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[§]

October 10, 2019

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 non-parametrically 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 2014-2016 and evaluated using data from 2017, based on log-likelihood and superthinned residuals. The impact of non-randomized 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 meters from the original reported crimes, and a reduction of 14.2% in retaliations within 130 meters.

Keywords: Gang violence, criminology, violence interruption, point processes, self-exciting.

[*]Department of Statistics, University of California, Los Angeles
[†]Corresponding Author, frederic@g.ucla.edu
[‡]Department of Mathematics, University of California, Los Angeles
[§]Department of Anthropology, University of California, Los Angeles

1 Introduction

Crime occurrences are highly localized in space and time. The observed clustering in the data has been theorized to be driven by two main effects: (1) spatial heterogeneity of local risk factors for crime and (2) the dependence on recent crimes which may incite retaliations and repeat offenses (Heckman 1991). This is in particular true of gang violence, though these two effects are difficult to distinguish in the observed data and often confounded (Johnson 2008).

Gang crimes are often retaliatory in nature (Decker 1996, Klein and Maxon 2006). Interactions between gangs that threaten geographic territory or gang reputation can easily escalate to a shooting, while a shooting or homicide often demands retribution in kind (Hughes and Short 2005, Jacobs and Wright 2006), ultimately driving a sequence of tit-for-tat reciprocal attacks (Bjerregaard and Lizotte 1995, Rosenfeld et al. 1999, Howell 2011). Such a process of self-excitation or contagion that triggers retaliation may be linked to a deep-seated drive for retaliatory aggression (Daly and Wilson 1998), street codes that demand quick and decisive retribution (Anderson 1999, Jacobs and Wright 2006), commitment to delinquent peers (Esbensen et al. 1993), the central role of reputation in gang life (Decker 1996), and social networks that promote the rapid spread of rumors (Hughes and Short 2005, Green, Horel, Papachristos 2017).

On the other hand, such crimes are more apt to occur and be reported in some portions of space-time simply due to heterogeneity of the physical and social environment. The rate of crime reporting and police surveillance may vary across space, and features such as housing density and economic disadvantage influence the number of gang-related crimes (Papachristos and Kirk 2006, Woodworth et al. 2014). Other known influences on space-time gang-related crime rates include street network structure (Beavon et al. 1994, Smith et al. 2012) and gang territorial footprints that have remained nearly constant for decades

(Patillo-McCoy 1999, Tita and Ridgeway 2007).

On the basis that reported crimes are driven by both spatial heterogeneity and retaliation, it has recently been argued that crime data can be modeled using Hawkes processes (Mohler 2011, Mohler 2014, Reinhart & Greenhouse 2018), suggesting that clustering in the data is *causal*: an event happening to occur in a particular location increases the likelihood that other events will occur in its vicinity in the near future, while some events occur exogenously due to a chronic, spatially inhomogeneous background component. Causal clustering may create the 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 this goal has been implemented in Los Angeles since 2009.

The City of Los Angeles Mayor's Office of Gang Reduction and Youth Development's (GRYD) Incident Response (IR) Program engages in violence interruption activities in an effort to decrease retaliatory gang-related violent crime (Tremblay, Herz and Kraus in prep, Cespedes & Herz 2011, Cespedes 2012, Skogan et al. 2009). Due to resource limitations, these activities occur following only a subset of reported gang-related violent crimes, and a question of interest is whether the GRYD IR Program is effective and how much, if at all, it reduces retaliatory crime.

Central to our study of the effect of violence interruption on gang-related violent crime is the discrimination between causal clustering and inhomogeneity in the data. Discriminating causal clustering and inhomogeneity is a difficult problem arising frequently in the study of spatial-temporal point processes (see chapter 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, section 4.2) suggests modeling the background rate for earthquakes using only the largest 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.

Therefore, in the case of reported gang-related crimes in South Los Angeles, we attempt to model the inhomogeneity non-parametrically 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 randomized due to practical and ethical considerations. GRYD Community Intervention Workers (CIWs) use highly specialized knowledge of the local gang dynamics and intervene in areas believed to be more prone to retaliations (Tremblay, Herz and Kraus in prep). As a result, excitation/retaliation rates of crimes that have been exposed to the GRYD IR Program are naturally biased upward 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 & 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 paper is organized as follows. A brief description of the data is provided in Section 2. Our proposed iterative method to incorporate non-parametric 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 2014-2017 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 (1,302 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 meter interval centered at each reported crime.

Demographic and socio-economic covariates are compiled at the census tract block level, which is currently the highest resolution published by the U.S. Census. These data are obtained from the American Community Survey (ACS, 2017), 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 essentially reported crime rate per unit of time and space while Kyriacou et al. (1999) studied reported crimes per 100,000 people.

