



Multivariate Hawkes processes with spatial covariates for spatiotemporal event data analysis

Chenlong Li¹ · Kaiyan Cui²

Received: 8 January 2023 / Revised: 29 November 2023 / Accepted: 5 December 2023
© The Institute of Statistical Mathematics, Tokyo 2024

Abstract

Spatiotemporal events occur in many disciplines, including economics, sociology, criminology, and seismology, with different patterns in space and time related to environmental characteristics, policing, and human behavior. In this paper, we propose a class of multivariate Hawkes processes with spatial covariates to consider the influence structure of spatial features in spatiotemporal events and the spatiotemporal patterns such as clustering. Baseline intensities are assumed to be a spatial Poisson regression model to explain spatial feature influence. The transfer functions are considered unknown but smooth and decreasing to explain the clustering phenomena. A semiparametric estimation method based on time discretization and local constant approximation is introduced. Transfer function estimators are shown to be consistent, and baseline intensity estimators are consistent and asymptotically normal. We examine the numerical performance of the proposed estimators with extensive simulation and illustrate the application of the proposed model to crime data obtained from Pittsburgh, Pennsylvania.

Keywords Hawkes process · Spatiotemporal event data · Mutually exciting · Semiparametric estimation

1 Introduction

As spatiotemporal event data frequently arise in many disciplines, models to analyze such data have received widespread attention. An essential pattern of these data is spatiotemporal clustering. For example, price fluctuations or transactions in finance, hot spots in crime data, and aftershocks in earthquakes exhibit clustering and are studied particularly by Vere-Jones and Davies (1966); Ogata (1988);

✉ Kaiyan Cui
cuiikaiyan@sxu.edu.cn

¹ College of Mathematics, Taiyuan University of Technology, Taiyuan 030024, China

² School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China

Bernasco and Nieuwbeerta (2004); Lewis et al. (2010); Mohler et al. (2011); Lewis (2012); Mohler (2013); Reinhart (2018).

Numerous models have been proposed to account for the clustering patterns observed in spatiotemporal event data and applied to various fields throughout economics, sociology, criminology, and seismology, etc. An instrumental class of stochastic processes in this context consists of self-exciting or Hawkes processes (Yuan et al., 2021). The general concept of the self-exciting point process has long been discussed and developed in the context of seismology with the Neymann-Scott model and other cluster processes (Vere-Jones and Davies, 1966). However, an explicit form for a self-exciting point process was not formally defined until the work of Hawkes (1971). This eventually led Ogata (1988) to develop the Epidemic Type Aftershock Sequence (ETAS) Model. Recently, self-exciting point processes have been developed to try to model human behavior in crime and terrorism in the work of Lewis et al. (2010); Egesdal et al. (2010); Mohler et al. (2011); Lewis (2012); Mohler and Short (2012); Mohler (2013); Clauset and Woodard (2013); Short et al. (2017); Park et al. (2021). The self-exciting point process is closely related to the branching process, i.e., each point of a self-exciting point process is either an immigrant (background) or a descendant (offspring or triggering) (Veen and Schoenberg, 2008). In this context, crimes can be divided into two classes, i.e., “main crimes” and their “offspring crimes.” For example, Mohler et al. (2011) proposed self-exciting point processes to describe random collections of crimes where the occurrence of one crime (“main crime”) increases the likelihood that another crime occurs (“offspring crime”) shortly after that.

In this paper we focus on modeling and analyzing the influence structure of spatial features in spatiotemporal event data and the spatiotemporal clustering patterns. The multivariate Hawkes processes with spatial covariates are proposed for this purpose. In the proposed multivariate Hawkes processes, the baseline intensities are assumed to be a spatial Poisson regression model to account for the spatial feature influence. The transfer functions are supposed to be unknown but smooth and decreasing to justify the clustering phenomena. A semiparametric estimation method based on time discretization and local constant approximation is introduced. The estimators for nonparametric transfer functions are shown to be consistent, and the estimators for the parameters of the baseline intensities are shown to be consistent and asymptotically normal. Using the proposed method, we seek to address the following questions:

- (a) Is the proposed semiparametric estimation method flexible and effective for multivariate Hawkes processes with spatial covariates?
- (b) Does the proposed multivariate Hawkes process with spatial covariates provide a significant explanation for the clustering features observed in spatiotemporal event data?

To solve the first issue, the simulation data generated from a specified model are considered. To generate the simulation data, we use a simple and efficient

multivariate Hawkes process simulation algorithm, which is an application of the simulation algorithm proposed by Zhuang et al. (2004). For the second issue, we use residual analysis to examine the fitting performance (Schoenberg, 2003). We discuss the numerical performance of the proposed estimators with extensive simulation and illustrate the application of the proposed model to crime data obtained from Pittsburgh, Pennsylvania.

The rest of this paper is organized as follows. Section 2 introduces the spatial heterogeneity and spatiotemporal dependence of spatiotemporal events. In Sect. 3, we propose the concepts of the multivariate point processes, the multivariate Hawkes processes with spatial covariates, the semiparametric estimation, the model diagnostics methods for goodness-of-fit assessment, and the simulation algorithm. Section 4 outlines the performance of the proposed estimation method used for fitting the simulation data. The utility of the proposed model is illustrated on the actual crime data from Pittsburgh, Pennsylvania. Finally, the results with a discussion are summarized in Sect. 5.

2 Spatial heterogeneity and spatiotemporal dependence

Spatiotemporal event data analysis requires a comprehensive understanding of the occurrence process of spatiotemporal events. From a social perspective, events may be correlated due to exogenous factors like the economic conditions, months, change in military operations, etc., rather than “caused” by endogenous factors such as repeat offender behavior (Lewis, 2012). For example, Bernasco and Nieuwbeerta (2004) show that burglary will repeatedly attack clusters of nearby targets because local vulnerabilities are well known to the offenders. A gang shooting may incite waves of retaliatory violence in the local set space (territory) of the rival gang (Tita and Ridgeway, 2007). Mohler et al. (2011) found that the local, contagious spread of crime leads to the formation of crime clusters in space and time. Zhuang and Mateu (2019) proposed a semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data, which is a powerful tool that combines the features of Hawkes process models with spatial and periodic components to analyze the pattern of criminal events in space and time. It has the advantage of being flexible, computationally efficient, and applicable to different types of criminal data. Yuan et al. (2021) provided estimates for multivariate spatiotemporal Hawkes processes that can effectively analyze spatiotemporal clustering characteristics of crime data. Therefore, spatiotemporal events have varying patterns in space and time, are related to the environment features, policing and human behavior, etc., and exhibit spatial heterogeneity.

On the other hand, there is spatiotemporal dependence in spatiotemporal events. A notable example is a spatiotemporal dependence between main shocks and aftershocks (Marsan and Lengline, 2008). The spatiotemporal reliance has also been noted in criminology (Lewis et al., 2010; Mohler et al., 2011; Lewis, 2012; Zhuang and Mateu, 2019). Unlike seismology where there is a clear physical explanation for the relationship between main shocks and aftershocks, there are innumerable factors that play roles in the connection between main crimes and their offspring crimes (Mohler et al., 2011;

Lewis, 2012). It is also unclear how much of the clustering is due to risk heterogeneity and how much is due to self-exciting (Mohler et al., 2011). Emergency calls, which are the consequences of actions of the insiders who know the emergency events, are another kind of spatiotemporal event data exhibiting spatiotemporal dependence. The insiders may be the natural triggers or the witnesses, even the indirect insiders who are told about the accidents or crimes by others. The connection between an emergency call and other emergency calls may be due to the spatiotemporal dependence of emergency events. Meanwhile, one emergency event may cause lots of emergency calls, which leads to strong spatiotemporal dependence of emergency calls in a small spatiotemporal region.

However, these spatiotemporal patterns are generally confounded (Reinhart and Greenhouse, 2017), requiring a method to take both into account. Figure 1 gives a simplified diagram of spatial heterogeneity and spatiotemporal dependence. In Figure 1, the arrow represents dependence between the events or between the spatial covariates and events.

3 Methods

3.1 General models

Spatiotemporal point processes are a class of available models for spatiotemporal event data, typically modeled by specifying their associated intensity functions. All finite-dimensional distributions of a spatiotemporal point process are uniquely characterized by its intensity, assuming it exists. To account for the spatiotemporal dependence of spatiotemporal event data, the intensity functions of the spatiotemporal point processes need to be conditioned by the history of spatiotemporal events occurring up to time t . Such intensity function is known as the conditional intensity function, which is defined as

$$\lambda(s, t | \mathcal{H}_t) = \lim_{\Delta s, \Delta t \downarrow 0} \frac{E[N(B(s, \Delta s) \times [t, t + \Delta t)) | \mathcal{H}_t]}{|B(s, \Delta s)| \Delta t},$$

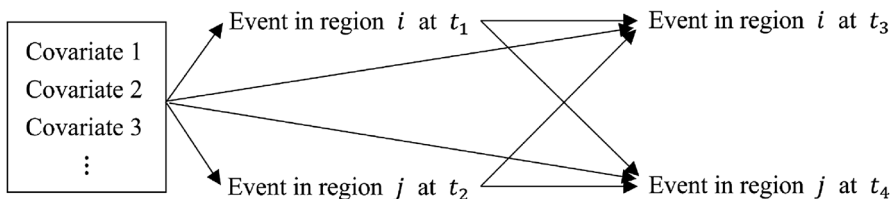


Fig. 1 Simplified diagram of spatial heterogeneity and spatiotemporal dependence: two events that occurred in i and j regions at time t_1 and t_2 affect the events that occurred in i and j regions at time t_3 and t_4 , respectively, and simultaneously, covariates affect these events that occur in i and j regions at times t_1 and t_2 , where $\max\{t_1, t_2\} < \min\{t_3, t_4\}$

where $s \in S \subset \mathbb{R}^2$ represents location, $t \in [0, T) \subset \mathbb{R}_+$ denotes time, $|B(s, \Delta s)|$ is the area of the circle $B(s, \Delta s)$ with center s and radius Δs , $N(\cdot)$ is the counting measure and $\mathcal{H}_t$ is the family of σ -algebras generated by the events occurring at times up to, but excludes t .

Modeling such spatiotemporal point processes is typically performed by specifying a particular structure for $\lambda(s, t|\mathcal{H}_t)$. The crucial problem is to determine how the conditional intensity functions depend on the history of events. This requires both physical and mathematical insights, and the scientist endeavor to find objective means of calculating probabilities of occurrence conditional on given precursory information, which is typically challenging. To address the above issue, we investigate spatiotemporal clustering patterns of the data using the multivariate point processes. The focus of our method is to reduce the difficulty of modeling, mitigate the potential for model inaccuracies, minimize model intricacy, and improve computational efficiency. Specifically, we divide the observation domain S into d disjoint areas of arbitrary shape such that $S = \cup_{i=1}^d S_i$ and consider the effect of events in one area on the occurrence of events in other areas. Let $N_i(A) := N(S_i \times A)$, $A \subset [0, T)$. N_i represents the counting measure of area S_i , and the conditional intensity functions are defined as

$$\lambda_i(t|\mathcal{H}_t) = \lim_{\Delta t \downarrow 0} \frac{E[N_i([t, t + \Delta t))|\mathcal{H}_t]}{\Delta t}, \quad i = 1, \dots, d.$$

Therefore, events occurring in different areas can be modeled by multivariate point processes $N = (N_1, \dots, N_d)'$ with conditional intensity functions $\lambda(t|\mathcal{H}_t) = (\lambda_1(t|\mathcal{H}_t), \dots, \lambda_d(t|\mathcal{H}_t))'$. Next, we shall specify a special structure for $\lambda(t|\mathcal{H}_t)$. In the remainder of this paper, we will write $N_i(t) = N_i([0, t))$ for simplicity.

3.2 Multivariate Hawkes processes with spatial covariates

A common effect between events occurring in different areas is ‘mutually exciting’: each event occurring in areas S_j can locally increase the event rate for a future period in area S_i , $j \neq i$, potentially exciting more events to happen the area S_i . To model the ‘mutually exciting’ effects, Hawkes introduced the following multivariate Hawkes processes.

Definition 1 A d -dimensional multivariate Hawkes process is a simple multivariate point process $N = (N_1, \dots, N_d)'$ such that N_i has a conditional intensity function

$$\lambda_i(t|\mathcal{H}_t) = \mu_i + \sum_{j=1}^d \int_{(-\infty, t)} g_{ij}(t-u) N_j(du), \quad i = 1, \dots, d, \quad (1)$$

where $\mu_i \in \mathbb{R}_+ = [0, +\infty)$ and $g_{ij} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and satisfy

$$\begin{cases} \int_0^\infty t g_{ij}(t) dt < \infty, \quad i, j = 1, \dots, d, \\ \text{Spr}(Q) < 1, \quad Q = (q_{ij}; q_{ij} = \int_0^\infty g_{ij}(t) dt, i, j \in \{1, \dots, d\})_{d \times d}, \end{cases} \quad (2)$$

$N_j(du) = 1$ if an infinitesimal element du includes an event t_k for some k ; otherwise $N_j(du) = 0$, $Spr(Q)$ is the spectral radius of Q .

μ_i is called the baseline/immigration intensity of N_i , which describes the dynamics of the ground process, $g_{ij}(\cdot)$ is called the transfer/excitation function, which captures the effect of past events that occurred in the j -th area on the events in the i -th area, and Q is called the branching matrix. Condition (2) ensures the existence and uniqueness of a multivariate Hawkes process with associated conditional intensity function as in **Definition 1** (Embrechts et al., 2011; Lewis, 2009). Existence here means that one can find a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that is rich enough to support such a multivariate point process. Uniqueness implies that any two multivariate point processes complying with **Definition 1** have the same distribution.

In applications, the baseline intensities are correlated with some exogenous variables such as population density (population per m^2), block area (m^2), and fraction of males (proportion from 0 to 1). We are interested in the influence structure of spatial features on the conditional intensity function. Therefore, in the spirit of the spatial Poisson regression, we assume the effect of the spatial covariates on the baseline intensities of N is log-linear. Let $\mathbf{x}_i \in \mathbb{R}^m$ be the column vector of spatial covariates, where m denotes the number of covariates. The conditional intensity functions of the multivariate Hawkes processes with spatial covariates are given by

$$\lambda_i(t|\mathcal{H}_t) = \exp(\beta \mathbf{x}_i) + \sum_{j=1}^d \int_{(-\infty, t)} g_{ij}(t-u) N_j(du), \quad i = 1, \dots, d, \quad (3)$$

where $\beta \in \mathbb{R}^m$ is row vector.

Remark 1 The model for the conditional intensity function can be extended to a general model. That is

$$\lambda_i(t|\mathcal{H}_t) = f_i(\beta \mathbf{x}_i) + \sum_{j=1}^d \int_{(-\infty, t)} g_{ij}(t-u) N_j(du), \quad i = 1, \dots, d, \quad (4)$$

where f_i is an unknown link function for the single index $\beta \mathbf{x}_i$. Specifically, the baseline intensities are assumed to be the single-index model instead of the spatial Poisson regression model. Such a structure is quite general and provides more flexibility to capture the influence structures of spatial features, while it becomes more challenging to obtain the estimators and their theoretical properties. Thus, we focus on proposed conditional intensity function (3) in this paper. Model (4) will be investigated elsewhere.

3.3 Semiparametric estimation

The transfer functions $g_{ij}(\cdot)$, $i = 1, \dots, d$, $j = 1, \dots, d$ play an important role for characterization of the ‘mutually exciting’ effects. The shape of transfer functions:

decreasing, increasing, U-shaped, or upside-down U-shaped presents information about the time-dependent reliability behavior of the exciting pattern. Existing work often restricts transfer functions to certain parametric forms, such as exponential function and polynomial function, and parameters are estimated by the maximum likelihood method. However, the specified parametric forms of transfer functions are not always justified, leading to model misspecification. A more flexible approach toward using the nonparametric methods, such as kernel density estimation and B-spline-based estimation, can be employed to explore the transfer functions. Here we consider a nonparametric method to estimate the transfer functions. Eichler et al. (2016) and Kirchner (2017) have used this method for fitting multivariate Hawkes processes, but they do not account for covariates. In the following we generalize this method to the case with covariates.

