Self-exciting point process modelling of crimes on linear networks

Nicoletta D'Angelo [1], David Payares [2], Giada Adelfio [1], and Jorge Mateu [3]

[1] Department of Economics, Business and Statistics, University of Palermo, Italy

[2] Department of Earth Observation Science, University of Twente, Netherlands

[3] Department of Mathematics, Universitat Jaume I, Spain

Address for correspondence: Nicoletta D'Angelo, University of Palermo, Italy.

E-mail: nicoletta.dangelo@unipa.it.

Phone: .

Fax: .

Abstract: Although there are recent developments for the analysis of first and second-order characteristics of point processes on networks, there are very few attempts in introducing models for network data. Motivated by the analysis of crime data in Bucaramanga (Colombia), we propose a spatio-temporal Hawkes point process model adapted to events living on linear networks. We first consider a non-parametric modelling strategy, for which we follow a non-parametric estimation of both the background and the triggering components. Then we consider a semi-parametric version, including a parametric estimation of the background based on covariates, and a non-

parametric one of the triggering effects. Our model can be easily adapted to multi-type processes. Our network model outperforms a planar version, improving the fitting of the self-exciting point process model.

Key words: Covariates; Crime data; Hawkes processes; Linear networks; Self-exciting point processes; Spatio-temporal point processes

1 Introduction

Point processes are stochastic processes defining a natural and convenient mathematical tool to describe the process of discrete events that occur in a continuous space, time or a space-time domain; examples, spanning many scientific branches, are found with forest fires, crimes, earthquakes, diseases, tree locations, animal locations or communication network failures, to name just a few. Depending on the domain where the events occur, we can talk about spatial, temporal or spatio-temporal point processes. When time is present, the process has an evolution in time, and the events can be sorted according to their chronological order, sharing some common features with time series. 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.

A number of papers have dealt with the analysis of crime data using self-exciting point process theory, after the analogy drawn by Mohler et al. (2011) between aftershock ETAS models and crime. In particular, several papers have proposed a Hawkes-type point process modelling framework for crime data, as this type of data is usually

clustered (Reinhart, 2018; Park et al., 2021). Recently, Zhuang and Mateu (2019) proposed a spatio-temporal Hawkes-type point process model, which includes a background component with daily and weekly periodisation, and a clustering component that is triggered by previous events. Their model is used to describe the occurrences of violence or robbery cases in a urban environment during two years, and their 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.

As crime events are naturally constrained to occur on the streets structure of a city, in this paper, we advocate the use of the theory of point processes on linear networks. A network, or a graph, is a collection of vertices joined by edges (Newman, 2010). A linear network is a union of finitely many line segments in the plane where different edges only possibly intersect with each other at one of their vertices (see Anderes et al. (2020) for spatio-temporal covariance functions on generalised linear networks). Point processes on linear networks are recently considered to analyse events occurring on particular network structures (e.g. traffic accidents). They were firstly introduced in the spatial context and then extended to the spatio-temporal case, focusing on the analysis of first- and second-order summary statistics (Ang et al., 2012; McSwiggan et al., 2017; Rakshit et al., 2017; Moradi et al., 2019; Rakshit et al., 2019a; Moradi and Mateu, 2019; D'Angelo et al., 2021a,b). Most of the literature about spatial and spatio-temporal point processes on networks is concerned with non-parametric estimation of the first-order intensity (Moradi et al., 2019; Moradi and Mateu, 2019; Mateu et al., 2019); however, only a recent paper by D'Angelo et al. (2021a) has dealt with parametric intensity specification of inhomogeneous first-order intensities on networks to analyse the spatio-temporal distribution of visitors' stops by touristic attractions

in Palermo (Italy). The authors fitted a Gibbs point process model with mixed effects for the purely spatial component, as well as a spatio-temporal log-Gaussian Cox process, adapting them to the underlying road network. This latter paper only considered inhomogeneous Poisson models and did not take into account the potential self-exciting behaviour of points. Other recent papers that deal with model building of non-Poisson models are the following. Baddeley et al. (2017) adapt to linear networks popular procedures for constructing a point process, such as the Switzer-type pseudostationary process, Cell process, and Cluster processes. van Lieshout (2018) defines nearest-neighbour point processes on graphs with Euclidean edges and linear networks, which can be seen as analogues of renewal processes on the real line. Rasmussen and Christensen (2021) adapt the so-called conditional intensity function used for specifying point processes on the time line to the setting of directed linear networks, considering specific classes of point process models as Poisson processes, Hawkes processes, non-linear Hawkes processes, self-correcting processes, and marked Hawkes processes, used in that paper to analyse simulated and neurological data.

Following these considerations, as none of the above-mentioned papers about point process modelling of crime have proposed models taking into account the road network geometry, the aim of this paper is to analyse crime data with self-exciting point processes, while also accounting for the underlying network structure where events occur. Statistical analysis of network data presents severe challenges (Baddeley et al., 2021). A network is not spatially homogeneous, which creates geometrical and computational complexities and leads to new methodological problems, with a high risk of methodological error. Real network data, as crime data, can also exhibit an extremely wide range of spatial scales. These problems pose a significant challenge to the classical methodology of spatial statistics based on stationary processes, which

is largely inapplicable to data on a network. Note also that the choice of distance metric on the network is pivotal in the theoretical development and in the analysis of real data. This is thus a key aspect when treating robberies that happen on the streets of a city to better model and understand the true spatial and temporal structures underpinning such type of crimes. Indeed, regarding the intensity estimation issue addressed in this paper, considering the geometry of the network into the fitting procedure represents a reasonable choice, given the increasing availability of road networks data and computational resources nowadays. For instance, from a practical point of view, the predicted intensity could be obtained only where events are driven to occur. Of course, crimes are an obvious example of constrained point patterns, and therefore, the impact of considering the road network structure into the inference procedure potentially reflects on more detailed information, which may then drive policy actions. In detail, we analyse robbery crimes occurred in the city of Bucaramanga (Colombia) in 2018. We fit a model similar to the semi-parametric specification employed in Zhuang and Mateu (2019) into the network case. Therefore, the first main contribution of our paper regards the proposal of an extension of the Hawkes model proposed by Zhuang and Mateu (2019). We do not only include the network geometry into the fitting procedure, but also draw conclusions on the scale of shortest-path distances, typically more appropriate when dealing with point processes occurring on linear networks. We find that our proposed model achieves a much better fit when compared to the planar counterpart, allowing us to better interpret the results.

Examples of applications incorporating the external information in self-exciting models can be found in Schoenberg (2016); Reinhart (2018); Adelfio and Chiodi (2020). In particular, Park et al. (2021) modelled gang-related violent crimes in Los Angeles

(California) using spatio-temporal Hawkes processes, proposing an algorithm to estimate the spatio-temporally varying background rate non-parametrically as a function of demographic covariates. Therefore, the second contribution of our paper concerns a further extension of our proposed model by including external covariates in the purely spatial background component, following the specification of self-exciting models that Meyer et al. (2012) proposed in an epidemiological context. In spatial point process theory, the spatial covariates are referred to as those variables with observable values, at least in principle, at each spatial location in the spatial window. For inferential purposes, their values must be known at each point of the data point pattern and at least at some other locations. Therefore, in the crime data context, those would comprehend socio-economic characteristics of the analysed regions, as in Park et al. (2021) such as education level, illiteracy, access to public services like water, electricity, sewerage, and unemployment, housing quality, socio-economic status. Other available covariates could be related to the distance to and density of facilities per street (such as police station, hospitals, schools), environmental characteristics (slope, PM10), and street characteristics (direction, type, structures). In this paper, we consider spatial covariates in the background component, even though the specification of the proposed model would also allow for the inclusion of marks in the triggered component (see Adelfio and Chiodi (2020) for their proposal in the seismic context, and Chiodi et al. (2021) for an application).

We find that the inclusion of some spatial covariates in the background component further improves the fitting of the model, and therefore it lays the bases for future developments in this promising direction, such as the inclusion of individual-related covariates into the triggering component.

All the codes are written in the R Core Team (2021) software language and are available through the link http://www.statmod.org/smij/archive.html, together with the data analysed.

The structure of the paper is as follows. Section 2 presents the data and the motivating problem. Section 3 provides an overview of self-exciting spatio-temporal point processes. The novel methods introduced in the paper are detailed in Section 4. Section 5 presents the data analysis, and Section 6 is devoted to conclusions.

2 Crimes in Bucaramanga

Reports of crimes from January 2010 to September 2020 in Bucaramanga (Colombia) were collected by the city Home Secretariat and the Mayor's Office of Security and Defense. The reports contain spatial location and time of occurrence of crimes that happened in Bucaramanga, together with some further information such as type of crime, severity, victim characteristics, victimiser's crime weapon, and transportation. The data aims primarily to study the behaviour in space and time of reported crimes in Bucaramanga to define location and population-based public policies to mitigate the city's reported crime rates. Further information about the data can be found at the Municipal Digital Observatory of Bucaramanga[1].

We note that in crime data, there are naturally unreported crimes and some false reported crimes. The latter is not our case (or at least there potentially are very few false cases) as the data we work with were double checked by the Home Secretary before releasing this info to public instances. It is an aspect that we have to live with

