



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

Jiancang Zhuang

*Research Organisation of Information and Systems, Tokyo, Graduate University for Advanced Studies, Tokyo, Japan, and London Mathematical Laboratory, UK*

and Jorge Mateu

*University Jaume I, Castellón, Spain*

[Received January 2018. Revised October 2018]

**Summary.** 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 non-parametric 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 secure the estimation process. This model is used to describe the occurrences of violence or robbery cases in Castellón, Spain, during 2 years. The results show that robbery crime is highly influenced by 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 shows that the model could be improved by considering the following ingredients: the daily occurrence patterns are different between weekends and working days; in the city centre, robbery activity shows different temporal patterns, in both weekly periodicity and long-term trend, from other suburb areas.

**Keywords:** Crime; Edge effect correction; Hawkes process; Kernel estimate; Periodicity; Spatiotemporal point process; Stochastic reconstruction

## 1. Introduction

Point process modelling 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 or communication network failures. Depending on the type of the domain where the events occur, point process models are classified into two classes: spatial point processes and spatiotemporal or temporal point processes. The difference between these two types of model is that the latter has 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.

*Address for correspondence:* Jiancang Zhuang, Institute of Statistical Mathematics, Research Organisation of Information and Systems, Tokyo 190-8562, Japan.  
E-mail: zhuangjc@ism.ac.jp

Among the different types of point process, clustered point processes have attracted much interest by 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 Bartlett–Lewis process to model the rainfall process (Bartlett, 1963; Lewis, 1964). Many spatiotemporal or temporal clustered point processes can be categorized into the Hawkes self-exciting process (Hawkes, 1971a, b; Hawkes and Oakes, 1974), including the epidemic-type aftershock sequence model for earthquake occurrence (e.g. Ogata (1988, 1989, 1998) and 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 non-stationary in time, and a triggered subprocess composed of 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, the expectation–maximization algorithm, first- and higher order residuals, and Bayesian analysis, as well as the theories that are associated with the asymptotic properties (see the review by Reinhart (2018))

The most common tools to predict crimes include ‘hotspotting’ (e.g. Bowers *et al.* (2004), Ratcliffe (2004) and 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 hotspotting 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 or when the risk decays back to the baseline level. Leading indicator regression looks for covariates that can be used as local risk indicators of future serious crimes. Risk terrain modelling identifies the risks that come from particular features of a landscape and models how they co-locate to create unique behaviour settings for crime.

Hawkes-type point process modelling 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, the model of Mohler and co-workers incorporates the time varying hotspots 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, non-parametric and semiparametric estimation methods for the Hawkes model have been developed. Marsan and Lengliné (2008) made use of the stochastic declustering technique that was proposed by Zhuang *et al.* (2002, 2004) and Zhuang (2006) and proposed the so-called model-independent stochastic declustering method, which is a non-parametric estimation method of an epidemic-type aftershock sequence type of model (Ogata, 1988, 1998) for earthquake occurrence. This method was introduced at the same time when point process modelling was used for analysing crime data for the first time (Mohler *et al.*, 2011) followed by improvements from other researchers (e.g. Johnson *et al.* (2018)). In a parallel line, several researchers 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 (2012)).

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 behaviours should be controlled by their biological clock and could be influenced by the periodic activity of 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 between crime events were not counted.

The aim of this study is to analyse crime data by using an extended semiparametric Hawkes model. Differently 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 behaviour 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 model proposed semiparametric since these two relaxation parameters can be estimated by using maximum likelihood.

This paper is organized as follows. Section 2 gives a brief description of the data. Section 3 provides the concepts and statistical modelling methodologies that are 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.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

<https://rss.onlinelibrary.wiley.com/hub/journal/1467985x/series-a-datasets>

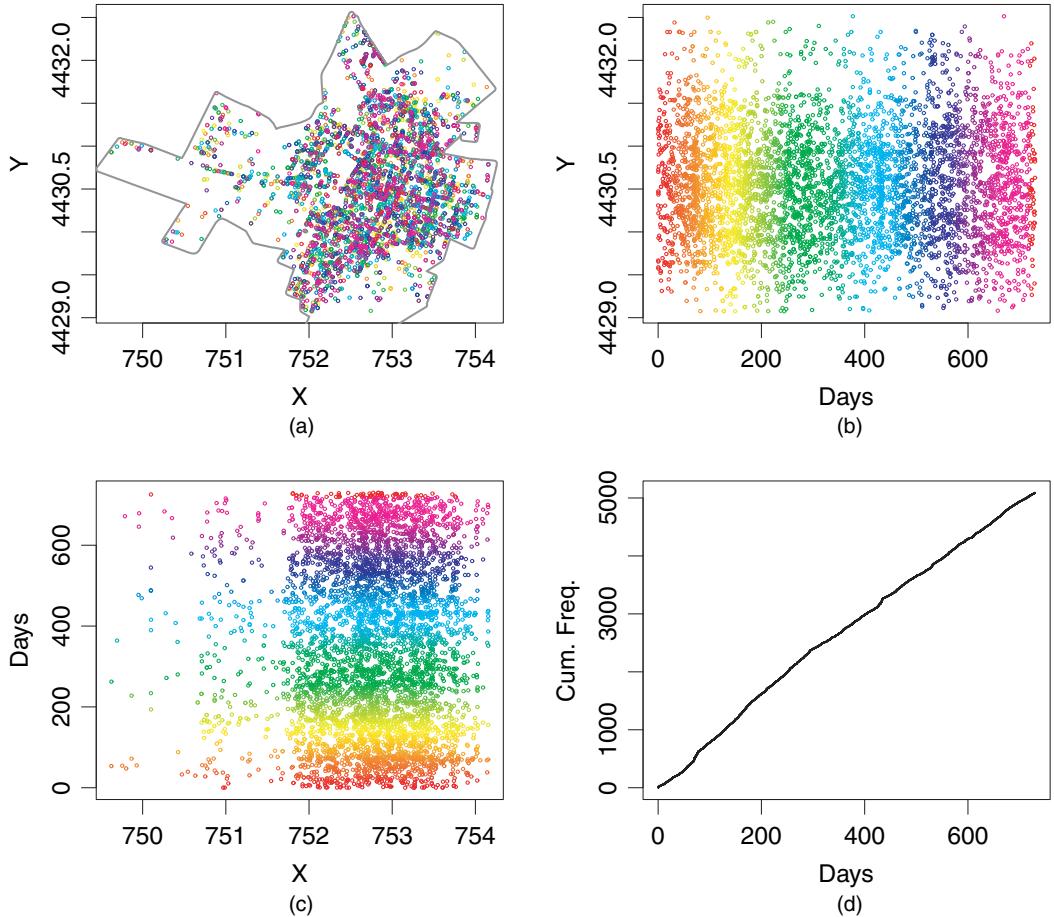
## 2. Data

In this study, we analyse the robbery-related violence data in Castellón city, Spain, during the years of 2012 and 2013. The data report georeferenced co-ordinates of phone calls received by the police station in the city of Castellón from January 2012 to December 2013. Castellón is a Mediterranean city of around 180000 inhabitants. The listed calls were received at the local police call centre or transferred by the '112' emergency service to the local police call centre. Geocodification was performed indirectly by local officials based on precise address information provided by the callers. The calls comprise up to nine different types of crime or antisocial behaviour categories, but we here focus only on robbery-related violence data, comprising a total number of 5089 events happening in the streets of Castellón. The city of Castellón is divided into 108 census tracts with an overall surface of 108.6 km<sup>2</sup>. Figs 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 analysing.

