UCLA
UCLA Electronic Theses and Dissertations

Title
Advances in Non-Parametric Spatial Temporal Point Process Models with Applications to Crimes and Infectious Diseases

Permalink
https://escholarship.org/uc/item/6xj9c3g7

Author
Park, Junhyung

Publication Date
2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

Advances in Non-Parametric Spatial Temporal Point Process Models with Applications to Crimes and Infectious Diseases

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Statistics

by

Junhyung Park

2020

© Copyright by
Junhyung Park
2020

ABSTRACT OF THE DISSERTATION

Advances in Non-Parametric Spatial Temporal Point Process Models with Applications to Crimes and Infectious Diseases

by

Junhyung Park
Doctor of Philosophy in Statistics
University of California, Los Angeles, 2020
Professor Frederic Paik Schoenberg, Chair

In recent decades there has been tremendous growth in new statistical methods and applications for modeling random events occurring in space and time. There is a positive outlook on the demand for research in this field for the coming decades as space- and time-referenced event data will become more commonly available and high in size and resolution.

This dissertation makes methodological contributions in non-parametric and semi-parametric self-exciting point processes models and their application to infectious disease spread and quantitative criminology. For infectious diseases we demonstrate that point process models can be an effective tool for real-time descriptions and forecasts of an outbreak by comparing its performance to traditional compartmental models. We introduce a purely infection-driven, non-stationary point process model and its estimation. We propose a non-parametric implementation of the Recursive Hawkes model. In dealing with gang crime event data, we address the long standing challenge of distinguishing spatial and temporal inhomogeneity with true triggering, as well as evaluating event-based treatments that are non-randomized due to practical and ethical considerations. To this end, we propose a new method to non-parametrically incorporate spatial covariates in the background rate of crimes. We also introduce a sub-sampling procedure to evaluate non-randomized, clustered treatments in order to generate synthetic controls to improve causal interpretation. We assess this procedure with simulation studies.

The dissertation of Junhyung Park is approved.

Mark Handcock

Chad Hazlett

Jeff Brantingham

Frederic Paik Schoenberg, Committee Chair

University of California, Los Angeles

2020

The heart of man plans his way, but the Lord establishes his steps.

Proverbs 16:9

The stone the builders rejected has now become the cornerstone.

Psalm 118:22

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Where I am and what I have seems nothing more than the product of time and the inspiration, grace and generosity of the following people. Prof. **Dukpa Kim** showed me the value of rigor, encouraged me to study probability and statistics deeply and pursue a Ph.D. The lucid and legendary teaching style of Prof. **Kenneth Elzinga** and his mentorship to so many students including myself inspired me to become a scholar and teacher. My manager at the IMF, Dr. **Paolo Mauro**, showed me the true meaning of strength and professionalism. I thank Prof. **Joseph Ostroy** for bringing me to Los Angeles with first year funding as well as supporting my transition out of the Economics Department with TA-ships in the interim; S.A.O. **Jessica Perez** was instrumental in facilitating everything. I thank Profs. **Rob Weiss** and **Tom Belin** for bringing me to the Department of Biostatistics. I am grateful to my master's advisor Prof. **Weng Kee Wong**, who at times felt like my only ally during a strange time in my journey. I am grateful to Profs. **Paavo Monkkonen** and **Michael Lens** for offering me a research position in my time of need as I transitioned from Biostatistics to Statistics. I thank Prof. **Honquan Xu** for seeing potential in me and bringing me to the Department of Statistics. S.A.O. **Glenda Jones** was always there to take care of me throughout the way.

I thank my teaching mentors Profs. **Mike Tsiang**, **Miles Chen**, **Vivian Lew**, **Nicolas Christou**, and especially **Rob Gould** for advising me during the job market trail and working with me in my duties as Head TA. I thank Prof. **Ryan Harrigan** and Dr. **Dan Kelly** for being a collaborator and domain expert in our work on infectious diseases. Many thanks to **Adam Chaffee** who implemented the SEIR portion of our work on Ebola in Chp. 3 and supporting us to the bitter end of the publication process when many would have dropped off the radar after landing a position in Silicon Valley. I also appreciate **Andy Kaplan** for helping with implementing the idea presented in Sec. 5.3. I am so grateful for Profs. **Andrea Bertozzi** and **Jeff Brantingham** for the funding and invaluable opportunity to work on LAPD crime data. This boosted my productivity and made my dissertation more interesting - both of which contributed to being awarded the Dissertation Year Fellowship. I thank

committee members Profs. **Mark Handcock**, **Chad Hazlett**, **Yingnian Wu** and **Jeff Brantingham** for their time, useful comments and suggestions that enriched my work. I thank my predecessor Dr. **Josh Gordon** for showing me the ropes on the computational side as I got acquainted to working with point processes. I am also indebted to other predecessors like Prof. **Eric Fox** and Dr. **Alejandro Veen**, whose work was very influential to me. I also would like to thank the faculty and staff at the **US Naval Academy** for welcoming me to their family and giving me a career to look forward to as I finished the last stretch of my degree.

I was sustained by the moral support of several that are dear to me. Biostatistics colleague and dear friends Dr. **Priscilla Yen** and her husband **Simp Wong** were always there for every hurdle and dilemma in my journey and helped me brave my fear of lack of parking and unruly mobs during many Costco trips. The virtual comradery, stress-relief, encouragement, and commiseration of my childhood friend Dr. **Bill Diplas** (aka `Vasilis`). My sister, (future Dr.) **Sophia Park** showed me that someone with the same DNA as me can have discipline. I thank my father and mother, Dr. **Taigyoo Park** and **Giyoung Park** who went through so much hardship as first generation immigrants and selflessly sacrificed everything for us to settle in the States. They taught me to value faith, persistence, education and hard work. I am so blessed to have the love and support of my wife Dr. **Esther Baik**, who accepts me in all my flaws and has endured so much for all these years. I am thrilled by the arrival of our daughter, baby **Evelyn** - you are the light at the end of a 9 year long tunnel and mark the beginning of a new phase in my life.

Finally, I am forever indebted to Prof. **Frederic Paik Schoenberg**, without whom none of this work would have been possible. I am so proud that he is my advisor and academic father. He has been instrumental to my development as a statistician; allowing me the right balance of autonomy and structure, treating me with so much respect and professionalism, and teaching me to seek simplicity, clarity, innovation and pragmatism in the research process. I will never forget his dedication to my success. All my flaws, mistakes and shortcomings are my own.

Chapter 2 contains work that appears as Kelly, Park, Harrigan, Schoenberg et al. (2019)

in Epidemics. Chapter 3 contains work forthcoming as Park, Chaffee, Harrigan & Schoenberg (2020) in the Journal of Applied Statistics. Chapter 4 contains work funded by the City of Los Angeles contract number C-132202, forthcoming as Park, Schoenberg, Bertozzi & Brantingham (2020) in the Journal of the American Statistical Association. Chapter 5 contains preliminary work to be published, tentatively titled 'Event-Based Interventions, Synthetic Controls and the Law of Crime Concentration' (joint with P.J. Brantingham and F.P. Schoenberg), 'Nonparametric Estimation of Recursive Point Processes with Application to Mumps in Pennsylvania' (joint with A. Kaplan and F.P. Schoenberg) and 'A Purely Epidemic Non-Stationary Point Process Model of Infectious Disease Spread'. The final year of my program was funded by the UCLA Dissertation Year Fellowship.

VITA

2009	B.A. *with Distinction*, Economics, Mathematics (minor), University of Virginia, Charlottesville
2009-2011	Research Assistant, International Monetary Fund, Washington D.C.
2012	M.A. Economics, University of California, Los Angeles
2012-2014	Teaching Assistant, Grader, Department of Economics, UCLA
2014	Teaching Assistant, Department of Biostatistics, UCLA
2014	M.S. Biostatistics, University of California, Los Angeles
2014–2015	Research Assistant, Department of Urban Planning, UCLA
2015–2017	Teaching Assistant, Department of Statistics, UCLA
2017–2018	Head TA, Department of Statistics, UCLA
2018	C.Phil. Statistics, University of California, Los Angeles
2017–2019	Graduate Student Researcher, Department of Mathematics, UCLA
2019–2020	Dissertation Year Fellow, University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Park, J., Schoenberg, F.P., Bertozzi, A. and Brantingham, P.J., "Investigating Clustering

and Interventions in Gang-Related Violent Crime Data using Spatial-temporal Point Processes with Covariates," (2020) *JASA*, forthcoming.

Park, J., Chaffe, A., Harrigan, R. and Schoenberg, F.P., "A Non-Parametric Hawkes Model of the Spread of Ebola in West Africa," (2020) *Journal of Applied Statistics*, forthcoming.

Kelly, J.D., **Park, J.**, Schoenberg, F.P., Harrigan, R., et al. "Real-time Predictions of the 2018-2019 Ebola Virus Disease Outbreak in the Northeastern Region of Democratic Republic of Congo using Hawkes Point Process Models," (2019) *Epidemics*, 28, 100354.

Park, J.,"Some Advances in Non-Parametric Modeling of Spatio-Temporal Point Processes," *Wake Forest University & US Naval Academy*, December 2019.

Park, J.,"Some Advances in Non-Parametric Modeling of Spatio-Temporal Point Processes," *UCLA Dept. of Statistics Seminar Series*, November 2019.

Park, J.,"Point Process Applications in Crime & Disease Spread with Proposed Non-Parametric Extensions," *Joint Statistical Meetings*, Vancouver, Canada, July 2018.

CHAPTER 1

Introduction

Some effort was made here to avoid overlap and redundancy with the many great overviews and introductions to point processes that have been written in the dissertations of former students working under Professor Schoenberg. Chapters 1 through 3 of Peng (2003) is a notable example. The reader is referred to Reinhart (2018) for an excellent survey of the literature on methods and applications. The following short note in Section 1.1 is an elementary primer that the author hopes will be useful for students approaching this area for the first time. Section 1.2 is a brief overview of the methodology in Hawkes models which were a foundational starting point for the author's research. Section 1.3 outlines the main contributions and organization of the remaining chapters.

1.1 A Brief Introduction to Point Processes

A spatio-temporal point process $\{p_1, p_2, \dots p_N\}$ is a random collections of points that represent the times and locations of events. Each point p_i is a vector $(t_i, \mathbf{x}_i')$, where t_i is the occurrence time and $\mathbf{x}_i \in \mathbb{R}^m$ is the spatial coordinate of point p_i. Most often $m = 2$ to represent the longitude and latitude of an event. Examples of events that can be modeled as point processes are 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), crimes (Mohler et al. 2011), email networks (Fox et al. 2016) and terrorist attacks (Porter and White, 2012).

To introduce the conceptual framework of point processes, it is easiest to consider a purely temporal point process which lacks $\mathbf{x}_i$, and is just an ordered collection of occurrence

times $\{t_1, t_2, \ldots t_N\}$. Its associated counting process $N(t)$ counts the number of events that have occurred up to and including time t. The number of events occurred over an interval is represented by $N(t_j, t_k) = N(t_k) - N(t_j)$ where $t_k > t_j$.

The simplest form of point process is the homogeneous Poisson process. Such a process with rate λ satisfies the following:

(a) For any interval $(t, t + \Delta t], N(t, t + \Delta t) \sim Poisson(\lambda \cdot \Delta t)$.

(b) For any non-overlapping intervals $(t, t + \Delta t]$ and $(u, u + \Delta u]$, the random variables $N(t, t + \Delta t)$ and $N(u, u + \Delta u)$ are independent.

Since the expectation of a Poisson random variable is equal to its parameter, the first condition states that the number of events over an interval scales linearly with the length of that interval, and because this does not depend on t, we say the process is stationary. Notice that the total number of events in the two non-overlapping intervals in condition (b) are Poisson distributed with the sum of the two rates, $\lambda(\Delta t + \Delta u)$. To see this, let X and Y be independent Poisson random variables with parameters λ_x and λ_y. The mass function for the random variable $Z = X + Y$ is,

$$\begin{aligned}
\Pr(Z = n) &= \sum_{k=0}^{n} \Pr(X = k) \Pr(Y = n - k) \\
&= \sum_{k=0}^{n} \frac{\lambda_x^k e^{-\lambda_x}}{k!} \frac{\lambda_y^{n-k} e^{-\lambda_y}}{(n-k)!} \\
&= \frac{e^{-(\lambda_x + \lambda_y)}}{n!} \sum_{k=0}^{n} \frac{n!}{n!(n-k)!} \lambda_x^k \lambda_y^{n-k} \\
&= \frac{e^{-(\lambda_x + \lambda_y)}}{n!} (\lambda_x + \lambda_y)^n,
\end{aligned}$$

where the last equality uses the binomial theorem.

In order to define a Poisson process whose rate *does* depend on t, we partition the observation interval into K regular bin sizes of Δt such that $(0, T] = (0, t_1, \ldots, t_{K-1}, t_K]$ and $t_k - t_{k-1} = \Delta t$, where $k = 1, \ldots, K$. If we assume that the Poisson rate may be vary for different bins of time, then the distribution of the number of events in any bin is

$N(t_k, t_k + \Delta t) \sim Poisson(\lambda(t_k)\Delta t)$. Now, write the random number of events occurring over an interval $(a, b]$ as $N(a, b) = \lim_{n\to\infty} \sum_{i=1}^{n} N(a + (i-1)\Delta t, a + i\Delta t)$, where $\Delta t = (b-a)/n$. Because we have shown that the sum of independent Poisson random variables is also Poisson with a rate as the sum of all rates, $N(a, b)$ must be Poisson distributed with parameter $\lim_{n\to\infty} \sum_{i=1}^{n} \lambda(a + i\Delta t)\Delta t$. Notice that this is the definition for the Riemann integral of $\lambda(t)$ over the range (a,b]. Therefore, an inhomogeneous Poisson process with rate $\lambda(t)$ satisfies:

(a) For any interval $(a, b], N(a, b) \sim Poisson(\int_a^b \lambda(t)dt)$.

(b) For any non-overlapping intervals (a, b) and (c, d), the random variables $N(a, b)$ and $N(c, d)$ are independent.

The quantity $\lambda(t)$ can be viewed as an instantaneous rate of an occurrence in a moment of time. To see this, we can make the bin sizes Δt small enough that the probability of observing more than one event in any bin becomes negligibly small, and the distribution of $N(t_k, t_k + \Delta t)$ in the limit becomes a Bernoulli distribution with probability of success $p_k = \lambda(t_k)\Delta t$. Let $o(\Delta t)$ be a function of Δt that goes to zero faster than Δt. Then the probability of observing no event, one event, and more than one event in any k^{th} bin is,

$$\begin{aligned}
\Pr(N(t_k, t_k + \Delta t) = 0) &= \exp(-\lambda(t_k)\Delta t) \\
&= 1 - \lambda(t_k)\Delta t + o(\Delta t) \\
\Pr(N(t_k, t_k + \Delta t) = 1) &= \lambda(t_k)\exp(-\lambda(t_k)\Delta t) \\
&= \lambda(t_k)\Delta t(1 - \lambda(t_k)\Delta t + o(\Delta t)) \\
&= \lambda(t_k)\Delta t + o(\Delta t) \\
\Pr(N(t_k, t_k + \Delta t) = c) &= \frac{(\lambda(t_k)\Delta t)^c e^{-\lambda(t_k)\Delta t}}{c!} = o(\Delta t), \quad c > 1
\end{aligned}$$

where the last equalities for each expression uses the Taylor expansion of $e^{-\lambda(t_k)\Delta t}$. So multiplying $p_k = \lambda(t_k)\Delta t$ by Δt and taken to the limit, we see that the Poisson rate function $\lambda(t) = \lim_{\Delta t \downarrow 0} \Pr(N(t, t + \Delta t) = 1)/\Delta t$ is the instantaneous expected rate of observing an event in a moment of time.

Because the count of events across different intervals are independent of each other, the Poisson process described above would not be appropriate to model processes that have a history dependent structure. A more general mathematical construct that allows for this is the conditional intensity, which is the infinitesimal rate at which an expected number of points are accumulating at time t, given the history $\mathcal{H}_t$ of all points occurring prior to time t (Daley & Vere-Jones, 2003)

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

To derive the probability distribution of the an event time t_i, suppose we observe the $(i-1)^{\text{th}}$ event at time t_{i-1} and its preceding history $\mathcal{H}_{t_{i-1}}$. The probability that the next event time t_i is greater than some time t is simply the probability that there are no events between t_{i-1} and t. Therefore

$$\Pr(t_i > t|\mathcal{H}_{t_{i-1}}) = \exp\Big(-\int_{t_{i-1}}^{t} \lambda(s|\mathcal{H}_t)ds\Big),$$

and the CDF of the i^{th} event time is

$$\Pr(t_i \leq t|\mathcal{H}_{t_{i-1}}) = 1 - \exp\Big(-\int_{t_{i-1}}^{t} \lambda(s|\mathcal{H}_t)ds\Big).$$

The PDF of the i^{th} event time is found by differentiating the CDF,

$$f_{t_i}(t|\mathcal{H}_{t_{i-1}}) = \frac{d}{dt}\left(1 - \exp\Big(-\int_{t_{i-1}}^{t} \lambda(s|\mathcal{H}_t)ds\Big)\right) = \lambda(t|\mathcal{H}_t)\exp\Big(-\int_{t_{i-1}}^{t} \lambda(s|\mathcal{H}_t)ds\Big).$$

Intuitively, this PDF can be seen to have two parts: $\exp(-\int_{t_{i-1}}^{t} \lambda(s|\mathcal{H}_t)ds)$ gives the probability of observing no events between times t_{i-1} and t, and $\lambda(t|\mathcal{H}_t)$ characterizes the instantaneous rate at which an event occurs right at time t.

The joint density of observing a sequence of events $(t_1, \ldots, t_N)$ over an observation window

$(0, T]$ is then,

$$
\begin{aligned}
f(t_1, \ldots, t_N) &= \prod_{i=1}^{N} \big(f(t_i|\mathcal{H}_{t_i})\big) \Pr(N(t_N, T) = 0) \\
&= \prod_{i=1}^{N} \lambda(t_i|\mathcal{H}_{t_i}) \exp\Big(-\int_{t_{i-1}}^{t_i} \lambda(s|\mathcal{H}_t)ds\Big) \exp\Big(-\int_{t_N}^{T} \lambda(s|\mathcal{H}_t)ds\Big) \\
&= \prod_{i=1}^{N} \lambda(t_i|\mathcal{H}_{t_i}) \exp\Big(-\int_{0}^{T} \lambda(s|\mathcal{H}_t)ds\Big).
\end{aligned}
$$

The above concepts are readily extended to include a spatial dimension. The conditional intensity of a space-time would be:

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

where $N(\ldots)$ is the number of events in a small space-time volume, and $\mathcal{H}_t$ contains the times, locations and associated marks/covariates of events occurring before time t. With a parametrized model for $\lambda(t, x, y)$, the log-likelihood of an observed sequence $\{(t_i, x_i, y_i); i = 1, \ldots, n\}$ over an interval $[0, T]$ and region A is,

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

The first term of the log-likelihood is straightforward to compute while the integral term must be numerically approximated. Ogata (1978) showed that under general conditions the MLE is consistent, asymptotically unbiased and efficient. Maximizing the likelihood requires numerical evaluation and can become very slow and numerically unstable depending on the complexity of λ. In light of this, Schoenberg (2013) gave an analytical approximation to the integral term in the log-likelihood that is exact if all after shock events are contained within the observation region. Alternatively, Veen and Schoenberg (2008) showed the likelihood can be maximized as an EM algorithm by introducing a latent variable u_i for each event i to indicate whether the event came from the background process ($u_i = 0$) or was triggered by a previous event j ($u_i = j$). This EM procedure reduces issues encountered with numerical

stability when directly maximizing the likelihood, but problems can still arise (see Mohler (2014) p. 494). Nonetheless this has become a popular way to fit parametric Hawkes models in the literature.

