



Semi-parametric Spatio-Temporal Hawkes Process for Modelling Road Accidents in Rome

Pierfrancesco ALAIMO DI LORO, Marco MINGIONE , and Paolo FANTOZZI

We propose a semi-parametric spatio-temporal Hawkes process with periodic components to model the occurrence of car accidents in a given spatio-temporal window. The overall intensity is split into the sum of a background component capturing the spatio-temporal varying intensity and an excitation component accounting for the possible *triggering* effect between events. The spatial background is estimated and evaluated on the road network, allowing the derivation of accurate risk maps of road accidents. We constrain the spatio-temporal excitation to preserve an isotropic behaviour in space, and we generalize it to account for the effect of covariates. The estimation is pursued by maximizing the expected complete data log-likelihood using a tailored version of the stochastic-reconstruction algorithm that adopts ad hoc boundary correction strategies. An original application analyses the car accidents that occurred on the Rome road network in the years 2019, 2020, and 2021. Results highlight that car accidents of different types exhibit varying degrees of excitation, ranging from no triggering to a 10% chance of triggering further events.

Keyword Hawkes process; Road accidents; Spatio-temporal; Kernel estimation; Point process; EM algorithm.

1. INTRODUCTION

According to the latest report of the National Statistical Institute, more than 150,000 road accidents occurred during 2021 in the Italian territory, of which 2875 were deaths within 30 days. Early estimates for the first semester in 2022 (January–June) highlight an increasing—though not alarming—number of road accidents compared to the pandemic period. The major causes of such events can be addressed as distracted driving, failure to give way, and speeding. Although the long-term trend of road accidents in Italy has been decreasing in the last two decades, the social costs of such events are still huge. The monitoring of road accident occurrences is then essential for the identification of crash hot spots and the timely implementation of prevention policies. Recent technological developments are finally

P. A. Di Loro · P. Fantozzi, Dipartimento di Giurisprudenza, Economia, Politica e Lingue moderne, LUMSA, Rome, Italy.

M. Mingione (✉), Dipartimento di Scienze politiche, Università Roma Tre, Rome, Italy
(E-mail: marco.mingione@uniroma3.it).

© 2024 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics
<https://doi.org/10.1007/s13253-024-00615-z>

enabling straightforward and cheap ways of recording the exact space-time location of the vehicles at the moment of the accident, favouring the implementation of advanced statistical techniques to model the dynamics of the accident occurrences. For example, since 2014, road authorities of the city of Rome have been assigned the task of recording the time and the location of every car accident on which their intervention is required. These geo-referenced records are collected monthly and published for public use at (<https://dati.comune.roma.it/catalog/dataset?tags=Incidenti&groups=sicurezza-urbana>). This huge gold mine of public data has already drawn some attention (Comi et al. 2018; Alaimo Di Loro et al. 2019), as many other similar sources focused on road accidents (Kalair et al. 2021; Borgoni et al. 2021; Gilardi et al. 2022).

The most natural way to model an observed point pattern is represented by spatio-temporal point processes (Daley and Vere-Jones 2003; Diggle 2013). A very interesting challenge in modelling road accidents is understanding whether the occurrence of an event increases the risk of other events in its proximity and, if so, quantifying the number of subsequent crashes triggered by the first occurrence (Acker and Yuan 2019). This sort of cascading effect might be due to the direct effect of the original crash or its indirect consequences: increased traffic congestion, lane reduction, reduced visibility, *rubbernecking*, etc. This particular dynamic in point processes is known as *self-excitation* and it is the defining property of the Hawkes process (Hawkes 1971; Reinhart 2018). It has been largely used in the literature to characterize the clustered point patterns arising from earthquakes (Ogata 1988; Zhuang et al. 2004) or financial shocks (Bacry et al. 2015; Hawkes 2018). Its application has been recently extended to many more contexts, e.g. crimes (Mohler et al. 2011; Zhuang and Mateu 2019), infectious diseases (Chiang et al. 2022; Briz-Redón et al. 2023) and road accidents (Li et al. 2018; Kalair et al. 2021). Kalair et al. (2021) is the first contribution to propose such modelling in a spatio-temporal setting, where the road accidents occurring on a single (uni-dimensional) one-way road are disentangled into *primary* and *secondary* accidents. We expand on this work to model the event occurrences on a whole road network, road segment by road segment. For estimation purposes, we take the lead from the EM-based stochastic-reconstruction algorithm proposed by Zhuang and Mateu (2019) and extend it (i) to account for a buffering area, (ii) to propose an alternative smoothing of the spatial excitation function that enforces isotropy (as for the histogram estimator by Fox et al. (2016)), (iii) to account for the effect of covariates in the excitation component, (iv) to use the *Expected Complete-Data Log-likelihood* (ECDL) for the branching structure of the Hawkes process (Hawkes and Oakes 1974). (i) It is especially useful to mitigate the edge effects that bias the kernel smoothing and the Hawkes process estimation near the boundaries of the domain (Zhuang et al. 2004; Silverman 2018); (ii) increases the model parsimony limiting its ability to over-fit on the available data; (iii) can be useful to determine what characteristics of an event make it more prone to trigger others; (iv) the use of ECDL eases the numerical maximization of the likelihood (Veen and Schoenberg 2008). The estimation algorithm presents several computational bottlenecks for increasing data size and domain resolution. Therefore, we implemented several strategies to expedite the estimation process.

The remainder of the paper is organized as follows: The available data are described in Sect. 2; the model is introduced in Sect. 3 and its estimation in Sect. 4; the application is shown in Sect. 5; finally, Sect. 6 provides the conclusions and some discussion.

Table 1. Basic summary statistics of car accidents occurred in the Roma Capitale territory from 2019 to 2021: total number of accidents (N), total number of accidents with at least one injured or dead (N_{DI}), the rate of accidents with at least one injured or dead, the total number of deaths (D), the total number of injured (I), and the death and injury rates

Year	N	N_{DI}	N_{DI} rate	D	I	D rate	I rate
2019	29,056	10,775	0.371	121	13,624	0.004	0.469
2020	20,100	7137	0.355	85	8896	0.004	0.442
2021	27,149	9939	0.366	103	12,210	0.004	0.449
Tot	76,305	27,851	0.365	309	34,730	0.004	0.455

2. AVAILABLE DATA

We use data that can be freely downloaded from the website of Roma Capitale (<https://dati.comune.roma.it/catalog/dataset?tags=incidenti>). This source contains all the road accidents that occurred on the streets under the jurisdiction of the city of Rome in which a patrol of the local police has intervened. In the present work, we consider all records referred to the latest three years fully available, i.e. from 2019 to 2021. The data contain one separate record for each subject involved in each road accident, with records referring to the same accident identifiable through a common *protocol identification number* (PID). A common issue of open data is that, in principle, they are not collected for statistical but for monitoring purposes only. Therefore, they are often keen to measurement errors and/or under-reporting. In our case, a major problem is represented by mistakes in the data entry of the accident's location, which presented altered records in $\approx 20\%$ of all records. Therefore, the raw data have undergone a pre-processing procedure consisting of three main steps: (i) grouping and combination by PID, (ii) correction and filtering of altered records, (iii) elimination of all accidents whose location is outside any of the Roman municipalities¹ or of those for which correction was not possible. This process reduces the original car accident count from 79,456 to 76,305. Further details of this data cleaning procedure are provided in Section A of the Appendix.

2.1. DATA DESCRIPTION

Every record of the dataset includes ancillary information on the road accident—in addition to the exact time of the event and the location (lon , lat)—such as the number of injured–dead–unharmed people involved in the accident and the type of accident. Table 1 contains basic summary statistics at the yearly level for the road accidents that occurred in this area.

Between 2019 and 2021, a total of 76,305 road accidents led to 34,730 injured and 309 dead people. The grand totals correspond to averages of 25,435 road accidents, 11,577 injured, and 103 dead per year, respectively. Notice that these averages are downward biased by the exceptionality of 2020, during which COVID-19 lockdown and restriction policies

¹Shape-file available at <http://www.datiopen.it/it/catalogo-opendata/daticomuneromait>

were in place. These numbers define the huge economic and social cost that car accidents have on the Roman community. The dataset includes also information about the type of accident, which is, however, categorized very roughly. Looking at the accident description, we were able to find 4 major categories that well-describe the main dynamic of the event. These new categories are: (i) *type 1*: observed for $\approx 8\%$ of all events, describes a situation where a car runs over a person or an animal; (ii) *type 2*: observed for $\approx 27\%$ of all events, describes a situation where a car hit a fixed obstacle; (iii) *type 3*: observed for $\approx 16\%$ of all events, describes the case of a rear-end collision between vehicles; *type 4*: the most frequent ($\approx 49\%$ of cases) and describes other crashes between two or more vehicles. Figure 1a compares the monthly trend of the three years. The effect of the COVID-19 lockdown is clear-cut visible from the sudden reduction in accident rates since March 2020. The rate started increasing again in June 2020 but reached the pre-pandemic levels only in July 2021 (end of the smart-working regulation in Italy). Figure 1b shows the average number of events during the weekdays and highlights how these present higher rates than weekends, with the minimum touched during Sundays. Figure 1c shows the hourly rates. The behaviour reflects the typical commuting patterns in the city of Rome, presenting very low levels during night-time and larger values during day-time, with peaks around 7:00–8:00, 10:00–11:00, and 17:00–18:00. It is interesting to notice how the morning peak around 7:00–8:00 is flattened out in both 2020 and 2021, a likely effect of the smart-working policy in place for all or part of the year. Figure 2 shows a preliminary (naive) intensity estimation of the car-accident occurrence in the analysed area. It is clear how the pattern is not homogeneous in space but has significant variations, with larger intensities towards the centre of the urban area (with a greater road density) and around the main roads. Figure 3 reports the joint distribution of spatial and temporal lags between pairs of car accidents. The evident peak occurring for lags lower than 0.1 km (100m) in space and than 0.025 days (≈ 40 mins) in time highlights the clustered pattern of the car accident under analysis.

All these aspects motivate the need for a flexible spatio-temporal point-process model that can account not only for spatial and temporal heterogeneity but also for the daily and weekly cyclical behaviour, and the potential of a primary car accident to trigger secondary ones.