[1] Municipal Observatory of Bucaramanga: Crimes in Bucaramanga

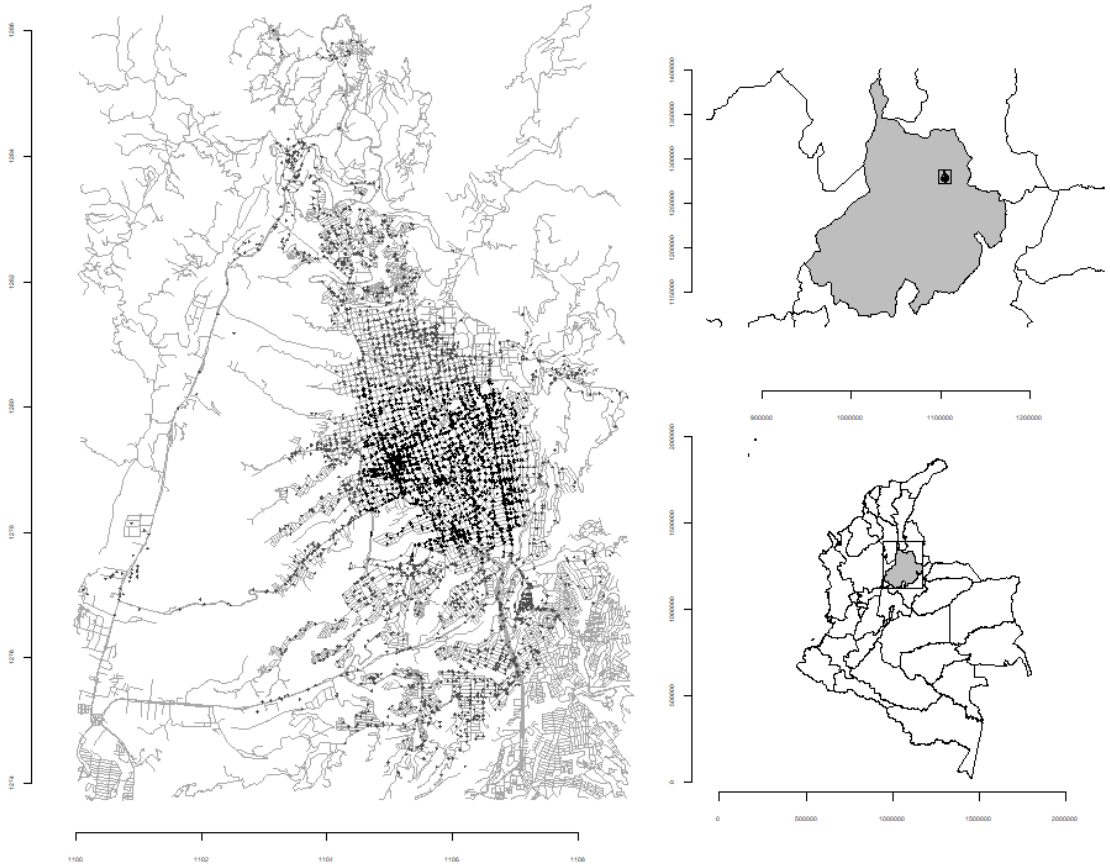


Figure 1: The black points represent the armed robberies in Bucaramanga, and the dark grey points the crimes in the city's downtown. In light grey, the segments of the streets of Bucaramanga city, obtained from OpenStreetMap

and assume that what we observe is just a noisy (subset) version of the real number of crimes. So we are dealing with reported crimes rather than with the total population of crimes. We also note that the events we focus on here (armed-robberies) recorded in the datasets often respond to complex historical police-community interactions, which means that the data collection process is possibly influencing the reported cases. Again, this could pose a biased sampling problem and this comes back to the fact that what we observe is a noisy, biased version of the reality.

Following Zhuang and Mateu (2019) analysis, we bounded our data to only robbery-related reported crimes, as these type of crimes are often encountered and they make the population and society being afraid and scare of moving freely within the city. Furthermore, we specified spatial and temporal windows to filter the original data for computational cost and statistical representativeness purposes. The subset data of this study comprises a spatio-temporal point pattern consisting of 2671 armed robberies in the downtown of Bucaramanga city from the 1st January 2018 to the 31st December 2018. Armed robberies have sharply increased in Bucaramanga over the last ten years, becoming the leading crime affecting the city's inhabitants. Given this crime frequency and situational characteristics, the local police constantly struggles to combat and prevent their occurrence. Efforts to reduce the armed robberies burden in Bucaramanga have been developed mainly in commercial and residential areas where robbery rates are intimidating. For instance, the city's downtown area accounted for 40-60% of annual reported armed robberies in Bucaramanga between 2010 and 2020. By 2018, the figures in downtown reached almost 55% of all robbery reported crimes. We selected armed robberies in the downtown region for 2018 as a representative sample of robbery-related crime dynamics in Bucaramanga, as the downtown region contains most of the city shops and facilities. Figure 1 displays the armed robberies

for the entire city of Bucaramanga (dark grey) and its downtown (black) in 2018. In this study, we focus on analysing the latter subregion.

Since reported crime's locations were slightly shifted from the street configuration, we relocated the points to the closest location on the linear network. Figure 2 shows the georeferenced coordinates of the crimes as reported in the original database (left panel) and the relocated coordinates matching the linear network L (right panel). Figure 3 displays the location of the armed reported crimes per month in 2018. We note how the number of reported crimes increases throughout the year and concentrates east of the city's downtown. The linear network L is composed of 3,136 nodes and 4,290 segments referring to Bucaramanga's downtown. The downtown area covers 23 commercial and ten residential neighbourhoods containing most of the city's shops and facilities. Likewise, the Mayor's Office of Bucaramanga estimates that at least 60% of its inhabitants commute daily to this area for professional, academic and touristic activities.

2.1 Spatial covariates

Demographic, socio-economic, geographical, and environmental variables were obtained at the block level for the city of Bucaramanga. The data was provided by the UN-Habitat Colombia, the United Nations program for human settlements and sustainable urban development in Colombia. The UN-Habitat Colombia develops multilayered indexes to assess socio-economic and demographic aspects related to urban planning. We used all 36 variables, including socio-economic factors such as unemployment rate, education level, literacy rate, public services coverage, socio-economic strata, and housing quality; demographic aspects, such as total population, gender

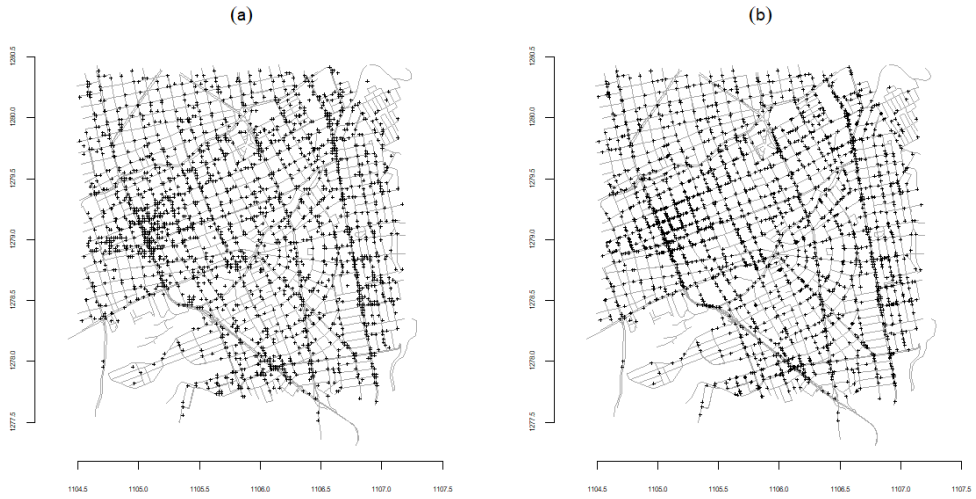


Figure 2: (a) Original locations of points and (b) locations relocated on the linear network

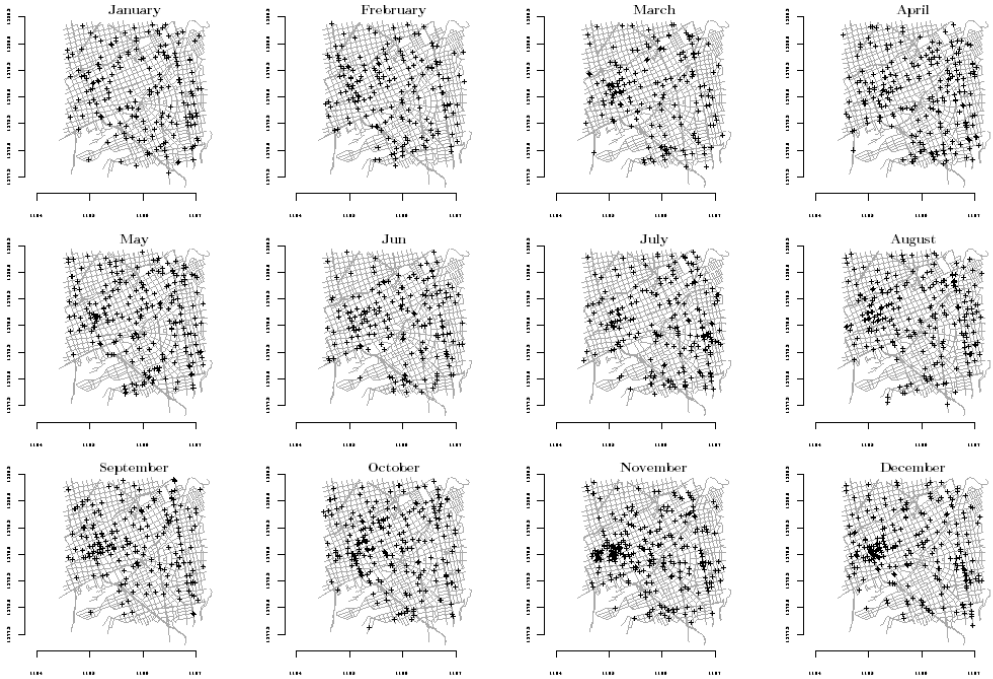


Figure 3: Armed robberies in Bucaramanga's downtown per month, 2018

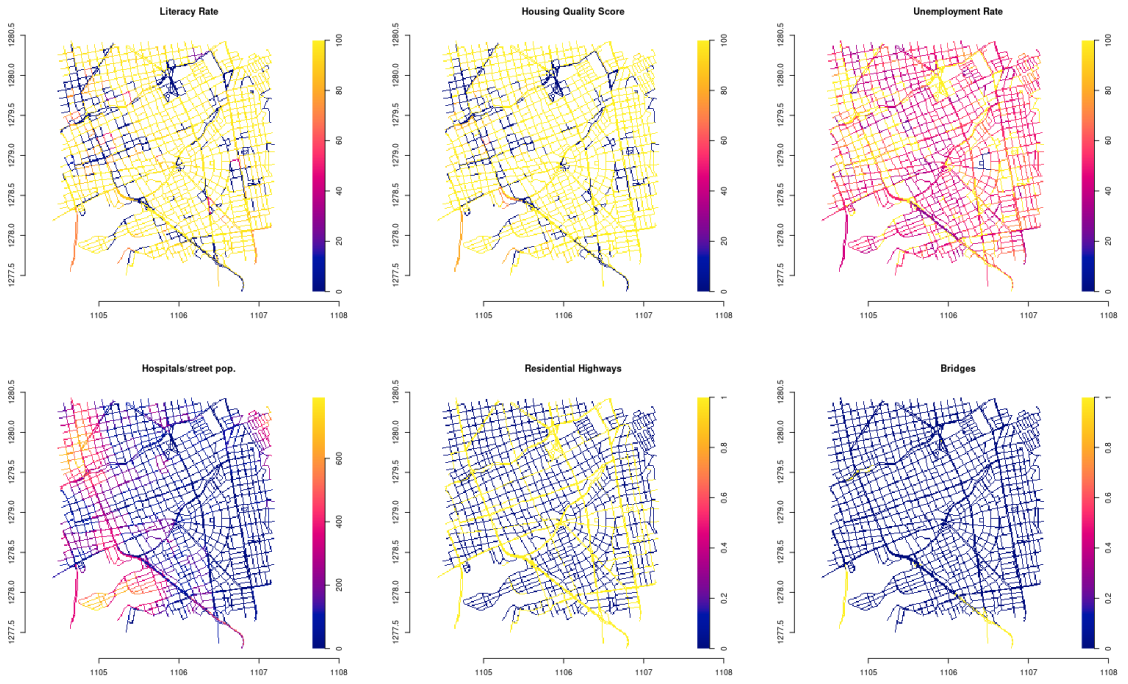


Figure 4: Some socio-economic, demographic and environmental spatial covariates used

ratio, and elderly population; environmental conditions, for instance, street slope and street type; and geographical covariates such as distances to facilities, the density of facilities per 10,000 inhabitants, among others. Some variables are of continuous nature (such as a percentage) or others are just informing of presence /absence of a characteristic (for example, the existence of a tunnel in a street). For reasons of space, Figure 4 displays six out of the thirty-six variables employed in our analysis. As shown in the figure, the covariate value is known in each segment of the linear network under analysis. Some of them are continuous variables, such as literacy rate in panel (a), and some are binary, such as bridges, in panel (f).