1.2 Overview of Hawkes Models

Several model formulations of λ have been proposed in the literature. An important class of point processes is called self-exciting, where the occurrence of an event increases the likelihood of observing another event nearby in time and space. An important parametric temporal model was proposed by Hawkes (1971):

$$\lambda(t) = \mu(t) + \sum_{i:t_i<t} g(t - t_i) = \mu(t) + \theta \sum_{i:t_i<t} \omega e^{-\omega(t-t_i)}.$$

In this model, the background process $\mu(t)$, which may or may not depend on time, generates 'main shock' events, and each of these contributes a kernel $\omega e^{\omega(t-t_i)}$ that generates offspring or 'after shock' events. Typically the triggering function g is constrained to be a density, so the parameter θ represents the expected number of new events directly attributable to each event. Since each event is posited to cause an expected number θ of additional events, any particular event is expected to be an ancestor to $\theta + \theta^2 + \theta^3 + ... = \frac{1}{1-\theta} - 1$ total events. Thus θ should satisfy $0 \leq \theta < 1$ in order for the process to be stable. A large ω means that offspring events are close to the main event and decay quickly, while a small ω corresponds to after shocks that are more persistent through time. Ogata (1988) incorporated marks into the conditional intensity,

$$\lambda(t) = \mu + \sum_{i:t_i<t} g(t - t_i) e^{\beta(m_i - M_r)}$$

where earthquakes of large magnitude have more productivity in the number of aftershocks through $e^{\beta(m_i-M_r)}$. Ogata (1998) extended this to incorporate both space and time with the

general form:

$$\begin{aligned}\lambda(t,x,y) &= \mu(x,y) + \sum_{i:t_i<t} g(t-t_i, x-x_i, y-y_i; m_i)] \\ &= \mu(x,y) + \sum_{i:t_i<t} \kappa(m_i) g(t-t_i) f(x-x_i, y-y_i)]\end{aligned}$$

and fitted various forms of g, κ and f.

A few methods have been introduced to estimate μ, κ, g, and f without assuming any parametric form. Marsan and Lengliné (2008) proposed Model Indepedent Stochastic Declustering (MISD), which is an iterative non-parametric method to estimate the triggering function and background rate as a piecewise constant step function by exploiting the cascading structure of main shocks and after shocks. Central to these methods is the probabilistic branching structure of earthquakes which is given by a lower triangular probability matrix,

$$p_{ij} = \begin{cases} \text{probability event } i \text{ was triggered by event } j, & j < i \\ \text{probability event } i \text{ is a background event}, & i = j \\ 0, & j > i \end{cases}$$

$$P = \begin{bmatrix} p_{11} & 0 & 0 & \cdots & 0 \\ p_{21} & p_{22} & 0 & \cdots & 0 \\ p_{31} & p_{32} & p_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NN} \end{bmatrix} \tag{1.1}$$

with the constraint that each row must sum to one since each event is either a main shock or an aftershock of a previous event. The sum of the diagonals and off-diagonals is in effect, the estimated number of main shocks and aftershocks, respectively. Given a correct model of $\lambda(t)$, the probability that a point t_j occurred exogenously due to the background rate is

$$p_{jj} = \frac{\mu}{\lambda(t_j)}, \tag{1.2}$$

and the probability that point t_i triggered point t_j is

$$p_{ij} = \frac{g(t_j - t_i)}{\lambda(t_j)}. \tag{1.3}$$

An initial guess, $p_{jj}^{(0)}$ and $p_{ij}^{(0)}$ where $1 \leq i < j \leq T$, gives the full probabilistic branching structure of the point process. This can be used to obtain initial estimates $\mu^{(0)}$ and $g(u)^{(0)}$ in two basic ways; a probability weighted non-parametric density estimation procedure using $\{(t_k, x_k, y_k, p_{kk})\}_{k=1}^{N}$ for μ and $\{(t_j - t_i, x_j - x_i, y_j - y_i, p_{ij})\}_{i<j}$ for g (see Mohler (2011) or Zhuang (2002)), or a histogram estimator can be used as proposed by Marsan and Lengliné (2008). Knowing these in turn allows updated probabilities $p_{jj}^{(1)}$ and $p_{ij}^{(1)}$ to be computed using (1.2) and (1.3). This is iterated until the largest update in any $p_{jj}^{(k)}$ or $p_{ij}^{(k)}$ is less than some small constant ϵ. Fox et al. (2016) extended this method to have a spatially varying background rate and a triggering function that is separable in time and space (*i.e.*, $h(t, x, y) = g(t)f(x, y)$), while Gordon et al. (2018) further allowed the spatial distribution of aftershocks to depend on magnitude and angular separation from the mainshock's local fault line.

The author's initial research approach was to extend MISD in order to apply non-parametric Hawkes processes to disease spread and crime. The larger goal was to contribute to the line of methodological literature stemming from Hawkes (1971) and Ogata (1988, 1998). The recent key works that have been influential to the author are Marsan & Lengliné (2008), Fox et al. (2016), Fox, Schoenberg & Gordon (2016), Mohler et al. (2011), Veen & Schoenberg (2008) and Zhuang, Ogata & Vere-Jones (2002).

1.3 Main Contributions and Organization

Chapter 2 demonstrates how effective Hawkes models can be in forecasting the spread of Ebola in *real time* by using data from the Democratic Republic of the Congo during 2018-2019. Chapter 3 to our knowledge is the first attempt to investigate the triggering function non-parametrically for infectious diseases and compare the fit and predictive performance of Hawkes modeling to compartmental modeling (SEIR). The *identical* data and SEIR modeling method from Althaus (2014) was used in the comparison. We demonstrate that *simple* Hawkes models can have a superior within-sample fit compared to basic SEIR models and that there is great potential for point process methods to generate quick and reliable real time forecasts during outbreaks that do not heavily rely on tuning parameters, initial values and artistic modeling choices of an expert for good performance. Chapter 4 confronts the long standing problem of distinguishing space-time inhomogeneity and true triggering by proposing an algorithm that allows spatial covariates to be used to non-parametrically estimate the background rate of a Hawkes model, where previous authors have used simple parametric forms or ignored the use of covariates altogether in favor of kernel smoothers. It also proposes a method to sub-sample event based interventions (for violent gang crimes) to generate *synthetic controls* in order to properly make causal interpretations of treatments that are non-randomized. Chapter 5 contains preliminary ongoing work to be published, and future directions. In one section we investigate, through simulation studies, the true efficacy of the so-called *synthetic controls* presented in Chapter 4 as well as quantify the amount of bias caused when non-randomized interventions are erroneously assumed to be a randomized control trial. We discover that Hawkes models can provide accurate productivity

estimates even when treatments are non-random and that bias is only caused when ground truth productivities vary in space and time. We show that the synthetic controls can be used to correctly recover the ground truth efficacy of the intervention/treatment. In another section we propose a branching-only point process model with finite average intensity and non-stationarity (productivity > 1 at times, roughly speaking) in order to better describe and forecast infectious diseases. Finally, we propose a non-parametric method of estimating the so-called recursive point process model in Schoenberg et al. (2019).

CHAPTER 2

Real-time Predictions of the 2018–2019 Ebola Virus Disease Outbreak in the Democratic Republic of the Congo using Hawkes Point Process Models

2.1 Introduction

As of June 16, 2019, 2136 confirmed and probable cases of Ebola virus disease (EVD) were reported in North Kivu and Ituri Provinces of the Democratic Republic of the Congo (DRC) (WHO, 2019). Security issues resulting from activities of over 100 rebel and other insurgent groups, including attacks on Ebola treatment centers in Butembo and Katwa, have likely contributed to the ongoing nature of this EVD outbreak (Damon et al., 2018). Of the>34 prior EVD outbreaks (CDC, 2019), none have occurred in a geographic region with a similar set of conflict issues. Moreover, the use of case counts from previous EVD outbreaks reported in the literature have proven unreliable in their ability to forecast an outbreak's size (Worden et al., 2018; Asher, 2018). It is likely that additional time and effort will be required before all of the contributing factors to this outbreak can be properly assessed, parameterized, and modeled.

The Hawkes point process model, however, offers the Ebola modeling community a novel, rapid option to forecast outbreak size and spread (Meyer et al., 2012). Using modern methods, one can rapidly and nonparametrically estimate short-term outbreak size and rely on minimal modeling assumptions to do so (Schoenberg et al., 2018; Hawkes, 1971; Park et al. (J. of Appld. Stat., rev. & resubm.)). Decomposing peak history effects into the contribution of previous events and an average background rate, this point process model has long been used in the context of seismology to describe earthquakes and their aftershocks as well as other environmental science and biological phenomena (Hawkes, 1971; Gerhard et al., 2017; Schoenberg, 2004; Marsan and Lengliné, 2008). In some cases, Hawkes point process models have also been used to forecast the spatial and temporal spread of infectious disease outbreaks (Schoenberg et al., 2018; Meyer and Leonard, 2014; Meyer et al., 2012), including the 2013–2016 EVD outbreak in West Africa (Park et al. (J. of Appld Stat., rev. & resubm.)).

There is an increasing body of evidence suggesting that short-term forecasts with few parameters are more reliable than long-term forecasts (particularly early in an outbreak) that determine the final outbreak size (Worden et al., 2018; Funk et al., 2018; Viboud et al., 2017; Chowell et al., 2017). In the context of an ongoing outbreak, many published statistical models have focused on long-term or final outbreak size (Meltzer et al., 2014; Kelly et al., 2018; Valdez et al., 2015; Chretien et al., 2015; Siettos et al., 2015). Given the advantages of the Hawkes model and the limitations of other statistical models in the ongoing EVD outbreak setting (Chowell et al., 2017), we fit the Hawkes point process model to daily EVD case counts to forecast case counts over subsequent weeks. It is our hope that this application of the Hawkes point process model may further engage outbreak responders on

the value of short-term forecasts when making important public health decisions related to resource allocations.

2.2 Methods

Data were collected from the Ministry of Health and World Health Organization (WHO) situation reports on EVD case counts occurring in the northeastern region of DRC. The Ministry of Health initially released daily case counts while WHO situation reports confirmed these case counts with weekly reports (WHO, 2019). Our dataset included probable and confirmed EVD cases that occurred from the start of the outbreak on May 3, 2018, until June 16, 2019 (Supplement 1). (We only included in our models case counts from the EVD outbreak in the northeastern region of DRC. In 2018, there was another EVD outbreak that occurred in the western region of DRC, and WHO declared the end of this outbreak on July 24. Although there was a temporal overlap of the EVD outbreaks in DRC, they occurred approximately 1500 miles apart and there has been no evidence of an epidemiological or viral genetic link between them).

We fit the Hawkes point process model to daily EVD case counts reported in the northeastern region of DRC. Details of this estimation method can be found elsewhere (Park et al. (J. of Appld. Stat., rev. & resubm.)). Briefly, for point processes, the expected rate at which points (or cases) accumulate at time t is characterized by the conditional intensity $\lambda(t)$. Although versions of Hawkes models have parameters that describe these types of data in space and time, to be comparable with the SEIR compartmental model here we consider

a purely temporal Hawkes process (Hawkes, 1971) here, where $\lambda(t)$ is written as:

$$\lambda(t) = \mu(t) + K \sum_{i:t_i<t} g(t - t_i).$$

The Hawkes model is estimated essentially by fitting a step function to the triggering density g, where the step heights and background rate μ are estimated by maximum likelihood, according to the method of Marsan and Lengliné (2008), and the step function is subsequently smoothed using a Gaussian kernel. The triggering density g indicates the rate at which infection is spread, and the fitted triggering density shows most secondary infections occurring within a week (Figure 2.1).

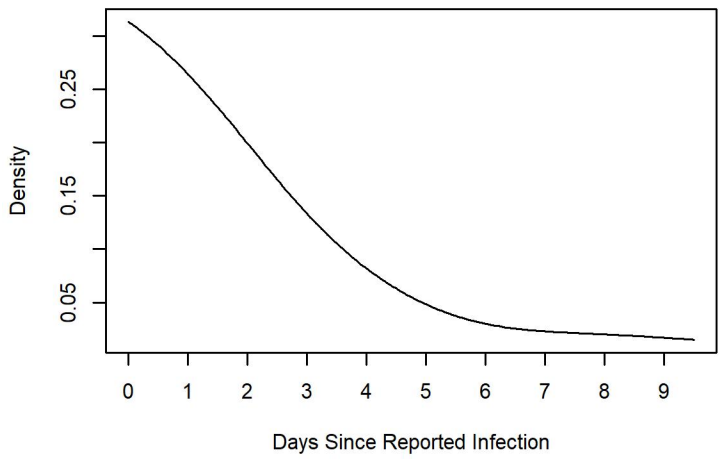


Figure 2.1: Fitted triggering density (smoothed) for the 2018–2019 EVD outbreak from May 3, 2018 to June 16, 2019. The x-axis represents days since infection as reported by WHO, where infection day was in some cases estimated based on how long patients were symptomatic.

The log-likelihood of an observed sequence of infections according to an estimated Hawkes

model is:

$$l(\hat{\theta}) = \sum_{i=1}^{n} log(\lambda(t_i; \hat{\theta})) - \int_0^T \lambda(t; \hat{\theta}) dt.$$

Here, $\hat{\theta}$ is the vector of parameter estimates. The log-likelihood can be computed on the data used to estimate the parameters or can be computed on data outside of the training sample. The log-likelihood is a measure of fit and is closely related to the entropy or information gain of the estimated model relative to a stationary Poisson model (Harte and Vere-Jones, 2005).

One application of the Hawkes model is to enable real-time forecasting of an EVD outbreak. Using the median of 1000 simulations of the fitted Hawkes model, we predicted the number of cases expected to occur over a nine-, six-, and three-week period, starting on April 14, 2019, May 5, 2019 and May 26, 2019, all ending on June 16, 2019, where each subsequent forecast uses model parameters re-estimated with updated data. Then using data up to June 16, we generated probabilistic projections of three-, six-, and nine- weeks based on prior research showing the degradation of epidemic forecasting accuracy over the long term (Worden et al., 2018; Chowell et al., 2017). We evaluated the accuracy of our probabilistic projections by comparing projected vs. actual outbreak sizes, the log-likelihood (information) score (Brocker and Smith, 2005) and the error per day in the median forecast. On April 14, 2019, May 5, 2019 and May 26, 2019 there were 1312, 1667 and 1956 reported EVD cases, respectively. We conducted all analyses using R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

2.2.1 A Word on the Sensitivity of Background Estimate to Start Date

We found that the estimate of μ can be sensitive to the choice of start date of the epidemic. Usually, the 'official' start date is agreed upon by public health experts, but this is an arbitrary decision and can be much after the date where observations are available. We find that if the data on the epidemic includes a long initial period of stagnation followed by a rapid increase in cases, then the background rate estimate will be necessarily small in order to fit the early part of the epidemic with low intensity. On the other hand, for the same outbreak, if the data begins with many cases, possibly in the early-middle part of an outbreak, the background rate estimate will be higher.

The estimation and interpretation of the background rate μ requires caution for diseases that are not readily contracted exogenously through environmental factors (i.e. no *endemic* component). In principle, it ought to be close zero for a disease driven mostly by person to person contact. As Reinhart (2018) notes, any departures from zero then, must be interpreted as cases caused by unobserved infections. Further research is needed on this topic.

2.3 Results

As of June 16, 2019, there were 2136 reported EVD cases across 22 health zones in the provinces of North Kivu and Ituri, DRC. Of these EVD cases, about 95.7% were confirmed and 4.3% were probable. We used the Hawkes model to generate nine-, six- and three-week probabilistic forecasts (all ending June 16, 2019) (Figure 2.2). The median simulated outbreak size on June 16 was 1892 (95% prediction interval [PI]: 1525, 2641), 2236 (95% PI:

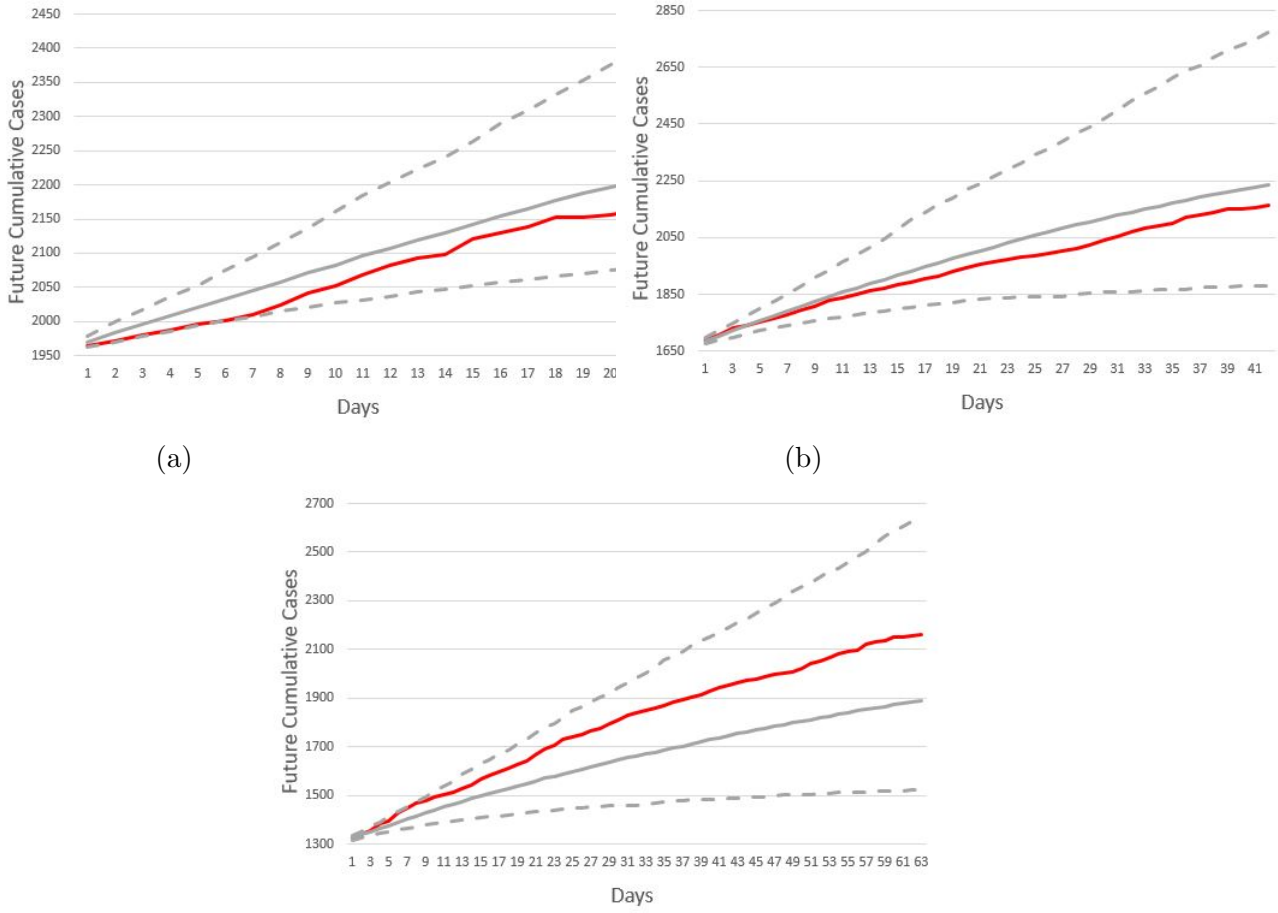
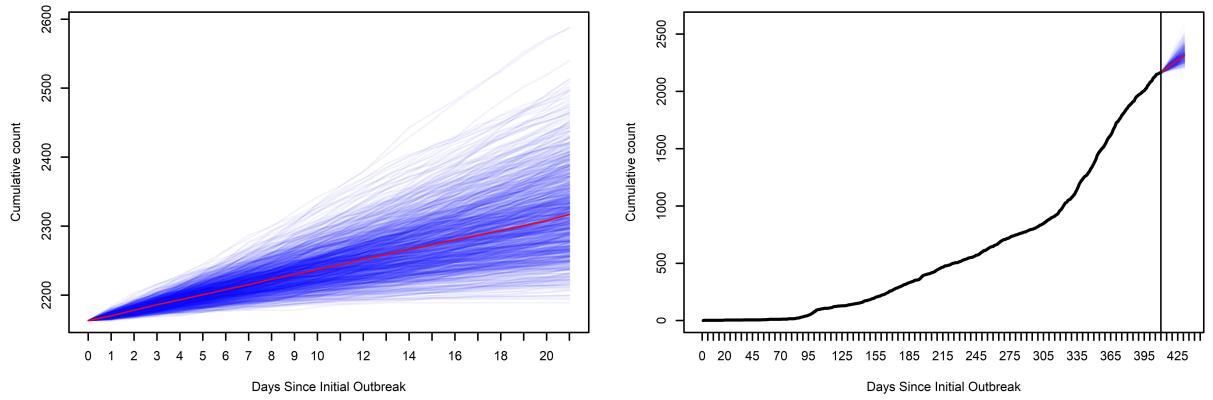
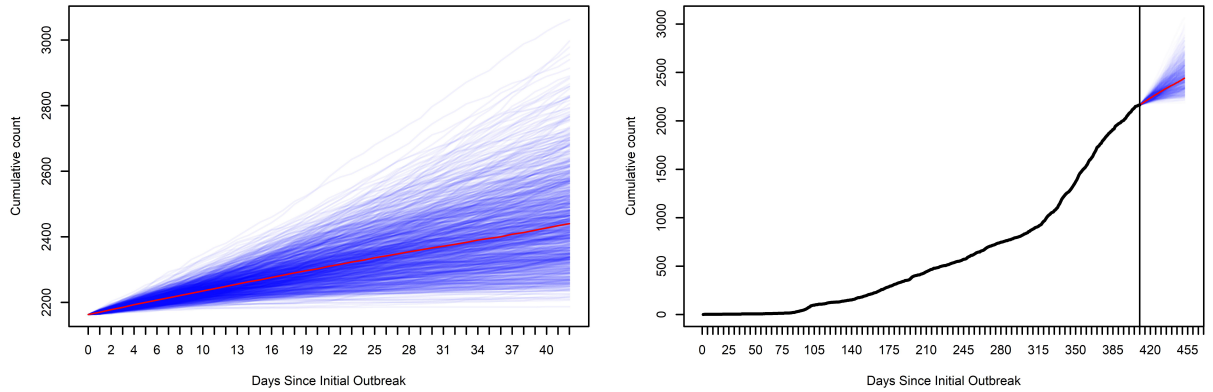


Figure 2.2: Median estimate of projected cumulative case counts (grey line) from April 14, 2019 (Fig. 2.2a), May 5, 2019 (Fig. 2.2b) and May 26, 2019 (Fig. 2.2c), all ending June 16, 2019, and the 95% prediction interval (dotted). Actual cumulative case counts are plotted for comparison (red line) but were not known at the time projections were made.