To estimate the transfer functions, we divide the real line into fixed intervals of length $h > 0$ and define $\mathbf{y}_k^h = (y_{i,k}^h; y_{i,k}^h = N_i(kh) - N_i(kh - h), i = 1, \dots, d)_{d \times 1}$ for all $k \in \mathbb{Z}$. $\{\mathbf{y}_k^h, k \in \mathbb{Z}\}$ is a d -dimensional integer-valued time series representing the numbers of jumps in a sequence of subintervals $\{[kh - h, kh), k \in \mathbb{Z}\}$. If the transfer functions $g_{ij}(\cdot), i = 1, \dots, d, j = 1, \dots, d$ are continuous, and h is small enough, one has the following approximation

$$\begin{aligned} E[y_{i,k+1}^h | \mathcal{H}_{kh}] &= E[N_i([kh, kh + h)) | \mathcal{H}_{kh}] \\ &\approx \lambda_i(kh | \mathcal{H}_{kh})h \\ &= h \exp(\beta \mathbf{x}_i) + h \sum_{j=1}^d \int_{(-\infty, kh)} g_{ij}(kh - u) N_j(du) \\ &= h \exp(\beta \mathbf{x}_i) + h \sum_{j=1}^d \sum_{s=1}^{\infty} \int_{[(k-s)h, (k-s+1)h)} g_{ij}(kh - u) N_j(du) \\ &\approx h \exp(\beta \mathbf{x}_i) + h \sum_{j=1}^d \sum_{s=1}^{\infty} g_{ij}(sh) y_{i,k-s+1}^h, \quad i = 1, \dots, d. \end{aligned} \quad (5)$$

Equation (5) can be rewritten as the following vector form

$$\begin{aligned} E[\mathbf{y}_{k+1}^h | \mathcal{H}_{kh}] &\approx \boldsymbol{\mu}^h + \sum_{s=1}^{\infty} \mathbf{g}(sh) \mathbf{y}_{k-s+1}^h, \\ &\approx \boldsymbol{\mu}^h + G_p^h \mathbf{Y}_{k,p}^h, \end{aligned} \quad (6)$$

where $\boldsymbol{\mu}^h = (\mu_i^h; \mu_i^h = h \exp(\beta \mathbf{x}_i), i = 1, \dots, d)_{d \times 1}$, $\mathbf{g}(sh) = (hg_{ij}(sh); i, j = 1, \dots, d)_{d \times d}$, $G_p^h = (\mathbf{g}(sh); s = 1, \dots, p)_{d \times dp}$, $\mathbf{Y}_{k,p}^h = (\mathbf{y}_{k-s+1}^h; s = 1, \dots, p)_{dp \times 1}$, p is the number of truncated terms in approximating the sum of infinite terms by partial sum. Let $[T_*, T^*]$ represent an observation period, where T_* denotes the observation start time, and T^* denotes the observation end time. Based on linear approximation (6), one can estimate the transfer functions $g_{ij}(\cdot), i = 1, \dots, d, j = 1, \dots, d$ and the parameters β by solving the following least squares problem

$$\sum_{k=p+1}^{T^h} ||\mathbf{y}_k^h - \boldsymbol{\mu}^h - G_p^h \mathbf{Y}_{k,p}^h||_2^2,$$

where $T^h = \lceil \frac{T^* - T_*}{h} \rceil$.

In particular, we first obtain the estimations of G_p^h and $\boldsymbol{\mu}^h$ as follows,

$$\hat{G}_p^h = \hat{\mathbf{Y}}_p^h (\hat{\Gamma}_p^h)^{-1} \quad \text{and} \quad \hat{\boldsymbol{\mu}}^h = \bar{\mathbf{y}}^h - \hat{G}_p^h \bar{\mathbf{Y}}_p^h, \quad (7)$$

where with $T_p^h = T^h - p$,

$$\begin{aligned} \bar{\mathbf{y}}^h &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \mathbf{y}_k^h, \quad \bar{\mathbf{Y}}_p^h = \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \mathbf{Y}_{k,p}^h, \\ \hat{\mathbf{Y}}_p^h &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} (\mathbf{y}_k^h - \bar{\mathbf{y}}^h)(\mathbf{Y}_{k,p}^h - \bar{\mathbf{Y}}_p^h)', \\ \hat{\Gamma}_p^h &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} (\mathbf{Y}_{k,p}^h - \bar{\mathbf{Y}}_p^h)(\mathbf{Y}_{k,p}^h - \bar{\mathbf{Y}}_p^h)'. \end{aligned}$$

Thus, based on the definition of G_p^h and $\boldsymbol{\mu}^h$, one has

$$\hat{g}_{ij}(sh) = \frac{1}{h} \hat{G}_p^h(i, j + (s-1)d), \quad i, j = 1, \dots, d, \quad s = 1, \dots, p, \quad (8)$$

$$\hat{\boldsymbol{\beta}} = \ln\left(\frac{\hat{\boldsymbol{\mu}}^h}{h}\right) X' (XX')^{-1}, \quad X = (\mathbf{x}_i; i = 1, \dots, d)_{m \times d}, \quad (9)$$

where $\ln(\frac{\hat{\boldsymbol{\mu}}^h}{h}) = (\ln(\frac{\hat{\mu}_i^h}{h}); i = 1, \dots, d)$, $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \dots, d$, m denotes the number of covariates. Further, one has the following step function approximation for $g_{ij}(t)$:

$$g_{ij}(t) = \hat{g}_{ij}(\lceil \frac{t}{h} \rceil h) \mathbf{1}_{[0, ph]}(t), \quad i, j = 1, \dots, d, \quad (10)$$

where $t \in \mathbb{R}$ and $\mathbf{1}_A(t)$ is the indicator function. In the simulations in Sect. 4.1, good results are obtained directly using the least squares solutions. However, the least squares method may lead to negative values in the estimates obtained from (10). Therefore, we use non-negative least squares solutions in Sects. 4.1-4.4.

Under some regularity conditions on the proposed multivariate Hawkes processes, the order p , and the time window h , one can show that estimators (9) and (10) of $\boldsymbol{\beta}$ and $g_{ij}(t)$ are consistent. We begin with the regularity conditions (Eichler et al., 2016). In the sequel, $|| \cdot ||_2$ denotes the Euclidean norm, and $|| \cdot ||$ denotes the spectral norm.

C1 Multivariate point processes $N = (N_1, \dots, N_d)'$ are stationary.

C2 Condition (2) in **Definition 1** is satisfied.

- C3 Order p and time window h are functions of $T = T^* - T_*$ such that $ph \rightarrow \infty$, $ph^2 \rightarrow 0$ and $\frac{p^2}{T} \rightarrow 0$ as $T \rightarrow \infty$.
- C4 The transfer functions $g_{ij}(t), i, j = 1, \dots, d$ are Lipschitz continuous and decrease to zero with $\|g_{ij}(t)\| \leq ct^{-1}$ and $\int_{[ph, \infty)} \|G(t)\| dt \xrightarrow{T \rightarrow \infty} 0$ where $G(t) = (g_{ij}(t); i, j = 1, \dots, d)_{d \times d}$.

Then based on Theorems 4.1 and 4.2 in Eichler et al. (2016), we obtain the following generalized theorems.

Theorem 1 (Consistency). *Under C1–C4, estimators (9) and (10) are consistent,*

$$\|\hat{\beta} - \beta\|_2 + \int_{[0, \infty)} \|\hat{G}(t) - G(t)\|_2 dt \xrightarrow{P} 0 \text{ as } T \rightarrow \infty,$$

where $\hat{G}(t) = (\hat{g}_{ij}(t); i, j = 1, \dots, d)_{d \times d}$, $T = T^* - T_*$.

Proof The proof of this theorem follows along the same lines as in Eichler et al. (2016), so we omit the proof here. $\square$

Remark 2 C1 is a common assumption. C2 guarantees the existence and uniqueness of a multivariate Hawkes process with associated conditional intensity function as in **Definition 1**. C3 and C4 balance the bias induced by the approximation in Eq. (5) and is similar to what is assumed by Eichler et al. (2016).

Remark 3 The piecewise constant approximation and least-squares estimation have been used for fitting multivariate Hawkes processes. See Eichler et al. (2016); Kirchner (2017) for details. Kirchner (2017) further pointed out that the number of points in each bin results in an integer-valued sequence, and the distribution of such sequence can be approximated by the INAR(p) model.

Remark 4 By replacing the least-squares method with the maximum likelihood method, one can obtain another semiparametric estimation. However, there is no analytical solution for estimating β and $g_{ij}(t)$ because of the nonlinearity of the likelihood function, and an iterative estimation method is required. See Marsan and Lengline (2008); Veen and Schoenberg (2008); Lewis and Mohler (2011) for details.

Theorem 2 (Asymptotic normality). *Assume that C1–C3 and the following conditions hold:*

- C5 Order p and time window h are functions of $T = T^* - T_*$ such that $\sqrt{T}p^2h^4 \rightarrow 0$ and $\frac{T}{p^9h^9} \rightarrow 0$ as $T \rightarrow \infty$.

C6 The transfer functions $g_{ij}(t), i, j = 1, \dots, d$ are Lipschitz continuous and decrease to zero with $\|g_{ij}(t)\| \leq ct^{-4}$ and $\int_{[ph, \infty)} \|G(t)\| dt \xrightarrow{T \rightarrow \infty} 0$ where $G(t) = (g_{ij}(t); i, j = 1, \dots, d)_{d \times d}$.

Then, for $T \rightarrow \infty$,

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} (XX')^{-1}XV\Sigma^{1/2}\mathcal{N},$$

where V is the diagonal matrix such that $V_{ii} = \frac{1}{\exp(\beta \mathbf{x}_i)}$, $i = 1, \dots, d$, $\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_d)'$ is a standard d -dimensional Gaussian random vector, $\Sigma = d \operatorname{diag}((Id - Q)^{-1}\boldsymbol{\mu})$ is a diagonal matrix, I is the unit matrix, Q is the branching matrix and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$, $\mu_i = \exp(\beta \mathbf{x}_i)$ is the baseline intensity of N_i , $i = 1, \dots, d$.

For the proof of Theorem 2, we need the following lemma.

Lemma 1 (Lemma 7 of Bacry et al. (2012), multivariate Hawkes process asymptotic normality). Let $\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \lambda(s|\mathcal{H}_s)ds$, where $\mathbf{N} = (N_1, \dots, N_d)'$ is a d -dimensional multivariate point process with conditional intensity function $\lambda(t|\mathcal{H}_t) = (\lambda_1(t|\mathcal{H}_t), \dots, \lambda_d(t|\mathcal{H}_t))'$, and let $\mathbf{W}(v) = (W_1(v), \dots, W_d(v))'$ be a standard d -dimensional Brownian motion. Then the martingales

$$T^{-1/2}\mathbf{M}(Tv), v \in [0, 1],$$

converge in law for the Skorokhod topology to

$$\Sigma^{1/2}\mathbf{W}(v)$$

as $T \rightarrow \infty$, where Σ is the same as defined in Theorem 2.

Proof of Theorem 2 Note that

$$\begin{aligned} \hat{\mu}^h - \mu^h &= \bar{\mathbf{y}}^h - \hat{\mathbf{G}}_p^h \bar{\mathbf{y}}_p^h - \mu^h \\ &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \left(\mathbf{y}_k^h - \hat{\mathbf{G}}_p^h \mathbf{Y}_{k,p}^h - \mu^h \right) \\ &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \left[\mathbf{y}_k^h - (\hat{\mathbf{G}}_p^h - \mathbf{G}_p^h) \mathbf{Y}_{k,p}^h - (\mathbf{G}_p^h \mathbf{Y}_{k,p}^h - \int_{(-\infty, kh)} \mathbf{g}(kh - u) \mathbf{N}(du)) \right. \\ &\quad \left. - \left(\int_{(-\infty, kh)} \mathbf{g}(kh - u) \mathbf{N}(du) + \mu^h - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s)ds \right) - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s)ds \right] \\ &= \frac{1}{T_p^h} \sum_{k=1}^{T^h} \left(\mathbf{y}_k^h - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s)ds \right) + U_{1,T} + U_{2,T} + U_{3,T} + U_{4,T} + U_{5,T} \end{aligned}$$

where

$$\begin{aligned}
 U_{1,T} &= -\frac{1}{T_p^h} \sum_{k=1}^p \left(\mathbf{y}_k^h - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s) ds \right), \\
 U_{2,T} &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \left(G_p^h - \hat{G}_p^h \right) Y_{k,p}^h, \\
 U_{3,T} &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \sum_{s=1}^p \int_{[(k-s)h, (k-s+1)h]} (\mathbf{g}(kh-u) - \mathbf{g}(sh)) N(du), \\
 U_{4,T} &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \int_{(-\infty, ph)} \mathbf{g}(kh-u) N(du), \\
 U_{5,T} &= \frac{1}{T_p^h} \sum_{k=p+1}^{T^h} \int_{(-\infty, kh)} \left(\frac{1}{h} \int_{(k-1)h}^{kh} \mathbf{g}(s-u) ds - \mathbf{g}((k-1)h-u) \right) N(du).
 \end{aligned}$$

According to Theorems 4.1 in Eichler et al. (2016), one can obtain that the last four terms $U_{2,T}$, $U_{3,T}$, $U_{4,T}$ and $U_{5,T}$ are of order $o_P(\frac{h}{\sqrt{T}})$. Further, using Lemma 1 and C3, one has

$$\begin{aligned}
 U_{1,T} &= -\frac{1}{T_p^h} \sum_{k=1}^p \left(\mathbf{y}_k^h - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s) ds \right) \\
 &= -\frac{\sqrt{ph}}{T_p^h} \frac{1}{\sqrt{ph}} \left(N(ph) - \int_0^{ph} \lambda(s|\mathcal{H}_s) ds \right) \\
 &= o_P\left(\frac{h}{\sqrt{T}}\right).
 \end{aligned}$$

and

$$\begin{aligned}
 \frac{1}{\sqrt{T_p^h h}} \sum_{k=1}^{T^h} \left(\mathbf{y}_k^h - \int_{(k-1)h}^{kh} \lambda(s|\mathcal{H}_s) ds \right) &= \sqrt{\frac{T^h}{T_p^h}} \frac{1}{\sqrt{T^h h}} \left(N(T^h h) - \int_0^{T^h h} \lambda(s|\mathcal{H}_s) ds \right) \\
 &\xrightarrow{d} \Sigma^{1/2} \mathbf{W}(1).
 \end{aligned}$$

Thus,

$$\sqrt{T} \frac{\hat{\mu}^h - \mu^h}{h} \xrightarrow{d} \Sigma^{1/2} \mathbf{W}(1).$$

This implies that

$$\begin{aligned}
 \sqrt{T}(\hat{\beta} - \beta) &= \sqrt{T} \left(\ln\left(\frac{\hat{\mu}^h}{h}\right) - \ln\left(\frac{\mu^h}{h}\right) \right) X'(XX')^{-1} \\
 &\xrightarrow{d} (XX')^{-1} X V \Sigma^{1/2} \mathbf{W}(1),
 \end{aligned}$$

where V is the diagonal matrix such that $V_{ii} = \frac{1}{\exp(\beta \mathbf{x}_i)}$, $i = 1, \dots, d$. $\square$

Remark 5 In order to meet conditions C3 and C5, there are many choices for the parameters p and h . For example, one can select $p = T^\alpha (\log(T))^{-\beta}$ and $h = T^{-\gamma}$ where $0 < \alpha \leq 0.5$, $\gamma, \beta > 0$, $1/9 + \gamma \leq \alpha \leq 2\gamma - 1/4$. In particular, if we let $\alpha = 0.5$, $\gamma = 3/8$ and $\beta = 0.1$, C3 and C5 hold. C6 is an enhancement of C4 to ensure the asymptotic normality of the proposed estimators. For instance, in many practical applications, the transfer functions usually have exponential delay, indicating that C6 is satisfied in some real scenarios.

3.4 The choice of d value and d disjoint areas

Different models can be established by selecting different d values and d disjoint areas. As discussed in Section 2, the purpose of this paper is to provide a fast and effective analytical method to study the spatial heterogeneity and dependence of spatiotemporal event data. In Section 3.1, we discuss that for data with spatial heterogeneity and spatiotemporal dependence, different spatial partitions correspond to different conditional intensity functions, i.e., each spatial partition corresponds to a multivariate Hawkes process model. Although these models are different, they all can show the spatial heterogeneity and spatiotemporal dependence of the studied data. Therefore, based on the research purpose of this paper, we can choose any spatial partition. On the other hand, differences in spatial partitioning can affect model estimation performance and interpretability (e.g., analysis of dependence and variability between data in different spaces). For example, the model estimation is poor when some spatial regions are partitioned too small or too dispersed. Also, considering that Hawkes processes are essentially a class of clustering processes, we choose the K-means method to cluster data with similar characteristics into one class and select the number of the spatial partitions with the best model estimation performance. Specifically, model selection is usually performed using the Akaike information criterion (AIC) and Bayes information criterion (BIC), which are defined as

$$\text{AIC} = -2 \text{LL} + 2q, \quad \text{and} \quad \text{BIC} = -2\text{LL} + 2q \ln(n)$$

where LL is the log-likelihood function of the model, q is the number of independent parameters, and n is the data size. For a d -dimensional multivariate Hawkes process, the log-likelihood function over a temporal interval $D = [T_*, T^*]$ is given by

$$\text{LL} = \sum_{j=1}^d \int_D \log \lambda_j(s | \mathcal{H}_s) N_j(ds) - \sum_{j=1}^d \int_D \lambda_j(s | \mathcal{H}_s) ds,$$

and the number of independent parameters $q = m + d^2 p$.

The cross-validation type technique has been proposed for selecting tuning parameters, e.g., Vere-Jones (1992); Adelfio and Chiodi (2015). In Vere-Jones (1992), two scoring methods, the K-L (Kullback–Leibler) score and the MISE (mean integrated square error) score, are proposed to determine the optimal values of the tuning parameters (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 tuning parameters. The FLP method is a special case of the K-L score. In this paper, we also use the FLP method to select the optimal tuning parameter values. To implement the FLP method, the data are first divided into two parts in time, a training period, $D_{\text{tra}} := [T_*, 0.9T^*]$, which is used to produce an intensity estimate, and a forecast forward period, $D_{\text{for}} := D - D_{\text{tra}} = (0.9T_*, T^*]$. Then the forecast is scored by using the log-likelihood of next point, i.e.,

$$\text{FLP} := \sum_{j=1}^d \int_{D_{\text{for}}} \log \hat{\lambda}_j(s|\mathcal{H}_s) N_j(ds) - \sum_{j=1}^d \int_{D_{\text{for}}} \hat{\lambda}_j(s|\mathcal{H}_s) ds,$$

where $\hat{\lambda}_j(s|\mathcal{H}_s)$ is the estimated conditional intensity function. The resulting scores can then be compared for different tuning parameters, and the optimal tuning parameter is chosen for forecast of next point (Vere-Jones, 1992; Adelfio and Chiodi, 2015).