3. THE PERIODIC SPATIO-TEMPORAL HAWKES PROCESS

We seek to model the occurrence of traffic collisions over a spatio-temporal domain $\mathcal{Q} = \mathcal{D} \times \mathcal{T}$, where $\mathcal{D} \subseteq \mathbb{R}^2$ denotes the spatial dimension and $\mathcal{T} = [0, T]$ the temporal dimension. We assume that the number of car-crashes $N(B \times [t_1, t_2])$, where $B \subset \mathcal{D}$ and $[t_1, t_2] \subset \mathcal{T}$, is the result of a simple and (locally) finite spatio-temporal point process. It can be defined through the suitable specification of the conditional intensity function:

$$\lambda_c(\mathbf{s}, t) = \lim_{d\mathbf{s}, dt \rightarrow 0} \frac{\mathbb{E}[N(d\mathbf{s}, dt) | \mathcal{H}_t]}{d\mathbf{s} \cdot dt}, \quad (1)$$

where $\mathcal{H}_t = \{(\mathbf{u}, \tau)\}_{\tau < t}$ is the history of the process up to time t (t excluded). As suggested in Sect. 1, point patterns arising from road accidents present an inhomogeneous and clustered

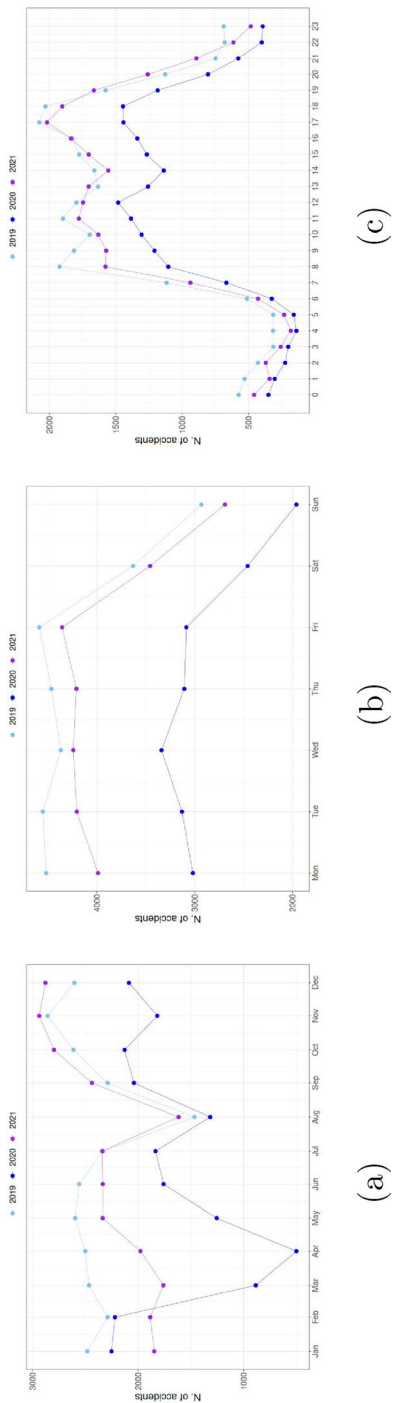


Figure 1. Temporal trends: monthly (a), weekly (b) and hourly (c).

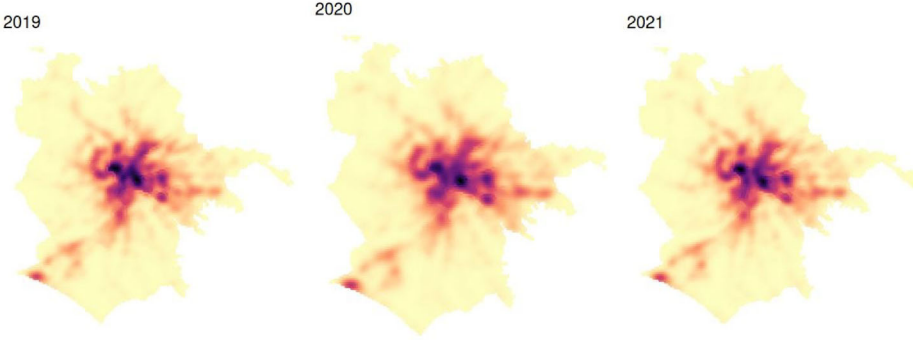


Figure 2. Spatial trend: simple kernel estimate of the intensity of the point process, separately for each year. Dark colours correspond to larger intensities (and vice versa) (color figure online).

pattern: (i) events are not spread uniformly along the observed region \mathcal{Q} , mainly because the risk of collision is affected by space-time-varying factors; (ii) events may exhibit self-excitation behaviour due to sudden and unexpected slowdowns or other consequential traffic events. These two components can be jointly captured by expressing Eq.(1) as one of a spatio-temporal Hawkes process $\lambda_c(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \int_0^t \int_{\mathcal{D}} g(\mathbf{s} - \boldsymbol{\sigma}, t - \tau) dN(d\boldsymbol{\sigma} \times d\tau)$. $\mu(\cdot, \cdot)$ is the background intensity, describing the space-time-varying general risk of a car accident, and $g(\cdot, \cdot)$ is a space-time excitation function, describing the space-time propagation of the excitation of each event. In particular, we express these two components as in the semi-parametric, periodic, and space-time separable version by [Zhuang and Mateu \(2019\)](#):

$$\lambda_c(\mathbf{s}, t) = \mu_0 \cdot \mu_s(\mathbf{s}) \cdot \mu_t(t) + A \cdot \int_0^t \int_{\mathcal{D}} g_s(\mathbf{s} - \mathbf{u}) \cdot g_t(t - \tau) N(d\mathbf{u} \times d\tau), \quad (2)$$

where $\mu_s(\cdot)$, $\mu_t(\cdot)$ are the spatial and temporal background intensities such that their average value over \mathcal{D} and \mathcal{T} is 1, $g_s(\cdot)$, $g_t(\cdot)$ are the spatial and temporal excitation functions such that their integral over \mathcal{D} and \mathcal{T} is 1, and μ_0 , $A > 0$ are two real-valued parameters that regulate the overall level of the background and the excitation. The purely spatial term $\mu_s(\cdot)$ accounts for proportional spatial variations in the risk of collision (due to e.g. generally higher traffic levels, low visibility, rough roads, dangerous intersections, etc.) and shall vary freely on the whole spatial domain \mathcal{D} . The purely temporal term $\mu_t(t)$ must capture the proportional variations in the long-term trend, and the daily and weekly periodicity of collisions. This need can be formalized by expressing $\mu_t(t)$ as the product of three components: $\mu_t(t) = \mu_d(d(t)) \cdot \mu_w(w(t)) \cdot \mu_{tr}(t)$, where $\mu_d(\cdot)$, $\mu_w(\cdot)$ model the daily and weekly periodicity, $\mu_{tr}(\cdot)$ represents the long-term trend, and $d(\cdot)$, $w(\cdot)$ match each time $t \in \mathcal{T}$ with the corresponding day-hour and week-day. Finally, the spatial and temporal excitation functions $g_s(\cdot)$, $g_t(\cdot)$ are two decreasing functions that describe the increase in risk occurring in the spatial and temporal proximity of a previous accident at different lags, respectively.

Isotropic excitation. The original parametrization by [Zhuang and Mateu \(2019\)](#) considers a potentially anisotropic excitation in space, with the excitation magnitude of event $\mathbf{s} = (x, y)$

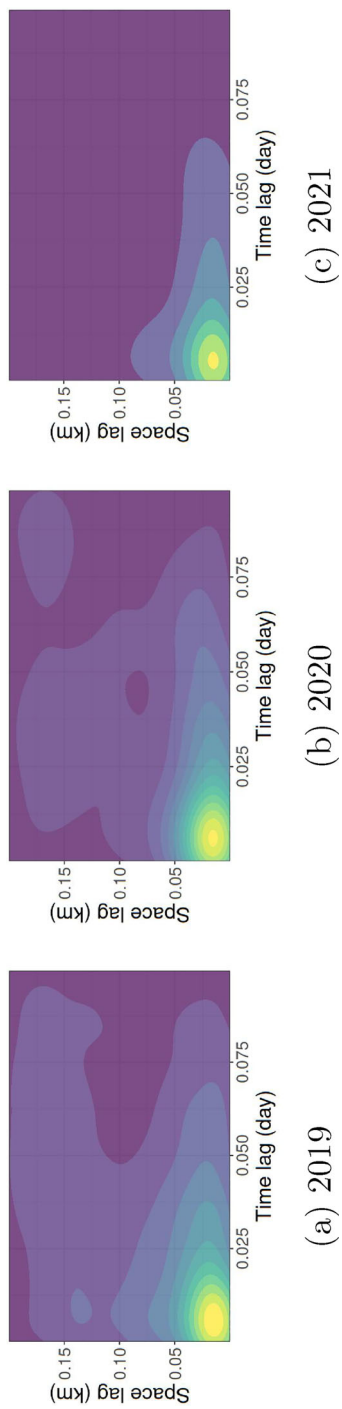


Figure 3. Spatio-temporal triggering density (events per unit-time per unit-area) of the lags between subsequent events (cut to 0.2 km and 0.1 days). The brighter the colour, the higher the density (Color figure online).

on location $s' = (x', y')$ depending on the whole separation vector $s' - s = (x' - x, y' - y)$. However, the typical expression of the spatio-temporal Hawkes process when adopted in a parametric framework considers an isotropic spatial excitation that is a function of the Euclidean distance between the primary event and nearby locations:

$$g_s(s' - s) = g_s(\|s' - s\|) = g_s\left(\sqrt{(x' - x)^2 + (y' - y)^2}\right), \quad (3)$$

so that $g_s(\cdot) : \mathbb{R}^+ \rightarrow [0, +\infty)$. This imposes additional structure on the excitation decay, which cannot behave freely across different directions, and induces a more parsimonious model specification. On a more general note, the freedom to decay differently across different directions is redundant in most applications and the unconstrained smoothing often results in an (approximately) isotropic excitation in space. Exploiting the isotropy as a preliminary assumption can be beneficial from multiple standpoints: (i) computational convenience, both in terms of computing runtime and storage memory; (ii) more robust estimation of the excitation at a certain distance, as the smoothing is performed on a lower-dimensional space; (iii) reduced liability to overfitting, as the resulting model is less flexible and less free in adapting to the current data configuration. More details are provided in Section C of the Appendix, where all statements are supported by a simulation study. Nevertheless, achieving a nonparametric kernel estimate of $g_s(\cdot)$ under the isotropic setting has some implications on the smoothing process, which is discussed in Sect. 4.

Generalized linear model on the excitation. The spatial and spatio-temporal Hawkes process is widely used in the seismological literature in the guise of the *Epidemic Type Aftershock Sequence* (ETAS) model (Ogata 1988, 1999; Zhuang et al. 2002, 2004; Veen and Schoenberg 2008). The parametric functional form given to the excitation function is based on scientific knowledge of the seismological process, also supported by decades of studies on earthquake aftershock sequences (Omori's law). In particular, the excitation effect is usually assumed to be (positively) associated with the magnitude of the primary event. In other contexts, what characteristics of the event and in what directions they should act are not as straightforward to determine. Nevertheless, we can express the effect of the available characteristics on the excitation coefficient A through a linear prediction and a suitable *link* function. Let \mathbf{x}_i be the $(k \times 1)$ vector of covariates available on each event, we can express:

$$\log(A_i) = \mathbf{x}_i^\top \cdot \boldsymbol{\beta}, \quad i = 1, \dots, n, \quad (4)$$

and the right-hand expression of Equation (2) becomes:

$$\int_0^t \int_{\mathcal{D}} A(\mathbf{u}) \cdot g_s(\mathbf{s} - \mathbf{u}) \cdot g_t(t - \tau) N(d\mathbf{u} \times d\tau) = \sum_{t_i < t} A_i \cdot g_s(\mathbf{s} - \mathbf{s}_i) \cdot g_t(t - t_i).$$

In the same way, given a set of ℓ covariates available at each space-time locations $\mathbf{z}(\cdot, \cdot)$, the background coefficient can be expressed as $\log(\mu_0(\mathbf{s}, t)) = \mathbf{z}(\mathbf{s}, t)^\top \cdot \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a $\ell + 1$ vector of coefficients. While straightforward to formulate, the inclusion of this term for space-time varying covariate can easily become unfeasible for large spatio-temporal domains and, in the sequel, we refer to the constant specification $\mu_0(\mathbf{s}, t) = \mu_0$.

