multivariate hawkes processes," Ph.D. dissertation, ETH Zürich, Zürich, Switzerland, 2009.

[20] D. Marsan and O. Lengliné, "Extending earthquakes' reach through cascading," *Science*, vol. 319, no. 5866, pp. 1076–1079, 2008.

[21] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382. Hoboken, NJ, USA: Wiley, 2007.

[22] G. Mohler, "Modeling and estimation of multi-source clustering in crime and security data," *Ann. Appl. Statist.*, vol. 7, no. 3, pp. 1525–1539, 2013.

[23] G. O. Mohler and M. B. Short, "Geographic profiling from kinetic models of criminal behavior," *SIAM J. Appl. Math.*, vol. 72, no. 1, pp. 163–180, 2012.

[24] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *J. Amer. Stat. Assoc.*, vol. 106, no. 493, pp. 100–108, 2011.

[25] J. Moller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL, USA: CRC Press, 2003.

[26] Y. Ogata, "On Lewis' simulation method for point processes," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 23–31, Jan. 1981.

[27] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *J. Amer. Statist. Assoc.*, vol. 83, no. 401, pp. 9–27, 1988.

[28] Y. Ogata, "Space-time point-process models for earthquake occurrences," *Ann. Inst. Stat. Math.*, vol. 50, no. 2, pp. 379–402, 1998.

[29] Y. Ogata, K. Katsura, and M. Tanemura, "Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis," *J. Roy. Stat. Soc. C, Appl. Statist.*, vol. 52, no. 4, pp. 499–509, 2003.

[30] B. D. Ripley, "Modelling spatial patterns," *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 2, pp. 172–212, 1977.

[31] F. P. Schoenberg, "Multidimensional residual analysis of point process models for earthquake occurrences," *J. Amer. Stat. Assoc.*, vol. 98, no. 464, pp. 789–795, 2003.

[32] M. B. Short, G. O. Mohler, P. J. Brantingham, and G. E. Tita, "Gang rivalry dynamics via coupled point process networks," *Discrete Continuous Dyn. Syst.-B*, vol. 19, no. 5, pp. 1459–1477, 2014.

[33] G. Tita and G. Ridgeway, "The impact of gang formation on local patterns of crime," *J. Res. Crime Delinquency*, vol. 44, no. 2, pp. 208–237, 2007.

[34] A. Veen and F. P. Schoenberg, "Estimation of space–time branching process models in seismology using an EM–type algorithm," *J. Amer. Stat. Assoc.*, vol. 103, no. 482, pp. 614–624, 2008.

[35] D. Vere-Jones, "Statistical methods for the description and display of earthquake catalogues," in *Statistics in the Environmental and Earth Sciences*, A. T. Walden and P. Guttorp, Eds. London, U.K.: Edward Arnold, 1992, ch. 11., pp. 220–246.

[36] D. Vere-Jones, "Forecasting earthquakes and earthquake risk," *Int. J. Forecasting*, vol. 11, no. 4, pp. 503–538, 1995.

[37] D. Vere-Jones and R. B. Davies, "A statistical survey of earthquakes in the main seismic region of New Zealand: Part 2—Time series analyses," *New Zealand J. Geol. Geophys.*, vol. 9, no. 3, pp. 251–284, 1966.

[38] W. Wang *et al.*, "Temporal patterns of emergency calls of a metropolitan city in China," *Phys. A, Stat. Mech. Appl.*, vol. 436, pp. 846–855, Oct. 2015.

[39] H. T. Xu, M. Farajtabar, and H. Zha, "Learning granger causality for hawkes processes," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1717–1726.

[40] Y. Yang, J. Etesami, N. He, and N. Kiyavash. (2018). "Nonparametric Hawkes processes: Online estimation and generalization bounds." [Online]. Available: https://arxiv.org/abs/1801.08273

[41] K. Zhou, H. Zha, and L. Song, "Learning triggering kernels for multidimensional hawkes processes," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1301–1309.

[42] J. C. Zhuang, Y. Ogata, and D. Vere-Jones, "Stochastic declustering of space-time earthquake occurrences," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 369–380, 2002.

[43] J. C. Zhuang, Y. Ogata, and D. Vere-Jones, "Analyzing earthquake clustering features by using stochastic reconstruction," *J. Geophys. Res., Solid Earth*, vol. 109, no. B5, pp. B05301-1–B05301-17, 2004.

CHENLONG LI was born in Luohe, Henan, China, in 1989. He received the B.Sc. degree in statistics and the M.Sc. degree in probability and mathematical statistics from the School of Science, Chongqing University of Technology, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the School of Mathematics, Tianjin University, China. His current research interests include multi-dimensional random signals and random fields.

ZHANJIE SONG was born in Hebei, China, in 1965. He received the Ph.D. degree in probability and mathematical statistics from Nankai University, China, in 2006. He was a Postdoctoral Fellow in signal and information processing with the School of Electronic and Information Engineering, Tianjin University.
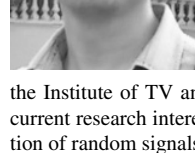
He is currently a Professor with the School of Mathematics, a Fellow with the Visual Pattern Analysis Research Lab, and a Vice-Director with the Institute of TV and Image Information, Tianjin University, China. His current research interest includes sampling, approximation, and reconstruction of random signals and random fields.

XU (SUNNY) WANG received the B.Sc. degree in applied mathematics and the M.Sc. degree in statistics from Tianjin University, China, in 1999 and 2002, respectively, and the Ph.D. degree in statistics from the University of Waterloo, Canada, in 2007.

She is currently an Associate Professor with the Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada. Her current research interests include statistical learning/data mining in drug discovery, mixture models, measurement errors in biostatistics, applied statistics to health-related research, and data mining in economic and business.

Dr. Wang is an active member of the Statistical Society of Canada (SSC) and has served in many SSC committees.

• • •