La Matematica (2026) 5:10
https://doi.org/10.1007/s44007-025-00177-7

ORIGINAL RESEARCH ARTICLE

Social Justice Implications of Model Selection for Point Process Models of Crime

Nicholas Clark[1] · **Jorge Mateu**[2]

Received: 1 October 2024 / Revised: 11 September 2025 / Accepted: 9 December 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2026

Abstract

Socioeconomically disadvantaged populations are often disproportionately subjected to over policing. While sometimes well intended, other times over policing is due to a belief that a response to crime must be swift in order to deter possible repeat actors. However, crime inspired by the action of another criminal is not the only reason why these events may be clustered in space and time. A competing theory states that spatio-temporal clustering occurs due to underlying socio-economic conditions rather than inspired actors. In quantitative criminology, repeat victimization attributed to copy-cat actors is often modeled through the use of a self-exciting, or Hawkes, process. This process is often assumed to exist prior to data analysis and alternative processes are rarely considered. In this manuscript, we will discuss how model selection, in particular model selection between a log Gaussian Cox process and a Hawkes process, is both a necessary as well as difficult step in statistical modeling of crime. We will provide a few techniques to conduct model selection between these processes and conclude, with a warning for researchers in this area, that sometimes these processes cannot be disentangled. In these instances, we suggest that modelers explicitly mention that their models rely on one theory of repeat victimization and that alternative theories may exist that lead to other forms of policing and may impact their interpretation of the root cause of why crime is spreading in space and time.

Keywords Spatio-temporal · Point process · Model selection · Criminology

✉ Nicholas Clark
clar91122@stthomas.edu

1 Department of Mathematics, University of St Thomas, 2115 Summit Avenue, 55105 St. Paul, Minnesota, USA

2 Department of Mathematics, University Jaume I, 12071 Castelló de la Plana, Castellon, Spain

Published online: 13 January 2026

1 Introduction

A long standing assumption in criminology is that crime can spread via a contagion-like process [1, 2]. Within the criminology literature, researchers describe this contagion-like process as arising from either flagged events or boosted offenses [3, 4]. Flagged events, or state heterogeneity, claims that targets appear to criminals as more attractive due to features of the particular location and are victimized due to these features. Boosted events, on the other hand, suggests that offenders learn which targets are easily victimized through the acts of other criminals and it is the occurrence of a previous event that boosts the probability that a new event will occur. Perpetrators of a crime, in this instance, are copying the actions of other individuals.

The boosted events theory is associated with prevention strategies such as quick responses to known areas of violence [5, 6]. Other responses to repeat victimization driven by a boosted event are quickly removing signs of property damage, removing or protecting targets, or short term regulating or controlling access [7]. Each of these response strategies involve an external police force acting quickly with a visible presence in the area of the crime.

Flagged event theory, on the other hand, suggests policing strategies that focus on addressing the underlying causes for the contagion. Here repeat victimization could be controlled for by hardening targets, or removing socio-economic conditions that exist in a given region. Clearly the responses to boosted events appear to be much more heavy handed and may negatively impact potentially vulnerable populations. Most concerning would be the misidentification of the root cause resulting in a response for a boosted event whereas the actual cause of the repeat victimization was a flagged event.

Regardless of the underlying cause, we can expect that the spatio-temporal pattern of data from a contagion-like process would appear to be clustered. However, modeling clustered data is not entirely straightforward. In fact, multiple processes can generate similar spatio-temporal cluster patterns [8]. In general, these models can be broadly classified as observation-driven models or parameter-driven models [9, 10]. Recently, observation-driven models, or Hawkes processes have become increasingly common to use in the modeling of criminal behavior [11–15]. However, as we will argue in this manuscript, using observation-driven models presupposes that crime spreads via the boosted theory of contagion. Assuming that crime spreads via flagged events would, in fact, assume a different generative process. Here we will argue that the natural process to generate data from the flagged theory would be more akin to a log Gaussian Cox process [16]. Critically, by picking the wrong underlying process, conclusions could be drawn that would negatively impact vulnerable populations. Specifically, as the response to boosted events is typically a heavy handed police response, misidentifying the models may result in overpolicing rather than addressing underlying reasons that some targets are more attractive than others.

In this manuscript we will discuss the basics of modeling spatio-temporal cluster processes focusing on Hawkes processes and log Gaussian Cox processes. We will then connect the processes to common theories in criminology and discuss how the proper theoretical underpinning is a social justice issue. We will close with a discussion on the difficulties in differentiating between the two processes and demonstrate how failure to disentangle the processes results in incorrect conclusions regarding the diffusion of crime in space-time and hence incorrect conclusions regarding proper police responses.

2 Spatio-temporal Clustering Point Processes

A spatio-temporal point process is defined on a bounded subset, say $B_{s,t}$ of $\mathbb{R}^2 \times \mathbb{R}$. For any subset $A \in B_{s,t}$ which is Lebesgue measurable we allow $Z(A)$ to denote the number of events in A. For completeness we assume $Z(A)$ is almost surely finite, then here $\{Z(A) : A \in B_{s,t}\}$ characterizes a point process (see e.g., p.205 of [17]).

In this manuscript we will consider Poisson point processes which extend straightforwardly through assuming $Z(A) \sim Po(\lambda^0|A|)$ where $Po(\cdot)$ denotes a Poisson distribution and $|A|$ is the three dimensional space-time volume under consideration. The stipulation above assumes a constant background intensity λ^0 which is often not realistic and can be relaxed through allowing a space-time varying background rate $\lambda(s,t)$. This assumes the intensity varies at location $s \in \mathbb{R}^2$ and $t \in \mathbb{R}$. Typically, we place an upper bound on the time, say T, so we can modify this as $t \in [0,T]$. In the case of a varying background rate we have $Z(A) \sim Po\left(\int_A \lambda(s,t)dsdt\right)$.

Spatio-temporal point processes are uniquely characterized by their conditional intensity [18] where the conditioning occurs conditional on the past, or conditional on the temporal component of the model. While a log likelihood function can be easily written out for a general spatio-temporal point process, inference is more difficult and sometimes relies on techniques such as minimum contrast estimation which essentially uses point estimates for higher order moments to functions of parameters minimizing the squared loss over a user specified range (for more see [19]). Alternative techniques for fitting spatio-temporal point processes are to create spatio-temporal grids and count the number of observations that occur in each grid. The process, then, can be represented through a vector, say $Z_t = (Z(s_1,t), Z(s_2,t), \dots, Z(s_n,t))$. The collection of vectors, $\{Z_1, \dots, Z_T\}$ then can be modeled using a parametric method. While this simplifies the calculations considerably, the choice of how to grid out the spatio-temporal field is arbitrary and different choices of grid structure may impact the results. In spatial statistics, this is referred as the modified areal unit problem (MAUP) [20]. However, some recent work in [21] suggests techniques for ensuring the grid choices does not significantly impact the analysis.

