ORIGINAL ARTICLE
Journal Section

# A semi-parametric spatiotemporal Hawkes-type point process model with periodic background for crime data

Jiancang Zhuang[1,2,3]* | Jorge Mateu[4]

[1]Institute of Statistical Mathematics, Research Organisation of Information and Systems, Tokyo, 190-8562, Japan
[2]Department of Statistical Science, the Graduate University for Advanced Studies, Tokyo, 190-8562, Japan
[3]London Mathematical Laboratory, London, UK
[4]Department of Mathematics, University Jaume I of Castellón, Castellón, E-12071, Spain

**Correspondence**
Jiancang Zhuang, Institute of Statistical Mathematics, Research Organisation of Information and Systems, Tokyo, 190-8562, Japan
Email: zhuangjc@ism.ac.jp

**Funding information**
J. Mateu: grants MTM2016-78917-R from the Spanish Ministry of Economy and Competitiveness and P1-1B2015-40 from University Jaume I.

Past studies have shown that crime events are often clustered. This study proposes a spatiotemporal Hawkes-type point process model, which includes a background component with daily and weekly periodization, and a clustering component that is triggered by previous events. We generalize the nonparametric stochastic reconstruction method so that we can estimate each component in the background rate and the triggering response that appears in the model conditional intensity: the background rate includes a daily and a weekly periodicity, a separable spatial component, and a long-term background trend. Two relaxation coefficients are introduced to stabilize and fasten the estimation process. This model is used to describe the occurrences of violence or robbery cases in Castellon, Spain, during two years. The results show that the robbery crime is highly influenced by the daily life rhythms, revealed by its daily and weekly periodicity, and that about 3% of such crimes can be explained by clustering. Further diagnostic analysis show that the model could be improved by considering the following ingredients: (1) The daily occurrence patterns are different between weekends and working days; (2) in the city center, robbery activity shows different temporal patterns, in both weekly periodicity and long-term trend, from

**Abbreviations:** M-L, Marsan and Lenglineé.

other suburb areas.

**KEYWORDS**
crime, edge effect correction, Hawkes process, kernel estimate, periodicity, spatiotemporal point process, stochastic reconstruction

## 1 | INTRODUCTION

Point process modeling is a natural tool when describing the process of discrete events that occur in a continuous space, time or a space-time domain, such as urban fires, wild forest fires, crimes, earthquakes, diseases, tree locations, animal locations, communication network failures, etc. Depending on the type of the domain where the events occur, point process models are classified into two classes: spatial point processes and spatiotemporal/temporal point processes. The difference between these two types of models is that the latter ones have a special evolutionary time axis, based on which events can be sorted according to their chronological order and share many common features as time series sequences. When a property or a characteristic can also be attached to each event, such as the magnitude of an earthquake or the burned area of a wild fire, the point process is then called a marked point process.

Among the different types of point processes, clustered point processes have attracted many interests of mathematicians and statisticians. Typical clustering processes include the Neyman-Scott process (Neyman and Scott, 1953, 1958), which has been used for describing the distribution of locations of galaxies in the universe, and the Barlette-Lewis process to model the rain fall process (Bartlett, 1963; Lewis, 1964). Many spatiotemporal/temporal clustered point processes can be categorized into the Hawkes self-exciting process (Hawkes, 1971b,a; Hawkes and Oakes, 1974), including the epidemic-type aftershock sequence model (ETAS) for earthquake occurrence (e.g., Ogata, 1988, 1989, 1998; Zhuang et al., 2002). The basic assumption of this type of models is that the process consists of two subprocesses, a background subprocess considered as a Poisson process, which can be inhomogeneous in space and/or nonstationary in time, and a triggered subprocess composed by the exciting effect from all the events that occurred in the past. In other words, once an event occurs in the process, no matter whether it is a background event or an event excited by others, it excites a process of its own direct offspring according to some probability rules. Many powerful tools have been developed for the Hawkes process, such as stochastic declustering, stochastic reconstruction, Expectation-Maximization algorithm, first- and higher-order residuals, and Bayesian analysis, as well as the theories associated with the asymptotic properties (see a review by Reinhart, 2018)

The most common tools to predict crimes include "hot-spotting" (e.g. Bowers et al., 2004; Ratcliffe, 2004; Levine, 2017), "near-repeats" (e.g., Townsley et al., 2003), "leading indicator" regression (e.g., Cohen et al., 2007), and "risk terrain" (e.g., Caplan and Kennedy, 2016 ) models. The "hot-spotting" models produce static maps of locations where crimes tend to occur. "Near-repeats" analysis uses methods borrowed from epidemiology to test whether the local risk of crime elevates at a location immediately after a crime occurs and how/when the risk decays back to the baseline level. The "leading indicator" regression looks for covariates that can be used as local risk indicators of future serious crimes. "Risk terrain" modeling identifies the risks that come from particular features of a landscape and models how they co-locate to create unique behavior settings for crime.

Hawkes-type point-process modeling of crime was proposed by Mohler and others in a series of papers (Mohler et al., 2011, 2015; Mohler, 2014; Rosser and Cheng, 2016). By adopting the formulation of the Hawkes process, Mohler et al.'s model incorporates the time-varying hot spots and near-repeats with the assumption that every crime induces a locally higher risk of crime which decays in space and time. Reinhart and Greenhouse (2018) considered a background with simple spatial covariates. Since parametric models are difficult to construct for data where empirical studies are insufficient, nonparametric and semi-parametric estimation methods for the Hawkes model have been developed. Marsan and Lengliné (2008) made use of the stochastic declustering technique proposed by Zhuang et al. (2002, 2004) and Zhuang (2006) and proposed a so-called "model-independent stochastic declustering (MISD)" method, which is a nonparametric estimation method of an ETAS-type model (Ogata, 1988, 1998) for the earthquake occurrence. This method has been introduced at the same time when point process modeling was used for analyzing crime data for the first time (Mohler et al., 2011) followed by improvements from other authors (e.g., Johnson et al., 2018). In a parallel line, several authors have followed the path of spatiotemporal log-Gaussian Cox processes to model crime data, with the main focus on surveillance analysis to detect emergent spatiotemporal clusters of crimes (e.g., Rodrigues and Diggle).

However, in these studies of crime data based on Hawkes-type point processes, the periodic components in the background rate are not considered. Since criminals are also human beings, their behaviors should be controlled by their biological clock and could be influenced by the periodic activity of the society (Felson and Boba, 2010). Thus, periodicity, for instance, daily periodicity and weekly periodicity, should be taken into account when building a more precise model. Shirota and Gelfand (2017) used a log-Gaussian Cox process with circular time to model the daily and weekly periodicities of crimes in the city of San Francisco. Since the Cox point process is only a first-order intensity model, interactions among crime events were not counted.

The aim of this study is to analyze crime data by using an extended semi-parametric Hawkes model. Different from past studies, where the excitation effect has been emphasized, we focus on disentangling the periodic components from the long-term trend in the background rate. The reason for such a separation is straightforward: crime behavior is influenced by the criminal's biological clock and the rhythms of our social life. Consequently, we generalize the stochastic reconstruction technique, which has been used to estimate Hawkes-type models with a simple background rate, by considering the theory of residual analysis for point processes, so that different periodic components can be extracted from the background rate. In this study, kernel estimation, which is straightforward to implement, is used for estimating all background and clustering components. In the estimation procedure, to stabilize the algorithm, we introduce two so-called relaxation parameters, which quantify the overall background rate and clustering effect. We call the proposed model semi-parametric since these two relaxation parameters can be estimated by using maximum likelihood.

This article is organized as follows. Section 2 gives a brief description of the data. Section 3 provides the concepts and statistical modeling methodologies related to the Hawkes process. The estimation procedure comes in Section 4. Section 5 presents the results of the statistical analysis, including model fitting and a diagnostic analysis to verify the hypotheses related to the model assumptions. Finally, the conclusions are summarized in Section 6.

## 2 | DATA

In this study, we analyze the robbery-related violence data in Castellon city, Spain, during the years of 2012 and 2013. The data reports geo-referenced coordinates of phone calls received by the Police station in the city of Castellon from January 2012 to December 2013. Castellon is a mediterranean city of around 180000 inhabitants. The listed calls were received at the local Police call center or transferred by 112 emergency service to the local Police call center. Geo-codification was performed indirectly by local officials based on precise address information provided by the callers. The calls comprise up to nine different types of crimes or anti-social behavior categories, but we here only focus on robbery-related violence data, comprising a total number of 5089 events happening in the streets of Castellon. The city of Castellon is divided into 108 census tracks with an overall surface of $108.6 \text{ km}^2$. Figures 1 and 2 show several two- and three-dimensional plots of the events in the city to provide a first rough idea of the type of data that we are analyzing.

# 3 | MODEL AND METHODOLOGY

## 3.1 | Hawkes process

The Hawkes process describes the excitation mechanisms among a series of events that occur in a continuous time domain or in a spatiotemporal domain. A point process can be completely defined by its conditional intensity. For the purely temporal case, the conditional intensity is defined by

$$\lambda(t) \quad = \quad \lim_{\Delta \downarrow 0} \frac{1}{\Delta_t} \Pr\left\{N\left([t, t+\Delta)\right) = 1 \mid \mathcal{H}_t\right\} \tag{1}$$

where $\mathcal{H}_t$ denotes the $\sigma$-algebra generated by the observational history of the process $N$ before time $t$ but not including $t$. A temporal Hawkes process, say $N = \{t_i : i \in \mathbb{Z}\}$ with $\mathbb{Z}$ being the set of all integers, has a conditional intensity of the form (Hawkes, 1971a,b)

$$\lambda(t) = \mu + \int_{-\infty}^{t_-} g(t-u) N(\mathrm{d}u) = \mu + \sum_{i:t_i<t} g(t-t_i), \tag{2}$$

where $\mu$ is the occurrence rate of spontaneous events (also called background events), and $g(t)$ is the occurrence rate of direct offspring generated by an event occurring at 0. Note this indicates that both $\mu$ and $g$ are nonnegative. The criticality parameter, which is the average number of direct offspring per ancestor, is given by

$$\rho = \int_0^\infty g(u)\,\mathrm{d}u. \tag{3}$$

If $\rho < 1$, this parameter is identical to the branching ratio, the proportion of non-spontaneous events in the whole process. In general, these two quantities are different (see Zhuang et al., 2013, for details).

