Title
Assessing the impact of interventions on retaliatory violent crimes using Hawkes models with covariates

Author
Wang, Huanchen

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Los Angeles

Assessing the impact of interventions on retaliatory violent crimes using Hawkes models with covariates

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Huanchen Wang

2018

ABSTRACT OF THE THESIS

Assessing the impact of interventions on retaliatory violent crimes using Hawkes models with covariates

by

Huanchen Wang
Master of Science in Statistics
University of California, Los Angeles, 2018
Professor Frederic Paik Schoenberg, Chair

Recent studies suggest the effective application of Hawkes point process models to describe the self-exciting nature of gang crimes. We focus on applying Hawkes point process models with spatial covariates incorporated in the background rate to assess the City of Los Angeles Mayors Office Gang Reduction and Youth Development (GRYD) intervention program. The data includes all officially recorded crimes that took place in South Los Angeles during 2014-2016. The fitted model is compared to a previous study (Brantingham et al. 2018) conducted on the same topic using Hawkes point process model with smoothing kernel background rate. The GRYD intervention program is found to be involved at locations with highly clustered crimes, which has led to difficulty in assessment. After accounting for this differential placement of GRYD events, the fitted model shows that the GRYD intervention program results in effective reduction of gang crime retaliations, although the results are not statistically sigificant.

The thesis of Huanchen Wang is approved.

Nicolas Christou

Hongquan Xu

Frederic Paik Schoenberg, Committee Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chair Rick Schoenberg for being a wonderful advisor and providing me much guidance. I would like to acknowledge the crucial contributions from Junhyung Park, who has helped with this work significantly along the way. I would also like to thank Hongquan Xu and Nicolas Christou, my committee members, for being great professors throughout my undergrad and grad school years at UCLA. I have learned so much from each one of you mentioned above.

I need to extend my love and thanks to my parents for being so patient with me and supporting me from half the world away over the past 8 years. I could not have done this and many other things without you. You are the best and the most selfless parents a child can ever have. Although we do not see each other often, I do think of you and cherish your love and support everyday.

I would like to thank Carmen, my sister and best friend, for always being there for me regardless the physical distance between us. I could not have gotten through the darkest days in the past years without you talking to me and making me laugh on the other side of the telephone. I truly hope that you will get everything beautiful in life.

Also, I want to thank Elissa for being the most caring friend I have ever had. We have had so many great adventures together and you have helped me in so many ways in life. You are my best friend and someone I know that I can rely on in this foreign city.

At last but not least, Bobby, I am so grateful and lucky to have you in my life. You are so kind and so unique. You are so fun and you make me feel that I am on the right track in life. Everytime thinking of you makes me smile.

Introduction

Gang crime and violence have been a major challenge for the city of Los Angeles given the fact that gang crimes can trigger highly clustered event sequences (Mohler et al. 2011). Particularly, a gang shooting can trigger waves of retaliatory violence of the rival gang in the local setting (Tita and Ridgeway, 2007; Cohen and Tita, 1999) , which could expose the civilians and neighborhoods to additional danger and disturbance. As an important strategy targeting gang crime and violence, since 2009 the City of Los Angeles Mayors Office Gang Reduction and Youth Development (GRYD) started engaging in disrupting the process of gang crime retaliations by providing community-based interventions.

To evaluate the effectiveness of GRYD Intervention Incident Response program, statistical research and studies are conducted. Existing studies have proposed using Hawkes process model to assess the dynamics of gang violence (Short et al. 2014). Particularly, in a recent study focusing on the evaluation of GRYD program, a multivariate Hawkes process model is proposed, and the analysis results show that GRYD intervention has brought down gang retaliations significantly (Brantingham et al. 2018). The model used in the above study is an extension of the Hawkes process model which partitions the cause of crime into background process and triggering process. However, the study has certain limitations. One common opinion of the critics is that the background component of the model did not take the varying spatial features, such as the economical and social differences of the neighborhoods, into account.

In this paper we focus on revising Brantingham et al. (2018) by including spatial features of the background environment in the model, and assessing how it changes the productivity estimates. More particularly, covariates that are meaningful in representing the zip codes and neighborhoods in both social and economical aspects are incorporated in the background component of the Hawkes process model. One existing study that has introduced including spatial features in analyzing the dynamics of crime in a similar approach is by Reinhart (Reinhart and Greenhouse 2017). The difference is that we compare the model with spatial covariates to a smoothing kernel background model, but no such comparison is studied in

Reinhart and Greenhouse (2017). In addition, we investigate how other certain aspects and choices in fitting of the model affect the estimates and results. The models are then assessed using AIC and superthinned residuals.

Data

GRYD currently provides services in 23 GRYD zones throughout Los Angeles (GRYD 2017 Evaluation Report). In this paper we focus on gang crimes took place during 2014-2016 in the ten GRYD zones in South Los Angeles ($87.2km^2$), which represent 6.7% of the total land of Los Angeles ($1,302km^2$) and approximately 15.5% of the total population (3.9 million). This region accounted for 45.3% of violent gang crimes in Los Angeles during 2014-2015 (Brantingham et al. 2018).

The data is collected by the Los Angeles Police Department (LAPD) as well as the City of Los Angeles Mayors Office of Gang Reduction Youth Development (GRYD). It includes all the officially reported gang crimes took place in the above region and time period, along with the crime location, time, and other details of the crime. The whole region covers 17 distinct zip codes, and the environmental covariates are incorporated at the zip code level. The covariates chosen for this study are population density 'Population density', fraction of high school graduate or higher 'Education', median income per household 'Income', and fraction of minority population 'Minority'. Such information was collected from data sources of the City of Los Angeles (www.laalmanac.com) and United States Census Bureau (www.census.gov).