In this manuscript we will consider two spatio-temporal point processes that have been used in the analysis of crime and violence. In particular, we will discuss the underlying assumptions and model formulation for both the Hawkes process as well as the log Gaussian Cox process.

2.1 Hawkes Processes

Hawkes processes, proposed in [22], fall into the category of observation-driven processes of [10]. Hawkes processes are also a type of cluster process that directly model the clustering behavior through parent processes which are drawn from a background rate and offspring events centered on the parent. However, unlike other observation-driven cluster processes such as Neyman-Scott processes or Matérn processes, Hawkes processes allow for offspring to generate additional offspring. These processes are also referred to as self-exciting processes. While initially formulated as temporal processes, they have recently been extended to spatio-temporal processes. For an overview of recent developments see [13].

To define a Hawkes process, we use the conditional intensity, conditioning on the history of the process up to point t, denoted as $\mathcal{H}_t$. The conditional intensity, then, is denoted as $\lambda(s, t|\mathcal{H}_t)$. The spatio-temporal Hawkes process has conditional intensity of the form

$$\lambda(s, t|\mathcal{H}_t) = \nu(s, t) + \sum_{i:t_i<t} g(s - s_i, t - t_i). \tag{1}$$

From here, we see that the parents are generated according to $\nu(s, t)$ and the offspring are generated according to the triggering kernel $g(., .)$. To simplify this, typically assumptions are made such as the background rate being constant in time and the triggering kernel to be separable in space-time, yielding

$$\lambda(s, t|\mathcal{H}_t) = \nu(s) + \sum_{i:t_i<t} f_1(s - s_i) f_2(t - t_i). \tag{2}$$

Quite often f_1 is assumed to be Gaussian and f_2 is exponential, giving

$$\lambda(s, t|\mathcal{H}_t) = \nu(s) + \sum_{i:t_i<t} \frac{\theta}{2\pi\phi\sigma^2} \exp\left(-(t - t_i)/\phi\right) \exp\left(\frac{-||s - s_i||}{2\sigma^2}\right). \tag{3}$$

An important parameter in (3) is θ which controls the expected number of offspring for a given point. This parameter, sometimes referred to as the self-excitation parameter, determines whether a system decays or grows exponentially. The parameter ϕ impacts the temporal decay and σ^2 controls the spatial decay for each triggering event. A larger ϕ and larger σ^2 would allow an event to impact a wider temporal and spatial region.

An example of a spatio-temporal Hawkes process with a constant background rate that has been margninalized over time is shown in Fig. 1. Here we can see substantial spatial clustering, however there are some parent events that never generate offspring. The process is ‘self-exciting’ in the extent that other clusters have multiple offspring that generate their own offspring as in a branching process.

Hawkes processes have had extensive use in seismology starting with [23], though with a different kernel created through the knowledge of how earthquakes propagate to aftershocks. Along with modeling crime, these processes have also been used to model financial transactions [24], social media [25], disease transmission [26], how protest activities spread across a country [27], and terrorism [28].

While the majority of these are driven by theoretical conceptualizations of the model, such as an initial social media post (parent process) followed by subsequent

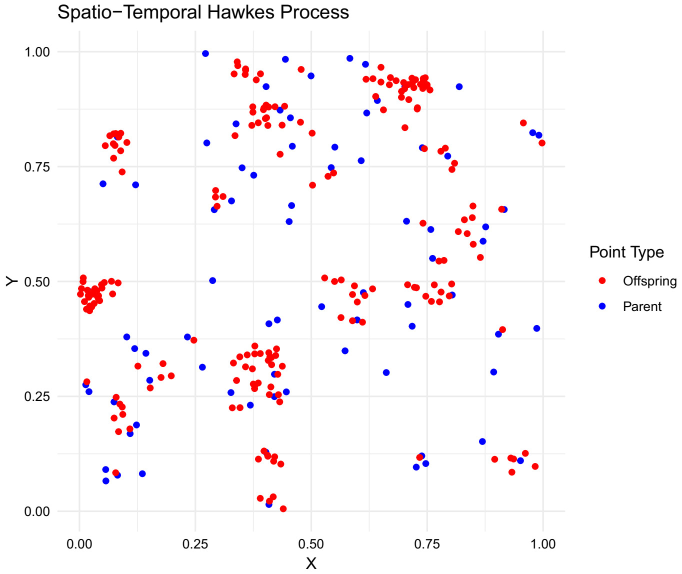

Fig. 1 Spatio-temporal Hawkes process with a constant background rate, an exponential kernel in time and a Gaussian kernel in space

re-posts (offspring process) as in [25], other times the process is more descriptive as in the case of epidemiological models where diseases must have a parent case at some level.

2.2 Log Gaussian Cox Processes

An alternative approach would be to use what [10] would refer to as a parameter-driven model. One common parameter-driven model that is used to model spatio-temporal point process data is a log Gaussian Cox process (LGCP). LGCPs can be viewed as latent Gaussian processes where the expectation of the Gaussian process represents both spatio-temporally varying observed covariates as well as unobserved variation. Initially developed as spatial models, they have been subsequently extended to spatio-temporal point process data, see e.g., [29]. LGCPs assume that the spatio-temporal correlation between regions in $\mathbb{R}^3$ is driven through exogeneous factors and can be expressed through both large-scale (observed) factors and small-scale (unobserved) variation.

To define an LGCP we need to define the structure of the observed factors as well as the structure of the unobserved factors. To do this, we assume that the log of the intensity is a Gaussian process, say $Y(s, t)$ with expectation $x(s,t)\beta$ and covariance function $C(\{s_i, t_i\}, \{s_j, t_j\})$. The covariance function must be positive-definite in order to ensure that a valid joint distribution exists.

The complicated nature of expressing an LGCP belies the straightforward assumptions that underlie the process. That is, if we consider any two spatio-temporal locations, say (s_1, t_1) and (s_2, t_2), an LGCP assumes that the intensity of each location is driven by a log Gaussian process, $\lambda(s_1, t_1) \equiv \exp(Y(s_1, t_1))$

and $\lambda(s_2, t_2) \equiv \exp(Y(s_2, t_2))$. Where $E[Y(s_1, t_1)] = x(s_1, t_1)'\beta$ and $E[Y(s_2, t_2)] = x(s_2, t_2)'\beta$ with Cov $(Y(s_1, t_1), Y(s_2, t_2)) = C((s_1, t_1), (s_2, t_2))$.

This formulation is familiar to practitioners who are used to using generalized linear mixed models. Typically $C(., .)$ is chosen to reflect higher correlation between spatially and/or temporally adjacent locations. Sometimes this is further simplified to assume separability in space and time, that is $C(\{s_i, t_i\}, \{s_j, t_j\}) = C_1(s_i, s_j)C_2(t_1, t_2)$ though this is not strictly necessary, see e.g., [30].

As discussed in [29], LGCP are typically used for point process phenomena that are driven through environmental factors rather than through the interaction between observations. Recent uses along these lines have been in modeling disease outbreaks with known socio-economic risk factors [31] and the modeling of pollution [32]. LGCPs have also been used to justify models of criminal behavior in [33–35] and others.