The Hawkes process can be easily extended to the spatiotemporal version

$$\lambda(t, x) = \mu(x) + \int_{\mathbb{R}^d \times (-\infty, t_-)} g(t-s, x-u)\, N(\mathrm{d}s \times \mathrm{d}u) \tag{4}$$

where $x$ denotes the locations in the space of $\mathbb{R}^d$, $\mu(x) \geq 0$, and $g(t, x) \geq 0$ for all $x$ and $t$. It is can also generalized to the multivariate case where, if we have $K$ types events in total, each type has a conditional intensity

$$\lambda_k(t, x) = \mu_k(x) + \sum_{\ell}^{K} \int_{\mathbb{R}^d \times (-\infty, t_-)} g_{\ell,k}(t, x; s, u)\, N_\ell(\mathrm{d}s \times \mathrm{d}u), \tag{5}$$

for $k = 1, \cdots, K$, where $\mu_k(x)$ represents the occurrence rate of spontaneous events (also called background) for type-$k$ events, and $g_{\ell,k}(t, x; s, u)$ is the occurrence rate of events that are excited by a type-$\ell$ event at $(s,\ u)$. Again we assume $\mu_k(x)$ and $g_{\ell,k}(x)$ are nonnegative for $k, \ell = 1, 2, \cdots, K$.

Given observation data of crime events in an observational space-time window $S \times T$, for a parametric Hawkes

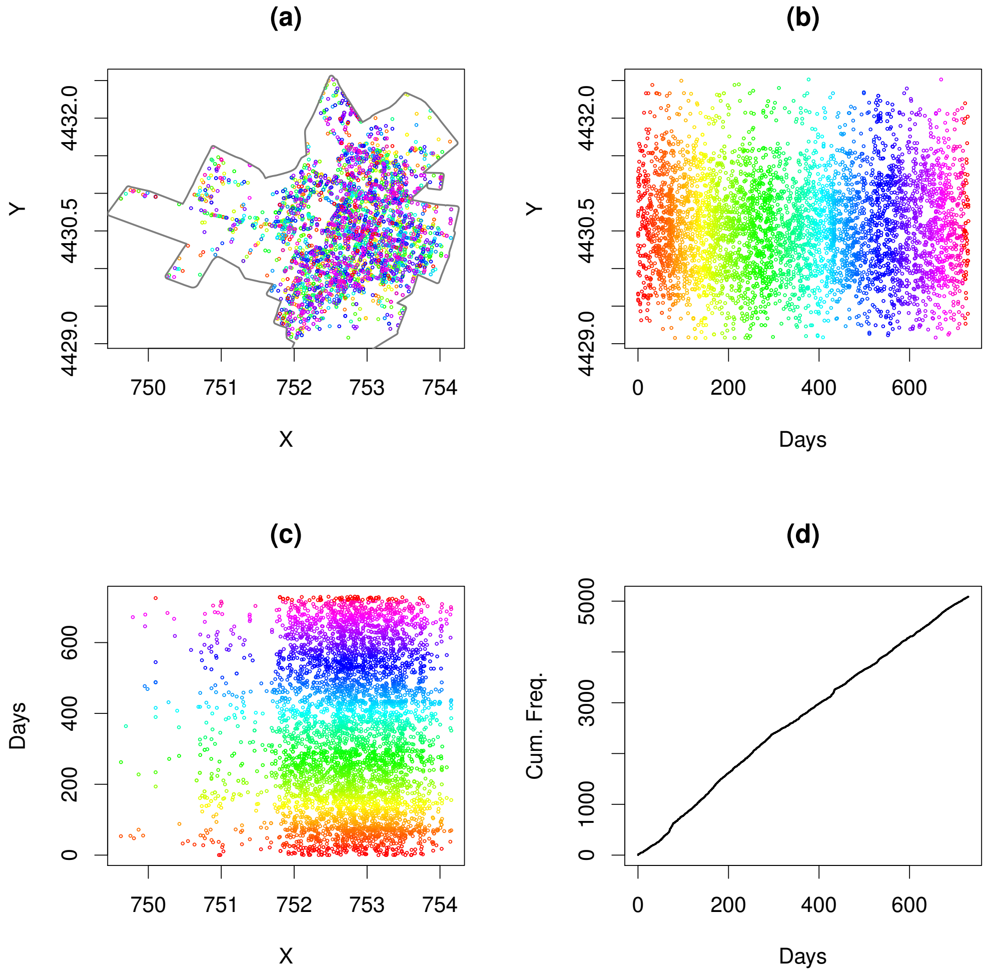


**FIGURE 1** Basic information of robbery-related violence in Castellon, Spain, 2012-2013: (a) Spatial locations, (b) $y$-$t$ coordinates, (c) $t$-$x$ coordinates, and (d) cumulative numbers against times. The rainbow colors show the occurrence times of the events, with red-colored points representing the earliest events and magenta ones the latest.

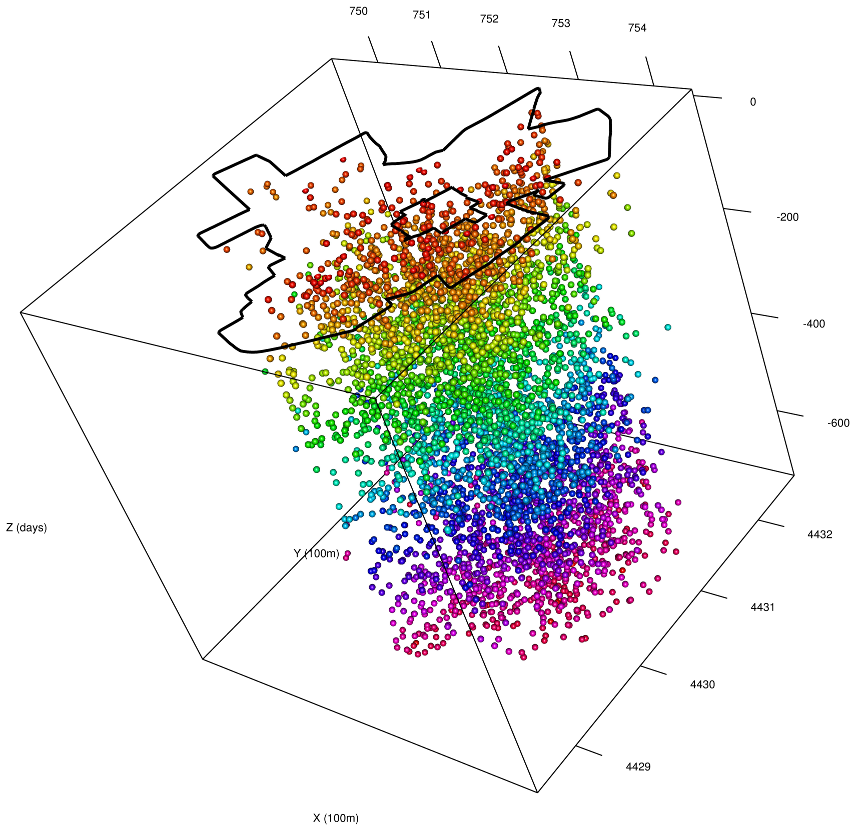


**FIGURE 2** A 3D plot of robbery-related violence in Castellon, Spain, 2012-2013. The rainbow colors show the occurrence times of the events, with red-colored points representing the earliest events and magenta ones the latest.

model, one can use maximum likelihood estimation to estimate the model parameters, i.e.,

$$\begin{aligned}\hat{\theta} &= \arg_\theta \max \log L(\cdot;\theta) \\ &= \arg_\theta \max \left[ \sum_{i:(t_i,x_i,y_i)\in S\times T} \log \lambda(t,x,y;\theta) - \int_T \iint_S \lambda(t,x,y;\theta)\,\mathrm{d}x\,\mathrm{d}y\ \mathrm{d}t \right]. \end{aligned} \tag{6}$$

Here we refer to Chapter 7 of Daley and Vere-Jones (2003) for the derivation of the standard likelihood function for point processes that are specified by conditional intensities.

## 3.2 | Stochastic declustering and reconstruction

Consider a Hawkes process with conditional intensity

$$\lambda(t,x) = \mu(t,x) + \sum_{k:t_k<t} g(t-t_k, x-x_k), \tag{7}$$

where $\mu(t,x)$ is the background rate, which is different from the corresponding term in (4) as it allows to be time dependent, and $g(t,x)$ is the occurrence rate triggered by an event at time 0 and location at the origin.

The probability that an event, say $j$, is a background event, i.e., *background probability*, is given by

$$\varphi_j = \Pr\{\text{Event } j \text{ is a background event}\} = \frac{\mu(t_j,x_j)}{\lambda(t_j,x_j)} \tag{8}$$

and the probability that event $j$ is triggered by another event $i$, $i<j$, is

$$\rho_{ij} = \Pr\{\text{Event } j \text{ is triggered by } i\} = \frac{g(t_j-t_i, x_j-x_i)}{\lambda(t_j,x_j)}. \tag{9}$$

It is easy to see

$$\varphi_j + \sum_{i=1}^{j-1} \rho_{ij} = 1, \ \text{ for all } j. \tag{10}$$

Another explanation for the above equation is that, once an event occurs at $(t,x)$, we can say that at $(t,x)$ we have observed $\varphi_j$ background events and that, for each $i = 1,\cdots,j-1$, event $i$ triggers $\rho_{ij}$ direct offspring at $(t_j,x_j)$. In this way, event $j$ is sliced into background and offspring from previous events (Zhuang et al., 2004). Consequently, the above treatment provides a nonparametric way to estimate functions $\mu(\cdot,\cdot)$ and $g(\cdot,\cdot)$. For example, $g(\cdot,\cdot)$ can be estimated by

$$\hat{g}(t,x) = \frac{\sum_{i,j} \rho_{ij}\, I(|t_j-t_i-t|<\delta_t)\, I(|x_j-x_i-x|<\delta_x)}{4\delta_t\,\delta_x\,\sum_{i,j}\rho_{ij}} \tag{11}$$

where the denominator is for normalizing purposes, and $\delta_t$ and $\delta_x$ are two small positive numbers. $\mu(\cdot,\cdot)$ can be also estimated through, e.g., a weighted kernel estimation as follows

$$\hat{\mu}(t,x) = \sum \varphi_i\, Z_{h_x}(x-x_i)\, Z_{h_t}(t-t_i), \tag{12}$$

where $Z_h$ is the Gaussian kernel with bandwidth $h$, $h_x$ and $h_t$ are bandwidths used for the smoothing in space and time, respectively, and $\varphi_i$ is defined in (8).

