Annals of the Institute of Statistical Mathematics (2020) 72:945–967
https://doi.org/10.1007/s10463-019-00715-5

Space–time inhomogeneous background intensity estimators for semi-parametric space–time self-exciting point process models

Chenlong Li[1,2] · Zhanjie Song[1,3] · Wenjun Wang[4]

Received: 29 April 2018 / Revised: 26 February 2019 / Published online: 5 April 2019

Abstract
Histogram maximum likelihood estimators of semi-parametric space–time self-exciting point process models via expectation–maximization algorithm can be biased when the background process is inhomogeneous. We explore an alternative estimation method based on the variable bandwidth kernel density estimation (KDE) and EM algorithm. The proposed estimation method involves expanding the semi-parametric models by incorporating an inhomogeneous background process in space and time and applying the variable bandwidth KDE to estimate the background intensity function. Using an example, we show how the variable bandwidth KDE can be estimated this way. Two simulation examples based on residual analysis are designed to evaluate and validate the ability of our methods to recover the background intensity function and parametric triggering intensity function.

Keywords Space–time point process models · Kernel density estimation · Expectation–maximization algorithm · Maximum likelihood

✉ Zhanjie Song
zhanjiesong@tju.edu.cn

Chenlong Li
lichenlong@tju.edu.cn

Wenjun Wang
wjwang@tju.edu.cn

1 School of Mathematics, Tianjin University, Tianjin 300072, China

2 Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada

3 Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China

4 School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

1 Introduction

Self-exciting point process models are proposed to describe random collections of events where the occurrence of one event increases the likelihood that another event occurs shortly thereafter. An explicit form of the self-exciting point process model was formally defined by Hawkes (1971). Ogata developed a class of important self-exciting point process models in seismology, the Epidemic Type Aftershock Sequence (ETAS) models, which were considered the main tool for the space–time analysis of earthquakes (Ogata 1988, 1998). Recently, the self-exciting point process models were used to describe crime and security, social networks and financial (Mohler et al. 2011; Bacry et al. 2012; Fox et al. 2016b).

ETAS and other space–time self-exciting models often restricted parametric forms of the triggering intensity function (Ogata 1998; Bacry et al. 2012, 2015; Fox et al. 2016b) and nonparametric forms of the background intensity function (Zhuang et al. 2002; Veen and Schoenberg 2008; Mohler et al. 2011; Fox et al. 2016a). Zhuang et al. (2002) proposed the variable bandwidth kernel density estimation (KDE) method to estimate the background intensity function of the ETAS model, and the parametric triggering intensity function was estimated by using the numerical optimization procedures. Veen and Schoenberg (2008) pointed out that the numerical optimization procedures can be unstable and computationally intensive, and then proposed a robust and accurate EM-type algorithm to estimate the background and triggering intensity functions. Specifically, the histogram estimators were used to estimate the background intensity function of the homogeneous time background process, then the mean background intensity estimation within each bin can be obtained, which implied the assumption of constancy in each bin. However, the shortcomings are that if a bin does not contain any events, then the estimated value of intensity function in that bin is zero, and if a small bin contains any events, then the estimated value of intensity function in that bin is big. Veen and Schoenberg (2008) were primarily interested in exploring the stable and computationally efficient of the EM-type estimation method, and the explicit assessment or validation of the proposed histogram estimators for estimating background intensity function was not addressed. Fox et al. (2016a) systematically studied the nonparametric estimation of the ETAS model and assessed the performance of the histogram estimators of triggering intensity function. Additionally, the background process is assumed to be homogeneous in most applications of self-exciting point process models such as Marsan and Lengline (2008), Bacry et al. (2012), Bacry et al. (2015) and Yang et al. (2018). While a homogeneous background process is unrealistic in real situations. Even if the truth background process is homogeneous, missing values contained in the observed data will lead to an inhomogeneous background process (Kagan 2003).

In this paper, we propose a novel EM-type estimation of the semi-parametric model to overcome the shortcomings of the histogram estimators. Firstly, we expand the semi-parametric models by incorporating an inhomogeneous background process in space and time. Secondly, we apply the variable bandwidth KDE to estimate the background intensity function. For the parametric form of the triggering intensity function, building on work done in modeling seismic datasets, crime and security datasets and social network datasets (Zhuang et al. 2002; Mohler et al. 2011; Fox et al. 2016b),

the anisotropy Gaussian kernel and exponential kernel are considered. Viewing the estimation of semi-parametric self-exciting point process models as incomplete data problems, the EM algorithm can be used to attain the maximum likelihood estimates (MLEs). By using the variable bandwidth KDE and EM algorithm, we are able to present a highly robust and accurate estimation procedure that can be used to estimate the proposed semi-parametric self-exciting point process model. We validate and assess the ability of each method to recover the true form of the inhomogeneous background process and the triggering intensity function using simulation datasets, and the bias and standard error are calculated by repeatedly simulating and re-estimating the simulation models (24) and (25). To demonstrate the fitting performance of our proposed method, we examine the variable bandwidth KDE and histogram estimators using the super-thinning method, which is a simple and efficient model diagnostics (Clements et al. 2012), and then we compare the results of the variable bandwidth KDE to the histogram estimators.

The rest of this paper is organized as follows. Section 2 describes the concepts of the space–time self-exciting point process models. Section 3 proposes the semi-parametric models, the EM-type MLEs, the model diagnostics, and the simulation algorithm. Section 4 outlines the performance of the proposed method using the simulation datasets. Finally, Sect. 5 summarizes the results with a discussion.

2 Space–time self-exciting point process models

A space–time point process X is a random collection of points with each point falling in the observed metric space $S \times T \subseteq \mathbb{R}^2 \times \mathbb{R}$. A space–time point process is typically determined by specifying its intensity process, i.e., all finite-dimensional distributions of the space–time point process are uniquely characterized by its intensity process if it exists (Liniger 2009). In the general case the intensity processes of the random point processes have to be conditioned, not only by the time since the last event, but by any additional information concerning the past history that may affect the distribution of the remaining time.

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

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

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

The crucial problem of modeling of such point processes is to determine how the conditional intensity process depends on such past variables (Vere-Jones 1995). Typically, it is performed by specifying the special structures for the conditional intensity processes. The self-exciting point process models are a kind of important conditional intensity process models. Given the space–time dataset of events with locations $\mathbf{s}_i$ and times t_i up to time t, we have the following definition of the conditional intensity process of a space–time self-exciting point process.

Definition 1 Given the observation dataset with locations $\mathbf{s}_i$ and times t_i up to time t, a space–time self-exciting point process is a simple point process N such that N follows conditional intensity process

$$\begin{aligned}\lambda(\mathbf{s}, t|\mathscr{H}_t) &= \mu(\mathbf{s}, t) + \int_{S\times(-\infty,t)} g(\mathbf{s}-\boldsymbol{\xi}, t-u)N(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}u) \\ &:= \mu(\mathbf{s}, t) + \sum_{i:t_i<t} g(\mathbf{s}-\mathbf{s}_i, t-t_i), \end{aligned} \tag{2}$$

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

The form of Eq. (2) is closely related to the branching process, i.e., each point of a self-exciting process is either an **immigrant (background)** or a **descendant (offspring or being triggered)** (Veen and Schoenberg 2008). 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., points arrive at a higher frequency for some time. This intensity increase 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 $g(\mathbf{s}, t)$.

Given the conditional intensity process in Eq. (2) over the observation period $D{:=}[t_*, t^*)$, estimates of the parameters may be obtained by maximizing the log-likelihood function (Daley and Vere-Jones 2003):

$$\begin{aligned}\log L &= \int\int\int_{S\times D} \log\lambda(\mathbf{s}, t)N(\mathrm{d}\mathbf{s}, \mathrm{d}t) - \int\int\int_{S\times D} \lambda(\mathbf{s}, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \\ &:= \sum_{i=1}^{n} \log\lambda(\mathbf{s}_i, t_i) - \int\int\int_{S\times D} \lambda(\mathbf{s}, t)\mathrm{d}x\mathrm{d}y\mathrm{d}t. \end{aligned} \tag{3}$$

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

3 Semi-parametric models and model diagnostics

In this section, we consider semi-parametric space–time self-exciting point process models with inhomogeneous background process. Following previous research including Zhuang et al. (2002), Veen and Schoenberg (2008) and Mohler et al. (2011),

we first give the proposed semi-parametric model and then explore the estimation method based on EM algorithm. To examine the goodness-of-fit of the proposed semi-parametric models, the residual analysis with the super-thinning method proposed in Clements et al. (2012) is considered.