## 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



**Fig. 1.** Basic information of robbery-related violence in Castellón, Spain, 2012–2013 (the rainbow colours show the occurrence times of the events, with red-coloured points representing the earliest events and magenta the latest): (a) spatial locations; (b)  $y$ - $t$  co-ordinates; (c)  $t$ - $x$  co-ordinates; (d) cumulative numbers against times

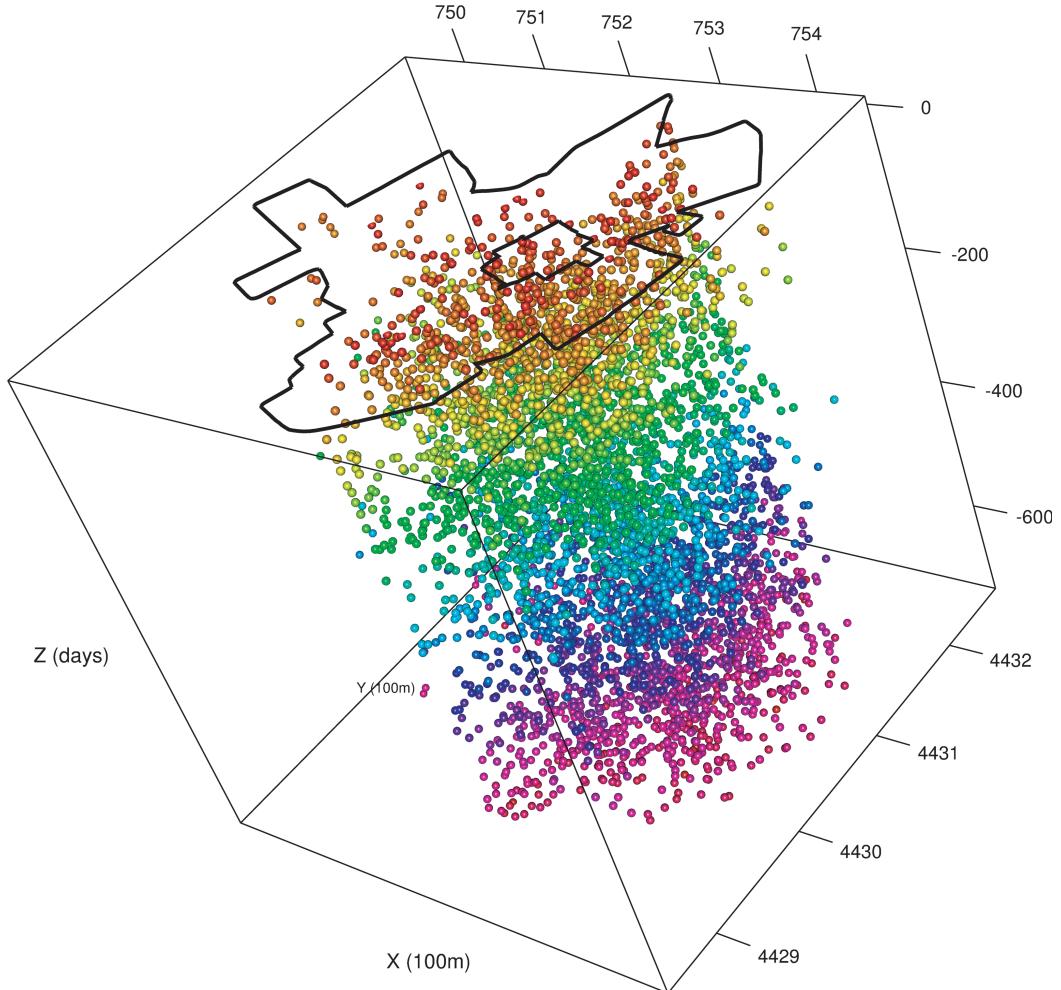
defined by its conditional intensity. For the purely temporal case, the conditional intensity is defined by

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

where  $\mathcal{H}_t$  denotes the  $\sigma$ -algebra that is 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)

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

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



**Fig. 2.** Three-dimensional plot of robbery-related violence in Castellón, Spain, 2012–2013 (the rainbow colours show the occurrence times of the events, with red-coloured points representing the earliest events and magenta the latest)

number of direct offspring per ancestor, is given by

$$\rho = \int_0^\infty g(u)du. \quad (3)$$

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

$$\lambda(t, x) = \mu(x) + \int_{\mathbb{R}^d \times (-\infty, t]} g(t-s, x-u) N(ds \times du) \quad (4)$$

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 be generalized to the multivariate case where, if we have  $K$  types of event in total, each type

has a conditional intensity

$$\lambda_k(t, x) = \mu_k(x) + \sum_{l=1}^K \int_{\mathbb{R}^d \times (-\infty, t_-)} g_{l,k}(t, x; s, u) N_l(ds \times du), \quad (5)$$

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

Given observation data of crime events in an observational space–time window  $S \times T$ , for a parametric Hawkes model, we can use maximum likelihood estimation to estimate the model parameters, i.e.

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \log \{L(\cdot; \theta)\} \\ &= \arg \max_{\theta} \left[ \sum_{i:(t_i, x_i, y_i) \in S \times T} \log \{\lambda(t, x, y; \theta)\} - \int_T \int_S \lambda(t, x, y; \theta) dx dy dt \right]. \end{aligned} \quad (6)$$

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

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

where  $\mu(t, x)$  is the background rate, which is different from the corresponding term in equation (4) as it allows for it to be time dependent, and  $g(t, x)$  is the rate of occurrence 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. the *background probability*, is given by

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

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

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

It is easy to see that

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

Another explanation for this equation is that, once an event has occurred at  $(t, x)$ , we can say that at  $(t, x)$  we have observed  $\varphi_j$  background events and that, for each  $i = 1, \dots, j-1$ , 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 non-parametric way to estimate functions  $\mu(\cdot, \cdot)$  and  $g(\cdot, \cdot)$ . For example,  $g(\cdot, \cdot)$  can be estimated by

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

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, for example, a weighted kernel estimation as follows:

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

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

In the above analysis, 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  $\{(t_i, x_i) : i = 1, \dots, n\}$  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  $i$  and  $j$ . Then we estimate the background rate  $\mu$  and each component in the clustering part  $g$  by using  $\varphi_i$  and  $\rho_{ij}$ , through some non-parametric methods, e.g. kernel estimation or histogram. Once  $\mu$  and  $g$  have been updated, we go back to the step of calculating  $\varphi$ , or we stop if convergence has been 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). Mohler *et al.* (2011) introduce it for the analysis of crime data. It is worthwhile to mention that in the Marsan–Lengliné algorithm, Marsan and Lengliné (2008) assumed that  $g$  is a stepwise constant function and the maximum likelihood estimate yields a histogram. In the Marsan–Lengliné 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) analysed the break-in burglary data from the Los Angeles Police Department. Their data set 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

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

In Mohler *et al.* (2011), the background rate was 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 paper, we improve the above algorithm by

- (a) introducing relaxing parameters and
- (b) considering periodic components in the background rates.