In Fig. 2 we can see clear clustering in space after we marginalize over time that is not obviously different than the clustering provided in Fig. 1. However, unlike the Hawkes processes, there are no parent/child events to differentiate. Of note, both Fig. 2 and Fig. 1 assume a constant expectation. Therefore, any clustering observed in Fig. 2 is due to the $C(\cdot, \cdot)$ function which captures spatio-temporal clustering not accounted for by covariates in the model and any clustering observed in Fig. 1 is due to self-excitation in the $g(\cdot, \cdot)$ function of Eq. (1).

2.3 Connections with Criminology Theories

Each of the processes described above have direct correlations with existing theories of criminal activity. Specifically, the Hawkes process relates to the boosted events

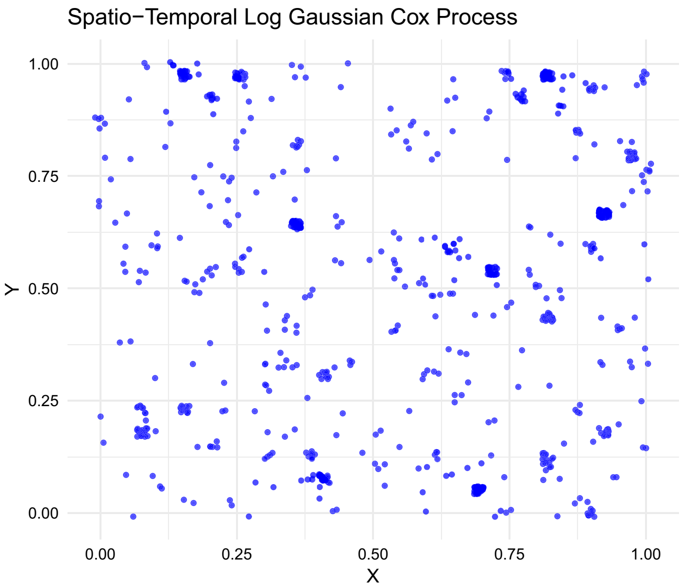

Fig. 2 A log Gaussian Cox process with constant expectation

theory whereas the LGCP relates to the flagged events theory. An early example of using these theories to motivate a statistical model was in [12] where the authors justified a Hawkes process through modeling near-repeat victimization and the broken windows effect which allowed the intensity to spread spatially from each house to its neighbors. Here the authors started with a discrete spatio-temporal model

$$\lambda(s_i,t) = \left(\lambda(s_i,t-1) + \frac{\eta\iota^2}{z}\triangle\lambda(s_i,t-1)\right)(1-\omega) + \theta Y(s_i,t-1). \quad (4)$$

A discrete spatio-temporal model, similar to a lattice model in spatial statistics, aggregates counts over a set of spatio-temporal regions. That is (s_i, t) indexes a unique geographic location region at a unique point in time. In Equation (4) the authors allow for a lagged intensity as well as repeat victimization through θ. In (4) $\triangle$ is a discrete spatial Laplacian operator, essentially looking at the difference between the intensity at s_i and the intensity at all spatially adjacent neighbors. While the authors here use an agent-based-model to generate a partial-differential-equation based model, we can slightly modify this to arrive at a discrete version of a Hawkes process.

If we first assume that $\lambda(s_i, 0) = 0$ and ignore the spatial spread by setting $\eta = 0$ in (4), we can arrive at:

$$\lambda(s_i,t) = \nu + \sum_{j<t}(1-\omega)^{t-j-1}\theta Y(s_i,t-j). \quad (5)$$

This is a discretization of a temporal Hawkes process with an exponential kernel for the temporal decay. That is, with a kernel of $\alpha e^{\beta(t-t_i)}$. To see this, consider three events that occur at times $t = 0.5, 0.7, 0.9$ if time has been discretized; then, each of these events have the same contribution to excitation at time $t = 1$. Here they would each contribute $\alpha e^{-\beta}$. Therefore the entire excitation would be $3\alpha e^{-\beta}$. Letting $\theta = \alpha e^{-\beta}$ we arrive at (5). While this model assumes a constant background rate, further modifications of the Hawkes process have been made to account for belief that differing spatial locations may have differing dynamics. In [36] the authors modified the Hawkes process as

$$\lambda(s,t|\mathcal{H}_t) = \exp\left(\beta X(s)\right) + \sum_{i:t_i<t} f_1(s-s_i)f_2(t-t_i), \quad (6)$$

where in (6) $X(s)$ are spatially varying covariates unique to cell s. In [37] the authors considered a similar model with $\exp\left(\beta X(s) + \gamma\right)$ where γ was a spatially structured error term attempting to capture both measured as well as unmeasured spatial structure in the background rate. An alternative approach was proposed in [11], where $\lambda(s_i, t)$ in (5) was multiplied by $\exp\left(\beta X\right)$.

However, each of these assumes that there exists a repeat victimization term driven through the boosted theory of criminology and while the structures in, say [37] and [36] may appear to be very similar, parameter estimates may vary widely when the structure of the models change slightly as demonstrated in [37].

If, instead, we assume a flagged event theory, the process that generates spatio-temporal clustering should be driven by external factors that can be captured as either measured covariates in the model or as a spatially varying component. In [38] these are helpfully referred to as large-scale variation, or structure in the mean component, and small-scale variation, or stochastic-dependence structure. For example, $E[Y(s_i, t)]$ from the LGCP would capture the large-scale variation and $C(., .)$ would capture the small-scale variation.

Examples of this appear in [39] where the authors assume a generalized additive model structure for the large-scale variation and an independent term for the small-scale structure. In [40] the authors justified their mean structure through environmental factors and further captured small-scale variation through a separable space-time covariance matrix similar to those described in Section 2.2.

3 Identifiability Issues

As the differing underlying processes relate to different theories in criminology, it would seem that differentiating between the two would be an important step in analyzing the cause of crimes. However, there exist identifiability issues when we look at the two processes. By identifiability we mean: given a set of data can we determine the generating statistical model? Or, in other words, can multiple models generate similar data patterns such that the data patterns are so similar that they cannot be differentiated. As an example we see in Fig. 2 and 1 both of these processes are able to generate substantial spatial clustering. While much has been done in literature on creating spatio-temporal point processes, only limited work appears to be done in model selection.

One of the issues is that while the distribution of an LGCP is completely determined by $E[Y(s_i, t)]$ and $C(\cdot, \cdot)$ [16], this is only true if we already know that the underlying process is an LGCP. That is, if we know the first and second-order-moments of the distribution can uniquely identify the generating model, it is then when we have restricted the model classes to only being LGCP.