Remark 6 We have considered different model evaluation criteria including AIC, BIC, residual sum of squares (RSS), and cross-validation (CV). Since the proposed estimators are based on the least squares, using RSS to select models may lead to overfitting. So it is not considered in this paper. In addition, since the log-likelihood function of observations in the application is large, the AIC and BIC showed similar results. A comparison of the AIC and BIC criterions for choosing d value and d disjoint areas is summarized in simulations.

3.5 Goodness-of-Fit test

Now we pay our attention to the examination of the goodness-of-fit of the used multivariate Hawkes processes for the spatiotemporal event data. Methods for evaluating the fit of multivariate Hawkes processes are presented using residual analysis methods, such as thinning, rescaling, and superposition, which involve transforming the point process utilizing a model for the conditional intensity function and then inspecting the uniformity of the result (Clements et al., 2012). Making use of residual analysis can help to identify defects in present multivariate Hawkes processes and to suggest ways in which the models may be improved (Schoenberg, 2003). In the following, we just introduce the simple and efficient rescaling residual analysis method.

Suppose we have observed a realization of a multivariate Hawkes process N with event times $\{\{t_{ik}, k = 1, \dots, n_i\}, i = 1, \dots, d\}$ over a temporal interval $D = [T_*, T^*]$, where sequence $\{t_{ik}, k = 1, \dots, n_i\}$ is the realization of the i -th component N_i for $i \in \{1, \dots, d\}$. Define the sequences $\{\tau_{i,k}, k = 1, \dots, n_i\}, i = 1, \dots, d$ of transformed times by

$$\tau_{i,k} = \exp(\beta \mathbf{x}_i)(t_{ik} - T_*) + \sum_{j=1}^d \int_{(-\infty, t_{ik})} [\bar{g}_{ij}(t_{ik} - u) - \bar{g}_{ij}(T_* - u)] N_j(du), \quad (11)$$

where $\bar{g}_{ij}(t) = \int_{(0,t)} g_{ij}(s) ds$, $j = 1, \dots, d$. Then the sequences $\{\tau_{i,k}, k = 1, \dots, n_i\}, i = 1, \dots, d$, are independent Poisson processes with unit intensity (Lewis, 2009).

We would expect that if the estimated conditional intensity function $\hat{\lambda}(t|\mathcal{H}_t)$ is correct, the transformed times

$$\hat{\tau}_{i,k} = \exp(\hat{\beta}\mathbf{x}_i)(t_{ik} - T_*) + \sum_{j=1}^d \int_{(T_*, t_{ik})} \bar{g}_{ij}(t_{ik} - u) N_j(du), \quad (12)$$

where

$$\begin{aligned} \bar{g}_{ij}(t) &= \int_{[0,t)} \hat{g}_{ij}(s) ds \\ &= \sum_{s=1}^{\lceil \frac{t}{h} \rceil - 1} h \hat{g}_{ij}(sh) + [t - (\lceil \frac{t}{h} \rceil - 1)h] \hat{g}_{ij}(\lceil \frac{t}{h} \rceil h), \quad i, j = 1, \dots, d, \end{aligned}$$

will be the Poisson processes with unit intensity, then some test statistics can be used to make a goodness-of-fit test. This will detect if the transformed times still have clustering not accounted for by the used multivariate Hawkes processes. In Sect. 4.2, we shall create a quantile-quantile (Q-Q) plot for $\{\hat{\tau}_{ik}\}$ using uniformly distribution to check $\{\hat{\tau}_{ik}\}$ from a Poisson process with unit rate. Otherwise a quantitative alternative is to apply a Kolmogorov–Smirnov test. On the other hand, to assess the independence of the transformed times $\{\hat{\tau}_{i,k}\}_{k=1}^{n_i}$, $i = 1, \dots, d$, we use Chi-square test. Let $\varepsilon_{i,k} := \hat{\tau}_{i,k+1} - \hat{\tau}_{i,k}$, $i = 1, \dots, d$. The test procedure is as follows: firstly, calculate the difference sequences $\{\varepsilon_{i,k}\}_{k=1}^{n_i-1}$, $i = 1, \dots, d$, then $\{\varepsilon_{i,k}\}_{k=1}^{n_i-1}$, $i = 1, \dots, d$ all obey the exponential distribution; secondly, select any two sequences $\{\varepsilon_{i,k}\}_{k=1}^{n_i-1}$ and $\{\varepsilon_{j,k}\}_{k=1}^{n_j-1}$, $1 \leq i < j \leq d$ and divide their value ranges into m_i and m_j mutually disjoint intervals $A_1, A_2, \dots, A_{m_i}$ and $B_1, B_2, \dots, B_{m_j}$; next, take M samples randomly from sequence $\{\varepsilon_{i,k}\}_{k=1}^{n_i-1}$ and M samples in the same order from sequence $\{\varepsilon_{j,k}\}_{k=1}^{n_j-1}$ to obtain the paired sequence $\{(\varepsilon_{i,\ell_k}, \varepsilon_{j,\ell_k})\}_{k=1}^M$; then, count the number $n_{k,l}$ for which the paired sequence $\{(\varepsilon_{i,\ell_k}, \varepsilon_{j,\ell_k})\}_{k=1}^M$ falls in the interval $A_k \times B_l$, $1 \leq k \leq m_i$, $1 \leq l \leq m_j$; finally, calculate the following Chi-square test statistic based on $n_{k,l}$,

$$\chi_{i,j}^2 = M \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \frac{n_{i,j} - (\sum_{k=1}^{m_i} n_{i,j})(\sum_{l=1}^{m_j} n_{i,j})/M}{(\sum_{k=1}^{m_i} n_{i,j})(\sum_{l=1}^{m_j} n_{i,j})},$$

and compute the corresponding p value. If the p value is less than a pre-determined significance level (e.g., 0.05), we can reject the null hypothesis and conclude that the two random sequences $\{\hat{\tau}_{i,k}\}_{k=1}^{n_i}$ and $\{\hat{\tau}_{j,k}\}_{k=1}^{n_j}$ are not independent; otherwise, we accept the null hypothesis and conclude that they are independent. In the simulations, we set $M = 1000$ and $m_i = m_j = 20$.

Note that the estimators of the branching matrix Q can be given by $\hat{q}_{ij} = \hat{g}_{ij}(\infty) = \sum_{s=1}^p h \hat{g}_{ij}(sh)$, $i, j = 1, \dots, d$. Using Theorem 1, we obtain the consistency of the estimators of the branching matrix Q .

Proposition 1 (Consistency). *Under C1–C4, the estimators $\hat{q}_{ij} = \bar{g}_{ij}(\infty) = \sum_{s=1}^p h\hat{g}_{ij}(sh)$, $i, j = 1, \dots, d$, are consistent.*

3.6 Simulation method

We use a simple and efficient simulation method to simulate a multivariate Hawkes process. This simulation method is an application of Zhuang's simulation algorithm in Zhuang et al. (2004), which is performed using the branching structure of the multivariate Hawkes process with time-independent baseline intensity as shown below algorithm. For more simulation methods and the rationality of the proposed method, see Ogata (1981); Moller and Waagepetersen (2003); Daley and Vere-Jones (2003).

Algorithm 1 (Simulation algorithm) Begin with a fully specified conditional intensity function $\lambda(t|\mathcal{H}_t) = (\lambda_1(t|\mathcal{H}_t), \dots, \lambda_d(t|\mathcal{H}_t))'$, and calculate branching matrix $Q = (q_{ij}; i, j \in \{1, \dots, d\})_{d \times d}$.

Step 1. Generate events from the background process using the baseline intensities $\exp(\beta \mathbf{x}_i)$, $i = 1, \dots, d$ over observation interval $D = [T_*, T^*]$, respectively, by using a simulation method for Poisson processes (e.g., Daley and Vere-Jones (2003)). Call these catalogs of events $G_i^{(0)}$, $i = 1, \dots, d$.

Step 2. Let $l = 0$.

Step 3. For each event k in $G_j^{(l)}$, $j = 1, \dots, d$, simulate its $N_{ij}^{(k)}$ offspring, where $N_{ij}^{(k)} \sim \text{Poisson}(q_{ij})$ (with q_{ij} defined as mean number of offspring), and the offspring's time are generated from the transfer function $g_{ij}(t - \cdot)$, normalized as a probability density. Call these offspring $O_{ijk}^{(l)}$.

Step 4. Let $G_i^{(l+1)} = \left(\bigcup_{j=1}^d \bigcup_{k \in G_j^{(l)}} O_{ijk}^{(l)} \right) \cap D$.

Step 5. If $\bigcup_{i=1}^d G_i^{(l)}$ is not empty, set $l = l + 1$ and return to Step 3. Otherwise, return $G_i = \bigcup_{j=0}^l G_i^{(j)}$, $i = 1, \dots, d$ as the final set of simulated events.

4 Numerical experiments

4.1 Simulation: multivariate cases

In this section, the simulations are carried out by algorithm 1 in observation intervals $D = [500, 2500]$, $[500, 4500]$ and $[500, 8500]$, respectively. In particular, the artificial data generated from conditional intensity function (13) with mean sample sizes 32500, 65000, and 130000 are used to verify the performance of the proposed semiparametric estimation method. Set $d = 6$, define

$$\lambda_i(t|\mathcal{H}_t) = \exp(\beta \mathbf{x}_i) + \sum_{j=1}^d \int_{(-\infty, t)} q_{ij} \omega_{ij} \exp(-\omega_{ij}(t-u)) N_j(du), \quad i = 1, \dots, d, \quad (13)$$

where $\beta = (\beta_0, \beta_1, \beta_2) = (-4, 2, 1)$, $\mathbf{x}_i = (1, x_{i1}, x_{i2})'$, $x_{i1} \sim U[1, 2]$, $x_{i2} \sim U[0, 1]$,

$$Q = (q_{ij}) = \begin{bmatrix} 0.2 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.5 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0.1 \\ 0 & 0 & 0.1 & 0 & 0 & 0.4 \end{bmatrix} \quad \text{and} \quad (\omega_{ij}) = \begin{bmatrix} 5 & 2 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 5 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 3 & 0 & 0 & 6 \end{bmatrix}.$$

Based on Remark 5, we let $p = c_1 T^{1/2} (\log(T))^{-1/10}$ and $h = c_2 T^{-3/8}$ where c_1 and c_2 are two tuning parameters, T is the length of the interval D . With each combination of observation interval D , order p , and time windows h , the semiparametric estimation method was repeated 100 times. The order p and time windows h are selected by using different tuning parameters c_1 and c_2 . For each sample size, the covariates were simulated once and held constant across the 100 simulated data.

Table 1 shows summaries of the estimates of the parameters β with several different orders p and time windows h for three observation intervals [500, 2500], [500, 4500], and [500, 8500]. From the results of Table 1, we can see that the semiparametric estimation method gave unbiased estimates of β when the proper tuning parameters are chosen. The bias and the standard deviation (SD) of the estimators of β decrease with increasing time window length for all tuning parameter settings, while for large tuning parameter c_2 , the bias and SD of the estimators of β are large. The reason is that the local constant approximation in Eq. (5) reduces the estimation accuracy.

Table 2 shows summaries of the estimates of the baseline intensity $\exp(\beta X)$ with several different orders p and time windows h for three observation intervals [500, 2500], [500, 4500], and [500, 8500]. From Table 2, we obtained a better performance of estimates of $\exp(\beta X)$ than that of β , that is, the bias and SD are smaller. The estimates of branching matrix Q for the observation interval [500, 4500] with tuning parameters $c_1 = 1$ and $c_2 = 3$ are shown in the following matrix.

$$\text{Bias (SD) of } \hat{Q} : \begin{bmatrix} -0.0458 & -0.0013 & -0.0001 & 0.0025 & -0.0019 & 0.0029 \\ (0.0320) & (0.0131) & (0.0114) & (0.0121) & (0.0377) & (0.0313) \\ -0.0080 & -0.0758 & 0.0039 & 0.0197 & 0.0031 & -0.0019 \\ (0.0578) & (0.0249) & (0.0201) & (0.0231) & (0.0608) & (0.0610) \\ -0.0003 & -0.0022 & -0.0238 & 0.0022 & 0.0076 & 0.0047 \\ (0.0373) & (0.0149) & (0.0217) & (0.0152) & (0.0457) & (0.0486) \\ 0.0002 & 0.0023 & 0.0170 & -0.0738 & 0.0199 & -0.0010 \\ (0.0646) & (0.0227) & (0.0228) & (0.0250) & (0.0648) & (0.0608) \\ -0.0012 & 0.0003 & 0.0032 & -0.0003 & -0.0224 & -0.0119 \\ (0.0249) & (0.0101) & (0.0087) & (0.0096) & (0.0275) & (0.0267) \\ 0.0030 & -0.0008 & 0.0054 & 0.0003 & -0.0014 & -0.0838 \\ (0.0237) & (0.0111) & (0.0114) & (0.0101) & (0.0269) & (0.0331) \end{bmatrix}.$$

Table 1 Estimates of β with simulated data with observation intervals $D_1 = [500, 2500]$, $D_2 = [500, 4500]$ and $D_3 = [500, 8500]$, using the proposed semiparametric estimation method with different orders p and time windows h . Note: Bias and SD stand for the bias and standard deviation

c_1	c_2	$\beta_0 = -4$						$\beta_1 = 2$						$\beta_2 = 1$					
		D_1		D_2		D_3		D_1		D_2		D_3		D_1		D_2		D_3	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
0.5	1	Bias	0.0214	0.0295	0.0333	0.0968	0.0869	0.0775	0.1190	-0.1368	-0.1358	-0.1368	-0.1358	-0.1368	-0.1358	-0.1368	-0.1358	-0.1368	-0.1358
		SD	0.2876	0.2144	0.1352	0.1586	0.1132	0.0723	0.1066	0.0738	0.0572	0.0738	0.0572	0.1066	0.0738	0.0738	0.0572	0.1066	0.0738
	3	Bias	-0.0338	-0.1299	-0.0429	0.0769	0.1153	0.0613	0.0654	0.0544	0.0297	0.0613	0.0297	0.0654	0.0544	0.0544	0.0297	0.0654	0.0544
		SD	0.4015	0.3012	0.1876	0.2202	0.1641	0.0924	0.1539	0.1065	0.0861	0.0924	0.0861	0.1539	0.1065	0.1065	0.0861	0.1539	0.1065
	5	Bias	-0.2511	-0.1029	-0.1583	0.2058	0.1155	0.1202	0.1604	0.1051	0.1221	0.1202	0.1221	0.1604	0.1051	0.1051	0.1221	0.1604	0.1051
		SD	0.4881	0.3826	0.2655	0.2554	0.2114	0.1422	0.1862	0.1297	0.1122	0.1422	0.1122	0.1862	0.1297	0.1297	0.1122	0.1862	0.1297
	10	Bias	-0.2599	-0.1917	-0.1501	0.2492	0.1962	0.1533	0.2805	0.2412	0.1935	0.1533	0.1935	0.2805	0.2412	0.2412	0.1935	0.2805	0.2412
		SD	1.0699	0.4975	0.3030	0.6160	0.2706	0.1669	0.2822	0.1763	0.1516	0.1669	0.1516	0.2822	0.1763	0.1763	0.1516	0.2822	0.1763
	15	Bias	-0.1588	-0.1867	-0.2335	0.2319	0.2258	0.2226	0.3067	0.2554	0.2666	0.2226	0.2666	0.3067	0.2554	0.2554	0.2666	0.3067	0.2554
		SD	0.8168	0.5919	0.3373	0.4589	0.3372	0.1886	0.3151	0.2369	0.1705	0.1886	0.1705	0.3151	0.2369	0.2369	0.1705	0.3151	0.2369
	25	Bias	-0.1660	-0.1629	-0.1607	0.2809	0.2528	0.2309	0.4001	0.3521	0.3186	0.2309	0.3186	0.4001	0.3521	0.3521	0.3186	0.4001	0.3521
		SD	1.1424	0.8229	0.6234	0.6379	0.4448	0.3421	0.4319	0.2463	0.2154	0.3421	0.2154	0.4319	0.2463	0.2463	0.2154	0.4319	0.2463
0.8	1	Bias	0.0366	0.0032	0.0112	0.0791	0.0547	0.0463	-0.0450	-0.0709	-0.0732	0.0463	-0.0732	-0.0450	-0.0709	-0.0709	-0.0732	-0.0450	-0.0709
		SD	0.3396	0.2530	0.1395	0.1875	0.1351	0.0761	0.1320	0.0853	0.0659	0.0761	0.0659	0.1320	0.0853	0.0853	0.0659	0.1320	0.0853
	3	Bias	-0.1748	-0.1086	-0.0998	0.1369	0.1020	0.0847	0.0966	0.0513	0.0528	0.0847	0.0528	0.0966	0.0513	0.0513	0.0528	0.0966	0.0513
		SD	0.4963	0.3272	0.2279	0.2721	0.1752	0.1192	0.2077	0.1320	0.1015	0.1192	0.1015	0.2077	0.1320	0.1320	0.1015	0.2077	0.1320
	5	Bias	-0.2650	-0.1607	-0.1144	0.2031	0.1328	0.1012	0.1721	0.1509	0.1177	0.1012	0.1177	0.1721	0.1509	0.1509	0.1177	0.1721	0.1509
		SD	0.5737	0.3882	0.3211	0.3222	0.2046	0.1730	0.2393	0.1846	0.1124	0.1730	0.1124	0.2393	0.1846	0.1846	0.1124	0.2393	0.1846
	10	Bias	-0.3141	-0.2067	-0.1895	0.2820	0.2022	0.1737	0.2413	0.2185	0.2073	0.1737	0.2073	0.2413	0.2185	0.2185	0.2073	0.2413	0.2185
		SD	1.1543	0.6876	0.4340	0.6668	0.3845	0.2456	0.3399	0.2646	0.1768	0.2456	0.1768	0.3399	0.2646	0.2646	0.1768	0.3399	0.2646
	15	Bias	-0.3572	-0.2447	-0.2152	0.3426	0.2356	0.2112	0.3102	0.3466	0.2586	0.2112	0.2586	0.3102	0.3466	0.3466	0.2586	0.3102	0.3466
		SD	1.2234	0.7001	0.5688	0.6675	0.3702	0.3091	0.4013	0.3113	0.2163	0.3091	0.2163	0.4013	0.3113	0.3113	0.2163	0.4013	0.3113
	25	Bias	-0.4921	-0.2614	-0.1794	0.4365	0.3134	0.2494	0.4226	0.3511	0.2884	0.2494	0.2884	0.4226	0.3511	0.3511	0.2884	0.4226	0.3511