### 3.4. Model formulation

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

$$\lambda(t, x, y) = \mu_t(t)\mu_d(t)\mu_w(t)\mu_b(x, y) + \int_{-\infty}^{t_-} \iint_S g(t - s, x - u, y - v) N(du \times dv \times ds), \quad (14)$$

where  $\mu_t(t)$ ,  $\mu_d(t)$  and  $\mu_w(t)$  represent the trend term, the daily periodicity and the weekly periodicity in the temporal components of the background rate respectively,  $\mu_b(x, y)$  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  $g$  non-parametrically by using the stochastic reconstruction method that was proposed in Zhuang (2006). First, we rewrite the conditional intensity as

$$\lambda(t, x, y) = \mu_0 \mu_t(t) \mu_d(t) \mu_w(t) \mu_b(x, y) + A \int_{-\infty}^{t_-} \iint_S g(t-s, x-u, y-v) N(du \times dv \times ds), \quad (15)$$

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 probability density functions, i.e.  $\int_0^\infty g(s) ds = 1$ , and  $\iint_S h(u, v) du dv = 1$ . Here we separate the spatiotemporal clustering response function into a temporal and a spatial component to avoid the non-parametric estimation of a three-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 that was developed in Zhuang (2006) to solve this problem. The key point of residual analysis for temporal or 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

$$\mathbf{E} \left[ \int_{[T_1, T_2] \times S} f(t, x) dN(dt \times dx) \right] = \mathbf{E} \left[ \int_{T_1}^{T_2} \int_S f(t, x) \lambda(t, x) dt dx \right], \quad (16)$$

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

##### 4.1. Reconstructing background components

Given a realization of the point process  $\{(t_i, x_i, y_i) : i = 1, 2, \dots, n\}$  in a time-space range  $[T_1, T_2] \times S$ , where  $t$  (days) and  $(x, y)$  (kilometres) 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 equation (16). Then, assuming that  $\mu_t$  is sufficiently smooth,

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

$$\begin{aligned} &\propto \int_{t-\Delta_t}^{t+\Delta_t} \mu_t(s) ds \\ &\approx 2\mu_t(t)\Delta_t, \end{aligned} \quad (17)$$

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

$$w_i^{(t)} = \mu_t(t_i)\mu_b(x_i, y_i)/\lambda(t_i, x_i, y_i); \quad (18)$$

then

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

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

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

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

and

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

where

$$w_i^{(d)} = \mu_d(t_i)\mu_b(x_i, y_i)/\lambda(t_i, x_i, y_i), \quad (23)$$

$$w_i^{(w)} = \mu_w(t_i)\mu_b(x_i, y_i)/\lambda(t_i, x_i, y_i), \quad (24)$$

$$\varphi_i = \mu_0 \mu_t(t_i) \mu_d(t_i) \mu_w(t_i) \mu_b(x_i, y_i) / \lambda(t_i, x_i, y_i), \quad (25)$$

and  $\Delta_t$ ,  $\Delta_x$  and  $\Delta_y$  are small positive numbers. In these equations the rescaled weights  $w_i^{(t)}$ ,  $w_i^{(d)}$  and  $w_i^{(w)}$  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 first

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

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

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

In the last step of this 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

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

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  $\mu_t$ ,  $\mu_d$ ,  $\mu_w$ ,  $\mu_b$ ,  $g$  and  $h$  have been estimated, we can update the relaxation coefficients  $\mu_0$  and  $A$  through maximizing the likelihood function:

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

Denote

$$U = \int_0^T \iint_S \mu_t(t) \mu_d(t) \mu_w(t) \mu_b(x, y) dx dy dt$$

and

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

The equations  $\partial \log(L) / \partial \mu_0 = 0$  and  $\partial \log(L) / \partial A = 0$  give

$$\sum_{i=1}^n \frac{\mu_t(t_i) \mu_d(t_i) \mu_w(t_i) \mu_b(x_i, y_i)}{\lambda(t_i, x_i, y_i)} - U = 0, \tag{32}$$

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

These equations can be solved by the following iteration system:

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

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

where

$$\varphi_i^{(k)} = \frac{\mu_0^{(k)} \mu_t(t_i) \mu_d(t_i) \mu_w(t_i) \mu_b(x_i, y_i)}{\mu_0^{(k)} \mu_t(t_i) \mu_d(t_i) \mu_w(t_i) \mu_b(x_i, y_i) + A^{(k)} \sum_{j:t_j < t_i} g(t_j - t_i) h(x_j - x_i, y_j - y_i)}. \quad (36)$$

#### 4.4. Smoothing estimates and correcting for edge effects

To obtain 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, equations (19)–(22), (28) and (30) become

$$\hat{\mu}_t(t) \propto \sum_i w_i^{(t)} Z(t - t_i; \omega_t), \quad (37)$$

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

$$\hat{\mu}_w(t) \propto \sum_i w_i^{(w)} \sum_{k=0}^{\lfloor T/7 \rfloor} Z(t - t_i + 7\lfloor t_i/7 \rfloor - 7k; \omega_d), \quad (39)$$

$$\hat{\mu}_b(x, y) \propto \sum_i \varphi_i Z(x - x_i; \omega_x) Z(y - y_i; \omega_y), \quad (40)$$

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

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

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 ‘?’ for the new estimates.

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

$$\hat{\mu}_t(t) \propto \sum_i w_i^{(t)} \frac{Z(t - t_i; \omega_t)}{\int_0^T Z(u - t_i; \omega_t) du}, \quad (43)$$

$$\hat{\mu}_d(t) \propto \sum_i w_i^{(d)} \frac{\sum_{k=0}^T Z(t - t_i + \lfloor t_i \rfloor - k; \omega_d)}{\int_0^T Z(u - t_i; \omega_d) du}, \quad (44)$$

$$\hat{\mu}_w(t) \propto \sum_i w_i^{(w)} \frac{\sum_{k=0}^{\lfloor T/7 \rfloor} Z(t - t_i + 7\lfloor t_i/7 \rfloor - 7k; \omega_w)}{\int_0^T Z(u - t_i; \omega_w) du}, \quad (45)$$

$$\hat{\mu}_b(x, y) \propto \sum_i \varphi_i \frac{Z(x - x_i; \omega_x) Z(y - y_i; \omega_y)}{\iint_S Z(u - x_i; \omega_x) Z(v - y_i; \omega_y) du dv}, \quad (46)$$

$$\hat{g}(t) \propto \frac{\sum_{i,j} \rho_{ij} \times Z(t - t_j + t_i; \omega_g) \left/ \int_0^{T-t_i} Z(u - t_j; \omega_g) du \right.}{\sum_i I(t_i + t \leq T)}, \quad (47)$$

$\hat{h}(x, y)$

$$\propto \frac{\sum_{i,j} \rho_{ij} \times Z(x - x_j + x_i; \omega_{h_x}) Z(y - y_j + y_i; \omega_{h_y}) \left/ \iint_S Z(u - x_j + x_i; \omega_{h_x}) Z(u - y_j + y_i; \omega_{h_y}) dx dy \right.}{\sum_i I\{(x_i + x, y_i + y) \in S\}}. \quad (48)$$

In each of these equations, the integral of the kernel function prevents ‘leaking out of masses’ outside the spatial or temporal range of interest. The denominators in equations (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.

#### 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.

*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 equations (18), (23), (24), (25) and (29) respectively.

*Step 3:* estimate  $\mu_t$ ,  $\mu_d$ ,  $\mu_w$ ,  $\mu_{bg}$ ,  $f$  and  $g$  by using equations (43)–(48).

*Step 4:* estimate  $\mu_0$  and  $A$  by using equations (34)–(36).

*Step 5:* stop if the results are convergent; otherwise, go to step 2.

