Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Investigating Clustering and Violence Interruption in Gang-Related Violent Crime Data Using Spatial–Temporal Point Processes With Covariates

Junhyung Park, Frederic Paik Schoenberg, Andrea L. Bertozzi & P. Jeffrey Brantingham

To cite this article: Junhyung Park, Frederic Paik Schoenberg, Andrea L. Bertozzi & P. Jeffrey Brantingham (2021) Investigating Clustering and Violence Interruption in Gang-Related Violent Crime Data Using Spatial–Temporal Point Processes With Covariates, Journal of the American Statistical Association, 116:536, 1674-1687, DOI: 10.1080/01621459.2021.1898408

To link to this article: https://doi.org/10.1080/01621459.2021.1898408

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1674–1687: Applications and Case Studies
https://doi.org/10.1080/01621459.2021.1898408

Investigating Clustering and Violence Interruption in Gang-Related Violent Crime Data Using Spatial–Temporal Point Processes With Covariates

Junhyung Park[a], Frederic Paik Schoenberg[a], Andrea L. Bertozzi[b], and P. Jeffrey Brantingham[c]

[a]Department of Statistics, University of California, Los Angeles, CA; [b]Department of Mathematics, University of California, Los Angeles, CA; [c]Department of Anthropology, University of California, Los Angeles, CA

ABSTRACT

Reported gang-related violent crimes in Los Angeles, California, from 1/1/14 to 12/31/17 are modeled using spatial–temporal marked Hawkes point processes with covariates. We propose an algorithm to estimate the spatial-temporally varying background rate nonparametrically as a function of demographic covariates. Kernel smoothing and generalized additive models are used in an attempt to model the background rate as closely as possible in an effort to differentiate inhomogeneity in the background rate from causal clustering or triggering of events. The models are fit to data from 2014 to 2016 and evaluated using data from 2017, based on log-likelihood and superthinned residuals. The impact of nonrandomized violence interruption performed by The City of Los Angeles Mayor's Office of Gang Reduction Youth Development (GRYD) Incident Response (IR) Program is assessed by comparing the triggering associated with GRYD IR Program events to the triggering associated with sub-sampled non-GRYD events selected to have a similar spatial–temporal distribution. The results suggest that GRYD IR Program violence interruption yields a reduction of approximately 18.3% in the retaliation rate in locations more than 130 m from the original reported crimes, and a reduction of 14.2% in retaliations within 130 m.

ARTICLE HISTORY
Received November 2019
Accepted February 2021

KEYWORDS
Criminology; Gang violence; Point processes; Self-exciting; Violence interruption

1. Introduction

Crime occurrences are highly clustered in space and time (Weisburd 2016, Mohler 2019). Theory suggests that the observed clustering in crime event data is driven by two main effects: (i) spatial heterogeneity in local risk factors and (ii) the dependence on recent crimes which may incite repeat offenses or retaliations (Heckman 1991). Unfortunately the two effects are difficult to disentangle in observed data and often confounded in statistical analyses (Diggle 2014, chap. 9.6).

Spatio-temporal clustering is particularly characteristic of gang violent crime (Martinez 2016; Valasik 2017). Explanations for the clustering in gang violent crime also point to a combination of stable, structural differences between neighborhoods (i.e., spatial heterogeneity) (Barton 2019) and the local dynamics of tit-for-tat attacks (i.e., statistical dependence) (Papachristos 2009; Brantingham 2019). Gangs tend to cluster in areas with high rates of poverty, high unemployment, under-performing schools, high rates of female-headed households, high residential instability, and high percentages of the population under the age of 18, all well-known indicators of concentrated disadvantage that change very slowly over time (Sampson 1997; Kubrin 2003; Papachristos and Kirk 2006). These neighborhood characteristics undermine a community's ability to exert social control and limit the activity of gangs (Curry 1988; Valasik 2017). Gangs are thus enduring features of the social landscape with territorial footprints that are very stable over time (Brantingham Patillo-McCoy 1999; Brantingham et al. 2019). Gang violent crime therefore tends to cluster where gangs are most active, particularly along gang territorial boundaries (Tita and Ridgeway 2007; Brantingham 2012). Moreover, since gang territories can be large (e.g., covering whole neighborhoods), or small (e.g., limited to a single street block), very fine-grained spatial heterogeneity may play a key role in the clustering of gang crime.

Superimposed on these structural generators of crime are gang social dynamics that operate both within and between neighborhoods. Gang crimes are often retaliatory in nature (Decker 1996; Klein and Maxon 2006). Interactions between gangs that threaten geographic territory or gang reputation can escalate to a shooting, while a shooting or homicide often demands retribution in kind (Hughes and Short 2005; Jacobs and Wright 2006), driving a sequence of tit-for-tat reciprocal attacks (Bjerregaard and Lizotte 1995; Rosenfeld et al. 1999; Howell 2011). Retaliatory aggression may be linked to a deep-seated moral instinct (Daly and Wilson 1988), street codes that demand quick and decisive retribution (Decker 1996; Anderson 1999; Jacobs and Wright 2006), commitment to delinquent peers (Esbensen et al. 1993), and social networks that promote the spread of rumors (Hughes and Short 2005; Green, Horel, and Papachristos 2017).

It has recently been argued that Hawkes process models offer a concise statement of these two important effects (Mohler et al. 2011; Mohler 2014; Reinhart and Greenhouse 2018). In Hawkes models clustering is attributed to both causal and noncausal mechanisms: an event occurring in a particular

CONTACT Frederic Paik Schoenberg *frederic@g.ucla.edu* Department of Statistics, University of California, Los Angeles, CA 90095.

location increases the likelihood that other events will occur in its vicinity in the near future (causal), while some events occur exogenously due to a chronic, spatially inhomogeneous background component (noncausal). While useful at a theoretical level, a Hawkes process model also has important practical implications. Specifically, causal clustering suggests an opportunity to prevent crime by disrupting the underlying, local dynamical processes (Mohler et al. 2011; Mohler et al. 2015; Green, Horel, and Papachristos 2017). One such program with the goal to disrupt retaliatory gang violence has been implemented in Los Angeles since 2009.

The City of Los Angeles Mayor's Office of Gang Reduction and Youth Development (GRYD) is a city-funded comprehensive gang prevention and intervention program to reduce the likelihood of retaliations when violent gang crimes do occur (Skogan et al. 2009; Cespedes and Herz 2011; Tremblay et al. 2020). Our research is focused on this latter effort, called the GRYD Incident Response (IR) program. In brief, GRYD IR tasks civilian community intervention workers (CIWs) with responding quickly to violent gang incidents as they occur. In the field, CIWs work to control rumors, proactively diffuse tensions and deliver services to victims and their families. CIWs coordinate with regional program coordinators (RPCs) in the GRYD Office, who remain in communication with the LAPD about gang suppression and investigative activities (Tremblay et al. 2020). However, due to both resource limitations and discretion in the reporting process, GRYD IR is usually only deployed for a subset of reported gang-related violent crimes. A central question of interest therefore is whether the GRYD IR Program is effective and how much, if at all, it reduces retaliatory crime. If it is effective, then the argument can be made that efforts should be made to ensure that GRYD IR is deployed more widely to cover more of the gang-related violent crime occurring on the streets.

We suspect that the effectiveness of GRYD IR is closely tied to how much causal triggering is present in gang related crime. In general, the more causal triggering there is, the greater the opportunity to disrupt retaliations with rapid response. Thus discriminating between causal clustering and inhomogeneity in the gang-related violent crime data is central to our study of the effect of GRYD IR. This is a difficult problem arising frequently in the study of spatial–temporal point processes (see Chp. 9.6 of Diggle 2014). Indeed, in fitting spatial–temporal Hawkes processes, it is often inadvisable to use identical data to estimate parameters governing the background rate (inhomogeneity) and the triggering density (causal clustering), as these parameters may not be jointly identifiable. For this reason, Ogata (1998, sec. 4.2) suggested modeling the background rate for earthquakes using only the largest magnitude events in the catalog, for instance. For crimes, there is no such natural partitioning of events based on magnitude to guide the estimation of the background rate. In the case of reported gang-related crimes in South Los Angeles, we attempt to model the inhomogeneity nonparametrically using generalized additive modeling. Specifically, we model spatially varying crime rates given observable covariates linked to social and economic variations in the urban environment. With these factors accounted for, additional clustering observed in the data may be more reasonably attributed to retaliatory criminal behavior.

Spatially varying covariates have previously been used to model the background spatial inhomogeneity in Hawkes processes by Reinhart and Greenhouse (2018). However, simple parametric forms of the background rate were required for tractable analytic maximization steps in the EM-algorithm. We propose an iterative procedure that allows for use of any supervised learning method using covariates. For the first time, we compare the fit and predictive performance between using covariates for estimating the background rate of crime and the more common method of kernel smoothing over all crimes as in Mohler (2011, 2014). Methodological choices in bandwidth selection for kernel smoothing are examined. We demonstrate through our results that kernel smoothing over all reported crimes in the dataset can lead to confounded estimates of background inhomogeneity and causal clustering/retaliation. We assess how this affects the estimated amount of retaliation and its space-time decay rates.

One challenge in evaluating the efficacy of the GRYD IR Program is that its response to violent events are not randomize; one cannot legitimately refuse to provide services to a victim's family for the purposes of experimental purity. Rather, GRYD CIWs use their specialized knowledge of local gang dynamics and intervene in areas believed to be more prone to retaliations (Tremblay et al. 2020). As a result, excitation/retaliation rates are naturally biased upward for crimes exposed to the GRYD IR Program compared to untreated crimes, even after controlling for spatial inhomogeneity of the background rate of reported crimes. We propose a simple method inspired by point process thinning (Lewis and Shedler 1979) to sample untreated crimes so that they are distributed similarly in space and time to the crimes exposed to GRYD IR Program efforts. This allows an approximate treatment vs. control comparison of the GRYD IR Program. The results reveal that the GRYD IR Program is effective, reducing rates of reported retaliations by an estimated 18.3% over two different spatial scales and reducing such retaliation rates within a spatial scale by an estimated 14.2% according to the fitted model.