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 (10) below, with background rate estimated using (9). 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 & 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 & Hautsch, 2009), neuron activity (Johnson, 1996), reported burglaries (Mohler et al. 2011), email networks (Fox et al. 2016) and terrorist

attacks (Porter & White, 2012). The Hawkes model can be specified as

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

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 $\kappa_2 + \kappa_2^2 + \kappa_2^3 + ... = \frac{1}{1-\kappa_2} - 1$ 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 $\int\int_A g_1(x, y)dA = \int_0^\infty \int_0^{2\pi} g_1(r\cos\theta, r\sin\theta)rdrd\theta = \int_0^\infty \int_0^{2\pi} h(r)rdrd\theta = 1$, 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 parametrized model for $\lambda(t, x, y)$, the log-likelihood of an observed sequence of reported crimes over an interval $[0, T]$ in region A is (Daley & Vere-Jones 2003)

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

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 (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) propose estimating μ using a time-invariant smoother over all events, using mark dependent weights, for example:

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

The smoothing bandwidth η is typically constrained to be identical to the triggering bandwidth σ in (2) in order to achieve numerical stability in optimization and identifiability of the parameters (Mohler et al. 2014). In our study, we choose not to impose these constraints on the smoothing nor triggering bandwidths and instead estimate them separately in what follows below.

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 over-smoothed estimates between dense and sparse areas (Zhuang et al., 2002). Thus, as an alternative to (4), we obtain a weighted, variable-bandwidth, stochastically de-clustered 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 variable bandwidth 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 parameters of h and updating estimates of $\{w_i\}_{i=1}^N$.

3.2.2 Temporal Variation in Background Rate

According to (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 non-stationary:

$$\hat{\mu}(x, y, t) = v(t) \cdot \sum_{i:(x_i,y_i)\neq(x,y)}^{N} \frac{w_i}{2\pi d_i^2} \exp\big(-\frac{(x-x_i)^2 + (y-y_i)^2}{2d_i^2}\big). \tag{7}$$

To estimate the temporally varying component $v(t)$ in the background rate (7), we follow the componentwise procedure of Fox et al. (2016), obtaining the estimate $\hat{v}(t) = c\hat{f}_1(h)\hat{f}_2(d)\hat{f}_3(y)$ where $h \in [0, 24)$, $d \in \{0, \ldots, 6\}$ and $y \in \{0, 1, 2, 3\}$ represent hours, days and years respectively. We estimate the daily cycle $\hat{f}_1(h)$ 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, i.e. $\hat{f}_2 = \sum_{i=1}^{N} I(d_i = d)/N$ and $\hat{f}_3 = \sum_{i=1}^{N} I(y_i = y)/N$. The estimated daily, weekly, and multi-year components of $v(t)$ 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 $\int_0^T \hat{v}(t)dt = 1$ and is accurately approximated by a Riemann sum.

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 covari-

ates to model the background rate of a Hawkes process for reported burglaries in Pittsburgh with the parametric form

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

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 using *gam*, $v(x, y)$ 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, e.g. by generalized additive modeling (GAM). 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 k^{th} 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 (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 $l \leftarrow 0$, $w_i^{(0)} \leftarrow$ `null` and $\mu^{(0)}(x, y) \leftarrow 10$.

2. Using maximum likelihood estimation, fit

$$\lambda(x, y, t) = c \cdot v(t) \cdot \mu^{(l)}(x, y) + \sum_{i:t_i<t} \kappa(i) g_1(x - x_i, y - y_i) g_2(t - t_i)$$

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.

3. Calculate w_i from (6) and update $w_i^{(l+1)} \leftarrow w_i$ for $i = 1 \ldots N$.

4. Fit

$$\hat{\mu}(x, y) = \hat{f}(v(x, y)), \tag{9}$$

where f is estimated by nonparametric regression of $\sum_{i \in B_k} w_i / a_k$ on census block level covariates and update $\mu^{(l+1)}(x, y) \leftarrow \hat{f}(v(x, y))$.

5. If $\max_i |w_i^{(l+1)} - w_i^{(l)}| > \epsilon$, where ϵ is a small positive number, then update $l \leftarrow l+1$ and go to step (2). Otherwise stop.

The function f can be estimated using any nonparametric regression method in step (4), and we estimate f via GAM in the application here for maximal flexibility.