## 5. Data analysis

We analyse the robbery-related violence data in Castellón city, Spain, during the years of 2012 and 2013, as presented in Section 2. See Figs 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:

- (a) a non-periodic but non-stationary Poisson model with  $\lambda(t, x, y) = \mu_0\mu_t(t)\mu_b(x, y)$ ,
- (b) a periodic Poisson model with  $\lambda(t, x, y) = \mu_0\mu_t(t)\mu_d(t)\mu_w(t)\mu_b(x, y)$ ,
- (c) a similar model to that in equation (15) but without daily and weekly periodic effects and
- (d) the model in equation (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 the 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 discuss only 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 Fig. 3. The general trend (Fig. 3(a)) indicates that there is a larger number of events in the first year than in the second year. Also the rate of occurrence of events keeps quite stationary throughout the second year. The weekly periodicity component (Fig. 3(b)) indicates that the robbery events steadily increase from Thursdays to Sundays, which is consistent with 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 (Fig. 3(c)), corresponding to 12–2 p.m. (lunch time) and 7–10 p.m. (dinner time), which are again the periods when more people are in the streets. In contrast the rate of occurrence of such crimes is relatively much lower around 4–10 a.m. in the morning, during which most people are resting. The reconstructed spatial and temporal response functions in the clustering component (Figs 3(d) and 3(e)) imply that, once a crime has occurred, it is likely to trigger another crime within the coming 3 days and within 100 m in distance.

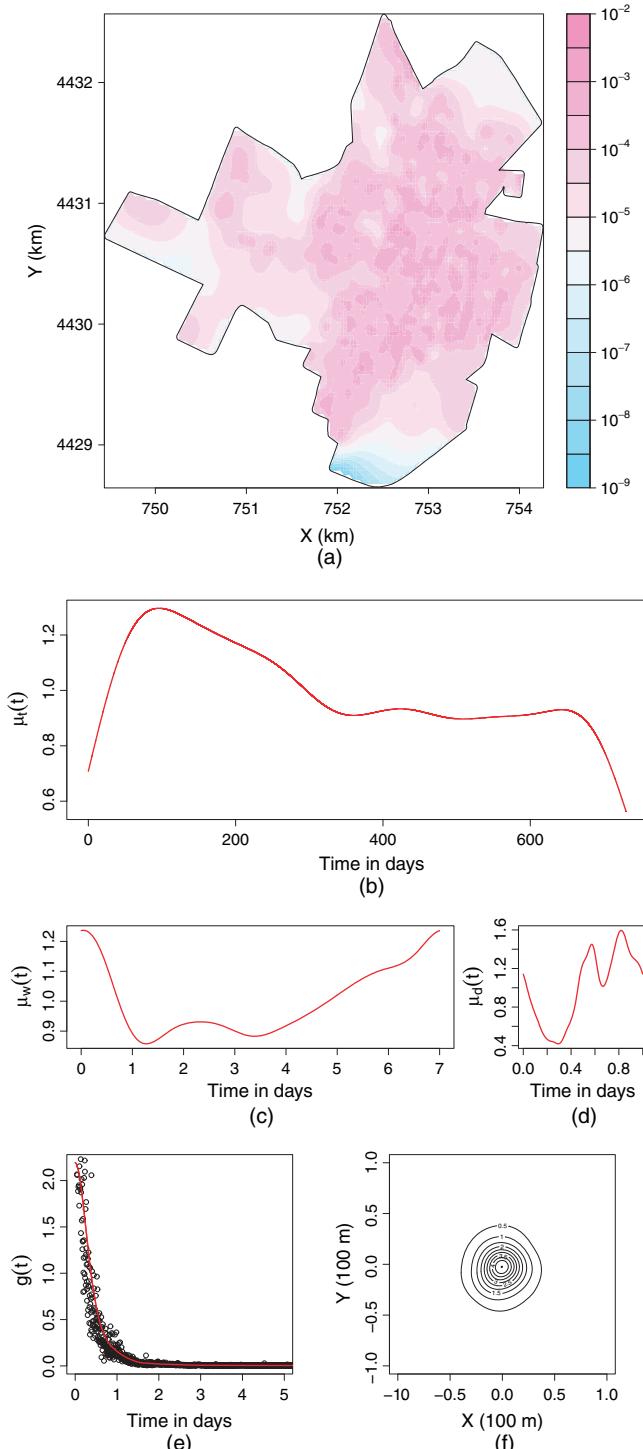
Here,  $A \approx 0.03$  implies that about 3% of the 5089 crime events (about 152 events), which should not be considered as a small number, can be explained by the triggering effect. Comparing with the results of the analysis of the burglary crimes in Los Angeles during the period of

**Table 1.** Results from fitting the model in equation (15) and three other models†

Model	$\hat{\mu}$	$\hat{A}$	$\log(L)$
Non-periodic Poisson	0.7920	—‡	-1335.07
Periodic Poisson	0.7927	—‡	-1050.26
Non-periodic and triggering	0.7710	0.02838	-1304.50
Periodic and triggering	0.7713	0.02913	-920.81

†See Section 5.1 for the details. Parameters  $\mu$  and  $A$  are the relaxation coefficients.

‡Not applicable.



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

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 2011–2016 amounts to 47%. The reason might be that the same burglar watches and visits several neighbouring houses within a short timespan, whereas 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 Fig. 4 in Mohler *et al.* (2011), there might be some periodicity in the marginal temporal response function, whereas our reconstructed function is monotone decreasing. A possible cause of this difference is that periodicity in the background is not considered in Mohler'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 modelling. 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:

- (a) how do we justify the goodness of fit of the model?;
- (b) do the data patterns vary with space and time?;
- (c) how do we 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 model, which is widely used for analysing, modelling 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.