Table 1 (continued)

c_1	c_2	$\beta_0 = -4$			$\beta_1 = 2$			$\beta_2 = 1$		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
1	SD	1.4360	0.9337	0.9095	0.7569	0.5232	0.4917	0.5577	0.3799	0.2811
	Bias	-0.0763	-0.0145	-0.0368	0.0901	0.0491	0.0606	-0.0376	-0.0436	-0.0482
	SD	0.4120	0.2525	0.1657	0.2251	0.1438	0.0856	0.1345	0.0991	0.0593
	Bias	-0.0582	-0.1385	-0.0934	0.0808	0.1113	0.0803	0.0903	0.0687	0.0546
5	SD	0.5726	0.3673	0.2401	0.3091	0.2040	0.1337	0.2043	0.1521	0.1258
	Bias	-0.2422	-0.1508	-0.1457	0.1926	0.1368	0.1310	0.1673	0.1156	0.0718
	SD	0.8660	0.5552	0.3279	0.5017	0.2989	0.1775	0.2951	0.1846	0.1331
	Bias	-0.5085	-0.3013	-0.1815	0.3743	0.2450	0.1675	0.2805	0.2611	0.1972
10	SD	1.1415	0.6468	0.3939	0.6342	0.3391	0.2119	0.4833	0.2710	0.1780
	Bias	-0.3954	-0.2698	-0.1903	0.3333	0.2528	0.1841	0.3838	0.3281	0.2939
	SD	1.3126	0.7938	0.6058	0.7097	0.4487	0.2970	0.5513	0.3588	0.2715
	Bias	-0.2871	-0.3408	-0.1641	0.3111	0.3329	0.2447	0.4415	0.3646	0.3081
25	SD	1.6225	1.3552	0.8845	0.8241	0.7127	0.4824	0.6587	0.5216	0.3615

Table 2 Estimates of $\exp(\beta \mathbf{x}_i)$ with simulated data with observation intervals $D_1 = [500, 2500]$, $D_2 = [500, 4500]$ and $D_3 = [500, 8500]$, using the proposed semiparametric estimation method with different orders p and time windows h

c_1	c_2	$\exp(\beta \mathbf{x}_1) = 0.9121$			$\exp(\beta \mathbf{x}_2) = 1.4311$			$\exp(\beta \mathbf{x}_3) = 0.4545$			$\exp(\beta \mathbf{x}_4) = 2.2070$			$\exp(\beta \mathbf{x}_5) = 0.5612$			$\exp(\beta \mathbf{x}_6) = 0.4342$			
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	
0.5	1	Bias	0.1637	0.1476	0.1333	0.2177	0.1838	0.1614	0.0111	0.0003	-0.0036	0.2159	0.1469	0.1151	0.0991	0.0916	0.0836	0.0089	-0.0015	-0.0051
		SD	0.0520	0.0425	0.0261	0.0648	0.0471	0.0286	0.0559	0.0411	0.0245	0.1330	0.0805	0.0609	0.0407	0.0341	0.0210	0.0551	0.0406	0.0242
	3	Bias	0.1226	0.0916	0.0735	0.2315	0.1832	0.1358	0.0625	0.0280	0.0266	0.4344	0.3440	0.2418	0.0625	0.0401	0.0367	0.0595	0.0258	0.0248
		SD	0.0663	0.0495	0.0397	0.0860	0.0662	0.0443	0.0799	0.0587	0.0314	0.2102	0.1440	0.1007	0.0514	0.0382	0.0323	0.0791	0.0581	0.0311
	5	Bias	0.1702	0.1352	0.0911	0.3723	0.2757	0.2122	0.0754	0.0678	0.0476	0.7778	0.5475	0.4638	0.0698	0.0625	0.0344	0.0708	0.0642	0.0448
		SD	0.0967	0.0718	0.0554	0.1235	0.0979	0.0676	0.1007	0.0763	0.0534	0.2801	0.1854	0.1405	0.0739	0.0516	0.0412	0.0994	0.0756	0.0529
10	1	Bias	0.2929	0.2414	0.1843	0.6501	0.5321	0.4045	0.2003	0.1480	0.1097	1.4189	1.1531	0.8697	0.1267	0.1055	0.0804	0.1923	0.1408	0.1041
		SD	0.1515	0.0934	0.0715	0.2293	0.1238	0.0857	0.1997	0.1212	0.0725	0.4775	0.2980	0.2116	0.1063	0.0702	0.0516	0.1972	0.1198	0.0720
	15	Bias	0.3926	0.3181	0.2564	0.8308	0.6750	0.5756	0.2649	0.1862	0.1493	1.7873	1.4244	1.2623	0.1841	0.1448	0.1082	0.2543	0.1774	0.1414
		SD	0.1877	0.1222	0.0912	0.2617	0.1677	0.1106	0.2327	0.1509	0.0903	0.6169	0.4156	0.2538	0.1292	0.0850	0.0632	0.2302	0.1497	0.0894
	25	Bias	0.5508	0.4513	0.3838	1.1755	0.9697	0.8280	0.4044	0.3000	0.2456	2.6118	2.1076	1.7941	0.2617	0.2108	0.1767	0.3899	0.2871	0.2342
		SD	0.2682	0.1739	0.1181	0.3960	0.2621	0.1734	0.3433	0.2140	0.1449	1.0685	0.6064	0.4549	0.1877	0.1259	0.0879	0.3401	0.2116	0.1434
0.8	1	Bias	0.0919	0.0796	0.0712	0.1353	0.1028	0.0879	0.0094	0.0008	-0.0022	0.1671	0.0917	0.0666	0.0517	0.0485	0.0442	0.0081	-0.0002	-0.0030
		SD	0.0555	0.0444	0.0270	0.0611	0.0603	0.0341	0.0682	0.0448	0.0279	0.1381	0.1205	0.0797	0.0445	0.0345	0.0208	0.0675	0.0444	0.0276
	3	Bias	0.1001	0.0883	0.0658	0.2164	0.1743	0.1363	0.0435	0.0298	0.0236	0.4486	0.3278	0.2671	0.0407	0.0400	0.0279	0.0410	0.0277	0.0219
		SD	0.0869	0.0597	0.0442	0.1040	0.0788	0.0476	0.0951	0.0607	0.0450	0.2648	0.1789	0.1099	0.0672	0.0461	0.0352	0.0942	0.0600	0.0445
	5	Bias	0.1560	0.1234	0.0999	0.3526	0.2771	0.2221	0.0779	0.0710	0.0579	0.7621	0.6032	0.4753	0.0608	0.0509	0.0420	0.0737	0.0674	0.0549
		SD	0.1208	0.0990	0.0618	0.1610	0.1098	0.0867	0.1232	0.0799	0.0605	0.3494	0.2037	0.1751	0.0843	0.0720	0.0453	0.1218	0.0793	0.0598
10	1	Bias	0.2890	0.2346	0.1873	0.6248	0.5062	0.4190	0.1798	0.1386	0.1126	1.3235	1.0813	0.9123	0.1246	0.1038	0.0789	0.1728	0.1321	0.1070
		SD	0.1725	0.1407	0.0909	0.2468	0.1863	0.1164	0.2214	0.1324	0.1004	0.5426	0.4003	0.2464	0.1224	0.1010	0.0635	0.2184	0.1308	0.0995
	15	Bias	0.4034	0.3069	0.2524	0.8704	0.7043	0.5623	0.2461	0.2180	0.1544	1.8896	1.6283	1.2342	0.1804	0.1321	0.1092	0.2363	0.2086	0.1469
		SD	0.2310	0.1743	0.1062	0.3215	0.2390	0.1459	0.2839	0.1830	0.1238	0.8565	0.5990	0.3720	0.1726	0.1242	0.0818	0.2803	0.1814	0.1229
	25	Bias	0.5213	0.4778	0.3960	1.1577	1.0223	0.8323	0.3272	0.2942	0.2466	2.6659	2.2193	1.7634	0.2288	0.2184	0.1868	0.3144	0.2814	0.2358
		SD	0.3481	0.2455	0.1563	0.5240	0.3617	0.2222	0.3773	0.2529	0.2014	1.4653	0.8425	0.5922	0.2391	0.1656	0.1246	0.3733	0.2495	0.1986
1	1	Bias	0.0748	0.0601	0.0568	0.1130	0.0826	0.0776	0.0016	0.0023	-0.0058	1.1404	0.0869	0.0748	0.0398	0.0350	0.0320	0.0006	0.0015	-0.0065

Table 2 (continued)

c_1	c_2	$\exp(\beta_{x_1}) = 0.9121$			$\exp(\beta_{x_2}) = 1.4311$			$\exp(\beta_{x_3}) = 0.4545$			$\exp(\beta_{x_4}) = 2.2070$			$\exp(\beta_{x_5}) = 0.5612$			$\exp(\beta_{x_6}) = 0.4342$		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
3	SD	0.0653	0.0509	0.0324	0.0784	0.0657	0.0409	0.0754	0.0483	0.0266	0.1495	0.1170	0.0751	0.0515	0.0360	0.0252	0.0747	0.0479	0.0264
	Bias	0.1154	0.0802	0.0653	0.2288	0.1699	0.1353	0.0704	0.0306	0.0257	0.4566	0.3407	0.2673	0.0571	0.0328	0.0280	0.0676	0.0286	0.0240
5	SD	0.1068	0.0586	0.0562	0.1289	0.0740	0.0636	0.1072	0.0747	0.0504	0.2492	0.1936	0.1326	0.0820	0.0452	0.0410	0.1063	0.0740	0.0500
	Bias	0.1601	0.1299	0.1099	0.3537	0.2744	0.2225	0.1017	0.0659	0.0369	0.7602	0.5626	0.4271	0.0655	0.0572	0.0481	0.0977	0.0625	0.0342
10	SD	0.1347	0.0876	0.0620	0.1678	0.1180	0.0857	0.1680	0.1015	0.0593	0.3789	0.2634	0.1935	0.0967	0.0700	0.0462	0.1657	0.1005	0.0588
	Bias	0.2799	0.2274	0.1805	0.6351	0.5238	0.4021	0.1469	0.1330	0.1074	1.4248	1.1798	0.8743	0.1083	0.0920	0.0767	0.1401	0.1262	0.1019
15	SD	0.2199	0.1309	0.0877	0.2835	0.1630	0.1180	0.2185	0.1315	0.0964	0.7320	0.4499	0.2963	0.1502	0.1012	0.0640	0.2153	0.1294	0.0950
	Bias	0.3738	0.3092	0.2383	0.8534	0.7012	0.5488	0.2666	0.2092	0.1714	2.0239	1.6022	1.2705	0.1583	0.1317	0.1034	0.2574	0.2003	0.1638
25	SD	0.2729	0.1836	0.1286	0.4283	0.2585	0.1622	0.3173	0.1882	0.1264	1.3154	0.6861	0.4797	0.1753	0.1194	0.1010	0.3156	0.1863	0.1251
	Bias	0.5075	0.4349	0.4179	1.1041	0.9419	0.8743	0.4201	0.2994	0.2753	2.6422	2.1419	1.8742	0.2458	0.1999	0.1985	0.4086	0.2888	0.2643
	SD	0.3799	0.2940	0.1914	0.5412	0.3821	0.2454	0.4799	0.3284	0.2480	1.5532	1.1454	0.6634	0.2782	0.2112	0.1387	0.4789	0.3256	0.2458

In the above matrix, the values outside the brackets represent the bias of the estimated branching matrix, and the intervals in the brackets represent the SD of the estimated branching matrix. From the results we note that the proposed semiparametric estimation method gave unbiased estimates of the branching matrix.

Figure 2 shows the 95% confidence limits and the mean value of the estimates of the transfer functions for observation interval [500, 4500], tuning parameters $c_1 = 1$ and $c_2 = 3$. The estimation results for the observation interval [500,4500] are shown because the length of the time interval observed in the application is about 4000. Although the optimal value of the tuning parameter c_2 is 1 as can be seen from Table 1, we choose $c_2 = 3$ to show that good estimation results are obtained even if the tuning parameter is not optimal. Figure 2 shows that the estimated transfer functions are unbiased. Figure 3 shows the quantile-quantile (QQ) plot of the residuals of the six different dimensions to assess the goodness of fit of the considered conditional intensity function. If a multivariate Hawkes process fits the data well, the corresponding residuals should be close to a homogeneous Poisson process with unit intensity. One can then graphically test the goodness of fit by comparing the quantiles of the residuals with that of the uniform random variable. The QQ plot in Fig. 3 shows that the estimated multivariate Hawkes process is close to the diagonal line. The bottom six QQ plots are plotted by 300 transformed times randomly sampled from all transformed times for visual effect.

4.2 Simulation: spatiotemporal cases with known partitions

In this section, the artificial data generated from following conditional intensity function (14) of the spatiotemporal Hawkes process with mean sample sizes

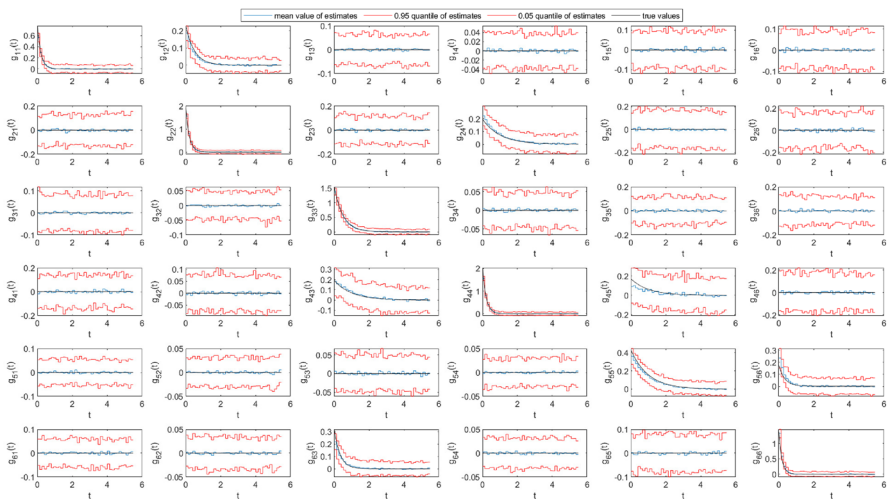


Fig. 2 Estimation of the transfer functions of the multivariate Hawkes process for simulated data: mean semiparametric estimate (blue solid), 0.95 quantile of semiparametric estimate (top red solid), 0.05 quantile of semiparametric estimate (bottom red solid), and true transfer functions (black solid)

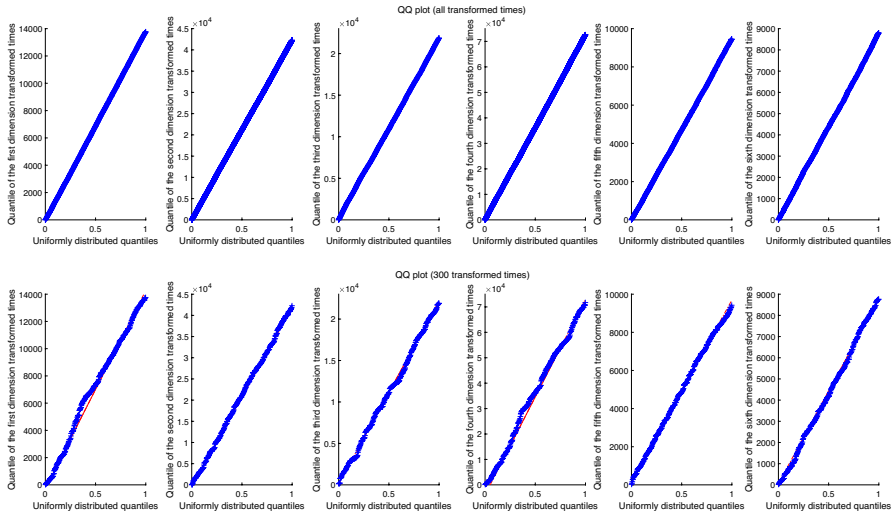


Fig. 3 Quantile-quantile plot of the residuals of the three different dimensions: all transformed times (top) and 300 transformed times randomly sampled from all transformed times (bottom)