In the above, when estimating $\mu(t, x)$ and $g(t, x)$, we need to know $\varphi_i$ and $\rho_{ij}$, and when estimating $\varphi_i$ and $\rho_{ij}$, we need to know $\mu$ and $g$. Such a loop can be solved by an iterative algorithm. Given an observed process of events $\{(t_i, x_i) : i = 1, \cdots, n\}$ in a time-space window $T \times S$, by assuming some initial guess of $\mu$ and $g$, we obtain $\varphi_i$ and $\rho_{ij}$, for all possible $i, j$. Then we estimate the background rate $\mu$ and each component in the clustering part $g$ by using $\varphi_i$ and $\rho_{ij}$, through some nonparametric methods, for example, kernel estimation or histogram. Once $\mu$ and $g$ are updated, we go back to the step of calculating $\varphi$, or stop if convergence is reached.

## 3.3 | On the Marsan-Lengliné estimation algorithm and Mohler's analysis of burglary data in Los Angeles

The idea of the stochastic reconstruction algorithm firstly appeared in Zhuang et al. (2004) and Zhuang (2006) and it was then used by Marsan and Lengliné (2008) (M-L). Mohler et al. (2011) introduce it for the analysis of crime data. It is worthwhile to mention that in the M-L algorithm, M-L assumed that $g$ is a stepwise constant function and the MLE yields a histogram estimation. In the M-L algorithm, $\mu$ is assumed to be constant throughout the whole observational space-time range, in order not to solve a non-fully-ranked equation system.

Mohler et al. (2011) analyzed the break-in burglary data from the Los Angeles Police Department. Their dataset consisted of 5376 reported residential burglaries in an 18 km × 18 km region of San Fernando Valley, Los Angeles during 2004–2005. They used a model with conditional intensity

$$\lambda(t, x, y) = \nu(t)\mu(x, y) + \sum_{k:t_k<t} g(t - t_k, x - x_k, y - y_k) \tag{13}$$

In Mohler et al. (2011), the background rate is assumed to be a function of space and time and they used kernel functions to smooth the estimates of both $\mu$ and $g$. In this article, we improve the above algorithm by (i) introducing relaxing parameters and (ii) considering periodic components in the background rates.

## 3.4 | Model formulation

We consider using the following space-time point process model to describe the crime data in Section 2, which is completely specified by a conditional intensity function

$$\lambda(t, x, y) = \mu_{\mathrm{t}}(t)\,\mu_{\mathrm{d}}(t)\,\mu_{\mathrm{w}}(t)\,\mu_{\mathrm{b}}(x, y) + \int_{-\infty}^{t_-} \iint_S g(t - s, x - u, y - v)\, N(\mathrm{d}u \times \mathrm{d}v \times \mathrm{d}s), \tag{14}$$

where $\mu_{\mathrm{t}}(t)$, $\mu_{\mathrm{d}}(t)$, and $\mu_{\mathrm{w}}(t)$ represent the trend term, the daily periodicity, and the weekly periodicity in the temporal components of the background rate, respectively, $\mu_{\mathrm{b}}(x, y)$ represents the spatial homogeneity of the background rate, and $g(t - s, x - u, y - v)$ represents the subprocess triggered by an event previously occurring at location $(u, v)$ and time $s$. Note that model (14) extends models (4), (7), and (13) by enabling the background rate to include a spatial background pattern that can be separated from the periodicity effects and the long term temporal trend.

# 4 | ESTIMATION METHOD AND ALGORITHM

We estimate $\mu_t, \mu_d, \mu_w, \mu_b$ and $g$ nonparametrically by using the stochastic reconstruction method proposed in Zhuang (2006). First, we rewrite the conditional intensity as

$$\lambda(t,x,y) = \mu_0\,\mu_t(t)\,\mu_d(t)\,\mu_w(t)\,\mu_b(x,y) + A\int_{-\infty}^{t-}\iint_S g(t-s)\,h(x-u,y-v)\,N(\mathrm{d}u\times\mathrm{d}v\times\mathrm{d}s), \tag{15}$$

where $A$ and $\mu_0$ are relaxation coefficients to be estimated, the average values of $\mu_t(t), \mu_d(t), \mu_w(t)$ and $\mu_b(x,y)$ are all normalized to 1, and $g$ and $h$ are p.d.f.s, i.e., $\int_0^\infty g(s)\,\mathrm{d}s = 1$, and $\iint_S h(u,v)\,\mathrm{d}u\,\mathrm{d}v = 1$. Here we separate the spatiotemporal clustering response function into a temporal and a spatial components in order to avoid the nonparametric estimation of a 3-dimensional function.

Since the periodic components of the background rate in our model formulation cannot be directly estimated by using the stochastic reconstruction method, we use the residual analysis method developed in Zhuang (2006) to solve this problem. The key point of residual analysis for temporal/spatiotemporal point processes is that the conditional intensity of a point process has the following property. Suppose that a spatiotemporal point process $N$ is equipped with a conditional intensity $\lambda(t,x)$; for a predictable process $f(t,x)$, we have

$$\mathbf{E}\left[\int_{[T_1,T_2]\times S} f(t,x)\,\mathrm{d}N(\mathrm{d}t\times\mathrm{d}x)\right] = \mathbf{E}\left[\int_{T_1}^{T_2}\int_S f(t,x)\,\lambda(t,x)\,\mathrm{d}t\,\mathrm{d}x\right], \tag{16}$$

for any given time interval $[T_1,\ T_2]$ and area $S$, provided that the integral on either side exists, or that $f$ is nonnegative.

## 4.1 | Reconstructing background components

Given a realization of the point process $\{(t_i,x_i,y_i): i=1,2,\cdots,n\}$ in a time-space range $[T_1,T_2]\times S$, where $t$ (day) and $(x,y)$ (km) denote time and location, respectively, the long-term trend term $\mu_t(t)$ in the background component can be reconstructed in the following way.

Let

$$w^{(t)}(t,x,y) = \mu_t(t)\,\mu_b(x,y)/\lambda(t,x,y)$$

and $f(t,x,y) = w^{(t)}(t,x,y)$ and substitute $f$ into (16). Then, assuming that $\mu_t$ is smooth enough,

$$\begin{aligned}
&\sum_i w^{(t)}(t_i,x_i,y_i)\,I(t_i\in[t-\Delta_t,\ t+\Delta_t])\\
\approx\ &\int_{T_1}^{T_2}\iint_S w^{(t)}(s,x,y)\lambda(s,x,y)\,I(s\in[t-\Delta_t,t+\Delta_t])\,\mathrm{d}s\,\mathrm{d}x\,\mathrm{d}y\\
=\ &\int_{t-\Delta_t}^{t+\Delta_t}\mu_t(s)\,\mathrm{d}s\iint_S\mu_b(x,y)\,\mathrm{d}x\,\mathrm{d}y\\
\propto\ &\int_{t-\Delta_t}^{t+\Delta_t}\mu_t(s)\,\mathrm{d}s\\
\approx\ &2\mu_t(t)\,\Delta_t,
\end{aligned} \tag{17}$$

where $\Delta_t$ is a small positive number. For ease of writing, define

$$w_i^{(\mathrm{t})} = \mu_{\mathrm{t}}(t_i)\,\mu_{\mathrm{b}}(x_i, y_i)/\lambda(t_i, x_i, y_i), \tag{18}$$

then

$$\hat{\mu}_{\mathrm{t}}(t) \propto \sum_i w_i^{(\mathrm{t})}\, I(t_i \in [t - \Delta_t,\ t + \Delta_t]). \tag{19}$$

Similarly, we can reconstruct the other components in the background rate as follows

$$\hat{\mu}_{\mathrm{d}}(t) \propto \sum_i w_i^{(\mathrm{d})}\, I\left(t_i \in \bigcup_{k\in\mathbb{Z}} [t + k - \Delta_t,\ t + k + \Delta_t]\right),\ \ t \in [0, 1], \tag{20}$$

$$\hat{\mu}_{\mathrm{w}}(t) \propto \sum_i w_i^{(\mathrm{w})}\, I\left(t_i \in \bigcup_{k\in\mathbb{Z}} [t + 7k - \Delta_t,\ t + 7k + \Delta_t]\right),\ \ t \in [0, 7], \tag{21}$$

and

$$\hat{\mu}_{\mathrm{b}}(x, y) \propto \sum_i \varphi_i\, I(x_i \in [x - \Delta_x,\ x + \Delta_x])\, I(y_i \in [y - \Delta_y,\ x + \Delta_y]), \tag{22}$$

where

$$w_i^{(\mathrm{d})} = \mu_{\mathrm{d}}(t_i)\,\mu_{\mathrm{b}}(x_i, y_i)/\lambda(t_i, x_i, y_i), \tag{23}$$

$$w_i^{(\mathrm{w})} = \mu_{\mathrm{w}}(t_i)\,\mu_{\mathrm{b}}(x_i, y_i)/\lambda(t_i, x_i, y_i), \tag{24}$$

$$\varphi_i = \mu_0\,\mu_{\mathrm{t}}(t_i)\,\mu_{\mathrm{d}}(t_i)\,\mu_{\mathrm{w}}(t_i)\,\mu_{\mathrm{b}}(x_i, y_i)/\lambda(t_i, x_i, y_i), \tag{25}$$

and $\Delta_t$, $\Delta_x$, and $\Delta_y$ are small positive numbers. In the above, the rescaled weights $w_i^{(\mathrm{t})}$, $w_i^{(\mathrm{d})}$, and $w_i^{(\mathrm{d})}$ are the key quantities for reconstructing the long trend, the daily periodicity, and the weekly periodicity in the background rate.