3.1 Semi-parametric model

Here, we consider the space–time self-exciting point process model with a background intensity function $\mu(x, y, t)$ and the parametric triggering intensity function $g(x, y, t)$. While the proposed estimation method allows for quite general parametric forms of the triggering intensity function $g(x, y, t)$, $g(x, y, t)$, in general, is chosen such that the conditional intensity process decreases in space and time away from each event. To distinguish the deduction proposed in Veen and Schoenberg (2008), we use the following parametric form of the triggering intensity function based on the studies in seismic, crime and security and social networks (Ogata 1998; Mohler et al. 2011; Fox et al. 2016b),

$$g(x, y, t) = \frac{\omega\beta}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2} - \beta t\right\}, \tag{4}$$

where $0 < \omega < 1, \beta > 0, \sigma_1 > 0, \sigma_2 > 0$. In Eq. (4), the anisotropic Gaussian kernel is considered in spatial dimension, and the exponential kernel is considered in time dimension. Parameter β controls the exponential kernel rate of decay, parameters σ_1 and σ_2 control the spatial distribution, parameter ω controls the number of triggering events.

Standard nonparametric methods for estimating the background intensity function $\mu(x, y, t)$ include spline, kernel smoothing, and Voronoi estimation (Mohler et al. 2011). As in Vere-Jones (1992), Zhuang et al. (2002) and Mohler et al. (2011), we assume that the space–time background intensity function $\mu(x, y, t)$ can be separately by

$$\mu(x, y, t) = \alpha u(x, y)v(t) \tag{5}$$

where α is a positive scaling factor controlling the overall background rate. Then, we use the variable bandwidth kernel smoothing to model $u(x, y)$ and $v(t)$ (Zhuang et al. 2002)

$$u(x, y) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_i^u}\left(x - x_i^b, y - y_i^b\right), \tag{6}$$

$$v(t) = \frac{1}{n_b} \sum_{i=1}^{n_b} K_{d_i^v}\left(t - t_i^b\right), \tag{7}$$

where $\{(x_i^b, y_i^b, t_i^b)\}_{i=1}^{n_b}$ represents background events, n_b represents the size of background events, $K_{d_i^u}(x, y)$ and $K_{d_i^v}(t)$ denote the Gaussian kernel functions

$$K_{d_i^u}(x, y) = \frac{1}{2\pi (d_i^u)^2} \exp\left\{-\frac{x^2}{2(d_i^u)^2} - \frac{y^2}{2(d_i^u)^2}\right\},$$
$$K_{d_i^v}(t) = \frac{1}{\sqrt{2\pi} d_i^v} \exp\left\{-\frac{t^2}{2(d_i^v)^2}\right\},$$

d_i^u and d_i^v represent the variable bandwidths of space and time calculated for each event i, respectively. In particular, the variable bandwidth d_i^u is computed by finding the radius of the smallest disk centered at (x_i, y_i) that contains at least n_p^u other events. Similarly, one can choose variable bandwidth d_i^v by finding the radius of the smallest disk centered at t_i that contains at least n_p^v other events. In addition, Zhuang et al. (2002) suggested setting d_i^u a threshold value ϵ if this distance is less than ϵ. This parameter becomes important when some points happen to overlap at one location caused by rounding the numbers or measurement errors (Zhuang 2011). This method makes **Algorithm 1** more stable and is used in Zhuang et al. (2002), Zhuang et al. (2004), Zhuang (2011) and Fox et al. (2016a).

Using Eqs. (4)–(7), we propose the semi-parametric models of the form

$$\lambda(x, y, t|\mathscr{H}_t) = \alpha u(x, y) v(t) + \sum_{i:t_i<t} g(x - x_i, y - y_i, t - t_i). \tag{8}$$

3.2 EM-type maximum likelihood estimates

Viewing the estimation of self-exciting point process models as incomplete data problems, we derive the EM-type MLEs of the proposed semi-parametric model (8).

Suppose now we have observed a realization of a space–time 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. Consider the random variables

$$\zeta_i = \begin{cases} i & \text{if event } i \text{ is a background event,} \\ j & \text{if event } i \text{ is triggered by event } j, i \neq j. \end{cases} \tag{9}$$

If the branching structure is incorporated, the complete data log-likelihood can be decomposed additively into likelihood functions for the background process and triggering processes, respectively:

$$\begin{aligned} \log L^c(\Theta) &= \sum_{i=1}^{n} \mathbf{1}_{\{\zeta_i = i\}} \log(\mu(\mathbf{s}_i, t_i)) + \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}_{\{\zeta_i = j\}} \log(g(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)) \\ &\quad - \int\int\int_{S\times D} \lambda(\mathbf{s}, t) \mathrm{d}x\mathrm{d}y\mathrm{d}t \\ &= \sum_{i=1}^{n} \mathbf{1}_{\{\zeta_i = i\}} \log(\mu(x_i, y_i, t_i)) - \int\int\int_{S\times D} \mu(x, y, t) \mathrm{d}x\mathrm{d}y\mathrm{d}t \end{aligned}$$

$$+\sum_{j=1}^{n}\left[\sum_{i:t_j<t_i}\mathbf{1}_{\{\zeta_i=j\}}\log(g(x_i-x_j,y_i-y_j,t_i-t_j))\right.$$
$$\left.-\int\int\int_{S\times[t_j,t^*)}g(x-x_j,y-y_j,t-t_j)\mathrm{d}x\mathrm{d}y\mathrm{d}t\right], \tag{10}$$

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

E-Step Calculating

$$p_{ij}^{(k)}:=E(\mathbf{1}_{\{\zeta_i=j\}})=\begin{cases}\frac{\mu^{(k)}(x_i,y_i,t_i)}{\lambda^{(k)}(x_i,y_i,t_i)} & j=i,\\ \frac{g^{(k)}(x_i-x_j,y_i-y_j,t_i-t_j)}{\lambda^{(k)}(x_i,y_i,t_i)} & t_i>t_j,\ j>0,\\ 0 & t_i<t_j,\ j>0.\end{cases} \tag{11}$$

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

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

The probabilities $p_{ij}^{(k)}$ allow estimating $u^{(k+1)}(x,y)$ and $v^{(k+1)}(t)$ by kernel smoothing as follows:

$$u^{(k+1)}(x,y)=\frac{1}{\sum_{i=1}^{n}p_{ii}^{(k)}}\sum_{i=1}^{n}p_{ii}^{(k)}K_{d_i^u}(x-x_i,y-y_i) \tag{13}$$

$$v^{(k+1)}(t)=\frac{1}{\sum_{i=1}^{n}p_{ii}^{(k)}}\sum_{i=1}^{n}p_{ii}^{(k)}K_{d_i^v}(t-t_i). \tag{14}$$

The last integral term of Eq. (12) can be approximated by ω using the fact that the space–time distances between triggering events and their parent events are usually much smaller than the study region $S\times D$. This approximation was also considered in Veen and Schoenberg (2008). Then we can calculate the rest parameters by making the partial derivatives with respect to the parameters of $g(x,y,t)$ be equal to zero, that is,

$$\alpha^{(k+1)} = \frac{\sum_{i=1}^{n} p_{ii}^{(k)}}{\int\int\int_{S\times D} u^{(k+1)}(x,y)v^{(k+1)}(t)\mathrm{d}x\mathrm{d}y\mathrm{d}t}, \tag{15}$$

$$\sigma_1^{(k+1)} = \sqrt{\frac{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}(x_i - x_j)^2}{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}}}, \tag{16}$$

$$\sigma_2^{(k+1)} = \sqrt{\frac{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}(y_i - y_j)^2}{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}}}, \tag{17}$$

$$\omega^{(k+1)} \approx \frac{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}}{n}, \tag{18}$$

$$\beta^{(k+1)} \approx \frac{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}}{\sum_{j=1}^{n}\sum_{i=j+1}^{n} p_{ij}^{(k)}(t_i - t_j)}. \tag{19}$$

For a similar reason, the integral term of Eq. (15) can be approximated by one. Then, we have

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

Furthermore, one has $\lambda^{(k+1)}(x,y,t) = \mu^{(k+1)}(x,y,t) + \sum_{i:t_i<t} g^{(k+1)}(x-x_i, y-y_i, t-t_i)$ and $\mu^{(k+1)}(x,y,t) = \alpha^{(k+1)}u^{(k+1)}(x,y)v^{(k+1)}(t)$.