13600, 20400, and 27600 are used to verify the performance of the proposed semiparametric estimation method. In particular, the simulations are performed in the common spatial observation interval $S = [0, 10] \times [0, 10]$ and time observation intervals $D = [0, 100]$, $[0, 150]$, and $[0, 200]$, respectively. Define

$$\lambda(x, y, t | \mathcal{H}_t) = \exp\{\beta \mathbf{s}(x, y)\} + \int_{S \times (-\infty, t)} g(x - u, y - v, t - w) N(dw, dv, du), \quad (14)$$

where $\beta = (\beta_0, \beta_1, \beta_2) = (-4, 2, 1)$, $\mathbf{s}(x, y) = (1, s_1(x, y), s_2(x, y))'$,

$$s_k(x, y) = \sum_{l=1}^4 \left\{ \sum_{\{0 \leq k_1, k_2 \leq 50: (z_{k_1}^\mu, z_{k_2}^\mu) \in S_l\}} \tilde{s}_k(z_{k_1}^\mu, z_{k_2}^\mu) / 625 \right\} I\{(x, y) \in S_l\},$$

$$\tilde{s}_k(x, y) = \sum_{i,j=0}^{49} v_k(z_i^\mu, z_j^\mu) I\{z_i^\mu \leq x < z_{i+1}^\mu, z_j^\mu \leq y < z_{j+1}^\mu\},$$

$$v_k(x, y) = \sum_{i,j=1}^2 \frac{50\theta_{i+(j-1)*2,k}}{2\pi\sigma_i^\mu\sigma_j^\mu} \exp\left\{-\frac{(x-\mu_i^\mu)^2}{2(\sigma_i^\mu)^2} - \frac{(y-\mu_j^\mu)^2}{2(\sigma_j^\mu)^2}\right\},$$

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

$\{z_i^\mu\}_{i=0}^{50}$ is an equally spaced division of the interval $[0, 10]$, $S_1 = [0, 5] \times [0, 5]$, $S_2 = [5, 10] \times [0, 5]$, $S_3 = [0, 5] \times [5, 10]$, $S_4 = [5, 10] \times [5, 10]$, $\theta_{i,1} \sim U[1, 2]$, $\theta_{i,2} \sim U[0, 1]$, $i = 1, 2, 3, 4$, $\mu_1^x = \mu_1^y = 2.5$, $\mu_2^x = \mu_2^y = 7.5$, $\sigma_1^\mu = 3$, $\sigma_2^\mu = 4$, $\omega = 0.1$,

$\alpha = 0.3$, $\sigma_1 = 0.01$, $\sigma_2 = 0.1$. The background intensity function in conditional intensity function (14) takes four different constant values over four regions S_l , $l = 1, 2, 3, 4$. In particular, the background intensity function values are obtained by a two-dimensional step function. Using a step function instead of a continuous function facilitates the calculation of the region of an arbitrary partition in the next simulation. Figure 4 shows the spatial distribution map of one realization of the spatiotemporal Hawkes process with conditional intensity function (14) during the observation time $[0, 100]$, where the points in different subspace regions of the space S are marked with different colors to visualize the spatial segmentation.

In this simulation, we assume the regions S_l , $l = 1, 2, 3, 4$ are known and define four temporal point processes on each of the four regions, thereby creating a 4-dimensional multivariate Hawkes process on region S . As in Sect. 4.1, we let $p = c_1 T^{1/2}(\log(T))^{-1/10}$ and $h = c_2 T^{-3/8}$ where c_1 and c_2 are two tuning parameters, T is the length of the interval D . With each combination of observation interval D , order p and time windows h , the semiparametric estimation method was repeated 100 times. Table 3 shows summaries of the estimates of the parameters β with several different orders p and time windows h for three observation intervals $[0, 90]$, $[0, 135]$, and $[0, 180]$. Observations in the intervals $(90, 100]$, $(135, 150]$, and $(180, 200]$ are used to calculate the FLPs. From the results of Table 3, we can see that the semiparametric estimation method gave high accurate estimates of β when the proper tuning parameters are chosen and the regions S_l , $l = 1, 2, 3, 4$ are known. The bias and the standard deviation (SD) of the estimators of β decrease with increasing time window length for almost all tuning parameter settings. While for large tuning parameter c_2 , the bias and SD of the

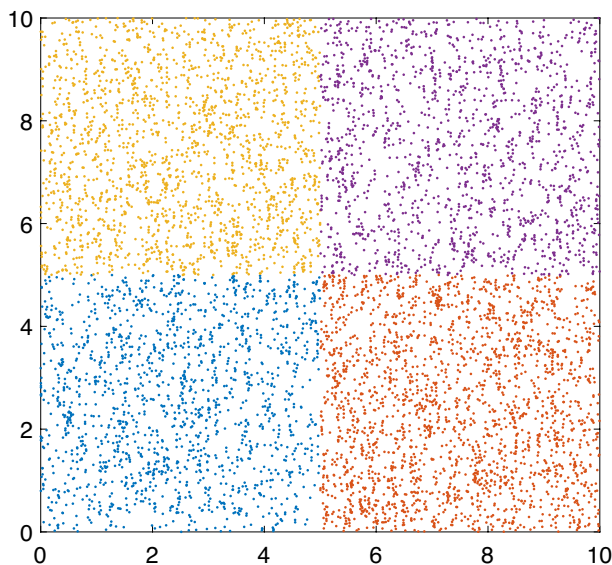


Fig. 4 Spatial distribution map of one realization of the spatiotemporal Hawkes process during the observation time $[0, 100]$

estimators of β are large. From Table 3, we can also see that the bias and standard deviation of the parameter estimates are small for observation interval D_3 when the tuning parameters $c_1 = 1$ and $c_2 = 0.01$. By comparing the results with the optimal tuning parameters obtained in Sect. 4.1, it can be seen that the tuning parameters differ from case to case.

In the following, we use the FLP, AIC, and BIC to evaluate the effect of the selection of tuning parameters. Figure 5 shows the 95% confidence limits and the mean values of the FLP, AIC, and BIC with different tuning parameters for observation interval $(180, 200]$, $[0, 180]$, and $[0, 180]$, respectively. From Fig. 5, it can be seen that the maximum FLP value is attained when $c_1 = 1$ and $c_2 \in \{0.01, 0.02, 0.05\}$, and both the minimum AIC value and BIC value are attained when $c_1 = 1$ and $c_2 \in \{0.01, 0.02, 0.05\}$. That means the FLP, AIC, and BIC show similar results for selecting the tuning parameters. Also, Table 3 shows that all three methods efficiently select the optimal tuning parameters.

Figure 6 shows the 95% confidence limits, the mean values of the estimates of the transfer functions, and the estimated transfer functions from one simulation for observation interval $[0, 180]$, tuning parameters $c_1 = 1$ and $c_2 = 0.01$. From Fig. 6, it can be seen that the estimated transfer functions from one simulation have a periodic trend, which may be due to the spatial superposition of spatiotemporal self-exciting functions (see the end of Sect. 4.3). Notice that the estimated transfer functions and branching matrix are too different from the spatiotemporal self-exciting functions to be comparable. We do not show the estimation of the branching matrix anymore. Figure 7 shows the QQ plot of the residuals of the four different dimensions to assess the goodness of fit of the considered conditional intensity function. The QQ plot in Fig. 7 shows that the estimated multivariate Hawkes process is close to the diagonal line. The bottom four QQ plots are plotted by 300 transformed times randomly sampled from all transformed times for visual effect. Table 4 shows the Chi-square test p values between 1000 random samples of the residuals of the different dimensions from one simulation for observation interval $[0, 180]$, tuning parameters $c_1 = 1$ and $c_2 = 0.01$. It can be seen that the transformed times for the different dimensions are statistically independent of each other when the significance level takes 0.05.

4.3 Simulation: spatiotemporal cases with unknown partitions

In this section, the artificial data generated from above conditional intensity function (14) of spatiotemporal Hawkes process are used to verify the performance of the proposed semiparametric estimation method. We assume the regions S_l , $l = 1, 2, 3, 4$ are unknown. The observations are divided into d areas 100 times repeatedly by the K-means method with random initial centers, where d is selected from $\{3, 4, 5, 6, 7\}$. Since there are three variables in parameter vector β , the minimum number of region divisions to ensure the invertibility of the matrix XX' is 3. Figure 8 shows the partition of the observations when d takes different values, with the observations within the partitioned regions indicated by different colors. From Fig. 8, we can see that the K-means method can cluster according to different background intensity function

Table 3 Estimates of β with simulated data with observation intervals $D_1 = [0, 90]$, $D_2 = [0, 135]$ and $D_3 = [0, 180]$, using the proposed semiparametric estimation method with different orders p and time windows h . Spatiotemporal cases with known partitions

c_1	c_2	$\beta_0 = -4$						$\beta_1 = 2$			$\beta_2 = 1$		
		D_1		D_2		D_3		D_1		D_2		D_1	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
1	0.01	0.0193	0.0151	0.0151	-0.0272	-0.0189	-0.0071	0.0182	0.0182	-0.0241	-0.0171	0.0218	0.0218
		1.1962	0.3520	0.3520	0.1994	0.1918	0.1918	0.1196	0.1196	0.5726	0.1577	0.2114	0.2114
	0.02	-0.1935	-0.2396	-0.2396	-0.1945	0.0072	0.0169	0.0084	0.0084	-0.0073	0.0307	0.0083	0.0083
		0.4001	0.2843	0.2843	0.3934	0.2613	0.1530	0.1956	0.1956	0.2812	0.1711	0.2894	0.2894
	0.05	-0.6206	-0.6728	-0.6728	-0.6389	0.0184	0.0557	0.0255	0.0255	0.0204	0.0189	-0.0129	-0.0129
		0.5734	0.4296	0.4296	0.4487	0.3251	0.2738	0.2401	0.2401	0.3443	0.3902	0.2260	0.2260
	0.1	-0.9451	-1.1382	-1.1382	-1.0582	0.0006	0.0512	0.0326	0.0326	-0.0411	-0.0339	0.0020	0.0020
		1.1585	0.7375	0.7375	0.5616	0.6285	0.3941	0.3022	0.3022	0.5750	0.3796	0.3173	0.3173
	0.2	-1.4132	-1.5814	-1.5814	-1.3411	-0.1670	-0.0154	0.0313	0.0313	0.2160	0.1143	-0.3654	-0.3654
		1.8217	1.0552	1.0552	2.0258	1.5407	0.6998	0.5637	0.5637	1.6546	1.0650	2.6506	2.6506
2	0.01	-0.2889	-0.2225	-0.2225	-0.2234	0.0161	0.0201	0.0274	0.0274	0.0743	0.0031	0.0260	0.0260
		0.6327	0.4684	0.4684	0.2633	0.3678	0.2474	0.1577	0.1577	0.3186	0.2476	0.1562	0.1562
	0.02	-0.7097	-0.6348	-0.6348	-0.5585	0.0949	0.0966	0.0525	0.0525	0.0840	0.0117	0.0135	0.0135
		0.4577	0.4669	0.4669	0.4226	0.2561	0.2404	0.2318	0.2318	0.2292	0.2690	0.2437	0.2437
	0.05	-1.0744	-1.1875	-1.1875	-1.1513	0.0019	0.1217	0.0867	0.0867	0.1341	-0.0469	0.0143	0.0143
		0.7244	0.6808	0.6808	0.5120	0.4879	0.3909	0.2481	0.2481	0.5447	0.4492	0.4062	0.4062
	0.1	-1.7518	-1.5275	-1.5275	-1.8016	0.0626	-0.0223	0.0918	0.0918	0.0568	0.0369	0.1369	0.1369
		1.1506	0.9642	0.9642	0.9152	0.6163	0.5319	0.4969	0.4969	0.5537	0.6232	0.5561	0.5561
	0.2	-2.3044	-2.6724	-2.6724	-2.0174	0.0066	0.2747	-0.0661	-0.0661	0.0652	0.1655	-0.0529	-0.0529
		2.1365	1.8958	1.8958	3.3934	1.1007	1.1769	1.4989	1.4989	1.0676	0.7571	1.3219	1.3219
3	0.01	-0.4243	-0.6017	-0.6017	-0.4470	0.1017	0.1371	0.0504	0.0504	-0.0653	0.0424	0.0164	0.0164
		0.8575	0.5264	0.5264	0.4398	0.2623	0.2668	0.2350	0.2350	1.0342	0.2916	0.1923	0.1923
	0.02	-0.9998	-1.0055	-1.0055	-0.8101	0.1457	0.1401	0.0479	0.0479	0.0862	0.2044	0.0038	0.0038
		0.6550	0.7097	0.7097	0.4827	0.3573	0.3460	0.2946	0.2946	0.2513	1.3125	0.2642	0.2642

Table 3 (continued)

c_1	c_2	$\beta_0 = -4$			$\beta_1 = 2$			$\beta_2 = 1$		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
4	0.05	Bias	-1.5740	-1.8773	-1.4954	0.1316	0.2990	0.1043	0.0253	0.1484
		SD	1.1323	1.8058	0.5758	0.7798	0.9835	0.3080	1.2765	0.6645
	0.1	Bias	-2.3145	-2.2398	-1.4146	0.3830	0.1306	-0.1414	-0.2640	0.1510
		SD	2.4268	1.5993	9.5268	1.4983	0.9873	4.4029	3.0030	0.9413
	0.2	Bias	-3.3106	-2.9585	-2.6937	0.2333	0.1521	-0.0854	0.3692	-0.0881
		SD	4.5744	4.0183	4.3537	2.6402	2.5797	3.0351	2.1002	1.8919
	0.01	Bias	-0.8527	-0.8453	-0.6501	0.1788	0.1682	0.0854	0.0923	0.0918
		SD	0.4797	0.7485	0.5209	0.2801	0.3730	0.2509	0.3410	0.4759
	0.02	Bias	-1.5731	-1.1474	-1.0764	0.3264	0.1662	0.1005	0.2348	0.0715
		SD	2.2973	0.6083	0.5230	1.1210	0.3255	0.3169	0.7694	0.3827
	0.05	Bias	-2.1070	-1.9461	-1.8681	0.2986	0.1114	0.1545	0.0692	0.2622
		SD	1.2252	0.8998	0.9681	0.7257	0.5071	0.5597	0.6827	1.1982
5	0.1	Bias	-2.7183	-2.7365	-2.6218	0.2707	0.2356	0.2154	-0.0051	0.1705
		SD	4.9786	1.8579	1.4239	2.5716	1.2255	0.7505	1.9353	1.2116
	0.2	Bias	-5.0795	-4.1996	-3.1586	0.9709	0.7399	0.0631	0.5704	-0.2581
		SD	6.3618	4.6370	2.0103	3.3827	2.5661	1.1323	4.4090	4.3001
	0.01	Bias	-0.9958	-0.9654	-0.8949	0.1995	0.1560	0.1305	0.0521	0.0687
		SD	0.5703	0.4887	0.3681	0.3492	0.2460	0.1736	0.3225	0.3868
	0.02	Bias	-1.7809	-1.4990	-1.3345	0.3753	0.2481	0.1189	0.1816	0.1388
		SD	1.3247	0.6446	0.6575	0.8263	0.3788	0.3294	0.8802	0.3969
	0.05	Bias	-2.0288	-1.9302	-1.6911	0.1380	0.0510	-0.0152	0.0271	0.0729
		SD	1.3071	1.5036	1.2318	0.7966	0.9455	0.6330	0.9306	0.6268
	0.1	Bias	-3.2543	-2.8814	-2.7982	0.3634	0.1469	0.1137	0.2529	0.3215
		SD	2.9700	3.6087	2.1052	1.7334	2.0341	1.3756	1.7103	1.4824

Table 3 (continued)

c_1	c_2	$\beta_0 = -4$			$\beta_1 = 2$			$\beta_2 = 1$		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
0.2	Bias	-2.1964	-3.9885	-3.1137	-0.7318	0.3677	-0.1431	-0.2820	-0.0773	0.0265
	SD	13.3864	5.9204	4.4479	5.1399	3.4097	2.6947	8.5954	3.5473	2.3836

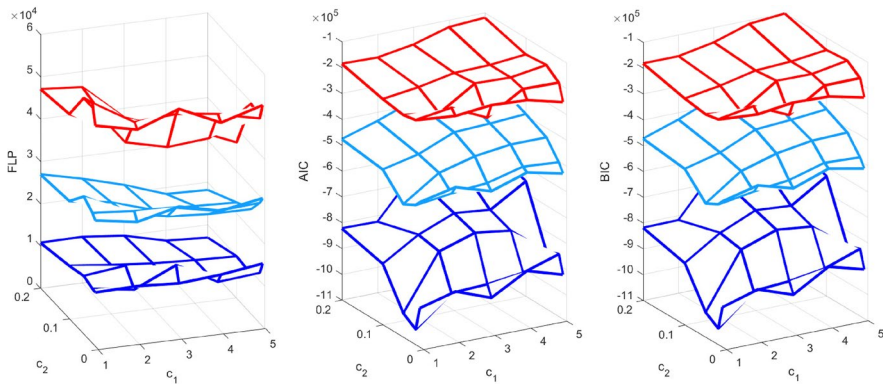


Fig. 5 FLP, AIC, and BIC values with different tuning parameter values for the observation intervals (180, 200], [0, 180], and [0, 180], respectively: mean (light blue solid), 0.95 quantile (top red solid), 0.05 quantile (bottom dark blue solid)