## 4.2 | Reconstructing excitation components

To estimate $g$ and $h$, we need to use the quadratic form in (**??**). First, let

$$\varrho\left(s^{(1)}, u^{(1)}, v^{(1)}, s^{(2)}, u^{(2)}, v^{(2)}\right) = \begin{cases} g\left(s^{(2)} - s^{(1)}\right)\, h\left(u^{(2)} - u^{(1)}, v^{(2)} - v^{(1)}\right)/\lambda\left(s^{(2)}, u^{(2)}, v^{(2)}\right), & s^{(2)} \geq s^{(1)}; \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

It is clear that $\varrho\left(s^{(1)},u^{(1)},v^{(1)},s^{(2)},u^{(2)},v^{(2)}\right)$ is a deterministic function for any fixed $\left(s^{(1)},u^{(1)},v^{(1)}\right)$, and, of course, predictable. Substituting $f\left(s^{(1)},u^{(1)},v^{(1)},s^{(2)},u^{(2)},v^{(2)}\right)=\varrho\left(s^{(1)},u^{(1)},v^{(1)},s^{(2)},u^{(2)},v^{(2)}\right)\,I(s^{(2)}-s^{(1)}\in[t-\Delta_t,t+\Delta_t]$ into (16) yields

$$\begin{aligned}
&\sum_j \varrho\left(t_i,x_i,y_i,s_j,u_j,v_j\right)\ I(t_j-t_i\in[t-\Delta_t,t+\Delta_t])\\
\approx\ & \int_{T_1}^{T_2}\iint_S \varrho\left(s^{(1)},u^{(1)},v^{(1)},s^{(2)},u^{(2)},v^{(2)}\right)\ I\left(s^{(2)}-s^{(1)}\in[t-\Delta_t,t+\Delta_t]\right)\lambda\left(s^{(2)},u^{(2)},v^{(2)}\right)\,ds^{(2)}\,du^{(2)}\,dv^{(2)}\\
\approx\ & 2\,g(t)\,\Delta_t\times\iint_S h\left(u^{(2)}-u^{(1)},v^{(2)}-v^{(1)}\right)\,du^{(2)}\,dv^{(2)}\\
\propto\ & g(t),
\end{aligned}\tag{27}$$

Note that in the last step of the above equation, the integrals are functions that do not depend on time $t$ or the spatial location, and thus they are independent of $(t_i,x_i,y_i)$. Therefore,

$$\sum_i\sum_j \varrho(t_i,x_i,y_i,s_j,u_j,v_j)\,I(t_j-t_i\in[t-\Delta_t,t+\Delta_t])$$

is approximately proportional to $g(t)$, i.e., $g(t)$ can be estimated by

$$\hat{g}(t)\propto\sum_{i,j}\rho_{ij}\ I(t_j-t_i\in[t-\Delta_t,t+\Delta_t])\tag{28}$$

where

$$\rho_{ij}=g(t_j-t_i)\,h(x_j-x_i,y_j-x_i)/\lambda(t_j,x_j,y_j),\quad i<j.\tag{29}$$

Similarly,

$$\hat{h}(x,y)\propto\sum_{i,j}\rho_{ij}\ I(x_j-x_i\in[x-\Delta_x,x+\Delta_x])\,I(y_j-y_i\in[y-\Delta_y,y+\Delta_y]),\tag{30}$$

where $\Delta_x$ and $\Delta_y$ are small positive numbers.

## 4.3 | Estimating relaxation coefficients

Once $\mu_{\mathrm{t}},\mu_{\mathrm{d}},\mu_{\mathrm{w}},\mu_{\mathrm{b}},g$ and $h$ are estimated, we can update the relaxation coeffiecients, $\mu_0$ and $A$, through maximizing the likelihood function

$$\log L=\sum_{i=1}^{n}\log\lambda(t_i,x_i,y_i)-\int_0^T\iint_S\lambda(t,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t.\tag{31}$$

Denote

$$U=\int_0^T\iint_S\mu_{\mathrm{t}}(t)\,\mu_{\mathrm{d}}(t)\,\mu_{\mathrm{w}}(t)\,\mu_{\mathrm{b}}(x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t$$

and

$$G = \sum_i \int_{t_i}^{T} \iint_S g(t - t_i)\, h(x - x_i, y - y_i)\, \mathrm{d}x\mathrm{d}y\, \mathrm{d}t.$$

The equations $\frac{\partial}{\partial \mu_0} \log L = 0$ and $\frac{\partial}{\partial A} \log L = 0$ give

$$\sum_{i=1}^{n} \frac{\mu_{\mathrm{t}}(t_i)\, \mu_{\mathrm{d}}(t_i)\, \mu_{\mathrm{w}}(t_i)\, \mu_{\mathrm{b}}(x_i, y_i)}{\lambda(t_i, x_i, y_i)} - U = 0, \tag{32}$$

$$\sum_{i=1}^{n} \frac{\sum_{j:t_j<t_i} g(t_j - t_i)\, h(x_j - x_i, y_j - y_i)}{\lambda(t_i, x_i, y_i)} - G = 0. \tag{33}$$

The above equations can be solved by the following iteration system

$$A^{(k+1)} = \frac{n - \sum_{i=1}^{n} \varphi_i^{(k)}}{G}, \tag{34}$$

$$\mu_0^{(k+1)} = \frac{n - A^{(k+1)} G}{U}, \tag{35}$$

where

$$\varphi_i^{(k)} = \frac{\mu_0^{(k)} \mu_{\mathrm{t}}(t_i)\, \mu_{\mathrm{d}}(t_i)\, \mu_{\mathrm{w}}(t_i)\, \mu_{\mathrm{b}}(x_i, y_i)}{\mu_0^{(k)} \mu_{\mathrm{t}}(t_i)\, \mu_{\mathrm{d}}(t_i)\, \mu_{\mathrm{w}}(t_i)\, \mu_{\mathrm{b}}(x_i, y_i) + A^{(k)} \sum_{j:\, t_j<t_i} g(t_j - t_i)\, h(x_j - x_i, y_j - y_i)}. \tag{36}$$

## 4.4 | Smoothing estimates and correcting for edge effects

To get robust reconstruction results and to ensure the convergence of the above iterative algorithm, instead of using histograms directly, we use kernel functions to smooth our estimates. That is to say, (19) to (22), (28) and (30) become

$$\hat{\mu}_{\mathrm{t}}(t) \propto \sum_i w_i^{(\mathrm{t})}\, Z(t - t_i;\, \omega_{\mathrm{t}}), \tag{37}$$

$$\hat{\mu}_{\mathrm{d}}(t) \propto \sum_i w_i^{(\mathrm{d})} \sum_{k=0}^{T} Z(t - t_i + \lfloor t_i \rfloor - k;\, \omega_{\mathrm{d}}), \tag{38}$$

$$\hat{\mu}_{\mathrm{w}}(t) \propto \sum_i w_i^{(\mathrm{w})} \sum_{k=0}^{\lfloor T/7 \rfloor} Z(t - t_i + 7 \cdot \lfloor t_i/7 \rfloor - 7k;\, \omega_{\mathrm{d}}), \tag{39}$$

$$\hat{\mu}_{\mathrm{b}}(x, y) \propto \sum_i \varphi_i\, Z(x - x_i;\, \omega_x)\, Z(y - y_i;\, \omega_y), \tag{40}$$

$$\hat{g}(t) \propto \sum_{i,j} \rho_{ij}\, Z(t - t_j + t_i; \omega_g), \tag{41}$$

$$\hat{h}(x, y) \propto \sum_{i,j} \rho_{ij}\, Z(x - x_j + x_i;\, \omega_{h_x})\, Z(y - y_j + y_i;\, \omega_{h_y}), \tag{42}$$

respectively, where $Z(x;\, \omega) = \frac{1}{\sqrt{2\pi}\omega} \exp\left(-\frac{x^2}{2\omega^2}\right)$ is the Gaussian kernel, and $\lfloor x \rfloor$ represents the largest integer not bigger than $x$. In the above equations, and when no confusion arises, we abuse the notation and use $\hat{\ }$ for the new estimates.