The following summarizes our proposed EM-type algorithm to estimate the proposed semi-parametric model:

Algorithm 1: Estimation Algorithm.

Step 1. Set $k = 0$ and initialize $P^{(0)} := (p_{ij}^{(0)})$ by

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

Step 2. Estimate background intensity function $u^{(k+1)}(x,y)$, $v^{(k+1)}(t)$ and $\alpha^{(k+1)}$ by using (13), (14) and (20).

Step 3. Estimate parameters of triggering intensity function by using (16)–(19).

Step 4. Update probabilities $p_{ij}^{(k+1)}$ by using (11).

Step 5. If $L_2 := \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(p_{ij}^{(k+1)} - p_{ij}^{(k)})^2} < \varepsilon$, then the algorithm has converged (in practice we take $\varepsilon = 1 \times 10^{-4}$). Otherwise, repeat steps 2–5 until

convergence or $k > B$ where B is the upper limit of iteration number (in practice we take $B = 200$).

Note: (1) The approximations in (18)–(20) can easy the computation. We show that one can obtain a robust and accurate estimation of background and triggering intensity functions with these approximations in the simulation. In addition, in **Algorithm 1**, $P^{(0)}$ is one possible initialization, one can take others, e.g., Fox et al. (2016a).

(2) The background intensity function can be more complex in real situations. For example, Mohler (2013) explored a 1-D discrete time Hawkes–Cox process model of crime and security. In this model, the background intensity function was determined by a log Gaussian Cox process (LGCP). This idea can be extended to our proposed semi-parametric models (8), and the proposed **Algorithm 1** can be used to estimate the background intensity function which is a realization of an LGCP. We verify the performance of the histogram estimators and the variable bandwidth KDE for this model in Sect. 4.2.

(3) Histogram maximum likelihood estimator proposed in Veen and Schoenberg (2008) modeled the background intensity function by subdividing the space–time observation window $S \times D$ into m cells each with constant intensity $\mu_k, k \in \{1, \ldots, m\}$. To simplify notations, one can define the expected number of background earthquakes in cell k, denoted as v_k (Veen and Schoenberg 2008):

$$v_k = \mu_k \times (\text{area of cell } k) \times (\text{length of time window}).$$

3.3 Bandwidth selection

The cross-validation-type technique has been proposed for selecting bandwidths, e.g., Vere-Jones (1992) and Adelfio and Chiodi (2015). In Vere-Jones (1992), the author proposed two scoring methods, the K–L (Kullback–Leibler) score and the MISE (mean integrated square error) score, to determine the optimal values of the bandwidths (i.e., the value giving the maximum score). In Adelfio and Chiodi (2015), the authors proposed the FLP (Forward Likelihood Predictive) method for estimating the optimal values of the bandwidths. The FLP method is a special case of the K–L score. In this paper, we use the FLP method to determine the optimal values of the bandwidths.

To implement the FLP method, the data are first divided into two parts in time, a training period, $\{(x_i, y_i, t_i)\}_{i=1}^{n_{\text{tra}}}$, which is used to produce an intensity estimate, and a forecast forward period, $\{(x_i, y_i, t_i)\}_{i=n_{\text{tra}}+1}^{n}$. Then the forecast is scored by using the log-likelihood of next point, i.e.,

$$CV(n_p^u, n_p^v, \epsilon_1, \epsilon_2) = \sum_{i=n_{\text{tra}}}^{n-1} \Big[\log \hat{\lambda}_{t_i}(x_{i+1}, y_{i+1}, t_{i+1}) - \int\int_S \int_{t_i}^{t_{i+1}} \hat{\lambda}_{t_i}(x, y, t) \mathrm{d}t \mathrm{d}x \mathrm{d}y \Big], \tag{21}$$

where $\hat{\lambda}_{t_i}(x, y, t)$ is the proposed intensity estimate by using the data up to t_i, ϵ_1 and ϵ_2 are the threshold values for the bandwidths d_p^u and d_p^v, respectively. The resulting

scores can then be compared for different bandwidths, and the optimal bandwidths are chosen for forecast of next point (Vere-Jones 1992; Adelfio and Chiodi 2015).

Note that

$$\begin{aligned}
&\int\int\int_{S\times[t_i,t_{i+1})} \hat{\lambda}_{t_i}(x,y,t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \\
&\quad= \int\int\int_{S\times[t_i,t_{i+1})} \hat{\mu}_{t_i}(x,y,t)\mathrm{d}x\mathrm{d}y\mathrm{d}t \\
&\qquad+ \hat{\omega}_{t_i}\sum_{j=1}^{i}\int\int\int_{S\times[t_i,t_{i+1})} \hat{g}_{t_i}(x-x_j,y-y_j,t-t_j)\mathrm{d}x\mathrm{d}y\mathrm{d}t \\
&\quad\approx \sum_{j=1}^{i}\Bigg[\hat{p}_{jj}\int_{[t_i,t_{i+1})} K_{d_j^v}(t-t_j)\mathrm{d}t \\
&\qquad+ \hat{\omega}_{t_i}(\exp\{-\hat{\beta}_{t_i}(t_i-t_j)\}-\exp\{-\hat{\beta}_{t_i}(t_{i+1}-t_j)\})\Bigg],
\end{aligned} \tag{22}$$

where $\hat{\mu}_{t_i}(x,y,t)$, $\hat{g}_{t_i}$, $\hat{\omega}_{t_i}$ and $\hat{\beta}_{t_i}$ represent the estimated values by using the data up to t_i. We can then use approximate (22) to improve computational efficiency of Equation (21).

Zhuang et al. (2002) suggested taking n_p^u from 15 to 100 for estimating background rate of self-exciting process for modeling earthquakes. Mohler et al. (2011) used the same parameter interval for estimating the space background rate of self-exciting model of crimes and suggested taking n_p^v from 15 to 100. Fox et al. (2016a) used the fixed $\epsilon = 0.02$ and chosen $n_p^u = 50$ by minimizing the root-mean-square error (RMSE) of the estimates over 200 simulated realizations of a specified ETAS model for $n_p^u \in \{10, 15, \ldots, 95, 100\}$. Based on Zhuang et al. (2002), Mohler et al. (2011) and Fox et al. (2016a), we set $n_p^u, n_p^v \in \{5, 10, \ldots, 100\}$ and $\epsilon_1, \epsilon_2 \in \{0.005, 0.01, \ldots, 0.05\}$.

Note: (4) The above method is also used for choosing the optimal bin width of the histogram estimators. We choose the optimal numbers of space bin over each edge of S, named n_{S1} and n_{S2}, and the optimal number of time bin over D, named n_D, instead of choosing the width of the histogram. And we set $n_{S1}, n_{S2}, n_D \in \{5, 6, \ldots, 20\}$. Furthermore, there are two cases, both of which lead to $\hat{\lambda}_{t_i}(x_{i+1}, y_{i+1}, t_{i+1}) = 0$. Firstly, if point $(x_{i+1}, y_{i+1}, t_{i+1})$ lies in the kth space bin and there is no other points in this bin, we set $\hat{\lambda}_{t_i}(x_{i+1}, y_{i+1}, t_{i+1}) = 10^{-15}$. Secondly, if the first case does not happen and points (x_i, y_i, t_i) and $(x_{i+1}, y_{i+1}, t_{i+1})$ lie in the kth and the $k+1$th time bins, respectively, we let the kth time bin, $(t^{k-1}, t^k]$, be $(t^{k-1}, t_{i+1}]$. In fact, in order to improve the estimation accuracy of the kth time bin, we always use $(t^{k-1}, t_{i+1}]$ as the kth time bin.

3.4 Model diagnostics

There are many model diagnostics for testing self-exciting point process models (Schoenberg 2003; Ogata et al. 2003; Bray and Schoenberg 2013; Bray et al. 2014).

The residual analysis with thinning method proposed in Schoenberg (2003) is a simple and efficient model diagnosis method. The thinning method used the property that any process characterized by its conditional intensity process may be thinned to obtain a homogeneous Poisson process (Ogata 1981). The following procedure is a standard thinning algorithm comes from Schoenberg (2003):

(1) Define $b = \inf_{(x_i, y_i, t_i)} \widehat{\lambda}(x_i, y_i, t_i)$;
(2) For each event i in the observed process, calculate the quantity $p_i = \frac{b}{\widehat{\lambda}(x_i, y_i, t_i)}$;
(3) Retain event i with probability p_i.