As the covariates were georeferenced at the block level, except for the geographical ones, we translate the values into the network by assigning the median value of its

neighboring blocks to each segment. We manually computed the values for highly heterogeneous areas of the city. For example, the Northwestern area of Bucaramanga's downtown possesses the richest and poorest zones of the city; the streets of wealthy neighborhoods can be connected to streets belonging to poverty-stricken blocks.

3 Self-exiting point processes on the Euclidean plane

3.1 Background concepts

Following Cressie (2015), we introduce point processes by a mathematical approach that uses the definition of a counting measure on a set $X \subseteq \mathbb{R}^d, d \geq 1$, with non-negative values in $\mathbb{Z}$: for each Borel set B this $\mathbb{Z}_+$-valued random measure gives the number of events falling in B.

Definition 1 ***Point process***

Let $(\Omega, \mathcal{A}, P)$ *be a probability space and* Φ *a collection of locally finite counting measures on* $X \subset \mathbb{R}^d$. *Define* $\mathcal{X}$ *as the Borel* σ-algebra of X, *and let* $\mathcal{N}$ *be the smallest* σ*-algebra on* Φ *generated by sets of the form* $\{\phi \in \Phi : \phi(B) = n\}$ *for all* $B \in \mathcal{X}$. *A point process* N *on* X *is a measurable mapping of* $(\Omega, \mathcal{X})$ *into* $(\Phi, \mathcal{N})$. *A point process defined on* $(\Omega, \mathcal{A}, P)$ *induces a probability measure* $\Pi_N(Y) = P(N \in Y), \forall Y \in \mathcal{N}$. *Then, for any set* $B \in \mathcal{X}$, $N(B)$ *represents the number of points falling in* B, *such that if* B *is the union of disjoint sets* $\tilde{B}_1, \tilde{B}_2, \ldots$, *then* $N(B) = \sum N(\tilde{B}_i)$.

The first step in analysing a point pattern is to learn about its first-order characteristics, studying the relationship of the points with the underlying environmental variables that describe the observed heterogeneity. When the purpose of the analysis

is to describe possible interactions among points, that is, if the given data exhibit spatial inhibition or aggregation, second-order properties of the process are analysed. Broadly speaking, the intensity function describes the rate at which the events occur in the given spatio-temporal region, while the second-order product densities are used when the interest is in describing spatio-temporal variability and correlations between pair of points of a pattern. They represent the point process analogues of the mean function and the covariance function of a real-valued process, respectively.

Point processes can be formally specified in several ways, for instance, by considering the joint distribution of the counts of points in arbitrary sets or by defining a conditional intensity function. Let N be a point process $\{(t_i, x_i, y_i) : i = 1, \dots, n\}$ on a spatio-temporal domain $X = W \times T \subseteq \mathbb{R}^2 \times \mathbb{R}_+$, with area $|W| > 0$ and length $|T| > 0$, and with t representing the time, and x and y the two spatial coordinates. Its conditional intensity function is defined by

$$\lambda(t, x, y|\mathcal{H}_t) = \lim_{\Delta t, \Delta x, \Delta y \to 0^+} \frac{E[N((t, t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)|\mathcal{H}_t)]}{|\Delta t \Delta x \Delta y|} \quad (3.1)$$

where $\mathcal{H}_t$ is the space-time occurrence history of the process up to time t, i.e. the σ-algebra of events occurring at times up to but not including t. $\Delta t, \Delta x, \Delta y$ are time and space increments respectively, and $E[N((t, t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)|\mathcal{H}_t)]$ is the history-dependent expected value of occurrence in the volume $\{(t, t + \Delta t) \times (x, x + \Delta x) \times (y, y + \Delta y)\}$. Assuming such a limit exists for each point (t, x, y) in the space-time domain and that the point process N is simple, the conditional intensity process uniquely characterises the finite-dimensional distributions of the point process (Daley and Vere-Jones, 2007).

To model events that are clustered, self-exciting point processes are often used. These models are largely used to describe earthquakes characteristics, assuming that the

occurrence of an event increases the probability of occurrence of others events in time and space. Examples of self-exciting point processes include Hawkes models (Hawkes, 1971a,b; Hawkes and Adamopoulos, 1973) and ETAS models (Ogata and Katsura, 1988; Adelfio and Chiodi, 2015). The conditional intensity function of a linear self-exciting process is defined by

$$\lambda(t,x,y|\mathcal{H}_t) = \mu(t,x,y) + \int_{-\infty}^{t}\int_{W} g(t-t^{'},x-x^{'},y-y^{'})N(\mathrm{d}x^{'}\times\mathrm{d}y^{'}\times\mathrm{d}t^{'}), \quad (3.2)$$

being the sum of two non-negative functions: $\mu(t,x,y) > 0$, that describes the large-scale variation of $\lambda(t,x,y|\mathcal{H}_t)$, and $g()$, such that $\iint g(v)dv < 1$, which describes its small-scale variation due to the interaction with the events in the past. This process can be interpreted as a generalised Poisson cluster process associating to centers, of rate μ, a branching process of descendants. The spatio-temporal Hawkes process has a conditional intensity of the form (Hawkes, 1971a,b)

$$\lambda(t,x,y) = \mu(t,x,y) + \sum_{i:t_i<t} g(t-t_i,x-x_i,y-y_i), \quad (3.3)$$

where $\mu(t,x,y)$ is the background rate, and $g(t,x,y)$ is the rate of occurrence triggered by an event at time 0 and location at the origin. The triggering density governs the spatial–temporal distance of triggered events from their antecedent events and is usually modelled to decay with distance from the origin over time and space (Park et al., 2021). In Mohler et al. (2011), the background rate $\mu(t,x,y)$ was assumed to be a function of space and time and they used kernel functions to smooth the estimates of both μ and g, introducing the idea of the stochastic reconstruction algorithm (Zhuang et al., 2004; Zhuang, 2006; Marsan and Lengline, 2008) for the analysis of crime data, by means of relaxing parameters and periodic components in the background rates.

3.2 Hawkes model for crime data

We consider the self-exciting model in Equation (3.2), following the semi-parametric specification proposed by Zhuang and Mateu (2019) for a spatio-temporal Hawkes process. Therefore, the full model, that we consider here, is specified as follows

$$\lambda(t,x,y) = \mu_0\mu_t(t)\mu_w(t)\mu_b(x,y) + A\int_{-\infty}^{t-}\int_W g(t-s)h(x-u,y-v)N(\mathrm{d}u\times\mathrm{d}v\times\mathrm{d}s), \tag{3.4}$$

where $\mu_t(t)$ and $\mu_w(t)$ represent the trend term and the weekly periodicity in the temporal components of the background rate, $\mu_b(x,y)$ represents the spatial background rate, and $g(t-s)h(x-u,y-v)$ represents the sub-process triggered by an event previously occurring at time s and location (u,v). Note that this model enables the background rate to include a spatial background pattern that can be separated from the periodicity effects and the long-term temporal trend. We estimate the two relaxation coefficients A and μ_0, normalise to 1 the average values of $\mu_t(t)$, $\mu_w(t)$ and $\mu_b(x,y)$, and define the probability density functions g and h such that $\int_0^\infty g(s)\mathrm{d}s = 1$, and $\int\int_X h(u,v)\mathrm{d}u\mathrm{d}v = 1$.

In the following, we outline the modified Estimation-Maximization (E-M) algorithm employed by Zhuang and Mateu (2019), as we do not to take into account the daily temporal resolution, not available for the data presented in Section 2 and further analysed in Section 5.

E-M estimation algorithm:

Step 1. Set initial values for $\{\mu_0, \mu_t, \mu_w, \mu_b, A, g, h\}$

Step 2. Compute background and excitation components by reconstructing for each i-th

event $(i = 1, \ldots, n)$

$$\begin{aligned} w_i^{(t)} =& \mu_t(t_i)\mu_b(x_i, y_i)/\lambda(t_i, x_i, y_i), \\ w_i^{(w)} =& \mu_w(t_i)\mu_b(x_i, y_i)/\lambda(t_i, x_i, y_i), \\ \varphi_i =& \mu_0\mu_t(t)\mu_w(t)\mu_b(x, y)/\lambda(t_i, x_i, y_i), \\ \rho_{i,j} =& g(t_j - t_i)h(x_j - x_i, y_j - y_i)/\lambda(t_i, x_i, y_i), \end{aligned}$$

and by estimating the background terms

$$\begin{aligned} \hat{\mu}_t(t) &\propto \sum_i w_i^{(t)} K(t - t_i; \omega_t) / \int_0^T K(u - t_i; \omega_t) du, \\ \hat{\mu}_w(t) &\propto \sum_i w_i^{(w)} \sum_{k=0}^{\lfloor T/7 \rfloor} K(t - t_i + 7\lfloor t_i/7 \rfloor - 7k; \omega_w) / \int_0^T K(u - t_i; \omega_t) du, \\ \hat{\mu}_b(x, y) &\propto \sum_i \varphi_i \frac{K(x - x_j + x_i; \omega_x) K(y - y_j + y_i; \omega_y)}{\int\int_X K(x - x_j + x_i; \omega_x) K(y - y_j + y_i; \omega_y) \mathrm{d}u\mathrm{d}v}, \end{aligned}$$

where $K(x; \omega) = \frac{1}{\sqrt{(2\pi)\omega}} \exp\left(-\frac{x^2}{2\omega^2}\right)$ is the Gaussian Kernel, and $\lfloor x \rfloor$ is the largest integer not bigger than x.