An important issue with kernel smoothing is the edge effect. To correct for the edge effect, we finally adopt the following estimates

$$\hat{\mu}_{\mathrm{t}}(t) \propto \sum_i w_i^{(\mathrm{t})} \frac{Z(t - t_i;\, \omega_{\mathrm{t}})}{\int_0^T Z(u - t_i;\, \omega_{\mathrm{t}})\, \mathrm{d}u}, \tag{43}$$

$$\hat{\mu}_{\mathrm{d}}(t) \propto \sum_i w_i^{(\mathrm{d})} \frac{\sum_{k=0}^{T} Z(t - t_i + \lfloor t_i \rfloor - k;\, \omega_{\mathrm{d}})}{\int_0^T Z(u - t_i;\, \omega_{\mathrm{d}})\, \mathrm{d}u}, \tag{44}$$

$$\hat{\mu}_{\mathrm{w}}(t) \propto \sum_i w_i^{(\mathrm{w})} \frac{\sum_{k=0}^{\lfloor T/7 \rfloor} Z(t - t_i + 7 \cdot \lfloor t_i/7 \rfloor - 7k;\, \omega_{\mathrm{w}})}{\int_0^T Z(u - t_i;\, \omega_{\mathrm{w}})\, \mathrm{d}u}, \tag{45}$$

$$\hat{\mu}_{\mathrm{b}}(x, y) \propto \sum_i \varphi_i \frac{Z(x - x_i;\, \omega_x)\, Z(y - y_i;\, \omega_y)}{\iint_S Z(u - x_i;\, \omega_x)\, Z(v - y_i;\, \omega_y)\, \mathrm{d}u\mathrm{d}v}, \tag{46}$$

$$\hat{g}(t) \propto \frac{\sum_{i,j} \rho_{ij} \frac{Z(t - t_j + t_i; \omega_g)}{\int_0^{T - t_i} Z(u - t_j; \omega_g)\, \mathrm{d}u}}{\sum_i I(t_i + t \le T)}, \tag{47}$$

$$\hat{h}(x, y) \propto \frac{\sum_{i,j} \rho_{ij} \frac{Z(x - x_j + x_i;\, \omega_{h_x})\, Z(y - y_j + y_i;\, \omega_{h_y})}{\iint_S Z(u - x_j + x_i;\, \omega_{h_x})\, Z(u - y_j + y_i;\, \omega_{h_y})\, \mathrm{d}x\, \mathrm{d}y}}{\sum_i I\left((x_i + x, y_i + y) \in S\right)}. \tag{48}$$

In each of the above equations, the integral of the kernel function prevents “leaking out of masses” outside the spatial or temporal range of interest. The denominators in (47) and (48) are for repetition corrections, i.e., for how many times the triggering effect at time lag $t$ or the spatial offset $(x, y)$ is observed.

**TABLE 1** Results from fitting the model in equation (15) and three other models (see Section 5.1 for the details). Parameters $\mu$ and $A$ are the relaxation coefficients.

| Model | $\hat{\mu}$ | $\hat{A}$ | $\log L$ |
|---|---|---|---|
| non-periodic Poisson | 0.7920 | NA | -1335.07 |
| periodic Poisson | 0.7927 | NA | -1050.26 |
| non-periodic & triggering | 0.7710 | 0.02838 | -1304.50 |
| periodic & triggering | 0.7713 | 0.02913 | -920.81 |

### 4.5 | Iterative algorithm

As explained in Section 3.1, when estimating $\mu$, $g$, and $h$, we need to know $\varphi_i$ and $\rho_{ij}$, and when estimating $\varphi_i$ and $\rho_{ij}$, we need to know $\mu$ , $g$, and $h$. To estimate them simultaneously together with the relaxation parameters $\mu_0$ and $A$, we have designed the following iterative algorithm.

**Algorithm:**

Step 1. Set up initial values of $\mu_t, \mu_d, \mu_w, \mu_{bg}, f, g, \mu_0$ and $A$.
Step 2. Calculate $w_i^{(t)}$, $w_i^{(d)}$, $w_i^{(w)}$, $\varphi_i$, and $\rho_{ij}$ for all possible $i$ and $j$, using (18), (23) – (25), and (29), respectively.
Step 3. Estimate $\mu_t, \mu_d, \mu_w, \mu_{bg}, f$, and $g$ using equations (43) to (48).
Step 4. Estimate $\mu_0$ and $A$ using (34) to (36).
Step 5. Stop if the results are convergent; otherwise, go to Step 2.

## 5 | DATA ANALYSIS

We analyze robbery-related violence data in Castellon city, Spain, during the years of 2012 and 2013, as presented in Section 2. See Figures 1 and 2 for graphical illustrations of the data set.

### 5.1 | Model fitting

We fit four models to the crime data that are given in Section 2: (1) a non-periodic but nonstationary Poisson model with $\lambda(t, x, y) = \mu_0\, \mu_t(t)\, \mu_b(x, y)$, (2) a periodic Poisson model with $\lambda(t, x, y) = \mu_0\, \mu_t(t)\, \mu_d(t)\, \mu_w(t)\, \mu_b(x, y)$, (3) a similar model as in Eq. (15) but without daily and weekly periodic effects, and (4) the model in (15). In our analysis, we adopt bandwidths of 0.03, 0.5, and 10, with days as the temporal unit, in the estimation of the daily periodicity, weekly periodicity, and long-term background rate, respectively, for all the four models. These bandwidths are selected according to resolution requirement of each component. The estimates of parameters and likelihoods are listed in Table 1. Since the model with daily and weekly periodic effects is much better than the others, with differences of 414.26, 129.55, and 383.69 in log-likelihood, we only discuss the full model in the following sections.

The corresponding estimated surface for the spatial background rate $\mu_b(x, y)$ and the other components are shown in Figure 3. The general trend (Figure 3a) indicates that there is a larger number of events occurring in the first year than in the second one. Also the occurrence rate of events keeps quite stationary throughout the second year. The weekly periodicity component (Figure 3b) indicates that the robbery events have a steady increase from Thursdays to Sundays which is consistent to reality as it is the time when more people are working and moving around the city. In addition, we can identify two significant peaks of occurrences within a day (Figure 3c), corresponding to 12h – 14h (lunch time) and 19h – 22h (dinner time), which are again the periods when more people are in the streets. On the other hand, the occurrence rate of such crimes is relatively much lower around 4h to 10h in the morning, during which most people are resting. The reconstructed spatial and temporal response functions in the clustering component (Figures 3d and 3e) imply that, once a crime occurs, it likely triggers another crime within the coming 3 days and within 100 meters in distance.

Here, $A \approx 0.03$ implies that about 3% of the 5089 crime events (about 152 events), which should not be considered as a small number, can be explained by the triggering effect. Comparing to the results of the analysis of the burglary crimes in Los Angeles during the period of 2010 – 2012 in Mohler et al. (2011), the clustering effect in the robbery violence data seems much lower. In Reinhart and Greenhouse (2018), the proportion of clustering events in all the burglary crimes in Pittsburgh during 2011 to 2016 amounts to 47%. The reason might be that the same burglar watches and visits several neighboring houses within a short time span, while a robber always escapes from the crime spot quickly to avoid being caught. Another difference is the reconstructed pattern of the temporal response function. In Figure 4 in Mohler et al. (2011), there might be some periodicity in the marginal temporal response function, while our reconstructed one is monotone decreasing. A possible cause of this difference is that periodicity in the background is not considered in Molher's model.

## 5.2 | Diagnostics of the model: Residual analysis

One must keep in mind that it is difficult to find an ideal model for the observations at the beginning stage of the modeling. Thus, finding the advantages and the shortcomings of the current model is important for improving the model formulation. Thus, after fitting a model to some observational data, we may ask some questions about the results. For example: (1) How to justify the goodness-of-fit of the model? (2) Does the data patterns vary with space and time? (3) How to improve the model formulation? Zhuang (2006) summarized the ideas of the residual analysis technique and provided some examples of finding the possible direction for improving the formulation of the Epidemic Type Aftershock Sequence (ETAS) model, which is widely used for analyzing, modeling and forecasting regional seismicity (Ogata, 1998; Zhuang et al., 2002). In this section, we carry out residual analysis to answer several questions related to the data.

### Transformed time sequence analysis

Traditionally, residual analysis is usually done in the following way. Given a point process $N = \{(t_i, x_i, y_i), i = 1, 2, \cdots, n\}$, which is determined by a conditional intensity $\lambda(t)$, the following transformation

$$t_i \;\to\; \tau_i = \int_0^{t_i} \int_S \lambda(u, x, y)\, \mathrm{d}x \mathrm{d}y\, \mathrm{d}u \tag{49}$$

transforms $N$ into a stationary Poisson process with a unit rate (standard Poisson process), namely, $N' = \{\tau_i : i = 1, 2, \cdots, n\}$. The process $N'$ is called the transformed time sequence (e.g., Ogata, 1988). The true $\lambda(t, x, y)$ is always unknown in real data analysis. If we replace $\lambda(t, x, y)$ by $\hat{\lambda}(t, x, y)$, which is a good approximation of the true model, in the above equation, we can also obtain a transformed time sequence that is approximately a Poisson process of rate 1 (the standard Poisson process). Thus, we can conclude that the model does not fit the data well unless the transformed time sequence deviates significantly from the standard Poisson process.

Confidence bands of the transformed time sequence have been studied by Ogata (1988, 1989). In this study, this problem is treated from another viewpoint: since such a transformed time sequence is a standard Poisson process for an ideal model, statistics related to the Poisson process can be used to construct the confidence band. Following Schoenberg (2002), the cumulative frequency curve ($\hat{\tau}_i = \int_0^{t_i} \int_S \hat{\lambda}(u, x, y)\,\mathrm{d}x\mathrm{d}y\,\mathrm{d}u,\ i$) always connects $(0,\ 0)$ and $(T,\ n)$, where $\hat{\lambda}(u, x, y)$ is the model estimated from the data in $[0,\ T]$ by using the maximum likelihood estimate and $n = N([0, T]\times S)$. For each positive integer $k$, if $k < n$, the confidence interval for $\tau_k$ is the same as $kZ$, where $Z$ is a random variable that obeys a beta distribution with parameter $(k+1, n-k+1)$; when $k > n$, $\tau_k$ can be approximated by a gamma distribution with a shape parameter $k - n$ and scale parameter 1. Here we refer Schoenberg (2002) for details.

The transformed time sequence for the analyzed data is plotted in Figure 4. Transformation in (49) approximately transforms the crime events into a stationary one. Around the transformed times of 530, 2400, 3250, and 4300, there seems to be some change point of the occurrence rate in the transformed time domain. This might be caused by the fact that kernel estimation is a bit over-smooth in detecting the change points of the long term background occurrence rate.

**Does the daily periodicity change in time or in space?**

To understand whether the daily periodicity pattern changes in time, we reconstruct the daily periodicity functions for each individual year of 2012 and 2013, as shown in Figure 5(a). We note that there are not significant differences between these two years. Similarly we reconstruct the daily periodicity for different seasons, different days of the week, and different areas in the city, as shown in Figures 5(b) to (d), respectively. These results do not show much difference among different seasons. The biggest difference is the effect of the days of the week. From Figure 5(c) we can see that the daily effect for Sundays is quite flat, a valley around 4am and two peaks around 1pm and 9pm. There exists slight differences in the city center and the suburb area (Figure 5(d)): the occurrence rate is relatively higher at noon and evening and relatively lower in the early morning and in the afternoon in the center of the city than in the other areas.