In order to evaluate the effectiveness of GRYD program, we compare results of two different intervention conditions. The first condition consists of gang crimes reported to LAPD and GRYD (GRYD event), and the second condition consists of gang crimes reported only to the Los Angeles Police Department (non-GRYD event). In the data, each crime is flagged as one of the above conditions. GRYD only engages in retaliation intervention when receiving crime notifications, and during 2014-2016 approximately 30.1% of all gang related crimes were reported to GRYD within the studied area.

Methods

Hawkes model with smoothing kernel background rate

Hawkes or self-exciting point processes (Hawkes, 1971) are widely used in modeling seismicity (Ogata, 1988, 1998). It is found that such models can also be applied to effectively describe the background rate and triggering process of crime (Mohler et al. 2011), and the conditional intensity often takes the form:

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

Here, $\mu(x, y)$ is the background process which represents, in the case of gang crimes, the expected background rate of crime in absence of retaliation. The triggering process g represents the expected rate of retaliation. In order to include the two intervention condition types in the model, a multivariate Hawkes model is used in Brantingham et al. (2018), and two particular parametric equations for μ and g are chosen based on prior research on self-exciting point process models of crime (Mohler et al. 2011, 2015):

$$g(x - x_i, y - y_i, t - t_i) = \alpha(i)\omega \exp\big(-\omega(t - t_i)\big)\frac{1}{2\pi\sigma^2} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2}\big) \tag{2}$$

$$\mu(x, y) = \sum_{i=1}^{N} \frac{\beta(i)}{2\pi\sigma^2 T} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2}\big) \tag{3}$$

Here, α is the productivity, where $\alpha(i) = \alpha_1$ if event i is a GRYD event, and $\alpha(i) = \alpha_2$ if it is a non-GRYD event. The productivity is the expected number of crimes directly triggered by a single crime, and any particular crime case is expected to be an ancestor to $\alpha + \alpha^2 + \alpha^3 + ... = \frac{1}{1-\alpha} - 1$ total crimes. ω stands for the temporal decay of self-excitation, σ is the spatial triggering bandwidth, and T is the total time period in the study. β is a weight scale for the degree to which events with type i contribute to the background rate, where $\beta(i) = \beta_1$ if event i is a GRYD event, and $\beta(i) = \beta_2$ if it is a non-GRYD event. The background component μ of the above model is a smoothing kernel equation, and this model will be referred to as the smoothing kernel model in the rest of the paper.

Maximum likelihood estimation (MLE) is used to estimate the productivities and other parameters throughout this study. The optimization algorithm uses the techniques developed by Nelder and Mead (1965) to find model parameters which minimize negative log-likelihood values.

Baseline model and alternative approaches

Some aspects of the smoothing kernel model and certain choices in fitting the model are found arguable, and alternative approaches are investigated.

Firstly, for kernel smoothing methods in density estimation, it is common to use the leave-one-out method, where for every event point it smoothes over all the other points but the original point. Otherwise if it smoothes over all the points including the original point, the density estimation becomes extremely spiky at the orginal points and the results are usually undesirable (Silverman 1986). However, the smoothing kernel model has the latter approach.

To address this difference and compare both approaches, we propose an alternative leave-one-out smoothing kernel background rate as following:

$$\mu(x, y) = \sum_{i:(x_i,y_i)\neq(x,y)}^{N} \frac{\beta(i)}{2\pi\eta^2 T} \exp\big(- \frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2}\big) \tag{I}$$

Here we use a different bandwidth η in the background instead of using the same σ for both the spatial triggering bandwidth and the background bandwidth as in the smoothing kernel model. Such choice in the original model is due to stability reasons (Mohler 2014).

This model with the leave-one-out smoothing kernel background component will be used as the baseline model to be compared to the models with spatial covariates. The log-likelihood function of this model is as below:

$$l(\alpha_1, \alpha_2, \beta_1, \beta_2, \omega, \sigma, \eta) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - \sum_{i=1}^{N} \big[\beta(i) + \alpha(i)[1 - \exp\big(- \omega(T - t_i)\big)]\big] \quad (4)$$

Another aspect that we approach differently from Brantingham et al. (2018) is the way how duplicated crimes are processed. The data originally contains 2849 crime events but many of the events have the same location and time, and they are recorded as separate events because multiple victims are involved. In Brantingham et al. (2018) all the events are analyzed but we propose a different approach that the events sharing the exact same location and time will be removed as duplicates. The total number of events reduces to 2216, and we use the reduced data for analysis with spatial covariates.

Hawkes model with spatial covariates

To address how incorporating spatial covariates will affect the productivity estimates, we replace the smoothing kernel background rate μ in model (I) with linear equation of the covariates. The covariates are demeaned (substracted the sample mean) for better interpretability. The background component of the model takes the following form:

$$\mu(x, y) = \gamma_0 + \sum_{k=1}^{4} \gamma_k V_k(x, y) \quad \text{(II)}$$

where $V_k(x, y)$ is the value of covariate k (four covariates included in this study) at location (x, y), γ_k is the coefficient of covariate k, and γ_0 is the intercept term. The log-likelihood function of this model is:

$$l(\alpha_1, \alpha_2, \gamma_0 ... \gamma_4, \omega, \sigma) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - T \sum_{j=1}^{m} \exp(\gamma_0 + \sum_{k=1}^{4} \gamma_k V_{kj}) A_j - \sum_{i=1}^{N} \big[\alpha(i)[1 - \exp\big(- \omega(T - t_i)\big)]\big] \quad (5)$$

where T is the total time preiod of the study, k indexes covariate variables, j indexes the m zip codes included in the study, A_j is the area of the j^{th} zip code, and V_{kj} is the value of the k^{th} covariate variable in zip code j.

The linear relationships between the crime case count per spatial temporal unit and the covariates are shown in Figure 1. The linear model captures the relationship well with 'Population density' and moderately well with 'Income', but it does not adequately capture the relationships with the other two covariates.