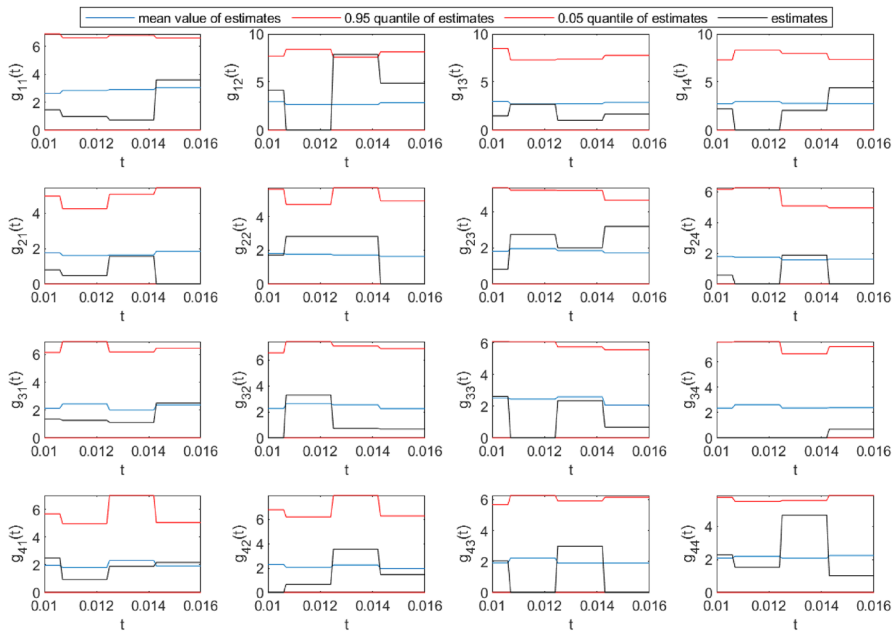


Fig. 6 Estimation of the transfer functions of the multivariate Hawkes process for simulated data: mean semiparametric estimate (blue solid), 0.95 quantile of semiparametric estimate (top red solid), 0.05 quantile of semiparametric estimate (bottom red solid), and estimated transfer functions from one simulation (black solid)

values. In particular, the results of the K-means method at $d = 4$ are consistent with the actual spatial partition.

As in Sect. 4.2, we let $p = c_1 T^{1/2} (\log(T))^{-1/10}$ and $h = c_2 T^{-3/8}$. With each combination of observation interval D , order p , time windows h , and partitions,

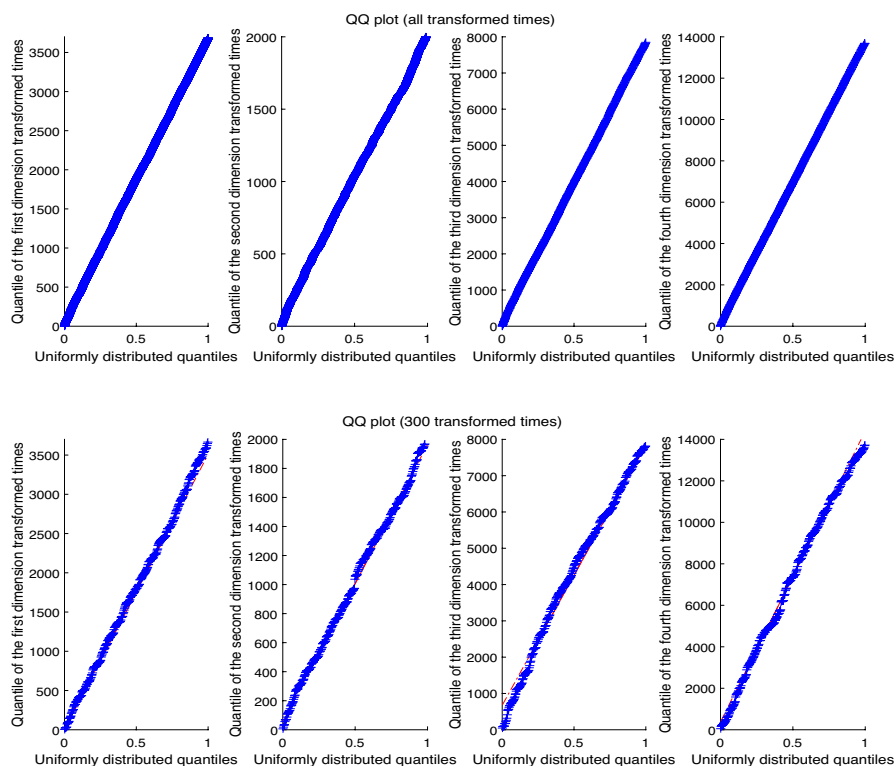


Fig. 7 Quantile-quantile plot of the residuals of the three different dimensions: all transformed times (top) and 300 transformed times randomly sampled from all transformed times (bottom)

Table 4 p value of Chi-square test for different regions

	$\chi^2_{1,2}$	$\chi^2_{1,3}$	$\chi^2_{1,4}$	$\chi^2_{2,3}$	$\chi^2_{2,4}$	$\chi^2_{3,4}$
p value	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000

the semiparametric estimation method was repeated 100 times, where the partitions were reacquired for each experiment using the K-means method. Tables 5–7 show summaries of the estimates of the parameters β with several different orders p , time windows h , and three observation intervals $[0, 90]$, $[0, 135]$, and $[0, 180]$ for d takes 4, 3, and 5, respectively. Observations in the intervals $(90, 100]$, $(135, 150]$, and $(180, 200]$ are used to calculate the FLPs. From the results of Table 5, we can see that the semiparametric estimation method gave acceptable accuracy for estimates of β under almost all tuning parameters. Specifically, the bias and the standard deviations of the estimators of β decrease with increasing time window length for some tuning parameters, such as $c_1 = 3$ and $c_2 = 0.02$, while for some tuning parameters, the bias and standard deviations of

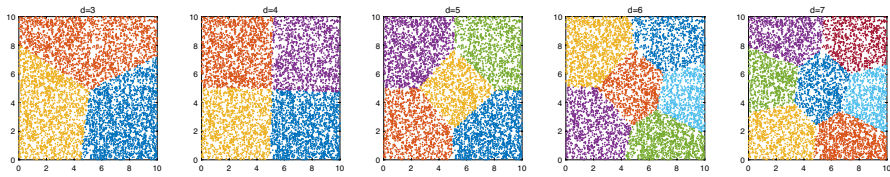


Fig. 8 Spatial distribution maps of one realization of the spatiotemporal Hawkes process during the observation time $[0, 100]$

the estimators of β are a little large. From Table 6, we can see that the standard deviations of the parameter estimates are significantly large for all observation intervals and tuning parameter settings. For Table 7, it can be seen that the biases of the parameter estimates are a little large for all observation intervals and tuning parameter settings.

In the following, we use the FLP, AIC, and BIC to evaluate the effect of the selection of tuning parameters. Figure 9 shows the 95% confidence limits and the mean values of the FLP, AIC, and BIC with different tuning parameters for observation interval $(180, 200]$, $[0, 180]$, and $[0, 180]$, respectively. From Fig. 9, we can see that the FLP, AIC, and BIC show similar results for selecting the tuning parameters. As can be seen from the black solid in Fig. 9, the maximum FLP value and both the minimum AIC value and BIC value are attained when $d = 4$, $c_1 = 2$ and $c_2 = 0.01$. In fact, in almost all simulations, the optimal d obtained based on FLP, AIC, and BIC takes values of 3, 4, or 5, the optimal tuning parameter c_1 takes values of 1, 2, or 3, and c_2 takes values of 0.01 or 0.02.

Figure 10 shows the 95% confidence limits, the mean values of the estimates of the transfer functions, and the estimated transfer functions from one simulation for observation interval $[0, 180]$, tuning parameters $c_1 = 3$ and $c_2 = 0.02$. From Fig. 10, it can be seen that the estimated transfer functions from one simulation have a strong periodic trend. The QQ plot and Chi-square test results are similar to those in Sect. 4.2, and the graphical presentation of the relevant results is omitted here. In particular, the QQ plot shows that the transformed times for all dimensions obey the Poisson processes with unit intensity, and the Chi-square test shows that the transformed times for the different dimensions are statistically independent of each other when the significance level takes 0.05.

The above results illustrate that for spatiotemporal data generated by the spatiotemporal Hawkes process, the multivariate Hawkes process can efficiently estimate the coefficients of the spatial covariates. Indeed, for intensity function (14), we have

Table 5 Estimates of β with simulated data with observation intervals $D_1 = [0, 90]$, $D_2 = [0, 135]$ and $D_3 = [0, 180]$, using the proposed estimation method with different orders p and time windows h . Spatiotemporal cases with unknown partitions. Number of partitions is 4

c_1	c_2	$\beta_0 = -4$									$\beta_1 = 2$									$\beta_2 = 1$								
		$\beta_0 = -4$			$\beta_1 = 2$			$\beta_1 = 2$			$\beta_2 = 1$			$\beta_2 = 1$			$\beta_2 = 1$			$\beta_2 = 1$			$\beta_2 = 1$			$\beta_2 = 1$		
		D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
1	0.01	Bias	0.6391	0.7125	0.6690	0.7125	0.6690	-0.2960	-0.4033	-0.3078	-0.1162	0.0541	-0.1503	-0.1162	0.0541	-0.1503	-0.1162	0.0541	-0.1503	-0.1162	0.0541	-0.1503	-0.1162	0.0541	-0.1503	-0.1162	0.0541	-0.1503
		SD	1.1807	1.3213	1.3037	1.3213	1.3037	0.6893	0.8456	2.1404	0.8857	0.9440	1.6902	0.8857	0.9440	1.6902	0.8857	0.9440	1.6902	0.8857	0.9440	1.6902	0.8857	0.9440	1.6902	0.8857	0.9440	1.6902
	0.02	Bias	1.0753	0.6344	0.4765	0.6344	0.4765	-1.3028	-0.3471	-0.3018	1.3675	-0.2449	-0.0470	1.3675	-0.2449	-0.0470	1.3675	-0.2449	-0.0470	1.3675	-0.2449	-0.0470	1.3675	-0.2449	-0.0470	1.3675	-0.2449	-0.0470
		SD	5.8794	2.1404	0.7318	2.1404	0.7318	8.2164	1.3295	1.2344	11.8998	1.2413	1.2321	11.8998	1.2413	1.2321	11.8998	1.2413	1.2321	11.8998	1.2413	1.2321	11.8998	1.2413	1.2321	11.8998	1.2413	1.2321
	0.05	Bias	-0.0552	0.5456	0.3066	0.5456	0.3066	-0.1614	-0.4663	-0.3725	-0.331	-0.4030	-0.2096	-0.331	-0.4030	-0.2096	-0.331	-0.4030	-0.2096	-0.331	-0.4030	-0.2096	-0.331	-0.4030	-0.2096	-0.331	-0.4030	-0.2096
		SD	1.918	1.7108	0.8188	1.7108	0.8188	1.3371	0.9293	1.8643	2.0426	1.4891	1.9303	2.0426	1.4891	1.9303	2.0426	1.4891	1.9303	2.0426	1.4891	1.9303	2.0426	1.4891	1.9303	2.0426	1.4891	1.9303
2	0.01	Bias	0.6658	0.8697	0.8264	0.8697	0.8264	-0.4777	-0.5369	-0.4640	-0.0187	-0.0625	-0.2287	-0.0187	-0.0625	-0.2287	-0.0187	-0.0625	-0.2287	-0.0187	-0.0625	-0.2287	-0.0187	-0.0625	-0.2287	-0.0187	-0.0625	-0.2287
		SD	1.8724	2.5439	1.5694	2.5439	1.5694	1.3611	1.6120	1.0687	1.5922	1.9133	1.6900	1.5922	1.9133	1.6900	1.5922	1.9133	1.6900	1.5922	1.9133	1.6900	1.5922	1.9133	1.6900	1.5922	1.9133	1.6900
	0.02	Bias	0.3969	0.3736	0.0326	0.3736	0.0326	-0.4025	-0.4065	-0.1977	-0.1764	-0.1150	-0.1720	-0.1764	-0.1150	-0.1720	-0.1764	-0.1150	-0.1720	-0.1764	-0.1150	-0.1720	-0.1764	-0.1150	-0.1720	-0.1764	-0.1150	-0.1720
		SD	1.4963	1.2302	2.2039	1.2302	2.2039	0.8995	0.7822	2.1026	1.1125	1.0890	2.4156	1.1125	1.0890	2.4156	1.1125	1.0890	2.4156	1.1125	1.0890	2.4156	1.1125	1.0890	2.4156	1.1125	1.0890	2.4156
	0.05	Bias	-0.0230	-0.1591	-0.5146	-0.1591	-0.5146	-0.4362	-0.4991	-0.1098	-0.3008	-0.2807	-0.1720	-0.3008	-0.2807	-0.1720	-0.3008	-0.2807	-0.1720	-0.3008	-0.2807	-0.1720	-0.3008	-0.2807	-0.1720	-0.3008	-0.2807	-0.1720
		SD	2.0801	1.8051	2.8069	1.8051	2.8069	1.3194	1.2156	1.9856	1.1752	1.4764	1.4709	1.1752	1.4764	1.4709	1.1752	1.4764	1.4709	1.1752	1.4764	1.4709	1.1752	1.4764	1.4709	1.1752	1.4764	1.4709
3	0.01	Bias	0.4027	0.6439	0.3263	0.6439	0.3263	-0.3488	-0.5801	-0.3176	-0.1138	0.1522	-0.0530	-0.1138	0.1522	-0.0530	-0.1138	0.1522	-0.0530	-0.1138	0.1522	-0.0530	-0.1138	0.1522	-0.0530	-0.1138	0.1522	-0.0530
		SD	1.3553	1.5885	2.1077	1.5885	2.1077	0.7800	1.0851	1.6433	0.9482	1.2095	1.9250	0.9482	1.2095	1.9250	0.9482	1.2095	1.9250	0.9482	1.2095	1.9250	0.9482	1.2095	1.9250	0.9482	1.2095	1.9250
	0.02	Bias	-0.1368	-0.0819	-0.0351	-0.0819	-0.0351	-0.1585	-0.2169	-0.2706	-0.2902	-0.1984	-0.2396	-0.2902	-0.1984	-0.2396	-0.2902	-0.1984	-0.2396	-0.2902	-0.1984	-0.2396	-0.2902	-0.1984	-0.2396	-0.2902	-0.1984	-0.2396
		SD	1.8114	1.4307	1.2379	1.4307	1.2379	1.2645	1.1840	0.7124	1.7774	1.5026	0.8495	1.7774	1.5026	0.8495	1.7774	1.5026	0.8495	1.7774	1.5026	0.8495	1.7774	1.5026	0.8495	1.7774	1.5026	0.8495
	0.05	Bias	-0.8714	-0.6883	-0.5287	-0.6883	-0.5287	-0.3252	-0.2092	-0.3774	0.1994	-0.5945	-0.0762	0.1994	-0.5945	-0.0762	0.1994	-0.5945	-0.0762	0.1994	-0.5945	-0.0762	0.1994	-0.5945	-0.0762	0.1994	-0.5945	-0.0762
		SD	1.4956	3.3466	1.4808	3.3466	1.4808	1.0723	1.5523	1.0552	1.8011	5.4157	1.5278	1.8011	5.4157	1.5278	1.8011	5.4157	1.5278	1.8011	5.4157	1.5278	1.8011	5.4157	1.5278	1.8011	5.4157	1.5278