**Does the weekly periodicity change in time or in space?**

We reconstruct the week periodicity for different years (Figure 6(a)) and different areas (Figure 6(b)). The results do not show much differences of weekly periodicity between years. However, the occurrence rate in the city center area gets much higher on Fridays.

**Does the long-term trend differ in different places?**

Figure 7 shows the reconstructed long-term trend component of the background rate. Even though there are two small rebounds about 420 and 640 days, the long-term background rate in the city center area decreases quicker in those two years than in the other suburb areas. Moreover, there is almost no difference among different suburb areas.

**Is the background rate separable in space and time?**

In the model formulation, we have assumed that the background rate is separable in space and time. We reconstruct $\mu_\mathrm{b}(x, y)$ for the years of 2012 and 2013, namely $\hat{\mu}_{b,'12}$ and $\hat{\mu}_{b,'13}$, respectively, and plot their difference ($\hat{\mu}_{b,'13} - \hat{\mu}_{b,'12}$) in Figure 8(a). For an easier comparison, we also plot the relative difference, $(\hat{\mu}_{b,'13} - \hat{\mu}_{b,'12})/\hat{\mu}_b$, in Figure 8(b), where $\hat{\mu}_b$ is the estimate in the model for the entire period. We see from these results that, even though it exists, the difference between $\hat{\mu}_{b,'13}$ and $\hat{\mu}_{b,'12}$ is negligible and that the assumption that the background rate is separable in space and time is reasonable.

**Is the clustering effect different in different places?**

It is also interesting to know whether the clustering effect differs between downtown and the suburb areas. A simple verification is to check whether the reconstructed $g(t)$ and $f(x, y)$ are different for the city center and other areas. These functions are plotted in Figures 9 and 10. The overall shapes of $g$ and $f$ are similar for the city center area and the suburb area. Taking into consideration the fact that there are not many triggered events, only less than 3% among all the events, for our estimation of these functions, it is not necessary to assume different temporal and spatial response functions for the city center and the suburb area, which might complicate our analysis. This also implies that our choice of using separable temporal and spatial response functions in our model (15) is reasonable.

Since the triggering effect is weak, we do not carry out the analysis of whether $f$ and $g$ vary in different time period.

## 6 | CONCLUSIONS AND DISCUSSION

In this study, we have proposed a spatiotemporal Hawkes model, whose background rate includes a long-term trend and periodicity, to describe the robbery-related violence in Castellon, Spain. To estimate the model, a semi-parametric method is used to reconstruct the background and clustering components and to estimate their relative contributions. Comparing with previous studies, we have introduced the periodic terms in the background rate and estimated them through kernel estimates.

The new stochastic reconstruction method developed in this study fits better to crime data and is simple to understand and to estimate, without requiring much prior knowledge of the studied phenomena. Using this method, we have analyzed and highlighted the existence of periodic components and the triggered effect in the process of the studied crime phenomena. In the estimation procedures of the background components and the excitation response functions, two relaxation parameters are adopted to stabilize and fasten the convergence.

The final results show the following features of the behaviors of the robbery-related violence in Castellon: (1) Background dominates the whole process while the clustering effect only contributes about 3%. (2) The periodicity effect is strong in the background. (3) Residual analysis shows that crime activity is different during weekends from working days. (4) Downtown has different characteristics in crime activities from suburb regions.

There are various possible ways of extending this research in the future. Here we list several possibilities. (1) We could consider the nonlinear Hawkes process (e.g.,Brémaud and Massoulié, 1996; Delattre et al., 2016; Torrisi, 2016, 2017; Zhu, 2013, 2014, 2015; Chevallier et al., 2018), whose temporal version has a conditional intensity in the form of

$$\lambda(t) = \Phi\left(\int_{-\infty}^{t_-} g(t-u)\, N(\mathrm{d}u)\right), \tag{50}$$

where $\Phi$ is a locally integrable and left-continuous nonnegative function. (2) In this study, we used kernel estimates with fixed bandwidths to obtain all the components in the model formulation. Also, in the comparison among results in Table 1, the model complexity is not accounted for. It is worthwhile to apply cross-validation to obtain the optimal bandwidths and to select the model that best fits the data. (3) Other nonparametric estimates, such as Bayesian procedures with smoothness priors, tessellation methods, etc., can be also incorporated into the proposed method. Careful and detailed comparisons should be done among these methods in order to find the best one for practical forecasting.

## ACKNOWLEDGEMENTS

The authors thank the Associate editor and two anonymous reviewers for their helpful comments and constructive suggestions.

# REFERENCES


Bartlett, M. S. (1963) The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, **25**, 264–296. URL: http://www.jstor.org/stable/2984295.

Bowers, K. J., Johnson, S. D. and Pease, K. (2004) Prospective hot-spottingthe future of crime mapping? *The British Journal of Criminology*, **44**, 641–658. URL: +http://dx.doi.org/10.1093/bjc/azh036.

Brémaud, P. and Massoulié, L. (1996) Stability of nonlinear hawkes processes. *The Annals of Probability*, **24**, 1563–1588. URL: http://www.jstor.org/stable/2244985.

Caplan, J. M. and Kennedy, L. W. (2016) *Risk Terrain Modeling: Crime Prediction and Risk Reduction*. Berkeley, CA, USA: University of California Press.

Chevallier, J., Duarte, A., Löcherbach, E. and Ost, G. (2018) Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, In press. URL: https://www.sciencedirect.com/science/article/abs/pii/S030441491830022X.

Cohen, J., Gorr, W. L. and Olligschlaeger, A. M. (2007) Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, **39**, 105–127. URL: http://dx.doi.org/10.1111/j.1538-4632.2006.00697.x.

Daley, D. D. and Vere-Jones, D. (2003) *An Introduction to Theory of Point Processes – Volume 1: Elementrary Theory and Methods (2nd Edition)*. New York, NY: Springer.

Delattre, S., Fournier, N. and Hoffmann, M. (2016) Hawkes processes on large networks. *Ann. Appl. Probab.*, **26**, 216–261. URL: https://doi.org/10.1214/14-AAP1089.

Felson, M. and Boba, R. (2010) *Crime and Everyday Life*. SAGE Publications. URL: https://books.google.co.jp/books?id=TI6xdKDLwtcC.

Hawkes, A. G. (1971a) Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **33**, 438–443. URL: http://links.jstor.org/sici?sici=0035-9246\%281971\%2933\%3A3\%3C438\%3APSOSME\%3E2.0.CO\%3B2-G.

— (1971b) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**, 83–90. URL: http://biomet.oxfordjournals.org/cgi/content/abstract/58/1/83.

Hawkes, A. G. and Oakes, D. (1974) A cluster process representation of a self-exciting process. *Journal of Applied Probability*, **11**, 493–503. URL: http://links.jstor.org/sici?sici=0021-9002\%28197409\%2911\%3A3\%3C493\%3AACPROA\%3E2.0.CO\%3B2-7.

Johnson, N., Hitchman, A., Phan, D. and Smith, L. (2018) Self-exciting point process models for political conflict forecasting. *European Journal of Applied Mathematics*, **29**, 685–707.

Levine, N. (2017) *CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents*, 381–388. Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-319-17885-1_229.

Lewis, P. A. W. (1964) A branching poisson process model for the analysis of computer failure patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**, 398–456. URL: http://www.jstor.org/stable/2984497.

Marsan, D. and Lengliné, O. (2008) Extending earthquakes' reach through cascading. *Science*, **319**, 1076–1079. URL: http://science.sciencemag.org/content/319/5866/1076.

Mohler, G. (2014) Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, **30**, 491 – 497. URL: http://www.sciencedirect.com/science/article/pii/S0169207014000284.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011) Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, **106**, 100–108. URL: http://dx.doi.org/10.1198/jasa.2011.ap09546.

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L. and Brantingham, P. J. (2015) Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, **110**, 1399–1411. URL: http://dx.doi.org/10.1080/01621459.2015.1077710.

Neyman, J. E. and Scott, E. L. (1953) Frequency of separation and interlocking of clusters of galaxies. *Proceedings of the National Academy of Sciences of the United States of America*, **39**, 737–743.

— (1958) A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B (Methodological)*, **20**, 1–43.

Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**, 9–27. URL: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560.

— (1989) Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, **169**, 159–174. URL: //www.sciencedirect.com/science/article/pii/0040195189901911.

— (1998) Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**, 379–402. URL: http://dx.doi.org/10.1023/A:1003403601725.

Ratcliffe, J. H. (2004) The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. *Police Practice and Research*, **5**, 5–23. URL: https://doi.org/10.1080/1561426042000191305.

Reinhart, A. (2018) A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, early view available.

Reinhart, A. and Greenhouse, J. (2018) Self-exciting point process with spatial covariates: modeling the dynamics of crime. *Journal of the Royal Statistical Society*, submitted.

Rodrigues, A. and Diggle, P. J. () Bayesian estimation and prediction for inhomogeneous spatiotemporal log-gaussian cox processes using low-rank models, with application to criminal surveillance. *Journal of the American Statistical Association*.

Rosser, G. and Cheng, T. (2016) Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. *Applied Spatial Analysis and Policy*. URL: https://doi.org/10.1007/s12061-016-9198-y.

Schoenberg, F. (2002) On rescaled Poisson processes and the Brownian bridge. *Annals of the Institute of Statistical Mathematics*, **54**, 445–457. URL: http://dx.doi.org/10.1023/A%3A1022494523519.

Shirota, S. and Gelfand, A. E. (2017) Space and circular time log gaussian cox processes with application to crime event data. *Ann. Appl. Stat.*, **11**, 481–503. URL: https://doi.org/10.1214/16-AOAS960.