We then compare model (I) and model (II) to a model with constant background rate:

$$\mu(x, y) = c \tag{III}$$

and the log-likelihood function of this model is:

$$l(\alpha_1, \alpha_2, c, \omega, \sigma) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - c * |s| * T - \sum_{i=1}^{N} \big[\alpha(i)[1 - \exp\big(-\omega(T - t_i)\big)]\big] \tag{6}$$

where $|s|$ is the area of the studied region.

Lastly, we modify the trigger function g (2) by using a distinct temporal decay parameter ω and a distinct spatial triggering bandwidth σ for each intervention condition type, and the function becomes:

$$g(x - x_i, y - y_i, t - t_i) = \alpha(i)\omega(i) \exp\big(-\omega(i)(t - t_i)\big) \frac{1}{2\pi\sigma(i)^2} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma(i)^2}\big) \tag{7}$$

where $\omega(i) = \omega_1$ if event i is a GRYD event and $\omega(i) = \omega_2$ if it is a non-GRYD event, and similarly where $\sigma(i) = \sigma_1$ if event i is a GRYD event and $\sigma(i) = \sigma_2$ if it is a non-GRYD event. We apply this modified trigger function to model (I) and model (II) to assess how it affects the estimates.

Evaluation techniques

Two main methods are applied in evaluation of the proposed models. The first one is to calculate the Akaike Information Criterion (AIC), which is calculated as $-2\cdot$log-likelihood $+\ 2p$ where p is the number of estimated parameters in the model (Ogata, 1988). AIC estimates the quality of each model relative to each of the other models, and the preferred model has lower AIC.

The models are also evaluated using superthinning. Superthinning involves both thinning the existing data points and superposing a new set of points, and it is an effective way to evaluate the fit of the model in point process analysis (Clements et al. 2013). Superthinning requires the choice of a tuning parameter b, and as the first step the existing data points are thinned where each point is randomly kept with probability $\min\left\{b/\hat{\lambda}(t), 1\right\}$. In the superposing step, a Poisson process with constant rate b is first generated over the time interval, and then each point is independently kept with probability $\max\left\{b - \hat{\lambda}(t)/b, 0\right\}$. If and only if the estimate of the conditional intensity, $\hat{\lambda}$, is correct, the resulting residual process, after superthinning, is a homogeneous Poisson process with rate b. Thus visually the resulting residual process plot can provide strong evidence regarding if the model is a good fit. Sparsity of points in the superthinned residuals corresponds to areas where the model over-predicted, whereas clustering in the superthinned residuals indicates areas where the model under-predicted the number of observed cases.

Results

Model Fitting and Estimates

The productivity estimates of the smoothing kernel model and the baseline model (model I) without removing the duplicated cases are presented in Table 1, the parameter estimates, log-likelihood, and AIC results of model (I), linear background model (model II) and constant background model (model III) with the duplicates removed are presented in Table 2, and the parameter estimates, log-likelihood, and AIC results of model (I) and model (II) with the modified g function are shown in Table 3.

Comparing the results of model (I) in Table 1 and Table 2, after removing duplicates the productivities of both GRYD and non-GRYD events increased significantly; α_1 changes from 0.1185 to 0.438 and α_2 changes from 0.2802 to 0.317. The increase in the overall productivities implies that the duplicated events are notably driving the results in Brantingham et al. (2018), and such events cause the model to distribute a larger portion of the crimes to the background process.

In Table 2, the GRYD event productivities of all three models are higher than the non-GRYD productivties. One explanation for this is that GRYD is more frequently involved at locations where crimes are already highly clustered. This point is evidenced by the fact that the GRYD productivity in model (III) is statistically significantly higher than the non-GRYD productivity. Further, the average number of crimes contained in a circle with radius of 500 meters centering at a GRYD event is 28.62, and the number for that of a non-GRYD events in the same radius is 26.28. Additionally, the average number of crimes is always higher within any given radius centering a GRYD event than that of a non-GRYD event, as shown in Figure 2.

Table 2 shows that model (I) has lower productivity estimates than model (II). This is because the estimated background rate in model (I) is non-parametric and thus fits the data more closely than the parametric estimate used in model (II), and thus a larger proportion of crimes are attributed to triggering in model (II). The temporal decay coefficient ω of all the models in Table 2 is relatively small, which implies gradual decay; it takes approximately 92 days for the triggering process to decay by 50% for the model (I), and it takes more than 300 days for the other two models.

The intercept and covariate coefficients of model (II) in Table 2 can be interpreted as following. The intercept term measures the expected number of background crimes given the average covariate values. For an area of $5.1km^2$, which is the average area of all the zip codes included in the study, with average covariate values, the expected number of background crimes per month is $(1.85E-02)*5.1*30 = 2.831$. The covariate coefficients measure the change in the expected number of background crimes when changing the values of the covariates. Take convariate 'Income' for example, all else held constant, an increase of \$3000 in yearly income is associated with a decrease of $3000*(5.01E-07)*5.1*30 = 0.23$ expected background crimes per month. The other covariate coefficients can be interpreted similarly. In general, estimates with model (II) suggest that areas with lower median income, higher population density, higher fraction of minority population, and higher fraction of high school graduate or higher tend to have more violent crimes.

To take into account the fact that GRYD events often take place at more crime clustered locations compared to non-GRYD events, the modified triggering function g is applied, in which we use a distinct temporal decay parameter ω and a distinct spatial triggering bandwidth σ for each intervention condition type. The results are shown in Table 3: while the GRYD productivity stays higher than the non-GRYD productivity for model (I), the GRYD productivity becomes lower than the non-GRYD productivity for model (II). However, the differences between the productivities for both models are not statistically significant. The above results shows that GRYD program is able to reduce gang crime retaliations with model (II), but the result is not statistically significant.