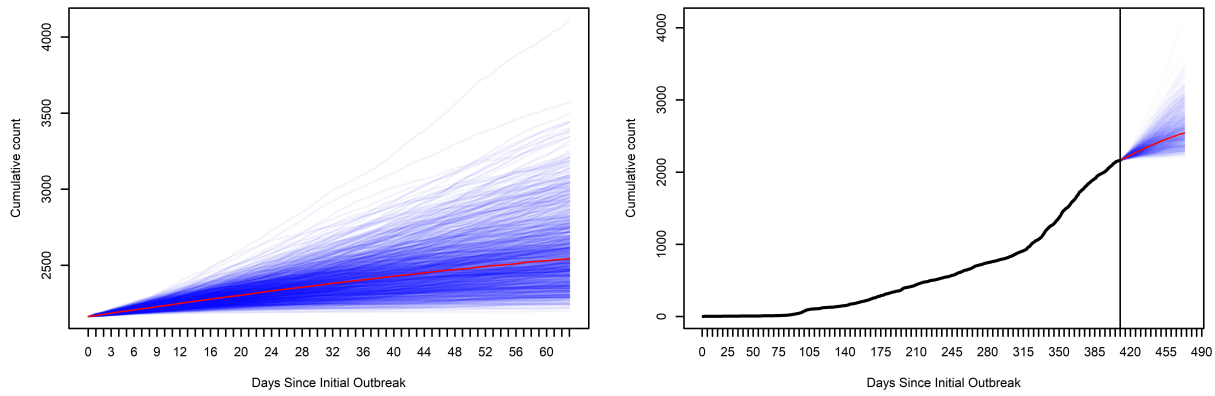
1881, 2773) and 2206 (95% PI: 2079, 2401) respectively. The errors in the median forecasts for the nine-, six-, and three-week forecasts were respectively 6.73 cases, 4.96 cases, and 4.85 cases per day. The log-likelihood (per case) evaluated on the data after the forecasts were made for the nine-, six- and three-week forecasts were 1.60, 1.43 and 1.24, respectively. The higher log-likelihood per day for the 9-week forecasts appears to be attributable to the increased number of observed cases during the first few weeks of the forecasting period causing a sharp increase in the sum of $\log(\lambda)$ term in the log-likelihood.



(a)



(b)



(c)

Figure 2.3: Epidemic curve, as of June 16 (cutoff at vertical line), followed by three- (Fig. 2.3a), six- (Fig. 2.3b), and nine-week (Fig. 2.3c) probabilistic projections (blue lines) of case counts, using the Hawkes model (median, red line), both with outbreak history and zoomed-in.

In our forecast of the unobserved period using data up to June 16, the three-, six-, and nine-week probabilistic projections of median outbreak size were respectively as follows: 2317 (95% PI: 2222, 2464); 2440 (95% PI: 2250, 2790); and 2544 (95% PI: 2273, 3205) (Figure 2.3). The log-likelihood score (per case) of the estimated Hawkes models in Figures 2.2a-2.2c and 2.3 are 0.69, 0.93, 1.04 and 1.06, respectively. Projected and actual outbreak sizes followed a near linear increase (Figure 2.2).

2.4 Discussion

We employed a non-parametrically estimated Hawkes point process model to generate multiple probabilistic projections of the ongoing 2018–2019 EVD outbreak size in DRC. As seen in Figure 2.2, the median nine-week projection experienced some degradation with forecast errors of 6.73 cases per day, whereas the six- and three-week projections were more reliable, with errors in the median forecasts of 4.96 and 4.85 cases per day, respectively, and with the observed number of cases falling well within the estimated 95% prediction window obtained using simulations of the fitted Hawkes model for the three-week period. These findings were consistent with other modeling studies that have shown how even short-term forecasts can degrade over longer periods of time (Funk et al., 2018; Worden et al., 2018). Our results support earlier work performed using Hawkes point process models to predict the size of infectious disease outbreaks; our models of the 2013–2016 EVD outbreak in West Africa reduced root mean squared error (RMSE) by as much as 38% when compared to traditional compartmental models (Park et al. (J. of Appld. Stat., rev. & resubm.)). Growing evidence, including the work presented here, suggests that point process models can provide accurate

estimates of caseloads for a wide variety of epidemics, including both ongoing and previous Ebola outbreaks (Schoenberg et al., 2017; Park et al. (J. of Appld. Stat., rev. & resubm.)).

The Hawkes model performed well during this outbreak with minimal modeling assumptions, and could be a valuable tool for real-time decision making amidst ongoing outbreak of EVD or other diseases. In its non-parametric form, a disadvantage of the Hawkes model may be its inability to parameterize contexts that may help explain the current epidemic trajectory. While these factors (e.g., contact tracing and clinical care) may be considered in future iterations of the Hawkes model (Funk et al., 2017), developing these parameters can also delay model development and application. Even with such parameters estimated, some factors in real biological epidemics, such as political unrest or armed conflict that affect disease transmission rates, can be challenging to parameterize in statistical models.

While the Hawkes model's simplicity has advantages, it can also be viewed as a limitation when, for instance, inhomogeneity of the background rate or changes in productivity lead to overestimation of fine scale clustering, leading to a triggering function estimate that may be less biologically plausible. Unanticipated shocks (e.g., introduction of EVD into a large metropolitan area) that occur after predictions may decrease the model's accuracy beyond our uncertainty estimates. In addition, these models estimate future cases via the triggering function, which requires scrutiny due to its tendency to underestimate secondary transmission rates. Should dynamics of the disease rapidly change at a time period for which data was not included (e.g. driving productivity to a value greater than 1), a relatively simple point process model may not account for such rapid shifts in triggering and may not be able to anticipate them in forecasts. While the Hawkes model might be able to adjust

to decreasing numbers of cases as data become available and parameter estimates change, it may well be that the Hawkes model fails to perform well as the disease cases wane near the end of an outbreak, and this behavior should be a major subject of future research in assessing the forecasts made here.

As such, we do not suggest here that these models replace traditional compartmental models (SIR and their relatives). Rather, we see these models as complementary, and in the particular case of requiring rapid response and prediction of caseloads, a valuable addition for efforts that attempt to limit the impacts of an outbreak. In an effort to continue to evaluate the efficacy of these models in predicting outbreak rates and cumulative cases in real-time (or near real-time, given the time it takes for corrected caseload data to be released), we have constructed a free, publicly-accessible website that can track this and other outbreaks, with purely prospective forecasts and results updated weekly as new data become available (for full details, see http://www.stat. ucla.edu/∼frederic/ebola).

In conclusion, we are encouraged by the ability of non-parametric Hawkes point process models to describe epidemic events over the short term and in real time that are consistent with the 2018–2019 EVD outbreak in DRC. The Hawkes point process is a relatively simple statistical model, and results suggest that statistical modelers in the public health community should consider the Hawkes model in their ensemble when engaging in decision-making and resource allocation of EVD and other emerging infectious disease outbreaks.

CHAPTER 3

A Non-Parametric Hawkes Model of the Spread of Ebola in West Africa

3.1 Introduction

Between March 2014 and June 2016, the West African countries of Guinea, Sierra Leone, and Liberia experienced a historical Ebola outbreak, one that eventually surpassed all other previous Ebola outbreaks combined in terms of total cases reported (WHO Ebola Response Team, 2014). The epidemic resulted in nearly 30,000 infections and more than 11,000 deaths (WHO, 2016), and also took a severe toll on the economy and quality of life in the region, due to decreased trade, border closures, and decreased foreign investment (United Nations Development Programme, 2015). To mitigate future outbreaks of highly infectious diseases, it is important that governments focus on improving detection and response capacity, among other important public health policy objectives (Spengler et al., 2016). To this aim, statistical models can play an important role in forecasting the spread of infectious diseases both during and after an outbreak, leading to more effective allocation and mobilization of public health resources.

One of the first major breakthroughs in epidemiological modeling was the development

of the compartmental model by Kermack and McKendrick (1927) which led to the SIR (Susceptible-Infected-Recovered) model and its variants. Such models involve classifying individuals according to disease status, and then modeling the changes in numbers of infected, susceptible, and recovered individuals in each group using a series of differential equation models. Compartmental modeling has grown to become a primary resource of the epidemiological community for modeling the establishment and spread of many infections such as HIV, SARS, and influenza. In recent decades, traditional SIR models have been modified with new parameters or more informed parameter estimates, to better fit individual disease characteristics (Britton, 2010).

Among these derivations, the SEIR (Susceptible-Exposed-Infected-Recovered) compartmental model has become especially popular for describing the dynamics of the Ebola virus, most notably by Chowell et al. (2004) and applied to the recent West African Ebola outbreak by Althaus (2014). While effective in predicting some aspects of outbreaks, SEIR models rely heavily on accurate parameter estimation and rely critically on the assumption of no community intervention in response to outbreaks as well as the mass action assumption that all susceptible members of the population are equally likely to be infected (Chowell et al., 2004). Although departures from these assumptions are common and can result in wildly inaccurate forecasts (Meyers, 2007), these models provided important insights about the potential progression of the disease without intervention, and played a critical role in eliciting a swift public health response.

While effective in predicting some aspects of outbreaks, SEIR models rely heavily on accurate parameter estimation and rely on the mass action assumption (Meyers, 2007) that

all susceptible members of the population are equally likely to be infected, which is typically violated in practice, often resulting in inaccurate forecasts and require intense field efforts to accurately classify infection status. For instance, compartmental SEIR models applied to the spread of SARS in China in 2003 estimated a high transmission rate and suggested 30,000 to 10 million SARS cases would occur in the first 4 months of the spread of disease in China, resulting in fears of a widespread pandemic (Meyers, 2007). Ultimately, only about 5,300 cases were reported in China (World Health Organization, 2003). These discrepancies in estimated as compared to observed cases suggests that there remains room for predictive improvement. An alternative lies in the use of non-parametrically estimated Hawkes point process models to characterize the dynamics of the spread of disease.

We put forth a simple, non-parametrically estimated Hawkes point process model as an addition to the many popular methods used in the disease modeling and forecasting toolbox. Hawkes models are currently widely used in seismology to describe earthquake catalogs. Though these models have outperformed their competitors in earthquake forecasting experiments (Schorlemmer et al. 2010, Zechar et al. 2013), rarely have they been applied to the emergence and spread of infectious diseases. In seismology, unlike for infectious diseases, the parametric form of the Hawkes model components has been well established over decades of research. Because infectious diseases is a relatively newer application of the Hawkes model, we choose to make no parametric assumptions about the model components.

Self-exciting point processes were used to model the occurrence of smallpox in Brazil by Becker (1977), and by Farrington et al. (2003) to describe the effect of vaccinations on cases of measles in the United States. Recently, Balderama et al. (2012) fit a modified Hawkes

model to sightings of one invasive species of red banana trees spreading in a Costa Rican rainforest, and Meyer et al. (2012) used a parametric Hawkes model for the incidence of invasive meningococcal disease (IMD) in humans, and the results proved useful for estimating spread rates and for the detailed description of properties of the outbreak. However, both Balderama et al. (2012) and Meyer et al. (2012) relied on parametric forms for the triggering function rather than more general non-parametric methods, and neither compared the fit or performance of their fitted model with those of more traditional compartmental models.

Here, for the first time, we compare the performance of non-parametrically estimated Hawkes point process models to more traditional compartmental models (SEIR) for estimating the spread of an infectious disease outbreak. While many variants and advances exist for SEIR modeling, including time-varying transmission rates and additional compartments for pending funerals, hospitals and exposed health care workers (see Viboud et al. 2018, Funk et al. 2018, Champredon et al. 2018), such modifications and parameterizations are often made retrospectively, leaving open the possibility of overfitting and questions about what choices to make regarding model specification and choices of parameters in situations where prospective forecasting is desired. We sought to test the most general, least parameterized versions of each model class to enable a baseline-to-baseline comparison.

Fortunately, the application to the spread of Ebola in West Africa by Althaus (2014) provides an ideal test case, where simple SEIR models have already been proposed, fitted, and shown by an expert to provide a good description of an outbreak in a well-vetted, large dataset with replicates across varying environmental and economic conditions in Guinea, Liberia and Sierra Leone. To follow Althaus (2014), separate models are fit for each of

the three study regions. There are other reasons to consider in support of estimating three separate models for this particular outbreak. Outbreaks may unfold very differently based on public health practices, social behavior, treatment of the dead, etc. In particular, the three countries developed outbreaks in different months, and Sierra Leone and Liberia closed borders within two weeks of their respective outbreaks, suggesting minimal mixing.

Comparisons between point process models and compartmental models are particularly illuminating, as the two types of models rely on different assumptions and operate under fundamentally different mathematical frameworks. As such, Hawkes models may provide different insights into the spread of epidemics and invasive species as compared to more typical models, including a description of the spread via an estimated triggering kernel.

Point process models and compartmental models may be used in complementary fashion, as the two types of models rely on different assumptions and operate under fundamentally different mathematical frameworks. In particular, Hawkes models may provide new insights into the spread of epidemics and invasive species as compared to more typical models, including a description of the spread via an estimated triggering kernel. Meanwhile, compartmental models have the advantage of providing estimates of an age structured basic reproductive number, as well as the dynamics between different sections of the population that are healthy, recovered, infected and dead. These are important in public health for helping to identify portions of the population at particularly high risk.

The structure of this paper is as follows. A description of the data is given in Section 3.2, followed by a brief review of Hawkes and SEIR models in Section 3.3, as well as methods for model fitting and assessing their fit. In Section 3.4, we compare the fit of the two models in

Guinea, Sierra Leone, and Liberia. A discussion and some concluding remarks are given in Section 3.5.

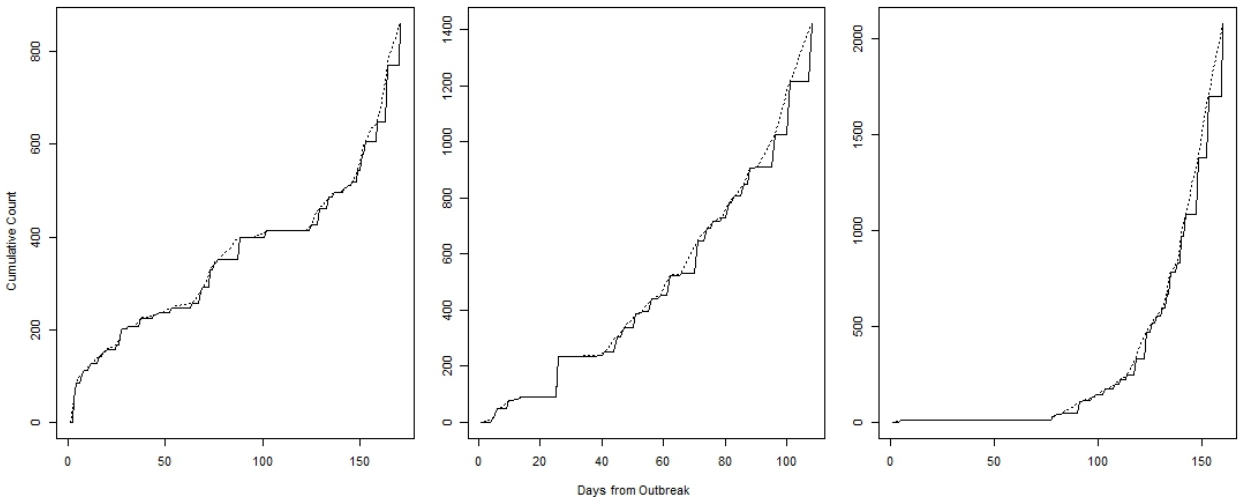
3.2 Data

Data were collected and aggregated from the World Health Organization (WHO) outbreak reports on Ebola during and after the outbreak period (WHO, 2016). These reports are typically released sub-weekly by WHO and include the country, geographic location within country (either by region, closest city, or village) as well as confirmed cases and deaths from Ebola virus. Following Althaus (2014), data were filtered to include only a count of infection cases from Ebola at regular, reported time points in three regions: Southeast Guinea, Eastern Sierra Leone, and Northwest Liberia. The time range of these observations begins on 2014-03-23 and ends on 2014-09-07, again to align with Althaus (2014). In fitting Hawkes models, estimated occurrence times were distributed uniformly within report dates. The first reported cases were distributed uniformly over the previous 2 days, which is half the average time between report dates thereafter. We emphasize that the results here are not sensitive to how this is done, and the prediction of the first week is removed in the overall evaluation of the models.

For a small number of report dates, the cumulative count of cases was subsequently revised downwards by WHO; these revisions are ignored in the current analysis, as in Althaus (2014). The cumulative count of cases reported by the WHO, and the data used to fit Hawkes models are plotted in Figure 3.1. The weekly occurrence of cases is displayed as solid lines in Figure 3.3. A copy of the code, the data used in Althaus (2014) and the data extending beyond

the scope of this study, including the peak and decline of the outbreak can be found at http://www.stat.ucla.edu/∼frederic/ebola.

Figure 3.1: Point process vs. WHO cumulative case counts



(Left to right): Guinea SE, Sierra Leone East, Liberia NW. Solid = cumulative number of cases reported by WHO, dashed = cumulative number of cases reported by WHO with times uniformly spread within WHO report dates. The start dates of outbreak from left to right are 2014-03-23, 2014-05-27 and 2014-04-05 respectively.

3.3 Methods

3.3.1 Hawkes Models and Their Non-parametric Estimation

A temporal point process is a collection of points occurring on the real line. Its associated counting process $N(t)$ counts the number of points occurring between time 0 and time t, inclusive. Such a process is usually characterized via its conditional intensity $\lambda(t)$, which is the instantaneous expected rate at which points are accumulating at time t, given the history $\mathcal{H}_t$ of all points occurring prior to time t (Daley & Vere-Jones, 2003):

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

Because the conditional intensity is a function of time, it is convenient for describing events that may happen with rates that change dynamically. Hawkes models are often extended to data occurring in both space and time. However, to provide an equivalent comparison to SEIR modeling where cases are aggregated over an entire spatial region as in Althaus (2014), we consider a purely temporal Hawkes process (Hawkes, 1971), where $\lambda(t)$ is written as:

$$\lambda(t) = \mu + K \sum_{i:t_i<t} g(t - t_i). \tag{3.1}$$

This is sometimes called a branching or epidemic process (Ogata, 1988) because every occurrence t_i contributes a secondary series of occurrences (aftershocks) occurring at a time-varying rate $Kg(t - t_i)$, which in turn produces its own aftershock sequence, and so on.

The triggering function g in equation (3.1) describes the temporal distribution of secondary infections and is constrained to be a density, so the parameter K represents the expected number of new infections directly attributable to each case. Since each case is posited to cause an expected number K of secondary infections, any particular case is expected to be an ancestor to $K + K^2 + K^3 + ... = \frac{1}{1-K} - 1$ total infections. Thus K should satisfy $0 \leq K < 1$ in order for the process to be stable.

For many processes, the triggering density $g(u)$ decays gradually as the time delay u increases. Model fitting typically involves choosing a parametric form for $g(u)$, and maximizing

the log-likelihood function (Daley & Vere-Jones, 2003),

$$l(\theta) = \sum_k \log[\lambda(t_k;\theta)] - \int_0^T \lambda(t;\theta)du \tag{3.2}$$

where θ is the vector of parameters governing the shape of g and $[0, T]$ is the time window of observation. All parameters and triggering densities are estimated separately for each region as in Althaus (2014).