4. SEMI-PARAMETRIC ESTIMATION

Let $\mathcal{O} = \{(s_i, t_i)\}_{i=1}^n \subset \mathcal{Q}$ be the realization spatio-temporal self-exciting point process. We want to achieve a nonparametric estimation of all components building up the conditional intensity function $\lambda_c(\cdot)$. Whilst the nonparametric estimation is a standard solution for the background rate in the point-process literature, the nonparametric estimation of the excitation function is a quite novel aspect originally introduced in [Marsan and Lengline \(2008\)](#) and further extended in [Mohler et al. \(2011\)](#); [Zhuang and Mateu \(2019\)](#); [Kalair et al. \(2021\)](#). As a matter of fact, in many real-world applications, there is no physical theory (or previous literature) to support any specific functional form of $g_s(\cdot)$ and $g_t(\cdot)$. Hence, the nonparametric estimation provides a fully data-driven approach to their estimation. Model-based inference for the conditional intensity function $\lambda_c(\cdot)$ and its components can be obtained by maximizing the log-likelihood function for a general point process (GPP):

$$\ell_{\lambda_c}(\mathcal{O}) = \sum_{i=1}^n \log(\lambda_c(s_i, t_i)) - \int_0^T \int_{\mathcal{D}} \lambda_c(\sigma, \tau) d\sigma d\tau \quad (5)$$

However, the conditional intensity function of a self-exciting process presents various discontinuities arising at each observation because of the resulting sudden excitation effect. This hinders the possibility of directly getting a nonparametric estimate of the spatio-temporal variations of $\lambda_c(\cdot)$ through *kernel smoothing* or other methods. Furthermore, the log-likelihood happens to be ill-behaved and nearly flat in large regions of the parameter space, troubling numerical maximization algorithms and making convergence extremely slow (or fail altogether).

The *cluster-process* representation of a Hawkes process by [Hawkes and Oakes \(1974\)](#) provides a possible solution to this issue. Events are split in *background events* and *triggered events* (effect of the excitation of previous events, see Fig. 4a). If labels denoting which events are background and which events are triggered are available, then the complete log-likelihood becomes:

$$\begin{aligned} \ell_{\lambda_c}^C(\mathcal{O}) = & \sum_{i=1}^n \mathbb{I}(u_i = 0) \cdot \log \mu(s_i, t_i) + \sum_{i=1}^n \sum_{j:t_j < t_i} \mathbb{I}(u_i = j) \cdot \log g(s_i - s_j, t_i - t_j) + \\ & - \int_0^T \int_{\mathcal{D}} \lambda_c(\sigma, \tau) d\sigma d\tau, \end{aligned} \quad (6)$$

where $u_i = 0$ if event i is a background event, $u_i = j$ if event i is triggered by event j , and $\mathbb{I}(\cdot)$ is the indicator function ([Veen and Schoenberg 2008](#)). The maximization of Eq. (7) is much easier as background events only contribute to the background part of the likelihood and vice versa. This cluster representation becomes extremely useful if we want to perform the non-parametric estimation of the background and excitation functions. In particular, all background events $\{i : u_i = 0\}$ can be used to estimate the background components; while all triggered pairs $\{i, j : u_i = j\}$ can be used to estimate the excitation components. Given that these labels are not available, we need a way to quantify how likely each event is a

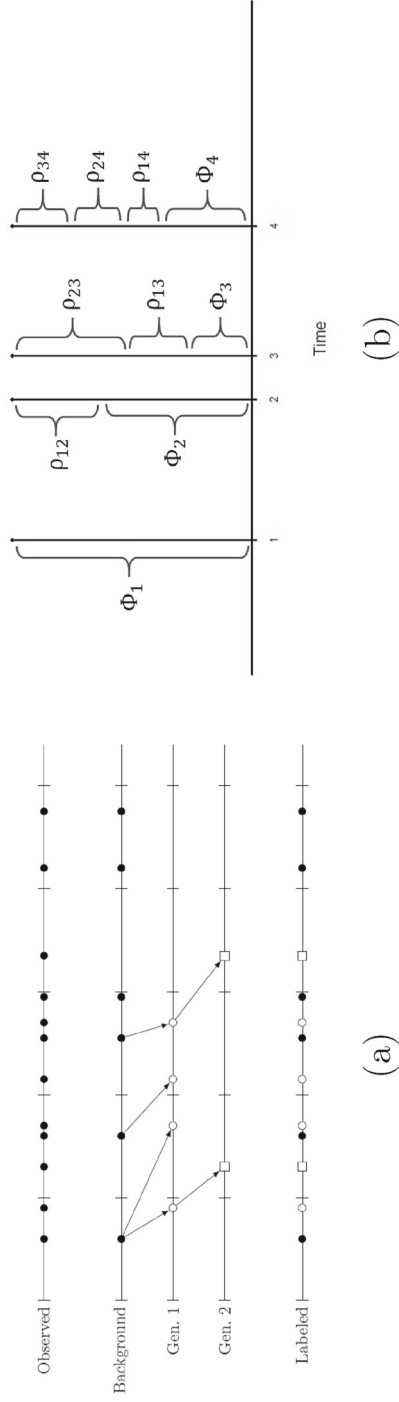


Figure 4. **a** Clustering representation of a Hawkes process: solid circles are background events, and open circles and squares are excited events (Reinhart 2018); **b** example of background/excitation contributions for events from a self-exciting process .

triggered or a background event. The corresponding probabilities are:

$$\Pr(u_i = j) = \rho_{ij} = \begin{cases} \frac{A_j \cdot g_s(\|s_i - s_j\|) \cdot g_t(t_i - t_j)}{\lambda_c(s', t')} & t_i > t_j \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$\Pr(u_i = 0) = \phi_i = 1 - \sum_j \rho_{ij} = \mu_0 \cdot \frac{\mu_s(s_i) \cdot \mu_t(t_i)}{\lambda_c(s_i, t_i)},$$

which can be used to evaluate the *Expected Complete-Data Log-Likelihood* (ECDL):

$$\begin{aligned} \ell_{\lambda_c}^E(\mathcal{O}) = & \sum_{i=1}^n \phi_i \cdot \log \mu(s_i, t_i) + \sum_{i=1}^n \sum_{j: t_j < t_i} \rho_{ij} \cdot \log g(s_i - s_j, t_i - t_j) + \\ & - \int_0^T \int_{\mathcal{D}} \lambda_c(\boldsymbol{\sigma}, \tau) d\boldsymbol{\sigma} d\tau. \end{aligned} \quad (8)$$

It is now self-evident how $\lambda_c(\cdot, \cdot)$ and all its components are necessary to evaluate these probabilities, and these probabilities are necessary to allocate each event in the background or triggered set. This circularity can be solved using an EM strategy where the update of Equation (7) is alternated with the maximization of the log-likelihood (Marsan and Lengline 2008; Marsan and Lengliné 2010). Mohler et al. (2011) considers the *Stochastic Declustering* (Zhuang et al. 2002), where events are randomly allocated to one of the two sets at each iteration according to the current values of (7). We here consider the stochastic reconstruction (SR) algorithm originally introduced (Zhuang et al. 2004) and later used in Zhuang and Mateu (2019) and Kalair et al. (2021).

4.1. THE STOCHASTIC RECONSTRUCTION

In the stochastic reconstruction, all components are estimated nonparametrically and all events contribute to the estimation of all components proportionally to their probability of being a background or triggered event (see Fig. 4b). The nonparametric estimation of the functions can then be achieved through weighted kernel smoothing. A major issue with kernel smoothing on limited domains is the bias occurring near the domain boundaries. This is especially true when the target intensity does not naturally decrease to 0 near the edges. Several strategies (mostly based on prior knowledge of the target intensity shape) have been proposed in the literature to mitigate this problem (Chiu 2000; Silverman 2018). Here, we suggest considering a buffering area, say $\mathcal{B}_b \times [-b, T + b] = \mathcal{Q}_b \supset \mathcal{Q}$, for smoothing purposes. All data points in $\{(s_i, t_i)\}_{i=1}^{n_b} \subset \mathcal{Q}_b$, with $n_b > n$, are used to smooth the functional forms of the background and excitation components. Instead, the conditional intensity function and the likelihood are evaluated only on points in the main area $\{(s_i, t_i)\}_{i=1}^n \subset \mathcal{Q}$. Not only does this overcome the kernel boundary bias in the spatial and long-term component of the background, but it also allows accounting for the excitation of points in $\mathcal{Q}_b/\mathcal{Q}$ on

points in \mathcal{Q} . As noted in (Zhuang et al. 2004), neglecting this effect would yield bias near the edge of the notwithstanding the way the functions are estimated.

Smoothing the Background Components The spatial background function $\mu_s(\cdot)$ is smoothed by averaging every single event (s_i, t_i) , $i = 1, \dots, n_b$ with weights proportional to the corresponding background weights ϕ_i :

$$\hat{\mu}_s(s) \propto \sum_{i=1}^{n_b} \phi_i \cdot \frac{k_{\omega_s}^2(s - s_i)}{\int_{\mathcal{D}} k_{\omega_s}^2(s - \sigma) d\sigma},$$

where $k_{\omega_s}^2(\cdot)$ is a two-dimensional kernel function with bandwidth ω_s and the denominator is needed for the domain normalization. The temporal background components can be reconstructed very similarly but rely on a slightly different set of weights:

$$w_i^a = \frac{\mu_a(a(t_i)) \cdot \mu_s(s_i)}{\lambda_c(s_i, t_i)}, \quad \hat{\mu}_a(t) \propto \sum_{i=1}^{n_b} w_i^a \frac{k_{\omega_a}(a(t) - a(t_i))}{\int_{\mathcal{T}_a} k_{\omega_a}(a(t) - \tau) d\tau},$$

where $a(\cdot)$ maps the raw time t_i to the corresponding temporal scale (e.g. raw, daily, or weekly) and $k_{\omega_a}(\cdot)$ is a one-dimensional kernel function with bandwidth ω_a (Zhuang 2006). The long-term trend does not require any specific boundary correction as the problem is already addressed by the buffering area. The daily and weekly periodic components, instead, can benefit from the adoption of a periodic correction (Silverman 2018; Kalair et al. 2021). All smoothed background components are then rescaled so that their average on \mathcal{Q} is equal to 1.