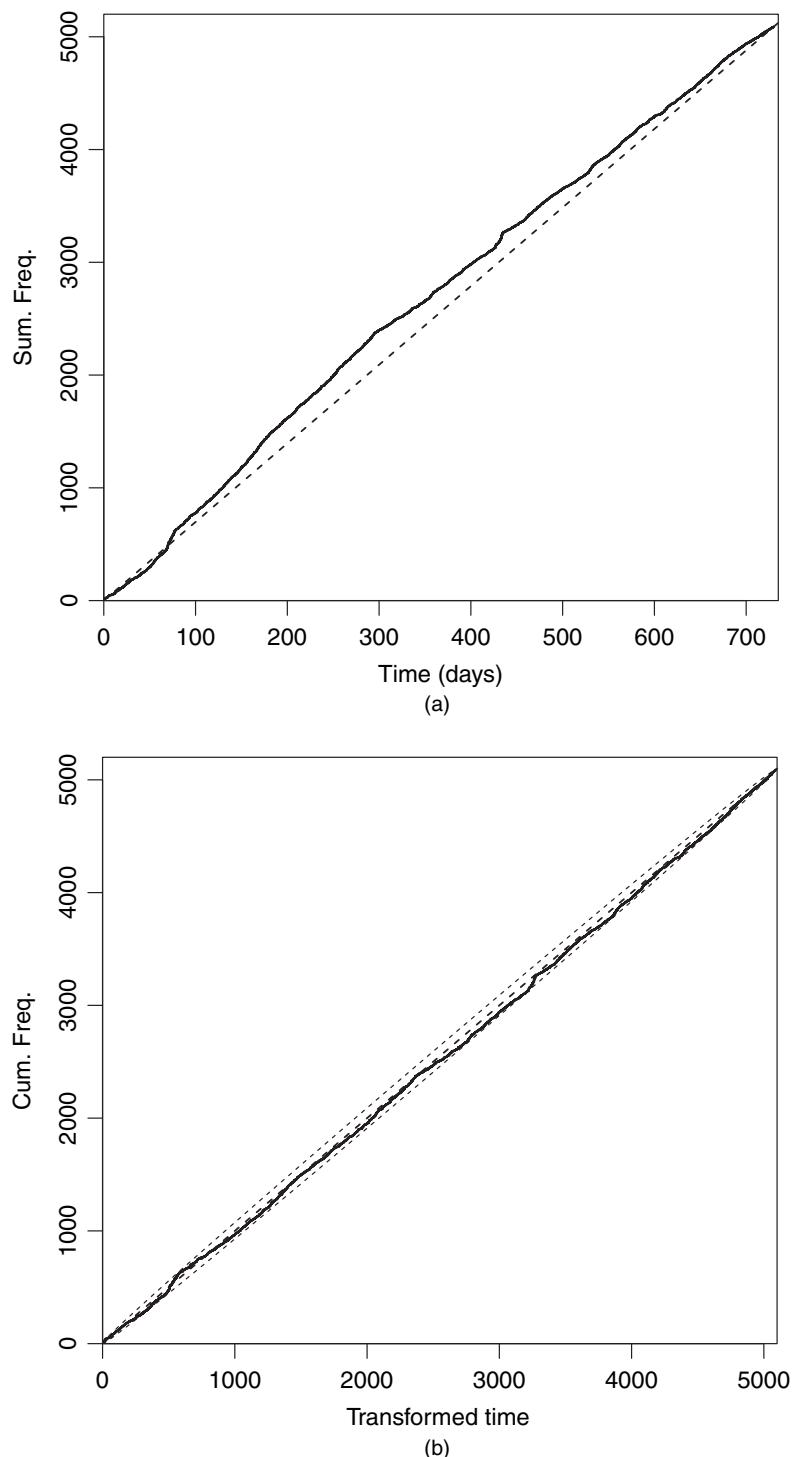
#### 5.2.1. Transformed time sequence analysis

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

$$t_i \rightarrow \tau_i = \int_0^{t_i} \int_S \lambda(u, x, y) dx dy du \quad (49)$$

transforms  $N$  into a stationary Poisson process with a unit rate (standard Poisson process), namely,  $N' = \{\tau_i : i = 1, 2, \dots, n\}$ . 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 equation (49), 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 that are related to the Poisson process can be used to construct the confidence band. Following Schoenberg (2002), the cumulative frequency curve ( $\hat{\tau}_i = \int_0^{t_i} \int_S \hat{\lambda}(u, x, y) dx dy du, i$ ) always connects  $(0, 0)$  and  $(T, n)$ , where  $\hat{\lambda}(u, x, y)$  is the model estimated from the data in  $[0, T] \times S$ . 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



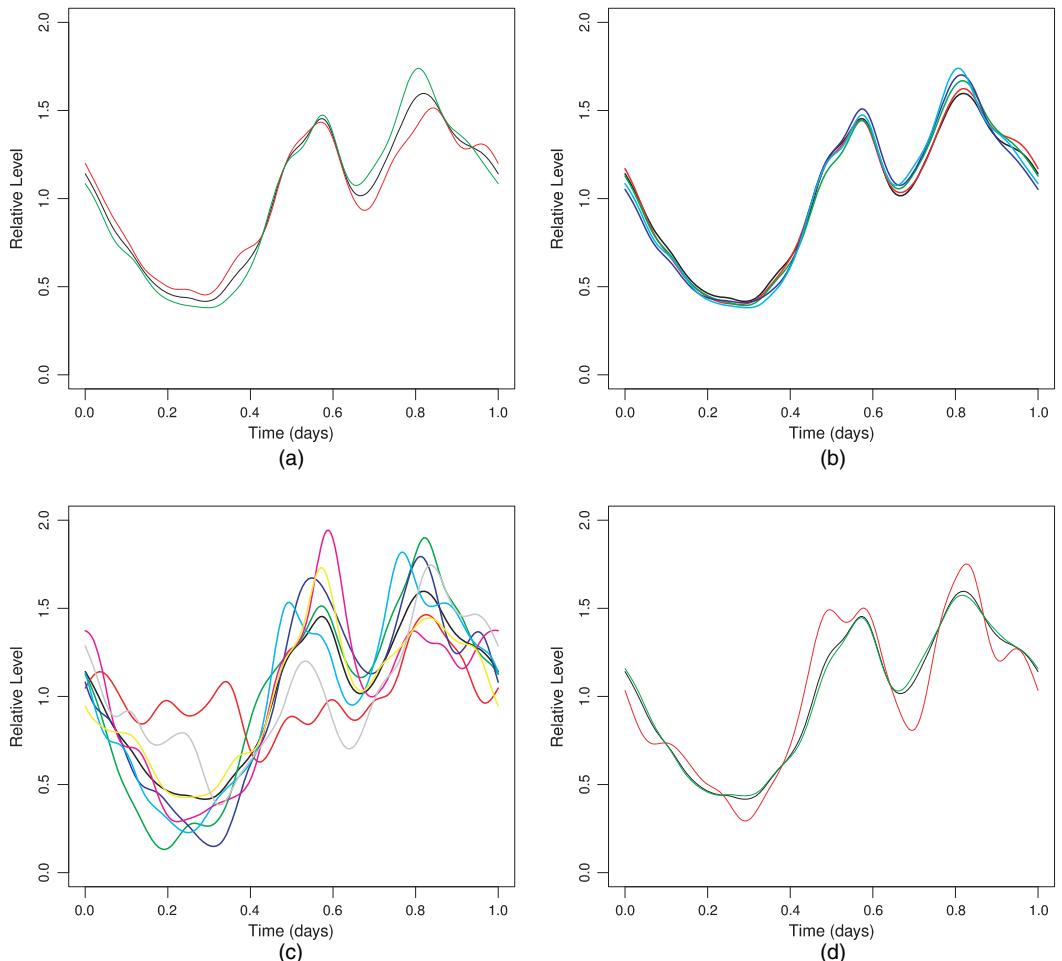
**Fig. 4.** Cumulative frequencies of crime events *versus* (a) original occurrence times and (b) transformed times: /, average occurrence rates; -----, 95% confidence bands for the transformed time sequence

$(k+1, n-k+1)$ ; 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 to Schoenberg (2002) for details.