The rest of this article is organized as follows. A brief description of the data is provided in Section 2. Our proposed iterative method to incorporate nonparametric regression for the background rate of a Hawkes process while simultaneously estimating the triggering component is explained in Section 3 along with a description of methods for covariate selection, out-of-sample prediction log likelihoods, residual analysis, and sampling controls to compare with the GRYD IR Program. The results and a discussion are given in Sections 4 and 5, respectively.

2. Data

Reports of gang-related violent crimes from 2014 to 2017 were collected by the Los Angeles Police Department (LAPD) and the City of Los Angeles Mayor's Office of Gang Reduction Youth Development (GRYD). GRYD operates in 23 zones throughout Los Angeles (GRYD 2017 Evaluation Report). We focus on ten GRYD Zones in South Los Angeles that represent 7% of the

total land of Los Angeles (1302 km^2) and approximately 15.5% of the total population (3.9 million). This region accounted for 44.7% of all officially reported gang-related violent crimes in Los Angeles between the beginning of 2014 to the end of 2017. Of the 3627 reported crimes in our study, 1100 were exposed to GRYD IR Program efforts. Multiple records, representing multiple victims of an identical crime, are collapsed to one report. LAPD officers record the locations of crimes at the level of street addresses or intersections. For privacy reasons, latitudes and longitudes are uniformly randomized over a 15 m interval centered at each reported crime.

Demographic and socio-economic covariates are compiled at the census block level, which is currently the highest resolution published by the U.S. Census. These data are obtained from the American Community Survey, publicly available at *https://factfinder.census.gov*. We use the same eight variables used in Kyriacou et al. (1999), who previously studied the relationship between socioeconomic factors and gang violence in the city of Los Angeles: per capita income, unemployment, percentage with high school degree, percentage of single-parent families, percentage of males, percentage under 20 years of age, percentage black, and percentage Hispanic. We also include population density as a potential covariate since in point process modeling our outcome variable is the reported crime rate per unit of time and space, while Kyriacou et al. (1999) studied reported crimes per 100,000 people. Ideally, we would also include covariates directly related the geography of gangs and the law enforcement response to gang violence. For example, gang territorial boundaries appear to be important generators of gang violence (Brantingham et al. 2012). Geographically based restraining orders against gangs, called criminal gang injunctions, are likely to have the opposite effect (Bichler et al. 2019). These unmeasured covariates likely impact spatial heterogeneity in the background rate as well as where GRYD IR chooses to intervene.

Latitudes and longitudes are geocoded to census block identifiers using *https://geocoding.geo.census.gov*. Data on the land mass of each census block uses the latest publicly available source, the 2010 Census of Population and Housing (U.S. Census 2012). Our study region consists of 410 census blocks. The average size of each block is approximately 0.22 km^2, and the median number of reported crimes in each census block over the 4 years of observation is 7.

3. Methods

We note at the outset that most of our inferences are based on the particular formulation of the Hawkes model in Equation (11) below, with background rate estimated using Equation (10). We will later refer to this as model (IV). For comparison and to motivate this model and estimation procedure, we also consider various alternatives described in what follows.

3.1. Overview of Hawkes Models

We consider crime data as a marked space-time point process $\{(t_i, x_i, y_i, m_i) : i = 1 \ldots .N\}$, representing the times, locations, and mark information associated with gang-related violent crimes. In our study, the marks recorded are indicators of whether crimes were exposed to GRYD IR Program efforts or not. The rate of occurrences of points with any mark is characterized via the conditional intensity,

$$\lambda(t,x,y|\mathcal{H}_t) = \lim_{\Delta t,\Delta x,\Delta y \downarrow 0} \frac{E[N\big((t,t+\Delta t)\times(x,x+\Delta x)\times(y,y+\Delta y)\big)|\mathcal{H}_t]}{\Delta t \Delta x \Delta y}.$$

Daley and Vere-Jones (2003) showed that all finite dimensional distributions of a simple point process (i.e., a process with almost surely no coincident points) are uniquely determined by its conditional intensity. In the study region S, where $(x,y) \in S \subset \mathbb{R}^2$ and $t \in [0,T)$, $N(A)$ counts the random number of occurrences over the set $A \subset S\times[0,T)$ given the history $\mathcal{H}_t$ of all points occurring prior to time t. The conditional intensity λ can be interpreted as the instantaneous expected rate of a reported crime per volume of space-time.

When the data features clustering over space and time, it is common to model λ using self-exciting point process models, where each event triggers further events by temporarily and locally boosting the conditional intensity λ. A Hawkes model is a particular formulation for a self-exciting process that has been successfully used to model the spread of invasive species (Balderama et al. 2012), epidemic disease spread (Meyer et al. 2012), earthquakes (Ogata 1998), financial transactions (Bauwens and Hautsch 2009), neuron activity (Johnson 1996), reported burglaries (Mohler et al. 2011), E-mail networks (Fox et al. 2016) and terrorist attacks. The Hawkes model can be specified as

$$\lambda(x,y,t) = \mu(x,y,t) + \sum_{i:t_i<t} \kappa(i) g(x-x_i, y-y_i, t-t_i), \quad (1)$$

where the triggering density g governs the spatial–temporal distance of triggered events from their antecedent events and is usually modeled to decay with distance from the origin over time and space. Previous authors have typically modeled the background rate μ as spatially varying but constant in time. The spatial–temporal distribution of triggered events is commonly assumed to be separable, meaning $g(x,y,t) = g_1(x,y)g_2(t)$. We scale g_1 and g_2 to be densities as suggested in Schoenberg (2013), which implies the productivity $\kappa(i) > 0$ represents the expected number of events triggered directly by event i, and we let $\kappa(i) = \kappa_1$ if crime i is exposed to GRYD IR Program efforts and κ_2 otherwise, where κ_1 and κ_2 are scalar parameters to be estimated. In the absence of the GRYD IR Program, any particular crime is expected to be an ancestor to $\kappa_2 + \kappa_2^2 + \kappa_2^3 + \ldots = \frac{1}{1-\kappa_2} - 1$ total retaliatory crimes. Productivities must be nonnegative and are constrained to be less than 1 in order for the process to be stable.

Our parametric specification of g follows: Mohler (2014) and Reinhart and Greenhouse (2018). Consider $g_2(t-t_i) = \omega e^{-\omega(t-t_i)}$ where ω controls the decay rate of triggering and $1/\omega$ is the average response time. The spatial distribution of triggered crimes, g_1, is assumed to be isotropic, that is, $g_1(x,y) = h(r)$ in polar coordinates where $r = \sqrt{x^2+y^2}$ and $h(r,\theta) = h(r)$. Given $\int\int_A g_1(x,y)dA = \int_0^\infty \int_0^{2\pi} g_1(r\cos\theta, r\sin\theta) r dr d\theta = \int_0^\infty \int_0^{2\pi} h(r) r dr d\theta = 1$, set $g_1(x,y) = h(r)/2\pi r$ so that $h(r)$

is the probability density function for the distance r between a reported crime and any reported retaliation it triggers. The function h of distance may be any density on the real half-line, such as the truncated Gaussian centered at zero,

$$h(r) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \exp(-\frac{r^2}{2\sigma^2}). \tag{2}$$

Given a parameterized model for $\lambda(t, x, y)$, the log-likelihood of an observed sequence of N reported crimes over an interval $[0, T]$ in region S is (Daley and Vere-Jones 2003)

$$l(\Theta) = \sum_{i=1}^{N} log(\lambda(t_i, x_i, y_i|\mathcal{H}_t)) - \int_0^T \int \int_S \lambda(t, x, y|\mathcal{H}_t) dx dy dt. \tag{3}$$

Ogata (1978) showed that under general conditions the maximum likelihood estimate (MLE) is consistent, asymptotically unbiased and efficient, with standard errors estimated using the square root of the diagonal elements of the inverse Hessian of the loglikelihood.

3.2. Background Rate Estimation

Accurate estimation of the background rate μ is critical for accurately estimating the parameters in (1), and is especially important for the discrimination between spatial–temporal inhomogeneity and causal clustering. The key idea is that after properly accounting for background inhomogeneity, the remaining clustering can be reliably attributed to retaliation. Background rate estimation is thus the subject of careful study here, and we consider two different estimates for the background rate $\mu(x, y, t)$ in Equation (1).

3.2.1. Kernel Smoothing With Stochastic Declustering

The background process $\mu(x, y, t)$ represents the expected rate of reported crimes in the absence of retaliation. In applying such Hawkes models to reported crimes, Mohler et al. (2014) proposed estimating μ using a time-invariant smoother over all events, using mark dependent weights $\beta(i)$, for example:

$$\begin{aligned}\hat{\mu}(x, y, t) &= \mu(x, y) \\ &= \sum_{i=1}^{N} \frac{\beta(i)}{2\pi\eta^2 T} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2}\big).\end{aligned} \tag{4}$$

Here T is the length of the observation period and N is the total number of observed points. The smoothing bandwidth η is typically constrained to be identical to the triggering bandwidth σ in Equation (2) in order to achieve numerical stability in optimization and identifiability of the parameters (Mohler et al. 2014; Yuan et al. 2019). In our study, we choose not to impose these constraints on the smoothing nor triggering bandwidths and instead estimate them separately.