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 meter 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:

$$\lambda(x,y,t) = v(t)\mu(x,y) + \sum_{i:t_i<t} \kappa_{near}(i)\frac{h_1(r)}{2\pi r}\omega_1 e^{-\omega_1(t-t_i)} + \sum_{\substack{i:t_i<t,\\ r\geq d}} \kappa_{far}(i)\frac{h_2(r)}{2\pi r}\omega_2 e^{-\omega_2(t-t_i)}, \quad (10)$$

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 $\kappa_{near}(i) = \kappa_1$ and $\kappa_{far}(i) = \kappa_3$ if crime i is associated with the GRYD IR Program, and otherwise $\kappa_{near}(i) = \kappa_2$ and $\kappa_{far}(i) = \kappa_4$, 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 (3) is straightforward to compute while the integral term must be numerically approximated, which can be a substantial computational challenge (Harte 2014). In all models investigated in this paper, 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 (3) and the sum in (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:

$$l(\kappa_1,\kappa_2,\beta_1,\beta_2,\omega,\sigma,\eta) = \sum_{i=1}^{N} \log[\lambda(x_i,y_i,t_i)] - \sum_{i=1}^{N} \big[w_i + \kappa(i)\big]. \quad (11)$$

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

rate

$$\mu(x, y, t) = c \cdot v(t), \tag{12}$$

whose log-likelihood is

$$l(\kappa_1, \kappa_2, c, \omega, \sigma) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - c \cdot |S| - \sum_{i=1}^{N} \kappa(i), \tag{13}$$

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

3.5 Sampling non-GRYD crimes as controls for GRYD IR Program crimes

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, it is well known that 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 (Skogan et al. 2009; Tremblay, Herz and Kraus in prep). 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 & Jones (1991). The 24 hour 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 $v(t)$ in (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 w_i given by

$$\nu_i = \frac{\hat{P}_1(x, y)\hat{P}_2(h)\hat{P}_3(d)}{\hat{Q}_1(x, y)\hat{Q}_2(h)\hat{Q}_3(d)}. \tag{14}$$

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.

3.6 Evaluation Methods

Four types of models and background rate estimation methods are investigated: (I) constant background model in (12), (II) kernel smoothed background model in (5), (III) covariate

background model in (9), all with triggering as in (2), and (IV) covariate background model (9) with near and far-field triggering as in (10). To assess the efficacy of the GRYD IR Program, we also evaluate the fit of the 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. 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 (Clements et al. 2013, Bray and Schoenberg 2013). The observations are first thinned, i.e. 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, i.e. 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 (14), 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, 3 and 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 modest deviations.

The following five variables are selected by the stepwise selection procedure in the R package gam (Hastie, 2018): income per capita, unemployment, population density, percent male and percent single parent families. These same variables were selected even when different initial models were used for the stepwise procedure. The estimated additive predictors for the GAM regression background rate are shown in Figure 5. Estimated background rates of reported gang-related violent crimes are higher in areas with high population density and in areas with low income per capita and high unemployment rates. 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.

The estimated spatial background rates (excluding the global non-stationary component $v(t)$) 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^o, 33.98^o) and Crenshaw (-118.35^o, 34.02^o) as well as along Normandie Avenue (longitude -118.3^o). 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 2014-2016 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). Between models (III) and (IV), the more complex model (IV) with independent triggering for near and far-field retaliations has higher in-sample log-likelihood.

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 meters, and is comparable to the bandwidth selected using Sheather & Jones (1991).

In Table 2, the estimated spatial triggering bandwidth σ in models (I), (II), and (III) are all very local, between 12 to 15 meters, 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 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 tract 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 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 meters 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 meters and the estimated median time to retaliation is 13 days. Thus, the fitted parameters in model (IV) suggest that there exists a small but non-trivial 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 2014-2016 and assessed on data from 2017 are listed in the bottom row of Table 1. Not surprisingly, 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 tracts for kernel smoothed background model (II) while the covariate model (IV) does not exhibit this feature.

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 130m 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 130m, 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 Program violence interruptions because they are not a random assignment over space and time, as seen in Figures 2,3 and 4. The parameter estimates in Table 3 are preferable for this purpose.

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 tract. For models (I), (II) and (III), this sub-census tract 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 meters away or more by approximately 18.3%, and appear to decrease reported retaliations within 130 meters of the original reported crime by 14.2%. Future research will explore spatially varying retaliation rates.