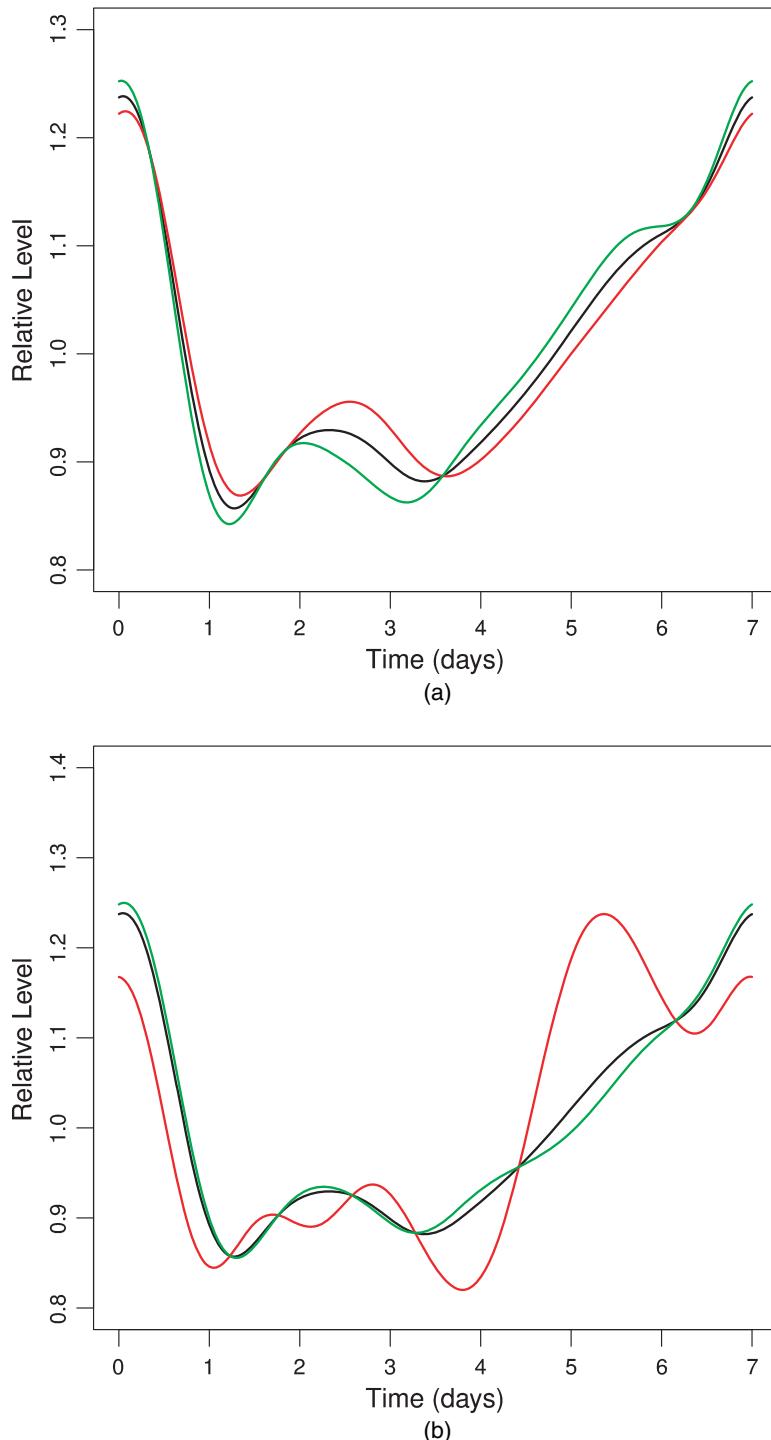
The transformed time sequence for the data analysed is plotted in Fig. 4. Transformation (49) approximately transforms the crime events into a stationary sequence. Around the transformed times of 530, 2400, 3250 and 4300, there seems to be some change point in the rate of occurrence in the transformed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

### 5.2.2. 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 Fig. 5(a). We note that there are no significant differences between these two years. Similarly we reconstruct the



**Fig. 5.** Reconstructed daily periodicity functions  $\hat{\mu}_d(t)$  for (a) 2012 and 2013 (—, all; —, 2012; —, 2013), (b) different seasons (—, all; —, January–March; —, April–June; —, July–September; —, October–December), (c) different days of the week (—, all; —, Sunday; —, Monday; —, Tuesday; —, Wednesday; —, Thursday; —, Friday; —, Saturday) and (d) different areas of the city (—, all; —, centre; —, suburb)



**Fig. 6.** Reconstructed weekly periodicity functions (a) for different years (—, all; —, 2012; —, suburb) and (b) for city centre and suburb areas (—, all; —, centre; —, suburb)

daily periodicity for different seasons, different days of the week and different areas in the city, as shown in Figs 5(b), 5(c) and 5(d) respectively. These results do not show much difference between different seasons. The biggest difference is the effect of the days of the week. From Fig. 5(c) we can see that the daily effect for Sundays is quite flat, and a valley around 4 a.m. and two peaks around 1 p.m. and 9 p.m. There are slight differences in the city centre and the suburb area (Fig. 5(d)): the rate of occurrence is relatively higher at noon and evening and relatively lower in the early morning and in the afternoon in the centre of the city than in the other areas.

#### 5.2.3. Does the weekly periodicity change in time or in space?

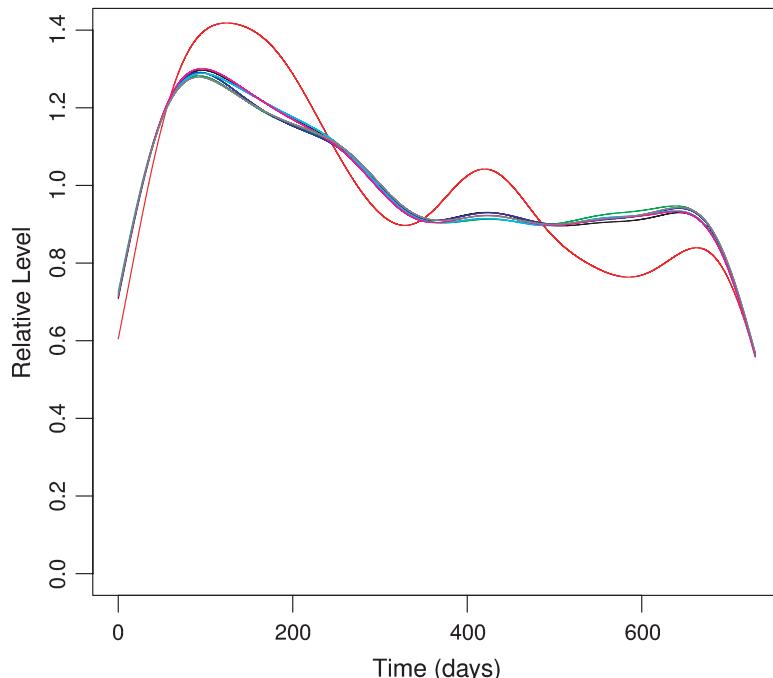
We reconstruct the week periodicity for various years (Fig. 6(a)) and various areas (Fig. 6(b)). The results do not show much difference in weekly periodicity between years. However, the rate of occurrence in the city centre area becomes much higher on Fridays.