Smoothing a spatially isotropic excitation function The reconstruction of the excitation components can be obtained by weighted kernel smoothing of all the pairwise space lags $\delta_{ij} = \|s_i - s_j\|$ and time lags $\tau_{ij} = t_i - t_j$ such that $t_i > t_j$. The number of possible pairs i, j scales quadratically with the size of the data n_b but most pairs lie far apart. In practice, the excitation must have a finite range, and only close pairs (in space or time) actually interact. One could decide on conservative cut-off points t^*, d^* and set the excitation functions to 0 for larger values while dropping all pairs such that $\tau_{ij} > \tau^*$ or $\delta_{ij} > \delta^*$. Thus, the two functions can be reconstructed on $(0, \tau^*)$ and $(0, \delta^*)$ by averaging each pair with weights proportional to the relative importance of the inter-event excitation at location (s_i, t_i) . The expressions are:

$$\hat{g}_t(t) \propto \frac{\sum_{i,j=1}^{n_b} \rho_{ij} \frac{k_{h_t}(t - \tau_{ij})}{\int_0^{\tau^*} k_{h_t}(\tau - \tau_{ij}) d\tau}}{\sum_{j=1}^{n_b} \mathbb{I}(t_j + t < T + b)}, \quad \hat{g}_s(d) \propto \frac{\sum_{i,j=1}^{n_b} \rho_{ij} \frac{k_{h_s}(d - \delta_{ij})}{\int_0^{\delta^*} k_{h_s}(\delta - \delta_{ij}) d\delta}}{\sum_{j=1}^{n_b} |\mathcal{C}(s_j, d) \cap \mathcal{D}_b|}, \quad (9)$$

where $k_{h_s}(\cdot), k_{h_t}(\cdot)$ are one-dimensional kernel function, the outer denominators are the *repetition correction*, and $\mathcal{C}(s_j, d)$ is the circle with centre s_j and radius d . In the case of the background smoothing, all n_b points contribute to the estimated value at all locations. However, in the case of lags between pairs of points, not all available n_b points can be

leveraged to estimate the occurrence of a target lag $t \in (0, t^*)$ or $d \in (0, d^*)$. The *repetition correction* term is needed to adjust for this disequilibrium and plays a key role in the proper smoothing of the isotropic spatial excitation. When averaging the time-lag t only the points t_j such that $t_j + t < T + b$ can actually contribute to the smoothing with a *potential* lag (that might or might not be observed). Points closer to the upper limit have no chance of contributing as that would occur in $t_j + t > T + b$, outside the observed window. For the spatial excitation, it is not only a matter of how many observed points have the potential of having another point at lag d . It is also a matter of the differing potential that each event j has to have another point at lag d within the domain \mathcal{D}_b . Indeed, we must consider that the larger the distance the wider the circle on which two points at the same distance may lie and the smoothing must account for the spreading of the mass across circles of different sizes. Therefore, the potential of each point must be proportional to $|\mathcal{C}(s_j, d)|$ properly intersected with the buffer spatial domain \mathcal{D}_b . Notice that if the spatial domain were the unlimited \mathbb{R}^2 , then $\sum_{j=1}^{n_b} |\mathcal{C}(\delta_j, d) \cap \mathbb{R}^2| = \sum_{j=1}^{n_b} |\mathcal{C}(s_j, d)| \propto 2 \cdot \pi \cdot d$, which corresponds to the scaling factor used for the *pair-correlation function* (Diggle 2013) and adopted by Fox et al. (2016) for their histogram estimator.

Let us point out that both excitation functions have a natural border in 0 and are expected to have a monotone-decreasing behaviour. For practical purposes, in the absence of a robust theoretical model, it would be desirable for the excitation functions to meet this left border with a flat form ($g'_t(0^+) = g'_s(0^+) = 0$). We can achieve that by applying the *reflection correction* (Silverman 2018) to the smoothing of the temporal excitation function while such adjustment is not necessary in the case of the isotropic spatial excitation function. Indeed, the density of points at decreasing spatial lags naturally decreases to 0 (as an effect of the reducing circumference), and this decrease shall balance out with the above-mentioned repetition correction. After smoothing, the temporal and spatial excitation functions are rescaled so that $\int_0^{\tau^*} \hat{g}_t(\tau) d\tau = \int_0^{\delta^*} 2 \cdot \pi \cdot \delta \cdot \hat{g}_s(\delta) d\delta = 1$.

4.2. ESTIMATION ALGORITHM AND COMPUTATIONAL DETAILS

Given the set of smoothed functions, the optimal values of μ_0 and A (or of the corresponding covariates' effects β), can be obtained by optimizing a target log-likelihood. All this procedure can be unified in an EM-type algorithm that alternates between the weights update, the function smoothing, and the log-likelihood optimization. We here modify the original versions in three major directions: (i) we consider all n_b events occurring in a buffering area $\mathcal{Q}_b \supset \mathcal{Q}$ for smoothing purposes but evaluate the log-likelihood only on the n points within the main area \mathcal{D} ; (ii) we optimize the ECDL (8) in place of the GPP log-likelihood (5); (iii) we introduce a weight update between the smoothing and the ECDL optimization. A sketch of the pseudo-code is provided in Algorithm 1 of the Appendix. For the sake of brevity, we write it for constant coefficients μ_0 , A and the set of background and excitation functions is denoted as $\{\mu(\cdot), g(\cdot)\}$.

It is well known that the EM algorithm may converge to local maxima of the log-likelihood function or singularities at the edge of the parameter space, where the log-likelihood is unbounded. To avoid such a problem, we pursue a multi-start strategy and run the EM algorithm from multiple random initializations keeping only the *best* solution. We stop the

optimization when the increase of two successive log-likelihoods falls below 10^{-4} . Another drawback of the stochastic-reconstruction algorithm lies in its computational complexity. This is a very common problem in the point-process literature and is exacerbated by the second-order characteristics of the Hawkes process. At each iteration, all functions must be evaluated on a grid of given space-time locations \mathcal{G} (or space-time lags \mathcal{H}) and integrated over (possibly irregular) domains. The finer the grid, the better the approximation, but these operations scale with $n \times |\mathcal{G}|$ in the background components and with $n^2 \times |\mathcal{H}|$ in the excitation components. The proposed *excitation tapering* (with cut-offs τ^* , d^*) can reduce the burden of the latter, but it is rarely sufficient to get a timely estimation for n large and fine grids (i.e. $|\mathcal{G}|$, $|\mathcal{H}|$ large). However, it is worth noting that many quantities (the kernels, the repetition correction, and the sizes of the integration domain) remain unchanged through all iterations and can be computed only once at the beginning. This can noticeably expedite the estimation process but requires the storage of large matrices, which might be unbearable in terms of needed storage.

In our implementation, we exploit the sparse allocation of such matrices to reduce the memory burden while retaining most of the computational advantage of prior computation. At the same time, we exploit parallel computing any time possible and code the more evident computational bottlenecks in C++. All these strategies greatly enhance the computational performances, which is a key aspect of our application in Sect. 5 developed on a quite large dataset and a fine grid. Codes with a snippet example are available at <https://github.com/PAIaimo/STHawkesSRISOroadNet>.

5. APPLICATION

The model presented in Sect. 3 strongly relies on the quality and coherency of the input data and on the stationarity of presumed dynamics (periodicity and excitation). The territory of the city of Rome is vast and remarkably heterogeneous in terms of landscape, road typology and condition, population density, etc. Such characteristics are a major source of heterogeneity that might hinder a proper estimation of the model. Therefore, we consider the three years separately and focus on the road accidents occurred within the *Great Ring Road* (GRA) surrounding Rome. This major road outlines the urban area of the city, which is still huge: 346 km^2 for a total of $\approx 2985 \text{ km}$ of streets.

Note that all computations for the results of this section have been executed on the HPC TeraStat2, a high-performance computing infrastructure developed by the Department of Statistics of Sapienza University of Rome. The computing environment is equipped with 24 modern computational nodes, having up to 256Gb of RAM, and counting a total of 1920 cores (<https://www.dss.uniroma1.it/en/HPCTeraStat2/Specifiche>).

5.1. MODEL FITTING AND COMPARISON

For the sake of brevity, we here report details on the model fitting and model comparison for the 2019 but the procedure follows the same rules and yields analogous conclusions on 2020 and 2021. Figure 5a provides a snapshot of the gridded approximation of the road

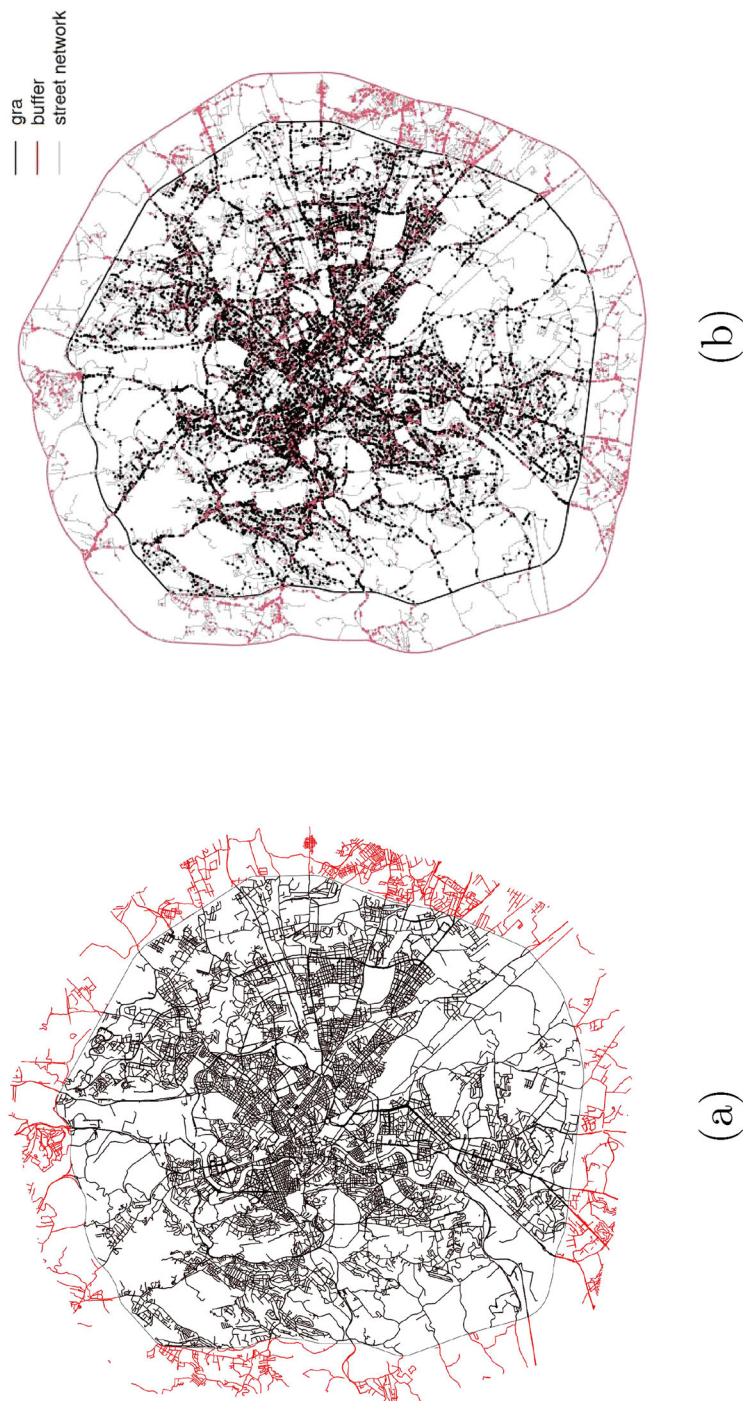


Figure 5. **a** Raster approximation ($10\text{ m} \times 10\text{ m}$) of the street network; **b** study area and observed point pattern (2019).

network and its spatial buffer. The latter is highlighted in red and is a 2-km buffering area in space surrounding the GRA. The $10 \text{ m} \times 10 \text{ m}$ grid approximation is able to well capture all road segments. Figure 5b reports all events of 2019 colour-coded for buffer or no buffer. The red dots inside the GRA are those that occurred in the two additional months in time, one preceding the beginning of the year (December 2018) and the other one following the end of the same year (January 2020). The n points reported in black directly contribute to the likelihood, while the red dots fall in the spatio-temporal buffer (the one inside the GRA are from the previous or following month) and are used for the smoothing process only. The smoothing is performed using Gaussian kernels with different bandwidths for each component. The bandwidth choice determines the *smoothness* of the resulting function, i.e. the resolution at which the function varies. The long-trend, weekly periodic, and daily periodic components have been smoothed with bandwidths of 7 (1 week), 1 (day), and 0.05 (1.1 h). For the spatial background, an isotropic Gaussian kernel with a varying-bandwidth approach has been chosen, with a minimum level set to 0.1 (i.e. 100 m), and individual bandwidths $b_s(i)$ such that at least 10 other events were within $b_s(i)$ distance from event i .