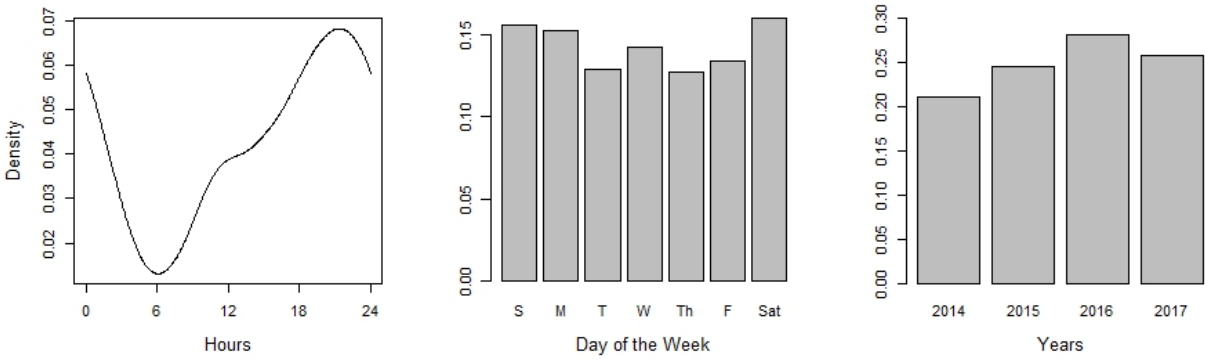
Methods for bandwidth selection are critical when using kernel smoothing methods for background rate estimates, which can in turn have a large impact on estimates of triggering, as any observations not attributed by the model to the background rate are necessarily attributed to retaliation. To allow 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 2014-2016 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.

Acknowledgements

Permission to use these data was provided by the City of Los Angeles Mayor's Office of Gang Reduction and Youth Development (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.

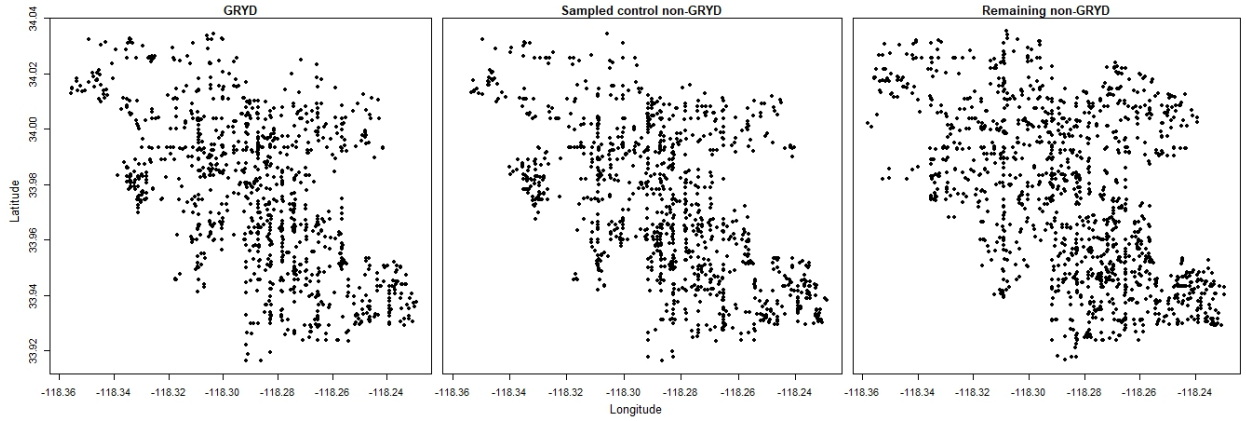
This research was funded by the City of Los Angeles contract number C-132202.

Figure 1: Temporal distribution of all gang crimes



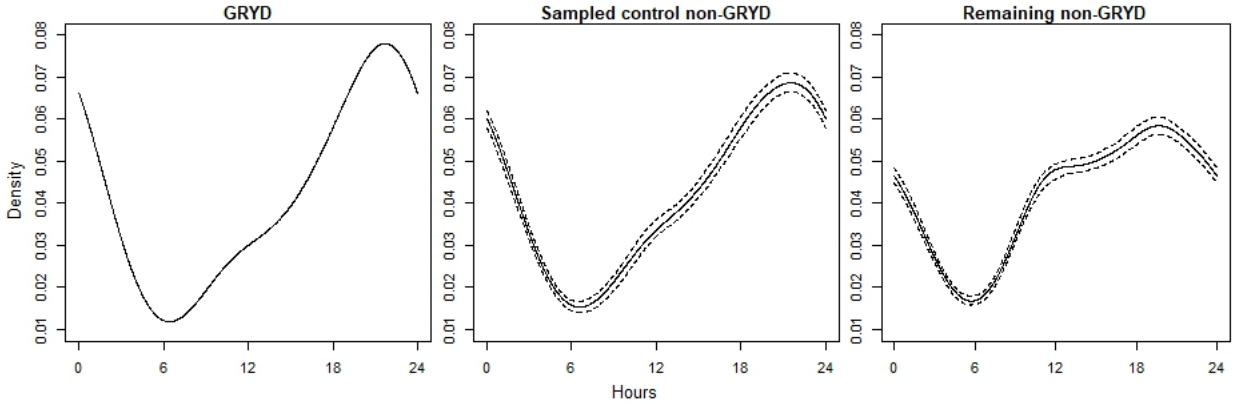
(Left to right): Daily cycle, weekly cycle, and year-to-year variations in crime used for components of $\hat{v}(t)$.