Table 6 Estimates of β with simulated data with observation intervals $D_1 = [0, 90]$, $D_2 = [0, 135]$ and $D_3 = [0, 180]$, using the proposed estimation method with different orders p and time windows h . Spatiotemporal cases with unknown partitions. Number of partitions is 3

c_1	c_2		$\beta_0 = -4$			$\beta_1 = 2$			$\beta_2 = 1$		
			D_1	D_2	D_3	D_1	D_2	D_3	D_1	D_2	D_3
1	0.01	Bias	12.7990	0.5695	0.6650	-6.9798	-0.5731	-0.3535	-0.3200	0.1330	-0.2402
		SD	117.6769	23.3671	4.4292	58.0839	14.6667	2.1128	19.5524	6.9236	1.9653
	0.02	Bias	0.7129	-0.3231	-1.6253	-0.5607	-0.0237	0.6140	-0.3256	0.3600	0.1406
		SD	7.2693	7.1626	12.6461	2.7915	3.9180	5.8994	5.6061	3.7931	3.2998
	0.05	Bias	1.8124	-2.2049	-1.1143	-1.0510	1.4004	0.1548	-0.5660	-2.1495	-0.1932
		SD	23.4631	19.6074	14.5444	10.1878	12.6171	7.6071	6.2878	18.8181	5.8402
2	0.01	Bias	-0.8584	-0.8517	1.2595	0.1590	0.3701	-0.9565	0.4486	0.0865	0.5105
		SD	15.4762	10.7339	9.7564	6.2041	5.1752	4.6821	9.7659	5.1671	5.4134
	0.02	Bias	3.9910	-0.1708	-2.1515	-2.7212	-0.2316	1.0207	0.1384	-0.1644	-0.8063
		SD	23.4655	2.9601	28.9618	15.2788	1.4949	21.5198	4.7979	1.3155	19.5937
	0.05	Bias	-0.6356	-2.7937	-2.5543	-0.0929	0.7655	0.7204	-0.5146	1.0564	-0.4130
		SD	4.1138	25.1119	25.4107	2.4649	13.1342	12.8926	1.7071	8.0716	2.5811
3	0.01	Bias	4.5401	-1.1638	0.7833	-2.6552	0.5151	-0.3853	-0.4736	-0.4469	-0.6043
		SD	34.1613	6.7007	7.0931	17.1199	4.3397	2.5008	8.3045	5.5081	8.6183
	0.02	Bias	10.0942	-0.8258	3.8059	-21.8244	-0.5021	-2.3052	36.6507	1.3404	0.2136
		SD	118.0386	11.5483	34.4486	218.2781	6.7749	17.7127	354.1629	8.8653	8.5422
	0.05	Bias	-1.2015	-4.1383	6.3161	-0.1047	1.7397	-4.8226	0.1199	-1.1606	2.4442
		SD	7.3531	33.3903	98.0239	3.6002	18.2915	54.1264	3.5855	5.3280	14.7658

Table 7 Estimates of β with simulated data with observation intervals $D_1 = [0, 90]$, $D_2 = [0, 135]$ and $D_3 = [0, 180]$, using the proposed estimation method with different orders p and time windows h . Spatiotemporal cases with unknown partitions. Number of partitions is 5

c_1	c_2	$\beta_0 = -4$						$\beta_1 = 2$						$\beta_2 = 1$					
				D_1		D_2		D_3		D_1		D_2		D_3		D_1		D_2	
				Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
1	0.01	Bias		2.6933	2.6314	2.4240		-1.4268		-1.4709		-1.2751		-0.5024		-0.3875		-0.4832	
		SD		1.0730	0.9367	1.0402		0.7065		0.6922		0.7282		1.2453		1.0083		1.1243	
	0.02	Bias		2.4677	2.3257	2.4138		-1.3926		-1.2674		-1.3992		-0.6606		-0.6462		-0.4195	
		SD		1.0280	1.0537	0.9703		1.0138		0.7343		0.6546		1.6489		1.4230		1.0826	
	0.05	Bias		1.8461	1.9743	2.1130		-1.3139		-1.3797		-1.4558		-0.3791		-0.4237		-0.4466	
		SD		0.8985	1.1078	1.0219		0.6569		0.9953		0.8621		1.0932		1.5021		1.2321	
2	0.01	Bias		2.4569	2.1871	2.5759		-1.4417		-1.3463		-1.5021		-0.5467		-0.2795		-0.4623	
		SD		1.1576	0.9410	1.1405		0.8013		0.7382		0.8584		0.9823		1.4349		1.1872	
	0.02	Bias		2.1640	1.9783	2.1294		-1.4537		-1.3438		-1.4367		-0.3328		-0.4031		-0.3447	
		SD		0.9878	0.9305	0.9806		0.6342		0.6403		0.6676		1.0337		0.9881		1.3522	
	0.05	Bias		1.6747	1.5428	1.4369		-1.3970		-1.3196		-1.3421		-0.6218		-0.5640		-0.3349	
		SD		1.1796	0.8833	1.4191		1.0283		0.7188		1.1304		1.8668		1.4831		1.7065	
3	0.01	Bias		1.9296	2.0383	1.9767		-1.2483		-1.2603		-1.2695		-0.3920		-0.6525		-0.3442	
		SD		0.8517	0.9798	0.8685		0.6717		0.7940		0.6894		1.0042		2.1750		1.1786	
	0.02	Bias		1.7639	1.8294	1.7743		-1.2782		-1.3250		-1.3490		-0.5948		-0.6121		-0.4020	
		SD		1.1514	1.0215	0.8880		0.7802		0.6228		0.6278		1.3725		1.2238		1.0576	
	0.05	Bias		1.2443	0.8771	1.2339		-1.4441		-1.1521		-1.4158		-0.2929		-0.5343		-0.3174	
		SD		1.0960	0.9582	1.1197		0.8556		0.6464		0.8801		1.5700		1.2814		1.3700	

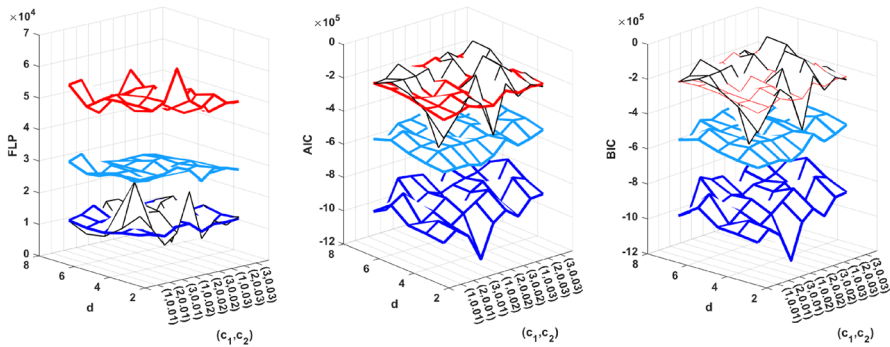


Fig. 9 FLP, AIC, and BIC values with different tuning parameter values for the observation intervals (180, 200], [0, 180], and [0, 180], respectively: mean (light blue solid), 0.95 quantile (top red solid), 0.05 quantile (bottom dark blue solid), and results from one simulation (black solid)

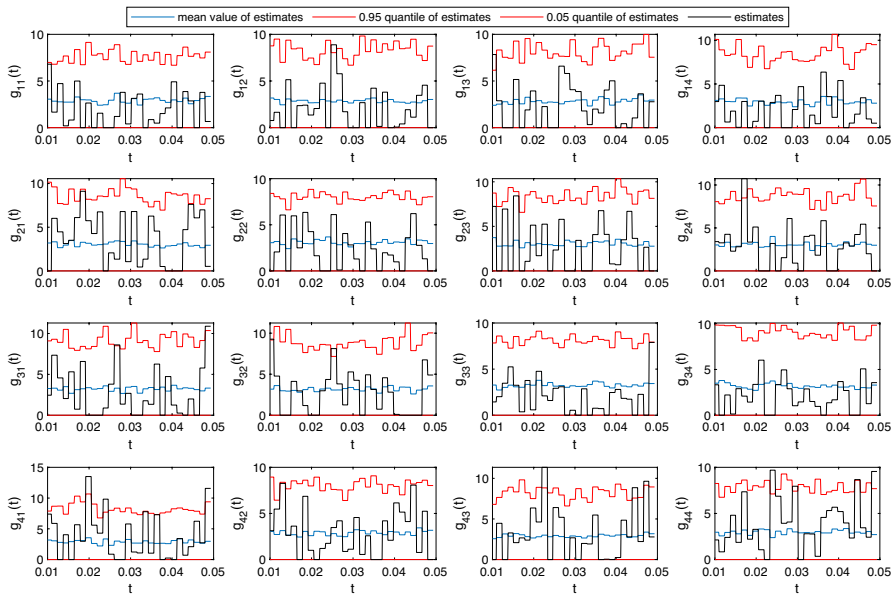


Fig. 10 Estimation of the transfer functions of the multivariate Hawkes process for simulated data: mean semiparametric estimate (blue solid), 0.95 quantile of semiparametric estimate (top red solid), 0.05 quantile of semiparametric estimate (bottom red solid), and estimated transfer functions from one simulation (black solid)

$$\begin{aligned}
\lambda_i(t|\mathcal{H}_t) &= \int \int_{S_i} \lambda(x, y, t|\mathcal{H}_t) dx dy \\
&= \int \int_{S_i} \exp\{\beta s(x, y)\} dx dy + \int \int_{S_i} \int_{S \times (-\infty, t)} g(x-u, y-v, t-w) N(dw, dv, du) dx dy \\
&= \int \int_{S_i} \exp\{\beta s(x, y)\} dx dy + \sum_{t_l \leq t, (x_l, y_l) \in S} \int \int_{S_i} g(x-x_l, y-y_l, t-t_l) dx dy \\
&= \int \int_{S_i} \exp\{\beta s(x, y)\} dx dy + \sum_{j=1}^d \sum_{t_l \leq t, (x_l, y_l) \in S_j} \int \int_{S_i} g(x-x_l, y-y_l, t-t_l) dx dy \\
&=: \int \int_{S_i} \exp\{\beta s(x, y)\} dx dy + \sum_{j=1}^d \sum_{t_l \leq t, (x_l, y_l) \in S_j} \tilde{g}_{ij}(t-t_l) \\
&= \exp\{\beta s(x, y) + \log |S_i|\} + \sum_{j=1}^d \int_{(-\infty, t)} \tilde{g}_{ij}(t-w) N_j(dw),
\end{aligned}$$

where the last equation holds when $S_i, i = 1, \dots, d$ are the true partitions. Note that the function $\tilde{g}_{ij}(t)$ is a spatial superposition of spatiotemporal self-exciting functions of intensity function (14), which may be the main reason for the periodicity of the estimated transfer functions. In particular, when the partition is unknown, the transfer functions exhibit stronger periodicity compared to when the partition is known.

4.4 Application

4.4.1 Pittsburgh burglary data and spatial covariates

We use the proposed model and semiparametric estimation method to fit the burglary data compiled by the Pittsburgh Bureau of Police between January 1, 2005, and September 9, 2015. The data include occurrence timestamp, type of each incident, city block, longitude, latitude, etc. The catalog is publicly available at <https://data.wprdc.org/dataset/uniform-crime-reporting-data>, which consists of several groups such as robbery, aggravated assault, burglary, and motor vehicle theft. The total data size is 495251. The sample size of burglary data is 31671. Almost all events of the catalog occurred in a rectangular area between longitudes -80.10° and -79.85° and latitudes 40.30° and 40.50° . By deleting data outside the above area and missing latitude and longitude records, the size of the rest burglary data is 31662. The burglaries are mapped in Fig. 11. In Fig. 11, the burglaries are divided into five areas-based K-means method, and burglaries in different areas are represented by different colors.

The spatial covariates for each census block used below are collect by Reinhart and Greenhouse (2017) from city and Census Bureau data, including (a) population density (per square meter), (b) fraction of residents who are male from ages 18–24 years, (c) fraction of residents who are black, and (d) fraction of homes that are occupied by their owners, rather than rented, etc. Using the above data, we calculated spatial covariates within the five areas divided based on the burglaries. The results are shown in Table 8. Below we built 5-dimension multivariate Hawkes

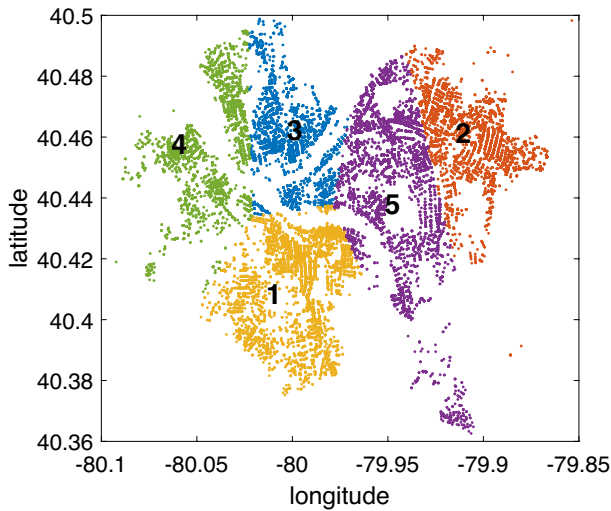


Fig. 11 Spatial distribution maps of burglaries with longitude and latitude coordinates recorded in Pittsburgh between January 1, 2005, and September 9, 2015

Table 8 Description of additional covariates

Covariate	x_1	x_2	x_3	x_4	x_5
constant (intercept)	1	1	1	1	1
population per m^2	0.0017	0.0021	0.0022	0.0027	0.0015
fraction male, 18-24 years	0.0657	0.0467	0.0781	0.1431	0.0521
fraction black	0.4096	0.4439	0.1571	0.1902	0.3075
fraction homes owned	0.4217	0.5117	0.5823	0.4020	0.5927

processes with four spatial covariates to analyze the burglary data and the corresponding spatial covariates.

4.4.2 Burglary data analysis

Burglary data are recorded in minutes, and in the following analyses, we have pre-processed it to be in days. Time window depends on the unit of the real data and should be chosen carefully. When the time window is too small, no event occurs in most intervals, making the estimated results deviate greatly. However, a large time window can lead to a large approximation error, which makes it difficult to accurately identify the mutually exciting pattern. We use the FLP, AIC, and BIC to select the tuning parameters. Figure 12 shows the values of the FLP, AIC, and BIC with different tuning parameters. From Fig. 12, we can see that the FLP, AIC, and BIC show similar results for selecting the tuning parameters. The maximum FLP

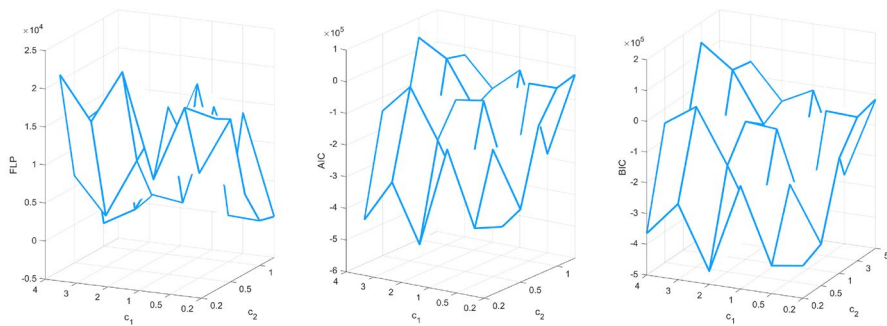


Fig. 12 FLP, AIC, and BIC values with different tuning parameters

value and both the minimum AIC value and BIC value are attained when $c_1 = 2$ and $c_2 = 0.2$. Table 9 shows a summary of the estimates of the parameter β and $\exp(\beta X)$.

The estimate of branching matrix $\hat{Q}$ is shown in the following matrix.

$$\hat{Q} : \begin{bmatrix} \mathbf{0.1659} & \mathbf{0.1094} & \mathbf{0.1190} & \mathbf{0.1208} & \mathbf{0.0928} \\ \mathbf{0.0847} & \mathbf{0.1317} & \mathbf{0.0778} & \mathbf{0.1151} & \mathbf{0.0980} \\ \mathbf{0.0821} & \mathbf{0.0883} & \mathbf{0.1097} & \mathbf{0.1002} & \mathbf{0.0657} \\ \mathbf{0.0631} & \mathbf{0.0605} & \mathbf{0.0609} & \mathbf{0.1062} & \mathbf{0.0597} \\ \mathbf{0.1280} & \mathbf{0.1168} & \mathbf{0.1242} & \mathbf{0.1282} & \mathbf{0.1921} \end{bmatrix}.$$

The spectral radius of $\hat{Q}$ is 0.5167. We can see that 14 branching coefficients marked in bold show significant values, meaning that burglaries occurring in the corresponding areas are mutually exciting. Meanwhile, 5 branching coefficients marked in italics (greater than 0.08 but less than 0.1) showed weak mutually exciting. We also note that 6 branching coefficients (less than 0.08) are close to zeros, which implies that burglaries occurring in these areas have no mutually exciting pattern. The mutually exciting relationship diagram of each region is shown in Fig. 13, where the wider arrow indicates the greater strength of mutual excitation. In Fig. 13, the bold arrows in the left panel indicate strong inter-regional ‘mutually exciting’ effect, the bold numbers indicate strong ‘mutually exciting’ effect within the region, and the italics number indicates weak ‘mutually exciting’ effect within the region; the pink arrows in the right panel indicate normal ‘mutually exciting’ effect between regions (the estimated branching coefficients are greater than 0.09 and less than 0.1),

Table 9 Predicting burglary using additional covariates

Parameter	Estimated value	Parameter	Estimated value
β_0	0.3839	$\exp(\beta x_1)$	0.9297
β_1	556.7690	$\exp(\beta x_2)$	0.7181
β_2	-6.9273	$\exp(\beta x_3)$	0.6527
β_3	-1.7896	$\exp(\beta x_4)$	0.5575
β_4	-1.4854	$\exp(\beta x_5)$	0.9464

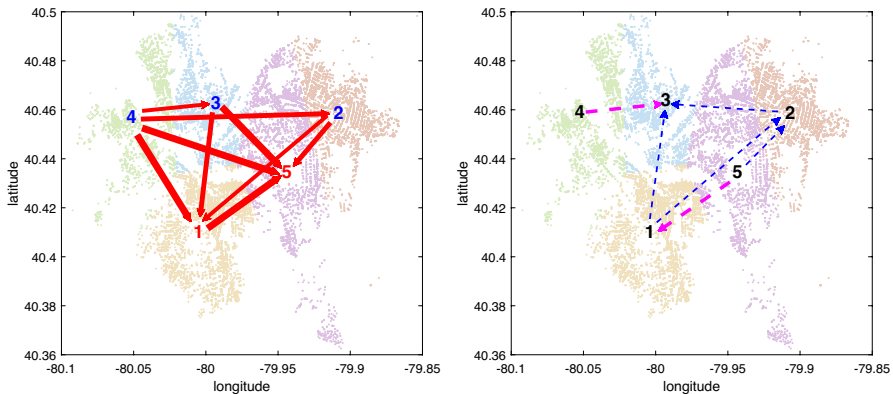


Fig. 13 Mutually exciting relationship diagram of each region

the italics arrows indicate weak ‘mutually exciting’ effect between regions (the estimated branching coefficients are greater than 0.8 and less than 0.9). From Fig. 13, it can be seen that the strong and normal ‘mutually exciting’ effects occur in the adjacent regions, and the weak adjacent regions ‘mutually exciting’ effects are found in the regions farther apart. Specifically, regions 1 and 5 are subject to strong ‘mutually exciting’ effects from other regions, and regions 2 and 3 are subject to weak ‘mutually exciting’ effects from other regions.