Clements et al. (2012) considered using "super-thinning" method to ensure there are enough points remained when b is small. Choosing a rate κ such that $b < \kappa < \sup_{x,y,t} \lambda(x, y, t)$, and thinning with probabilities $p_i = \min\{\frac{\kappa}{\widehat{\lambda}(x_i, y_i, t_i)}, 1\}$. Then by adding a simulated inhomogeneous Poisson process with intensity $\max\{\kappa - \widehat{\lambda}(x, y, t), 0\}$ to the thinned process. The sum process will be a homogeneous process with rate κ when the estimated intensity $\widehat{\lambda}(x, y, t)$ is correct. The parameter κ is actually a tuning parameter that needs to be chosen in a way that optimizes the power of formal tests of homogeneity of the residuals (Clements et al. 2012). We follow the example of Clements et al. (2012) and choose

$$\kappa = \frac{1}{|S||D|} \int \int \int_{S \times D} \hat{\lambda}(\mathbf{s}, t) \mathrm{d}\mathbf{s}\mathrm{d}t,$$

where $|S|$ is the area of the space region S and $|D|$ is the length of the time interval D.

For the thinned process, one can use Ripley's K-function (Ripley 1977), which calculates the proportion of events per unit area which are within a given distance, to make model diagnostics. This will detects if the thinned process still has clustering not account by the model (Schoenberg 2003; Clements et al. 2012). The most commonly used K-function with edge-corrected estimator is given as follows (Ripley 1977):

$$\widehat{K}(d) = |S| n^{-2} \sum_i \sum_{j \neq i} w(\mathbf{s}_i, \mathbf{s}_j)^{-1} \mathbf{1}_{\{d_{ij} < d\}}, \tag{23}$$

where $|S|$ is the area 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 that circle that falls in the study area. Alternatively, the L-function, estimated by $L(d) = \sqrt{\frac{\widehat{K}(d)}{\pi}}$, has more stable variance than Ripley's K-function. Thus, we use L-function to make model diagnostics. Note that it takes a lot of time to calculate L-function for large κ. Therefore, we further thinned the sum process with probability $\frac{\tilde{\kappa}}{\kappa}$. The following algorithm gives the diagnostics procedure based on Monte Carlo method.

Algorithm 2: Diagnostics Algorithm.

Begin with the parameter κ, the estimated intensity $\widehat{\lambda}(\mathbf{s}, t)$, the number of simulation M and the discrete d.

Step 1. Using "super-thinning" method to obtain the sum process, and thinned the sum process with probability $\frac{\tilde{\kappa}}{\kappa}$.

Step 2. Calculate the L-functions.
Step 3. Repeat Step 1 and 2 M times.
Step 4. Calculate the empirical mean of L-functions.
Step 5. Simulate homogeneous Poisson process with intensity $\tilde{\kappa}$ over the observed region S and repeat Step 2 M times.
Step 6. Calculate the empirical 95% bounds for L-functions corresponding to M realizations of the homogeneous Poisson process.

3.5 Simulation method

Here, we introduce the simple and efficient simulation method proposed by Zhuang et al. (2004). The original simulation method is modified to generate space–time points over region $S \times D$. For a thorough introduction to simulation methods and the rationality of the thinning method, see Ogata (1981), Moller and Waagepetersen (2003) and Daley and Vere-Jones (2003).

Algorithm 3: Simulation Algorithm.
Begin with estimated intensity processes $\widehat{\mu}(x, y, t)$ and $\widehat{g}(x, y, t)$, and calculate $\widehat{m} = \int_{\mathbb{R}^2} \int_{\mathbb{R}} \widehat{g}(x, y, t)\mathrm{d}t\mathrm{d}x\mathrm{d}y$.

Step 1. Generate events from the background process with the intensity function $\widehat{\mu}(x, y, t)$ over region $S \times D$ by using a thinning method for nonhomogeneous space–time Poisson point processes (e.g., Ogata 1998). Call this catalog of events $G^{(0)}$.
Step 2. Let $l = 0$.
Step 3. For each event i in $G^{(l)}$, simulate its $N^{(i)}$ offspring where $N^{(i)} \sim$ Poisson$(\widehat{m})$, and the offspring's location and time are generated from the triggering intensity function $\widehat{g}$, normalized as a probability density. Call these offspring $O_i^{(l)}$.
Step 4. Let $G^{(l+1)} = \bigcup_{i \in G^{(l)}} O_i^{(l)}$.
Step 5. If $G^{(l)}$ is not empty, set $l = l + 1$ and return to Step 3. Otherwise, return $G = \bigcup_{j=0}^{l} G^{(j)}$.
Step 6. Disregard the event in G falling in $(\mathbb{R}^2 \times \mathbb{R})\backslash(S \times D)$, and keep the rest as the final set of simulated events.

4 Numerical experiments

4.1 Experiment 1

4.1.1 Artificial data 1

In this experiment, the artificial datasets generated from conditional intensity process (24) with data size around 1×10^4 are used to verify the performance of the proposed method. The artificial conditional intensity process (24) is modified from Mohler et al. (2011) to obtain the space–time self-exciting point process with inhomogeneous background process.

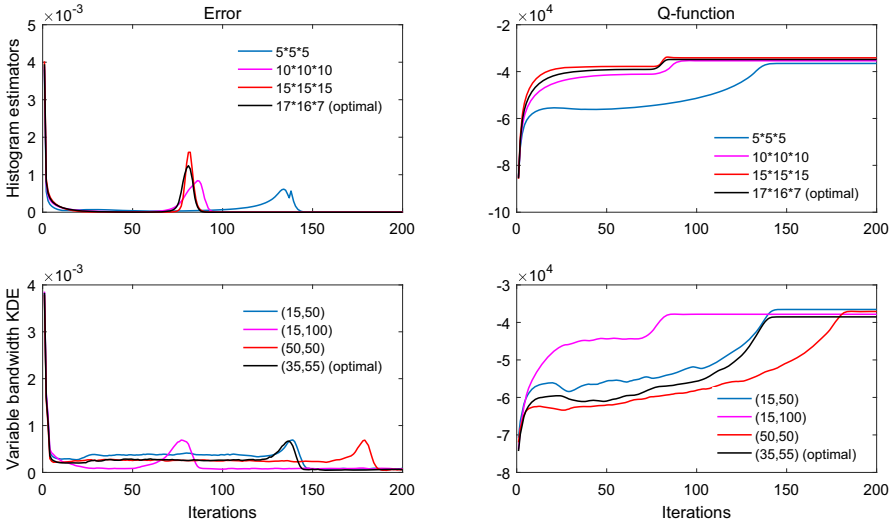

Fig. 1 Experiment 1: error L_2 (top left) and Q-function (top right) at the $k+1$th iteration for histogram estimators. Error L_2 (bottom left) and Q-function (bottom right) at the $k+1$th iteration for variable bandwidth KDE

$$\begin{cases} \mu(x,y,t) = \frac{\overline{\mu}(\cos(t/120)+2)}{2\pi\sigma_\mu^2} \exp\left(-\frac{(x-c)^2}{2\sigma_\mu^2} - \frac{(y-d)^2}{2\sigma_\mu^2}\right), \\ g(x,y,t) = \frac{\theta\omega}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2} - \omega t\right), \end{cases} \tag{24}$$

where $\overline{\mu} = 5.71$, $\sigma_\mu = 4.5$, $c = 10$, $d = 10$, $\theta = 0.2$, $\omega = 0.1$, $\sigma_1 = 0.01$ and $\sigma_2 = 0.1$. The simulation is carried out using **Algorithm 3** in an 20 by 20 region of space. In order to have a realization of the point process at steady state, the first and last 2000 points are disregarded in each simulation.

4.1.2 Fitting artificial data 1

We fit the artificial datasets generated from (24) to the conditional intensity (8). The variable bandwidth KDEs depend on the smoothing parameters n_p^u, n_p^v and threshold values ϵ_1, ϵ_2. Here we choose $n_p^u = 35$, $n_p^v = 55$, $\epsilon_1 = 0.02$, and $\epsilon_2 = 0.04$, since these values give the largest average CV (21) of ten artificial datasets for $n_p^u, n_p^v \in \{5, 10, \ldots, 100\}$ and $\epsilon_1, \epsilon_2 \in \{0.005, 0.01, \ldots, 0.05\}$. For efficiently and effectively calculating $CV(n_p^u, n_p^v, \epsilon_1, \epsilon_2)$, we set $n_{\text{tra}} = [\frac{19n}{20}] \approx 500$. The results of $(n_p^u, n_p^v) \in \{(15, 50), (15, 100), (50, 50)\}$ are also shown for comparison. Meanwhile, for the histogram estimators, we choose $n_{S1} = 17$, $n_{S2} = 16$, $n_D = 7$ based on the average CV (21) of ten artificial datasets for $n_{S1}, n_{S2}, n_D \in \{5, 6, \ldots, 20\}$. We also show other three cases where the space–time observation window $S \times D$ are subdivided into $5 \times 5 \times 5$, $10 \times 10 \times 10$, $15 \times 15 \times 15$ cells, respectively.