As LGCPs are characterized by their first two moments, it is potentially useful to examine statistics relating to the second moment. One of the most common second-order-measures for point process data is Ripley's K-function [41]. In broad terms, Ripley's K-function measures the extra events (above that which could be generated by spatial randomness) within a distance h of an arbitrary event. Non-parametric estimators of the K-function are given on pages 210-213 of [17] which can be extended to spatio-temporal data in the obvious way by defining the neighborhood in $\mathbb{R}^3$ instead of $\mathbb{R}^2$. Essentially, the K-function captures the amount of spatial or spatio-temporal clustering in the data. Given an empirically generated K-function, parameters for an LGCP can be estimated through minimizing a distance between the empirical function and a theoretical function, as in Section 10.1 of [19].

Unfortunately, if we don't know the class of model in the first place, the first and second-order-moments do not uniquely define a process [42]. In [29] the authors demonstrate how LGCP and Thomas processes (a Hawkes process that does not allow for offspring to generate new offspring) can share the same second-order-properties. Therefore, first and second-order-processes such as K-functions [41] cannot be used to differentiate between some processes. This was further studied in [43] where the authors found that spatio-temporal LGCP and Hawkes processes could generate similar second-order-processes.

Outside of examining summary statistics such as Ripley's K-function and comparing them to reference distributions, alternative model selection techniques generally rely on calculating a penalized likelihood, considering nested models, or restricting the class of models to the same functional form, see e.g., [44]. Each of these techniques, however, are not able to meaningfully differentiate between LGCP and Hawkes processes.

One of the most common methods is computing a penalized version of a likelihood function. Typically these are methods like Akaike's information criterion (AIC) or Bayesian information criterion (BIC). Both of these techniques rely on being able to compute a likelihood from the data. The difficulty for computing a likelihood for an LGCP is that the density function involves an unobserved latent spatio-temporal random error term. Therefore, in order to compute a likelihood, the latent spatio-temporal error term would need to be integrated out, this is generally a high dimensional integral that is approximated. Therefore the likelihood computed is not a true likelihood. Hawkes processes, on the other hand, do have a tractable likelihood that can be maximized, however without a true likelihood from an LGCP any comparison of AIC or BIC between the two processes would not be valid. Alternatively, statistical models often rely on examining nested models and computing a statistical test. Nested models are statistical models where the parameters of one model are a subset of the parameters of the second. However, clearly the structure of the two processes is not nested, in this case. This further limits the ability of the modeler to rely on variable selection methods to differentiate the models as well.

While it is not straightforward to differentiate between Hawkes and LGCP when modeling the spatio-temporal spread of violence, failure to do so may attribute flagged events to boosted events or vice versa. The policy conclusions, then, may be to increase police presence as to combat a boosted event whereas the correct policy might be to address the underlying socio-economic risk factors as in a flagged event. Further, it is possible that events within a close proximity are both flagged and boosted which may further complicate matters. Regardless, the misidentification of underlying processes is much more apt to impact lower income regions.

3.1 Some Potential Ways Out

Here we discuss some possible techniques for differentiating between the two processes that do not rely on typical second-order analysis or methods outlined above. However, arguably none of these are a single answer that researchers can currently rely on using in all cases. While some of them rely on using higher-order statistics, beyond second-order, this offers a possibility in using machine learning methods that, again, rely on the automatic selection of higher-order features that may not be as immediately obvious as the second-order-processes discussed above. Other methods involve creating novel point process models that combine the multiple theories into a single framework.

In [37] the authors used statistics such as variance-to-mean ratio, the maximum observed count in a given cell, spatial correlation, and temporal correlation to differentiate between two competing models. This technique relied on posterior predictive checks [45]. Posterior predictive checks, typically done in a Bayesian framework, involve picking key characteristics from the data, fitting a model to the data, simulating from the fitted model, and then comparing the key characteristic calculated from the data to the key characteristic calculated from the simulated model. The researcher then finds the proportion of simulations that have a higher value than the actual data. This is sometimes referred to as a posterior predictive P-value, and both high or low values of the posterior predictive P-value would indicate that the model was not compatible with the data.

While the posterior predictive checks offer a way to use non-conventional statistics, they still rely on the analyst to know, a-priori, which statistics would be appropriate to use. An alternative method, presented in [46] relies on using convolutional neural networks to automatically extract key features from spatial point processes that are then used to differentiate between two, or more, competing models.

Convolutional neural networks (CNNs) potentially may be used by gridding out the spatio-temporal region. The number of points, then, can be counted in each region yielding a vector of length $s_1 \times s_2 \times t$ where $s_1 \times s_2$ are the length of each spatial grid and t is the length of each temporal grid. The vector, then, is passed through a convolutional neural network (for more on CNNs see [47]). The CNN, then, classifies the vector as either coming from, say, the Hawkes process or the LGCP. CNNs appear to have first been used to differentiate spatial point patterns in [46] where the authors demonstrated that these algorithms could disentangle some common spatial patterns.

While these techniques have started to be used in the spatial literature, there do not appear to be many uses, so far, in the spatio-temporal cases. One potential way to use CNNs would be to simulate from, say, a Hawkes process then fit the data to a LGCP. The fitted models then should share similar statistical characteristics. This could be done a number of times to generate a training set of LGCPs and Hawkes processes that can then be combined and fed through a CNN. Pseudo-code for this is given in Algorithm 1.

Algorithm 1 Simulation and evaluation of spatio-temporal Hawkes and log Gaussian Cox processes

```
Set parameters for the spatio-temporal Hawkes process.
for i = 1 to 50 do
    Simulate data from the spatio-temporal Hawkes process using parameters.
    Aggregate data into a spatio-temporal grid.
    Fit the aggregated data to a log Gaussian Cox process.
    Simulate data from the log Gaussian Cox process.
end for
Combine the 50 sets of Hawkes process data and 50 sets of log Gaussian Cox process data.
Fit the combined data to a convolutional neural network.
Evaluate the ability of the network to discriminate between the two processes.
```

The main benefit of using a CNN is that it does not require the practitioner to a-prior select the key features that differentiate processes from each other. Rather, the algorithm selects the features based on their ability to discriminate between the multiple processes under consideration. While these techniques are promising, much more work needs to be done on assessing the classification ability of various architectures and automating the process to make these methods practical for use by analysts and criminologists. As shown in [46] there are many instances that even the CNN fails to fully disentangle two similar processes. It remains unclear whether this is due to the underlying architecture or whether the processes just cannot be differentiated.

An alternative technique that, while seemingly logical, may be indeed be problematic is to build statistical models that combine the two processes. An early example of this is in [48] where the author considered a temporal point process model that additively combined the expectation from a log Gaussian Cox process with that of a Hawkes process. This was extended to a spatio-temporal case in [37], however with a simplified Hawkes process involving a point-mass kernel. This was more generally presented in the spatio-temporal case in [15] for the discrete case and [49] in the continuous case.