Figure 2: Spatial distribution of marks



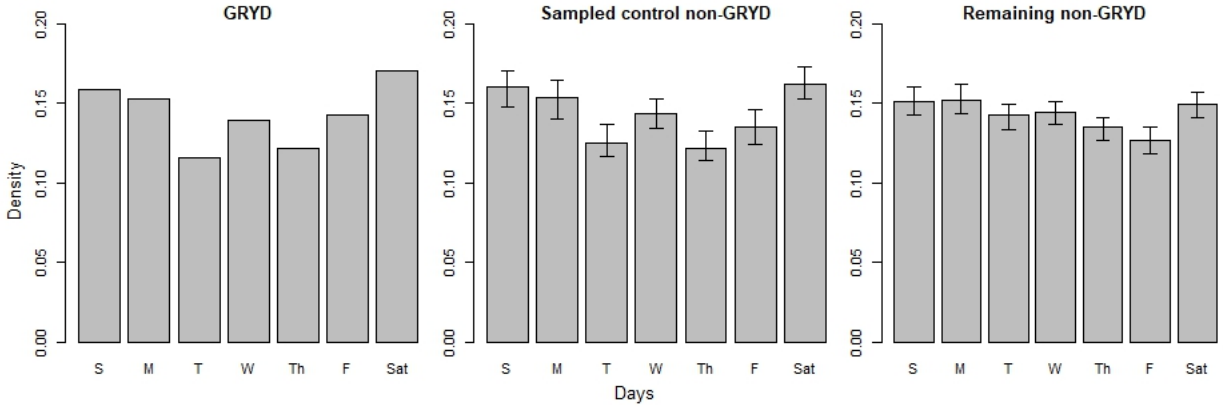
(Left to Right): Locations of GRYD IR Program crime events, one sample of non-GRYD crimes and remaining unsampled non-GRYD crimes. The union of crimes in these three panels are used to estimate models in Table 1.

Figure 3: Temporal distribution of marks (hourly)



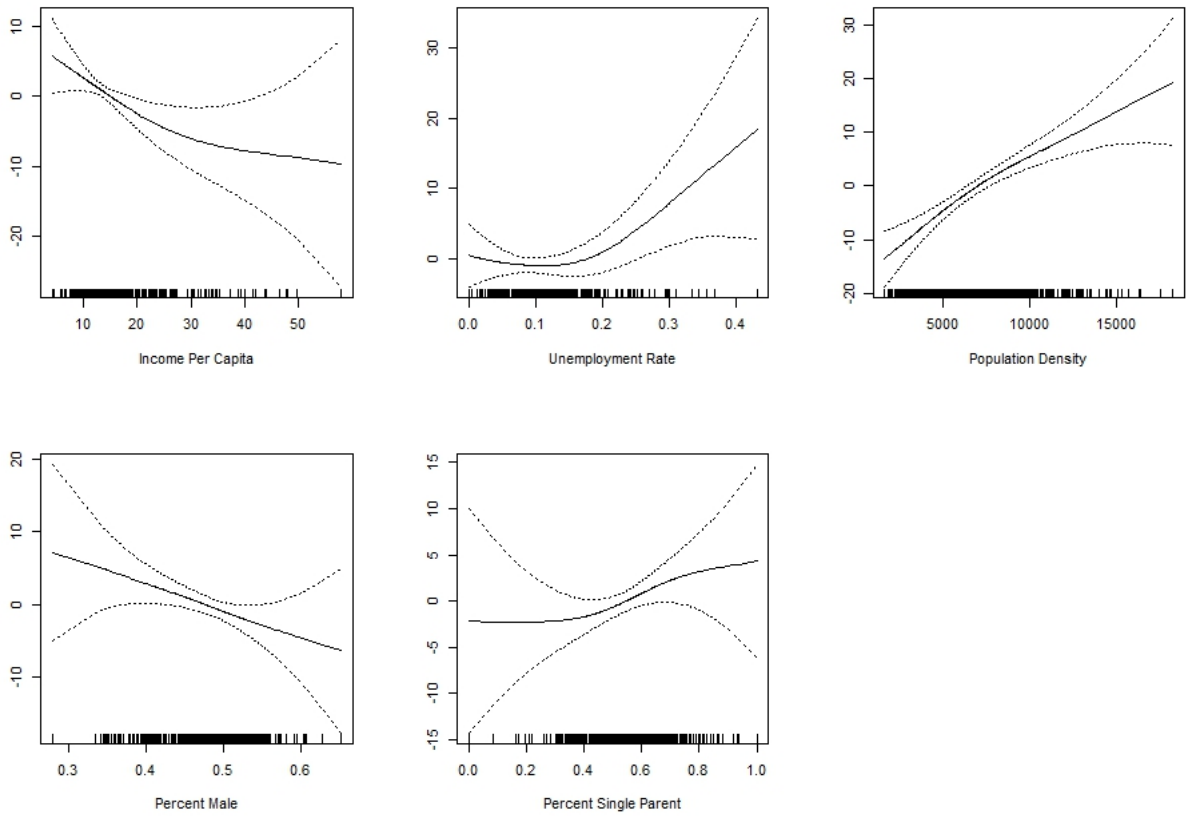
(Left to Right): Kernel density of GRYD IR Program crime events, average kernel density of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Dotted: 5th and 95th percentile of 50 estimated kernel densities.