The triggering terms are estimated as follows

$$\begin{aligned} \hat{g}(t) &\propto \frac{\sum_{i,j} \rho_{i,j} \times K(t - t_j + t_i; \omega_g) / \int_0^{T - t_i} K(u - t_j; \omega_g) \mathrm{d}u}{\sum_i I(t_i + t \leq T)}, \\ \hat{h}(x, y) &\propto \frac{\frac{\sum_{i,j} \rho_{i,j} \times K(x - x_j + x_i; \omega_{h_x}) K(y - y_j + y_i; \omega_{h_y})}{\int\int_S K(u - x_j + x_i; \omega_{h_x}) K(v - y_j + y_i; \omega_{h_y}) \mathrm{d}u\mathrm{d}v}}{\sum_i I\{(x_i + x, y_i + y) \in S\}}, \end{aligned}$$

Step 3. Estimate the relaxation parameters

$$\begin{aligned} A^{(k+1)} &= \frac{n - \sum_{i=1}^n \varphi_i^{(k)}}{G}, \\ \mu_0^{(k+1)} &= \frac{n - \sum_{i=1}^n A^{(k)} G}{U}, \end{aligned}$$

where

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

and G and U are the arguments that maximise the likelihood function with respect to μ_0 and A, respectively.

Indeed, since the likelihood function takes the form

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

where

$$\lambda(t, x, y) = \mu_0 \mu(t, x, y) + A \sum_{i:\, t_i < t} g(t_i - t, x_i - x, y_i - y),$$

and denoting $U = \int_0^T \iint_X \mu(t, x, y)\, \mathrm{d}x\mathrm{d}y\, \mathrm{d}t$ and $G = \int_0^T \iint_X \sum_{i:\, t_i < t} g(t - t_i, x - x_i, y - y_i)\, \mathrm{d}x\mathrm{d}y\, \mathrm{d}t$, then the equations $\frac{\partial}{\partial \mu_0} \log L = 0$ and $\frac{\partial}{\partial A} \log L = 0$ can be written as

$$\begin{aligned} \sum_{i=1}^{n} \frac{\mu(t_i, x_i, y_i)}{\lambda(t_i, x_i, y_i)} - U &= 0 \\ \sum_{i=1}^{n} \frac{\sum_{j: t_j < t_i} g(t_j - t_i, x_j - x_i, y_j - y_i)}{\lambda(t_i, x_i, y_i)} - G &= 0. \end{aligned}$$

Step 4. If convergence is reached, we obtain the estimates of μ_0 and A. Otherwise, we go back to Step 2.

4 New modelling approach on linear networks

Formally, a linear network $L = \cup_{i=1}^{n} l_i \subset \mathbb{R}^2$ is commonly taken as a finite union of line segments $l_i \subset \mathbb{R}^2$ of positive length Ang et al. (2012). The endpoints of the segments are called *nodes* and the *degree* of a node is the number of line segments that share the same node (Okabe and Sugihara, 2012). A line segment is defined as $l_i = [\mathbf{u}_i, \mathbf{v}_i] = \{k\mathbf{u}_i + (1-k)\mathbf{v}_i : 0 \leq k \leq 1\}$, where $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{R}^2$ are the endpoints of l_i. For any $i \neq j$ the intersection of l_i and l_j is either empty or an endpoint of

both segments. The total length of all line segments in L is denoted by $|L|$. The distance between two locations $\mathbf{u}$ and $\mathbf{v}$ in the network L is usually computed by the shortest-path distance $d_L(\mathbf{u}, \mathbf{v})$ which is the minimum of the length of all possible paths between $\mathbf{u}$ and $\mathbf{v}$. However, different possible distances have been discussed in Rakshit et al. (2017), see also Baddeley et al. (2021) for a recent review on statistical methods for analysing spatial point patterns on a network of lines.

4.1 Intensity estimation on linear networks

To adapt model in Equation (3.4) on the underlying spatial network, the main issue is to choose estimators for the spatial components $\hat{\mu}_b(x, y)$ and $\hat{h}(x-u, y-v)$, taking properly into account the underlying network structure.

In this section, we review the most recent proposals for intensity estimation on linear networks. A natural first step in analysing spatial patterns is to form a kernel density estimate (Silverman, 1986) of the spatially-varying events rate. However, this is not straightforward on a linear network.

Concerning intensity estimation, several kernel-based methods (Borruso, 2005, 2008; Xie and Yan, 2008; Okabe et al., 2009; Okabe and Sugihara, 2012; McSwiggan et al., 2017; Moradi et al., 2018; Rakshit et al., 2019a) and a resample-smoothing technique applied to Voronoi intensity estimators (Moradi et al., 2019; Mateu et al., 2019) have been proposed.

Obtaining good estimates for intensity functions of point processes on linear networks has been a challenging task due to geometrical complexities and unique methodological problems (Cronie et al., 2020). Nevertheless, there have been a few particularly

interesting proposals. Baddeley et al. (2021) reviews the most recent approaches for spatial kernel density estimation on linear networks. We summarise here the main contributions

Approaches based on sums of kernels (Xie and Yan, 2008) basically assume that the statistical basis for kernel estimation can be transferred from the real line to the linear network. However, several authors (Okabe et al., 2009; Okabe and Sugihara, 2012; McSwiggan et al., 2017) have highlighted that translating the kernel density estimate on the one-dimensional real line directly to the linear network, is a fallacious estimate because it does not conserve mass. This basically means that the kernel $K(\cdot, x_i) = k(d_L(\cdot, x_i))$ is not a probability density on the linear network, and therefore the corresponding density estimate is not a a probability density. The true probability density would be roughly overestimated in the denser parts of the network.

Then, Borruso (2005, 2008) and Okabe et al. (2009) proposed several modifications of the kernel sums estimator. Moradi et al. (2018) and McSwiggan et al. (2020) pointed out that the bias can be removed by adapting classical edge-corrections from spatial statistics, thus proposing some *edge-corrected classical kernel-based intensity estimators*.

As per *path enumeration methods*, (Okabe et al., 2009; Sugihara et al., 2010; Okabe and Sugihara, 2012) considered computational algorithms which start with a kernel k on the real line, and progressively redistribute the mass of this kernel over the network. These authors proposed two kernel estimators satisfying many desired properties: (a) the "equal-split continuous" rule, which has excellent properties (symmetric, conserves mass, and is unbiased when the true intensity is uniform) but it is extremely slow to compute; (b) the "equal-split discontinuous" rule, which is faster, but has less

desirable properties (see Okabe and Sugihara (2012) and McSwiggan et al. (2017) for the algorithms).

Another proposal is represented by the *heat kernel*: it is the network counterpart of the Gaussian kernel. It was developed by McSwiggan et al. (2017) by exploiting the connection between kernel smoothing and diffusion. Since the heat kernel estimator is mathematically equivalent to the "equal-split continuous" estimator extended to the Gaussian kernel, it has the same properties.

Finally, a *fast kernel smoothing using two-dimensional convolutions* has been recently proposed by Rakshit et al. (2019b). It is a computationally efficient and statistically principled method for kernel smoothing of point pattern data on a linear network. The point locations, and the network itself, are convolved with a two-dimensional kernel and then combined into an intensity function on the network. This can be computed rapidly using the Fast Fourier Transform (FFT), even on large networks and for large bandwidths, and is robust against errors in network geometry. The estimator is consistent, and its statistical efficiency is only slightly suboptimal.

A crucial problem with kernel estimation is that if there are wide variations in intensity across the spatial domain, it may be impossible to find a single fixed bandwidth value which is satisfactory for smoothing every part of the spatial domain. Consequently, the bandwidth can be spatially-varying, giving rise to a spatially "adaptive" kernel estimator at the cost of increased complexity. Recently, there has been an increasing interest in point patterns on linear networks. Here the matter of kernel estimation is even more delicate due to the geometry of the underlying network. For example, street crimes and traffic accidents tend to happen particularly in busy streets, which may be surrounded by quiet neighbourhoods. In such cases, classical

kernel estimation approach is often unsuitable for such types of data, and alternatives based on *resample-smoothing of addaptive Voronoi intensity estimators* have been proposed (Moradi et al., 2019).

4.2 A Hawkes point process on linear networks

The spatio-temporal Hawkes process model on the linear network L that we propose has the following specification

$$\lambda(t,x,y) = \mu_0\mu_t(t)\mu_w(t)\mu_L(x,y) + A\int_{-\infty}^{t-}\int\int_L g(t-s)h_L(x-u,y-v)N(\mathrm{d}u\times\mathrm{d}v\times\mathrm{d}s), \tag{4.1}$$

where $\mu_L(x,y)$ and $h_L(x-u,y-v)$ are computed using the 2D convolutional Gaussian Kernel of Rakshit et al. (2019b), which represents the best alternative for our purposes. Indeed, Rakshit et al. (2019b) proposed a kernel estimator on a linear network based on a 2D smoothing kernel. The original motivation was speed: this estimator can be expressed in terms of 2D convolutions of the kernel, so it can be computed very rapidly using the FFT.

Definition 2 ***Convolution Kernel Estimator***

Let $\mathbf{x} = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ *be a point pattern on a linear network* L. *Let* κ *denote a bivariate kernel function, that is, a probability density on* $\mathbb{R}^2$. *The convolution kernel estimator of the intensity is, with the uniform correction,*

$$\hat{\lambda}^U(\mathbf{u}) = \frac{1}{c_L(\mathbf{u})}\sum_{i=1}^{n}\kappa(\mathbf{u}-\mathbf{x}_i), \quad \mathbf{u}\in L, \tag{4.2}$$

and with Jones–Diggle correction, it becomes

$$\hat{\lambda}^{JD}(\mathbf{u}) = \sum_{i=1}^{n}\frac{\kappa(\mathbf{u}-\mathbf{x}_i)}{c_L(\mathbf{x}_i)}, \quad \mathbf{u}\in L, \tag{4.3}$$

where

$$c_L(\mathbf{u}) = \int_L \kappa(\mathbf{u} - \mathbf{v}) \mathrm{d}_1\mathbf{v}, \quad \mathbf{u} \in L. \tag{4.4}$$

The denominator $c_L(\mathbf{u})$, defined in Equation (4.4), is the convolution of the kernel with the arc-length measure on the network. This function is evaluated only at locations on the network.

It is theoretically possible to choose a kernel κ that is not isotropic, that is, not invariant under rotation. This seems undesirable in practice, except in situations where the coordinate system is not isometric, such as the latitude–longitude coordinates on a globe. We thus assume here that κ is isotropic. Unlike estimators of the intensity based on path distances in the network, the convolution estimators are robust against errors in the geometry of the network. If κ is uniformly continuous, the quantities (4.2) and (4.3) are continuous functions of the point pattern $\mathbf{x}$ and the linear network L, and the sums in (4.2) and (4.3), and the integral (4.4) can be recognised as convolutions of the kernel κ with different measures M on $\mathbb{R}^2$ (Rakshit et al., 2019b).

Therefore, the fitting procedure for our model in Equation (4.1) follows the same strategy outlined in Section 3.2 with the main modification concerning the planar spatial intensities $\mu_b(x, y)$ and $h(x, y)$ that are substituted by their network counterparts $\mu_L(x, y)$ and $h_L(x, y)$, computed using the 2D convolutional Gaussian Kernel of Rakshit et al. (2019b), and implemented in Baddeley and Turner (2005) package of the software R Core Team (2021).