Estimating the background rate by smoothing over all points with equal weights, regardless of whether each point is more likely to be a background point or a retaliation, may lead to mis-attribution of triggering as background and vice versa. In addition, in the presence of intense spatial clustering, a fixed bandwidth may yield noisy estimates in sparse areas and oversmoothed estimates between dense and sparse areas (Zhuang et al. 2002; Yuan et al. 2019). Thus, as an alternative to Equation (4), we obtain a weighted, variable-bandwidth, stochastically declustered background rate estimate

$$\hat{\mu}(x, y, t) = \sum_{i:(x_i,y_i)\neq(x,y)}^{N} \frac{w_i}{2\pi d_i^2 T} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2d_i^2}\big). \tag{5}$$

In (5), the d_i is the radius of the smallest disk centered at point (x_i, y_i, t_i) that includes at least n_p other events; each d_i is constrained to be at least some minimal value ϵ representing the approximate size of errors in location estimates. The weight w_i is the estimated probability, according to the fitted model (1), that crime i is a background event, and is computed as

$$w_i = \frac{\hat{\mu}(t_i, x_i, y_i)}{\hat{\lambda}(t_i, x_i, y_i)}. \tag{6}$$

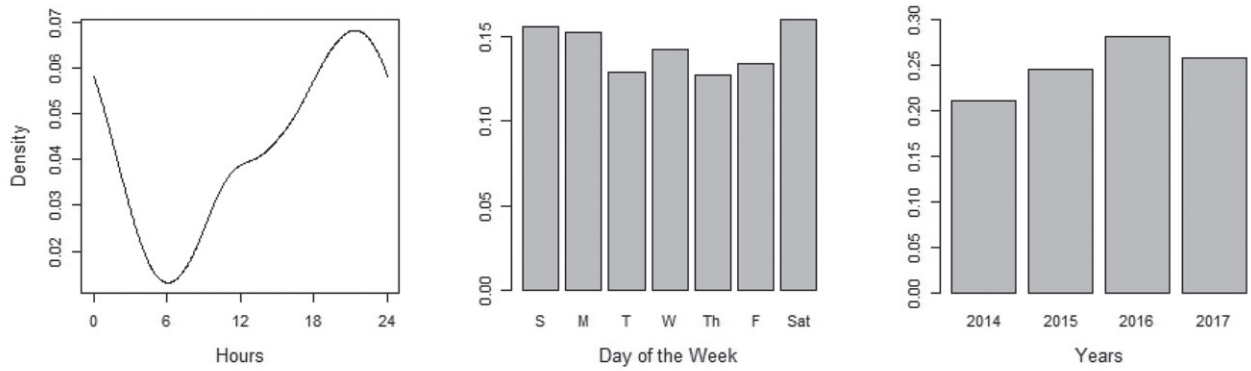
The algorithm, originally proposed in Zhuang et al. (2002), works by iteratively estimating the triggering parameters and updating estimates of $\{w_i\}_{i=1}^N$.

3.2.2. Temporal Variation in Background Rate

According to Equation (5), the temporal density of background events is stationary $(1/T)$, and the spatial distribution does not change over time. However, given the pronounced temporal fluctuations in reported gang-related violent crimes shown in Figure 1, a constant temporal background rate is unrealistic and may lead to inflated estimates of productivity (i.e. misattribution of background crimes as triggered events). Therefore, as in Fox et al. (2016), we allow the temporal distribution of background crimes to be nonstationary

$$\begin{aligned}&\hat{\mu}(x, y, t) \\ &= \nu(t) \cdot \sum_{i:(x_i,y_i)\neq(x,y)}^{N} \frac{w_i}{2\pi d_i^2} \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2d_i^2}\right).\end{aligned} \tag{7}$$

Estimating periodic components of $\nu(t)$ while incorporating weights such as w_i would be optimal, especially in the presence of strong clustering (Zhuang and Mateu 2019). But for simplicity and because of the added computational costs, we opt to estimate the periodic components of $\nu(t)$ once using all the data with equal weights as in Fox et al. (2016), obtaining the estimate $\hat{\nu}(t) = c\hat{f}_1(h(t))\hat{f}_2(d(t))\hat{f}_3(y(t))$ where $h(t) \in [0, 24)$, $d(t) \in \{0, \ldots, 6\}$ and $y(t) \in \{0, 1, 2, 3\}$ represent the hour, day of the week, and year, respectively, corresponding to time t. We estimate the daily cycle $\hat{f}_1(h(t))$ via kernel smoothing the times of the days of the reported events; the repeating weekly cycle $\hat{f}_2$, and year-to-year variations $\hat{f}_3$ are simply estimated via histogram estimators representing the proportion of crimes occurring on certain days d and years y, that is, $\hat{f}_2 = \sum_{i=1}^N I(d_i = d(t))/N$ and $\hat{f}_3 = \sum_{i=1}^N I(y_i = y(t))/N$. The estimated daily, weekly, and multi-year components of $\nu(t)$ are shown in Figure 1. No substantial annual cycle was observed for reported gang-related violent crimes in this study period. The constant c is chosen to ensure that $\int_0^T \hat{\nu}(t)dt = 1$ and is accurately approximated by a Riemann sum.



(Left to right): Daily cycle, weekly cycle, and year-to-year variations in crime used for components of $\hat{v}(t)$.

Figure 1. Temporal distribution of all gang crimes. (Left to right): Daily cycle, weekly cycle, and year-to-year variations in crime used for components of $\hat{v}(t)$.

3.2.3. Generalized Additive Modeling of Covariates

Rather than estimating the background rate μ by smoothing over observed events, an alternative way to estimate the background rate is to use information on the spatial heterogeneity in demographic and socio-economic covariates. Reinhart and Greenhouse (2018) used covariates to model the background rate of a Hawkes process for reported burglaries in Pittsburgh with the parametric form

$$\mu(x,y) = \exp\left(v(x,y)'\gamma\right), \tag{8}$$

where γ is a vector of coefficients to estimate and $v(x,y)$ is a vector of covariates measured at location (x,y). In our application, following covariate selection, $v(x,y)$ represents the covariates per capita income, population density, male percentage, single parent rate and unemployment rate for the census block containing the location (x,y).

Instead of requiring the background rate to follow an exponential or some other particular functional form, we propose allowing $\hat{\mu}(x,y) = \hat{f}(v(x,y))$, where $\hat{f}$ is estimated nonparametrically, for example, by generalized additive modeling (GAM),

$$l(\mu(x,y)) = b_0 + \sum_{j=1}^{p} h_j(v_j(x,y)) \tag{9}$$

where p and l represent the number of covariates and a link function, respectively. Here, the additive components h are smoothing splines. A simplistic approach would be to first estimate f by nonparametric regression of the observed crimes on the covariates $v(x,y)$. The problem with such an approach, however, is that both background and triggered crimes would be used in estimating f, though in principle only background crimes should be used.

We propose the following iterative solution. Suppose the study region is divided into 410 census blocks $\{B_k\}_{k=1}^{410}$, where B_k is a set of indices of crimes belonging to the kth census block. Given a fitted model, we estimate the background crime rate of census block k as $\sum_{i\in B_k} w_i/a_k$ where w_i is defined in Equation (6) and a_k is the area in km^2. We propose to estimate f via nonparametric regression of $\sum_{i\in B_k} w_i/a_k$ on covariates in the following algorithm:

Algorithm 1

1. Initialize $m \leftarrow 0$, $w_i^{(0)} \leftarrow Unif(0,1)$.
2. Fit
$$\mu^{(m)}(x,y) = \hat{f}(v(x,y)) \tag{10}$$
where f is estimated by nonparametric regression of $\sum_{i\in B_k} w_i/a_k$ on census block level covariates.
3. Using maximum likelihood estimation, fit
$$\lambda(x,y,t) = c\cdot v(t)\cdot \mu^{(m)}(x,y) + \sum_{i:t_i<t} \kappa(i) g_1(x-x_i, y-y_i) g_2(t-t_i)$$
where $v(t)$, κ, g_1, g_2 are as defined previously and c is an estimated parameter governing the proportion of events attributed to the background rate.
4. Calculate w_i from (6) and update $w_i^{(m+1)} \leftarrow w_i$ for $i = 1\ldots N$.
5. If $\max_i |w_i^{(m+1)} - w_i^{(m)}| > \epsilon$, where ϵ is a small positive number, then update $m \leftarrow m+1$ and go to step (2). Otherwise stop.

The function f can be estimated using any nonparametric regression method in Step 2, and we estimate f via GAM in the application here for flexibility and interpretability.

3.3. Near and Far-Field Triggering

The smoothness of the estimated background rate using spatial covariates depends not only on the variability of the covariates across different spatial units, but also on the resolution of the spatial units themselves. In practice, the observed spatial covariates are piecewise constant. Even when using the highest available spatial resolution kept by the U.S. Census where the average size of a census block is equivalent to an area of a 470 by 470 m square containing only a median of 7 crimes over 4 years, clustering in the reported crimes is still evident within the scale of a census block, especially in census blocks with high 4-year crime counts.

Therefore, when estimating models with background rates using covariates, we allow different parameters for the near and

far-field triggering, using the following modification to the total triggering rate:

$$\lambda(x,y,t) = \nu(t)\mu(x,y) + \sum_{i:t_i<t} \kappa_{\text{near}}(i)\frac{h_1(r)}{2\pi r}\omega_1 e^{-\omega_1(t-t_i)} + \sum_{\substack{i:t_i<t,\\ r\geq d}} \kappa_{\text{far}}(i)\frac{h_2(r)}{2\pi r}\omega_2 e^{-\omega_2(t-t_i)}, \quad (11)$$

where h_1 is a half-normal density over the positive real line with triggering bandwidth σ_1 and h_2 is a half-normal density centered at d (d km's away from the originating reported crime) with support $[d,\infty)$ and with triggering bandwidth σ_2. We estimate d using the median distance from the observed crimes to their nearest neighbors in different census blocks (130m). We use the notation $\kappa_{\text{near}}(i) = \kappa_1$ and $\kappa_{\text{far}}(i) = \kappa_3$ if crime i is associated with the GRYD IR Program, and otherwise $\kappa_{\text{near}}(i) = \kappa_2$ and $\kappa_{\text{far}}(i) = \kappa_4$, where $\kappa_1, \kappa_2, \kappa_3$ and κ_4 are scalar parameters to be estimated by maximum likelihood.