Finally, the temporal and spatial excitation bandwidths have been set to 0.03 (≈ 45 min) and 0.05 (50 m) as they must capture smaller scale variations than the background components.

We consider the possibility that different types of accidents have different excitation abilities. Indeed, different accident dynamics might have varying degrees of severity (on average) or correspond to different consequences (e.g. modify the road conditions, the passing drivers' behaviour, etc.). We consider the $k = 1, \dots, K$ categories introduced in Sect. 2 and use Eq. (4) to express the excitation coefficient as a function of the type of accident. This results in $K = 4$ different excitation effects for the different accident types. We fit the proposed full model (Full) and other several competitors with various degrees of complexity on this same data. We consider a full model including only one common A (Full—One A), a model without the periodic component of the background (NP), a model without the excitation effect (NE), and a model with neither the periodic nor the excitation components (NENP). We compare the model adaptations in terms of the *GPP* log-likelihood of Eq. (5). Table 2 shows that when the periodicity is taken into account, the model including the excitation effect provides a sensible improvement in terms of in-sample adaptation. The same data have been used to fit the anisotropic spatial excitation version of the full model. The results do not differ significantly and are reported in Section D of the Appendix. This proves the explanatory role of the excitation component in describing the realized point pattern. Hence, all the following analysis and discussion is focused on the full model.

5.2. MODEL DIAGNOSTIC: RESIDUAL ANALYSIS

A large value of the log-likelihood does not guarantee that the model fits the data adequately. We perform additional diagnostics to verify if the estimated conditional intensity rate $\hat{\lambda}_c(\cdot)$ produces a residual analysis that complies with the model assumptions. For instance, it is established that for a given realization $\mathcal{O} = \{(s_i, t_i)\}_{i=1}^n \subset \mathcal{Q}$ of a spatio-temporal point process N with conditional intensity rate $\lambda_c(\cdot, \cdot)$, the transformation of the observed time-sequence $\tau_i = \Lambda_c(t_i) = \int_0^{t_i} \int_{\mathcal{D}} \lambda_c(\sigma, \tau) d\sigma d\tau$ yields a stationary Poisson process with

Table 2. Log-likelihood values at the parameter estimates of the different models for 2019

Model	Full	Full—One A	NE	NP	NENP
$-\log(\mathcal{L})$	- 8911.04	- 8844.17	- 6936.835	- 3984.941	- 3717.216

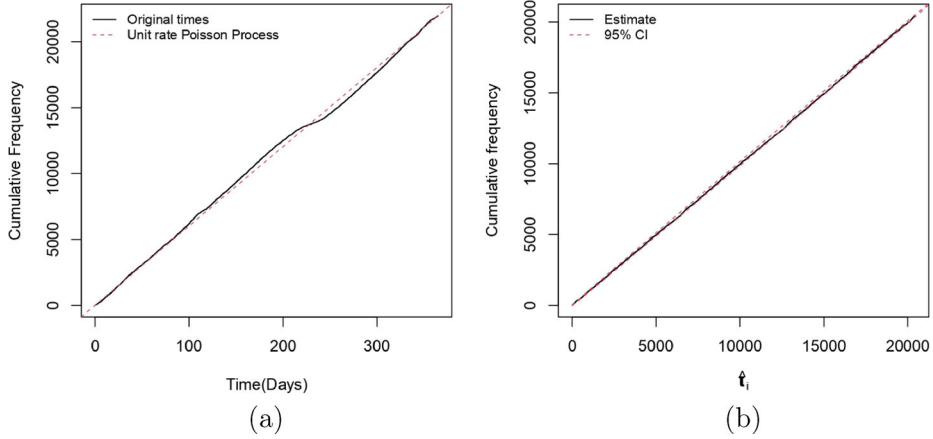


Figure 6. Residual diagnostic on the original and transformed times for 2019 .

unit rate in time (Daley and Vere-Jones 2007). From a qualitative point of view, this transformation is expanding or compressing arrival times according to the cumulative intensity $\Lambda_c(\cdot)$. We consider a diagnostic based on this property (Ogata 1988), where the estimated conditional intensity rate is plugged in to get:

$$\hat{\tau}_i = \hat{\Lambda}_c(t_i) = \int_0^{t_i} \int_{\mathcal{D}} \hat{\lambda}_c(\sigma, \tau) d\sigma d\tau. \quad (10)$$

If the model is well-specified, the scaled transformed time sequence $\hat{\tau}_i$ must be distributed as $i \cdot Z$ where $Z \sim Beta(i + 1, n - i + 1)$ for each positive integer $i = 1, \dots, n$. This provides a tool to build up a confidence interval and check for deviation from the model assumptions. Figure 6a shows the counting measure of the original times compared to the bisector (unit-rate Poisson process) and 6b shows the estimated transformed time sequence with the corresponding confidence bounds. Since it is poorly visible from Fig. 6b, we here report that 95.3% of the $\hat{\tau}_i$ are included within the nominal 0.95 confidence bounds.

5.3. ESTIMATED PARAMETERS ACROSS THE YEARS

The MLE of the coefficients for all years is reported in Table 3. Given the large sample size under analysis, the uncertainty has been evaluated relying on the asymptotic normality

of the MLE by evaluating $\Sigma_{\hat{\mu}, \hat{\beta}} = -\mathcal{H}(\ell_{\lambda_c}(\hat{\mu}, \hat{\beta}))^{-1}$ (Rathbun 1996; Wang et al. 2010).

Background components The estimate $\hat{\mu}_0$ represents the average intensity of road accidents across space and time and can be interpreted as the average number of primary road accidents occurring in each unit of space (10 m \times 10 m) and unit of time (1 day) in the Rome urban road network. The yearly estimates confirm and corroborate the preliminary discussion in Sect. 2 as the coefficient for 2020 is lower than for 2021 and 2019. The functional form of the estimated background components is reported in Figs. 7a–c and 8a. The estimated long-term trend component of 2019 (Fig. 7a) highlights the typical seasonal variations, with the major peaks in January, March and October, and a tight and deep valley in August. The latter is easily explained by the greatly reduced presence of people (and traffic) during the Italian summer holidays. The long-term trend of 2020 clearly captures the abrupt reduction during the strict lockdown period that lasted from March to May and the second lockdown of October 2020. Finally, the long-term trend of 2021 shows a clear increasing trend from April 2021 onwards. Figure 7c shows the estimated weekly cyclical pattern that confirms the original impressions from Sect. 2. The high values on Thursday and Friday, at the edge of the weekend, may be explained by the combination of the standard commuting patterns and the night traffic associated with leisure activities. It is worth noticing that the peak is slightly shifted backwards in 2020 (Thursday instead of Friday), reflecting once again the restrictions (e.g. lockdown, green pass, etc.) imposed by the government for a large part of the year to reduce the spread of COVID-19. Figure 7b shows the estimated daily periodic components in the three years that do not vary sensibly and do not add much to what has already been said in Sect. 2. Figure 8a shows the estimated spatial background for 2019 only, for the sake of brevity. Indeed, this is once again almost identical across the three years, suggesting that the spatial dynamic of road accident occurrence within the GRA did not change significantly. It highlights some major hot spots and we report in the following just three examples, zooming in on these areas of major interest for the Roman community. The first area (see Fig. 8b) belongs to the VIII municipality, where the bridge (Ponte Marconi) separates the northwest side of the city from the southern part; the major road in the centre of the figure (Viale G. Marconi) collects all people that commute from the north and either go to Roma Tre University (one of the largest universities in Rome) or go to work in the southern area (EUR neighbourhood) that hosts central offices of large companies. Figure 8c instead shows the risk close to the second largest train and bus station of Rome (stazione Roma Tiburtina); the risk is larger on the major road in the centre of the figure (Via Tiburtina) that collects the commuting pattern of people living in the northeastern part of the city instead. Finally, Fig. 8d pinpoints a larger area where the risk of collision is high and includes a couple of major roads placed in the northern part of the city (behind Vatican City) that are usually very busy during the whole day.

Excitation components The estimates of the β coefficients of the excitation parameter A in Table 3 must be interpreted with respect to the baseline category 1 and reported on the exponential scale. They correspond to $\hat{A}_1 = 0.058, 0.097, 0.028$, $\hat{A}_2 = 0.197, 0.168, 0.218$, $\hat{A}_3 = 0.074, 0.038, 0.076$, and $\hat{A}_4 = 0.036, 0.035, 0.023$, for 2019, 2020 and 2021, respectively. The accidents of type 2—“vehicle hit fixed obstacle”—are those with the larger

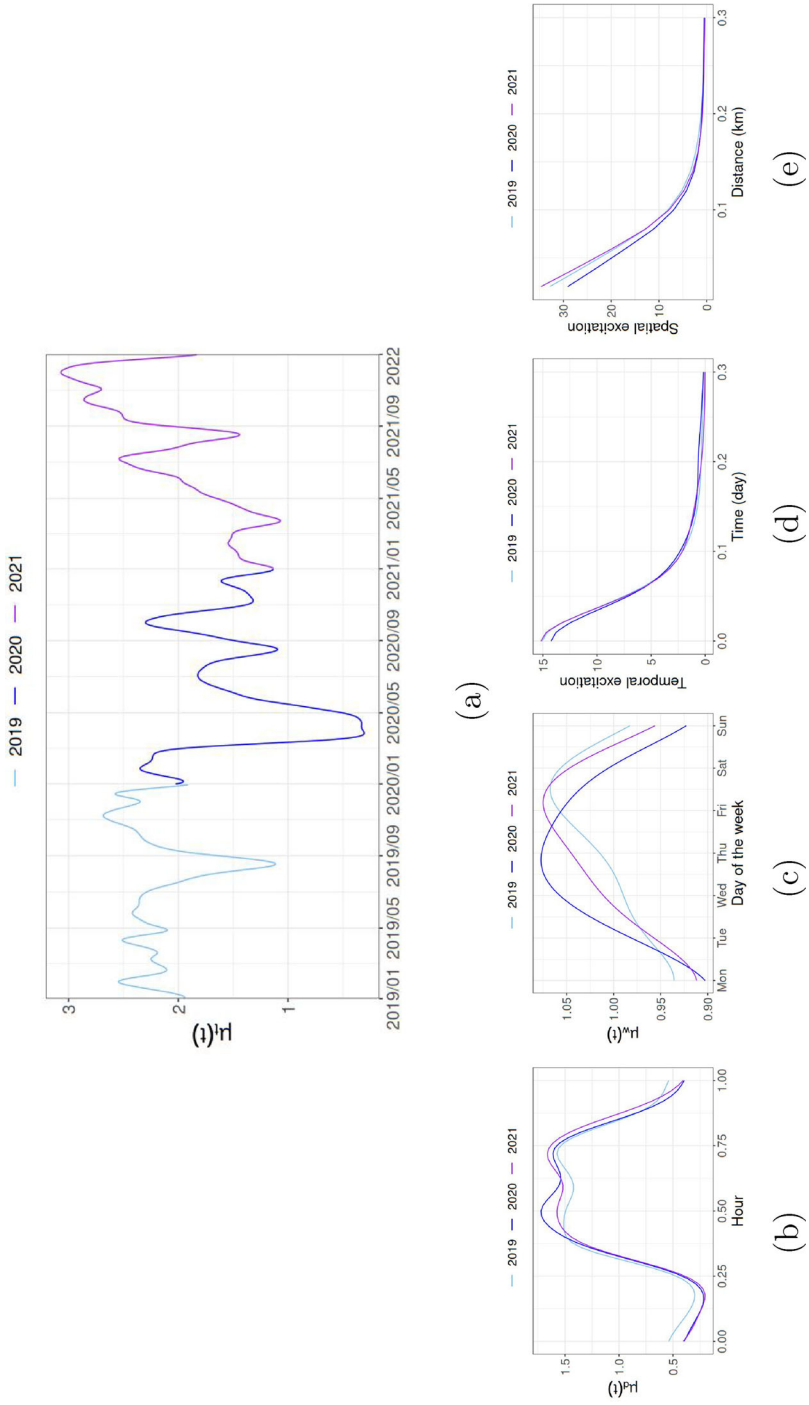


Figure 7. **a** Cyclical components of the background: hour (left), day of the week (centre) and long term (right); **b** Excitation: temporal (left) and spatial (right).