We non-parametrically estimate the triggering function g, constant background rate μ and productivity constant K using the method proposed by Marsan and Lengliné (2008). This method assumes that g is a piecewise constant step function with user-defined number of steps and unknown heights estimated by approximate maximum likelihood. Although the number of steps is set by the user and the step heights are estimated, the method is non-parametric in that no specific parametric model is assumed for the shape of the triggering function. A key principle driving this methodology is that, given a model for $\lambda(t)$, the probability that a point t_j occurred exogenously due to the background rate is

$$p_{jj} = \frac{\mu}{\lambda(t_j)}, \tag{3.3}$$

and the probability that point t_i triggered point t_j is

$$p_{ij} = \frac{Kg(t_j - t_i)}{\lambda(t_j)}, \tag{3.4}$$

as noted in Zhuang, Ogata and Vere-Jones (2002). Initial estimates, $p_{jj}^{(0)}$ and $p_{ij}^{(0)}$ for $1 \leq$

$i < j \leq N(T)$, dictate the full probabilistic branching structure of the point process, and this is used to obtain estimates $\mu^{(0)}$, $K^{(0)}$, and the steps heights of $g(u)^{(0)}$ following the non-parametric procedure described in Marsan and Lengliné (2008) and Fox et al. (2016). Using these estimates, one computes updated estimates of the probabilities $p_{jj}^{(1)}$ and $p_{ij}^{(1)}$ using (3.3) and (3.4). This is iterated until the largest update in any $p_{jj}^{(k)}$ or $p_{ij}^{(k)}$ is less than some small constant ϵ.

The Hawkes model uses occurrence times that are uniformly randomized between report dates. To examine the sensitivity of the estimates to this randomization, for each country we estimate the model on 50 randomized datasets and plot the distribution of the resulting estimated triggering density in Figure 3.2b.

In simulating a Hawkes process, an infection can be categorized into one of two types: (1) An infection triggered by a previously occurring infection, (2) a background infection resulting from the constant rate μ not attributed by the model to any previous parent infection. The number of first generation triggered infections is a Poisson(K) random variable, and they are randomly placed in time according to the temporal distribution given by g in equation (3.1).

3.3.2 SEIR Models and Their Estimation.

The SEIR (Susceptible-Exposed-Infected-Recovered) compartmental model embodies the idea that the infected population spreads the disease at time t with rate $\beta(t)$, but can only spread the disease to the proportion of the population still susceptible, and these rates and proportions can change as an outbreak proceeds. It has been frequently used to describe

Ebola disease dynamics and is characterized by the following set of ordinary differential equations (Chowell et al., 2004):

$$\frac{dS}{dt} = -\beta(t)\frac{SI}{N} \tag{3.5}$$

$$\frac{dE}{dt} = \beta(t)\frac{SI}{N} - \sigma E \tag{3.6}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{3.7}$$

$$\frac{dR}{dt} = \gamma I. \tag{3.8}$$

Here S is the size of the susceptible population, E is the size of the population that has contracted Ebola but is not yet infectious ("latent population"), I is the size of the infectious population, and R is the size of the recovered/deceased population. These four quantities sum to N, the total population. The populations were treated as discrete entities as in Althaus (2014).

When modeling the infectious phase, the primary quantity of interest in this model is $\beta(t)$, the transmission rate. Under this model, it is assumed to decline exponentially at rate κ:

$$\beta(t) = \beta_0 e^{-\kappa t}, \tag{3.9}$$

where t is the number of days from the start of the outbreak (Lekone and Finkenstädt, 2006). Other parameters in the SEIR model include the rate of infectious onset, σ, and the rate of death or recovery, γ. In model fitting, γ and σ are typically assumed constant, as in Althaus (2014).

A central feature to compartmental SIR/SEIR modeling is the reproductive number, $R_0(t)$. In the model, $R_0(t)$ at any time is given by the transmission rate, $\beta(t)$, multiplied by the average duration of infectiousness, $1/\gamma$. Here $R_0(t)$ represents the average number of new infections generated by an infected person until the infected person dies or recovers. The critical threshold for $R_0(t)$ is 1: if $R_0(t)$ is above 1, the epidemic can spread to infect a large proportion of the population. When $R_0(t)$ drops below 1, the epidemic is unsustainable (Diekmann and Heesterbek 2000, Lipsitch et al. 2003).

As in Althaus (2014), parameter estimates for SEIR models were obtained using maximum likelihood estimation (MLE) assuming that occurrences of new cases follow a Poisson distribution. In simulating the model forward, we assume new cases follow a Poisson distribution. SEIR models were fit separately for Guinea, Sierra Leone, and Liberia using the original discretely reported Ebola outbreak data containing only cases and deaths at each reporting date, as in Althaus (2014). The theoretical SEIR model outlined above is purely deterministic, so to convert this process into a stochastic model for simulating real-world outbreaks forward, and to facilitate model evaluation using statistical methods, the tau-leaping approximation of Cao et al. (2007) was applied. Under the Tau-leaping simulation method applied to SEIR, transitions from each of the S, E, I, and R populations are simulated based on a Markov-Chain Monte Carlo (MCMC) process. Transition probabilities from S to E, E to I, and I to R are all calculated using the current state populations and the fitted $R_0(t)$, κ, σ, and γ. Under each round of simulation, at random one person at one time point will transfer to a new population based on these probabilities. The transition probabilities are then recalculated based on the new state populations. The process continues until the end

of the observed time window is reached.

Under this process, new exposures, infections, and recoveries occur randomly as a Poisson arrival process at probabilities based on R_0, κ, σ, and γ. With each new transition, the probabilities of these events update to reflect the new S, E, I, and R populations. This approach is identical to the SEIR model used by Althaus (2014), and as a result, we were able to recreate the parameter estimations and model output of this work for each country to allow for direct comparison with Hawkes models.

As in Althaus (2014), the Nelder-Mead algorithm and the deSolve R package (Soetaert et al., 2010) are used for SEIR parameter estimation, and the tau-leaping method in the adaptivetau R package (Johnson, 2016) is used to simulate SEIR forward in time.

3.3.3 Evaluation Techniques

Simulations of both SEIR and Hawkes models were used in order to assess statistically the compatibility of the observations with forecasts made with each model. SEIR and Hawkes parameter estimates obtained using all available data for each country were used for each of 1,000 simulations, and the mean of simulations for each week and each country were recorded and compared to the actual number of infections per week. Here, different scoring rules are possible as surveyed in Gneiting and Katzfuss (2014), and we focus primarily on the root mean square error (RMSE) of weekly predictions and note that our main findings do not appear to be substantially influenced by the choice of scoring rule. SEIR simulations require assumed values for the starting populations on day 0. Hawkes models do not require such assumptions. However, according to the fitted Hawkes model, simulated infections near the

beginning of the recording period are assumed to result exclusively from the background process rather than contagion, which may lead to under-predictions in week 1. Since both models are impractical for estimating cases in week 1, only estimates from week 2 and beyond are used for comparison.

The susceptible population was set based on the most recently published census data from Guinea (National Institute of Statistics, 2015), Sierra Leone (Sierra Leone Statistics, 2016), and Liberia (LISGIS, 2009), under the assumption that everyone in the population is susceptible. The infectious populations were set based on the observed data from the WHO.

In a separate analysis to account for the possibility of over-fitting in a retrospective analysis such as this, and to assess the ability of the models to forecast spread during an actual outbreak, parameter estimates were fitted using only the infections occurring in the first 75% of the observation window, and the resulting fitted models were then used to project cumulative infections for the remaining period. This analysis also involved 1,000 simulations per country for both SEIR and Hawkes.

The SEIR and Hawkes models were also evaluated using superthinning (Clements et al., 2013). In superthinning, the existing data points are first thinned where each point is randomly kept independently of the others with probability $\min\{b/(\hat{\lambda}(t)), 1\}$, and then new points are superposed according to a Poisson process over the observed time window $[0, T]$, with rate $(b - \hat{\lambda}(t))^+$. Superthinning requires an initial choice of the tuning parameter, b, and as suggested in Clements et al. (2013), we used the simple default value of the total number of cases divided by the length, in days, of the observation period. For the SEIR model, the value of $\hat{\beta}(t)$ multiplied by the size of the infectious population at time t was used as the

estimated rate function $\hat{\lambda}(t)$ to calculate thinning and superposing probabilities. For the Hawkes model, $\hat{\lambda}(t)$ corresponds directly to the estimated conditional intensity function in equation (3.1), which is calculated with estimates $\hat{\mu}$, $\hat{K}$, $\hat{g}(\cdot)$. After superthinning, the resulting residual process is a homogeneous Poisson process with rate b if and only if the estimate of the conditional intensity, $\hat{\lambda}$, is correct (Clements et al., 2013), and thus one may examine the superthinned residuals for uniformity. 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.

3.4 Results

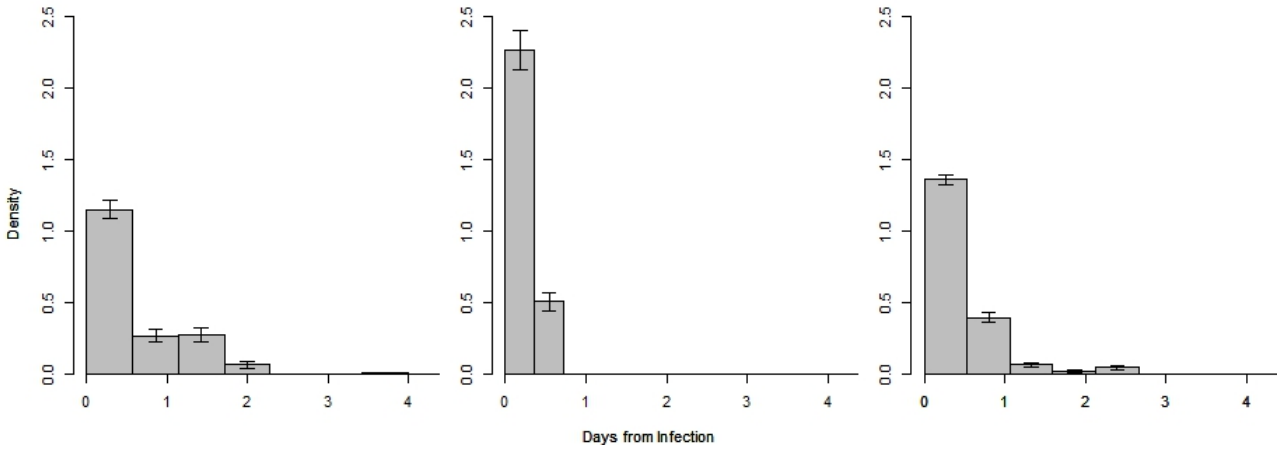
3.4.1 Model Fitting and Weekly Estimates

Figure 3.2a displays non-parametric estimates of the Hawkes triggering density in (3.1) for each country. Figure 3.2b plots the distribution of estimated triggering densities for 50 different random imputations of occurrence times. According to the fitted model, in Guinea and Liberia, an infected individual directly triggered new infections on the scale of up to 3 days. In Sierra Leone, this density appeared to decay somewhat faster, with most triggering occurring within 1 day, according to the fitted Hawkes model. Note that the estimated triggering times in Figure 3.2a are times between *recorded diagnoses* of cases of Ebola; such recordings may be substantially more clustered than actual transmissions of the disease, for instance, due to small reporting delays common in epidemic data.

Table 3.1 shows the estimated parameters of the Hawkes model (3.1) for each country.

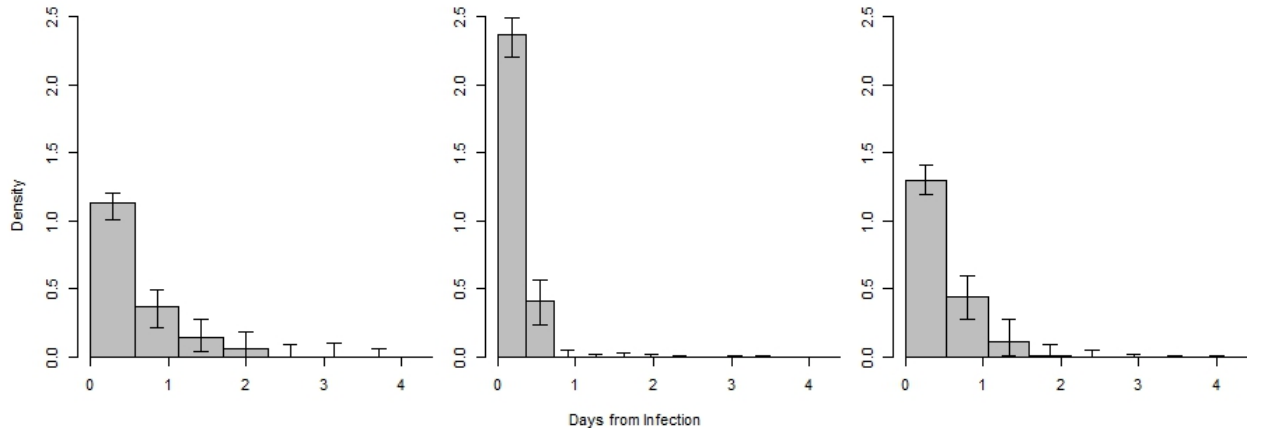
Figure 3.2: Estimate and sensitivity of triggering

(a) Estimated Hawkes triggering density



(Left to Right): Guinea SE, Sierra Leone East, Liberia NW. Whiskers represent $\pm$ 2 standard errors computed as in Fox et al. (2016).

(b) Sensitivity of triggering histogram to randomly imputed data



(Left to Right): Guinea SE, Sierra Leone East, Liberia NW. Bar heights represent the median. Whiskers represent the 95th percentile interval.

More intense clustering was observed in Liberia and a significantly smaller percentage of the points are attributed to the background rate according to the fitted Hawkes model. According to the fitted Hawkes model (3.1), 89% of cases in Guinea were attributable to contagion from other observed cases in Guinea, whereas in Sierra Leone an estimated 93% of cases were attributed to contagion from other observed cases, and in Liberia the corresponding estimate was 99%.

Table 3.1: Hawkes parameter estimates and standard errors

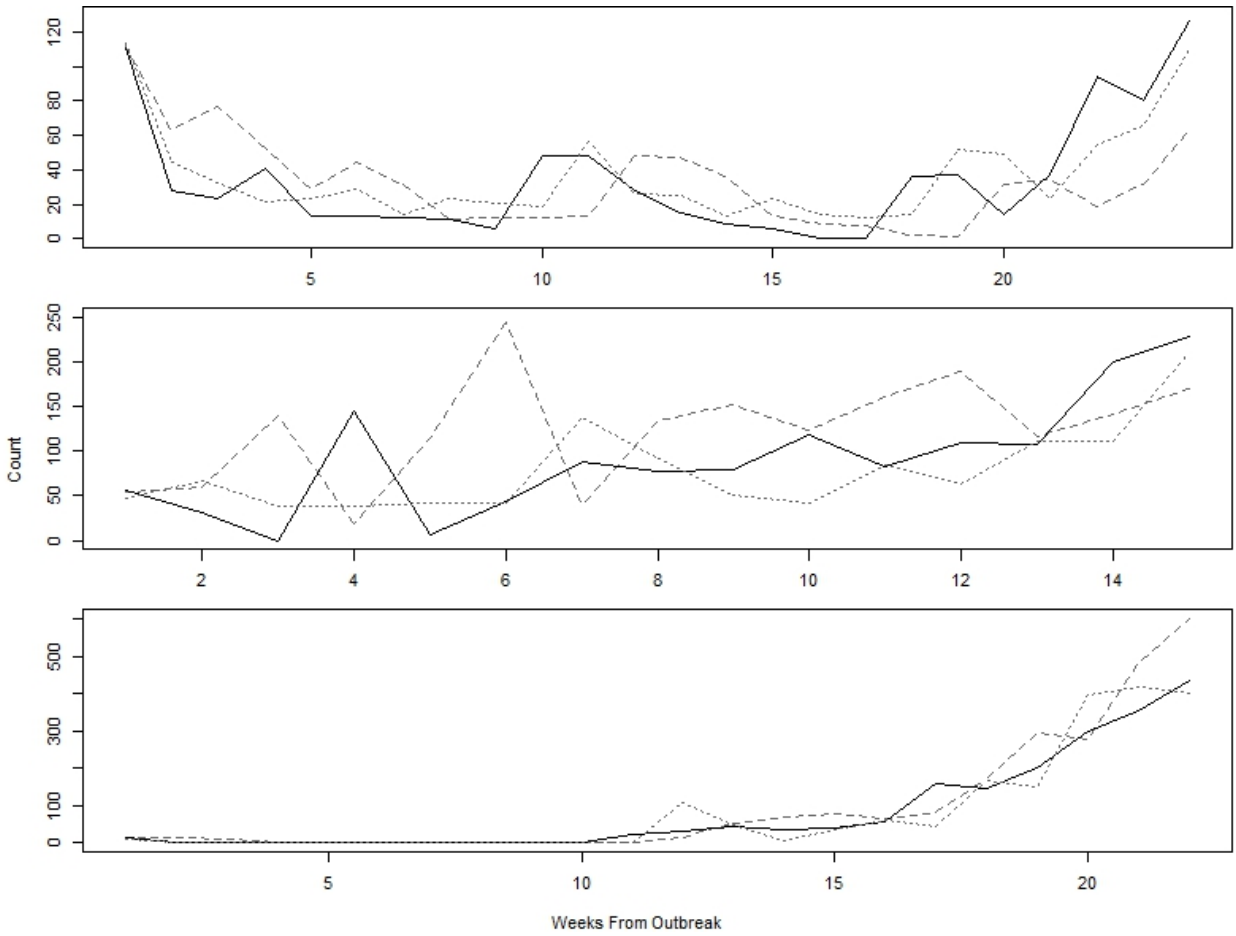
	Guinea	**Sierra Leone**	**Liberia**
Background rate $\hat{\mu}$	0.544	0.91	0.037
	(0.053)	(0.089)	(0.015)
Productivity constant $\hat{K}$	0.893	0.931	0.997
	(0.011)	(0.0067)	(0.0012)

Standard errors in parentheses are calculated following Fox et al. (2016).

The first row of Table 3.2 contains the RMSE's corresponding to the simulation based goodness-of-fit study in Figure 3.3. Hawkes models had a correspondingly lower RMSE in weekly predictions compared to SEIR models, for all three countries. The total RMSE across all countries combined was 59.8 cases/week using SEIR and 37.1 cases/week using the Hawkes model (3.1), which represents a 38% decrease in the RMSE. Weekly estimates of total infections per week based on the mean of 1000 simulations of the Hawkes and SEIR models are displayed along with the observed number of infections in Figure 3.3. The weekly simulation means for the fitted SEIR model resemble a lagged version of the observed weekly counts with a lag of two weeks. This is not surprising given that the rate of infectiousness for SEIR modeling is a function of the current infectious population, and the incubation time of the Ebola virus is approximately 8-12 days (Chowell et al. 2004).

Hawkes models show a similar dependence on past fluctuations in cases due to the cascading nature of Hawkes processes; a large number of cases in the previous week will be expected to trigger more simulated infections into the following week. However, the week to week dependence is weaker and more complex than in the fitted SEIR model. For example, in weeks 18 through 20 in Guinea SE, the Hawkes weekly projection behaves like a one week lag of the actual data. At other times, however, the Hawkes estimate appears to be more

Figure 3.3: Weekly forecasts of new infections from SEIR and Hawkes models



(Top to bottom): Guinea SE, Sierra Leone East, Liberia NW. Solid curve = observed new case incidence per week as reported in WHO data, dashed curve = SEIR forecast, dotted curve = Hawkes forecast. As in Figure 3.4, the start dates of outbreak from top to bottom are 2014-03-23, 2014-05-27 and 2014-04-05 respectively. Each weekly forecast is the mean of 1000 simulations. For each week, simulations of new cases were conducted using model parameters fitted over each country's entire data set. Each week's simulations began with the same number of initial infected cases based on the history of reported infections preceding each week's simulation start date.