Torrisi, G. L. (2016) Gaussian approximation of nonlinear hawkes processes. *Ann. Appl. Probab.*, **26**, 2106–2140. URL: https://doi.org/10.1214/15-AAP1141.

— (2017) Poisson approximation of point processes with stochastic intensity, and application to nonlinear Hawkes processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **53**, 679–700. URL: https://doi.org/10.1214/15-AIHP730.

Townsley, M., Homel, R. and Chaseling, J. (2003) Infectious burglaries. a test of the near repeat hypothesis. *The British Journal of Criminology*, **43**, 615–633. URL: +http://dx.doi.org/10.1093/bjc/43.3.615.

Zhu, L. (2013) *Nonlinear Hawkes Processes*. Ph.D. thesis, Department of Mathematics New York University, New York.

— (2014) Process-level large deviations for nonlinear Hawkes point processes. *Annales de l'I.H.P. Probabilités et statistiques*, **50**, 845–871. URL: http://www.numdam.org/item/AIHPB_2014__50_3_845_0/.

— (2015) Large deviations for Markovian nonlinear Hawkes processes. *Ann. Appl. Probab.*, **25**, 548–581. URL: `https://doi.org/10.1214/14-AAP1003`.

Zhuang, J. (2006) Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 635–653.

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

— (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, **109**, B05301.

Zhuang, J., Werner, M. J. and Harte, D. S. (2013) Stability of earthquake clustering models: Criticality and branching ratios. *Phys. Rev. E*, **88**, 062109. URL: `http://link.aps.org/doi/10.1103/PhysRevE.88.062109`.

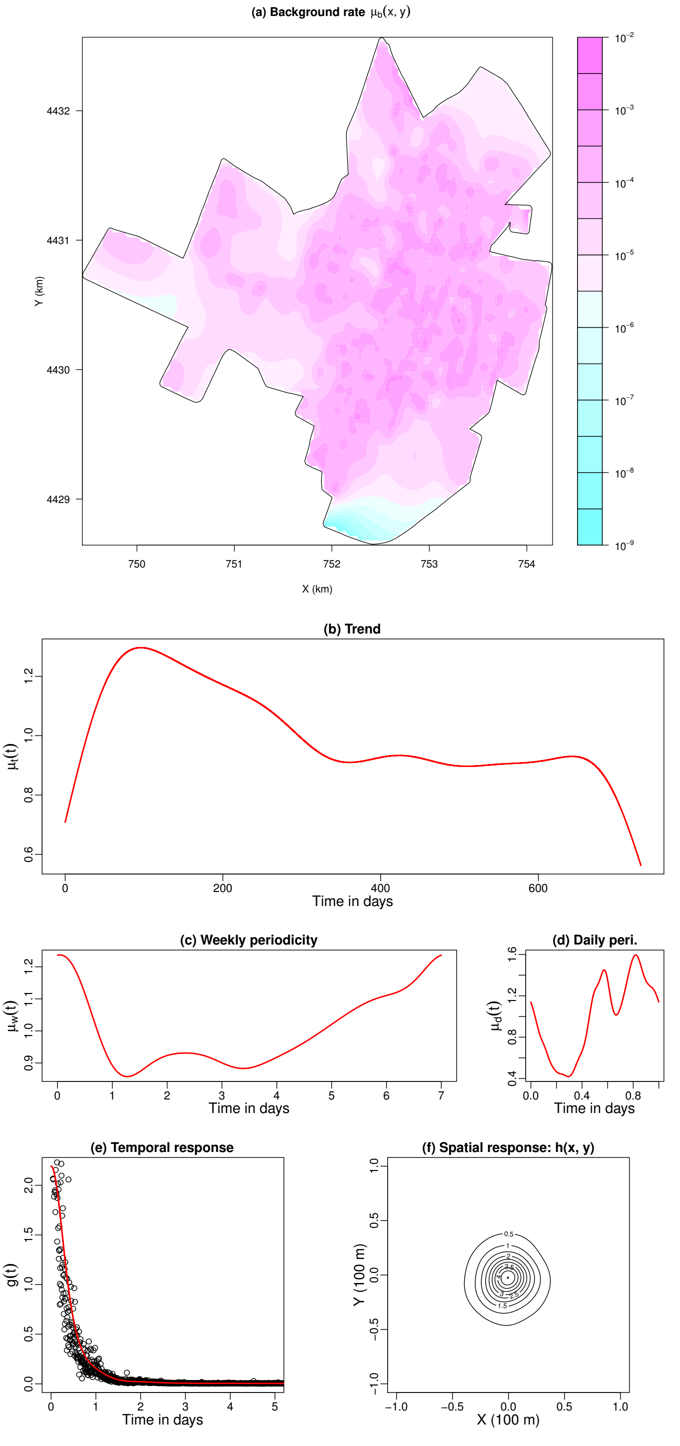


**FIGURE 3** Output results: (a) spatial background rate $\mu_b(x, y)$ (b) trend function, (c) weekly periodicity, (d) daily periodicity, (e) temporal response function, and (f) spatial response function.

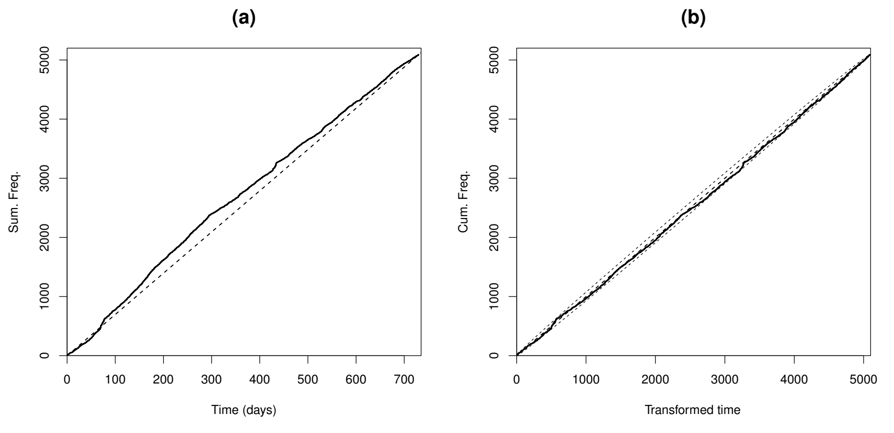


**FIGURE 4** Cumulative frequencies of crime events versus (a) original occurrence times and (b) transformed times. The slopes of the dashed straight lines in both panels represent the average occurrence rates. The dashed curves in (b) mark the 95% confidence bands for the transformed time sequence.

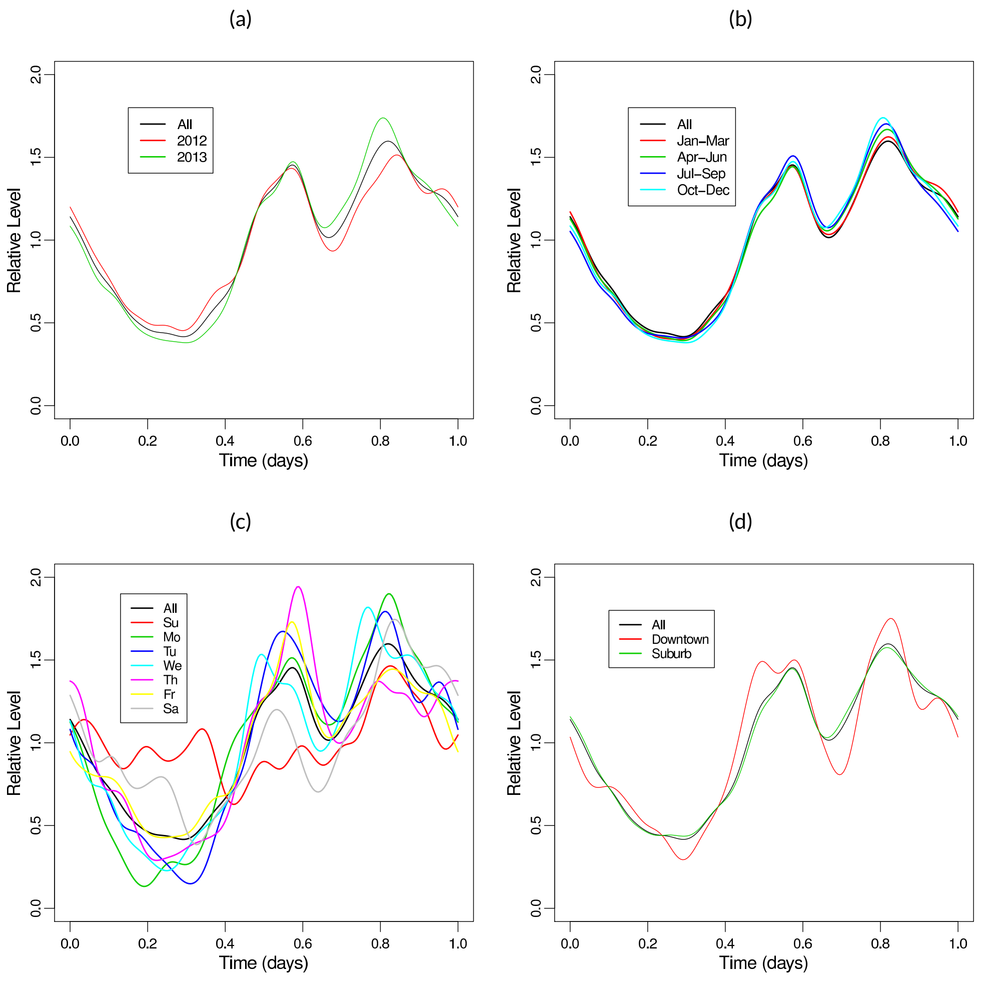


**FIGURE 5** Reconstructed daily periodicity functions $\hat{\mu}_{\mathrm{d}}(t)$ for (a) different years of 2012 and 2013, (b) different seasons, (c) different days of the week, and (d) different areas of the city-center and the suburb.