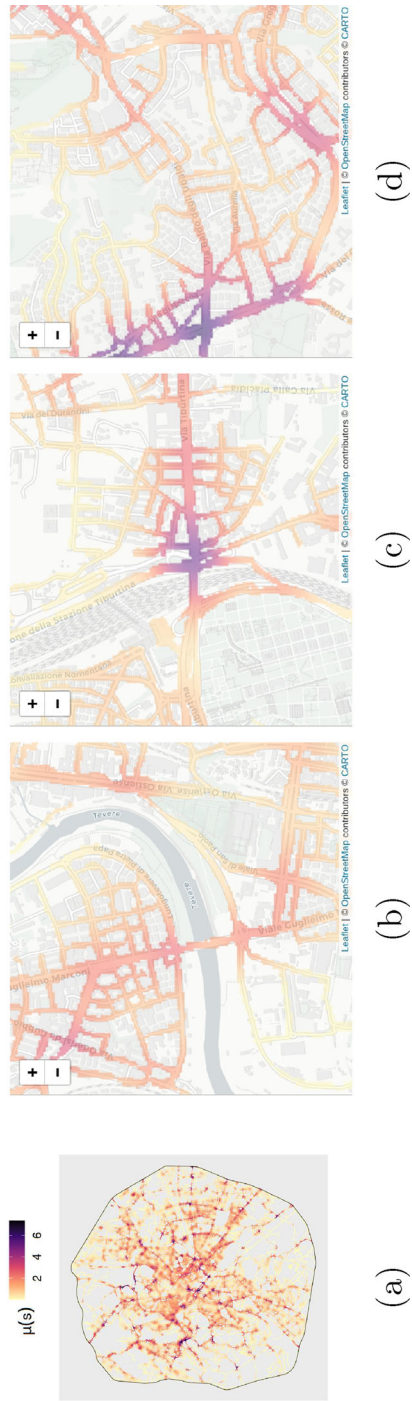


Figure 8. **a** Spatial background in 2019; Zoom: **b** Ponte Marconi, **c** Ponte Tiburtino, **d** Baldo degli Ubaldi.

Table 3. Estimated background and triggering coefficients from 2019 to 2021 (standard error). The triggering coefficients represent deviations from the baseline category (1)

Year	$\hat{\mu}_0$	$\hat{\beta}_0$ (1)	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
2019	2.21 (0.007)	- 2.85 (0.23)	1.22 (0.24)	0.24 (0.27)	- 0.47 (0.26)
2020	1.49 (0.008)	- 2.32 (0.22)	0.55 (0.24)	- 0.94 (0.34)	- 1.02 (0.27)
2021	2.03 (0.007)	- 3.56 (0.34)	2.04 (0.35)	0.99 (0.37)	- 0.20 (0.37)

excitation effect.² They include eye-catching dynamics such as the vehicle overturn, which might capture the attention of the passing driver and make him lose focus on driving. If we want to get an estimate of the average number of observed road accidents that were actually due to the excitation effect, we can evaluate $\sum_{ij} \hat{\rho}_{ij} \approx 328, 190, 293$, corresponding to percentages of 1.5%, 1.2%, 1.4% triggered secondary accidents. Figure 7d, e shows the estimated temporal and spatial excitation in the left and right panels, respectively. These do not differ significantly across the years. In particular, the excitation decreases steeply getting away in space or time from the original event. The effect is limited to $\approx 200m$ radius in space and ≈ 2 h (0.1 of a day) in time. This can be a very useful indication for traffic authorities to delimit for how long and how far the area surrounding a primary accident should be kept under control.

6. CONCLUSION AND DISCUSSION

This research paper introduces the analysis of the spatio-temporal distribution of road accidents in the city of Rome. We consider all data from the year 2019 to 2021, as they are the most recent data available, with a particular focus on 2019. The final objective of the analysis is to verify if there is any excitation effect between road accidents occurring on Rome urban road network. Section 2 provides a deep exploratory analysis of the temporal and spatial features of the analysed point patterns. It motivates the need to include in the model a temporal trend with (weekly and daily) cyclical components, a non-homogeneous spatial intensity, and an excitation effect to capture the clustered point pattern. Therefore, we consider a semi-parametric and periodic spatio-temporal Hawkes process and extend the previous proposal in various directions. We restrict the domain to an approximation of the actual road network, where each road segment is represented by a $10 \text{ m} \times 10 \text{ m}$, so that the process intensity is evaluated only on the portion of space where the accident can actually happen. We introduce a buffer area and extend the original stochastic reconstruction algorithm to take into account the shape of the domain and to smooth an isotropic spatial excitation. This is a particularly delicate aspect as in irregular domains such as the road network the specific road configuration might affect the estimation of the excitation in some directions. The simulation study in Appendix C substantiates the practical advantages of considering a constrained isotropic spatial excitation function. Finally, we generalize the

²These values represent the average number of secondary car accidents that would be caused in the Euclidean space. When restricting the excitation on the road network the real triggering potential depends on the road density in its proximity.

excitation effect to depend on an arbitrary set of covariates. Previous preliminary attempts to analyse the same set of data using a similar model were hindered by the overwhelming computational cost to evaluate all components for such a large set of points (Alaimo Di Loro 2021). We here attain improved performances by implementing the core of the algorithm in C++ and adopting a number of computational strategies that go from multi-core parallelization to sparse-matrix representation.

The application of our model on this original set of data allowed identifying road accidents hot spots (in space and time) and verifying the presence (and strength) of the triggering effect between road accidents in an urban road network. The results suggest that the excitation has a sensible impact on the accident dynamics, with $> 1\%$ of triggered events (> 200 per year), and decays smoothly in space and time. In particular, it appears that some types of road accidents are more likely to cause others and trigger a cascading effect. All this gives policy-makers and local authorities precious indications for preventing and managing road accident events and reducing the chance of both primary and secondary collision.

There is still room for improvement in the proposed model and methods. Indeed, even if the domain on which the intensity is evaluated is an approximation of the road network, the excitation effect is still a function of the Euclidean distance between pairs of points/locations. Further developments will focus on respecting the topology of the linear road network in all the model aspects. (D'Angelo et al. 2022,b) consider the linear network in a similar model specification, and we are currently working to combine their and our ideas using efficient tools to make the computations feasible. By then, the full application could be adapted to the linear road network of Rome to provide more reliable inference. We are also working on the possibility of including auxiliary covariates to explain differences in the background components and developing feasible cross-validation strategies for model and bandwidth selection, such as the forward predictive likelihood by Adelfio and Chiodi (2015a,b). All this could yield a deeper understanding of the statistical properties of the road accident process and provide more accurate guidance in traffic management to local authorities.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funding Open access funding provided by Università degli Studi Roma Tre within the CRUI-CARE Agreement. For this work, Pierfrancesco Alaimo Di Loro was partially supported by PON "Ricerca e Innovazione" 2014–2020 (PON R & IFSE-REACT EU), Azione IV.6 "Contratti di ricerca su tematiche Green", grant number 60-G-34690-1.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

APPENDIX

A: DATA CLEANING

Due to some issues with the software used by the local police to locate the accidents, longitude and latitude coordinates should be checked and (eventually) corrected before proceeding with the analysis. We classified the records into different categories that have been processed differently:

1. Records marked using house numbers: 45.14%
2. Records marked using intersections: 38.79%
3. Records marked using a textual description: 9.83%
4. Records marked using a light pole/bus stop: 3.80%
5. Records marked using the kilometre number (i.e. distance from road beginning): 2.44%

House numbers. For the records marked using the house number, we used the geocoding API provided by MapBox.³ Whenever the name of a street was also present in close cities or regions, we provided the API with specific parameters to avoid an incorrect result. We also used the same parameters to fix errors in the parsing by the API, which sometimes returned the location in a country different from Italy that was however present in the name of the street (e.g. Viale Somalia, Corso Francia, etc.). The parameters we used are the following:

- Limit: 1
- Country: IT
- Types: address
- Proximity: 12.482478, 41.895543

Intersections. For the records marked using the intersection of two or more streets, we used data provided by OpenStreetMap.⁴ We searched for a textual matching based on the street names, and we searched for an intersection between the lines representing the streets. When no intersection was detected, we searched for the two nearest points between the sets of sections (one for each street), and we took the middle point along the segment joining the points. This may happen because the representation of the streets as a network (LINESTRING) is simplified with respect to the *true* street network.

Length of the road. For the records marked with the kilometre number (e.g. Via Appia km 1.200), we had to make some assumptions as there is no freely accessible dataset including such information. Therefore, we used OpenStreetMap to retrieve the starting point of the street: we defined the starting point of a street as the most extreme point among all the points composing the segments of the street, which is also the nearest point to the city centre

³<https://www.mapbox.com/geocoding>

⁴<https://www.openstreetmap.org/>

(note that all major roads exit from Rome). Then, we draw a circle of radius equal to the reported mileage centred on the starting of the street and took the intersection of them as an approximation of the original location.

Textual description. For the records containing a textual description, the information on the location was completely unstructured. Therefore, we identified repeated patterns (mimicking all previous situations) or operated manually to obtain the largest possible number of records represented in a structured form.

Light pole. For the records marked with a light pole, does not exist a public dataset (free or paid) containing such information. Therefore, when available, we trusted the geolocation provided by official sources. These records, however, constitute only the 2.44% of the whole records from 2019 to 2021.

B: SCHEMATIC SR ALGORITHM

Algorithm 1 reports the EM scheme without explicit mention of the periodic components, their weights, and the eventual covariates' coefficients. Without any loss of generality, the set of background and excitation components is denoted as $\{\mu(\cdot), g(\cdot)\}$, with corresponding sets of weights uniquely denoted as $\phi_i, \rho_{i,j}$, and all scalar parameters denoted as μ_0, A .