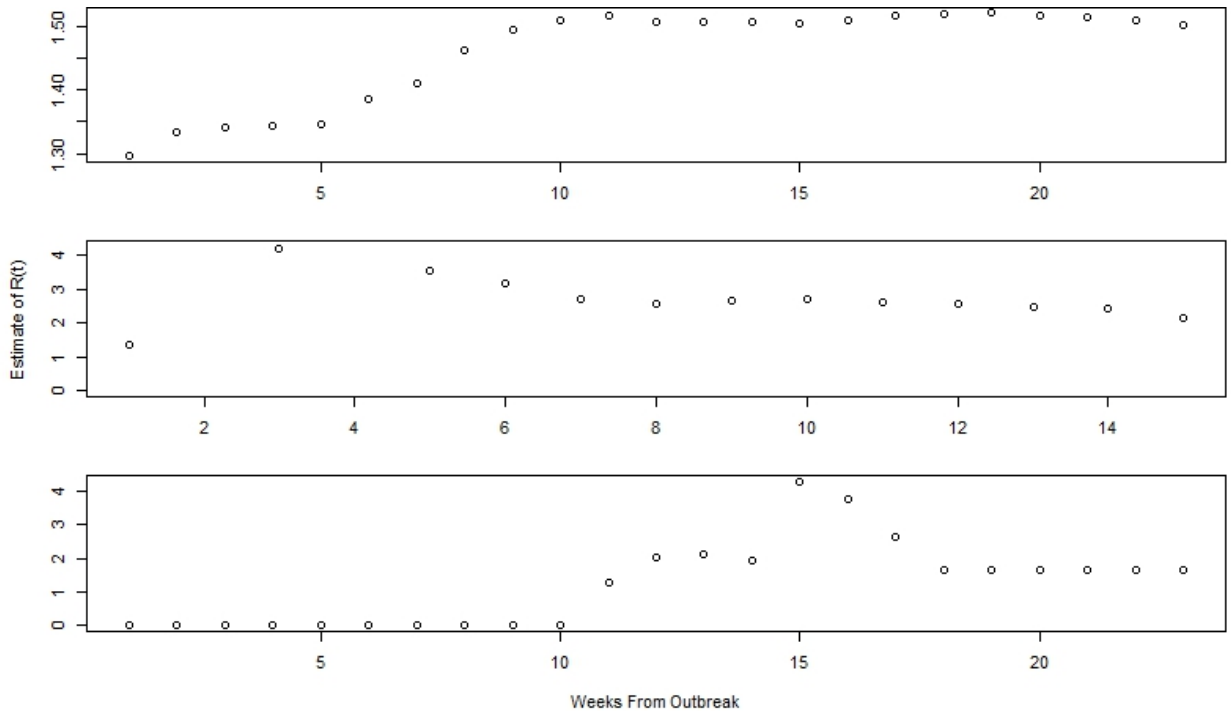
adaptive than a mere one week lag, such as in forecasting the sudden decrease in infections from week 11 to 12 in Guinea SE, or the rise in infections from week 5 to 7 in Sierra Leone East. Across all three countries the Hawkes models tend to produce more accurate weekly estimates with less extreme errors.

Table 3.2: RMSE of weekly forecasts for SEIR and Hawkes

	Guinea		**Sierra Leone**		**Liberia**	
	SEIR	Hawkes	SEIR	Hawkes	SEIR	Hawkes
in-sample	33.25	17.85	91.84	49.91	55.03	42.31
out-of-sample	26.63	2.87	46.75	46.50	357.10	93.88

First row contains RMSE corresponding to Figure 3.3, in units of cases. Second row contains RMSE corresponding to the first two weeks of Figure 3.5. Guinea, Sierra Leone, Liberia observed 861, 1434 and 2081 total cases respectively, of which 403, 641 and 1189 occurred in the final 25% of the observation window.

Figure 3.4: Weekly estimates of R_0 over time, denoted $R(t)$.



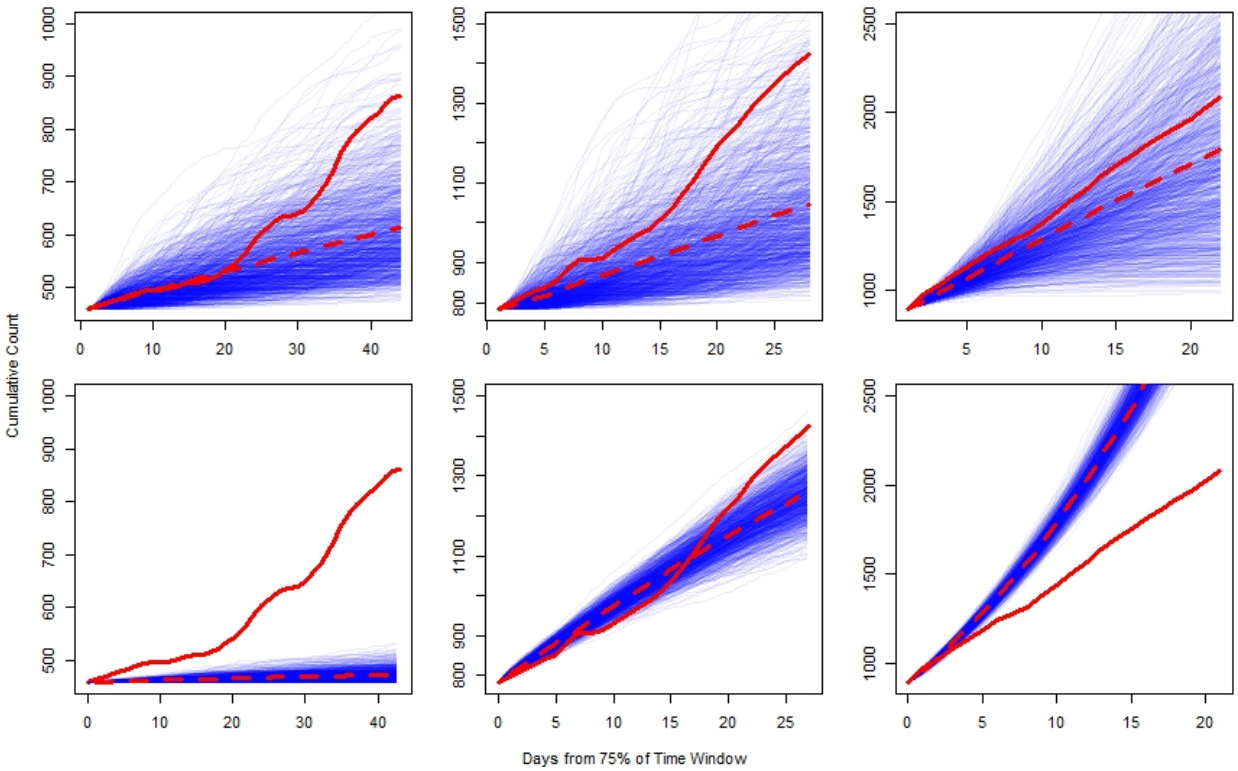
(Top to bottom): Guinea SE, Sierra Leone East, Liberia NW. Point estimates of the SEIR reproductive number, R_0, by week. As in Figure 3.3, the start dates of outbreak from top to bottom are 2014-03-23, 2014-05-27 and 2014-04-05 respectively. Each weekly point estimate is based on the actual case counts observed up until that point in time. Note that for Sierra Leone East, estimates are off the charts at weeks 2 and 4 with 12.75 and 40.46, respectively.

3.4.2 Prospective Out-of-Sample Prediction

Figure 3.5 shows simulations of SEIR and Hawkes models fitted using only the first 75% of data for each outbreak and simulated for the remaining 25% of the time. For Sierra

Leone, the SEIR model simulations forecast the trajectory of the number of new infections quite accurately, especially during the first 16 days of the simulations. The SEIR model significantly underestimated the number of new infections in Guinea throughout the course of the simulation. For Liberia, the simulations of the fitted SEIR model initially tended to overestimate the number of observed infections, due to the approximately exponential predicted acceleration characteristic of the SEIR model.

Figure 3.5: SEIR and Hawkes projections using first 75% of data for fitting



(Left to right): Guinea SE, Sierra Leone East, Liberia NW. Starts dates of simulations from left to right are 2014-07-28, 2014-08-13 and 2014-08-19 respectively. Thin curves in top panels show 1,000 simulations of Hawkes model (2) with parameters fit using first 75% of the data for the corresponding country and simulated forward for the last 25% of the observed time period. Thin curves in bottom panels show 1,000 simulations of SEIR model with parameters fit using first 75% of the data for the corresponding country and simulated forward on the last 25%. Dashed curve = mean of simulations. Solid curve = actual cumulative total number of observed cases as reported by WHO.

Simulated Hawkes processes for Guinea were remarkably accurate for the first 20 days of the simulations, and tended on average to underpredict the number of new infections after

30 days. The simulations of the Hawkes model in Sierra Leone and Liberia also tended to slightly underpredict the number of infections. In all three countries, the variation in the Hawkes simulations was much greater than that of the SEIR model. Using the average of the simulations as a forecast, the weekly RMSE in each country for the first two weeks is shown in the second row of Table 3.2. The total RMSE of the forecasts in the first two weeks for SEIR was 208.5 cases/day, compared to 60.5 cases/day for the Hawkes model, representing a 71% reduction in error. The results suggest that overfitting is not responsible for the improved performance of the Hawkes model for describing the spread of Ebola in these 3 countries, since in these comparisons the models were assessed using data not used in the parameter estimation.

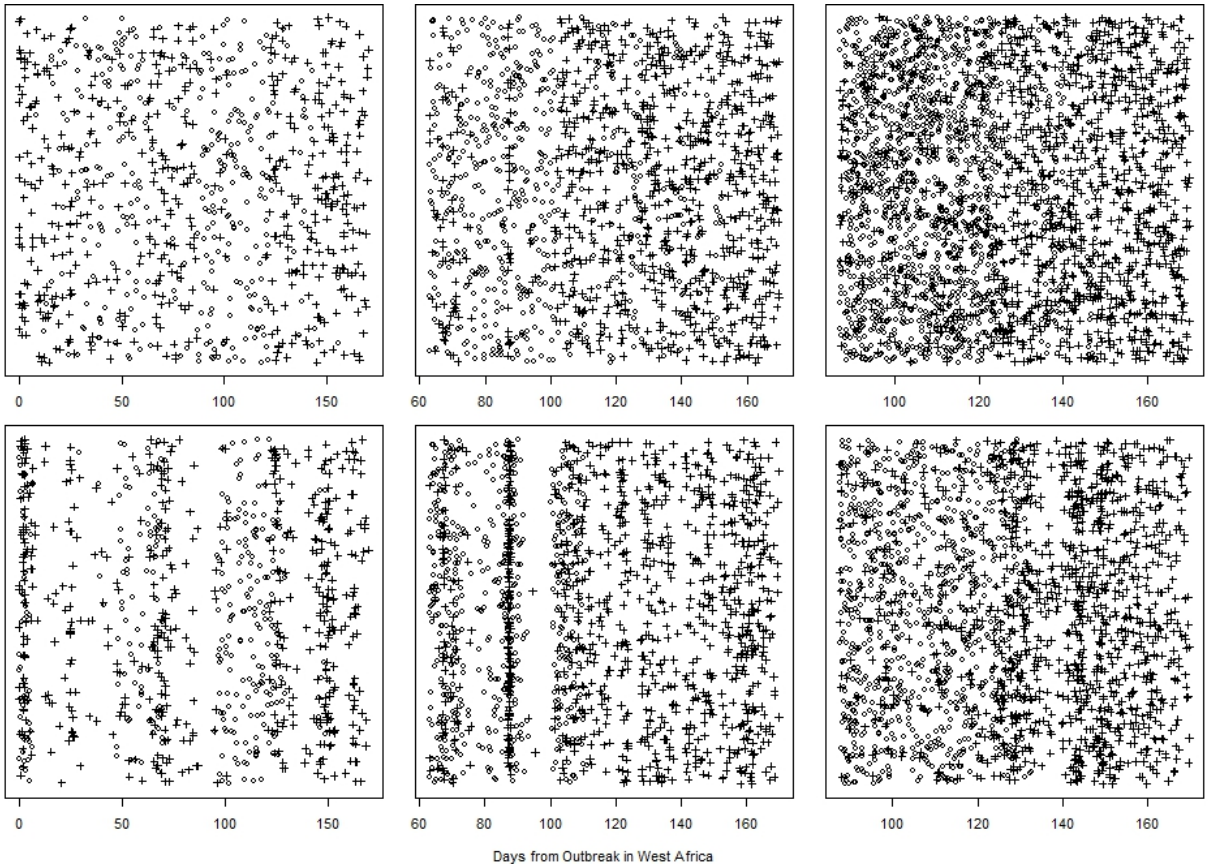
The amount of data used to fit both SEIR and Hawkes models has a large impact in determining the direction and variability of simulations. For SEIR, simulations were affected by the estimated populations of infected individuals for each country. If a model was fit at a time point right after a reported sharp increase in case load, such as in Liberia when estimating using only the first 75% of the data, the simulations would project large increases. Given a time period culminating in few new infections, the SEIR model tended to forecast a continued period of very few new infections, which was inconsistent with the actual observations. For all Hawkes simulations, the average simulated path was approximately linear because according to the fitted Hawkes model, the background rate of new infections is constant and the number of triggered infections stemming from any single infection is finite.

3.4.3 Superthinning Analysis

Superthinning results are displayed in Figure 3.6 for all three regions. The superthinned residuals corresponding to the SEIR model for Guinea have obvious clustering during the first week and around 2014-06-01 and 2014-08-20, indicating times when the model underestimated the number of new infections. This likely occurs because the estimated conditional intensity is heavily dependent on the current infectious population which can be highly variable. Immediately after an unexpected surge in observed infections, the modeled rate according to the SEIR model tends to remain relatively low for approximately 2 weeks, and as a result, most of the observed points are retained after superthinning, resulting in intense clustering in the residuals.

The gaps in Figure 3.6 indicate time periods where the model overestimates the rate at which infections are occurring. An example is the large gap in Figure 3.6 for Sierra Leone East around 95 days after the start of the overall West African Ebola outbreak (2014-03-23), which corresponds to the large overestimate by SEIR in Sierra Leone East 6 weeks after 2014-05-27 as seen in Figure 3.3. The SEIR superthinned residuals for Guinea also have noticeable gaps and regions of sparsity of points, particularly around 2014-07-01, corresponding to the SEIR model overestimating the number of new cases around week 12 and 13 in Figure 3.3. Overestimation from a lack of points is seen in several places in Guinea but most noticeable around June 26th. The superthinned residuals corresponding to Sierra Leone indicate poor fit in many places as well and most notably underestimation of the number of new infections around 2014-05-21, as well as overestimation around 2014-07-01. The superthinned residuals for Liberia corresponding to the SEIR model show few noticeable departures from uniformity

Figure 3.6: Superthinning using Hawkes and SEIR infection rate parameters



(Left to right): Guinea SE, Sierra Leone East, Liberia NW. Top = Hawkes, bottom = SEIR. Thinned original points are marked with plus signs, and superposed points are marked with circles. X-axis indicates days from 2014-03-23, the beginning of the West African Ebola outbreak. The y-coordinates are uniform(0,1) random coordinates.

despite some sparsity of points around 2014-08-05.

The superthinned residuals corresponding to the fitted Hawkes models, shown in Figure 3.6, show no substantial departures from homogeneity for all three countries. The Hawkes model does not have any gross inaccuracies in Guinea, Sierra Leone, or Liberia and appears to more closely describe the spread of new infections in these three countries.

3.5 Discussion

The application of non-parametric Hawkes point process models to predict the spread of Ebola virus in West Africa indicates that these novel methods have the potential to be a useful addition to the pallet of available methods for disease forecasting. The root-mean-square errors in forecasts made using data used in parameter estimation as well as in forecasts of data not used in parameter estimation, as well as the uniformity of superthinned residuals all clearly suggested that in each case the simple Hawkes models fit as well as, or better than, the corresponding SEIR models. Our results do not suggest traditional SEIR models should be discounted; rather, they highlight the utility of Hawkes processes as an alternative and novel framework with which to predict disease spread to reveal new insights to how outbreaks may evolve. SEIR and Hawkes models provide two very different descriptions of an outbreak and focus on different aspects of the disease spread. We suggest that using both types of forecasts may provide additional insights to inform public health decision making.

For instance, in weekly forecasts, Hawkes models appeared to provide more accurate predictions than SEIR. This is possibly because a Hawkes model, unlike SEIR, works by estimating the temporal distribution of times between *individual* reported infections via the triggering kernel. This provides new information about the dynamics of an outbreak that is highly localized in time.

One might object that, in a retrospective analysis such as this, the improvement in fit might be due to fitting a more complex model with more free parameters, in which case overfitting might be a problem and the improvement would be unlikely to be replicated in further applications, particularly if the model were used in forecasting. However, we found

that even when models were fit using one portion of the data and assessed on a different portion of the data, simulating performance *during* an outbreak, Hawkes models were more accurate than SEIR by a 71% reduction in RMSE for the first two weeks of forecasting. This suggests that Hawkes models may in the future be used for more accurately forecasting the spread of epidemic diseases, which could help facilitate and inform surveillance and mitigation efforts to help curb outbreak spread as it occurs.

One limitation of our Hawkes models is their reliance on a constant supply of exogenous infections from the background rate for the continued propagation of the disease. According to the fitted model, subsequent triggered infections eventually die out since each infection only directly triggers a Poisson(K) number of new infections where $K < 1$. Such a Hawkes model would fit poorly to data exhibiting intense clustering of observed cases. This may also lead to difficulty in properly modeling the first few weeks of a contagious outbreak when the total cumulative number of cases is low or the latter parts of an outbreak where disease spread is nearly saturated and slowing down due to human containment or intervention. Some modifications to Hawkes models have been proposed to account for this issue and show promising results (Schoenberg et al., 2019). In addition, where the background rate is known, rather than estimated by maximum likelihood, the triggering kernel of a Hawkes process could perhaps be estimated more accurately and this is a subject of ongoing research. The results in this paper demonstrate that basic Hawkes modeling can be effective in forecasting caseloads up to a few weeks after an outbreak has emerged. A conservative estimate of when forecasts based on the Hawkes model lose most of their predictive power is around 2 weeks, as evidenced not only by the results in this paper but also particularly in Kelly & Park et

al. (2019).

It should be noted that the WHO data considered here consist of periodically updating counts of cases (on average 4 days between updates). A higher temporal resolution will be particularly desirable for future research in order to improve the accuracy and assessment of point process models. The data are likely not comprehensive in accounting for every case of Ebola at the correct time due to limits on human resources in managing the large area and population of the three study regions. Because SEIR modeling is heavily dependent on the current population of infected and susceptible individuals, it may be hypothesized to be more sensitive to missing data and errors in reporting, though this should be studied and quantified in future research. With improved surveillance and reporting measures, as well as more frequent measurements, it is possible that SEIR models would show significant improvements. While the spread of Ebola in West Africa in 2014 is one case study that demonstrates the effectiveness of Hawkes modeling, SEIR and Hawkes models may perform differently for other diseases, regions, and time periods. It is recommended that future investigations compare the fit of Hawkes and SEIR models to data on other diseases and in other regions, and to perform prospective analyses to evaluate the forecasting performance of the two types of models. Such work could help determine whether Hawkes modeling will provide a generalizable framework for prediction across a variety of infectious diseases, with highly disparate outbreak periods.

Another important area for future study is to use spatial-temporal triggering densities in Hawkes models, to describe the detailed spatial-temporal distribution of infections when sufficient spatial precision is available. Whether classic exponential and power law kernels work just as well as the non-parametric approach can be checked as well. In this paper,

we limited the Hawkes model to purely temporal triggering in each spatial region, in order for the Hawkes and SEIR models to be comparable and so both models could be estimated using the data of Althaus (2014). However, one advantage of Hawkes models is their natural generalization to the case of further spatial precision. By contrast, compartmental modeling is generally limited due to its assumption of spatial homogeneity of each compartment's population. In the case of an outbreak across adjacent countries, a single SEIR model that incorporates some dependencies from neighbors may be an important topic of future study for expert SEIR modelers. Other attempts have been made in this regard through meta-population modeling and spatial compartmental modeling by Keeling & Rohani (2007) and Guofo et al. (2014), respectively. The latter, for example, propose a fractional SEIR model using separate S, E, I, and R compartments for each neighboring major metropolitan region in New Zealand with additional terms for the spread between these regions, such models still spatially aggregate the observations resulting in the loss of some information and resolution compared with spatial-temporal point process models such as Hawkes models. These considerations should provide impetus for future model refinement and improvements for each model class that are likely to further improve our understanding of the evolutionary nuances of disease outbreaks.

CHAPTER 4

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

4.1 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: (1) spatial heterogeneity in local risk factors and (2) 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 (Johnson 2008, Chp. 9.6 Diggle 2014).