3.4. Integral Approximation

The first term of the log-likelihood in Equation (3) is straightforward to compute while the integral term must be numerically approximated, which can be a substantial computational challenge (Harte 2013). In all models investigated in this article, we use the analytic integral approximation in Schoenberg (2013), and find parameter estimates by MLE using the quasi-Newton method developed by Broyden, Fletcher, Goldfarb, and Shanno (1970). The integral approximation is based on interchanging the order of the integral in Equation (3) and the sum in Equation (1); this approximation is perfect if all triggering is confined to the spatial–temporal region being observed (Schoenberg 2013).

For example, the approximate log-likehood for the model with background rate (7) is

$$l(\kappa_1,\kappa_2,\beta_1,\beta_2,\omega,\sigma,\eta) = \sum_{i=1}^{N}\log[\lambda(x_i,y_i,t_i)] - \sum_{i=1}^{N}\left[w_i+\kappa(i)\right]. \quad (12)$$

As a baseline for comparison, we also consider a model with a spatially constant background rate

$$\mu(x,y,t) = c\cdot\nu(t), \quad (13)$$

whose log-likelihood is

$$l(\kappa_1,\kappa_2,c,\omega,\sigma) = \sum_{i=1}^{N}\log[\lambda(x_i,y_i,t_i)] - c\cdot|S| - \sum_{i=1}^{N}\kappa(i), \quad (14)$$

where $|S|$ is the area of the observation region being studied.

3.5. Sampling Non-GRYD Crimes as Controls

After attempting as carefully as possible to distinguish inhomogeneity from causal clustering, we seek to evaluate whether the GRYD IR Program effectively reduces retaliations by comparing the estimated productivity for reported crimes with exposure to violence interruption with the productivity of reported crimes without such exposure. If the GRYD IR Program violence interruption efforts were randomly assigned over space and time, this comparison would be straightforward. However, the decision by the GRYD IR Program when and whether to intervene is made based on attempts to maximize the effect of violence interruption with limited resources, using specialized knowledge of local gang dynamics in an attempt to intervene following crimes believed most likely to spark retaliation. Thus, the reported crimes associated with the GRYD IR Program are more likely to occur in areas of high reported gang-related activity, for instance, and thus to occur in areas of higher subsequent reported crime incidence despite the possible effectiveness of the violence interruption.

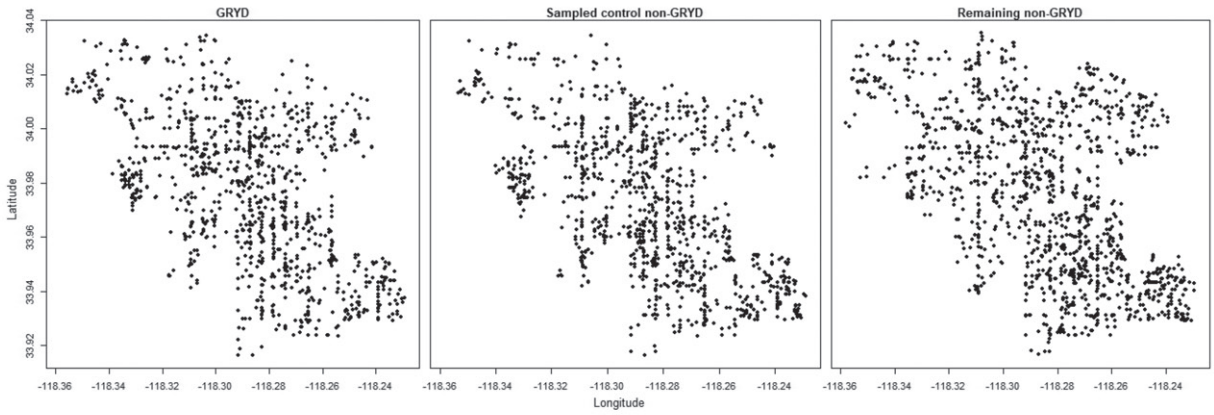
Instead of using just two marks for the GRYD IR Program and non-GRYD, we introduce a third mark which is sampled from non-GRYD crimes that are spatially temporally distributed similarly to GRYD IR Program crimes. Our aim is to obtain a sample of non-GRYD crimes with similar spatial–temporal characteristics as the GRYD IR Program crimes to isolate the effect of the GRYD IR Program.

We suppose that reported crimes exposed to GRYD IR Program efforts occur with an intensity varying over space, hour of the day and day of the week, given by $P(x,y,h,d) = P_1(x,y)P_2(h)P_3(d)$, and that non-GRYD crimes follow $Q(x,y,h,d) = Q_1(x,y)Q_2(h)Q_3(d)$. Spatial distributions P_1 and Q_1 are estimated using kernel density estimation with Gaussian kernels and default bandwidths along each dimension given by Sheather and Jones (1991). The 24 h cycles P_2, Q_2 and day-to-day weekly cycles P_3, Q_3 are estimated in the same manner as the components of $\nu(t)$ in Equation (7) and displayed in Figures 3 and 4. We then sample the same number of non-GRYD crimes as there are reported crimes for the GRYD IR Program, without replacement, using sampling weights ν_i given by

$$\nu_i = \frac{\hat{P}_1(x,y)\hat{P}_2(h)\hat{P}_3(d)}{\hat{Q}_1(x,y)\hat{Q}_2(h)\hat{Q}_3(d)}. \quad (15)$$

This results in a sample of non-GRYD crimes whose spatial–temporal distribution is similar to that of GRYD IR Program crimes, and this sampling can be performed repeatedly. The results of one such sampling are shown in Figure 2. This sampling is repeated 50 times, and the associated productivities are estimated by maximum likelihood each time. We then compare the average estimated productivity of GRYD IR Program crimes with the average estimated productivity of the sampled control crimes to evaluate the efficacy of the violence interruption efforts.

We note that there may be potential confounding that could lead to an underestimation (or overestimation) of GRYD effectiveness. For example, the "specialized" information used by GRYD IR to decide whether to intervene or may be unrelated to the exact time and location of a reported crime. This would result in an underestimation of the program's effectiveness. However, evidence suggests that gang social processes are tightly coupled to the time and locations where gang crimes occur (Tita and Griffiths 2005; Tita and Ridgeway 2007; Brantingham et al. 2012; Huebner et al. 2016; Valasik and Tita 2018). Thus, the outcome is likely to be similar whether GRYD uses the location and time of an event to make a decision, or more detailed information.



(Left to Right): Locations of GRYD IR Program crime events, one sample of non-GRYD crimes and remaining unsampled non-GRYD crimes. The union of crimes in these three panels are used to estimate models in Table 1.

Figure 2. Spatial distribution of marks. (Left to Right): Locations of GRYD IR Program crime events, one sample of non-GRYD crimes and remaining unsampled non-GRYD crimes. The union of crimes in these three panels are used to estimate models in Table 1.

3.6. Evaluation Methods

Four types of models and background rate estimation methods are investigated: (I) constant background model in Equation (13), (II) kernel smoothed background model in Equation (7), (III) covariate background model in (10), all with triggering as in Equation (2), and (IV) covariate background model (10) with near and far-field triggering as in Equation (11). To assess the efficacy of the GRYD IR Program, we also evaluate the fit of model (IV) with sampled non-GRYD control marks as detailed in Section 3.5. Log-likelihood scores are used to compare the goodness of fit on training data from 1/1/14 to 12/31/16, the same data used in the fitting. To investigate possible over-fitting, out-of-sample log-likelihood scores for each model are also computed, using data from 1/1/14 to 12/31/16 in the fitting and data from 1/1/17 to 12/31/17 for evaluation. In the rare instance of negative predicted crime rates, the prediction is coerced to zero. Superthinned point process residuals, described below, are used to examine the model forecasts from 1/1/17 to 12/31/17.

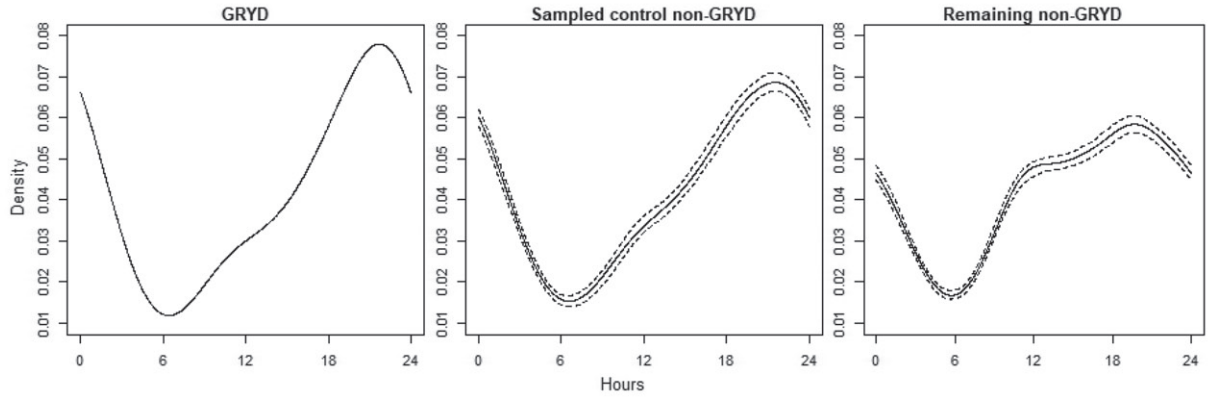
Superthinning involves both thinning the original data points and superposing a new set of points, and is an effective way to evaluate the fit of a point process model (Bray and Schoenberg 2013; Clements et al. 2013). The observations are first thinned, that is, each observation is randomly kept with probability $\min\left\{b/\hat{\lambda}(t), 1\right\}$, where b is a tuning parameter chosen by the user. Next, a Poisson process with constant rate b is generated over the space-time observation region, each point of this Poisson process is independently kept with probability $\max\left\{(b - \hat{\lambda}(t))/b, 0\right\}$, and these remaining points are superposed, that is, added to the collection of thinned observations. The resulting residual process should be a homogeneous Poisson process with rate b if and only if the modeled conditional rate is correct (Clements et al. 2013), and thus departures from homogeneity in the residuals can be detected as evidence of lack of fit of the model. Sparsity of points in the superthinned residuals corresponds to areas where the model over-predicted, whereas clustering in the residual points indicates areas where the model under-predicted the number of observed events. For all models considered here, we use identical candidate points to be superposed, so that our comparisons are not impacted by random fluctuations in the superposition step, and we use the mean number of observed points per unit of space-time as the default estimate of b, as suggested by Clements et al. (2013).