Input: Initialization $\{\mu^{(0)}(\cdot), g^{(0)}(\cdot)\}, \mu_0^{(0)}, A^{(0)}$

Set $k = 1$

while not converged do

a) Compute $\phi_i^{(k)}, \rho_{i,j}^{(k)}$ given $\{\mu^{(k-1)}(\cdot), g^{(k-1)}(\cdot)\}$ for all $(s_i, t_i) \in \mathcal{Q}_b$

b) Reconstruct $\{\mu^{(k)}(\cdot), g^{(k)}(\cdot)\}$ using all $(s_i, t_i) \in \mathcal{Q}_b$

c) Evaluate $\tilde{\phi}_i^{(k)}, \tilde{\rho}_{i,j}^{(k)}$ given $\{\mu^{(k)}(\cdot), g^{(k)}(\cdot)\}$ for all $(s_i, t_i) \in \mathcal{Q}$

d) Get $\mu_0^{(k)}$ and $A^{(k)}$ by optimizing $\ell_{\lambda_c}^E(\mathcal{Q})^{(k)}$ given $\{\mu^{(k)}(\cdot), g^{(k)}(\cdot)\}, \tilde{\phi}_i^{(k)}, \tilde{\rho}_{i,j}^{(k)}$

e) Compute $r = \ell_{\lambda_c}^E(\mathcal{Q})^{(k)} - \ell_{\lambda_c}^E(\mathcal{Q})^{(k-1)}$

 If $r > \epsilon$ go to **Step a)**, otherwise **Stop**

end

Output: Optimized components $\hat{\mu}_0, \hat{A}, \{\hat{\mu}(\cdot), \hat{g}(\cdot)\}$

Algorithm 1: Overview of the modified stochastic reconstruction algorithm.

C: SIMULATION STUDY: ISOTROPIC VERSUS ANISOTROPIC EXCITATION

Our work suggests enforcing the smoothing of an isotropic excitation function in place of an unconstrained (anisotropic) smoothing [as in [Zhuang and Mateu \(2019\)](#)]. This additional structure is motivated by the typical parametric specification of the excitation function in a spatio-temporal Hawkes process. Adopting our constrained specification can be beneficial from multiple points of views, which we list here below in detail.

1. It is computationally convenient both in terms of runtime and storage memory. The unconstrained smoothing (anisotropic) requires getting a smoothed estimate for each

possible pair of distances across the easting (x) and northing (y) axes, which scales quadratically with the number of points in each direction. On the contrary, the smoothing on the Euclidean distances (isotropic) boils down to getting a value for each possible Euclidean distance (> 0), that will then hold for all directions.

2. It is more parsimonious in terms of degrees of freedom. The model with anisotropic spatial excitation is more flexible and can recover the isotropic model as a particular case, not vice versa. Under mild assumptions, this statement can be further corroborated by mathematical arguments as the trace of the large kernel matrix \mathbf{K}_{xy} in the unconstrained *bivariate* smoothing is larger than that of the kernel matrix \mathbf{K}_d in the *univariate* smoothing of an isotropic spatial excitation function.
3. It leverages more data for each smoothed excitation value, yielding more robust estimates under the model-driven assumption of isotropy. When performing the smoothing on the 2d Euclidean space of pairwise lags, the smoothed estimate at each x - y pair benefits from the contribution of the lags lying close to that x - y location. This means that if one lag direction has more data than another the unconstrained smoothing might suffer from over-fitting and adjust accordingly. Instead, when performing the smoothing on the 1d space of Euclidean distances among points, the smoothed estimate of the excitation function at a certain distance d benefits from the contribution of all lags at a distance close to d (i.e. lying around the circle of radius d).

While point 1. is true in general, points 2. and 3. are positive if and only if the actual behaviour of the underlying excitation is isotropic. To see the competitive advantage of our isotropic modelling of the excitation function in a well-specified scenario, we conducted a simulation study.

We simulated $B = 50$ different data configurations from a spatio-temporal Hawkes Process with conditional intensity function (2) in the rectangular spatio-temporal domain $\mathcal{Q} = \mathcal{D} \times \mathcal{T}$, where $\mathcal{D} = \{x, y \in \mathbb{R} : |x| < 10, |y| < 10\}$ and $\mathcal{T} = (-5, 65)$. Given that the focus lies on the ability to recover the spatial excitation function, all data have been simulated assuming a constant background rate in space and time without any periodic component, such that $\mu_0 \cdot \mu_s(s) \cdot \mu_t(t) = \mu$ with $\mu = 0.05$. This corresponds to an expected number of ≈ 1400 background events in the given spatio-temporal window. The overall excitation level has been set to $A = 0.4$, which corresponds to the average number of offspring generated by each event. The spatial excitation function has been set equal to an exponential of parameter $\lambda = 9$ and the temporal excitation function to a half-Gaussian with mean $\mu = 0$ and variance $\sigma^2 = 0.04$. Without any loss of generality, both the excitation functions have been normalized to integrate to 1 in the interval $(0, 1)$. (In both cases, the value of the excitation over 1 would be negligible.)

We fit both the unconstrained anisotropic and the constrained isotropic version of the spatio-temporal Hawkes process through Algorithm 1 on each of the $B = 50$ simulated sets. Full estimation is performed on the restricted domain $\mathcal{Q}^* = \mathcal{D}^* \times \mathcal{T}^*$, where $\mathcal{D}^* = \{x, y \in \mathbb{R} : |x| < -7, |y| < 7\}$ and $\mathcal{T}^* = (0, 6)$, with the remaining area used as a buffering window. Bandwidths have been set to the same values. We compare the performances of the two estimation strategies according to three major aspects: (i) average computing time,

Table 4. Details on the estimation performances of the nonparametric estimation with anisotropic and isotropic spatial excitation: average computing time, bias and root-mean-squared-error (RMSE) in parameter estimation, average log-likelihood value

Model	Runtime (min)	$\hat{\mu}$		\hat{A}		$-\log(\mathcal{L})$
		Bias	RMSE	Bias	RMSE	
Iso	1.04	-7×10^{-4}	2×10^{-3}	-2×10^{-2}	2×10^{-2}	2548.355
Aniso	3.65	-3×10^{-3}	4×10^{-3}	3×10^{-2}	3×10^{-2}	2492.232

(ii) estimation accuracy of the parameters μ and A , (iii) ability to recover the true isotropic spatial excitation function shape.

The estimation performances of the two approaches are compared in Table 4. We can see how the isotropic model is computationally three times faster than the anisotropic one, needing ≈ 1 min (vs. > 3 min) to run every single estimation. Both approaches return slightly biased estimates of the true μ and A value, as expected in a kernel smoothing-based estimation, but the isotropic model has significantly lower bias and estimation error (RMSE). In particular, let us point out how the true values of μ and A are included in the 90% confidence interval of the isotropic estimates ((0.046, 0.052) for μ and (0.352, 0.402) for A), while A is not included in the one for the anisotropic estimates ((0.043, 0.050) for μ and (0.401, 0.451) for A).

Last but not least, Fig. 9 reports the spatial excitation recovered under the isotropic model. Figure 9a shows the spatial excitation as a function of the Euclidean distance, where the shaded area is the envelope of the estimated spatial excitation functions across the $B = 50$ simulated sets, while Fig. 9b shows the average estimated excitation function on the 2d Euclidean space.

On the other hand, Fig. 10 shows the spatial excitation recovered under the anisotropic model. In particular, Fig. 10a reports the average estimated excitation function on the 2D Euclidean space, while Figs. 10b–d report three examples of estimated spatial excitation functions on three specific sets. We can clearly notice how, on average, the anisotropic model can recover the true isotropic behaviour of the spatial excitation function, even if it is not able to capture fully the peak of the excitation close to $(0, 0)$ (it is downward biased). Furthermore, the recovery of the isotropic shape is not true for the single estimation on a single data configuration. The unconstrained smoothing tends to adapt to the specific configuration and if, by any chance, one direction happens to have more data points than another, then the model adapts and captures this as anisotropy in the excitation. This kind of behaviour disappears on average and is soothed in a data-rich context such as the one described in Section D.

D: ANISOTROPIC EXCITATION ON THE REAL-DATA APPLICATION

This section is dedicated to comparing our model proposal and its counterpart with anisotropic spatial excitation. Indeed, we argue that an isotropic spatial excitation could

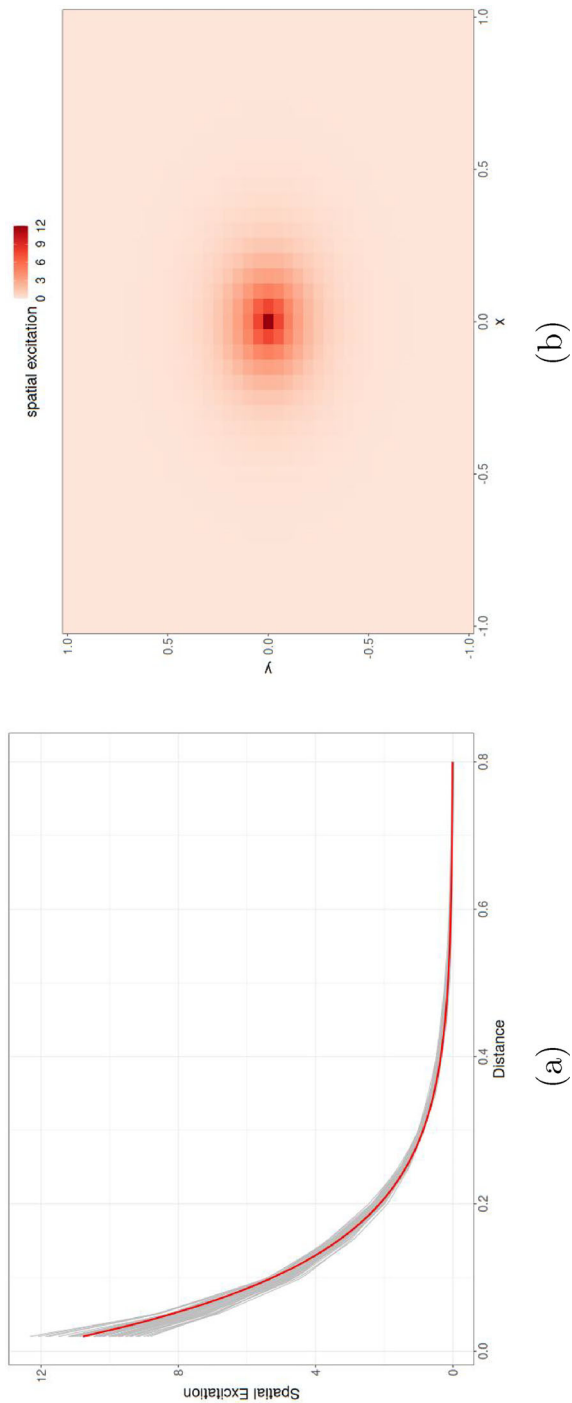


Figure 9. **a** Estimated spatial excitation for each simulated dataset (grey line) and true spatial excitation (red line); **b** average isotropic spatial excitation reported on a regular grid (Color figure online).