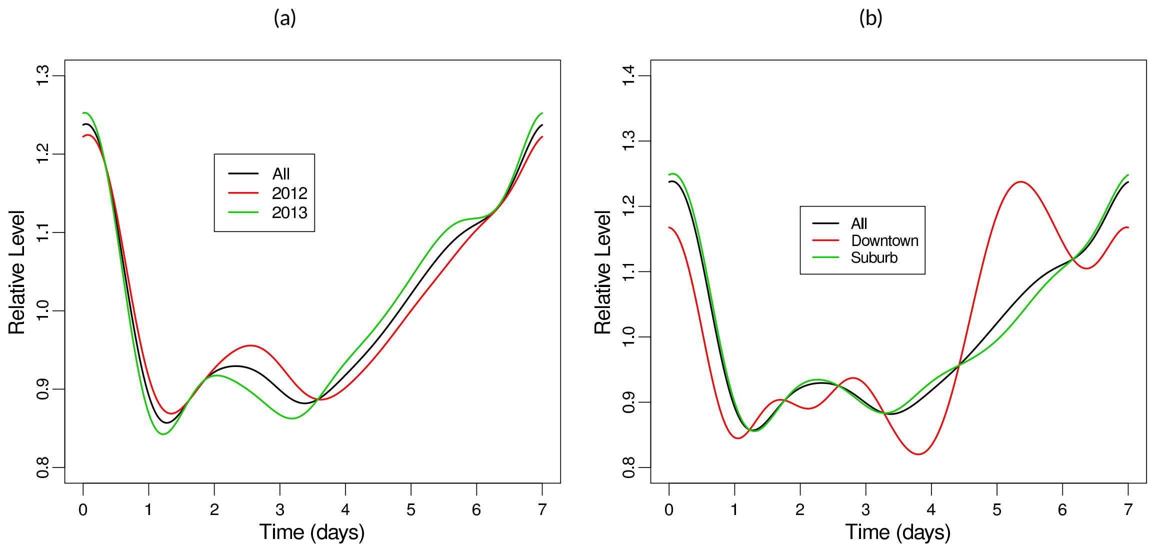


**FIGURE 6** Reconstructed weekly periodicity functions (a) for different years and (b) for city center and suburb areas.

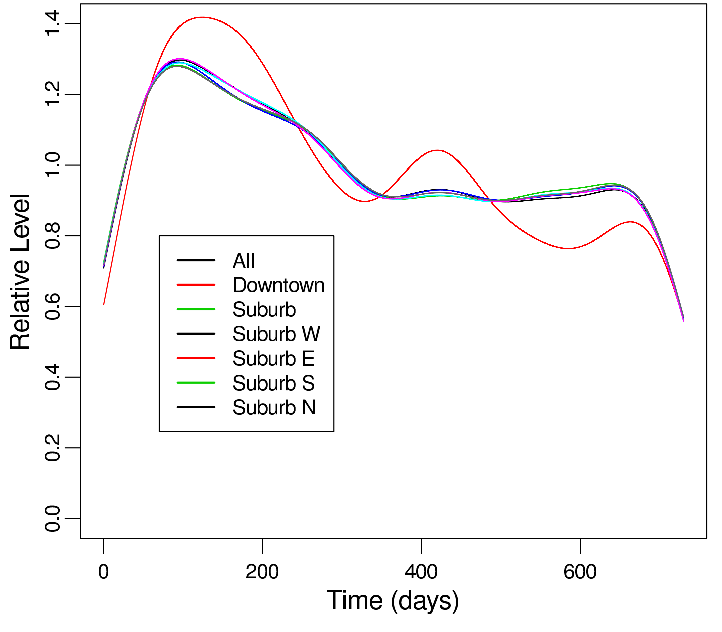


**FIGURE 7** Reconstructed long-term trend for different areas.

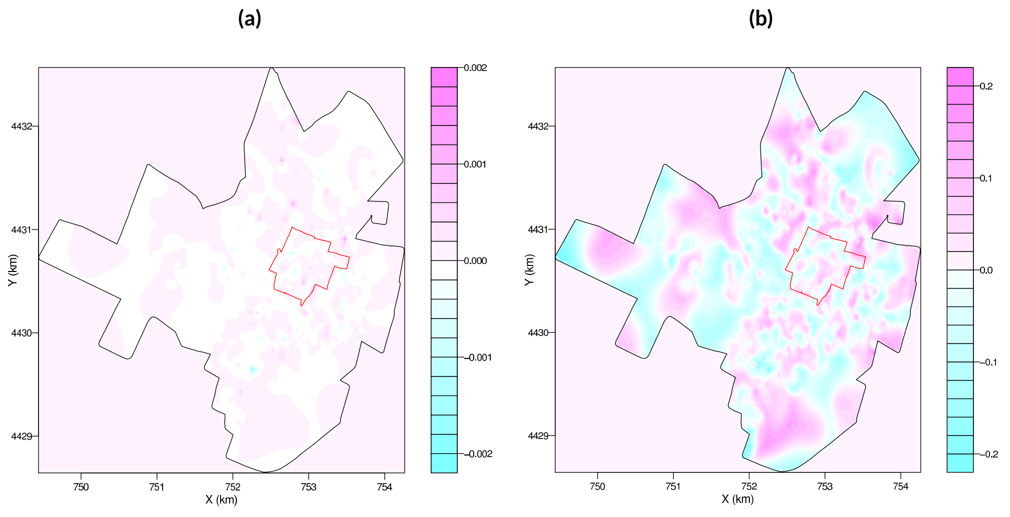


**FIGURE 8** Diagnostics of space-time separability of background rate: (a) Absolute difference between the reconstructed background rates estimated by using data from 2012 and 2013 (the latter minus the former) and (b) relative difference between them (the latter minus the former then divided by the background rate for the entire dataset).

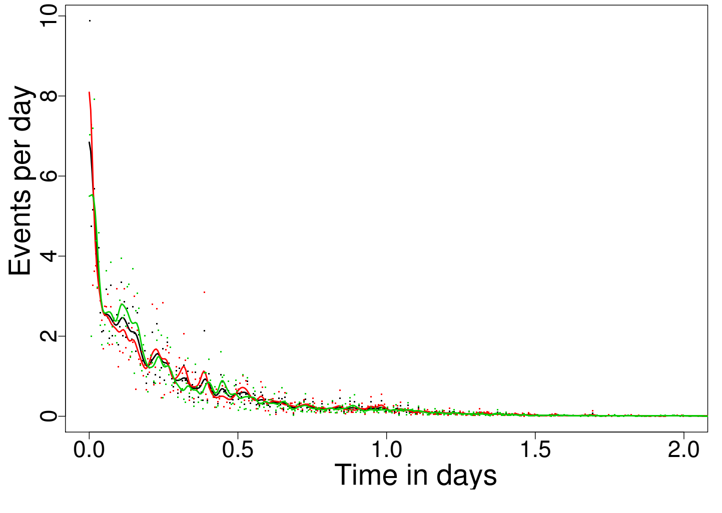


**FIGURE 9** Reconstructed temporal response of the triggering effect, $\hat{g}(t)$. The black, red and green curve are for all the region, the city center area, and the suburb areas, respectively.

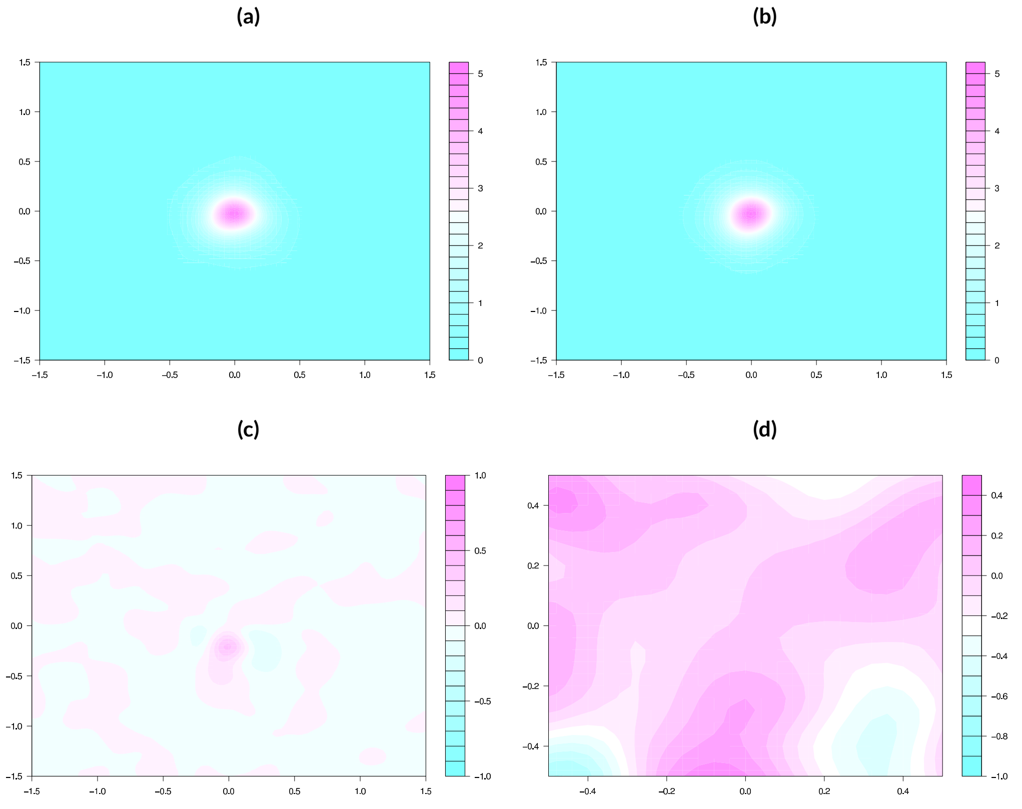


**FIGURE 10** Diagnostics of regional difference of the spatial response between city center and suburb areas. (a) Reconstructed $f$ for the city center area. (b) Reconstructed $f$ for the suburb areas. (c) Related difference between the spatial response function in (a) and (b). (d) Absolute difference between the spatial response function in (a) and (b).