In Fig. 1, we plot the errors L_2 (see **Algorithm 1**) and the Q-function at the $k+1$th iteration. Here, we observe that, from the left of Fig. 1, the errors converge quickly for

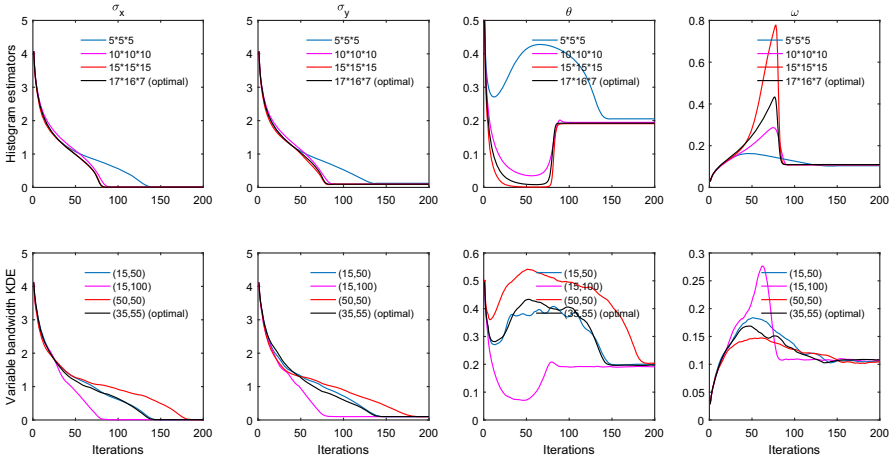

Fig. 2 Experiment 1: estimated parameter values of $\sigma_x, \sigma_y, \theta$ and ω (top) at the $k+1$th iteration for histogram estimators. Estimated parameter values of $\sigma_x, \sigma_y, \theta$ and ω (bottom) at the $k+1$th iteration for variable bandwidth KDE method

the first 10 iterations and then a fluctuation appears before the errors stabilize. From the right of Fig. 1, it can be seen that the Q-functions converge to some maximums. Both histogram estimators and variable bandwidth KDE have similar results except that the fluctuations appeared in the errors of variable bandwidth KDE are smaller than that of histogram estimators.

In Fig. 2, we plot the estimated parameter values of triggering intensity function at the $k+1$th iteration. From Fig. 2, we can see that the fluctuations appeared in the estimated parameter values of variable bandwidth KDE are smaller than that of histogram estimators. We also observe that the estimated σ_x converges to the true value without fluctuation. Furthermore, we observe that, from Figs. 1 and 2, the fluctuations of the errors and estimated θ and ω emerge when the estimated σ_x and σ_y are about to get to the true values, respectively. In Table 1, we list the exact parameter values, the estimates averaged over 100 times fitting results with 100 datasets generated from (24) and the standard errors calculated from the above fitting results. Here we observe that both the estimated parameter values are close to the exact parameter values.

In Fig. 3, we plot the estimated space marginals. It can be seen, from the top right of Fig. 3, that the space background intensity is averaged within each bin for histogram estimators. We observe that the estimated space marginal $u^{(200)}(t)$ for variable bandwidth KDE is consistent with the true space marginal. In Fig. 4, we plot the estimated time marginals. From the top of Fig. 4, we can see that the estimated time background intensity is piecewise constant for histogram estimators. From the bottom of Fig. 4, we observe that there is a boundary issue in the estimated space marginal $\widehat{\alpha} v^{(200)}(t)$ for variable bandwidth KDE.

4.1.3 Model diagnostics

Now we compare the model diagnostics of histogram estimators and variable bandwidth KDE method using residual introduced in Sect. 3.2. In this paper, we set

Table 1 Experiment 1: Estimates of parameters $\sigma_x, \sigma_y, \theta, \omega$

Tuning parameter	Parameters	$\widehat{\sigma}_x$	$\widehat{\sigma}_y$	$\widehat{\theta}$	$\widehat{\omega}$
$5 \times 5 \times 5$	Mean	.0104	.1032	.2033	.1033
	Standard errors ($\times 10^{-3}$)	.1293	1.1153	1.5234	1.6128
	Bias in % of true value	4.2842%	3.2060%	1.6421%	3.2934%
$10 \times 10 \times 10$	Mean	.0102	.0950	.2014	.1057
	Standard errors ($\times 10^{-3}$)	.1036	1.0387	1.8522	1.4995
	Bias in % of true value	2.0692%	−5.0004%	.7309%	5.6903%
$15 \times 15 \times 15$	Mean	.0097	.0994	.1856	.1088
	Standard errors ($\times 10^{-3}$)	.1193	1.2388	1.7008	1.5152
	Bias in % of true value	−2.6319%	.6203%	−7.1734%	8.8194%
$17 \times 16 \times 7$ (optimal)	Mean	.0099	.0970	.1924	.1095
	Standard errors ($\times 10^{-3}$)	.1040	1.2291	1.8836	1.7298
	Bias in % of true value	−.7115%	−2.9660%	−3.7913%	9.5335%
	Variable bandwidth KDE				
(15, 50)	Mean	.0098	.1009	.1980	.1017
	Standard errors ($\times 10^{-3}$)	.1307	1.1943	1.6900	1.6342
	Bias in % of true value	−1.7617%	.8657%	−1.0197%	1.6863%
(15, 100)	Mean	.0095	.1007	.1973	.1050
	Standard errors ($\times 10^{-3}$)	.1093	1.2538	1.7782	1.3919
	Bias in % of true value	−5.0109%	.7321%	−1.3451%	4.9627%
(50, 50)	Mean	.0098	.0975	.1947	.0995
	Standard errors ($\times 10^{-3}$)	.1206	1.5928	1.5910	1.2666
	Bias in % of true value	−2.0224%	−2.4609%	−2.6405%	−.5022%
(35, 55) (optimal)	Mean	.0100	.0994	.1983	.1066
	Standard errors ($\times 10^{-3}$)	.1375	1.3636	1.9092	1.2819
	Bias in % of true value	−.4038%	−.5755%	.8434%	6.5963%

$\tilde{\kappa} = 1500$. In Fig. 5, we plot the estimated centered L-function, $L(d) - d$, and 95% confidence bounds of homogeneous Poisson processes. For a homogeneous Poisson process, $L(d) - d = 0$, so departures from zero indicates inhomogeneity. It can be seen that the thinned residuals for variable bandwidth KDE method are evidently homogeneous, as seen in the bottom of Fig. 5 by the estimated centered L-functions, which are entirely within the 95% bounds. On the contrary, we observe that the thinned residuals for histogram estimators are inhomogeneous for $5 \times 5 \times 5$ cells, as seen in the top left of Fig. 5 by the estimated centered L-function, which is entirely outside the 95% bounds. In addition, from Fig. 5, we can observe that the estimated centered L-functions departure from zero for the other results of histogram estimators. The shortcomings of histogram estimators introduced in Sect. 1 lead to the inaccurate estimates of background intensity function. Furthermore, we study other three cases where $\tilde{\kappa} = 1000$, 2000 and 3000 and obtain similar results.