4. Results

4.1. Spatial–Temporal and Covariate Effects

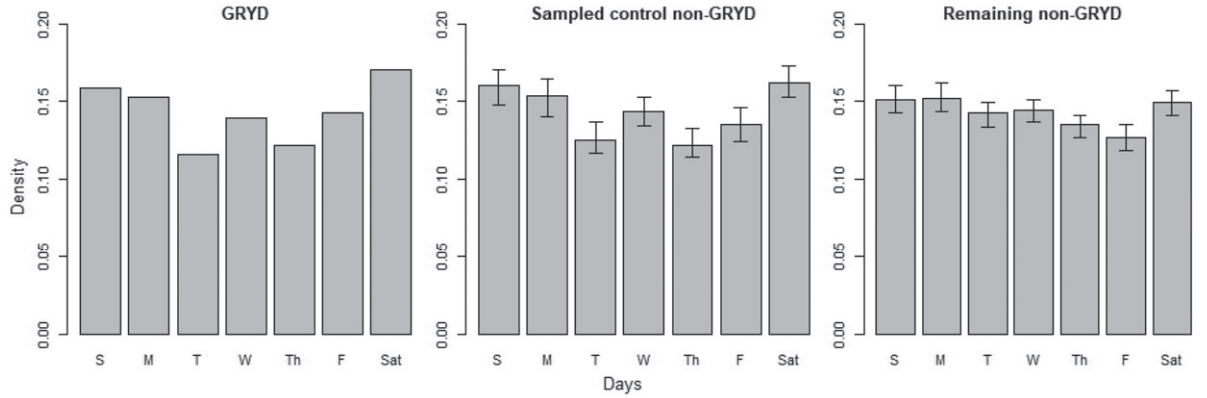
The spatial distribution of the GRYD IR Program events, non-GRYD events sampled according to Equation (15), and other non-GRYD events are shown in Figure 2, and the temporal distributions of the three classes of events are shown in Figures 3 and 4. If GRYD IR Program events were assigned purely at random, then the point patterns shown in the three panels of Figures 2–4 would be distributed identically. As expected, however, the GRYD IR Program events are substantially more clustered than the non-GRYD events depicted in the rightmost panel of Figure 2. The temporal distributions of GRYD IR Program and non-GRYD events in Figures 3 and 4 show the modest deviations.

Variable selection is done once, using a GAM regression of total crimes on the covariates listed in Section 2. The following five variables are selected by the stepwise selection procedure in the *R* package gam (Hastie 2018) in order to reduce prediction error and avoid overfitting: income per capita, unemployment, population density, percent male and percent single parent families. The estimated additive predictors for the GAM regression background rate are shown in Figure 5. Population density and income per capita account appear to be the most important predictors for this dataset according to the fitted GAM model for the background rate, with higher estimated background rates of reported gang-related violent crimes in areas with high population density and in areas with low income per capita. A slight increase in the estimated background rate of reported crimes is associated with areas where the proportion of males is lower and the percentage of single parent families is higher, though these effects appear to be rather minimal.



(Left to Right): Kernel density of GRYD IR Program crime events, average kernel density of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Dotted: 5th and 95th percentile of 50 estimated kernel densities.

Figure 3. Temporal distribution of marks (hourly). (Left to Right): Kernel density of GRYD IR Program crime events, average kernel density of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Dotted: 5th and 95th percentile of 50 estimated kernel densities.



Proportion of occurences each day. (Left to Right): GRYD IR Program crime events, average of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Whiskers: 5th and 95th percentile proportion of 50 samples.

Figure 4. Temporal distribution of marks (weekly). Proportion of occurrences each day. (Left to Right): GRYD IR Program crime events, average of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Whiskers: 5th and 95th percentile proportion of 50 samples.

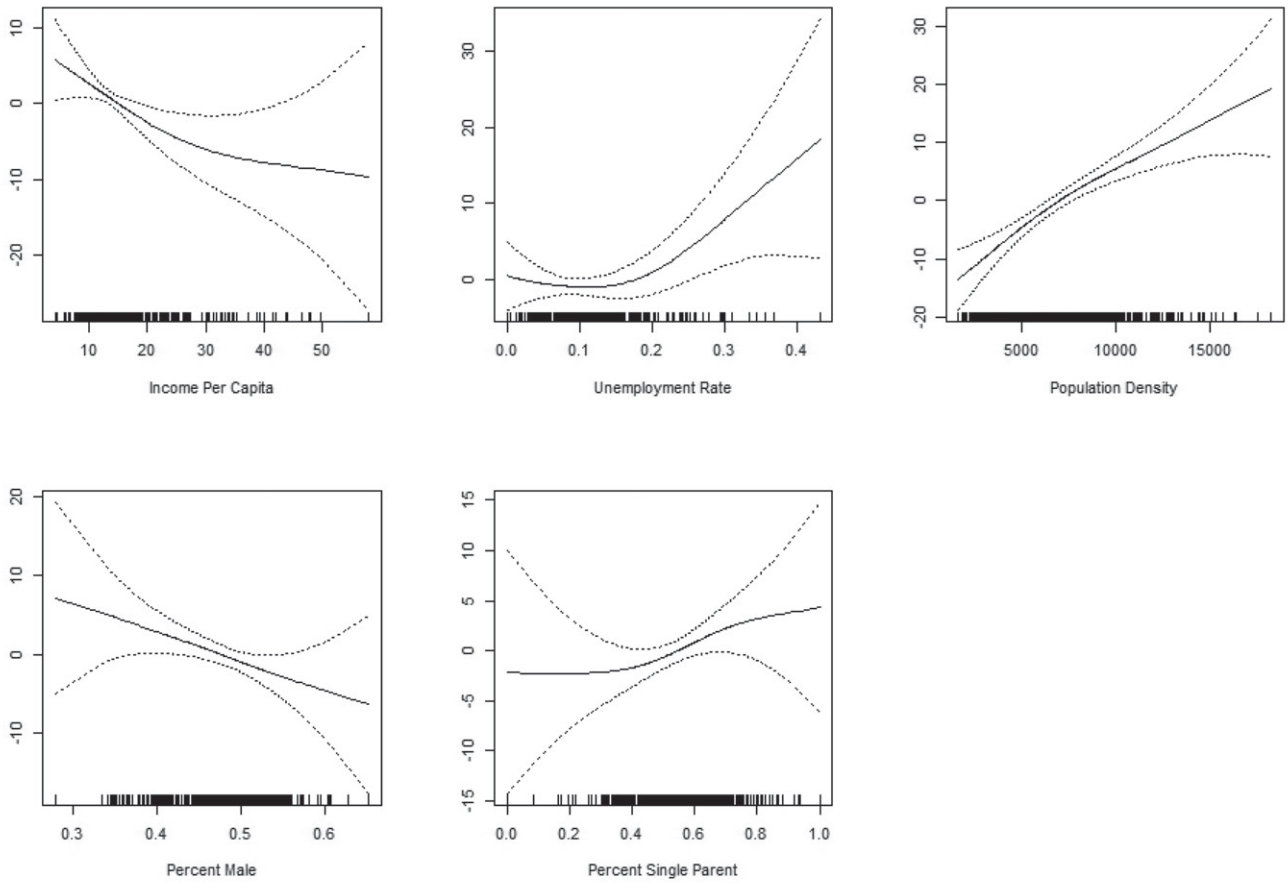
The estimated spatial background rates (excluding the non-stationary component $\nu(t)$) for models (II) and (IV) are depicted in Figure 6. With all three models, the estimated background rates indicate substantial inhomogeneity. Certain hot spots are noticeable, such as near Hyde Park (−118.335°, 33.98°) and Crenshaw (−118.35°, 34.02°) as well as along Normandie Avenue (longitude −118.3°). The eastern half of the study region generally appears to have a higher background rate.

4.2. Model Fit and Estimates

Parameter estimates and log-likelihood scores for models fit using data from 2014 to 2016 are reported in Table 1. Comparison of the fit of models (I), (II), and (III) reveals that all three have serious inadequacies in untangling causal clustering from inhomogeneity. Model (I) fits worse than the others as indicated by its much lower log-likelihood in-sample (for 2014–2016). Model (I) also fits the worst on the out-of-sample testing data from 2017. Model (IV) has considerably higher log-likelihood than the other models, indicating superior fit to the in-sample, 2014–2016 data. In contrast to model (II), models (III) and (IV) have a higher log-likelihood while their estimated background rates attribute more reported crimes to triggering (18% and 22% respectively).

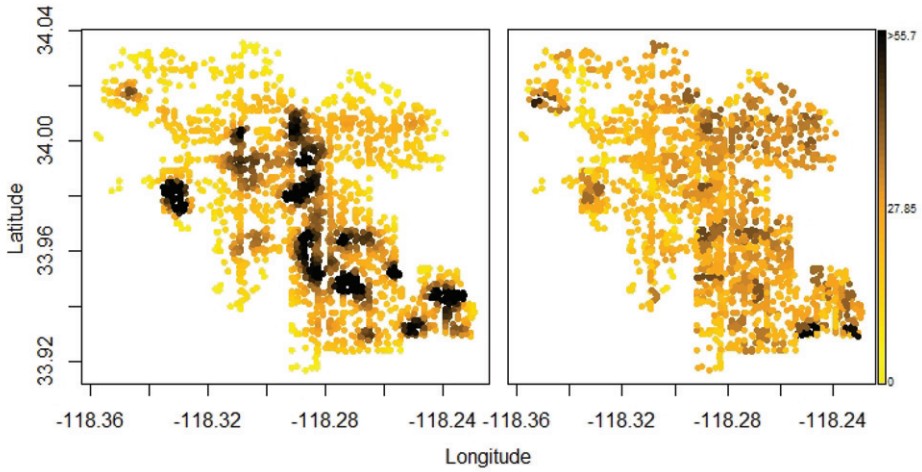
The variable bandwidth estimate (5) used in model (II) appears less smooth than model (IV) in Figure 6, and as a result attributes only 16% of reported crimes to triggering. The background rate in model (II) uses $n_p = 15$, which is the minimum recommended number by Fox et al. (2016). The 25th, 50th, and 75th percentile of the varying bandwidths are respectively 275, 351, and 437 m, and is comparable to the bandwidth selected using Sheather and Jones (1991).