#### 5.2.4. Does the long-term trend differ in different places?

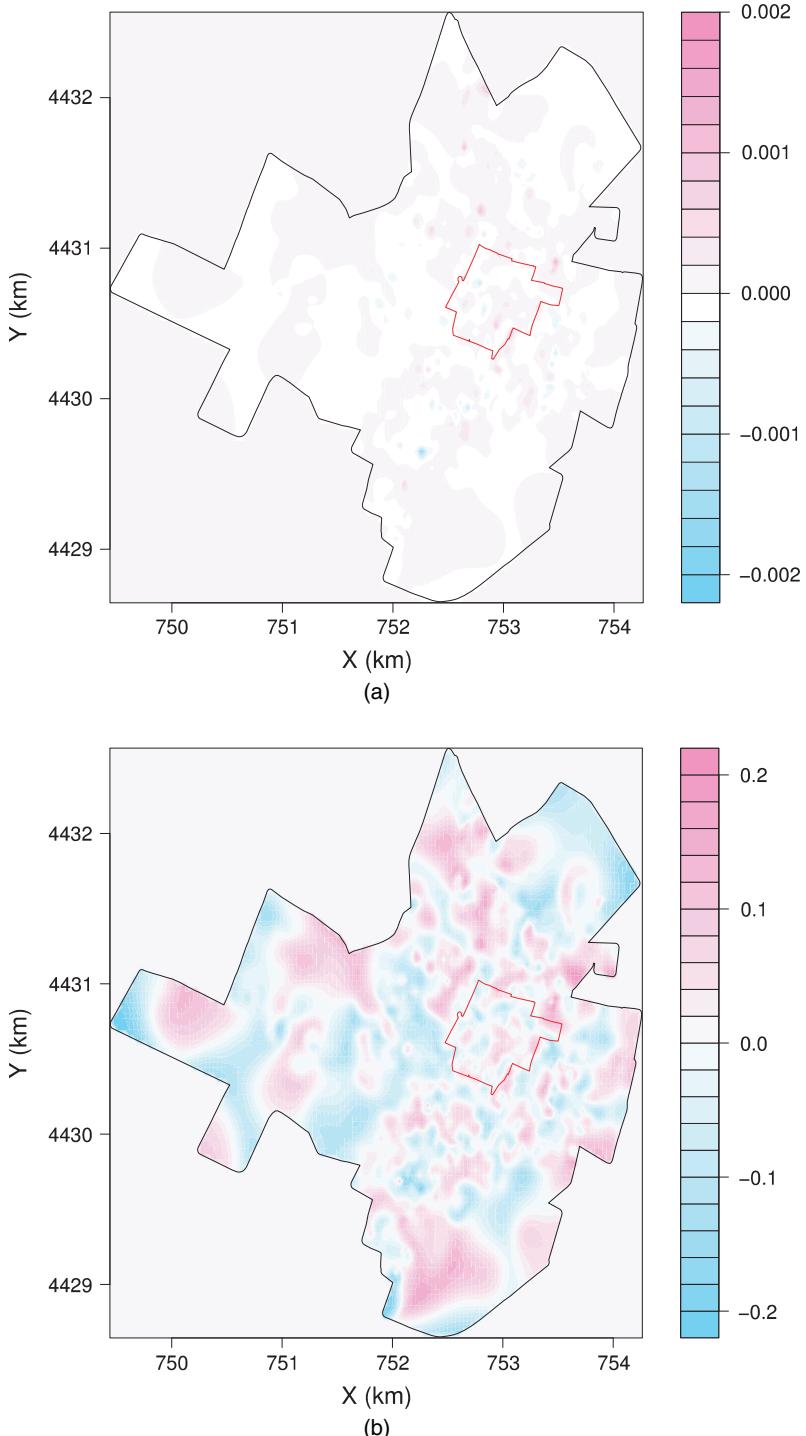
Fig. 7 shows the reconstructed long-term trend component of the background rate. Even though there are two small rebounds about 420 and 640 days, the long-term background rate in the city centre area decreases quicker in those two years than in the other suburb areas. Moreover, there is almost no difference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

#### 5.2.5. 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  $\mu_b(x, y)$  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 Fig. 8(a). For an easier comparison, we also plot the



**Fig. 7.** Reconstructed long-term trend for different areas: —, all; —, city centre; —, suburb; —, suburb, west; —, suburb, east; —, suburb, south; —, suburb, north



**Fig. 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); (b) relative difference between the reconstructed background rates in 2012 and 2013 (the latter minus the former then divided by the background rate for the entire data set)

relative difference,  $(\hat{\mu}_{b'13} - \hat{\mu}_{b'12})/\hat{\mu}_b$ , in Fig. 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.

#### 5.2.6. Is the clustering effect different in different places?

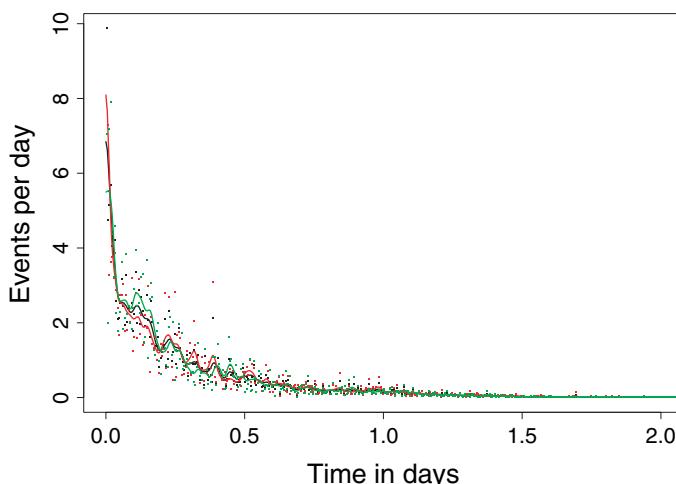
It is also interesting to know whether the clustering effect differs between downtown and the suburb areas. A simple verification is to check whether the reconstructed  $g(t)$  and  $f(x, y)$  are different for the city centre and other areas. These functions are plotted in Figs 9 and 10. The overall shapes of  $g$  and  $f$  are similar for the city centre 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 centre 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 periods.

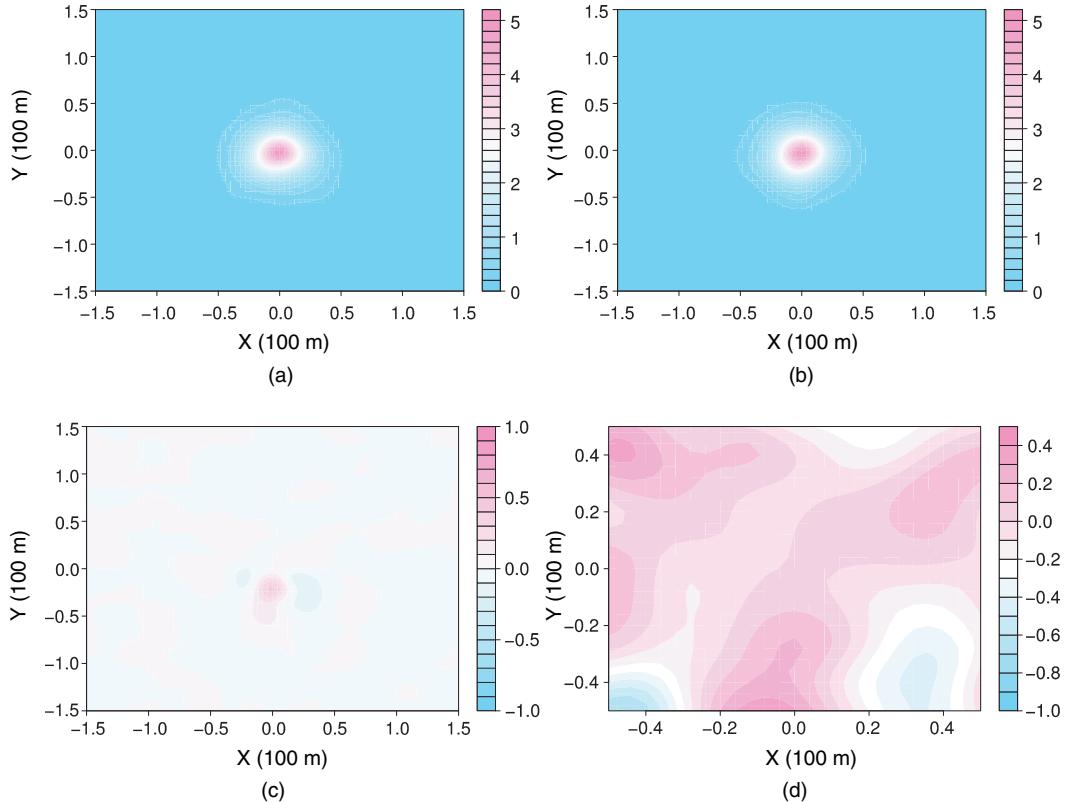
## 6. Conclusions and discussion

In this study, we have proposed a spatiotemporal Hawkes model, whose background rate includes a long-term trend and periodicity, to describe robbery-related violence in Castellón, Spain. To estimate the model, a semiparametric 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 that is developed in this study fits better to crime data and is simple to understand and to estimate, without requiring much prior knowledge of the phenomena studied. Using this method, we have analysed and highlighted the existence of



**Fig. 9.** Reconstructed temporal response of the triggering effect  $\hat{g}(t)$ : —, all the region; —, city centre area; —, suburb area



**Fig. 10.** Diagnostics of regional difference of the spatial response between the city centre and suburb areas: (a) reconstructed  $f$  for the city centre 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)

periodic components and the triggered effect in the process of the crime phenomena studied. In the estimation procedures of the background components and the excitation response functions, two relaxation parameters are adopted to stabilize and secure convergence.