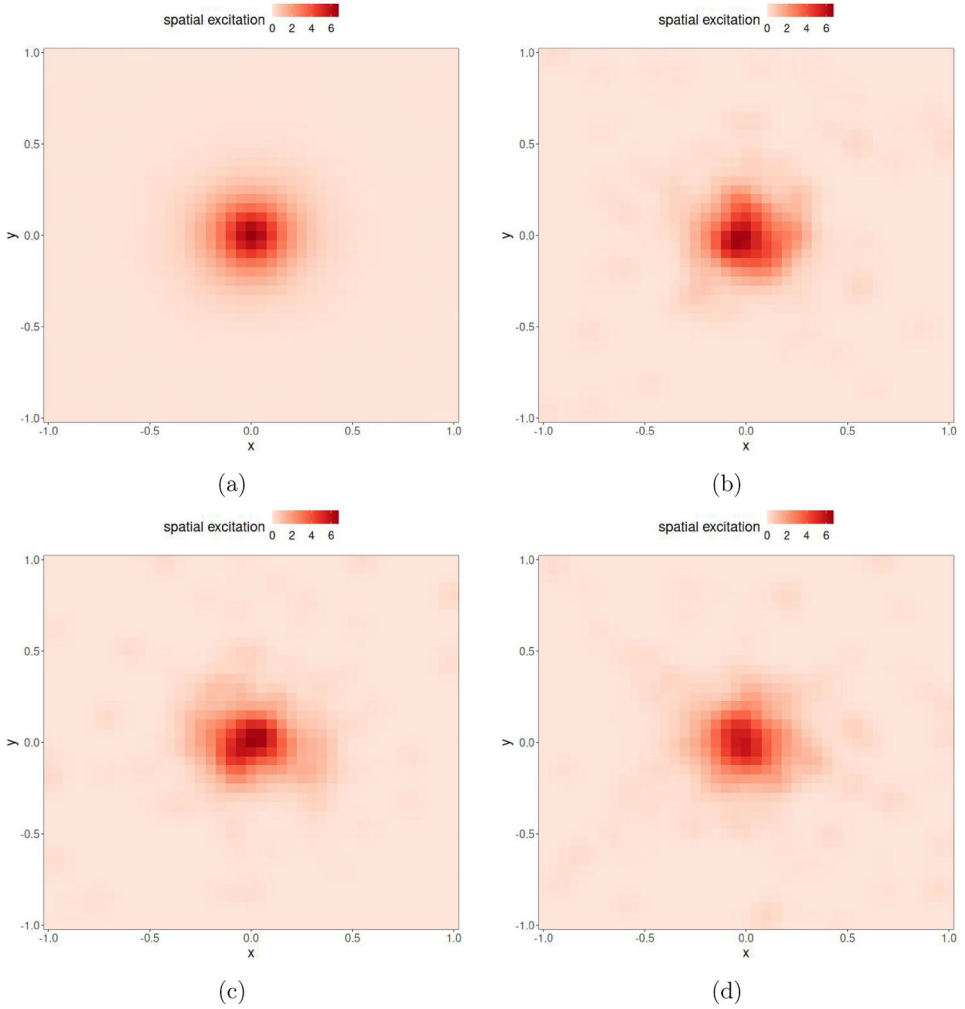
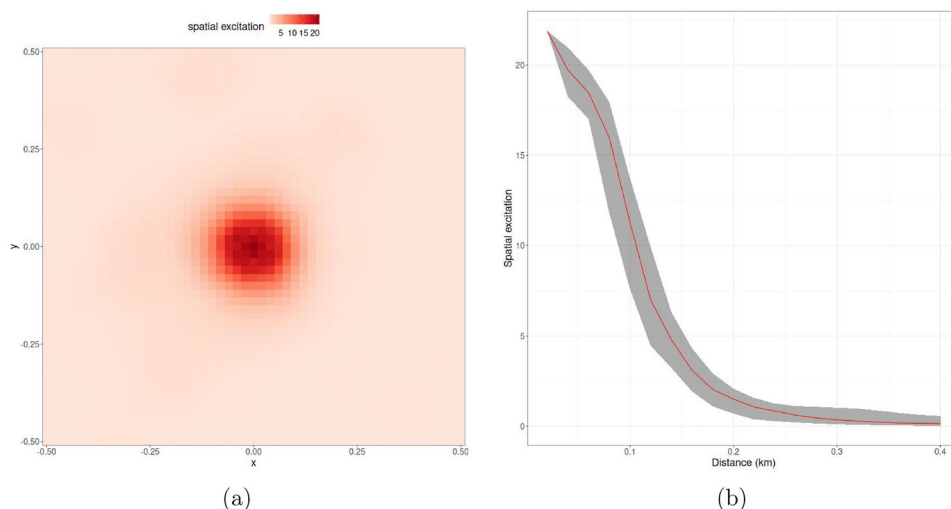


Figure 10. Estimates of the anisotropic spatial excitation: **a** average across the $B = 50$ simulated datasets; **b–d** three examples of single estimates.

better fit the data under consideration and provide a more robust parameters estimation, while also reducing the computational burden (see Appendix C). For the sake of brevity, we only report the results of the comparison for the year 2019, but notice that the conclusions apply also for 2020 and 2021. In particular: (i) the estimates of the long-term, weekly and daily trends are comparable; (ii) parameters' estimates are slightly different, but again comparable, with the triggering coefficient of the accident types 3 and 4 that are larger in absolute value compared to the ones resulting from the isotropic model (see Table 5); (iii) the estimated anisotropic spatial excitation (Fig. 11a) is practically isotropic (see Fig. 11b). It is important to notice that the negative log-likelihood and diagnostic performances under the anisotropic specification are extremely close to that of the isotropic one, with $-\log(\mathcal{L}) = -8911.468$ (vs. -8911.04) and 94.6% of the $\hat{\tau}_i$ included within the nominal 0.95 confidence bounds (vs. 95.3%).

Table 5. Estimated background and triggering coefficients from 2019 to 2021 (standard error). The triggering coefficients represent deviations from the baseline category (1)

Year	$\hat{\mu}_0$	$\hat{\beta}_0$ (1)	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Iso	2.21 (0.007)	- 2.85 (0.23)	1.22 (0.24)	0.24 (0.27)	- 0.47 (0.26)
Aniso	2.21 (0.007)	- 2.72 (0.22)	1.24 (0.23)	0.42 (0.26)	- 0.72 (0.26)

Figure 11. **a** Estimated anisotropic spatial excitation; **b** average spatial excitation (red line) and range (grey area) obtained from the anisotropic estimate for increasing distances (Color figure online).

REFERENCES

- Acker B, Yuan M (2019) Network-based likelihood modeling of event occurrences in space and time: a case study of traffic accidents in Dallas, Texas, USA. *Cartogr Geogr Inf Sci* 46(1):21–38. <https://doi.org/10.1080/15230406.2018.1515037>
- Adelfio G, Chiodi M (2015) Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stoch Environ Res Risk Assess* 29:443–450
- Adelfio G, Chiodi M (2015) Flp estimation of semi-parametric models for space-time point processes and diagnostic tools. *Spat Stat* 14:119–132
- Alaimo Di Loro P, Ciminello E, Tardella L (2019) Hidden Markov Model estimation via Particle Gibbs. *Book Short Pap SIS 2019*:829–835
- Alaimo Di Loro P (2021) Innovative approaches in spatio-temporal modeling: handling data collected by new technologies
- Bacry E, Mastromatteo I, Muzy J-F (2015) Hawkes processes in finance. *Market Microstruct Liq* 1(01):1550005
- Borgoni R, Gilardi A, Zappa D (2021) Assessing the risk of car crashes in road networks. *Soc Indic Res* 156:429–447
- Briz-Redón Á, Iftimi A, Mateu J, Romero-García C (2023) A mechanistic Spatio-temporal modeling of Covid-19 data. *Biom J* 65(1):2100318
- Chiang W-H, Liu X, Mohler G (2022) Hawkes process modeling of Covid-19 with mobility leading indicators and spatial covariates. *Int J Forecast* 38(2):505–520

- Chiu S-T (2000) Boundary adjusted density estimation and bandwidth selection. *Stat Sin* 10:1345–1367
- Comi A, Persia L, Nuzzolo A, Polimeni A (2018) Exploring temporal and spatial structure of urban road accidents: some empirical evidences from rome. In: *The 4th conference on sustainable urban mobility*. Springer, pp 147–155
- Daley DJ, Vere-Jones D (2007) *An introduction to the theory of point processes: volume II: general theory and structure*. Springer, New York
- Daley DJ, Vere-Jones D (2003) *An introduction to the theory of point processes: volume I: probability and its Applications*. Springer, New York
- D'Angelo N, Adelfio G, Abbruzzo A, Mateu J (2022) Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *Ann Appl Stat* 16(2):791–815
- D'Angelo N, Payares D, Adelfio G, Mateu J (2022) Self-exciting point process modelling of crimes on linear networks. *Stat Model* 24:1471082X221094146
- Diggle PJ (2013) *Statistical analysis of spatial and Spatio-temporal point patterns*. CRC Press, Boca Raton
- Fox EW, Schoenberg FP, Gordon JS (2016) Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *Ann Appl Stat* 10:1725–1756
- Gilardi A, Mateu J, Borgoni R, Lovelace R (2022) Multivariate hierarchical analysis of car crashes data considering a spatial network lattice. *J R Stat Soc Ser A Stat Soc* 185(3):1150–1177
- Hawkes AG (1971) Point spectra of some mutually exciting point processes. *J R Stat Soc Ser B (Methodol)* 33(3):438–443
- Hawkes AG (2018) Hawkes processes and their applications to finance: a review. *Quant Finance* 18(2):193–198
- Hawkes AG, Oakes D (1974) A cluster process representation of a self-exciting process. *J Appl Probab* 11:493–503
- Kalair K, Connaughton C, Loro PAD (2021) A non-parametric Hawkes process model of primary and secondary accidents on a UK smart motorway. *J R Stat Soc Ser C Appl Stat* 70(1):80–97
- Li Z, Cui L, Chen J (2018) Traffic accident modelling via self-exciting point processes. *Reliab Eng Syst Safe* 180:312–320. <https://doi.org/10.1016/j.ress.2018.07.035>
- Marsan D, Lengline O (2008) Extending earthquakes' reach through cascading. *Science* 319(5866):1076–1079
- Marsan D, Lengliné O (2010) A new estimation of the decay of aftershock density with distance to the mainshock. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2009JB007119>
- Mohler GO, Short MB, Jeffrey BP, Schoenberg FP, Tita GE (2011) Self-exciting point process modeling of crime. *J Am Stat Assoc* 106(493):100–108
- Ogata Y (1999) *Seismicity analysis through point-process modeling: a review*. Seismicity patterns, their statistical significance and physical meaning. Springer, New York, pp 471–507
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83(401):9–27
- Rathbun SL (1996) Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *J Stat Plan Inference* 51(1):55–74
- Reinhart A et al (2018) A review of self-exciting spatio-temporal point processes and their applications. *Stat Sci* 33(3):299–318
- Silverman BW (2018) *Density estimation for statistics and data analysis*. Routledge, Oxfordshire
- Veen A, Schoenberg FP (2008) Estimation of space-time branching process models in seismology using an em-type algorithm. *J Am Stat Assoc* 103(482):614–624
- Wang Q, Paik SF, Jackson David D (2010) Standard errors of parameter estimates in the etas model. *Bull Seismol Soc Am* 100(5A):1989–2001
- Zhuang J (2006) Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J R Stat Soc Ser B (Stat Methodol)* 68(4):635–653
- Zhuang J, Mateu J (2019) A semiparametric spatiotemporal hawkes-type point process model with periodic background for crime data. *J R Stat Soc A Stat Soc* 182(3):919–942

Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. *J Am Stat Assoc* 97(458):369–380

Zhuang J, Ogata Y, Vere-Jones D (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2003JB002879>

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