In Table 2, the estimated spatial triggering bandwidth σ in models (I), (II), and (III) are all very local, between 12 and 15 m, and the respective estimates of the temporal decay ω, are consistently small with a median time to response of almost 180 days. This would suggest that triggered crimes are near-repeat



Dotted=95% confidence intervals. The y-axes are the individual additive contributions of each covariate towards the output which is in units of background crimes per square km. All covariates are estimated with 2 degrees of freedom.

Figure 5. Estimated additive predictors of GAM background. Dotted=95% confidence intervals. The y-axes are the individual additive contributions of each covariate toward the output which is in units of background crimes per square km. All covariates are estimated with 2 degrees of freedom.



(Left, Right): Kernel smoothed background rate with variable bandwidth and weights of model (II), generalized additive model (GAM) background rate of model (IV).

Figure 6. Estimated spatial background rates. (Left, Right): Kernel smoothed background rate with variable bandwidth and weights of model (II), generalized additive model (GAM) background rate of model (IV).

and chronic, and there are few swift retaliations across gang territories. Model (IV) investigates whether there exists any additional triggering beyond the scale of census blocks. The estimated percentage of crimes attributed to background, non-triggered crimes (78.2%) in model (IV) is smaller than models (I), (II), and (III). According to the fitted model (IV), an

Table 1. Productivity and background rate parameter estimates, log-likelihood.

	(Model number): Background type			
	(I): Constant	(II): Variable bandwidth/weights	(III): Covariate	(IV): Covariate
GRYD IR Program, κ_1	0.184	0.144	0.172	0.170
	(0.018)	(0.017)	(0.018)	(0.020)
non-GRYD, κ_2	0.197	0.170	0.187	0.186
	(0.012)	(0.011)	(0.012)	(0.013)
Constant background, c	24.644			
	(0.560)			
GRYD IR Program, κ_3 (far-field)				0.102
				(0.030)
non-GRYD, κ_4 (far-field)				0.00893
				(0.0022)
Percent Background	0.807	0.838	0.818	0.782
Log-likelihood	0	143.17	177.64	187.21
Out-of-sample log-likelihood	0	59.8	73.4	76.0

NOTE: The standard errors of the parameter estimates are in parentheses. Spatial units are in kilometers and temporal units are in days. Log-likelihoods are the difference with respect to model (I), where the log-likelihood was −11148.84 and the out-of-sample log-likelihood was −3964.9.

Table 2. Triggering shape parameter estimates.

	(Model number): Background type			
	(I): Constant	(II): Variable bandwidth/weights	(III): Covariate	(IV): Covariate
Temporal decay ω_1	0.00391	0.00389	0.00391	0.00391
	(0.00019)	(0.00021)	(0.00020)	(0.00022)
Temporal decay ω_2 (far-field)				0.0519
				(0.010)
Spatial triggering bandwidth σ_1	0.0151	0.0121	0.0139	0.0138
	(0.001)	(0.0009)	(0.0010)	(0.0011)
Spatial triggering bandwidth σ_2 (far-field)				0.200
				(0.010)

NOTE: The standard errors of the parameter estimates are in parentheses. Spatial units are in kilometers and temporal units are in days.

estimated 18.1% of crimes in this dataset are triggered within the scale of a census block and 3.7% are triggered by preceding crimes occurring at least 130 m away. These estimates are found by a weighted average of (κ_1, κ_2) and (κ_3, κ_4), respectively. The estimated spatial bandwidth for the far-field triggering is 200 m and the estimated median time to retaliation is 13 days. Thus, the fitted parameters in model (IV) suggest that there exists a small but nontrivial amount of triggering which occurs across distances of several hundred meters within short inter-event times.

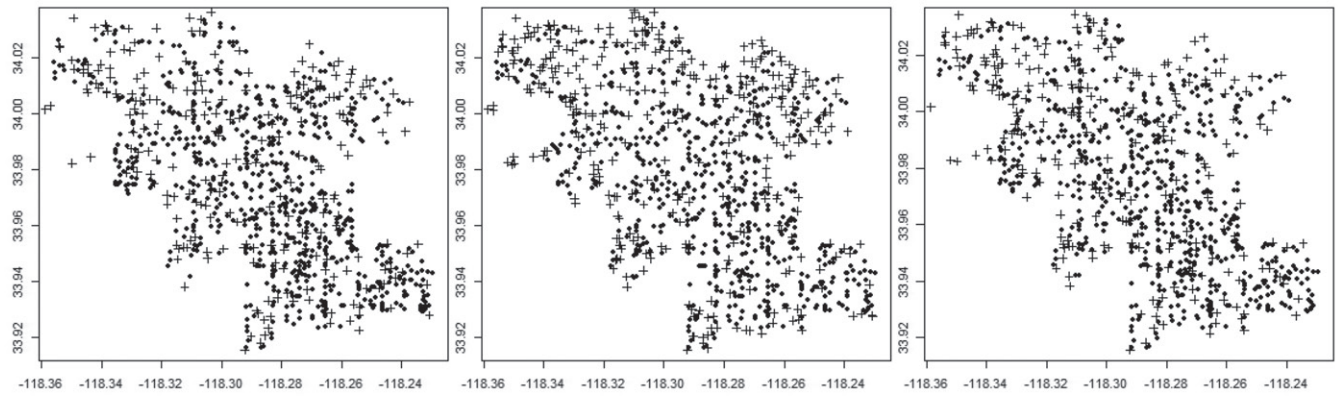
4.3. Out-of-Sample Evaluation

The log-likelihood evaluated on the testing data using models with parameters estimated using only data from 2014 to 2016 and assessed on data from 2017 as a proper predictive score are listed in the bottom row of Table 1. The constant background model (I) offers a poor fit compared to all models. Between models (III) and (IV), the more complex model (IV) has slightly higher out-of-sample log-likelihood while both out-performed model (II). The results suggest that the superior in-sample fit of models (III) and (IV) relative to (II) is not a result of over-fitting.

Superthinned residuals are shown in Figure 7. Model (I) shows clustering of residual points east of Normandie Avenue and sparsity in the North Western quarter of the observation region. There is identifiable clustering (underprediction) in northern census blocks for model (II) while the covariate model (IV) does not exhibit this feature. The temporal distribution of the residuals, shown in Figure 8, shows significant departures from uniformity for models (I) and (II), which appear to significantly underpredict crimes in June and July, 2017. The temporal distribution of residuals for model (IV), meanwhile, has no significant departures from uniformity. No noticeable differences between the spatial distribution of residuals were observed for any of the models between the first half of 2017 and the second half of 2017. The spatial distribution of the residuals are examined in Figure 9. Significant departures from homogeneity are seen for model (I). Model (IV) appears to be slightly more close to homogeneity compared to model (II).

4.4. Efficacy of the GRYD Program

Table 3 shows the average of 50 estimates of model (IV) using all 4 years of data with sampled non-GRYD control marks detailed in Section 3.5. The estimated productivities show that the GRYD IR Program appears to have an effect on reducing triggered reported gang-related violent crimes. For distances less than 130 m within census blocks, the GRYD IR Program appears to reduce retaliation rates from 0.240 to 0.206 retaliations per crime, for a decrease of 14.2%, compared with events in similar locations but without the GRYD IR Program. Over distances greater than 130 m, the GRYD IR Program appears to reduce retaliatory triggering rates from 0.197 to 0.161 retaliations per crime, for a decrease of 18.3%. Note that the estimated productivities for GRYD IR Program and non-GRYD crimes in Table 1 offer a biased estimate of the impact of GRYD IR



(Left to right): Constant background model (I), variable bandwidth - variable weighted kernel smoothed background model (II). GAM background model (IV). Dots: Kept original points. Crosses: superposed points.

Figure 7. Superthinned residuals. (Left to right): Constant background model (I), variable bandwidth - variable weighted kernel smoothed background model (II). GAM background model (IV). Dots: Kept original points. Crosses: superposed points.

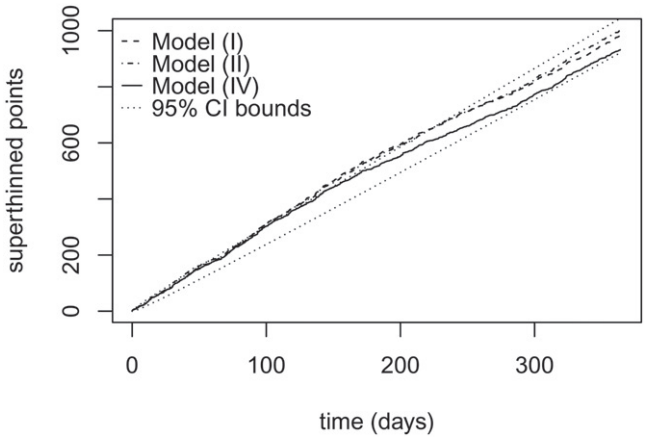


Figure 8. Temporal distribution of superthinned residuals. Cumulative number of superthinned residuals over time in 2017, for models (I), (II), and (IV). Dotted curves are 95% confidence bounds for a stationary Poisson process.