4.3 Further extensions: dependence on external covariates

The model in Equation (4.1) can be extended by including external spatial covariates in $\mu_L(x, y)$. In the context of spatial point processes, if $Z(x, y)$ is referred to a spatial covariate, this means that its value is assumed to be observable, at least in principle, at each location (x, y) in the region of interest. For inferential purposes, its values must be known at each point of the data point pattern and at least at some other locations. This specification gives rise to

$$\lambda(t, x, y) = \mu_0 \mu_t(t) \mu_w(t) \mu_L(x, y, \boldsymbol{\beta}_{back}) + A \int_{-\infty}^{t-} \int \int_L g(t-s) h_L(x-u, y-\nu) N(\mathrm{d}u \times \mathrm{d}\nu \times \mathrm{d}s), \tag{4.5}$$

where $\boldsymbol{\beta}_{back}$ denotes the parameters associated to the spatial covariates $Z(x, y)$ included in the model.

Note that if the main interest would be focused on the selection of the covariates, the algorithm in Section 3 would have been modified in order to maximise the likelihood also with respect to the parameters of the external covariates. Otherwise, as we are interested in achieving the best fitting, we use the same algorithm, and the selection of the variables can be performed during the setting of the first guesses, and those can be held fixed in the E-M.

As all the available covariates are continuous in space, these can be included linearly by choosing a basis function. If $b_i(x)$ is the i^{th} such basis function, then the function f is assumed to have a representation

$$f(x) = \sum_{i=1}^{q} b_i(x) \beta_i, \tag{4.6}$$

for some values of the unknown parameters, β_i.

The degree of smoothing is controlled by the basis dimension. In this paper, we keep the basis dimension fixed at a size a bit larger than it is believed could reasonably be necessary. After having compared different number of knots, these are chosen to be equal to 5 for the available external covariates, and equal to 30 for the spatial coordinates, estimated here through thin plate regression splines.

5 Data analysis

5.1 Model selection

As presented in Section 2, we analyse 2671 armed robberies in the city of Bucaramanga, Colombia, in 2018. We fitted Zhuang and Mateu (2019) model (recall this is on the Euclidean plane) and our proposed extensions (Equations (4.1) and (4.5)) to the data. Table 1 displays the bandwidths of the background events and the windows of the triggered events that were adopted for the models. These values are identical in every model to ensure comparability and inference. Note that we omitted the daily periodicity term $\mu_t(t)$, and thus its bandwidth, as our data lacks hourly crime data information. The weekly periodicity and long-term bandwidths were selected to secure the resolution requirements of the temporal components. We chose the spatial and temporal extent (15 days and 2.5 km, respectively) to restrict the spatio-temporal domain in which we expect non-spontaneous events to occur.

We estimated the relaxation coefficients $\hat{\mu}$ and $\hat{A}$ through a 40-loops iterative algorithm as in Zhuang and Mateu (2019). We also computed the log-likelihood to assess the fit of the three space-time point process models. Table 2 reports the estimates

Table 1: Temporal bandwidths for the background events and spatial and temporal windows for the triggered events

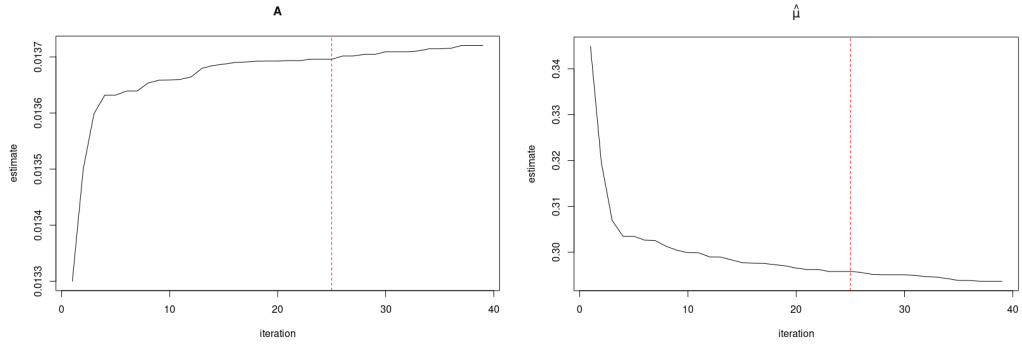
Background	Bandwidth
Weekly periodicity	0.4 days
Long-term	0.6 days
Triggered	Window
Temporal	15 days
Spatial	2.5 km

of the relaxation coefficients and the corresponding log-likelihood. $\hat{\mu}$ and $\hat{A}$ values reached convergence after the 25th iteration in all approaches (Figure 5). The convergence threshold was established as the iteration on which the difference between the values of the k^{th} and $k^{th} - 1$ iteration is smaller than 0.0001 units.

The estimates of $\hat{\mu}$ have similar magnitudes in the planar model and linear network model without background covariates. The effect of the covariates in the estimation of $\hat{\mu}$ is evident; the value in the model with covariates, model in Equation (4.5) decreases significantly compared to its counterparts. A higher value of $\hat{\mu}$ for the linear network model in Equation (4.1) compared to the model in the planar case, reflects the change of spatial support, that is, the expected number of events increases when estimated per linear segment instead of per surface. Nonetheless, the inclusion of covariates adjusts for the change of support forcing $\hat{\mu}$ to drop. Note that the convergence rate of $\hat{\mu}$ for model in Equation (4.5) is smoother than the one of the model without covariates.

$\hat{A}$ changes noticeably from specifications (3.4) to specifications (4.1) and (4.5). In the

(a)



(b)

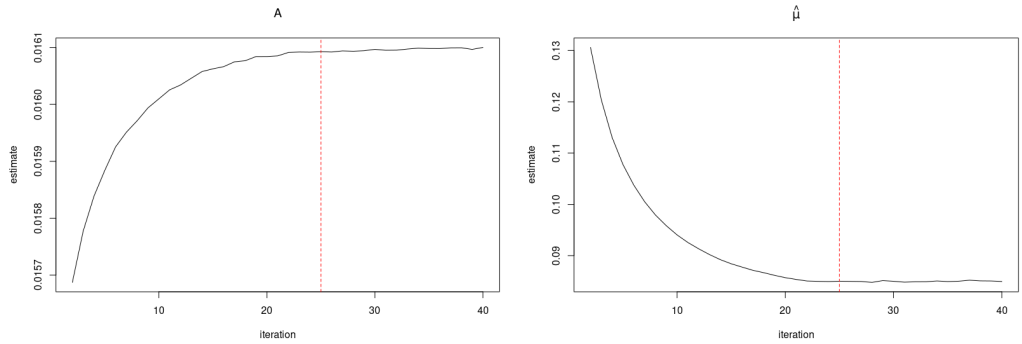


Figure 5: Relaxation coefficients convergence: (a) Model (4.1) and (b) Model (4.5)

Table 2: Comparative results for the three fitted models

Model	$\hat{\mu}$	$\hat{A}$	$\log(L)$
Model on the plane (3.4)	0.204	0.639	-706.47
Model on the network (4.1)	0.292	0.014	-353.35
Model on the network plus covariates (4.5)	0.085	0.016	-319.70

former, almost 64% of the crimes are triggered, while the latter inform that only 1.4% and 1.6% (respectively) of armed robberies are provoked by previous crimes. The $\hat{A}$ estimate of our models better resembles Bucaramanga's robbery offenses scenario: robberies typically obey the crime opportunity theory, to wit, a robber's motivation arises from the context and factors involved in the environment they are situated (Larrota et al., 2017). One would expect a low triggering effect as the offenders rather exploit the situational factors (spontaneous events) than premeditate their crimes' circumstances (non-spontaneous events).

The log-likelihood shows that our models fit better the armed robberies data with a difference of 353.12 and 386.77 units for models (4.1) and (4.5), respectively. We fit both the background and the triggering effect directly on a linear network, and we get accurate estimates and model the data appropriately. Furthermore, introducing covariates in the background rate improves the estimation of the relaxation coefficients and the overall fitting of the model.

To demonstrate the advantages of the linear network models (4.1) and (4.5), we compare their spatial results with the ones of the planar model (Equation (3.4)). Figure 6 displays the spatial background rate for the three self-exciting point process models. The spatial background rates share similarities in every model. Although the models'

background rates present close resemblance, the rates in the linear network models are higher than those in the planar one. As we discussed before, the rates increase in the linear network models due to the reduction of the area unit. The distribution of the intensities in Bucaramanga's city center is equivalent in all models. However, the linear network models provide larger estimated intensities in the central and north-western regions of the study area. In particular, the specification from Equation (4.5) displays larger and rougher high-intensity areas. This is primarily attributed to the covariates effects as these affect the estimates per segment rather than per unit of space. Also, note that the planar model indeed captures the "network" nature of the data, but the clustering effect extends to areas instead of street segments. This might lead to misinterpreting low-crime rate streets as dangerous ones when these connect to neighbouring streets with high crime rates. Furthermore, as the planar model is defined over continuous space, it generates artificial intensities in locations lacking events.

One of the main differences between Zhuang and Mateu (2019) model and our proposed approaches is the definition of the spatial triggering component. Figure 7 compares the spatial response function of the planar and network models. The spatial response function of the planar model remains constant across events. The occurrence of an armed robbery in a specific location will excite future armed robberies within a radial distance of approximately 100 meters. This is an unrealistic assumption for events located in a linear network as the extent of the spatial triggering effect depends intrinsically on the street topology. Figure 7(b) shows the triggering function for four armed robberies. In our approach, $h_L(x, y)$ adapts to the street configuration on which each event lies. The triggering window on the linear network varies from event to event, preserving the geometrical properties of network structure, for example, the