While these models are logical on the surface, there are some subtle issues with parameter interpretation that do not appear to be fully realized. As shown in [15] the choice of covariance function in the log Gaussian Cox process impacts parameters associated with the Hawkes process and vice-versa. This would mean that a conclusion about whether repeat-victimization is present in the data may be biased through the choice of what spatial covariates are included in the statistical model. While this may seem counter-intuitive, it is important to note, and to mention in any analysis, that the parameter estimates are conditional on all other terms in the model. That is, when we allow a Hawkes process to exist alongside a log Gaussian Cox process in a model, the parameter estimates from the LGCP portion can only be interpreted as conditional given the presence of the Hawkes process. Therefore, by hoping to disentangle the processes we have, in fact, more closely tied them together by no longer being able to cleanly interpret the meaning of any parameters in the model.

Perhaps the most logical path forward currently is to ensure that any analysis done using criminal data is explicit in the assumptions that are built into the model. For instance, when using a Hawkes process analysts need to ensure they explicitly state that they are assuming repeat victimization brought about through boosted events theory. In these instances, it is imperative that researchers also consider competing hypotheses and fit multiple models to show the sensitivity of any finding prior to making conclusions about why violence or crime is spreading. At a minimum, quantitative criminologists should consider log Gaussian Cox processes alongside Hawkes processes when building out inferential models for crime.

4 Example

In this section we illustrate how differing processes, which can both be justified by the data, if we only rely on summary statistics, can generate differing conclusions on the spread of violence. In particular, we will look at homicides in the city of Chicago in 2015 that is available from the R library ‘crimedata’ [50]. All results may be replicated by following the R Code available at https://github.com/nick3703/SocialJusticeCrime/tree/main.

The data, marginalized over space, is shown in Fig. 3. Here we’ve rotated the grid and standardized the longitude and latitude data.

From Fig. 3 we see some evidence of spatial clustering. This can be further analyzed by calculating Ripley’s K-function and comparing it to a homogeneous Poisson process which denotes complete spatial randomness. This is depicted in Fig. 4. Here we see that there is substantially more spatial clustering than would be expected if there were complete spatial randomness.

To examine the temporal pattern we aggregated the number of events over each week and plotted a smoothed rolling 3 week window of events and also computed the autocorrelation function for the aggregated data, as shown in Fig. 5. From here we do see an upward trend in summer months and a slight autocorrelation in the data. While this could be captured in large scale covariates in the model, for example by using temperature or another proxy measure as a covariate, here we leave this as part of the excitement or small-scale (unexplained) variation in the model

Prior to considering a spatio-temporal model, we first look at potential models for the data marginalized over time. We consider a spatial only LGCP and a Hawkes process with only a spatial spread. Specifically in (2) we let the kernels $f_2(\cdot) = 1$ and $f_1(\cdot) = \kappa \mathbb{I}_{||s-s_i||<\delta||}$. A Hawkes process with these kernels is also called a Matérn cluster process [19]. The Matérn cluster process is a spatial only parent child process where the patterns are distributed according to a homogeneous Poisson process and the children are uniformly clustered around the parent.

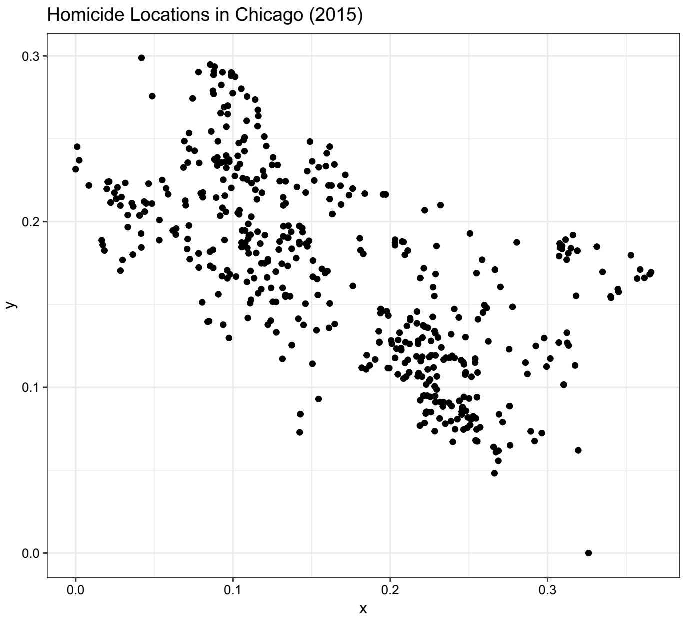

Fig. 3 Homicide locations in Chicago in 2015 placed on a regular grid

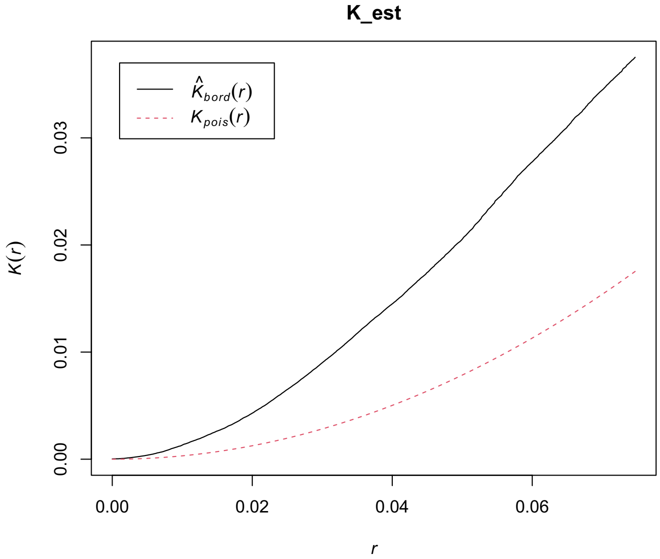

Fig. 4 Border corrected Ripley's K-function compared to theoretic K-function from a homogeneous Poisson process

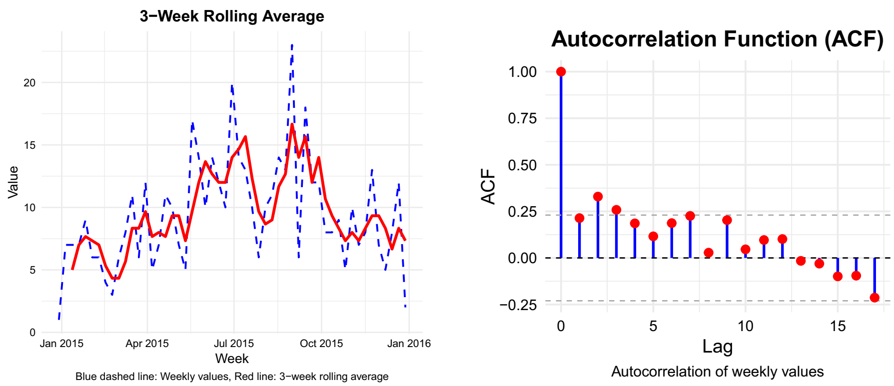

Fig. 5 Three week rolling average and an ACF plot of the weekly aggregated number of homicides in Chicago in 2015. Dashed lines from ACF plot represent approximate 95% confidence intervals

To demonstrate the difficulty in model selection for the temporally marginalized process, we can look at fitting the estimate of the K-function to the theoretic K-functions from both an LGCP and a Matérn cluster process as seen in Fig. 6.