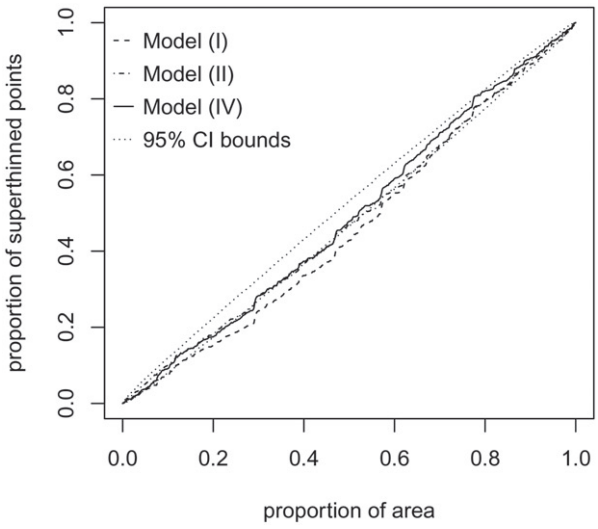


Figure 9. Spatial distribution of superthinned residuals. Cumulative proportion of superthinned residuals over space, sweeping horizontally from left to right for models (I), (II), and (IV). Dotted curves are 95% confidence bounds for a homogeneous Poisson process.

Program violence interruptions because they are not a random assignment over space and time, as seen in Figures 2–4. The parameter estimates in Table 3 are preferable for this purpose.

A measure of precision for the percentage decrease in productivity can be obtained using standard errors reported in Table 3 and simulating percentage differences. However, this is only approximate as we assume that parameter estimates are normally distributed (Ogata 1978) and independent. Also, parameter estimates and standard errors were subject to further variability as the values in Table 3 are an average of estimates from 50 resampled marks from weights (15). Only 8.5% of simulated percentage changes were positive for the program's effect on the productivity of retaliations within 130 m, while 22% were positive for retaliations over 130 m away.

Table 3. Estimated productivity and smoothing weights for sampled controls.

	Model (IV) with sub-sampled control marks
GRYD IR Program, κ_1	0.206
	(0.017)
Sampled non-GRYD controls, κ_2	0.240
	(0.018)
Remaining non-GRYD, κ_3	0.196
	(0.014)
GRYD IR Program, κ_4 (far-field)	0.161
	(0.033)
Sampled non-GRYD controls, κ_5 (far-field)	0.197
	(0.033)
Remaining non-GRYD, κ_6 (far-field)	0.0002
	(0.0028)
Log-likelihood	−14610.16

Average standard errors of the parameter estimates are in parentheses. Data from 1/1/2014 to 12/31/2017 are used.

5. Discussion

This paper proposes an algorithm to non-parametrically estimate the background rate of a marked spatial–temporal point

process model using spatial covariates. After fitting a variety of models designed to describe the inhomogeneity in the background rate as accurately as possible, we find evidence of chronic, near-repeat clustering within the scale of a census block. For models (I), (II), and (III), this sub-census block clustering dominated the estimated triggering parameters, which suggested that almost no retaliations occur swiftly across gang territories. Model (IV) performed better both within and out-of-sample, and its fitted parameters suggest that an estimated 18.1% of reported crimes in this dataset occur in a slow and chronic response to preceding crimes occurring within the scale of a census block and 3.7% are swift retaliations to preceding crimes occurring at least 130 meters away.

To evaluate the efficacy of the GRYD IR Program, we propose a sampling method to find a subset of un-intervened crimes to serve as controls. This revealed that, after accounting for the fact that GRYD IR Program violence interruption efforts occurred in locations of generally high rates of gang-related violent crimes, the GRYD IR Program appears to reduce reported retaliations occurring 130 m away or more by approximately 18.3%, and appear to decrease reported retaliations within 130 m of the original reported crime by 14.2%. The uncertainty in these reductions is an important subject for future research. While we have reported the standard errors for the parameter estimates, future work will focus on further quantifying uncertainty, using simulation-based approaches as well as methods based on the Fisher information for point process parameters (Ogata 1978), in such functions of these parameters, including estimated differences between two parameter estimates that might be correlated.

Methods for bandwidth selection are critical when using kernel smoothing methods for background rate estimates, which can in turn have a large impact on estimates of triggering, as any observations not attributed by the model to the background rate are necessarily attributed to retaliation. To allow for more accurate estimation, we use variable bandwidth kernel smoothing, allowing the bandwidths used in the estimation of the background rate to be different from those governing the triggering kernel. Over-fitting is also a serious concern, and we find no evidence here of over-fitting for model (III) and (IV), which offer superior fit to the data from 2014 to 2016 used in the model fitting, as well as high log-likelihoods evaluated on external data from 2017 used for testing compared to models (I) and (II). Therefore, using covariates to estimate the background rate, rather than simply smoothing over the observed points, appears to be preferable. An important subject for future research is the detailed investigation into the relative importance of covariates on the estimated background rate and predictive performance of the model.

The results hold important policy implications for designing, implementing and evaluating comprehensive gang violence prevention programs (see Spergel 1995; Office of Juvenile Justice and Delinquancy Prevention 2009; Gravel et al. 2013). GRYD is organized around community engagement, gang prevention and intervention, and violence interruption efforts (Tremblay et al. 2020). Our results show that GRYD's violence interruption efforts, termed the GRYD Incident Response (IR) Program, do work. Using the productivities estimated for Model IV, every 100 reported gang-related violent crimes produces on average 43.7 retaliations (24 near-field and 19.7 far-field), when there is no GRYD response. When GRYD Incident Response is notified, the estimated number of retaliations falls to 36.7 for every 100 reported gang-related violent crimes (20.6 near-field and 16.1 far-field). Cities can benefit by including violence interruption in their comprehensive gang violence prevention plans. Our results also suggest that there is room to expand violence interruption in Los Angeles. Nearly 70% (2527 of 3627) of the gang-related violent crimes nominally eligible for violence interruption were not exposed to GRYD IR efforts. While the exposure imbalance created unique opportunities to evaluate the GRYD IR Program, it also suggests that the overall reduction in retaliations could have been even greater had more events been similarly exposed. However, GRYD IR Program resources are limited and choices therefore need to be made about which events are critical to respond to. Policy makers can consider incrementally expanding violence interruption efforts and assessing their effects. Our results suggest that violence interruption is unlikely to eliminate all reported retaliations. We therefore might expect to reach a point of diminishing returns. Where this point may lie is an open question.

Our analyses also highlight a key distinction between background and triggered crimes. Model IV estimated that 78.2% of reported gang violent crimes can be attributed to background processes. The remainder is split between near-field triggering (18.1%), within 130 m of a parent crime, and far-field triggering (3.7%), beyond 130 m from a parent crime. That retaliations appear to happen more often close-by is consistent with the idea that gang-related violence clusters in contested spaces such as near territorial boundaries (Brantingham et al. 2012). These are locations where rivals are more likely to meet, encouraging spontaneous attacks, and also are known places to seek retribution. However, the median time to a reported retaliation in the near-field is close to 180 days. Thus, while the risk of retaliation close-by a parent crime is quite high, it is a chronic, rather than an acute risk. There is much less far-field triggering (beyond 130 m), but it tends to occur quickly, with a median time of 13 days. This is consistent with gang rivalries extending across neighborhoods (Papachristos 2009; Brantingham et al. 2019). These observations are important for the tactical implementation of violence interruption programs. Specifically, events that prompt violence interruption may require resources to be deployed immediately to contend with fast retribution in far-field spaces (i.e., beyond 130 m from the parent). However, the bulk of these resources need to deal with the chronic threat to the immediate area that can last over months. Thus, immediate incident response should naturally extend to proactive peacemaking (Tremblay et al. 2020).

GRYD IR is focused on stopping retaliations. This leaves open the question of how to deal with the estimated 78.2% of reported gang violent crimes that arise from background processes. It is impractical to think that community intervention workers (CIWs) can intervene in all of the situations that might trigger a spontaneous violent crime. Rather, these are events where we expect GRYD community engagement and gang prevention and intervention efforts to have an important effect. Comprehensive violence prevention programs are about changing the so-called root causes of gang violence in addition to dealing with the acute consequences. The "root causes"

ultimately contribute to the background processes that drive spontaneous gang related crimes. However, the impact of these efforts is certainly indirect and likely to play out over relatively long periods of time. In other words, comprehensive gang violence prevention programs can hope to reduce retaliations relatively quickly, but are likely to reduce background crimes relatively slowly.

Acknowledgments

Permission to use these data was provided by the City of Los Angeles Mayor's Office of GRYD. Any opinions, findings, conclusions or recommendations expressed in this study, however, are those of the author(s) and do not necessarily reflect the views of the GRYD Office.

Funding

This research was funded by the City of Los Angeles contract number C-132202.

References

Anderson, E. (1999), *Code of the Street: Decency Violence and the Moral Life of the Inner City.* New York: Norton and Company. [1674]

Balderama, E., Schoenberg, F. P., Murray, E., and Rundel, P. W. (2012), " Application of Branching Point Process Models to the Study of Invasive Red Banana Plants in Costa Rica," *Journal of the American Statistical Association*, 107, 467–476. [1676]

Barton, M. S., Valasik, M. A., Brault, E., and Tita, G. E. (2019), "Gentefication" in The Barrio: Examining the Relationship Between Gentrification and Homicide in East Los Angeles, *Crime & Delinquency*, doi:0011128719860835. [1674]

Bauwens, L., and Hautsch, N. (2009), *Modelling Financial High Frequency Data Using Point Processes* (pp. 953–979). Berlin: Springer Berlin Heidelberg. [1676]

Bichler, G., Norris, A., Dmello, J. R., and Randle, J. (2019), "The Impact of Civil Gang Injunctions on Networked Violence Between the Bloods and the Crips," *Crime and Delinquency*, 65, 875–915. [1676]

Bjerregaard, B., and Lizotte, A. J. "Gun Ownership and Gang Membership," *The Journal of Criminal Law and Criminology*, 86, 37–58, 1995. [1674]

Brantingham, P. J., Tita, G. E., Short, M. B., and Reid, S. E. (2012), "The Ecology of Gang Territorial Boundaries," *Criminology*, 50, 851–885. [1674,1676,1679,1685]

Brantingham, P. J., Valasik, M., and Mohler, G. (2018), "Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial," *Statistics and Public Policy*, 5, 1–8.