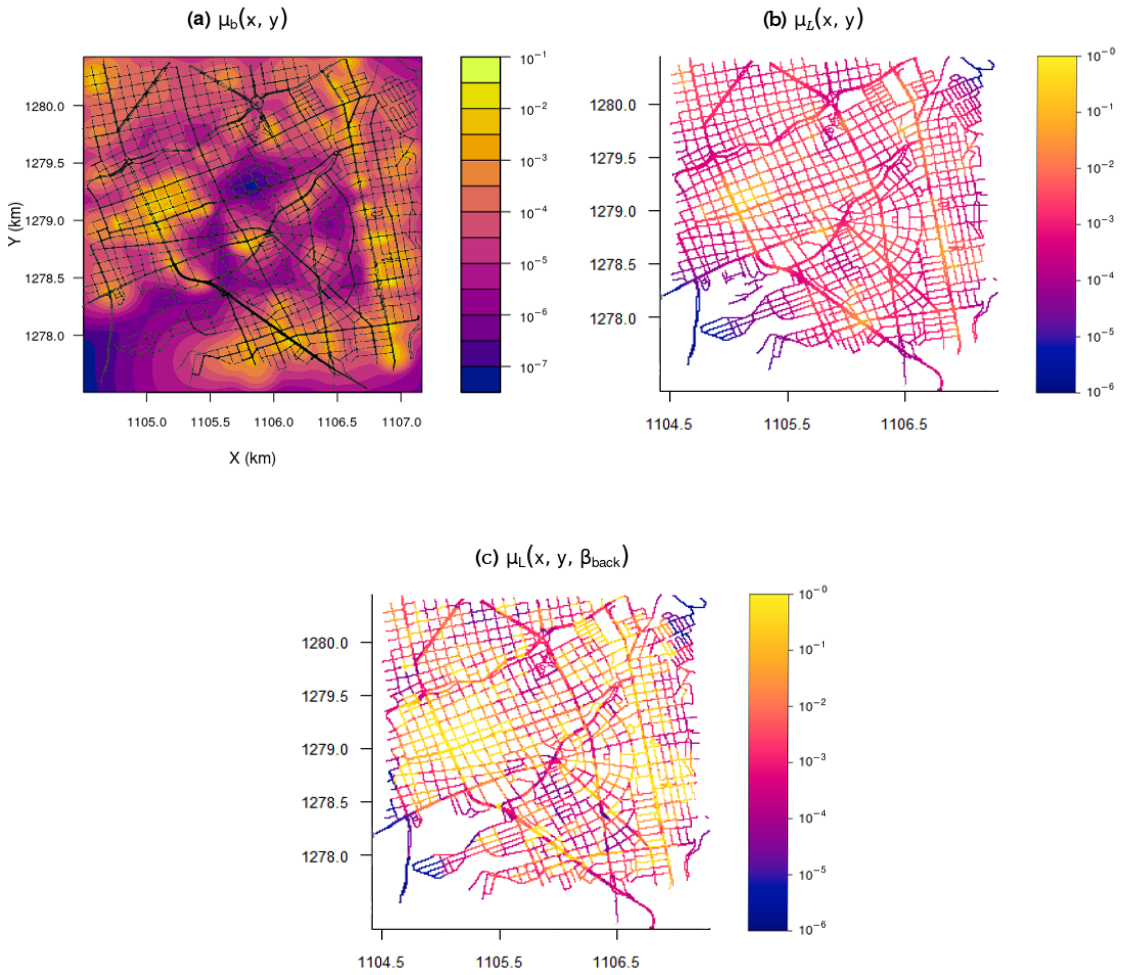


Figure 6: Spatial background rate: (a) planar model $\mu_b(x, y)$ (Equation (3.4)), (b) linear network model $\mu_L(x, y)$ (Equation (4.1)) and (c) linear network model with background covariates $\mu_L(x, y, \boldsymbol{\beta}_{back})$ (Equation (4.5)). Rates are reported on the log-scale

shortest path distance between events.

Figure 8 displays the temporal components of the linear network models. The temporal results in both models were identical. The long-term trend (Figure 8(a)) indicates that the armed robberies occur primarily in the second semester of the year. The crime rates increase from June onwards and reach their maximum value in December. In Colombia, the period from June to December comprises most of the annual public and school holidays. As the number of tourists increases in the commercial and touristic areas of the country's main cities (such as Bucaramanga), local surveillance and security weaken, promoting higher crime rates. The weekly periodicity implies that robberies occur mainly on Wednesdays, Fridays, and Saturdays. The original data also reports these days as the days with the highest frequencies in armed robberies. On Wednesdays, the city center's main venues such as restaurants, bars, and cinemas offer discounts on their services, attracting many locals. Fridays and Saturdays are naturally the days when most people commute to the city center for recreational activities. People leaving venues in late hours and limited public transport provide the offenders the ideal circumstances to take advantage of. The temporal response function (Figure 8(c)) suggests that after an armed robbery occurs, subsequent armed robberies are triggered within the coming ten days. However, the triggering effect lessens after one week.

5.2 Covariates results

As previously introduced, we select the covariates to include in model in Equation (4.5) prior to the E-M procedure. Among the significant external covariates affecting the background spatial intensity, we have cycleway, pedestrian, primary, residential,

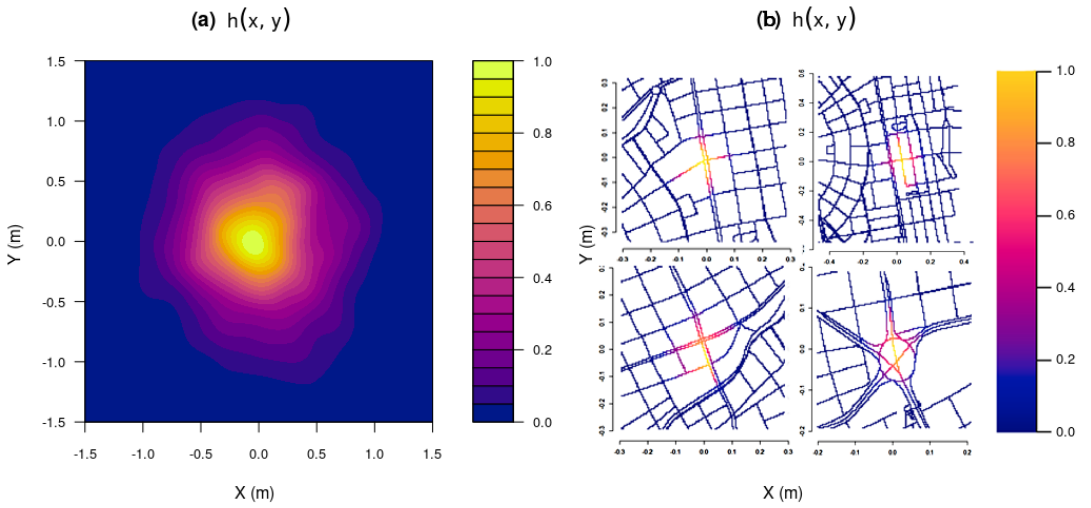


Figure 7: Spatial triggering function: (a) planar model $h(x, y)$ and (b) linear network model $h_L(x, y)$

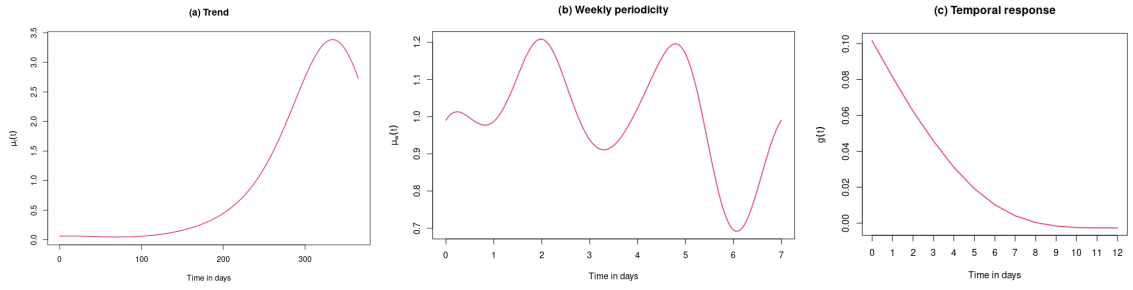


Figure 8: Temporal outputs: (a) Long-term function $\mu_t(t)$; (b) weekly periodicity $\mu_w(t)$; (c) temporal response function $g(t)$

secondary, service, and tertiary highways, bridges and steep slope. Indeed, all of these binary variables are significant in explaining the crime pattern and their estimated coefficients are all positive meaning that the crimes are most likely to occur in those places. Only oneway street and flat slope coefficients are estimated as negative values, meaning that crimes are less likely to occur there.

In Figure 9, we show the marginal relationship among the significant continuous covariates and the intensity of the point pattern under study. The non-parametric combination of these variables plus the event coordinates best explains (based on the Akaike criteria) the linear network point pattern.

5.3 Residual analysis

As stated in some previous papers (Adelfio et al., 2020; Adelfio and Schoenberg, 2009), the main problem when dealing with residual analysis for point processes is to find a correct definition of residuals, since the one used in dependence models cannot be used for point processes.

For the temporal domain diagnostics of the fitted model, the marginal time process can be obtained by integrating the estimated intensity function with respect to the observed spatial region (Adelfio and Chiodi, 2015). Indeed, given a point process $N = \{(t_i, x_i, y_i), i = 1, \ldots, n\}$ which is determined by a conditional intensity $\lambda(t, x, y|\mathcal{H}_t)$, the transformation

$$t_i \to \tau_i = \int_0^{t_i} \int_X \lambda(u, x, y|\mathcal{H}_t) \mathrm{d}x \mathrm{d}y \mathrm{d}u$$

transforms N into a stationary Poisson process N' with unit rate (Vere-Jones and Schoenberg, 2004). The resulting process is called the transformed time sequence

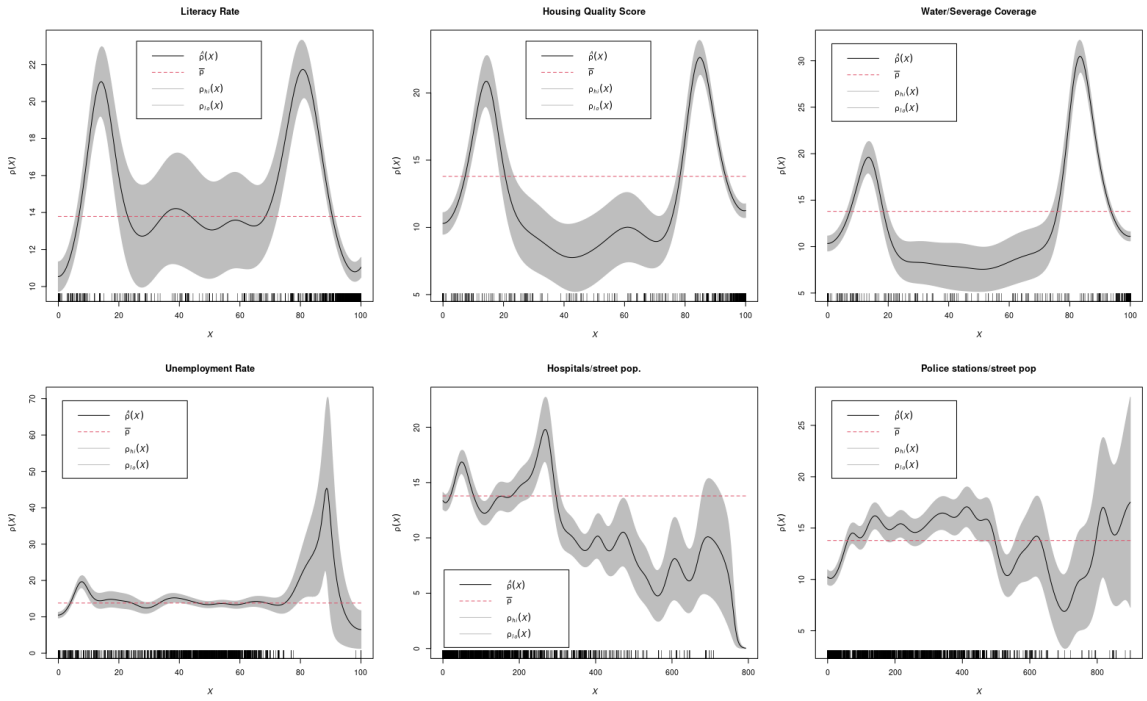


Figure 9: Non-parametric estimate of the intensity of the analysed point pattern, as a function of the available variables for the crime data: literacy rate (a), housing quality score (b), water coverage (c), unemployment rate (d), number of hospitals per street population (e), and number of police stations per street population (f). The considered smoothing procedure is based on fixed-bandwidth kernel density estimation

(Ogata and Katsura, 1988). Then, a plot of τ_i versus i can give insight about the quality of the fitting in time, concluding that the model has a good fit to the data if the transformed time sequence does not deviate significantly from the standard Poisson process. In particular, this plot, together with a plot of the estimated time intensities, informs on the time at which departures from model assumptions are more evident. See Schoenberg (2002) for details for the bands construction.