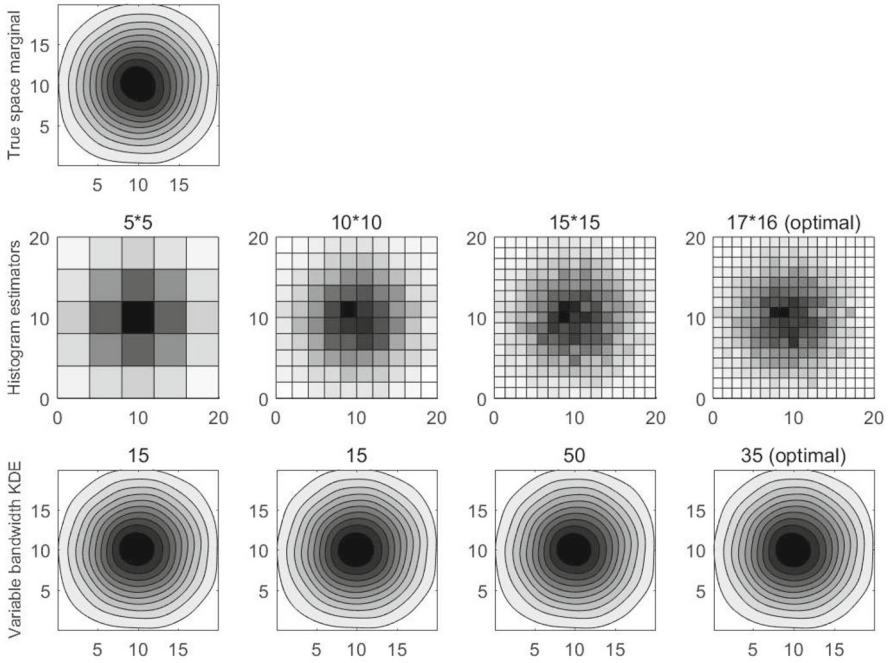

Fig. 3 Experiment 1: actual space marginals (top left). Estimated space marginals (top right) for histogram estimators. Estimated space marginals (bottom right) for variable bandwidth KDE

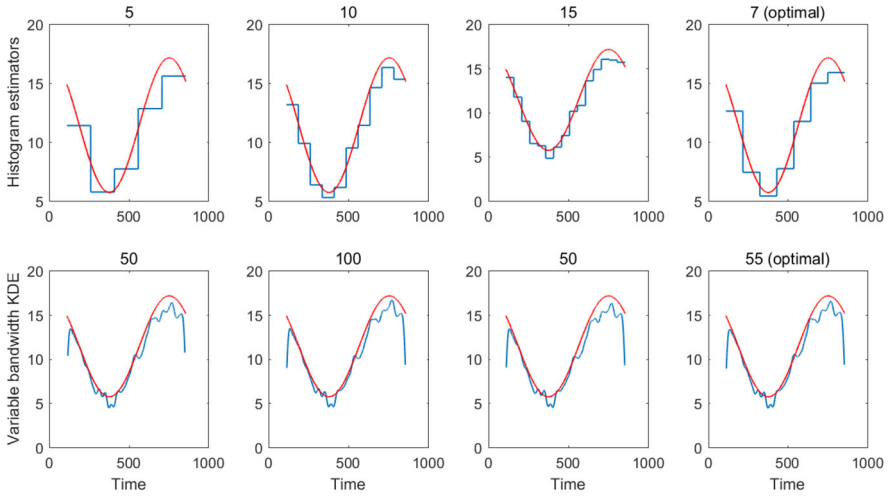

Fig. 4 Experiment 1: estimated (solid blue) and actual (solid red) time marginals (top) for histogram estimators. Estimated (solid blue) and actual (solid red) time marginals (bottom) for variable bandwidth KDE method

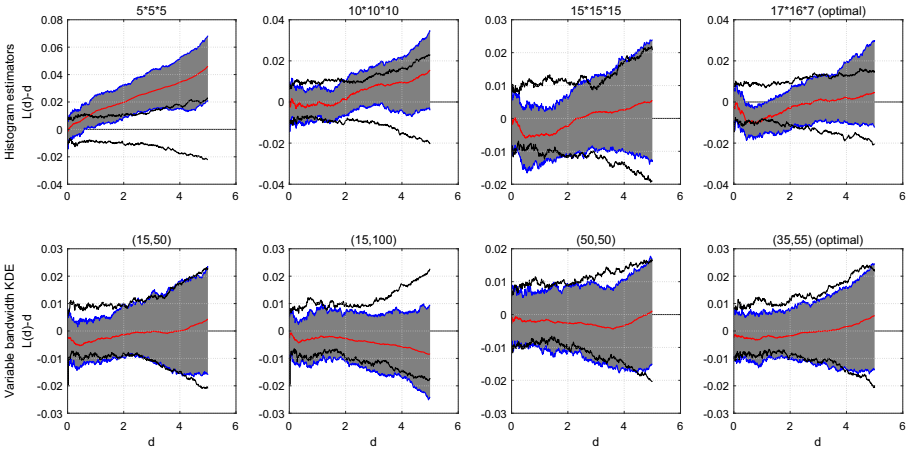

Fig. 5 Experiment 1: thinning residuals for histogram estimators (top) and variable bandwidth KDE method (bottom). Lower red: 5% bounds of the estimated centered L-function, $L(d) - d$, of homogeneous Poisson process; Upper red: 95% bounds of the estimated centered L-function of homogeneous Poisson process; Lower green: 5% bounds of the estimated centered L-function of the thinning residuals; Upper green: 95% bounds of the estimated centered L-function of the thinning residuals; Middle black: the empirical mean of the estimated centered L-function of the thinning residuals; Gray region: confidence region of the estimated centered L-function of the thinning residuals

4.2 Experiment 2

4.2.1 Artificial data 2

Based on Brix and Diggle (2001), Diggle (2006) and Diggle et al. (2013), we propose the following conditional intensity process (25). The artificial datasets generated from (25) with data size around 1×10^4 are used to verify the performance of the proposed method.

$$\begin{cases} \mu(x, y, t) = \frac{\overline{\mu}(\cos(t/120)+2)}{2\pi\sigma_\mu^2} \exp(G(\mathbf{s})), \\ g(x, y, t) = \frac{\theta\omega}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \omega t\right), \end{cases} \tag{25}$$

where $\mathbf{s} = (x, y)$, $G(\mathbf{s})$ is a spatial Gaussian process with expectation $E[G(\mathbf{s})] = -\frac{(x-c)^2}{2\sigma_\mu^2} - \frac{(y-d)^2}{2\sigma_\mu^2}$, variance $\mathrm{Var}\{G(\mathbf{s})\} = 0.5$ and exponential correlation function, $\rho(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Var}\{G(\mathbf{s})\} \exp(-||\mathbf{s}_1 - \mathbf{s}_2||_2/12)$, and the other parameters are the same as the parameters used in (24). The simulation are carried out using **Algorithm 3** in the $[0, 20] \times [0, 20] \times (0, 2500]$ space–time region. The first and last 2000 points are disregarded in each simulation.

4.2.2 Fitting artificial data 2

As the experiment 1, we choose $n_p^u = 20$, $n_p^v = 90$, $\epsilon_1 = 0.015$, and $\epsilon_2 = 0.045$ for the variable bandwidth KDE and $n_{S1} = 17$, $n_{S2} = 17$, $n_D = 11$ for the histogram

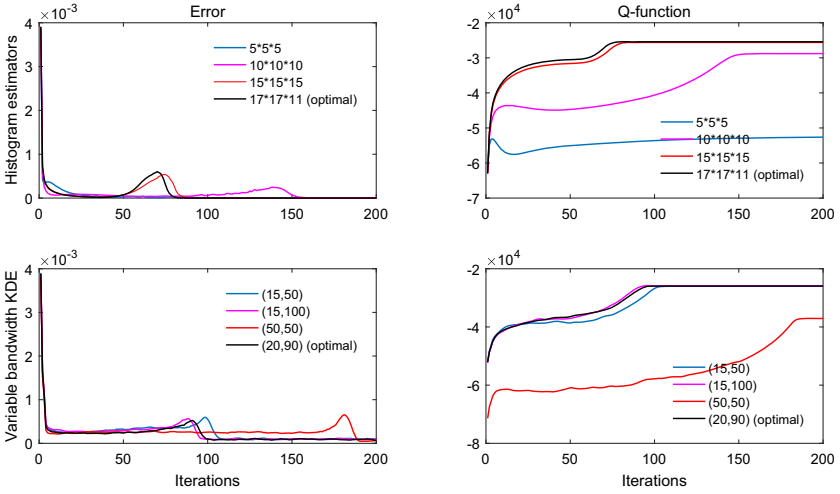

Fig. 6 Experiment 2: error L_2 (top left) and Q-function (top right) at the $k+1$th iteration for histogram estimators. Error L_2 (bottom left) and Q-function (bottom right) at the $k+1$th iteration for variable bandwidth KDE

estimators. The results of $(n_p^u, n_p^v) \in \{(15, 50), (15, 100), (50, 50)\}$ for variable bandwidth KDE and $5 \times 5 \times 5$, $10 \times 10 \times 10$, $15 \times 15 \times 15$ cells for histogram estimators are also shown for comparison.