Here we visually can see very little difference between the theoretic K-functions and the border corrected estimates of the second-order-process. It would seem, in this instance, that researchers could justify either process based upon fit to K-function. Here, the Matérn cluster process would be more akin to the boosted event theory of crime and the LGCP would be more like the flagged event theory. If, rather, we look at spatio-temporal processes we see a similar phenomenon occur.

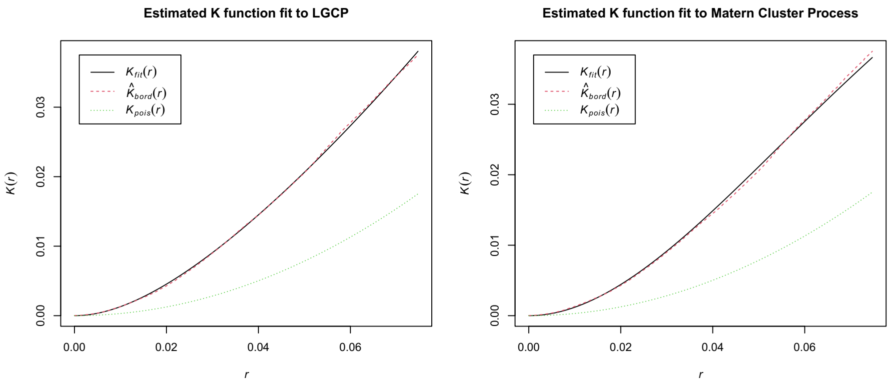

Fig. 6 Estimate of K-function from chicago data fit to both LGCP and a matérn cluster process

Next, we fit an LGCP, a spatio-temporal Hawkes process with an exponential decay in time and a Gaussian spread in space for $f_2(\cdot)$ and $f_1(\cdot)$ respectively, as well as a combined spatio-temporal Hawkes process with the addition of an LGCP. To fit these processes we use Integrated Nested Laplace Approximation (INLA) [51]. To fit the LGCP we create a gridded mesh following [21]. For more on mesh selection, see Chapter 6 of [52]. All models were fit using the stelfi [53] package in R.

The LGCP fitting function within the stelfi [53] package relies on the user inputting not only a spatial mesh as we discussed above but also a temporal mesh, which we created using 10 knots. We placed diffused priors on each of the parameters and the model took approximately 5 min to converge to parameter estimates.

The model fit here is

$$\lambda(s_i, t) = \exp(\beta + G_t(s_i) + \epsilon), \tag{7}$$

Where $G_t(s_i)$ is a Gaussian Markov Random Field (GMRF) defined on a grid as in [21]. This is an approximation to a Matérn covariance function that takes two parameters, τ and κ, which sometimes are expressed as the range and standard deviation, given by $r = \frac{\sqrt{8}}{\kappa}$ and $\sigma = \frac{1}{\sqrt{4\pi\kappa^2\tau^2}}$. Temporal structure is assumed here through an auto-regressive structure on the GMRF terms, $G_t(s_i) = \rho G_{t-1}(s) + \epsilon_i$. Here we are choosing an AR(1) type of model as it is commonly used in time series and typically offers an easy way to interpret ρ, however we note that this may not be the best model as the ACF plot suggests only weak temporal structure.

In Table 1 we see extreme temporal structure with ρ near the edge of the parameter space and some spatial structure as evident through the value of the range.

We next repeated this analysis assuming that the spatio-temporal spread was driven through a Hawkes process. This model is given by

$$\lambda(s_i, t) = \mu + \alpha \sum_{i:\tau_i < t} f_1(||s - s_i||) f_2(t - t_i), \tag{8}$$

Table 1 Parameter estimates for a spatio temporal LGCP

Parameter	Estimate	Standard error
ρ	0.998	0.00125
β	-3.34	2.45
$\log(\tau)$	-5.66	0.106
$\log(\kappa)$	2.89	0.224
Range	0.158	0.0353
Standard Deviation	4.53	0.887

Table 2 Parameter estimates with standard errors

Parameter	Estimate	Standard error
μ	6.89	0.98
α	0.013	0.002
β	0.015	0.003
x_σ	0.011	0.0.001
y_σ	0.0.011	0.0016
ρ	-0.63	0.13

where $f_1(||s - s_i||)$ is a multivariate Gaussian density with parameters $(x_\sigma, y_\sigma, \rho)$ and $f_2(t - t_i)$ is an exponential density with parameter β. Here α is the self-excitation parameter that yields the expected number of offspring from a given observation.

This model suggests that there is very little self-excitation as $\hat{\alpha} = 0.013$ in the data and perhaps no repeat victimization (see Table 2). However, to demonstrate how conclusions can change if we consider combined models as outlined in Section 3.1, we next repeated this by fitting a model that allows for both excitation as well as latent Gaussian spatial spread. That is, a combined Hawkes and log Gaussian Cox process.

The model fit here is

$$\lambda(s_i, t) = \mu + \exp(\gamma_i) + \alpha \sum_{i:\tau_i<t} f_1(||s - s_i||) f_2(t - t_i), \tag{9}$$

where $f_1(||s - s_i||)$ is a multivariate Gaussian density with parameters $(x_\sigma, y_\sigma, \rho)$ and $f_2(t - t_i)$ is an exponential density with parameter β. Here α is the self-excitation parameter that yields the expected number of offspring from a given observation and κ and τ control the spatial random effect term, γ_i in (9).

From the parameter estimates in Table 3, we can see that a conclusion from this model might be that self-excitation is present in the data due to the value of the α parameter, though the standard error may give us some concern. In Fig. 7 we overlay the expectation with the observed counts and can see that around the 175-225 day mark there may exist some times of increased excitation.

Any of the structures considered could be justified through the flagged events or the boosted events theory or a combination of the two. As all fitting methods rely on approximations, any model selection method outlined in Section 3.1 would be problematic other than clearly stating the assumptions that went into the model. For example, we could conclude that if repeat victimization is due to the flagged event theory combined with the boosted event theory, then each event would, on average, generate 0.21 additional events. However, we cannot tell, from the data and analysis

Table 3 Parameter estimates with standard errors

Parameter	Estimate	Standard error
μ	10.3	4.40
α	0.214	0.108
β	0.0505	0.175
x_σ	0.00261	0.000798
y_σ	0.00844	0.00256
ρ	0.974	0.0251
κ	33.2	9.48
τ	0.00673	0.00119