We thus consider this type of residual analysis for the best fitted model in the previous section, i.e. the model based on Equation (4.5). Figure 10 depicts both the cumulative frequencies of the original time sequence, and the transformed time for the data for all three models for comparison purposes. We, in particular, follow the recommendation by Zhuang and Mateu (2019) which compare the real and the transformed rates of event occurrences.

Inspecting Figure 10, we note that model (4.5) provides the best fit, with the cumulative frequencies lying within the confidence interval at 95%. Although the transformed times deviate from the average rate of occurrence in some points, the overall trend remains consistent. Transformed times diverge from the actual time, mostly in low crime rates periods, for example, in the first months of the year when just a few armed robberies were reported. Furthermore, the fluctuation in the rate of occurrence of the transformed times could be attributed to the kernel estimation.

Similarly to the time component, one of the mostly used methods for diagnostics of spatial point processes are the smoothed raw residuals, which follow the same line of reasoning. For an inhomogeneous Poisson process model, with fitted intensity $\hat{\lambda}(\mathbf{u})$ in space, the predicted number of points falling in any region D is $\int_D \hat{\lambda}(\mathbf{u})\mathrm{d}\mathbf{u}$. Hence, the residual in each region $D \subset \mathbb{R}^2$ is the observed minus predicted number of points

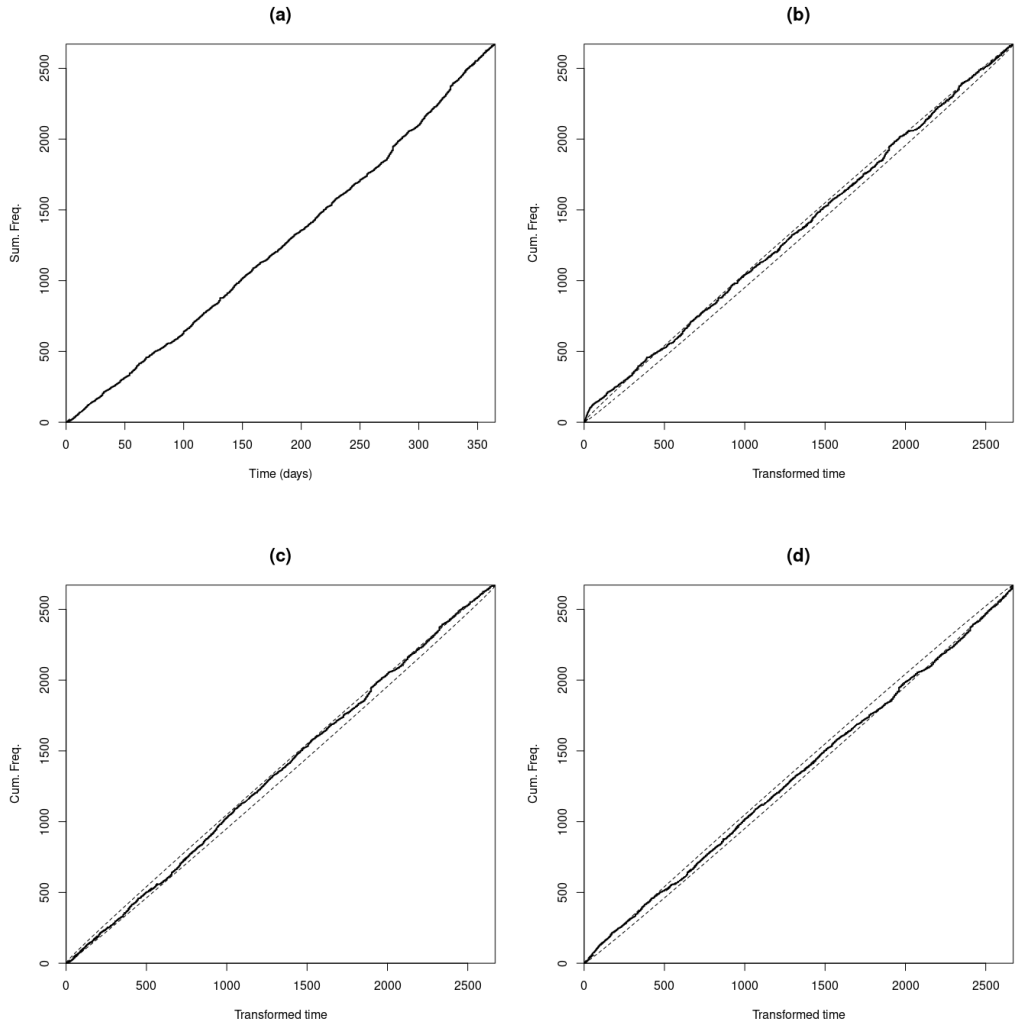


Figure 10: Cumulative frequencies of armed robberies events: (a) original occurrence times; (b) transformed times for model (3.4); (c) transformed times for model (4.1); and (d) transformed times for model (4.5). The 95% confidence bands under a Poisson process are also shown in dashed lines

falling in D (Alm, 1998), that is, $R(D) = n(\mathbf{x} \cap D) - \int_D \hat{\lambda}(\mathbf{u})d\mathbf{u}$, where $\mathbf{x}$ is the observed point pattern, $n(\mathbf{x} \cap D)$ the number of points of $\mathbf{x}$ in the region D, and $\hat{\lambda}(\mathbf{u})$ is the intensity of the fitted model. A simple visualisation of these residuals can be obtained by smoothing. The smoothed raw residual fields are defined as

$$s(\mathbf{u}) = \tilde{\lambda}(\mathbf{u}) - \lambda^{\dagger}(\mathbf{u}) \tag{5.1}$$

where $\tilde{\lambda}(\mathbf{u}) = e(\mathbf{u}) \sum_{i=1}^{n(\mathbf{x})} \kappa(\mathbf{u} - \mathbf{x}_i)$ is the non-parametric, kernel estimate of the fitted intensity $\hat{\lambda}(\mathbf{u})$, while $\lambda^{\dagger}(\mathbf{u})$ is a correspondingly smoothed version of the (typically parametric) estimate of the intensity of the fitted model, $\lambda^{\dagger}(\mathbf{u}) = e(\mathbf{u}) \int_D \kappa(\mathbf{u} - \mathbf{v})\hat{\lambda}(\mathbf{v})d\mathbf{v}$. Here, κ is the smoothing kernel and $e(\mathbf{u})$ is the correction for edge effects in the window D given by $e(\mathbf{u})^{-1} = \int_D \kappa(\mathbf{u} - \mathbf{v})d\mathbf{v}$ (Baddeley et al., 2005). The smoothing bandwidth for the kernel estimation of the raw residuals is selected by cross-validation, as such value that minimises the Mean Squared Error criterion defined by Diggle (1985), by the method of Berman and Diggle (1989). See Diggle (2013) for further details. We employ the same reasoning, knowing that $\lambda^{\dagger}(\mathbf{u})$ is the smoothed version of the estimate of the intensity of the fitted models, and of course $\mathbf{u} \in L$. The difference in Equation (5.1) should be approximately zero when the fitted model is close to the real one. Therefore, the best model is the one with the lowest values of the smoothed raw residuals.

Figure 11 depicts the smoothed raw residuals for the two models fitted on the network, and for the model in the planar case, along with their distributions. Even though smoothed raw residuals are most appropriate for parametric specifications of fitted models, we still manage to obtain useful information. Indeed, it appears evident that both models on the network achieve a good fit to the data, if compared to their planar version, which tends to overestimate the intensity in several areas. Furthermore, the