Figure 6 shows the errors L_2 and the Q-function at the $k+1$th iteration. It can be seen that, from the left of Fig. 6, the errors also converge quickly for the first 10 iterations and then a fluctuation appears before the errors stabilize. Meanwhile the Q-functions converge to a maximum. These results are similar to the results of Fig. 1. Figure 7 shows the estimated parameter values of triggering intensity function at the $k+1$th iteration. Here we obtain the similar results except when the number of cells m is small. That is, the estimated parameters θ, ω, σ_x and σ_y departure from the true values in the cases of $5 \times 5 \times 5$ cells and $10 \times 10 \times 10$ cells. Table 2 lists the exact parameter values, the estimates averaged over 100 times fitting results with 100 datasets generated from (25) and the standard errors calculated from the above fitting results. We eliminate the results of histogram estimators for the case of $5 \times 5 \times 5$ cells because of the poor performance. From Table 2, it can be seen that the estimated parameter values are close to the exact parameter values. Figures 8 and 9 show the estimated space and time marginals, respectively. We observe that the results are similar to the results shown in Figs. 3 and 4 of experiment 1.

4.2.3 Model diagnostics

The model diagnostics of histogram estimators and variable bandwidth KDE using residual are shown in Fig. 10. We observe that the thinned residuals for variable bandwidth KDE are evidently homogeneous as same as the results shown in Fig. 5. On the contrary, we observe that the thinned residuals for histogram estimators are

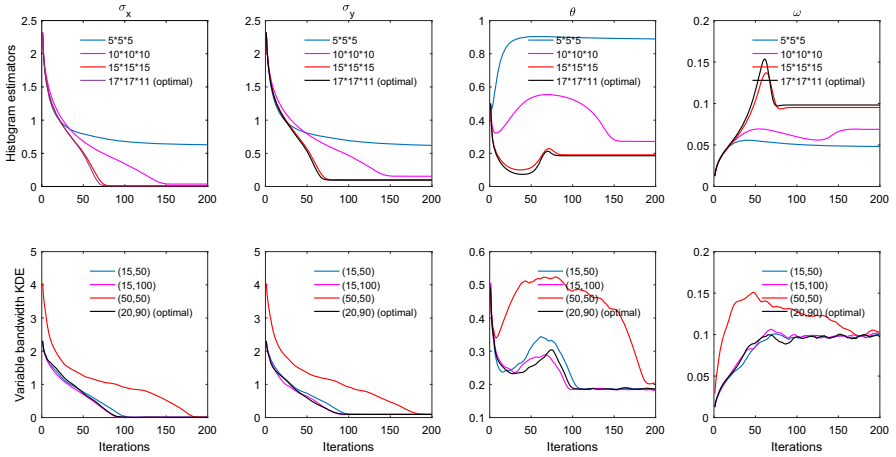

Fig. 7 Experiment 2: estimated parameter values of $\sigma_x, \sigma_y, \theta$ and ω (top) at the $k+1$th iteration for histogram estimators. Estimated parameter values of $\sigma_x, \sigma_y, \theta$ and ω (bottom) at the $k+1$th iteration for variable bandwidth KDE method

evidently inhomogeneous for $5 \times 5 \times 5$ cells, as seen in the top left of Fig. 10 by the estimated centered L-function, and the estimated centered L-function for the case of $10 \times 10 \times 10$ cells departures from zero. Different from the results of the cases of $15 \times 15 \times 15$ cells and the optimal cells for histogram estimators shown in Fig. 5, the thinned residuals are homogeneous. Which means the histogram estimators can obtain well performance for some cases with a proper selection of the number of bin.

5 Discussion

We present a semi-parametric model based on the variable bandwidth KDE for learning the background intensity function of a space–time self-exciting point process model given the space–time observations. We expand the semi-parametric methods by incorporating the inhomogeneous background process into the method. The advantages of this method compared to other semi-parametric methods are substantial for the case of estimating background intensity function, in terms of robustness and accuracy. The results in Sect. 4 demonstrate that the proposed estimation method based on variable bandwidth KDE and EM algorithm performs remarkably well at estimating semi-parametric space–time self-exciting process models. Specifically, by repeatedly simulating and re-estimating the parametric models (24) and (25), we show the bias and variance of the estimated semi-parametric model approximate zero, and using the residual analysis as the model diagnostic, we assess the goodness-of-fit of the estimated semi-parametric. The results show that the thinned residuals for variable bandwidth KDE method are evidently homogeneous, i.e., the estimated centered L-function is entirely within the 95% bounds. On the contrary, the histogram estimators can be biased when the background process is inhomogeneous in some

Table 2 Experiment 2: Estimates of parameters $\sigma_x, \sigma_y, \theta, \omega$

Tuning parameter	Parameters	$\widehat{\sigma}_x$	$\widehat{\sigma}_y$	$\widehat{\theta}$	$\widehat{\omega}$
	True values	.0100	.1000	.2000	.1000
	Histogram estimators				
$10 \times 10 \times 10$	Mean	.0349	.1549	.2685	.0696
	Standard errors ($\times 10^{-3}$)	.5446	3.1166	3.0901	1.2592
	Bias in % of true value	249.1692%	54.9977%	34.2650%	−30.3765%
$15 \times 15 \times 15$	Mean	.0105	.1001	.1897	.0954
	Standard errors ($\times 10^{-3}$)	0.1802	1.4712	2.2452	2.0443
	Bias in % of true value	−5.0002%	0.1041%	−5.1531%	−4.6041%
$17 \times 17 \times 11$ (optimal)	Mean	.0103	.0951	.1869	.0965
	Standard errors ($\times 10^{-3}$)	.1512	1.2798	2.5345	1.6937
	Bias in % of true value	3.0561%	−4.8722%	−6.5151%	−3.4988%
	Variable bandwidth KDE				
(15, 50)	Mean	.0102	.0954	.1862	.0979
	Standard errors ($\times 10^{-3}$)	.1526	1.7150	2.5225	2.4047
	Bias in % of true value	2.3748%	−4.5992%	−6.9013%	−2.0838%
(15, 100)	Mean	.0102	.0950	.1842	.0992
	Standard errors ($\times 10^{-3}$)	.1302	1.4068	2.4145	2.2439
	Bias in % of true value	1.8187%	−4.9698%	−7.9062%	−.7883%
(50, 50)	Mean	.0101	.1002	.2023	.1046
	Standard errors ($\times 10^{-3}$)	.1641	.9335	1.4597	1.5822
	Bias in % of true value	1.0627%	.1956%	1.1424%	4.6499%
(20, 90) (optimal)	Mean	.0103	.0953	.1863	.0974
	Standard errors ($\times 10^{-3}$)	.1639	1.7796	2.6050	2.5500
	Bias in % of true value	2.5507%	−4.7130%	−6.8283%	−2.5275%

cases such as experiment 1. Specifically, the thinned residuals for histogram estimators are evidently inhomogeneous when the number of the bin is small, i.e., the estimated centered L-function is entirely outside the 95% bounds. The assumption of constancy within each bin leads to the inaccurate estimates of background intensity function.

The number of cells m influences the estimation results of the histogram estimators. We further study the estimation of the partition from $20 \times 20 \times 20$ cells to $25 \times 25 \times 25$ cells (the results are not shown in this paper), and we find that the EM-type algorithm based on histogram estimators may fail to recover the background and triggering intensity functions. The reason is that, in each small bin, the estimated histogram estimators for background intensity function is too big. In addition, we note that the number of partition on any dimension used in Fox et al. (2016a) was less than 25.

The limited number of partition makes it impossible for us to improve the accuracy of the estimation of background intensity function based on histogram estimators.