The background rate μ of model (I) and model (II) are estimated and shown in Figure 3. The overall estimated background rate of model (I), with mean of 0.036, is higher than model (II), with mean of 0.019.

Evaluation: AIC and superthinning

To evaluate the goodness of fit of the models, AIC is compared. Since a lower AIC implies better fit of the model, model (II) moderately outperforms the model (I), and model (III) shows the least preferred fit. However, the log-likelihoods are not comparable between model (I) and model (II). It is because that the integral approximation in model (I) is done over space as in Schoenberg (2013) , while no approximation is needed for model (II). As a result, estimated log-likelihoods of the models with no approximation are higher compared to models of the equal fit that require the approximation on the background rate. This could potentially explain why model (I) has a lower estimated log-likelihood and a higher AIC than model (II).

The models are also assessed using superthinning. After being thinned and superposed, the preferred model should have a residual plot close to a homogeneous Poisson process. Superthinning was applied and analyzed both spatially and temporally. Spatially, Figure

4 shows that model (I) and model (II) demonstrate roughly homogeneous residual plot, implying that both models are farily good fits. Model (III) performs poorly as the plot shows multiple areas with gaps, which means that the model is over-predicting in these areas. Temporally, Figure 5 shows that all three models have fairly good performance.

Discussion

Throughout this study, Brantingham et al. (2018) on the effectiveness of GRYD is revisited. Varied approaches to modify the model are investigated, with a special interest in how incorporating spatial covariates in the background process of the Hawkes model affects productivity estimates. It is important to note that naive estimates of the effect of GRYD interventions are biased due to the fact that GRYD events are disproportionately located in areas where violent crimes tend to cluster, as shown in Table 2 and Figure 2.

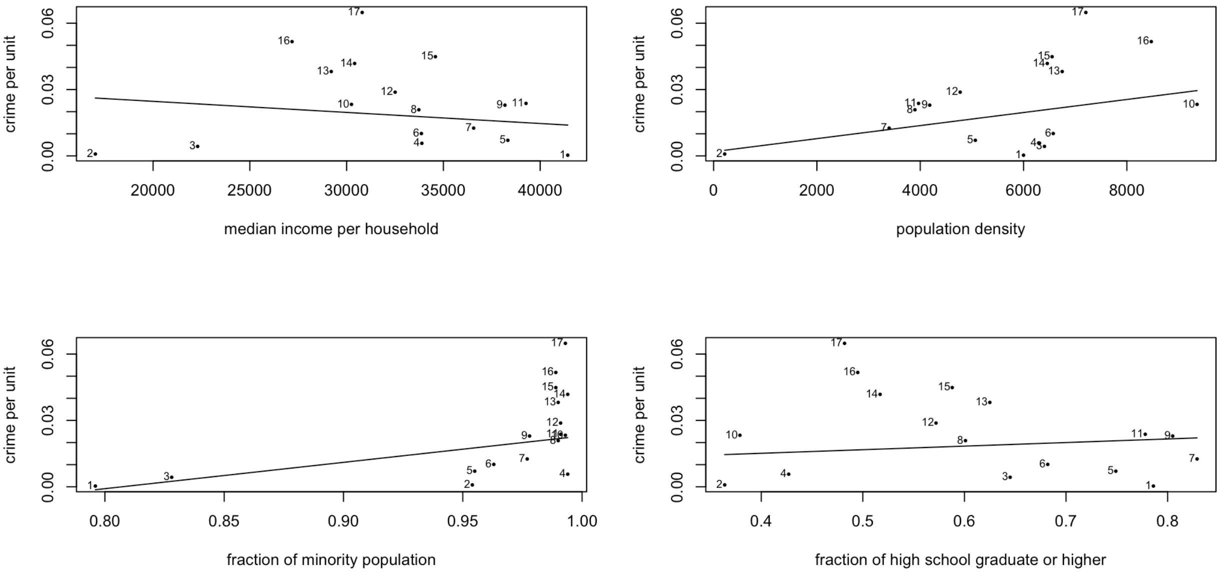
After taking this issue into account, model (II) has shown a lower GRYD productivity, implying that GRYD is able to reduce gang crime retaliations. With this model, the estimated GRYD productivity is 0.601, and the estimated non-GRYD productivity is 0.659. Based on such results, assuming there was no GRYD intervention program, one non-GRYD crime is expected to trigger $0.659+0.659^2+0.659^3+... = \frac{1}{1-0.659}-1 = 1.933$ gang crimes. One GRYD crime is expected to trigger $0.601+0.659*0.601+0.659^2*0.601+0.659^3*0.601+... = \frac{0.601}{1-0.659} = 1.762$ gang crimes. Thus, model (II) suggests that 0.171 gang crimes can be prevented with each GRYD intervention. However, the difference between the productivities is not statistically significant, and further study is required. Nonetheless, such results have shown the potential to analyze retaliatory violent crimes using self-exciting point process models with spatial covariates incorporated.

The spatial covariates included in the study are 'Population density', 'Education', 'Income', and 'Minority'. These covariates are selected because they reflect the environment and neighborhoods from varied social and economical aspects. Without covariates, the problem with the reference model (III) is that it does not take the inhomogeneity of the background into account, and it is not able to reflect the differences in background crime rate due to environmental features such as the covariates above. Model (I) takes the inhomogeneity of the background into account, but it does not do as explicitly as including the spatial covariates. Thus it is relatively difficult to tell apart if the cause of the crime is due to clustering or the inhomogeneity of the background. With spatial covariates included, model (II) is able to separate the two causes of crime explicitly.