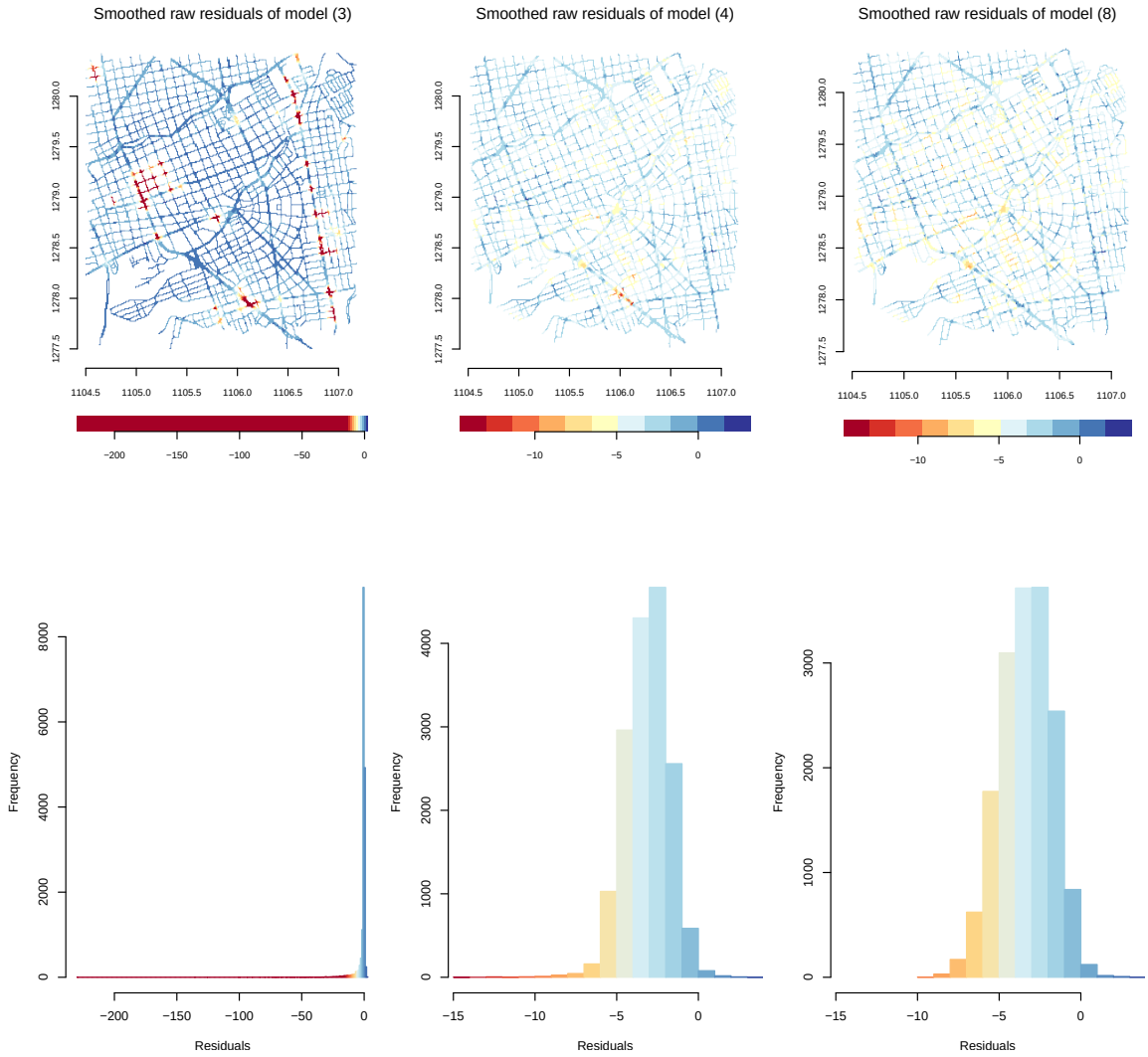


Figure 11: Maps and histograms of smoothed raw residuals for models (3.4), (4.1), and (4.5), respectively

model with covariates provides residuals with a lower range and magnitude compared to the model wit no covariates. All in all, the smoothed raw residuals of the model with covariates are overall better than the ones of the model with no covariates, even though both models still represent good competitors if compared to the planar version. We note that there were other options for carrying out diagnostics in this paper. First, superthinned residuals (Clements et al., 2012), which have the disadvantage of being less straightforwardly applicable, as they require a tuning parameter to be chosen. Another possible option could have been using the weighted spatio-temporal second order statistics, as in Adelfio et al. (2020), which do not require any transformation of the data. However, we opted for diagnostic procedures separated in space and time, due to the separable specification of the models employed in the paper. Nevertheless, the chosen diagnostics allows us to interpret separately the contribution of the space specification into the network, as evident from Figure 11.

6 Conclusions and future developments

In this paper, we analyse robbery crimes as events of a spatio-temporal point pattern living on a linear network structure. We first fit a Euclidean planar model following Zhuang and Mateu (2019) providing two parameters related to the background and triggered occurrence rates. We further propose an extension of that model in order to take into account the linear network. This allows us both to include the network geometry into the fitting procedure, but most importantly, to draw conclusions on the scale of shortest-path distances, more appropriate when dealing with point processes occurring on linear networks. Furthermore, starting from the specification of this

second model, we are also able to include the dependence on external covariates rightly constrained onto the spatial support of the road network. We find that our proposed models on the network achieve a much better fit when compared to the planar counterpart.

As future work, we draw different lines of research, both methodological and applied. From a methodological perspective, the proposed self-exciting model could be extended by including external covariates in the triggered components, as proposed by Meyer et al. (2012) in an epidemiological context, and Adelfio and Chiodi (2020) in the seismic one. Indeed, the model could include individual-specific covariates in the triggered component of the conditional intensity function, by means of an exponential linear predictor. This would allow for including the dependence on some specific characteristics of the event i.e. the crime, such as gender, or age of the victim.

Another main issue is related to the fitting procedure. Indeed, the proposed algorithm could be modified in order to include the estimation of the covariates parameters at each iteration of the Expectation-Maximization procedure. This would allow to consider only the crimes most likely to be background events when selecting the spatial covariates, as done by Park et al. (2021), since, in principle, only background crimes should be used. The best model, and therefore the best set of covariates, could be selected by comparing some information criterion of the fitted competitor models.

Finally, we could follow a multivariate specification analysing several types of crimes in only one modelling strategy.

References

Adelfio, G. and Chiodi, M. (2015). FLP estimation of semi-parametric models for space-time Point Processes and diagnostic tools. *Spatial Statistics*, **14**, 119–132.

Adelfio, G. and Chiodi, M. (2020). Including covariates in a space-time point process with application to seismicity. *Statistical Methods & Applications*. ISSN 1613-981X. doi: 10.1007/s10260-020-00543-5. URL https://doi.org/10.1007/s10260-020-00543-5.

Adelfio, G. and Schoenberg, F. P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Annals of the Institute of Statistical Mathematics*, **61**(4), 929–948.

Adelfio, G., Siino, M., Mateu, J., and Rodríguez-Cortés, F. J. (2020). Some properties of local weighted second-order statistics for spatio-temporal point processes. *Stochastic Environmental Research and Risk Assessment*, **34**(1), 149–168.

Alm, S. E. (1998). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes*, **1**(1), 111–126.

Anderes, E., Møller, J., and Rasmussen, J. G. (2020). Isotropic covariance functions on graphs and their edges. *Annals of Statistics*, **48**, 2478–2503.

Ang, Q. W., Baddeley, A., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, **39**(4), 591–617.

Baddeley, A. and Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, **12**(6), 1–42. URL https://www.jstatsoft.org/v12/i06/.

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(5), 617–666.

Baddeley, A., Nair, G., Rakshit, S., and McSwiggan, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat*, **6**(1), 68–78.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. (2021). Analysing point patterns on networks — A review. *Spatial Statistics*, **42**, 100435. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2020.100435. URL https://www.sciencedirect.com/science/article/pii/S2211675320300294. Towards Spatial Data Science.

Berman, M. and Diggle, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society: Series B (Methodological)*, **51**(1), 81–92.

Borruso, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications*, pages 126–132. Springer.

Borruso, G. (2008). Network density estimation: a GIS approach for analysing point patterns in a network space. *Transactions in GIS*, **12**(3), 377–402.

Chiodi, M., Nicolis, O., Adelfio, G., D'angelo, N., and Gonzàlez, A. (2021). Etas space–time modeling of chile triggered seismicity using covariates: Some preliminary results. *Applied Sciences*, **11**(19), 9143.

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

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Cronie, O., Moradi, M., and Mateu, J. (2020). Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and Computing*, **30**(5), 1221–1239.

Daley, D. J. and Vere-Jones, D. (2007). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Springer-Verlag, New York, second edition.

D'Angelo, N., Adelfio, G., Abbruzzo, A., and Mateu, J. (2021a). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The annals of applied statistics*, **Forthcoming**.

D'Angelo, N., Adelfio, G., and Mateu, J. (2021b). Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. *Spatial Statistics*, **45**, 100534. ISSN 2211-6753.

Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **34**(2), 138–147.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC.

Hawkes, A. and Adamopoulos, L. (1973). Cluster models for erthquakes-regional comparison. *Bulletin of the International Statistical Institute*, **45**(3), 454–461.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, **33**(3), 438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**(1), 83–90.

Larrota, R., Gómez, M., and Beltrán, C. (2017). Modus operandi en personas recluidas por el delito de hurto en prisiones de bucaramanga, colombia [Modus operandi in persons detained for the crime of theft in prisons of Bucaramanga, Colombia]. *INFORMES Psicologicos [REPORTS in Psychology]*, **17**(2), 107–118.

Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, **319**(5866), 1076–1079.

Mateu, J., Moradi, M., and Cronie, O. (2019). Spatio-temporal point patterns on linear networks: Pseudo-separable intensity estimation. *Spatial Statistics*, page 100400.

McSwiggan, G., Baddeley, A., and Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, **44**(2), 324–345.

McSwiggan, G., Baddeley, A., and Nair, G. (2020). Estimation of relative risk for events on a linear network. *Statistics and Computing*, **30**(2), 469–484.

Meyer, S., Held, L., and Hohle, M. (2012). Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance. *Journal of Statistical Software*, **77** (11).

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**(493), 100–108.

Moradi, M. M. and Mateu, J. (2019). First-and Second-Order Characteristics of Spatio-Temporal Point Processes on Linear Networks. *Journal of Computational and Graphical Statistics*, pages 1–21.

Moradi, M. M., Rodríguez-Cortés, F. J., and Mateu, J. (2018). On kernel-based intensity estimation of spatial point patterns on linear networks. *Journal of Computational and Graphical Statistics*, **27**(2), 302–311.

Moradi, M. M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., and Baddeley, A. (2019). Resample-smoothing of Voronoi intensity estimators. *Statistics and computing*, **29**(5), 995–1010.

Newman, M. (2010). *Networks: An Introduction.* New York: Oxford University Press.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. *Annals of the Institute of Statistical Mathematics*, **40**(1), 29–39.

Okabe, A. and Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods.* John Wiley & Sons.

Okabe, A., Satoh, T., and Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science*, **23**(1), 7–32.

Park, J., Schoenberg, F. P., Bertozzi, A. L., and Brantingham, P. J. (2021). Investigating clustering and violence interruption in gang-related violent crime data

using spatial–temporal point processes with covariates. *Journal of the American Statistical Association*, pages 1–14.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rakshit, S., Nair, G., and Baddeley, A. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, **22**, 129–154.

Rakshit, S., Baddeley, A., and Nair, G. (2019a). Efficient Code for Second Order Analysis of Events on a Linear Network. *Journal of Statistical Software*, **90**(1), 1–37.

Rakshit, S., Davies, T., Moradi, M. M., McSwiggan, G., Nair, G., Mateu, J., and Baddeley, A. (2019b). Fast Kernel Smoothing of Point Patterns on a Large Network using Two-dimensional Convolution. *International Statistical Review*, **87**(3), 531–556.

Rasmussen, J. G. and Christensen, H. S. (2021). Point processes on directed linear networks. *Methodology and Computing in Applied Probability*, **23**(2), 647–667.

Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, **33**(3), 299–318.

Schoenberg, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica*, **26**, 861–879.

Schoenberg, F. P. (2002). On rescaled Poisson processes and the Brownian bridge. *Annals of the Institute of Statistical Mathematics*, **54**(2), 445–457.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Sugihara, K., Satoh, T., and Okabe, A. (2010). Simple and unbiased kernel function for network analysis. In *2010 10th International Symposium on Communications and Information Technologies*, pages 827–832. IEEE.

van Lieshout, M. (2018). Nearest-neighbour Markov point processes on graphs with Euclidean edges. *Advances in applied probability*, **50**(4), 1275–1293.

Vere-Jones, D. and Schoenberg, F. (2004). Rescaling Marked Point Processes. *Australian & New Zealand Journal of Statistics*, **46**, 133–143. doi: 10.1111/j.1467-842X.2004.00319.x.

Xie, Z. and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, **32**(5), 396–406.

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**(4), 635–653.

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

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, **109**(B5).