Figure 4: Temporal distribution of marks (weekly)



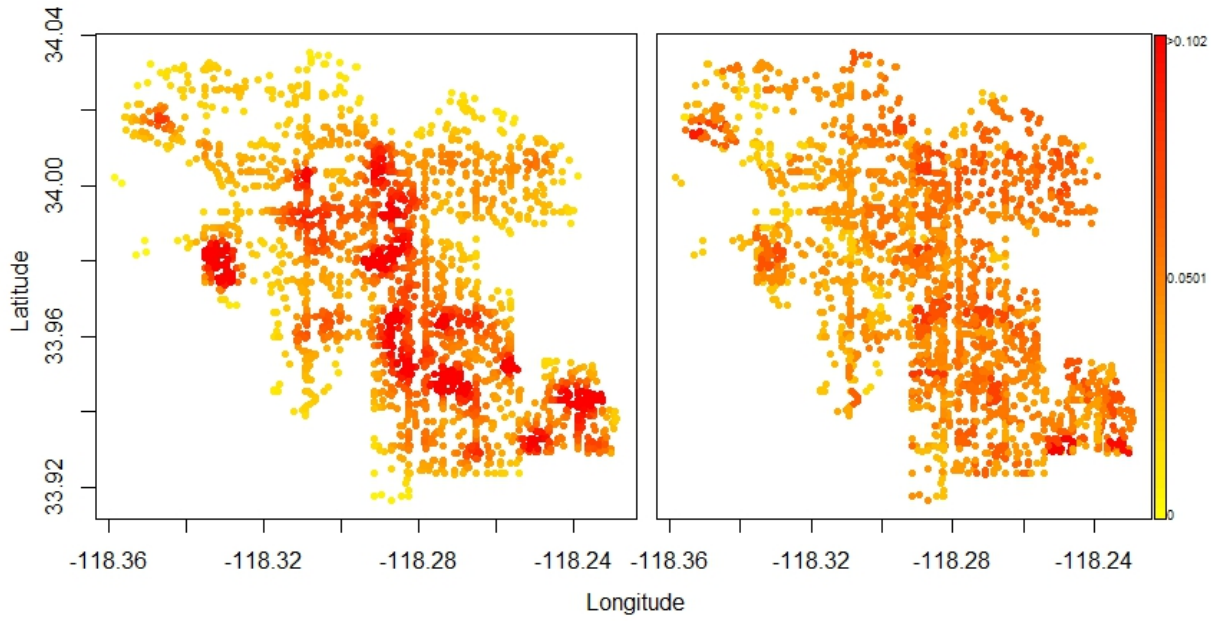
Proportion of occurences each day. (Left to Right): GRYD IR Program crime events, average of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Whiskers: 5th and 95th percentile proportion of 50 samples.

Figure 5: Estimated additive predictors of GAM background



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 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).

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)	(0.014)
non-GRYD, κ_2	0.197	0.170	0.187	0.186
	(0.012)	(0.011)	(0.012)	(0.011)
Constant background, c	24.644			
	(0.560)			
GRYD IR Program, κ_3 (far-field)				0.102
				(0.027)
non-GRYD, κ_4 (far-field)				0.00893
				(0.0014)
Percent Background	0.807	0.838	0.818	0.782
Log-likelihood	0	143.17	177.64	187.22
Out-of-sample log-likelihood	0	59.8	73.3	75.9