The final results show the following features of the behaviours of robbery-related violence in Castellón.

- Background dominates the whole process whereas the clustering effect contributes only about 3%.
- The periodicity effect is strong in the background.
- Residual analysis shows that crime activity is different during weekends from during working days.
- 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.

- We could consider the non-linear Hawkes process (e.g. Brémaud and Massoulié (1996), Delattre *et al.* (2016), Torrisi (2016, 2017), Zhu (2013, 2014, 2015) and Chevallier *et al.* (2018)), whose temporal version has a conditional intensity of the form

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

where  $\Phi$  is a locally integrable and left continuous non-negative function.

- (b) In this study, we used kernel estimates with fixed bandwidths to obtain all the components in the model formulation. Also, in the comparison between 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.
- (c) Other non-parametric estimates, such as Bayesian procedures with smoothness priors and tessellation methods, can be also incorporated in the method proposed. Careful and detailed comparisons should be done between these methods to find the best for practical forecasting.

## Acknowledgements

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

## References

- Bartlett, M. S. (1963) The spectral analysis of point processes (with discussion). *J. R. Statist. Soc. B*, **25**, 264–296.
- Bowers, K. J., Johnson, S. D. and Pease, K. (2004) Prospective hot-spotting the future of crime mapping? *Br. J. Crimin.*, **44**, 641–658.
- Brémaud, P. and Massoulié, L. (1996) Stability of nonlinear Hawkes processes. *Ann. Probab.*, **24**, 1563–1588.
- Caplan, J. M. and Kennedy, L. W. (2016) *Risk Terrain Modeling: Crime Prediction and Risk Reduction*. Berkeley: 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. *Stoch. Processes Appl.*, to be published.
- Cohen, J., Gorr, W. L. and Olligschlaeger, A. M. (2007) Leading indicators and spatial interactions: a crime-forecasting model for proactive police deployment. *Geog. Anal.*, **39**, 105–127.
- Daley, D. D. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*, vol. 1, *Elementary Theory and Methods*, 2nd edn. New York: Springer.
- Delattre, S., Fournier, N. and Hoffmann, M. (2016) Hawkes processes on large networks. *Ann. Appl. Probab.*, **26**, 216–261.
- Felson, M. and Boba, R. (2010) *Crime and Everyday Life*. New York: Sage.
- Hawkes, A. G. (1971a) Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. B*, **33**, 438–443.
- Hawkes, A. G. (1971b) Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**, 83–90.
- Hawkes, A. G. and Oakes, D. (1974) A cluster process representation of a self-exciting process. *J. Appl. Probab.*, **11**, 493–503.
- Johnson, N., Hitchman, A., Phan, D. and Smith, L. (2018) Self-exciting point process models for political conflict forecasting. *Eur. J. Appl. Math.*, **29**, 685–707.
- Levine, N. (2017) *CrimeStat: a Spatial Statistical Program for the Analysis of Crime Incidents*, pp. 381–388. Cham: Springer.
- Lewis, P. A. W. (1964) A branching Poisson process model for the analysis of computer failure patterns (with discussion). *J. R. Statist. Soc. B*, **26**, 398–456.
- Marsan, D. and Lengliné, O. (2008) Extending earthquakes' reach through cascading. *Science*, **319**, 1076–1079.
- Mohler, G. (2014) Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.*, **30**, 491–497.
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011) Self-exciting point process modeling of crime. *J. Am. Statist. Ass.*, **106**, 100–108.
- 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. *J. Am. Statist. Ass.*, **110**, 1399–1411.
- Neyman, J. E. and Scott, E. L. (1953) Frequency of separation and interlocking of clusters of galaxies. *Proc. Natn. Acad. Sci. USA*, **39**, 737–743.
- Neyman, J. E. and Scott, E. L. (1958) Statistical approach to problems of cosmology (with discussion). *J. R. Statist. Soc. B*, **20**, 1–43.
- Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Statist. Ass.*, **83**, 9–27.

- Ogata, Y. (1989) Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, **169**, 159–174.
- Ogata, Y. (1998) Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.*, **50**, 379–402.
- Ratcliffe, J. H. (2004) The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. *Police Pract. Res.*, **5**, 5–23.
- Reinhart, A. (2018) A review of self-exciting spatio-temporal point processes and their applications. *Statist. Sci.*, to be published.
- Reinhart, A. and Greenhouse, J. (2018) Self-exciting point process with spatial covariates: modeling the dynamics of crime. *Appl. Statist.*, **67**, 1305–1329.
- Rodrigues, A. and Diggle, P. J. (2012) Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. *J. Am. Statist. Ass.*, **107**, 93–101.
- Rosser, G. and Cheng, T. (2016) Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. *Appl. Spat. Anal. Pol.*, to be published.
- Schoenberg, F. (2002) On rescaled Poisson processes and the Brownian bridge. *Ann. Inst. Statist. Math.*, **54**, 445–457.
- Shirota, S. and Gelfand, A. E. (2017) Space and circular time log Gaussian Cox processes with application to crime event data. *Ann. Appl. Statist.*, **11**, 481–503.
- Torrisi, G. L. (2016) Gaussian approximation of nonlinear Hawkes processes. *Ann. Appl. Probab.*, **26**, 2106–2140.
- Torrisi, G. L. (2017) Poisson approximation of point processes with stochastic intensity, and application to nonlinear Hawkes processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **53**, 679–700.
- Townsley, M., Homel, R. and Chaseling, J. (2003) Infectious burglaries. a test of the near repeat hypothesis. *Br. J. Crimin.*, **43**, 615–633.
- Zhu, L. (2013) Nonlinear Hawkes processes. *PhD Thesis*. Department of Mathematics, New York University, New York.
- Zhu, L. (2014) Process-level large deviations for nonlinear Hawkes point processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **50**, 845–871.
- Zhu, L. (2015) Large deviations for Markovian nonlinear Hawkes processes. *Ann. Appl. Probab.*, **25**, 548–581.
- Zhuang, J. (2006) Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. R. Statist. Soc. B*, **68**, 635–653.
- Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002) Stochastic declustering of space-time earthquake occurrences. *J. Am. Statist. Ass.*, **97**, 369–380.
- Zhuang, J., Ogata, Y. and Vere-Jones, D. (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.*, **109**, article B05301.
- Zhuang, J., Werner, M. J. and Harte, D. S. (2013) Stability of earthquake clustering models: criticality and branching ratios. *Phys. Rev. E*, **88**, article 062109.

Copyright of Journal of the Royal Statistical Society: Series A (Statistics in Society) is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.