Spatio-temporal clustering is particularly characteristic of gang violent crime (Valasik 2017, Martinez 2016). 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) (Brantingham 2019, Papachristos 2009). 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 (Kubrin 2003, Sampson 1997, Papachristos & 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 2019, Patillo-McCoy 1999). Gang violent crime therefore tends to cluster where gangs are most active, particularly along gang territorial boundaries (Brantingham 2012, Tita & Ridgeway 2007). 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 easily escalate to a shooting, while a shooting or homicide often demands retribution in kind (Hughes & Short 2005, Jacobs & Wright 2006), ultimately 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 1998), street codes that demand quick and decisive retribution (Anderson 1999, Jacobs and Wright 2006, Decker 1996), commitment to delinquent peers (Esbensen et al. 1993), and social networks that promote the rapid spread of rumors (Hughes & Short

2005, Green, Horel, 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 & Greenhouse 2018). In Hawkes models clustering is attributed to both causal and non-causal mechanisms: an event occurring in a particular 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 (non-causal). 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. GRYD is guided by four foundational approaches: (1) community engagement through educational campaigns, events and public-private partnerships; (2) prevention programming aimed at youth ages 9-14 at risk for joining gangs; (3) intervention programming to assist youth and young adults 15-25 in leaving gangs; and (4) violence interruption that seeks to reduce the likelihood of retaliations when violent gang crimes do occur (Tremblay, Herz & Kraus 2019, Cespedes & Herz 2011, Cespedes 2012, Skogan et al. 2009). 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 close communication with the LAPD about gang suppression and investigative activities. This Triangle Partnership ensures that effective lines of communication remain open and allows CIWs to work separately from, but in parallel with the Los Angeles Police Department (Tremblay, Herz & Kraus 2019). However, due to both resource limitations and discretion in the reporting process, violence interruption 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. Discriminating between causal clustering and inhomogeneity 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, 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. 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 serious practical and ethical considerations;

One cannot legitimately refuse to provide services to a victim's family for the purposes of experimental purity. Rather, GRYD CIWs use their highly specialized knowledge of the local gang dynamics and intervene in areas believed to be more prone to retaliations (Tremblay, Herz & Kraus 2019). 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 & 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 4.2. 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 4.3 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 4.4 and 4.5 respectively.

4.2 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.

4.3 Methods

We note at the outset that most of our inferences are based on the particular formulation of the Hawkes model in equation (4.10) below, with background rate estimated using (4.9). 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.

4.3.1 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

$$\lambda(x, y, t) = \mu(x, y, t) + \sum_{i:t_i<t} \kappa(i) g(x - x_i, y - y_i, t - t_i), \tag{4.1}$$

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,

$$h(r) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \exp(-\frac{r^2}{2\sigma^2}). \tag{4.2}$$

Given a parametrized 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 & Vere-Jones 2003)

$$l(\Theta) = \sum_{i=1}^{N} log(\lambda(t_i, x_i, y_i|\mathcal{H}_t)) - \int_0^T \int\int_S \lambda(t, x, y|\mathcal{H}_t) dx dy dt. \tag{4.3}$$

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.

4.3.2 Background Rate Estimation

Accurate estimation of the background rate μ is critical for accurately estimating the parameters in (4.1), 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 (4.1).

4.3.2.1 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 $\beta(i)$, for example:

$$\hat{\mu}(x, y, t) = \mu(x, y) = \sum_{i=1}^{N} \frac{\beta(i)}{2\pi\eta^2 T} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2}\big). \tag{4.4}$$

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 (4.2) 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 (4.4), we obtain a weighted, variable-bandwidth, stochastically de-clustered background rate estimate:

$$\hat{\mu}(x, y, t) = \sum_{i:(x_i, y_i) \neq (x, y)}^{N} \frac{w_i}{2\pi d_i^2 T} \exp\big(-\frac{(x - x_i)^2 + (y - y_i)^2}{2d_i^2}\big). \tag{4.5}$$

In (4.5), 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 (4.1), that crime i is a back-

ground event, and is computed as

$$w_i = \frac{\hat{\mu}(t_i, x_i, y_i)}{\hat{\lambda}(t_i, x_i, y_i)}. \tag{4.6}$$

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

4.3.2.2 Temporal Variation in Background Rate

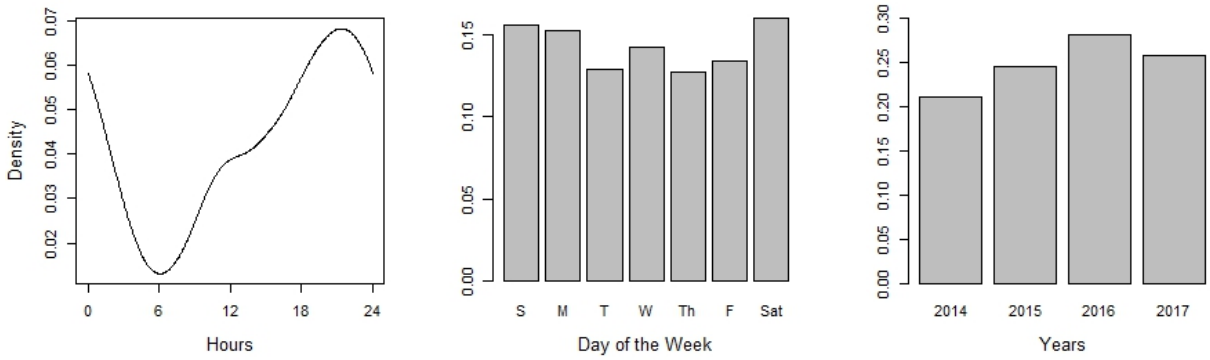
According to (4.5), the temporal density of background events is stationary $(1/T)$, and the spatial distribution does not change over time. However, given the pronounced temporal fluctuations in reported gang-related violent crimes shown in Figure 4.1, a constant temporal background rate is unrealistic and may lead to inflated estimates of productivity (*i.e.* mis-attribution 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:

$$\hat{\mu}(x, y, t) = v(t) \cdot \sum_{i:(x_i,y_i)\neq(x,y)}^{N} \frac{w_i}{2\pi d_i^2} \exp\big(-\frac{(x-x_i)^2 + (y-y_i)^2}{2d_i^2}\big). \tag{4.7}$$

To estimate the temporally varying component $v(t)$ in the background rate (4.7), we follow the component-wise 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 4.1. 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.

Figure 4.1: Temporal distribution of all gang crimes



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

4.3.2.3 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

$$\mu(x, y) = \exp\left(v(x, y)'\gamma\right), \tag{4.8}$$

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 (4.6) 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 (4.6) and update $w_i^{(l+1)} \leftarrow w_i$ for $i = 1 \dots N$.

4. Fit

$$\hat{\mu}(x, y) = \hat{f}(v(x, y)), \tag{4.9}$$

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.

4.3.3 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:

$$\lambda(x,y,t) = v(t)\mu(x,y) + \sum_{i:t_i<t} \kappa_{near}(i)\frac{h_1(r)}{2\pi r}\omega_1 e^{-\omega_1(t-t_i)} + \sum_{\substack{i:t_i<t,\\ r\geq d}} \kappa_{far}(i)\frac{h_2(r)}{2\pi r}\omega_2 e^{-\omega_2(t-t_i)}, \quad (4.10)$$

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.

4.3.4 Integral Approximation

The first term of the log-likelihood in (4.3) 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 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 (4.3) and the sum in (4.1); 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 (4.7) is:

$$l(\kappa_1, \kappa_2, \beta_1, \beta_2, \omega, \sigma, \eta) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - \sum_{i=1}^{N} \big[w_i + \kappa(i)\big]. \tag{4.11}$$

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

$$\mu(x, y, t) = c \cdot v(t), \tag{4.12}$$

whose log-likelihood is

$$l(\kappa_1, \kappa_2, c, \omega, \sigma) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - c \cdot |S| - \sum_{i=1}^{N} \kappa(i), \tag{4.13}$$

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

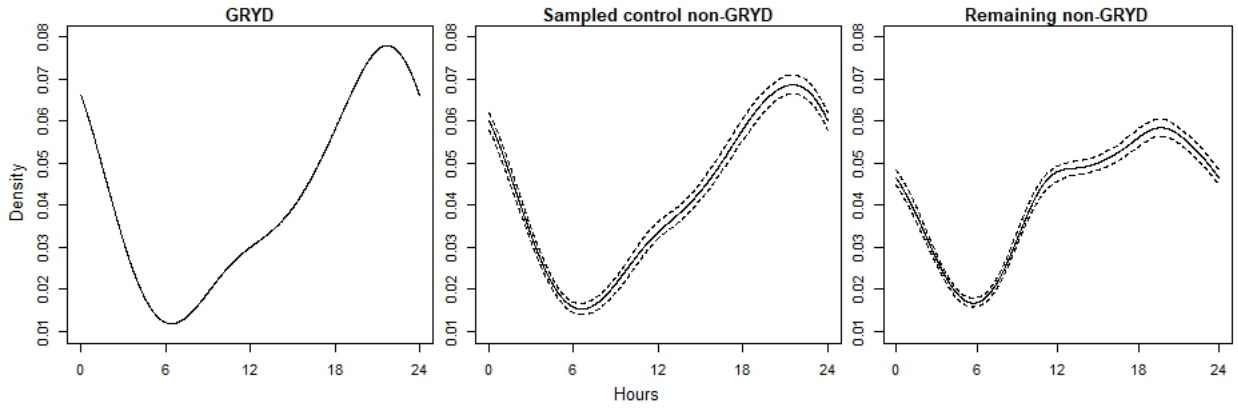
4.3.5 Sampling non-GRYD crimes as controls for GRYD IR Program crimes

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 & Kraus 2019). 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.

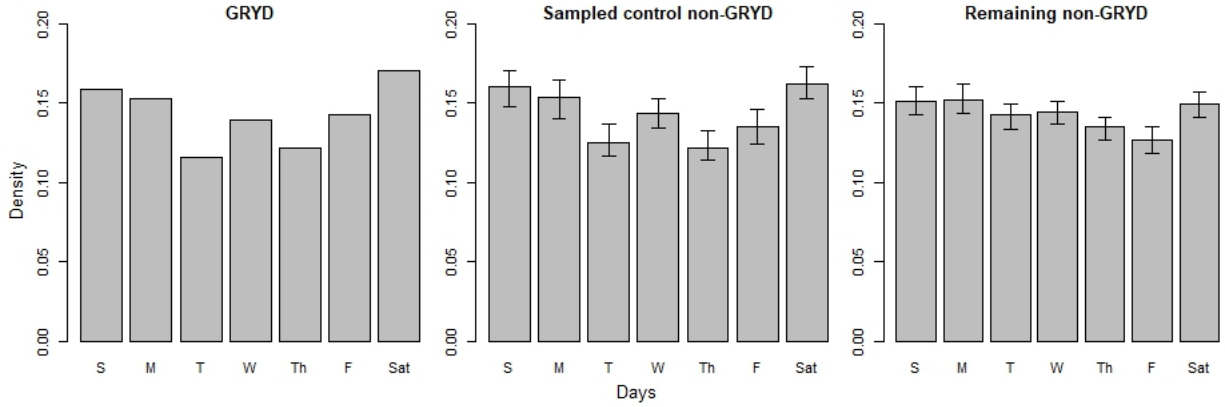
Figure 4.2: Temporal distribution of marks (hourly)



(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.

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

Figure 4.3: Temporal distribution of marks (weekly)



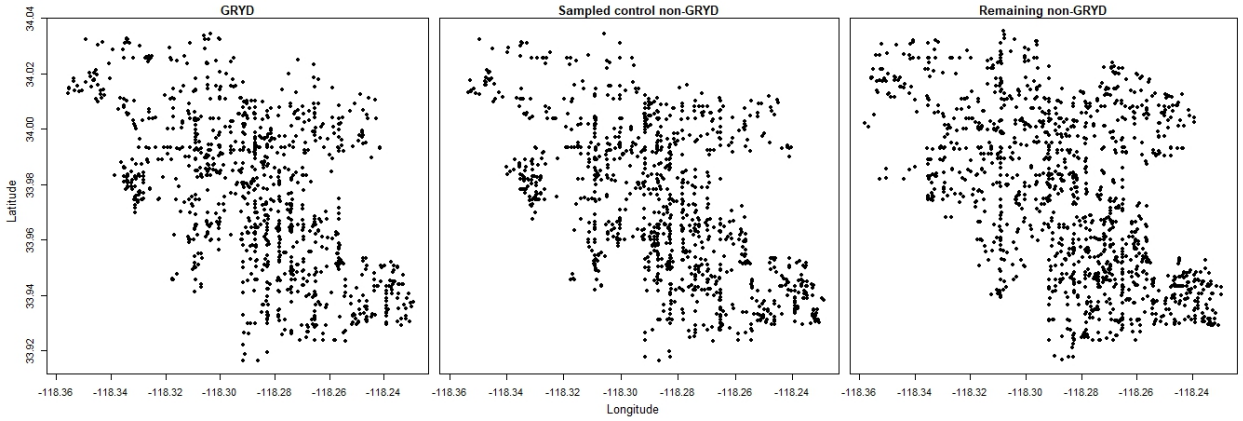
Proportion of occurences each day. (Left to Right): GRYD IR Program crime events, average of 50 samples of non-GRYD crimes and remaining unsampled non-GRYD. Whiskers: 5th and 95th percentile proportion of 50 samples.

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 (4.7) and displayed in Figures 4.2 and 4.3. 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

$$\nu_i = \frac{\hat{P}_1(x,y)\hat{P}_2(h)\hat{P}_3(d)}{\hat{Q}_1(x,y)\hat{Q}_2(h)\hat{Q}_3(d)}. \tag{4.14}$$

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 4.4. 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.

Figure 4.4: 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 4.1.

4.3.6 Evaluation Methods

Four types of models and background rate estimation methods are investigated: (I) constant background model in (4.12), (II) kernel smoothed background model in (4.7), (III) covariate background model in (4.9), all with triggering as in (4.2), and (IV) covariate background model (4.9) with near and far-field triggering as in (4.10). 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 4.3.5. 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).

4.4 Results

4.4.1 Spatial-Temporal and Covariate Effects

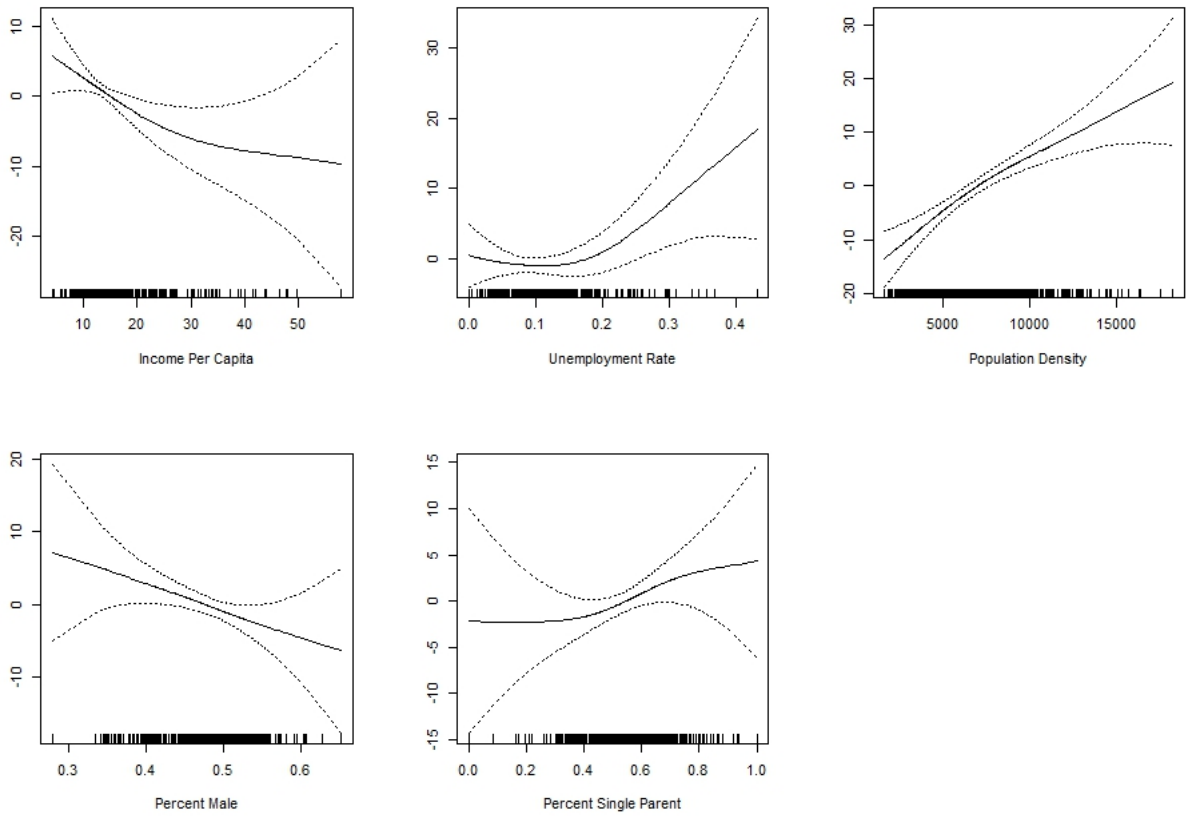
The spatial distribution of the GRYD IR Program events, non-GRYD events sampled according to (4.14), and other non-GRYD events are shown in Figure 4.4, and the temporal distributions of the three classes of events are shown in Figures 4.2 and 4.3. If GRYD IR

Program events were assigned purely at random, then the point patterns shown in the three panels of Figures 4.4, 4.2 and 4.3 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 4.4. The temporal distributions of GRYD IR Program and non-GRYD events in Figures 4.2 and 4.3 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 4.5. 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 4.6. 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.

Figure 4.5: Estimated additive predictors of GAM background

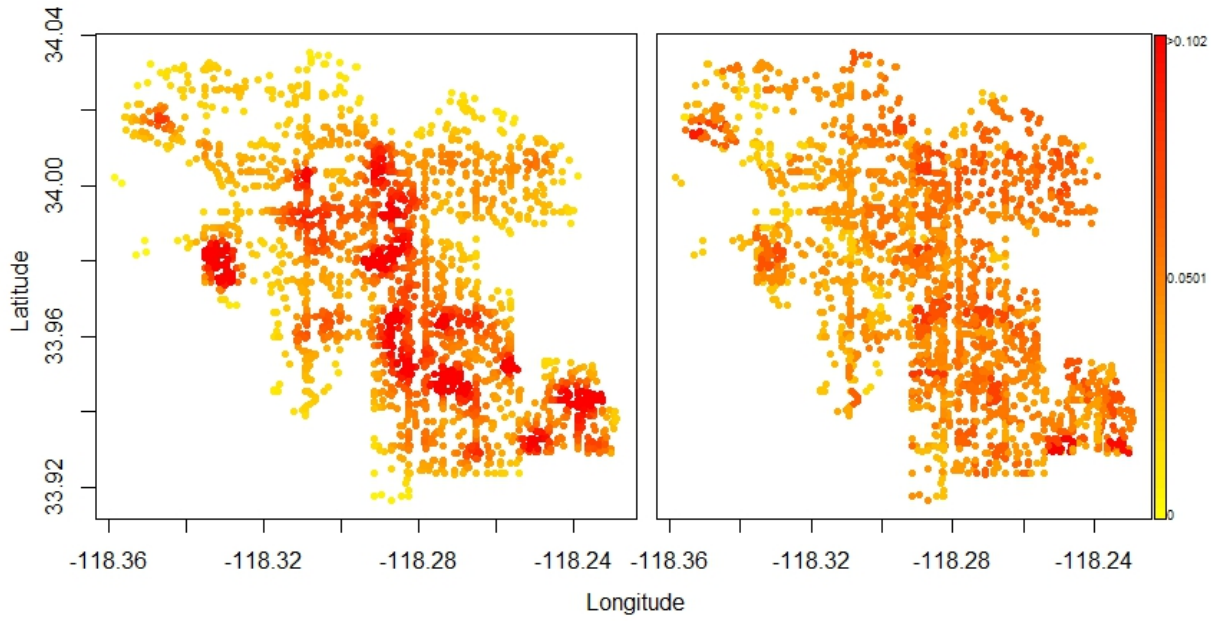


Dotted=95% confidence intervals. The y-axes are the individual additive contributions of each covariate towards the output which is in units of background crimes per square km. All covariates are estimated with 2 degrees of freedom.

4.4.2 Model Fit and Estimates

Parameter estimates and log-likelihood scores for models fit using data from 2014-2016 are reported in Table 4.1. 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

Figure 4.6: 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).

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 (4.5) used in model (II) appears less smooth than model (IV) in Figure 4.6, 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).

Table 4.1: 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.

In Table 4.2, 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.

Table 4.2: 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.