It should be noted that on average GRYD is informed of gang crime 72 hours after the event taking place. Therefore, one potential approach for future study is to use the same productivity for the two intervention condition types within the first 72 hours, and only use different poductivities after this time window. Another disadvantage of this study is the limitation of the data. First of all, the data is collected during a 3-year period, and it is not sufficient for tracking development over time. Also, some zip codes in the studied area have extremely low count of violent crime cases. For example, zip code 90302 only has 1 violent crime, and zip code 90058 has 8 violent crimes over the three year time period. This could have led to unreliable estimates. Further more, Figure 1 shows that the linear model does not fit perfectly with the given data. The potential way of improvement could be, when more data is collected in the future, to remove the zip codes that have too extremely low or high counts of gang crimes, or apply more complex models, instead of linear model, to achieve better fit.

In conclusion, using linear model of the covariates as the background rate is proven to be a competitive approach in studying the effectiveness of GRYD intervention program and can be likely extended to similar studies of violent crime with self-exciting point process models.

Figure 1: Linear relationships between the number of crimes and covariates



The solid lines stand for the linear relationships. The x-axis indicates the values of the covariates, and the y-axis indicates the values of the number of crimes per spatial temporal unit. Note that the covariates are incorportaed at the zip code level, and the zip codes are numbered from 1 to 17 based on the number of crimes took place, with 1 being the zip code with the lowest number of crimes

Table 1: Before removing duplicates: productivity estimates

	Smoothing kernel model	**Baseline model**
GRYD productivity α_1	0.1175	0.1185
	(0.0145)	(0.0145)
Non-GRYD productivity α_2	0.2815	0.2802
	(0.0149)	(0.0149)

The standard errors of the productivity estimates are in parentheses.

Table 2: Parameter estimates, Log-Likelihood, and AIC

	Baseline model	**Linear background model**	**Constant background model**
GRYD productivity α_1	0.438	0.627	0.811
	(0.078)	(0.072)	(0.069)
Non-GRYD productivity α_2	0.317	0.623	0.650
	(0.052)	(0.065)	(0.051)
GRYD smoothing weight β_1	0.691		
	(0.051)		
Non-GRYD smoothing weight β_2	0.668		
	(0.040)		
Temporal decay ω	0.0075	0.0023	0.0021
	(0.0013)	(0.0004)	(0.0002)
Spatial triggering bandwidth σ	0.432	0.074	0.071
	(0.031)	(0.004)	(0.003)
Background smoothing bandwidth η	0.071		
	(0.003)		
Intercept γ_0		1.85E-02	
		(6.899E-04)	
Income γ_1		-5.01E-07	
		(1.16E-07)	
Population density γ_2		2.95E-06	
		(1.76E-07)	
Minority γ_3		0.119	
		(0.006)	
Education γ_4		1.62E-02	
		(5.28E-03)	
Constant background c			0.013
			(0.001)
Log-Likelihood	-9869.603	-9798.459	-10080.61
AIC	19753.21	19614.92	20171.22
Percentage background	0.677	0.602	0.561

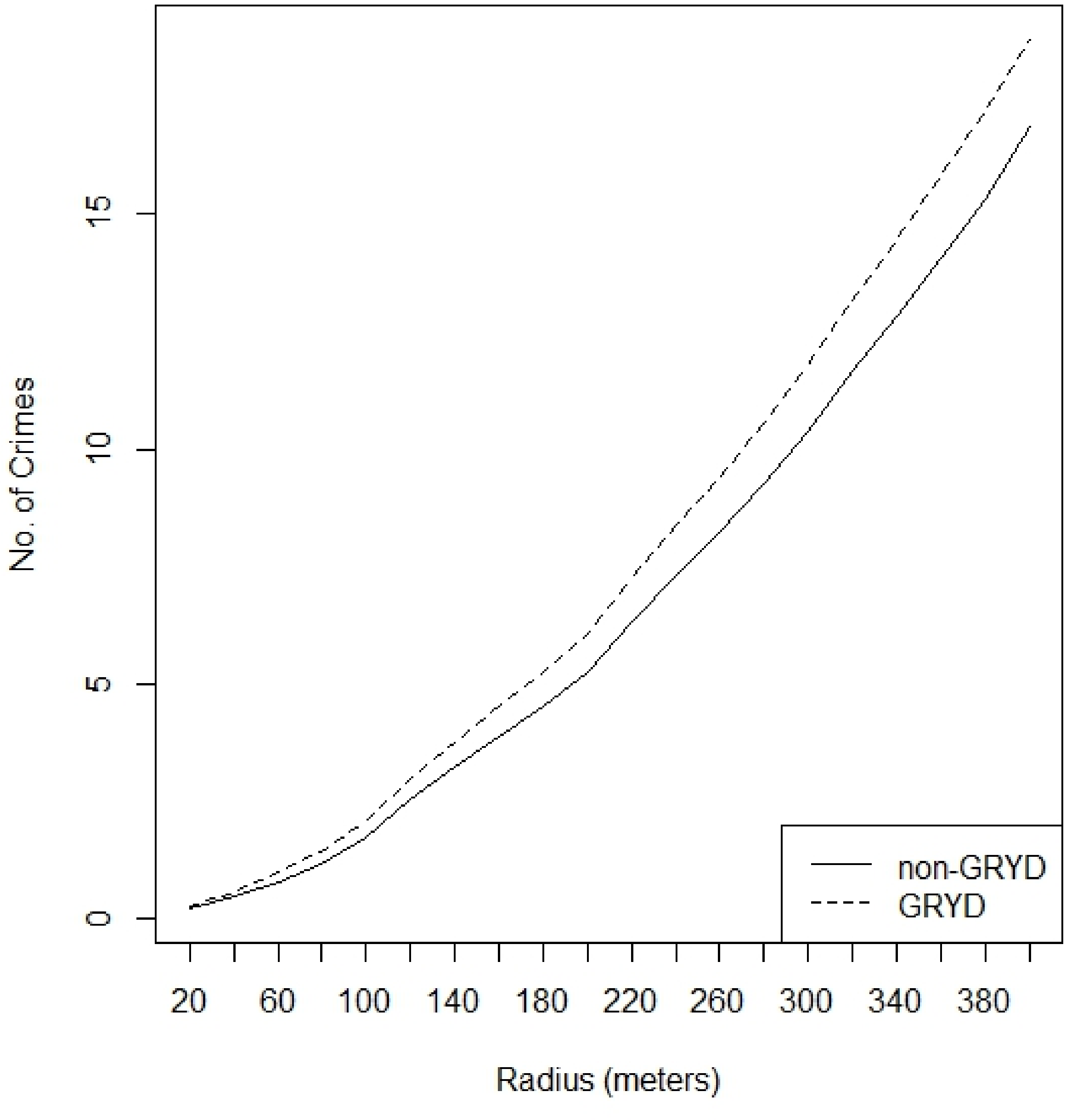
The standard errors of the parameter estimates are in parentheses.