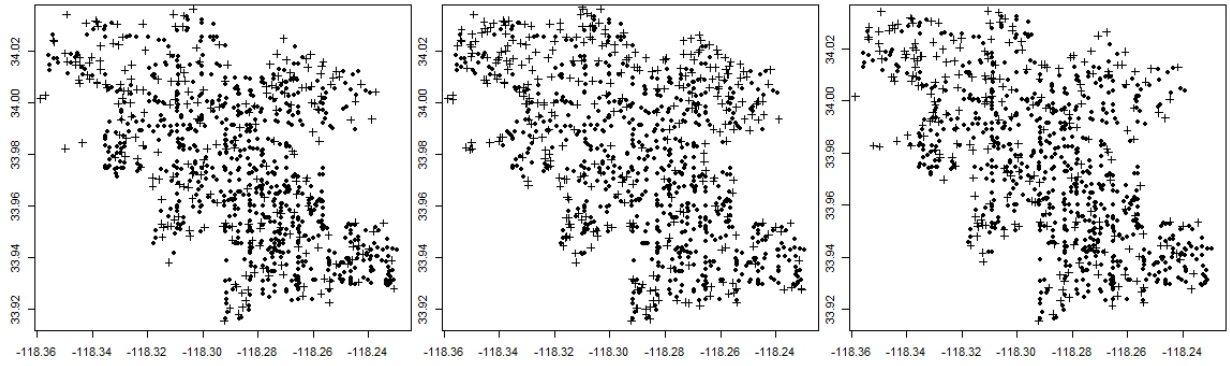
The standard errors of the parameter estimates are in parentheses. Spatial units are in kilometers and temporal units are in days. Log-likelihoods are the difference with respect to model (I), where the log-likelihood was -11148.84 and the out-of-sample log-likelihood was -3964.9.

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)	(0.00021)
Temporal decay ω_2 (far-field)				0.0519
				(0.0099)
Spatial triggering bandwidth σ_1	0.0151	0.0121	0.0139	0.0138
	(0.001)	(0.0009)	(0.0010)	(0.001)
Spatial triggering bandwidth σ_2 (far-field)				0.200
				(0.012)

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

Figure 7: Superthinned Residuals



(Top, 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.

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

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

References

[1] C.L. Althaus. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLOS Current Outbreaks*, 2014.

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

[3] E. Balderama, F.P. Schoenberg, E. Murray, and P.W. Rundel. Application of branching point process models to the study of invasive red banana plants in Costa Rica. *Journal of the American Statistical Association*, 107(498):467–476, 2012.

[4] L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. *SSRN Electronic Journal*, 01 2006.

[5] D. J. K. Beavon, P. J. Brantingham, and P. L. Brantingham. *The Influence of Street Networks on the Patterning of Property Offenses.*, volume 2. Criminal Justice Press, Monsey, 1994.

[6] B. Bjerregaard and A.J. Lizotte. Gun ownership and gang membership. *The Journal of Criminal Law and Criminology*, 86:37–58, 1995.

[7] P.J. Brantingham, N. Sundback, B. Yuan, and K. Chan. GRYD Intervention Incident Response and gang crime. *2017 Evaluation Report.*, pages 1–45, 2017.

[8] P.J. Brantingham, M. Valasik, and G. Mohler. Does predictive policing lead to biased arrests? results from a randomized controlled trial. *Statistics and Public Policy*, 5(1), 2018.

[9] C. G. Broyden. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6:76–90, 1970.

[10] G. Cespedes and D. Herz. The City of Los Angeles Mayors Office of Gang Reduction and Youth Development (GRYD) comprehensive strategy. *Los Angeles: Mayors Office of Gang Reduction and Youth Development.*, 2011.

[11] G. Cespedes and D. Herz. Los Angeles youth development and gang reduction comprehensive model. *City of Los Angeles*, 2012.

[12] R.A. Clements, F.P. Schoenberg, and A. Veen. Evaluation of space-time point process models using super-thinning. *Environmetrics*, 23(7):606–616, 2013.

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

[14] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure.* New York: Springer, 2007.

[15] M. Daly and M. Wilson. *Homicide.* Aldine de Gruyter, New York., 1988.

[16] S.H. Decker. Collective and normative features of gang violence. *Justice Quaterly*, 13:243–264, 1996.

[17] P.J. Diggle. *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. CRC Press, Boca Raton, 3 edition, 2014.

[18] R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, 13(3):317–322, 1970.

[19] E.W. Fox, K. Coronges, F.P. Schoenberg, M.B. Short, and A.L. Bertozzi. Modeling e-mail networks and inferring leadership using self-exciting point processes . *Journal of the American Statistical Association*, 111(514):1–21, 2016.

[20] E.W. Fox, F.P. Schoenberg, and J.S. Gordon. Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *Annals of Applied Statistics*, 10(3):1725–1756, 2016.

[21] D. Goldfarb. A family of variable metric updates derived by variational means. *Mathematics of Computation*, 24(109):23–26, 1970.