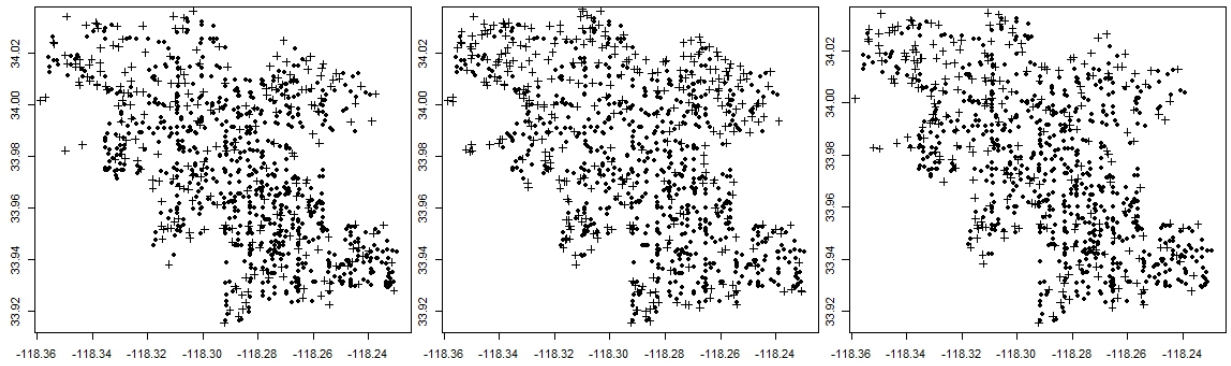
4.4.3 Out-of-Sample Evaluation

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 4.1. 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 4.7. 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.

Figure 4.7: 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.

4.4.4 Efficacy of the GRYD Program

Table 4.3 shows the average of 50 estimates of model (IV) using all 4 years of data with sampled non-GRYD control marks detailed in Section 4.3.5. 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 4.1 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 4.4,4.2 and 4.3. The parameter estimates in Table 4.3 are preferable for this purpose.

Table 4.3: 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.

4.5 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.

CHAPTER 5

Preliminary Work on Ongoing Extensions

This chapter contains some ongoing work to be published. Section 5.1 validates and assesses the procedure in Section 4.3.5. In Section 5.2 we introduce a point process model of disease spread with no background rate of infections, and smoothly varying productivities that are allowed to be greater than 1. Section 5.3 proposes a non-parametric algorithm to estimate the Recursive Point Process model introduced in Schoenberg et al. (2019).

5.1 Event-Based Interventions, Synthetic Controls and the Law of Crime Concentration

5.1.1 Introduction

Violence interventions may be non-randomized because a controlled trial would pose both practical and ethical problems. We saw evidence of this in Chapter 4, Figure 4.2 and 4.3. This caused biased in the estimates of the productivity associated with the intervened crimes, where naive results showed that violence interventions had no effect or even worsened the productivity of retaliations. The so-called *synthetic controls* introduced in Section 4.3.5 allowed an approximate treatment-to-control comparison of intervened and un-intervened crimes.

Through simulation studies in the following sections, we recreate conditions in which estimated probabilities become biased due to the non-random nature of treatments. We discover that simple Hawkes models can provide accurate productivity estimates even when treatments are non-random and that bias is only caused when ground truth productivities vary in space and time. We quantify the amount of bias caused when non-randomized interventions are erroneously assumed to be a randomized control trial and we show that the synthetic controls can be used to correctly recover the ground truth efficacy of the intervention/treatment even with non-random assignments.

5.1.2 Methods

5.1.2.1 Simulation 1

We simulate a space-time Hawkes process over the space $(x_i, y_i) \in S = [0,1] \times [0,1]$ and the window $t_i \in [0, T]$, where $T = 365$. The ground truth background rate is set to be

$$\mu(x, y, t) = N \cdot v(t)\big(a_1 G_1 + a_2 G_2\big) \tag{5.1}$$

where $a_1 = a_2 = \frac{1}{2}$ and $v(t) = (\frac{cos(t \cdot 2\pi)}{2} + 1)/T$ and G_1, G_2 are bivariate Gaussian densities with means $\mu_1 = (\frac{1}{3}, \frac{1}{3})'$ and $\mu_2 = (\frac{2}{3}, \frac{2}{3})'$ respectively and covariances $\Sigma_1 = \Sigma_2 = 0.1 * \mathrm{I}$. Here, $v(t)$ represents the diurnal pattern commonly observed in crime data and is set to integrate to 1 over $[0, T]$ such that $\int_T \int_S \mu(t) dS dt = N$. The spatial distribution of one simulation is plotted in Figure 5.1. The corresponding temporal distribution is in Figure 5.2.

The overall intensity $\lambda(x, y, t)$ follows a simple Hawkes process

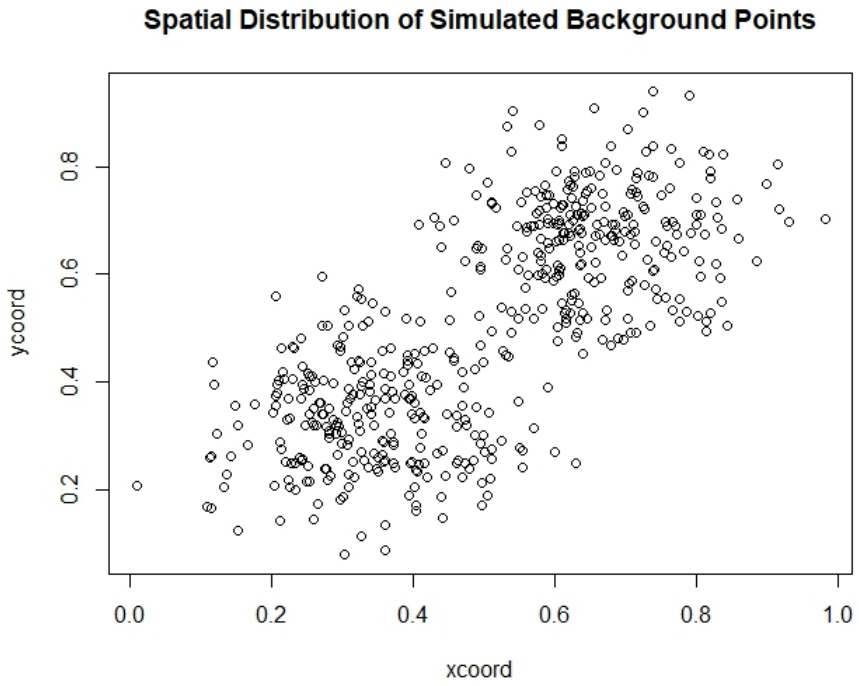


Figure 5.1: An even mixture of two Gaussian intensities with unit variance and zero covariance.

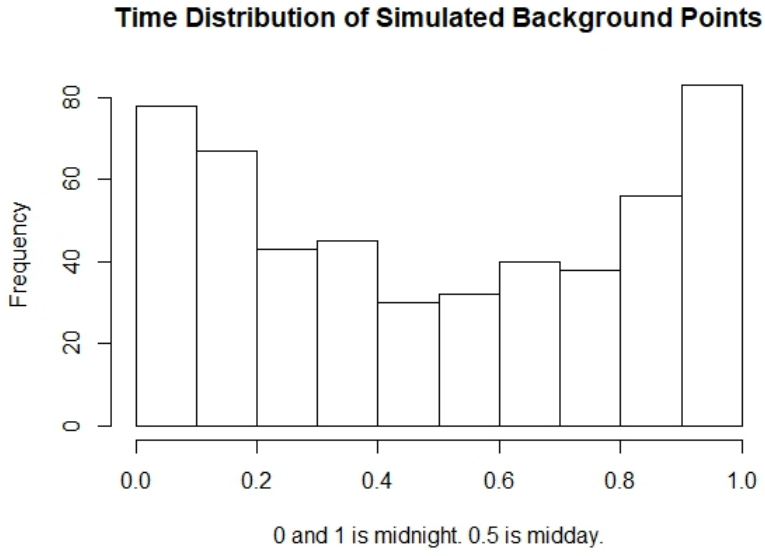


Figure 5.2: One unit of time is equal to 24 hours. The simulation roughly mimics diurnal cycles. In Los Angeles data, crimes bottom out at 6am and peak around 11pm.

$$\lambda(x, y, t) = \mu(x, y, t) + \sum_{i:t_i<t} \kappa(i) g_1(x - x_i, y - y_i) g_2(t - t_i). \tag{5.2}$$

As in Park et al. (JASA, rev. & resubm.), the temporal triggering density g_2 is set to decay exponentially, $g_2(t - t_i) = \omega e^{-\omega(t-t_i)}$, and $g_1(x, y) = h(r)/2\pi r$ where $h(r)$ is the probability density function for the distance r between a crime and its triggered crimes and h is the truncated Gaussian density over the real half-line with spatial triggering bandwidth σ. The productivities $\kappa(i) = \kappa_1 \mathbb{I}(\iota_i = 1) + \kappa_2 \mathbb{I}(\iota_i = 0)$. The simulation of the $(n \times 1)$ vector of interventions, $(\iota_1 \dots \iota_n)$, is detailed below (ι_i=1 if a crime i receives intervention). Here we set $\kappa_1 = 0.2$, $\kappa_2 = 0.4$, $\sigma = 0.01$ and $\omega = 1/7$ (i.e. average time to retaliation is one week).

Given a simulated set of n points, each intervention $\iota_i \in (\iota_1 \dots \iota_n)$ is a Bernoulli random variable with success $p_i = \mu(x_i, y_i, t_i)/D$ such that times and areas with higher rates of crime are more likely to receive an intervention. We set $D = 50$ which roughly equates to about 11% of crimes to be intervened on. In Figure 5.3 and 5.5 we plot the spatial and temporal (respectively) distribution of a simulated set of points and its interventions under this setting. With such interventions we hypothesize that biased productivity estimates will be produced. Specifically, that the estimate of k_1 will be biased upward. On the other hand, if interventions are perfectly randomized, we hypothesize that estimates of κ_1 and κ_2 will be unbiased. For comparison, we also simulated a perfectly randomized intervention and plot its distribution in Figure 5.4. We repeat the simulation 50 times and examine the median of 50 estimates.

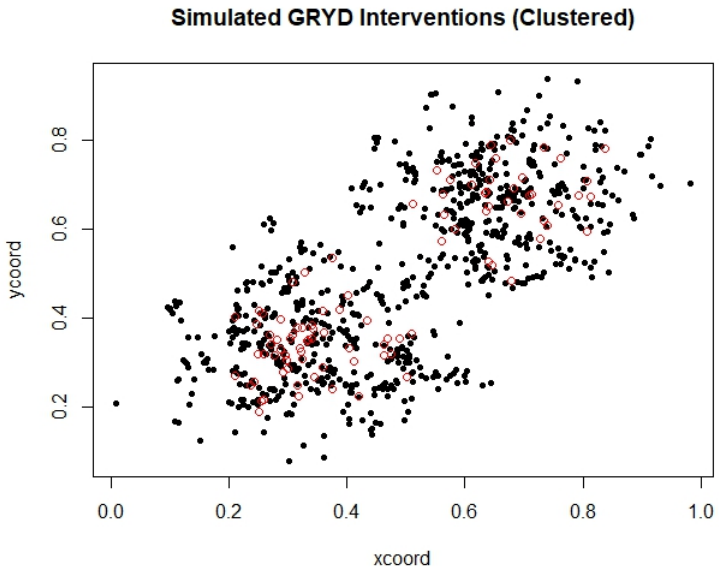


Figure 5.3: Clustered intervention probability $p_i = \mu(x_i, y_i, t_i)/D$. Black represents unintervened simulated crimes. Red represents interventions.

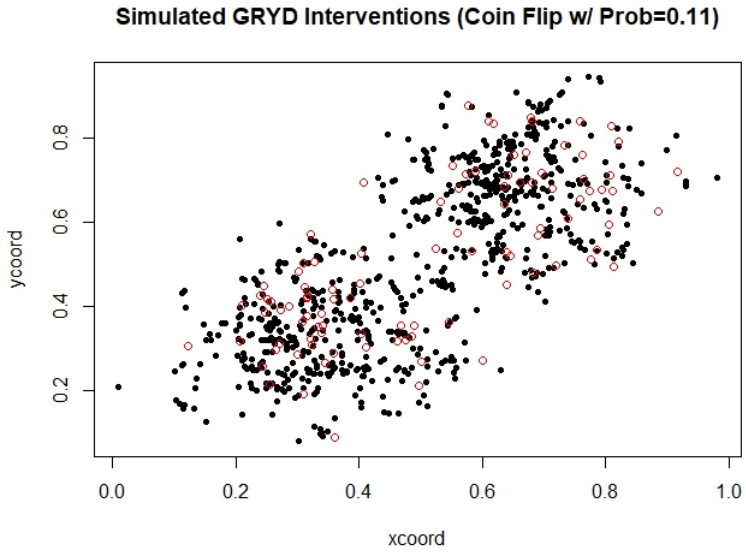


Figure 5.4: Random intervention where probability p_i set to 0.11. Black represents unintervened simulated crimes. Red represents interventions.

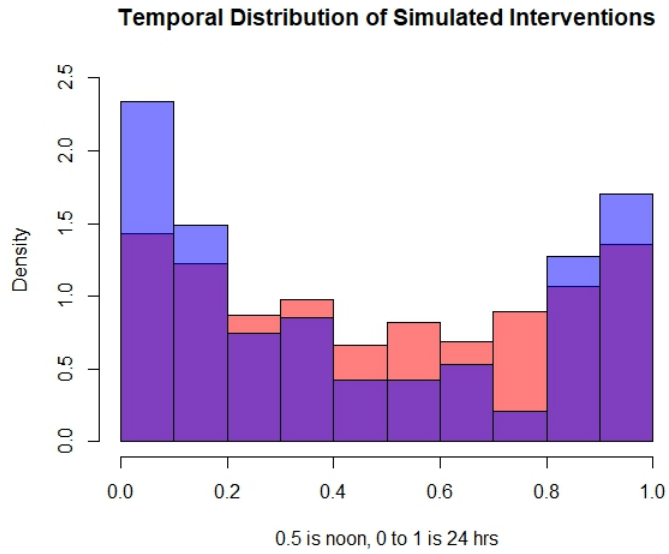


Figure 5.5: Transparent blue represents temporal distribution of interventions with probability $p_i = \mu(x_i, y_i, t_i)/D$. Red represents the same for un-intervened crimes.

5.1.2.2 Simulation 2: Time and Space Dependent Productivities

Here we maintain all ground truth parameters set in Section 5.1.2.1, except that productivities κ_i not only depend on the intervention but also on space and time. Specifically,

$$\lambda(x, y, t) = \mu(x, y, t) + \sum_{i:t_i<t} \kappa(x_i, y_i, t_i, \iota_i) g_1(x - x_i, y - y_i) g_2(t - t_i).$$

Here, we define a point's community membership to A or B. We call any descendant of a point generated from the process G_1 or G_2 in (5.1) to belong to community A or to B, respectively. Let $\mathbb{I}_A(i)$ and $\mathbb{I}_B(i)$ indicate community membership. We also define a point i to have occurred during the 'day' with the indicator $\mathbb{I}_{\text{day}}(i) = 1$ if $(t_i \mod 1) \in [0.2, 0.8]$. Also consider the treatment indicator $\mathbb{I}(\iota_i = 1)$ from (5.2) and define r to be the rate of violence

reduction due to an intervention. The ground truth productivity is set to be

$$\kappa(x_i, y_i, t_i, \iota_i) = \big(1 - r \cdot \mathbb{I}(\iota_i = 1)\big)\Big(\alpha_1 * \Big[\mathbb{I}_B(i) + \mathbb{I}_A(i)\mathbb{I}_{\text{day}}(i)\Big] + \alpha_2 * \mathbb{I}_A(i)\Big[1 - \mathbb{I}_{\text{day}}(i)\Big]\Big). \quad (5.3)$$

Simply put, all un-intervened points have a productivity of $\alpha_1 = 0.2$ unless if it belongs to community A and occurs at night, where $\alpha_2 = 0.4$. Hence, community B is safer than A because in the latter, there is potential for crimes to occur at night and have a high productivity of $\alpha_2 > \alpha_1$. All interventions reduce the productivity by $r = 0.25$.

We simulate the interventions to occur more frequently in times and places where it is less safe, similar to what is thought to have been observed in Park et al. (JASA, rev. & resubm.). Therefore night crimes belonging to community A receive an intervention with probability 1/3 and all other crimes with probability 1/9. This equates to about 16% of crimes receiving an intervention.

After simulating such a point process, we estimate the model (50 repetitions) in (5.2), where κ_1 represents the productivity associated to crimes receiving an intervention and κ_2 for all other crimes. Since around 30% of points are night crimes in community A, we expect the productivity estimate of un-intervened crimes to be a weighted average of roughly $(1 - 0.3)\alpha_1 + 0.3\alpha_2 = 0.26$. Since interventions reduce the productivity of all crimes by r%, we expect the estimate to be roughly $(1 - r) * 0.26 \approx 0.2$ if there is no bias. We also repeat the simulation and estimation where interventions have no effect, $r = 0$.

5.1.2.3 Simulation 3: Synthetic Controls

We hypothesize that productivity estimates will be biased upward for interventions that are non-randomly assigned as they are in Simulation 1 or Simulation 2. In order to reduce this bias and better evaluate the efficacy of interventions, we implement the procedure introduced in Park et al. (JASA, rev. & resubm.) to randomly subsample un-intervened crimes as synthetic controls. The resulting space-time distribution of the synthetic controls are similar to the actual interventions.

For a simulated dataset of n points, the number of un-intervened crimes is $n - \sum_{i=1}^{n} \iota_i = J$. For indices $j \in (1 \dots J)$ we compute weights $\nu_j = \hat{P}_1(x,y)\hat{P}_2(t)/\hat{Q}_1(x,y)\hat{Q}_2(t)$ where $\hat{P}_1$ and $\hat{P}_2$ is an estimate of the space-time distribution of interventions and $\hat{Q}_1$ and $\hat{Q}_2$ is an estimate of the space-time distribution of un-intervened crimes. The set of un-intervened crimes is sampled using weights ν_j without replacement to obtain a set of synthetic controls. Then the model in (5.2) is estimated with three marks instead of two, where κ_1 represents productivities for interventions, κ_2 for synthetic controls and κ_3 for all remaining crimes. For a given simulated catalog of crimes, this can be repeated multiple times. We simulate 50 point processes and repeat this sub-sampling procedure 30 times, for a total of 1500 estimations.

5.1.2.4 Simulation 4: No Triggering and No Interventions

A major source of criticism against the application of Hawkes models for crimes is the assertion that crimes are purely driven by spatial and temporal inhomogeneity and that Hawkes models are over-specified and mis-attributing the inhomogeneity as triggering. Therefore the

reliability of the magnitudes of estimated productivities is also a subject of interest here. We simulate a process that is purely driven by the background rate in (5.1) and estimate a misspecified model with a triggering term and a single productivity parameter.

5.1.3 Estimation

We use the same basic kernel smoother for all background rate estimates in Simulation 1-4,

$$\hat{\mu}(x, y, t) = v(t) \sum_{i=1}^{N} \frac{\beta}{2\pi\eta^2} \exp\big(- \frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2}\big) \tag{5.4}$$

where $v(t)$ is estimated as in Fox et al. (JASA, 2016) and Park et al. (JASA, rev. & resubm.). Parameter estimates are obtain by maximizing the log-likelihood using the quasi-Newton method developed by Broyden, Fletcher, Goldfarb and Shanno (1970). For instance, the log-likelihood for models estimated for Simulations 1 and 2 is approximated as

$$l(\kappa_1, \kappa_2, \beta, \omega, \sigma, \eta) = \sum_{i=1}^{N} \log[\lambda(x_i, y_i, t_i)] - \sum_{i=1}^{N} \Big[\beta + \kappa(i) * \big[1 - \exp\big(- \omega(T - t_i)\big)\big]\Big]. \tag{5.5}$$

5.1.4 Synthetic Controls Without Repeated Simulation

After computing sampling weights ν_j as detailed in Section 5.1.2.3, surrogate controls are sub-sampled from un-intervened crimes to simulate indicators $\iota_i^{control} = 1$ if the i^{th} is selected to serve as an approximate control. The productivity of the model in Simulation 3 is $\kappa(i) = \kappa_1\iota_i + \kappa_2\iota_i^{control} + \kappa_3(1 - \iota_i^{control})$ where ι_i are indicators representing the actual treatments, consistent in the above sections.

This requires multiple draws of $\iota_i^{control}$ and estimations for each draw and can be time consuming. The productivity of treated crimes κ_1 and the productivity of an approximate control crime κ_2 could be obtained *without* repeated simulation and estimation by defining the productivity as $\tilde{\kappa}(i) = \kappa_1\iota_i + \kappa_2\nu_i + \kappa_3(1-\nu_i)$, hence using the sampling weights defined in Section 5.1.2.3 as covariates where the productivity of an untreated crime is a weighted average of κ_2 and κ_3; The weight is higher for crime i if it is similar in space and time distribution to treated crimes. This may seem like an arbitrary formulation but it derives from the fact that $E(\kappa(i)) = \tilde{\kappa}(i)$.