Table 3: Parameter estimates, Log-Likelihood, and AIC with modified g function

	Baseline model	**Linear background model**
GRYD productivity α_1	0.401	0.601
	(0.088)	(0.110)
Non-GRYD productivity α_2	0.347	0.659
	(0.064)	(0.082)
GRYD smoothing weight β_1	0.691	
	(0.051)	
Non-GRYD smoothing weight β_2	0.668	
	(0.040)	
GRYD temporal decay ω_1	0.0092	0.0019
	(0.0033)	(0.0006)
Non-GRYD temporal decay ω_2	0.0063	0.0028
	(0.0018)	(0.0006)
GRYD spatial triggering bandwidth σ_1	0.419	0.059
	(0.057)	(0.005)
Non-GRYD spatial triggering bandwidth σ_2	0.436	0.090
	(0.052)	(0.006)
Background smoothing bandwidth η	0.071	
	(0.003)	
Intercept γ_0		1.84E-02
		(6.738E-04)
Income γ_1		-6.85E-07
		(1.22E-07)
Population density γ_2		3.06E-06
		(1.71E-07)
Minority γ_3		0.137
		(0.008)
Education γ_4		2.25E-02
		(5.40E-03)
Log-Likelihood	-9869.232	-9792.285
AIC	19756.46	19604.57
Percentage background	0.677	0.587

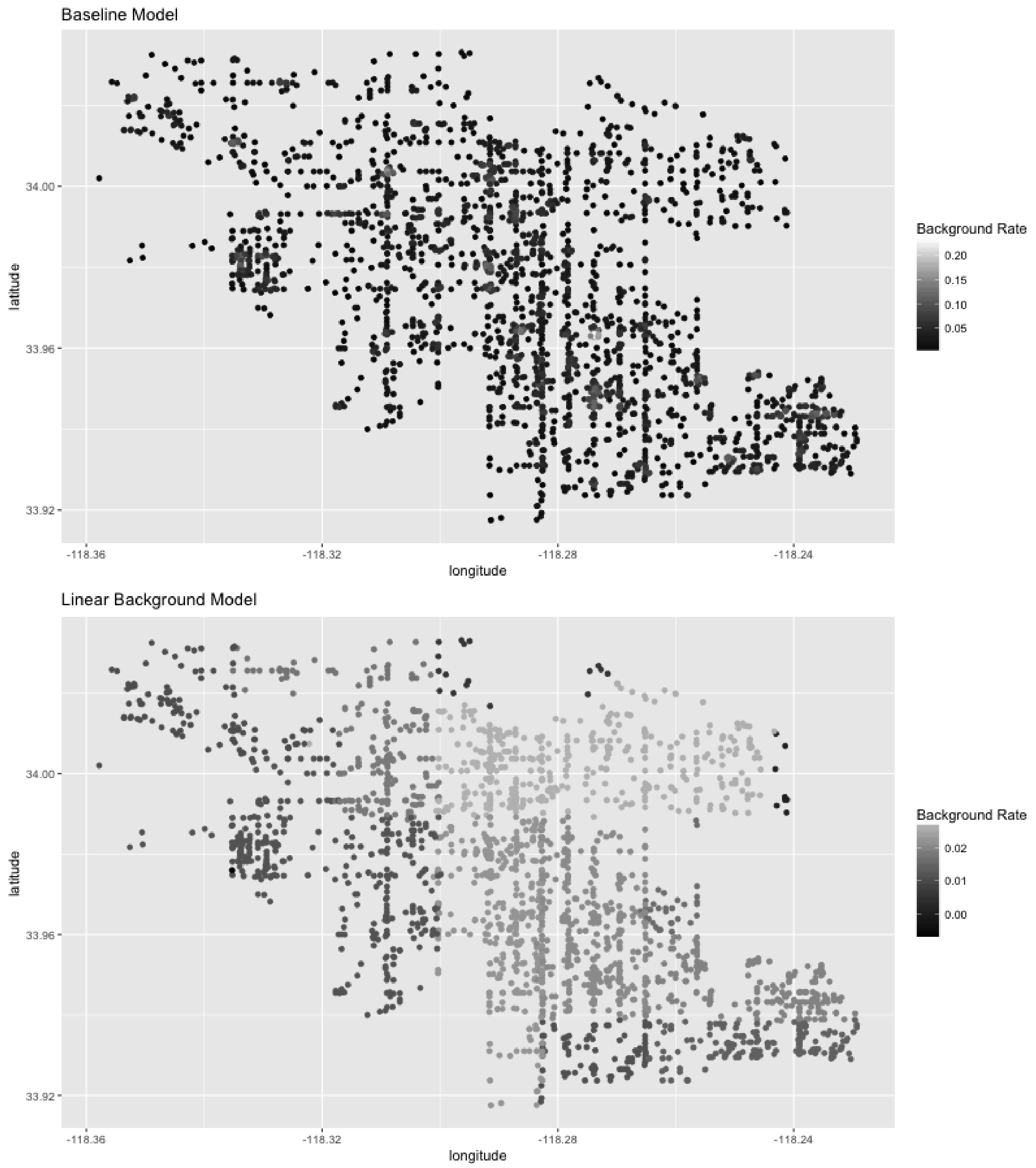
The standard errors of the parameter estimates are in parentheses.