Brantingham, P. J., Valasik, M., and Tita, G. E. (2019), "Competitive Dominance, Gang Size and the Directionality of Gang Violence," *Crime Science*, 8, 7. [1674,1685]

Bray, A., and Schoenberg, F. P. (2013), "Assessment of Point Process Models for Earthquake Forecasting," *Statistical Science*, 28(4), 510–520. [1680]

Broyden, C. G. (1970), "The Convergence of a Class of Double-rank Minimization Algorithms," *Journal of the Institute of Mathematics and Its Applications*, 6, 76–90. [1679]

Cespedes, G., and Herz, D. (2011), *The City of Los Angeles Mayor's Office of Gang Reduction and Youth Development (GRYD) Comprehensive Strategy*. Los Angeles, CA: Mayor's Office of Gang Reduction and Youth Development. [1675]

Clements, R. A., Schoenberg, F. P., and Veen, "A. Evaluation of Space-time Point Process Models Using Super-thinning," *Environmetrics*, 23, 606–616. [1680]

Curry, G. D., and Spergel, I. A. (1988), "Gang Homicide, Delinquency, and Community," *Criminology*, 26, 381–406, . [1674]

Daley, D., and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes* (2nd ed.). New York: Springer. [1676,1677]

Daly, M., and Wilson, M. (1988), *Homicide.* New York: Aldine de Gruyter. [1674]

Decker, S. H. (1996), "Collective and Normative Features of Gang Violence," *Justice Quaterly*, 13, 243–264. [1674]

Diggle, P. J. (2014), *Statistical Analysis of Spatial and Spatio-temporal Point Patterns* (3rd ed.). Boca Raton, FL: CRC Press. [1674,1675]

Esbensen, F.-A., Huizinga D., and Weiher A. W. (1993), "Gang and Non-Gang Youth: Differences in Explanatory Factors," *Journal of Contemporary Criminal Justice* 9(2), 94–116. [1674]

Fox, E. W., Coronges, K., Schoenberg, F. P., Short, M. B., and Bertozzi, A. L. (2016), "Modeling e-mail Networks and Inferring Leadership Using Self-exciting Point Processes," *Journal of the American Statistical Association*, 111, 1–21. [1676,1677,1681]

Gravel, J., Bouchard, M., Descormiers, K., Wong, J. S., and Morselli, C. (2013), "Keeping Promises: A Systematic Review and a New Classification of Gang Control Strategies," *Journal of Criminal Justice*, 41, 228–242. [1685]

Green, B., Horel, T., and Papachristos, A. V. (2017), "Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014," *JAMA Internal Medicine*, 177, 326–333. [1674,1675]

Harte, D. S. (2013), "Bias in Fitting the ETAS Model: A Case Study Based on New Zealand Seismicity," *Geophysical Journal International*, 192, 390–412. [1679]

Hastie, T. (2018), *gam: Generalized Additive Models*, R package version 1.16. [1680]

Heckman, J. J. (1991) "Identifying the Hand of Past: Distinguishing State Dependence From Heterogeneity," *American Economic Review*, 81, 75–79. [1674]

Howell, J. C. (2011), *Gangs in America's Communities.* Thousand Oaks, CA: Sage Publications. [1674]

Huebner, B. M., Martin, K., Moule Jr, R. K., Pyrooz, D., and Decker, S. H. (2016), "Dangerous Places: Gang Members and Neighborhood Levels of Gun Assault," *Justice Quarterly*, 33, 836–862. [1679]

Hughes, L. A., and Short, J. F. (2005), "Disputes Involving Youth Street Gang Members: Micro-social Contexts," *Criminology*, 43, 43–75. [1674]

Jacobs, B. A., and Wright, R. T. (2006), *Street Justice: Retaliation in the Criminal Underworld.* Cambridge, UK: Cambridge University Press. [1674]

Johnson, D. H. (1996), "Point Process Models of Single-neuron Discharges," *Journal of Computational Neuroscience*, 3, 275–299. [1676]

Klein, M. W., and Maxson, C. L. (2006), *Street Gang Patterns and Policies*, New York: Oxford University Press. [1674]

Kubrin, C. E., and Weitzer, R. "Retaliatory Homicide: Concentrated Disadvantage and Neighborhood Culture," *Social Problems*, 50, 157–180. [1674]

Kyriacou, D. N., Hutson, H. R., Anglin, D., Peek-Asa, C., and Kraus, J. F. (1999), "The Relationship Between Socioeconomic Factors and Gang Violence in the City of Los Angeles," *The Journal of Trauma: Injury, Infection, and Critical Care*, 46, 334–339. [1676]

Lewis, P. A. W., and Shedler, G. S. (1979), "Simulation of Nonhomogeneous Poisson Processes by Thinning," *Naval Research Logistics Quarterly*, 26, 403–413. [1675]

Martinez, C. (2016), *The Neighborhood Has Its Own Rules: Latinos and African Americans in South Los Angeles*. NYU Press. [1674]

Meyer, S., Elias, J., and Höhle M. (2012), "A Space-time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence," *Biometrics*, 68, 607–616. [1676]

Mohler, G. O. (2014), "Marked Point Process Hotspots Maps for Homicide and Gun Crime Prediction in Chicago," *International Journal of Forecasting*, 30, 491–497. [1674,1675,1676,1677]

Mohler, G. O., Brantingham, P. J., Carter, J., and Short, M. B. (2019), "Reducing Bias in Estimates for the Law of Crime Concentration," *Journal of Quantitative Criminology*, DOI: 10.1007/s10940-019-09404-1, 1–19. [1674]

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011), "Self-exciting Point Process Modeling of Crime," *Journal of the American Statistical Association*, 106, 100–108. [1674,1675,1676]

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., and Brantingham, P. J. (2015), "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association*, 110, 512, 1399–1411. [1675]

Office of Juvenile Justice and Delinquency Prevention. (2009), *OJJDP Comprehensive Gang Model: Planning for implementation.ce.* Washington, DC: Institute for Intergovernmental Research, U.S. Department of Justice. [1685]

Ogata, Y. (1978), "The Asymptotic Behavior of Maximum Likelihood Estimators for Stationary Point Processes," *Annals of the Institute of Statistical Mathematics*, 30(Part A):243–261. [1677,1684,1685]

——— (1998) "Space-time Point-process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402. [1675,1676]

Papachristos, A. V. (2009), "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide," *American Journal of Sociology*, 115, 74–128. [1674,1685]

Papachristos, A. V., and Kirk, D. S. (2006), *Neighborhood Effects on Street Gang Behavior* (pp. 63–84). New York: AltaMira. [1674]

Patillo-McCoy, M. (1999), *Black Picket Fences: Privilege and Peril among the Black Middle Class*. Chicago, IL: University of Chicago Press. [1674]

Reinhart, A., and Greenhouse, J. (2018), "Self-exciting Point Processes With Spatial Covariates: Modeling the Dynamics of Crime," *Journal of the Royal Statistical Society*, Series C, 67, 1305–1329. [1674,1675,1676,1678]

Rosenfeld, R., Bray, T. M., and Egley, A. (1999), "Facilitating Violence: A Comparison of Gang-motivated, Gang-affiliated, and Nongang Youth Homicides," *Journal Of Quantitative Criminology*, 15, 495–516. [1674]

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997), "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science*, 277, 918–924. [1674]

Schoenberg, F. P. (2013), "Facilitated Estimation of ETAS," *Bulletin of the Seismological Society of America*, 103, 601–605. [1676,1679]

Sheather, S. J., and Jones, M. C. (1991), "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society*, Series B, 53, 683–690. [1679,1681]

Skogan, W. G., Hartnett, S. M., Bump, N., and Dubois, J. (2009), *Evaluation of CeaseFire-Chicago*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. [1675]

Spergel, I. A. (1995), *The Youth Gang Problem: A Community Approach*, Oxford University Press. [1685]

Tita, G., and Griffiths, E. (2005), "Traveling to Violence: The Case for a Mobility-based Spatial Typology of Homicide," *Journal of Research in Crime and Delinquency*, 42, 275–308. [1679]

Tita, G., and Ridgeway, G. (2007), "The Impact of Gang Formation on Local Patterns of Crime," *Journal of Research in Crime and Delinquency*, 44, 208–237. [1674,1679]

Tremblay, A., Herz, D. C., and Kraus, M. (2020), *The Los Angeles Mayor's Office of Gang Reduction and Youth Development Comprehensive Strategy*, Los Angeles, CA: The Los Angeles Mayor's Office of Gang Reduction and Youth Development. [1675,1685]

Valasik, M., Barton, M. S., Reid, S. E., and Tita, G. E. (2017), "Barriocide: Investigating the Temporal and Spatial Influence of Neighborhood Structural Characteristics on Gang and Non-gang Homicides in East Los Angeles," *Homicide Studies*, 21, 287–311. [1674]

Valasik, M., and Tita, G. (2018), "Gangs and Space," in *The Oxford Handbook of Environmental Criminology*, Oxford: Oxford University Press, pp. 843–871. [1679]

Yuan, B., Hao, L., Bertozzi, A. L., Brantingham, P. J., and Porter, P. (2019), " Multivariate Spatiotemporal Hawkes Processes and Network Reconstruction," *SIAM J. Mathematics of Data Science*, 1, 356–382. [1677]

Zhuang, J., and Mateu, J. (2019), "A Semi-parametric Spatiotemporal Hawkes-type Point Process Model With Periodic Background for Crime Data," *JRSS A*, 182, 919–942. [1677]

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002), "Stochastic Declustering of Space-time Earthquake Occurrences," *Journal of the American Statistical Association*, 97, 369–380. [1677]