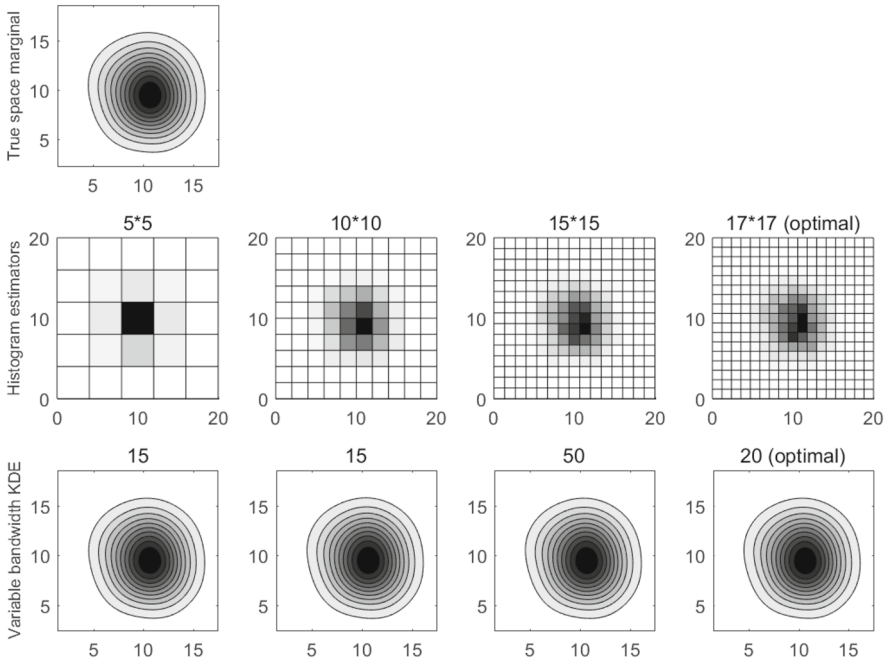

Fig. 8 Experiment 2: actual space marginals (top left). Estimated space marginals (top right) for histogram estimators. Estimated space marginals (bottom right) for variable bandwidth KDE method

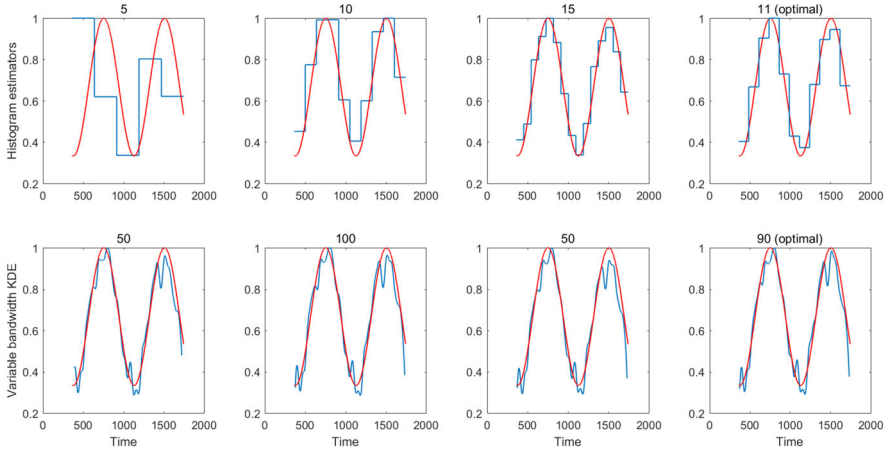

Fig. 9 Experiment 2: estimated (solid blue) and actual (solid red) time marginals (top) for histogram estimators. Estimated (solid blue) and actual (solid red) time marginals (bottom) for variable bandwidth KDE method

The variable bandwidth KDE discussed in this paper can serve as an estimation method for estimating the semi-parametric space–time self-exciting point process model.

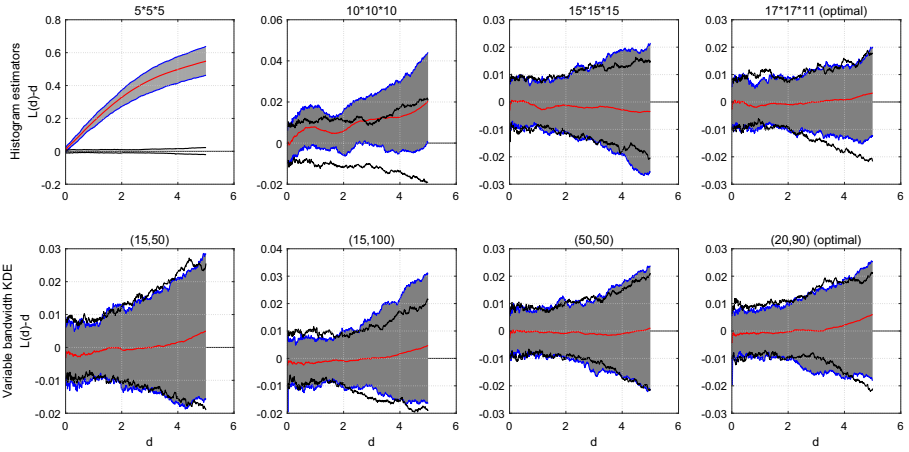

Fig. 10 Experiment 2: thinning residuals for histogram estimators (top) and variable bandwidth KDE method (bottom). Lower red: 5% bounds of the estimated centered L-function, $L(d) - d$, of homogeneous Poisson process; Upper red: 95% bounds of the estimated centered L-function of homogeneous Poisson process; Lower green: 5% bounds of the estimated centered L-function of the thinning residuals; Upper green: 95% bounds of the estimated centered L-function of the thinning residuals; Middle black: the empirical mean of the estimated centered L-function of the thinning residuals; Gray region: confidence region of the estimated centered L-function of the thinning residuals

Acknowledgements The authors would like to express their sincere gratitude to the reviewers and editor for very helpful suggestions and comments which greatly improved this paper. The authors would like to thank the National Natural Science Foundation of China (No. 91746107, 91746205); the State Scholarship Fund of China Scholarship Council (CSC); and the National Science and Engineering Research Council (NSERC) of Canada, for their funding and support.

References

Adelfio, G., Chiodi, M. (2015). Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stochastic Environmental Research and Risk assessment*, *29*(2), 443–450.

Bacry, E., Dayri, K., Muzy, J. F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B*, *85*(5), 1–12.

Bacry, E., Gaïffas, S., Muzy, J. F. (2015). A generalization error bound for sparse and low-rank multivariate Hawkes processes. arXiv:1501.00725.

Bray, A., Schoenberg, F. P. (2013). Assessment of point process models for earthquake forecasting. *Statistical Science*, *28*(4), 510–520.

Bray, A., Wong, K., Barr, C. D., Schoenberg, F. P. (2014). Voronoi residual analysis of spatial point process models with applications to California earthquake forecasts. *The Annals of Applied Statistics*, *8*(4), 2247–2267.

Brix, A., Diggle, P. J. (2011). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *63*(4), 823–841.

Clements, R. A., Schoenberg, F. P., Veen, A. (2012). Evaluation of space-time point process models using super-thinning. *Environmetrics*, *23*(7), 606–616.

Daley, D. J., Vere-Jones, D. (2003). *An introduction to the theory of point processes*: *Volume I*: *Elementary theory and methods*. New York: Springer.

Diggle, P. J. (2006). Spatio-temporal point processes: Methods and applications. In *Semstat2004* (pp. 1–45). London: CRC Press.

Diggle, P., Moraga, B., Rowlingson, B. M., Taylor, et al. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistical Science*, *28*(4), 542–563.
Fox, E. W., Schoenberg, F. P., Gordon, J. S. (2016a). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *The Annals of Applied Statistics*, *10*(3), 1725–1756.
Fox, E. W., Short, M. B., Schoenberg, F. P., Coronges, K. D., Bertozzi, A. L. (2016b). Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, *111*(514), 564–584.
Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.
Kagan, Y. Y. (2003). Accuracy of modern global earthquake catalogs. *Physics of the Earth and Planetary Interiors*, *135*(2–3), 173–209.
Liniger, T. J. (2009). *Multivariate Hawkes processes*. Ph.D. Thesis, ETH Zurich.
Marsan, D., Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, *319*(5866), 1076–1079.
Mohler, G. O. (2013). Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, *7*(3), 1525–1539.
Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.
Moller, J., Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. London: CRC Press.
Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, *27*(1), 23–31.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27.
Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402.
Ogata, Y., Katsura, K., Tanemura, M. (2003). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis. *Journal of the Royal Statistical Society*: *Series C (Applied Statistics)*, *52*(4), 499–509.
Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(2), 172–212.
Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, *98*(464), 789–795.
Veen, A., Schoenberg, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624.
Vere-Jones, D. (1992). Statistical methods for the description and display of earthquake catalogues. In A. T. Walden, P. Guttorp (Eds.), *Statistics in the environmental and earth sciences*, pp. 220–246. London: Edward Arnold.
Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, *11*(4), 503–538.
Yang, Y. X., Etesami, J., He, N., Kiyavash, N. (2018). Nonparametric Hawkes processes: Online estimation and generalization bounds. arXiv:1801.08273.
Zhuang, J. C. (2011). Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth, Planets and Space*, *63*(3), 207–216.
Zhuang, J. C., Ogata, Y., Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, *97*(458), 369–380.
Zhuang, J. C., Ogata, Y., Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*: *Solid Earth*, *109*(B5), B05301.