Figure 2: Average number of crimes contained in circle of X meters



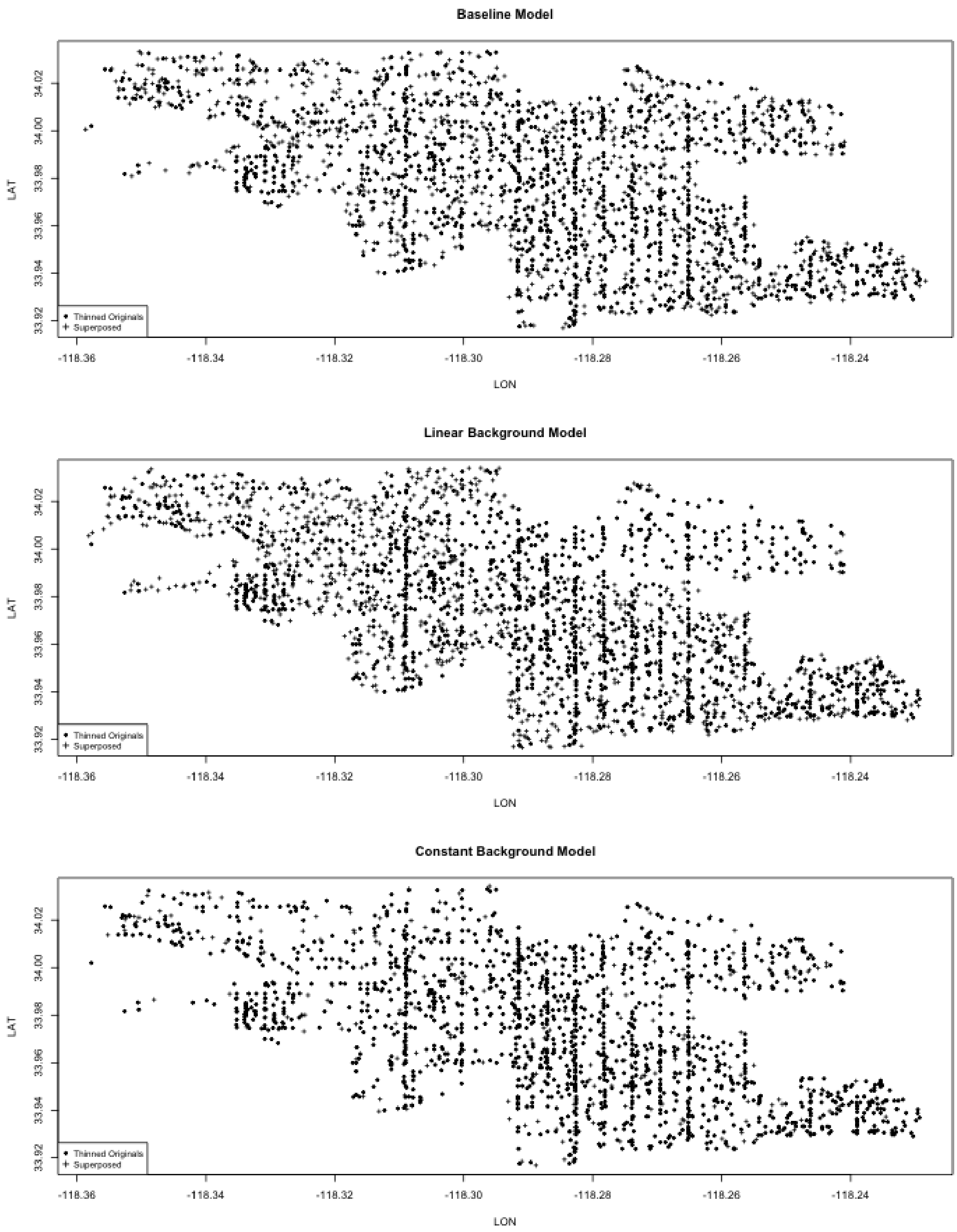
Dashed curves stand for the average number of crimes in circle of X meters centered at a GRYD event, and solid curves stand for that of a non-GRYD event. The x-axis indicates the values of the radius of the circle, and the y-axis indicates the values of the average number of crimes.