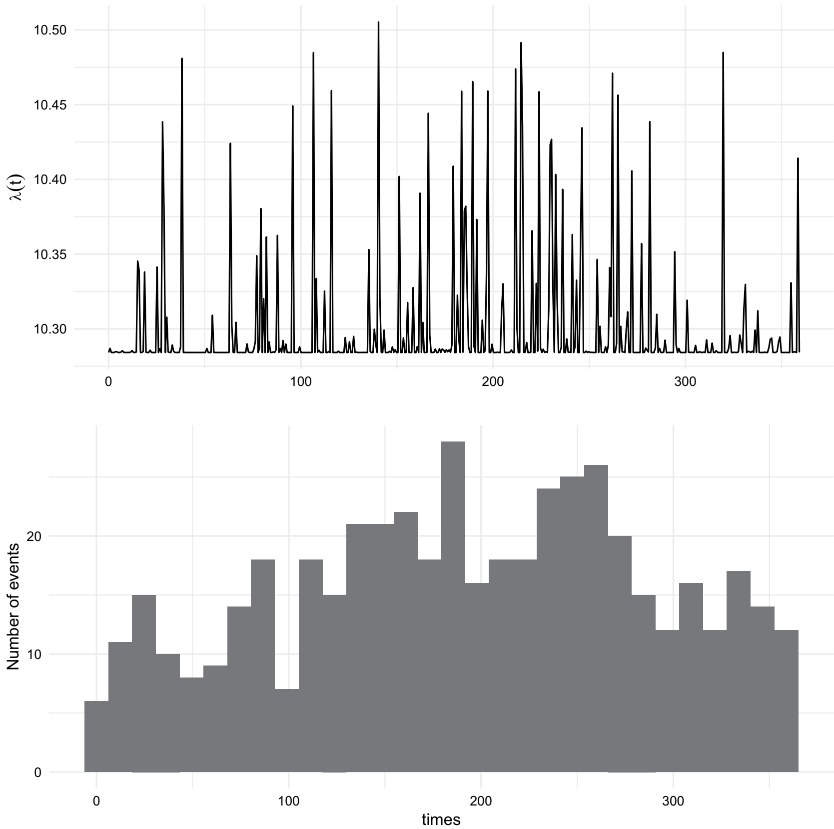

Fig. 7 Plot of the intensity of the spatio-temporal Hawkes process over time compared to the observed events

presented, whether this repeat victimization was instead due to underlying small-scale spatial or temporal exogeneous factors.

An alternative approach to determining which is the 'correct' model may be to employ a CNN as outlined above. Here we could follow the process outlined in Algorithm 2. While this may yield a more satisfying conclusion, we do not attempt this here as we remain uncertain how the aggregation, the choice of architecture, or the parameters estimated may impact the final results.

Algorithm 2 Algorithm applied to chicago data

---

for $i = 1$ to 50 **do**
    Simulate data from the spatio-temporal Hawkes process using parameter estimates.
    Aggregate data into a spatio-temporal grid.
    Fit the aggregated data to an LGCP.
    Simulate data from the LGCP.
end for
Combine the 50 sets of Hawkes process data and 50 sets of LGCP data.
Fit the combined data to a convolutional neural network.
Take original data, aggregate, and feed into the CNN.
return class probabilities.

---

5 Conclusions

In this manuscript, we argue that model selection is not only an important statistical step in building out spatio-temporal models for crime or violence, it also is a socially responsible step. While lately statistical models for crime have relied on using a Hawkes process to justify the formulation of the model, we discuss that the Hawkes process relies on the boosted theory of spatio-temporal clustering. We further discuss how an alternatively theory, the flagged theory, would be more appropriately represented as a log Gaussian Cox process with a spatio-temporal covariance matrix.

While differentiating the two process is difficult, it is necessary to either attempt to do so, or to explicitly state that the model is based off of an assumption about how violence is spreading. While there are some potentially new methods of differentiating spatio-temporal point patterns using machine learning methods [54] it is unclear whether such methods would work over the entire parameter space or what the uncertainty in these methods would be.

Most importantly, we have discussed how improper or missing model selection may impact marginalized communities and lead to incorrect conclusions about the root cause of the spatio-temporal spread of violence. We continue to stress the importance of explicitly stating the modeling assumptions and ensuring that readers of quantitative criminology are aware that the use of a Hawkes process, or self-exciting spatio-temporal process, is built off of the assumption that repeat victimization is driven by the boosted event theory and the presence of self-excitation in these models does not confirm this assumption but rather is a result of it. We are excited about the future of using advanced techniques to continue to differentiate these previous entangled processes, however we currently urge caution in their application as techniques such as model selection of point process data using convolutional neural networks is still in its infancy.

Acknowledgements We sincerely thank two anonymous reviewers whose feedback greatly improved this manuscript.

Data Availability Please note that all data and code used to support this article can be accessed at https://github.com/nick3703/SocialJusticeCrime.

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

References

1. Johnson, S.D.: Repeat burglary victimisation: a tale of two theories. J. Exp. Criminol. **4**, 215–240 (2008)
2. Sagovsky, A., Johnson, S.D.: When does repeat burglary victimisation occur? Australian & New Zealand Journal of Criminology **40**(1), 1–26 (2007)
3. Grove, L.E., Farrell, G., Farrington, D.P., Johnson, S.D.: Preventing repeat victimization: a systematic review. Brottsförebyggande rådet/The Swedish National Council for Crime Prevention, Stockholm, Sweeden (2012)
4. Lee, Y., O, S.: Flag and boost theories for hot spot forecasting: An application of nij's real-time crime forecasting algorithm using colorado springs crime data. International journal of police science & management **22**(1), 4–15 (2020)
5. Farrell, G.: Preventing repeat victimization. Crime Justice **19**, 469–534 (1995)
6. Johnson, S.D., Summers, L., Pease, K.: Offender as forager? a direct test of the boost account of victimization. J. Quant. Criminol. **25**(2), 181–200 (2009)
7. Weisel, D.L.: Analyzing repeat victimization. US Department of Justice, Office of Community Oriented Policing Services, Washington, D.C. (2005)
8. Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S.: In: Maimon, O., Rokach, L. (eds.) Spatio-temporal clustering, pp. 855–874. Springer, Boston, MA (2010). https://doi.org/10.1007/978-0-387-09823-4_44
9. Davis, R.A., Dunsmuir, W.T., Streett, S.B.: Observation-driven models for Poisson counts. Biometrika **90**(4), 777–790 (2003)
10. Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., Lauritzen, S.L.: Statistical analysis of time series: Some recent developments [with discussion and reply]. Scandinavian Journal of Statistics, 93–115 (1981)
11. Olinde, J., Short, M.B.: A self-limiting Hawkes process: interpretation, estimation, and use in crime modeling. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 3212–3219 (2020). IEEE
12. Short, M.B., D'Orsogna, M.R., Pasour, V.B., Tita, G.E., Brantingham, P.J., Bertozzi, A.L., Chayes, L.B.: A statistical model of criminal behavior. Math. Models Methods Appl. Sci. **18**(supp01), 1249–1267 (2008)
13. Reinhart, A.: A review of self-exciting spatio-temporal point processes and their applications. Stat. Sci. **33**(3), 299–318 (2018)
14. D'Angelo, N., Payares, D., Adelfio, G., Mateu, J.: Self-exciting point process modelling of crimes on linear networks. Stat. Model. **24**(2), 139–168 (2024)
15. Clark, N.J., Dixon, P.M.: A class of spatially correlated self-exciting statistical models. Spatial Statistics **43**, 100493 (2021)
16. Møller, J., Syversveen, A.R., Waagepetersen, R.P.: Log Gaussian Cox processes. Scand. J. Stat. **25**(3), 451–482 (1998)
17. Cressie, N., Wikle, C.K.: Statistics for spatio-temporal data. John Wiley & Sons, Hoboken, New Jersey (2011)
18. Fishman, P., Snyder, D.: The statistical analysis of space-time point processes. IEEE Trans. Inf. Theory **22**(3), 257–274 (1976)