Figure 14 shows the estimated transfer functions of the multivariate Hawkes process based on burglaries data. As can be seen from Fig. 14, the transfer functions have a periodic trend. Further, it should be noted that some of the estimated transfer functions are equal to 0 since the least-squares estimation $\hat{G}_p^h$ is restricted to $[0, \infty)$, indicating that there may be an inhibitory effect rather than a mutually exciting

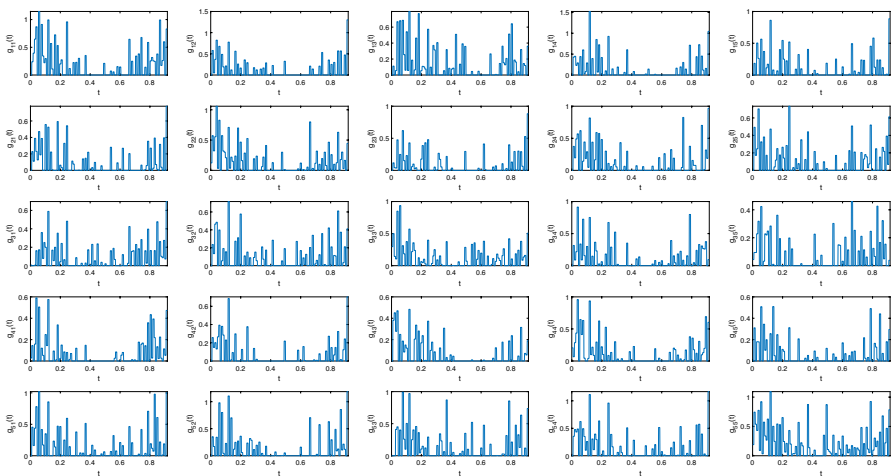


Fig. 14 Estimated transfer functions of multivariate Hawkes process based on burglaries data

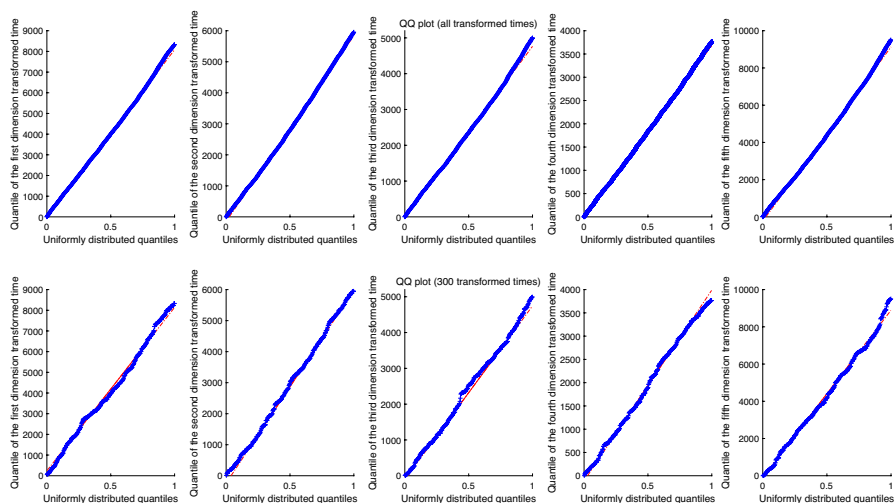


Fig. 15 Quantile-quantile plot of all transformed times (top) and 300 transformed times randomly sampled from all transformed times based on the estimated multivariate Hawkes process

Table 10 p value of Chi-square test for different regions

	$\chi^2_{1,2}$	$\chi^2_{1,3}$	$\chi^2_{1,4}$	$\chi^2_{1,5}$	$\chi^2_{2,3}$	$\chi^2_{2,4}$	$\chi^2_{2,5}$	$\chi^2_{3,4}$	$\chi^2_{3,5}$	$\chi^2_{4,5}$
p value	0.9905	1.0000	0.9765	0.9999	1.0000	0.9967	0.9999	1.0000	1.0000	1.0000

effect. In fact, the least-squares estimation can be generalized to the inhibitory point processes.

Figure 15 shows the QQ plot of all transformed times and 300 transformed times randomly sampled from all transformed times, respectively, of the five different dimensions to assess the goodness of fit of the estimated conditional intensity function. It can be seen from Fig. 15 that the estimated multivariate Hawkes process is close to the diagonal line. This is evidence that the estimated multivariate Hawkes process with spatial covariates produces an acceptable fit to the burglary data. Table 10 shows the Chi-square test p values between 1000 random samples of the residuals of the different dimensions from one simulation. It can be seen that the transformed times for the different dimensions are statistically independent of each other for a significance level of 0.05.

4.4.3 Effect analysis of the selection of d value and the d disjoint areas

The larger the d value, the more mutually exciting patterns between different areas can be found. We set $d = 5$ in Subsection 4.4.2. In the following we use the FLP, AIC, and BIC to evaluate the effect of the selection of d value and the d disjoint areas. The results are shown in Fig. 16. For Fig. 16, d is selected from set $\{5, \dots, 10\}$, and the burglaries are divided into d areas 20 times repeatedly

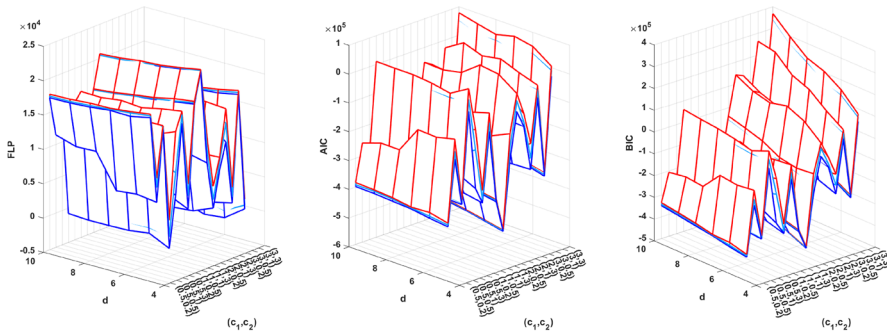


Fig. 16 FLP, AIC, and BIC values with different tuning parameters

by K-means method with random initial centers. Since there are 5 variables in parameter β , the minimum number of region divisions to ensure the invertibility of the matrix XX' is 5. From Fig. 16, it can be seen that when $d = 5$, the FLP value is maximum, and both AIC and BIC values are minimum. Furthermore, Fig. 16 shows that the optimal tuning parameters c_1 and c_2 are chosen to be the same on different partitions, given d .

For Fig. 17, the burglaries are divided into 5 areas 100 times repeatedly by K-means method with random initial centers. From Fig. 17, it can be seen that the FLP, AIC, and BIC values vary very little for different divisions of 5 disjoint areas. Further, we compared the estimated Q over the 20 simulations. The results are shown in Fig. 18. Figure 18 shows the boxplots of estimated q_{ij} , $i, j = 1, \dots, 5$. The horizontal axis represents the positions of the elements in the vector formed by stretching the branching matrix Q by rows. From Fig. 18, it can be seen that the branching coefficients q_{ii} , $i = 1, \dots, 5$ show significant values, meaning that burglaries occurring in the corresponding areas are self-exciting; branching coefficients q_{ij} , $i \neq j$, $i, j = 1, \dots, 5$ vary with different divisions of 5 disjoint areas. For instance, q_{12} can show significant value in some division and be close to zero in other division. This may be due to the fact that the regions with mutually exciting pattern are divided from area 2 to other areas.

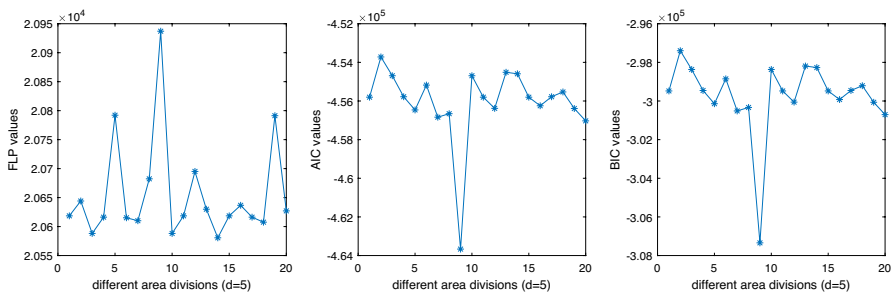


Fig. 17 FLP, AIC, and BIC values with different area divisions ($d = 5$)

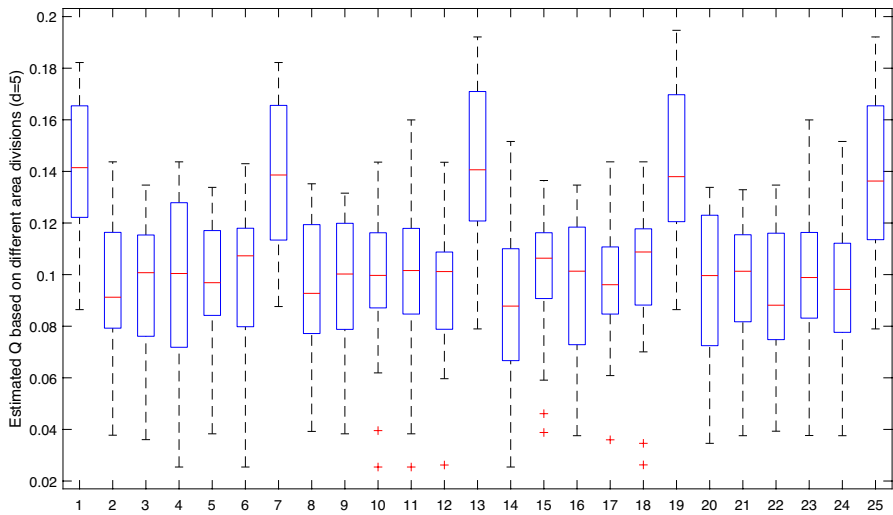


Fig. 18 Boxplots of estimated Q with different area divisions ($d = 5$)

5 Conclusion

We present multivariate Hawkes processes with spatial covariates that take into account the influence structure of spatial features in spatiotemporal event data and the spatiotemporal clustering patterns. Specifically, the baseline intensities are assumed to be a spatial Poisson regression model to explain spatial feature influence. The transfer functions are considered to be unknown but smooth and decreasing to explain the clustering phenomena. A semiparametric estimation method is derived by using time discretization and local constant approximation. The advantages of this estimation method compared to other estimation methods based on the maximum likelihood method are substantial for the case of estimating the transfer functions of multivariate Hawkes processes, in terms of flexible and rapid calculation. In addition, the proposed semiparametric estimation method can serve as a diagnostic to assess the performance of different parameterizations of the multivariate Hawkes processes. For example, in other applications of multivariate Hawkes processes, such as economics and sociology, there is less established literature on suitable parametric models. In such applications, a semiparametric estimation can be a powerful exploratory tool in determining an appropriate parameterization of the transfer functions.

Numerically, we quantify the variability of the estimates over multiple realizations of a synthetic process. Judging from the simulations described in Sects. 4.1–4.3, the proposed semiparametric estimation method has good performance in estimating branching matrix, transfer functions and parameter β when appropriate tuning parameters are chosen. In Sect. 4.4, we show how multivariate Hawkes processes with spatial covariates can be used for modeling crime data. In this application, the estimated transfer functions may be rough and insufficiently smooth to meet the assumptions of the proposed theorem. Additionally, the model's explanation of

the temporal triggering effect is limited since it is convoluted with different spatial aspects that have been aggregated, which makes it difficult to rigorously separate the spatial and temporal triggering effects, limiting the model's ability to explain complex spatiotemporal clustering phenomena. These are the drawbacks of the proposed multivariate Hawkes process model. To address these problems, a spatiotemporal model and spatiotemporal data with latitude and longitude coordinates would be necessary. However, such data are not readily available. In contrast, local average data or counting data are more commonly available. Therefore, the proposed multivariate Hawkes process model can be used for modeling and preliminary analysis of covariate and temporal triggering effects for local average data or counting data. Specifically, Table 5 in Sect. 4.3 shows that the proposed multivariate Hawkes process model gives an accurate estimate of the covariate coefficients for local counting data. This suggests that the proposed multivariate Hawkes process model is valid for analyzing covariate effects even when the data do not satisfy the proposed theorem's assumptions. On the other hand, for spatiotemporal data with accurate latitude and longitude coordinates, a rapid preliminary study can be conducted with the proposed multivariate Hawkes process model. If the temporal triggering effect displays cyclic fluctuations, it can be further modeled and analyzed with the spatiotemporal Hawkes process. Overall, the provided methodology is intended to provide a fast and efficient method for determining spatiotemporal clusteredness. After confirming the spatiotemporal clusteredness of the actual data, the spatiotemporal clustering patterns can be further analyzed using some refinement model, such as Park et al. (2021). In conclusion, the proposed model will enable new research on spatiotemporal event data, revealing the influence structure of spatial features and spatiotemporal clustering patterns of spatiotemporal events and contributing to improved policy strategies in different areas.

Funding The first author was supported by the Applied Basic Research Programs of Shanxi Province (Grant No. 201901D211105). The second author was supported by the Fundamental Research Program of Shanxi Province (Grant No. 202103021223023).

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., Delattre, S., Hoffmann, M., Muzy, J. F. (2012). Scaling limits for Hawkes processes and application to financial statistics. *Stochastic Processes and Their Applications*, 123(7), 251–265.
- Bernasco, W., Nieuwbeerta, P. (2004). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *British Journal of Criminology*, 45(3), 296–315.
- Clauset, A., Woodard, R. (2013). Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics*, 7(4), 1838–1865.
- 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-Verlag.

- Egesdal, M., Fathauer, C., Louie, K., Neuman, J. (2010). Statistical modeling of gang violence in Los Angeles. *Society for Industrial and Applied Mathematics Undergraduate Research Online*.
- Eichler, M., Dahlhaus, R., Dueck, J. (2016). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 38, 225–242.
- Embrechts, P., Liniger, T., Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48A, 367–378.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90.
- Kirchner, M. (2017). An estimation procedure for the Hawkes process. *Quantitative Finance*, 17(4), 571–595.
- Lewis, E., Mohler, G. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics*, 1(1), 1–20.
- Lewis, E., Mohler, G., Brantingham, P. J., Bertozzi, A. (2010). Self-exciting point process models of insurgency in Iraq. *Security Journal*, 25(3), 244–264.
- Lewis, E. A. (2012). *Estimation techniques for self-exciting point processes with applications to criminal behavior*. PhD thesis, University of California, Los Angeles.
- Liniger, T. J. (2009). *Multivariate Hawkes processes*. PhD 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. (2012). Geographic profiling from kinetic models of criminal behavior. *SIAM Journal on Applied Mathematics*, 72(1), 163–180.
- 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.
- Møller, 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.
- Park, J., Schoenberg, F. P., Bertozzi, A. L., Brantingham, P. J. (2021). Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. *Journal of the American Statistical Association*, 116(536), 1674–1687.
- Reinhart, A. (2018). A review of self-exciting Spatio-temporal point processes and their applications. *Statistical Science*, 33(3), 299–318.
- Reinhart, A., Greenhouse, J. (2017). Self-exciting point processes with spatial covariates: modeling the dynamics of crime. *Journal of the Royal Statistical Society*, 67(5), 1305–1329.
- Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464), 789–795.
- Short, M. B., Mohler, G. O., Brantingham, P. J., Tita, G. E. (2017). Gang rivalry dynamics via coupled point process networks. *Discrete and Continuous Dynamical Systems-Series B*, 19(5), 1459–1477.
- Tita, G., Ridgeway, G. (2007). The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquency*, 44(2), 208–237.
- 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, *Statistics in the Environmental and Earth Sciences* (eds. A. T. Walden and P. Guttorp), 220–246, Edward Arnold, London.
- Vere-Jones, D., Davies, R. B. (1966). A statistical survey of earthquakes in the main seismic region of New Zealand: Part 2-Time series analyses. *New Zealand Journal of Geology and Geophysics*, 9(3), 251–284.
- Yuan, B., Schoenberg, F. P., Bertozzi, A. L. (2021). Fast estimation of multivariate spatiotemporal Hawkes processes and network reconstruction. *Annals of the Institute of Statistical Mathematics*, 73(6), 1127–1152.

- Zhuang, J. C., Mateu, J. (2019). A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society Series A*, 182(3), 919–942.
- 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.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.