Figure 3: Background rate estimates



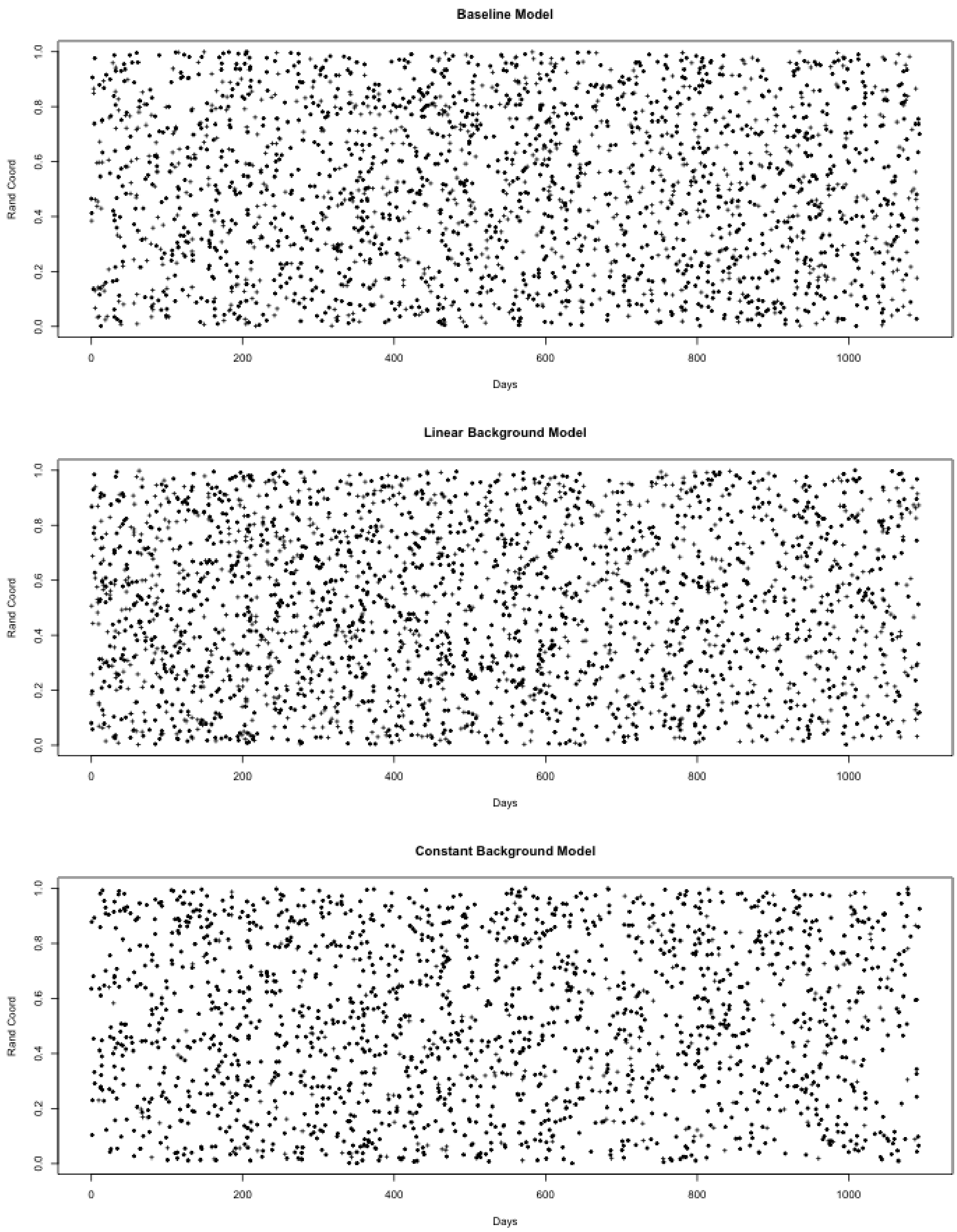
Every point stands for one crime event and the grey scale of the point represents the background rate estimated at the crime event location. The x-axis and y-axis indicates the longitude and latitide of the point respectively.

Figure 4: Superthinning (Spatial)



Thinned original points are marked with dots, and superposed points are marked with plus signs. The x-axis and y-axis indicates the longitude and latitide of the point respectively.

Figure 5: Superthinning (Temporal)



Thinned original points are marked with dots, and superposed points are marked with plus signs. The x-axis indicates the timeline of the events, and the y-axis indicates a randomly generated coordinate.

REFERENCES

[BSY17] P. J. Brantingham, N. Sundback, and B. Yuan. "GRYD Intervention Incident Response Gang Crime 2017 Evaluation Report." 2017.

[BYS18] P. J. Brantingham, B. Yuan, and N. Sundback. "Does Violence Interruption Work?" 2018.

[CSV13] Robert Alan Clements, Frederic Paik Schoenberg, and Alejandro Veen. "Evaluation of space-time point process models using super-thinning." *Environmetrics, 23(7), 606-616*, 2013.

[CT99] J. Cohen and George Tita. "Diffusion in Homicide: Exploring a General Method for Detecting Spatial Diffusion Processes." *Journal of Quantitative Criminology, 15(4): 451–493*, 1999.

[Haw71] Alan G. Hawkes. "Spectra of some self-exciting and mutually exciting point processes." *Biometrika, Volume 58, Issue 1*, 1971.

[Moh14] George Mohler. "Marked point process hotspot maps for homicide and gun crime prediction in chicago." *International Journal of Forecasting 30(3):491–497*, 2014.

[MSB11] G. O. Mohler, M. B. Short, and P. J. Brantingham. "Self-exciting point process modeling of crime." *Journal of The American Statistical Association, 106(493):100–108*, 2011.

[MSM15] G. O. Mohler, M. B. Short, and Sean Malinowski. "Randomized controlled field trials of predictive policing." *Journal of the American Statistical Association 110(512):1399–1411*, 2015.

[NM65] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization." *The Computer Journal, 7(4): 308-313*, 1965.

[Oga88] Yosihiko Ogata. "Statistical models for earthquake occurrences and residual analysis for point process." *Journal of the American Statistical Association, 83: 9-27*, 1988.

[Oga98] Yosihiko Ogata. "Space-Time Point-Process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics, 50: 379-402*, 1998.

[RG17] A. Reinhart and J. Greenhouse. "Self-exciting point processes with spatial covariates: modeling the dynamics of crime." 2017.

[Sch13] Frederic Paik Schoenberg. "Facilitated Estimation of ETAS." *Bulletin of the Seismological Society of America,103(1): 601-605*, 2013.

[Sil86] Bernard. W. Silverman. "Density Estimation For Statistics and Data Analysis." 1986.

[SMB14] M. B. Short, G. O. Mohler, and P. J. Brantingham. "Gang Rivalry Dynamics via Coupled Point Process Networks." *Discrete Continuous Dynamical Systems-Series B 19(5)*, 2014.

[TR07] George Tita and Greg Ridgeway. "The Impact of Gang Formation on Local Patterns of Crime." *Journal of Research in Crime and Delinquency 44(2):208–237*, 2007.