19. Moller, J., Waagepetersen, R.P.: Statistical Inference and Simulation for Spatial Point Processes. CRC Press, Boca Raton, Florida (2003)
20. Wong, D.W.S.: In: Janelle, D.G., Warf, B., Hansen, K. (eds.) The Modifiable Areal Unit Problem (MAUP), pp. 571–575. Springer, Dordrecht (2004). https://doi.org/10.1007/978-1-4020-2352-1_93
21. Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H.: Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Biometrika **103**(1), 49–70 (2016)
22. Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. Biometrika **58**(1), 83–90 (1971)
23. Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes. J. Am. Stat. Assoc. **83**(401), 9–27 (1988)
24. Bacry, E., Mastromatteo, I., Muzy, J.-F.: Hawkes processes in finance. Market Microstructure and Liquidity **1**(01), 1550005 (2015)
25. Rizoiu, M.-A., Lee, Y., Mishra, S., Xie, L.: Hawkes processes for events in social media, pp. 191–218. Association for Computing Machinery and Morgan & Claypool, New York (2017). https://doi.org/10.1145/3122865.3122874
26. Chiang, W.-H., Liu, X., Mohler, G.: Hawkes process modeling of covid-19 with mobility leading indicators and spatial covariates. Int. J. Forecast. **38**(2), 505–520 (2022)
27. Bahid, Y., Kutsenko, O., Rodríguez, N., White, D.: The statistical and dynamic modeling of the first part of the 2013–2014 Euromaidan protests in Ukraine: the revolution of dignity and preceding times. PLoS ONE **19**(5), 0301639 (2024)
28. White, G., Porter, M.D., Mazerolle, L.: Terrorism risk, resilience and volatility: a comparison of terrorism patterns in three southeast asian countries. J. Quant. Criminol. **29**, 295–320 (2013)
29. Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M.: Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. Stat. Sci. 28(4), 542–563 (2013)
30. Cressie, N., Huang, H.-C.: Classes of nonseparable, spatio-temporal stationary covariance functions. J. Am. Stat. Assoc. **94**(448), 1330–1339 (1999)
31. Li, Y., Brown, P., Gesink, D.C., Rue, H.: Log Gaussian Cox processes and spatially aggregated disease incidence data. Stat. Methods Med. Res. **21**(5), 479–507 (2012)
32. Chao, C.-W., Ma, H.-W., Heijungs, R.: The green economy mirage? examining the environmental implications of low carbon growth plans in Taiwan. J. Ind. Ecol. **17**(6), 835–845 (2013)
33. Li, G., Haining, R., Richardson, S., Best, N.: Space-time variability in burglary risk: a bayesian spatio-temporal modelling approach. Spatial Statistics **9**, 180–191 (2014)
34. Shirota, S., Gelfand, A.E.: Space and circular time log Gaussian Cox processes with application to crime event data. Annal. Appl. Stat. **11**(2), 481–503 (2017)
35. Shirota, S., Gelfand, A.E., Mateu, J.: Analyzing car thefts and recoveries with connections to modeling origin-destination point patterns. Spatial Statistics **38**, 100440 (2020)
36. Reinhart, A., Greenhouse, J.: Self-exciting point processes with spatial covariates: modelling the dynamics of crime. J. R. Stat. Soc.: Ser. C: Appl. Stat. **67**(5), 1305–1329 (2018)
37. Clark, N.J., Dixon, P.M.: Modeling and estimation for self-exciting spatio-temporal models of terrorist activity. The Annals of Applied Statistics **12**(1), 633–653 (2018)
38. Cressie, N.: Statistics for Spatial Data. John Wiley & Sons, New York, New York (1991)
39. Wang, X., Brown, D.E.: The spatio-temporal modeling for criminal incidents. Security Informatics **1**, 1–17 (2012)
40. Yang, B., Liu, L., Lan, M., Wang, Z., Zhou, H., Yu, H.: A spatio-temporal method for crime prediction using historical crime data and transitional zones identified from nightlight imagery. Int. J. Geogr. Inf. Sci. **34**(9), 1740–1764 (2020)
41. Dixon, P.M.: Ripley's k function. Encyclopedia of environmetrics (2006)
42. Baddeley, A., Silverman, B.W.: A cautionary example on the use of second-order methods for analyzing point patterns. Biometrics **40**(4), 1089–1093 (1984)
43. Clark, N.J., Watts, K.: Identification of latent structure in spatio-temporal models of violence. In: 2021 winter simulation conference (WSC), pp. 1–8 (2021). IEEE
44. Guttorp, P., Thorarinsdottir, T.L.: Bayesian inference for non-Markovian point processes. In: advances and challenges in space-time modelling of natural events, pp. 79–102 (2012). Springer
45. Gelman, A., Meng, X.-L., Stern, H.: Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica **6**(4), 733–760 (1996)
46. Jalilian, A., Mateu, J.: Assessing similarities between spatial point patterns with a siamese neural network discriminant model. Adv. Data Anal. Classif. **17**(1), 21–42 (2023)

47. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans. Neural Network. Learn. Syst. **33**(12), 6999–7019 (2021)
48. Mohler, G.: Modeling and estimation of multi-source clustering in crime and security data. Annal. Appl. Stat. **7**(3),1525–1539 (2013)
49. Miscouridou, X., Bhatt, S., Mohler, G., Flaxman, S., Mishra, S.: Cox-Hawkes: doubly stochastic spatiotemporal poisson processes. In: 2023 IMS international conference on statistics and data science (ICSDS), p. 564 (2023)
50. Ashby, J.M.P.: Studying crime and place with the crime open database. Research Data Journal for the Humanities and Social Sciences **4**(1), 65–80 (2019). https://doi.org/10.1163/24523666-00401007
51. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. J. R. Stat. Soc. Ser. B Stat Methodol. **71**(2), 319–392 (2009)
52. Blangiardo, M., Cameletti, M.: Spatial and spatio-temporal Bayesian Models with R-INLA. Wiley (2015)
53. Jones-Todd, C.M., van Helsdingen, A.: Stelfi: Hawkes and Log-Gaussian cox point processes using template model builder (2023). R package version 1.0.1. https://CRAN.R-project.org/package=stelfi
54. Rodrigues, A., González, J.A., Mateu, J.: A conditional machine learning classification approach for spatio-temporal risk assessment of crime data. Stoch. Env. Res. Risk Assess. **37**(7), 2815–2828 (2023)

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

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