[22] B. Green, T. Horel, and A.V. Papachristos. Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Internal Medicine*, 177(3):326–333, 2017.

[23] D.S. Harte. Bias in fitting the etas model: a case study based on New Zealand seismicity. *Geophysical Journal International*, 192(1):390–412, 2013.

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

[25] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.

[26] J.J. Heckman. Identifying the hand of past: Distinguishing state dependence from heterogeneity. *American Economic Review*, 81:75–79, 1991.

[27] J.C. Howell. *Gangs in America's Communities.* Sage Publications, Thousand Oaks, 2011.

[28] L.A. Hughes and J.F. Short. Disputes involving youth street gang members: micro-social contexts. *Criminology*, 43:43–75, 2005.

[29] B.A. Jacobs and R.T. Wright. *Street justice: Retaliation in the criminal underworld.* Cambridge University Press., 2006.

[30] D.H. Johnson. Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 3(4):275299, 1996.

[31] M.W. Klein and C.L. Maxson. *Street gang patterns and policies.* Oxford University Press, New York, 2006.

[32] D.N. Kyriacou, H.R. Hutson, D. Anglin, C. Peek-Asa, and J.F. Kraus. The relationship between socioeconomic factors and gang violence in the city of Los Angeles. *The Journal of Trauma: Injury, Infection, and Critical Care*, 46(2):334–339, 1999.

[33] D. Marsan and O. Lengliné. Extending earthquakes reach through cascading. *Science*, 319:1076–1079, 2008.

[34] S. Meyer, J. Elias, and Höhle M. A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68:607616, 2012.

[35] G.O. Mohler. Marked point process hotspots maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30:491–497, 2014.

[36] G.O. Mohler, P.J. Brantingham, J. Carter, and M.B. Short. Reducing bias in estimates for the law of crime concentration. *Journal of Quantitative Criminology*, DOI: 10.1007/s10940-019-09404-1., 2019.

[37] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, and G.E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[38] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[39] Y. Ogata. The asymptotic behavior of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(Part A):243–261, 1978.

[40] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

[41] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50:379–402, 1998.

[42] A.V. Papachristos and D.S. Kirk. *Neighborhood effects on street gang behavior*, pages 63–84. AltaMira, New York, 2006.

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

[44] A. Reinhardt and J. Greenhouse. Self-exciting point processes with spatial covariates: modeling the dynamics of crime. *Journal of the Royal Statistical Society: Series C*, Forthcoming., 2018.

[45] R. Rosenfeld, T.M. Bray, and A. Egley. Facilitating violence: A comparison of gang-motivated, gang-affiliated, and nongang youth homicides. *Journal Of Quantitative Criminology*, 15:495–516, 1999.

[46] F.P. Schoenberg. Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1):601–605, 2013.

[47] F.P. Schoenberg, M. Hoffman, and R. Harrigan. A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287, 2019.

[48] D. Schorlemmer, J. D. Zechar, M. J. Werner, E. H. Field, D. D. Jackson, T. H. Jordan, and THE RELM WORKING GROUP. First results of the regional earthquake likelihood models experiment. *Pure and Applied Geophysics*, 167:859–876, 2010.

[49] D.F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656, 1970.

[50] S.J. Sheather and M.C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B.*, 53(3):683–690, 1991.

[51] B.W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.

[52] W.G. Skogan, Susan M. Hartnett, Natalie Bump, and Jill Dubois. *Evaluation of CeaseFire-Chicago.* U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, D.C., 2009.

[53] L.M. Smith, A.L. Bertozzi, P.J. Brantingham, G.E. Tita, and M. Valasik. Adaptation of an animal territory model to street gang spatial patterns in Los Angeles. *Discrete and Continuous Dynamical Systems*, 32:3223–3244, 2012.

[54] G. Tita and G. Ridgeway. The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquency*, 44(2):208–237, 2007.

[55] A. Tremblay, D.C. Herz, and M. Kraus. The Los Angeles Mayors Office of Gang Reduction and Youth Development comprehensive strategy. *Los Angeles, CA: The Los Angeles Mayors Office of Gang Reduction and Youth Development.*, 2019.

[56] T.J. Woodworth, G.O. Mohler, A.L. Bertozzi, and P.J. Brantingham. Nonlocal crime density estimation incorporating housing information. *Philosophical Transactions of the Royal Society A*, 372(20130403), 2014.

[57] J.D. Zechar, D. Schorlemmer, M.J. Werner, M.C. Gerstenberger, D.A. Rhoades, and T.H. Jordan. Regional earthquake likelihood models i: First-order results. *Bull. Seismol. Soc. Amer.*, 103(2A):787–798, 2013.

[58] J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.