5.1.5 Results

The median estimated productivity of intervened and un-intervened crimes in Simulation 1 was 0.213 and 0.376 respectively. Contrary to our hypothesis, this is reasonably close to the ground truth of 0.2 and 0.4. This means that a simple Hawkes model can correctly estimate the productivity of interventions that are non-random and concentrated in times and places where the background rate is high.

The median estimated productivity of intervened and un-intervened crimes in Simulation 2 was 0.240 and 0.222 respectively. In the ground truth simulation, interventions were reducing the productivity by 25% but these biased estimates would suggest that the intervention has no effect or causes more crime. In the other ground truth where interventions reduced the productivity by 0%, the median estimated productivity of intervened and un-intervened crimes was 0.290 and 0.232. Simply as a result of the non-random assignment that favored times and areas with high productivity, the estimated of productivity of the interventions

was 25% higher than crimes that had the same ground truth productivity.

In Simulation 3, after 1500 simulations and estimations, the median productivity of interventions, their synthetic controls, and other crimes was (0.230, 0.306, 0.214). The percent decrease in the intervention productivities relative to the synthetic controls is 25% which is consistent with the ground truth reduction of 25%.

The median of 50 productivity estimates resulting from Simulation 4 was arbitrarily close to zero. This shows evidence that productivity estimates will be close to zero for point processes generated with only spatial and temporal inhomogeneity, even when using a simple background rate estimator.

5.1.6 Discussion

Through simulation studies, we have validated and assessed the sub-sampling procedure in Park et al. (JASA, rev. & resubm.). We found that simple Hawkes processes can accurately recover the true intervention effect even when interventions are clustered and non-random and erroneously assumed to be a randomized trial. We discovered that the bias from non-random interventions only appears when the ground truth productivity varies in space-time and when interventions favor areas with higher productivity. Even with a simple model where productivities are constant in space and time, we are able to accurately recover true intervention effect. Further work on providing measures of uncertainty and examining the sensitivity to simulation parameters is in order.

5.2 A Purely Epidemic Non-Stationary Point Process Model of Infectious Disease Spread

One constraint of a Hawkes model in the context of disease spread is that events are generated through two sources: (1) an *endemic* background component and (2) a self-exciting, *epidemic* component. This is appropriate for diseases that can be contracted both exogenously from the environment or endogenously from an infected neighbor, such as in Meyer et al. (2012). But this may be inappropriate for many types of diseases which are spread purely from infections beginning with one 'patient-zero'. Brémaud & Massoulié (2001) studied a Hawkes model only sustained by the epidemic component and found that for a finite non-degenerate long run average intensity, the average number of off-springs per event (i.e. 'productivity') must not only be constant in time but strictly equal to one. During a actual outbreak, the productivity is time-variant and at times greater than one.

This section outlines preliminary work on a purely epidemic point process model of disease spread and propose an algorithm to estimate the productivity as a smooth function of time that is not restricted to be equal or less than one, and also propose/derive several estimates of the temporal distribution of off-springs. The accuracy of the estimation procedure can be examined using simulated catalogs of infections. A simulation based approach for uncertainty quantification of the estimates is proposed.

Consider a sequence of infections $0 < t_1 < t_2 < \ldots < t_N$. The intensity of occurrences at any point t is time variant and depends on the history $\mathcal{H}_t$ of all points leading up to time t. The conditional intensity is the infinitesimal rate at which points occur at time t:

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

We formulate a conditional purely driven by self-excitation as:

$$\lambda(t) = \sum_{j:t_j<t_N} \theta(t_j) g(t - t_j).$$

Any point occurring at time t_j causes secondary infections into the future at a rate of $\theta(t_j)g(t-t_j)$. Here g integrates to 1 and describes the temporal distribution of secondary infections. The expected number of cases caused by any infection occurring at t_j is represented by $\theta(t_j)$.

Besides t_1 (i.e. patient zero), all other infections must be triggered by some other prior infection. A latent branching structure χ_{ij} can be written according to this, where

$$\chi_{ij} = \begin{cases} 1, & \text{if case } i \text{ was caused by a prior case } j, \quad j < i \\ 0, & \text{otherwise.} \end{cases}$$

Given χ_{ij}, the complete data log-likelihood can be written as,

$$l_c(t_1, \ldots, t_N; \theta, g) = \sum_{j=1}^{N} \left[\sum_{i>j} \chi_{ij} \log \left(\theta(t_j) g(t_i - t_j) \right) - \int_{t_j}^{t_N} \theta(t_j) g(t - t_j) dt \right].$$

Although χ_{ij} are unobserved, it can be probabilistically defined with the matrix P,

$$P = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & 0 \\ p_{31} & p_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ p_{N1} & p_{N2} & \cdots & p_{N,N-1} & 0 \end{bmatrix}. \tag{5.6}$$

The elements p_{ij} represent the probability that an infection i was caused by a prior infection j. Note that the second observed case t_2 is necessarily caused by t_1, hence $p_{21} = 1$. Since all points must have been triggered by a prior case, $\sum_{j=1}^{N-1} p_{ij} = 1$ for any i.

Given the probabilistic branching structure P and a model for λ, the E-step at some iteration v of an EM-algorithm can be written as

$$E\big[\chi_{ij}|\theta^{(v)}, g^{(v)}\big] = p_{ij}^{(v)} = \frac{\theta^{(v)}(t_j)g^{(v)}(t_i - t_j)}{\lambda^{(v)}(t_i)},$$

where $\theta^{(v)}$ and $g^{(v)}$ represent the current estimates of the productivity and triggering function. The expected complete data log-likelihood is then given by

$$E\big[l_c(\theta, g)|\theta^{(v)}, g^{(v)}\big] = \sum_{j=1}^{N} \sum_{i>j} p_{ij}^{(v)} \Big(\log \theta(t_j) + \log g(t_i - t_j) \Big) - \sum_{j=1}^{N} \theta(t_j) \cdot G(t_N - t_j) \tag{5.7}$$

where $G(x) = \int_0^x g(s)ds$. Note that there is no integral approximation in the second summation of (5.7). In many applications like in Fox et al. (2016), an integral approximation allows tractable estimates in the M-step at the expense of some acceptable amount of boundary

error. But as will be seen below, in our study, this would cause underestimation of the productivity of infections towards the end of the dataset. It is crucial to correct this in order to generate prospective forecasts of the productivity.

The simple approach proposed here for finding smooth estimates of θ and g is to smooth over point-wise optimums. The first order condition for the θ at some point t_j is

$$\frac{\partial E}{\partial \theta(t_j)} = \sum_{i>j} p_{ij}^{(v)} \frac{1}{\theta(t_j)} - G(t_N - t_j) = 0 \tag{5.8}$$

$$\Rightarrow \tilde{\theta}(t_j) = \frac{1}{G(t_N - t_j)} \sum_{i>j} p_{ij}^{(v)}. \tag{5.9}$$

Simply put, the estimated productivity of point t_j is the j^{th} column sum of P. For most points, $G \approx 1$, but for points closer to t_N, the estimated productivity is a column sum of P with an inflation factor of $1/G(t_N - t_j)$. In similar fashion to θ, we find point-wise optimums of g at inter-time differences s_k, where $k = 1 \dots N(N-1)/2$. We replace the integral G in (5.7) with its Riemann approximation in order to take the partial derivative of the expected complete data log-likelihood with respect to g at some inter-time distance s_k,

$$\frac{\partial E}{\partial g(s_k)} = p_{s_k}^{(v)} \frac{1}{g(s_k)} - \left(\sum_{j: s_k < t_N - t_j} \theta(t_j) \right) g(s_k) \Delta s_k = 0. \tag{5.10}$$

The notation p_{s_k} refers to a certain ij^{th} element of the matrix P where $t_i - t_j = s_k$. The superscript v denotes the estimate of p_{s_k} at the v^{th} iteration. The summation in (5.10) is due to a given $g(s_k)$ appearing in only a subset of the summands in the second sum of (5.7). Riemann intervals are denoted by Δs_k.

Due to not having integrated away G in (5.7) with an approximation, it is difficult to

analytically separate and solve for both θ and g. The maximum at $g(s_k)$ depends on the maximized value of θ at various time points, which in turn depends on the integral G in (5.9). Since g is generally expected to decay over time and $G \approx 1$ for most points at the beginning of a dataset, we suggest substituting G in (5.9) with the estimate of G from the previous iteration. For $v = 1$, a reasonable guess such as an exponential decay can be used and the sensitivity of the results can be checked

$$\tilde{\theta}(t_j) = \frac{1}{G^{(v-1)}(t_N - t_j)} \sum_{i>j} p_{ij}^{(v)}.$$

Plugging in $\tilde{\theta}$ and solving for $g(s_k)$ in (5.10) yields,

$$\tilde{g}(s_k) = \sqrt{\frac{p_{s_k}^{(v)}}{\Delta s_k \sum_{j:s_k<t_N-t_j} \theta(t_j)}}.$$

Finally, kernel regression can be used to smooth over $\tilde{\theta}$ and $\tilde{g}$ to obtain smooth estimates $\hat{\theta}$ and $\hat{g}$.

There are some alternative proposals for the estimator of g that would yield similar results. The distribution g can be specified as a histogram with fine bins containing only one or no points. These bin heights can be maximized by the first order condition in (5.10) where the partial derivative of the second summand in (5.7) can be treated as negligibly close to zero with fine Riemann bins and a Lagrange multiplier for the constraint that the histogram integrates to one can be added. The noisy bin heights can be smoothed using kernel regression. Alternatively, we could find the kernel density estimate of all inter-time differences s_k over the support $(0, \infty)$ with weights $p_{s_k}^{(v)}$.

This estimation procedure can be validated by simulating catalogs of infections using a ground truth for θ and g then examining the estimates. Uncertainty of any one estimate from a given dataset can be found by repeatedly simulating catalogs of infections using $\hat{\theta}$ and $\hat{g}$ and tabulating the variability of each re-estimate. Prospective forecasts can be obtained by extrapolating $\hat{\theta}$ with some polynomial fit and simulating catalogs into the future.

5.3 Non-Parametric Estimation of the Recursive Point Process Model

Schoenberg et al. (2019) introduced the Recursive Hawkes model

$$\lambda(t) = \mu + \int_0^t H(\lambda_{t'}) g(t - t') \mathrm{d}N(t'),$$

where $\mu > 0$, $g > 0$ is a density function, and $\lambda_{t'}$ means $\lambda(t')$. The novel feature of this model is that the productivity of a point t' not only varies in time but depends on the conditional intensity at time t' through the parametric function H.

In this section, the author of this thesis proposes a non-parametric estimation of Schoenberg's Recursive Hawkes model. This methodology is implemented as joint work in Kaplan et al. (subm.) in a comparison to the SVEILR (Susceptible, Vaccinated, Exposed, Infected, Lightly infected, Recovered) compartmental model applied to data on mumps in Pennsylvania.

The primary challenge in implementing an EM-styled estimation here is that the elements of P cannot be used to identify the bin heights of both g and H simultaneously; in order

to estimate the heights of g, we need to have updated elements of P, which requires having updated λ and this cannot be done without an estimate of H which in turn depends on λ. The key 'trick' in this methodology is to begin the iterations with a naive guess of H as a constant. We suggest $\sum_i \sum_{i>j} \hat{p}_{ij}/n$. In subsequent iterations, the bin heights of g are updated using 'fresh' updated values of P while the heights of H are updated with 'stale' or trailing estimates of λ from the previous iteration.

Suppose a purely temporal point process of the form $\tau_1, \tau_2, \ldots, \tau_n$ is observed over the time interval $[0, T]$ where n is th number of events observed. Define P as a matrix of probabilities such that elements p_{ij} is the estimated probability that event i was triggered by a prior event j. This probability matrix is lower-triangular since an event j can only trigger later events. Each diagonal entry p_{ii} represents the probability that infection i is a background event stemming from the process μ. The sum of any row of P must therefore equal 1 since each event must either be a background event or have been triggered by some prior event. The temporal distribution of triggered events, g, is assumed to be a step function and H is similarly estimated as a step function with bins corresponding to intervals of the conditional intensity λ. The background rate μ is constant and is an estimated parameter.

After randomly initializing P and guessing H as a constant, the M-step update of a certain bin height of g is given as

$$\hat{g}_l = \frac{\sum_j \sum_{i>j} \mathbb{I}(\tau_i - \tau_j \in B_l)\hat{p}_{ij}}{w_l \sum_i \sum_{i>j} \hat{p}_{ij}}$$

where B_l represent a bin corresponding to an interval of the real line containing some of the

interevent times and w_l is the bin width. The background rate μ is updated as

$$\hat{\mu} = \frac{1}{T}\sum_{i}^{n} \hat{p}_{ii}.$$

With estimates of g, μ and a *previous* trailing estimate of H, we can calculate λ as

$$\hat{\lambda}(\tau_i) = \hat{\mu} + \sum_{j=1}^{i} \hat{H}_j \hat{g}(\tau_i - \tau_j).$$

With the newly calculated λ, the bin heights of $H_i = H(\lambda(\tau_i))$ can be updated. For each bin C_k corresponding to an interval of the real line containing some values of $\hat{\lambda}(\tau_j)$, the k^{th} bin height given by

$$\hat{H}_k = \frac{\sum_j \sum_{i>j} \mathbb{I}(\hat{\lambda} \in C_k)\hat{p}_{ij}}{\sum_j \mathbb{I}(\hat{\lambda} \in C_k)}$$

provided $\sum_j \mathbb{I}(\hat{\lambda} \in C_k) > 0$ and $\hat{H}_k = 0$ otherwise.

After updating g, μ and H from the M-step, we can update the matrix P in the E-step.

$$\hat{p}_{ij} = \frac{\hat{g}(\tau_i - \tau_j)\hat{H}_j}{\hat{\mu} + \sum_{k=1}^{i-1} \hat{g}(\tau_i - \tau_k)\hat{H}_k}.$$

These iterations are repeated until a convergence criteria is met, such as the maximum change update of an element of P being less than ϵ.

CHAPTER 6

Conclusion and future work

We have presented advances in the methodology and application of point process models. For disease spread, we showed that point process models have great potential in becoming a useful tool for forecasting diseases and performed residual analysis to assess its descriptive performance. Further testing and real time validation is needed in order to increase the popularity and credibility of these methods. Further research on how to non-parametrically model contagion through space will be an important topic, but will depend on the availability of high resolution surveillance data. In modeling crimes, we have tackled the deep problem of identifying true inhomogeneity and clustering. In doing so we introduced a method to non-parametrically estimate the background rate as a function of spatial covariates. In evaluating non-randomized event-based interventions, we have introduced a simple method to generate synthetic controls to improve causal interpretation of estimates. Through simulation studies we have quantified the potential bias caused by erroneously treating non-random interventions as a randomized-trial and demonstrated that synthetic controls can effectively be used to recover the true intervention effect. Future studies in modeling non-separable marks and variable space-time productivities will be important.

The author hopes to continue research in these regards as well as in other outstand-

ing challenges in the field such as: studies in marked spatio-temporal point processes and predictability of marks; dealing with large N and high dimensional point processes; non-likelihood based inference; advancing methods and applications for neuro-science and network science, residual analysis for space-time models, and filling in the theoretical gap in the statistical properties of modern non-parametric models and estimators for point process models.

REFERENCES

[1] C. 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] J. Asher. Forecasting ebola with a regression transmission model. *Epidemics*, 22:50–55, 2018.

[4] E. Balderama, F. Schoenberg, E. Murray, and P. 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.

[5] M. Barton, M. Valasik, E. Brault, and G. Tita. "gentefication" in the barrio: Examining the relationship between gentrification and homicide in east los angeles. *Crime & Delinquency*, page 0011128719860835, 2019.

[6] L. Bauwens and N. Hautsch. *Modelling Financial High Frequency Data Using Point Processes*, pages 953–979. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[7] 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.

[8] N. Becker. Estimation for discrete time branching processes with application to epidemics. *Biometrics*, 33(3):515–522, 1977.

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

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

[11] P. Brantingham, G. Tita, M. Short, and S. Reid. The ecology of gang territorial boundaries. *Criminology*, 50(3):851–885, 2012.

[12] P. 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.

[13] P. Brantingham, M. Valasik, and G. Tita. Competitive dominance, gang size and the directionality of gang violence. *Crime Science*, 8(1):7, 2019.

[14] T. Britton. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225(1):24–25, 2010.

[15] J. Brocker and L. A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388, 2007.

[16] 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.

[17] Y. Cao, D. Gillespie, and L. Petzold. Adaptive explicit-implicit tau-leaping method with automatic tau selection. *J Chem Phys*, 126(22):224101, 2010.

[18] Centers for Disease Control & Prevention. Ebola (ebola virus disease): History of ebola virus disease: 2014–2016 ebola outbreak in west africa: Case counts. 2019.

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

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

[21] D. Champredon, M. Li, B. Bolker, and J. Dushoff. Two approaches to forecast ebola synthetic epidemics. *Epidemics*, 22:36–42, 2018.

[22] G. Chowell, N. Hengartner, C. Castillo-Chavez, P. Fenimore, and J. Hyman. The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. *J Theor Biol*, 229(1):119–126, 2004.

[23] G. Chowell, C. Viboud, L. Simonsen, S. Merler, and A. Vespignani. Perspectives on model forecasts of the 2014–2015 ebola epidemic in west africa: lessons and the way forward. *BMC Med.*, 15(1):42, 2017.

[24] J. Chretien, S. Riley, and D. George. Mathematical modeling of the west africa ebola epidemic. *Elife*, 4, 2014.

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

[26] G. D. Curry and I. A. Spergel. Gang homicide, delinquency, and community. *Criminology*, 26(3):381–406, 1988.

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

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

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

[30] I. Damon, P. Rollin, M. Choi, R. Arthur, and R. Redfield. New tools in the ebola arsenal. *N. Engl. J. Med.*, 379(21):1981–1983, 2018.

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

[32] O. Diekmann and J. Heesterbek. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. *Chichester, UK: Wiley*, 2000.

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

[34] C. Farrington, M. Kanaan, and N. Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, 4(2):279–295, 2003.

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

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

[37] E. Fox, F. Schoenberg, and J. 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.

[38] S. Funk, A. Camacho, A. Kucharski, R. Eggo, and W. Edmunds. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, 22:56–91, 2018.

[39] S. Funk, A. Camacho, A. Kucharski, R. Lowe, R. Eggo, and J. Edmunds. Assessing the performance of real-time epidemic forecasts: a case study of the 2013–16 ebola epidemic. *BioRxiv*, 2018.

[40] S. Funk, I. Ciglenecki, A. Tiffany, E. Gignoux, A. Camacho, R. Eggo, et al. The impact of control strategies and behavioural changes on the elimination of ebola from lofa county, liberia. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.*, 372(1721), 2017.

[41] F. Gerhard, M. Deger, and W. Truccolo. On the stability and dynamics of stochastic spiking neuron models: nonlinear hawkes process and point process glms. *PLoS Comput. Biol.*, 13(2):e1005390, 2017.

[42] T. Gneiting and M. Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Applications*, 1(1):125–151, 2014.

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

[44] B. Green, T. Horel, and A. 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.

[45] E. Guofo, S. Noutchie, and S. Mugisha. A fractional seir epidemic model for spatial and temporal spread of measles in metapopulations. *Abstract and Applied Analysis*, 781028:1–6, 2014.

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

[47] D. Harte and D. Vere-Jones. The entropy score and its uses in earthquake forecasting. *Pure Appl. Geophys.*, 162(6-7):1229–1253, 2005.

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

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

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

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

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

[53] L. Hunt, A. . Gupta-Wright, V. Simms, F. Tamba, V. Knott, K. Tamba, S. Heisenberg-Mansaray, E. Tamba, A. Sheriff, S. Conteh, T. Smith, S. Tobin, T. Brooks, C. Hoolihan, R. Cummings, and T. Fletcher. Clinical presentation, biochemical, and haematological parameters and their association with outcome in patients with ebola virus disease: an observational cohort study. *The Lancet Infectious Diseases*, 15(11):1292 – 1299, 2015.

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

[55] D. Johnson. Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 3(4):275–299, 1996.

[56] P. Johnson. *adaptivetau: Tau-Leaping Stochastic Simulation*, 2016. R package version 2.2-1.

[57] M. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals.* Princeton University Press, Princeton, N.J., 1 edition, 10 2007.

[58] J. Kelly et al. Real-time projections of ebola outbreak size and duration with and without vaccine use in equateur, democratic republic of congo. *BioRxiv*, page 10.1101/331447, 2018.

[59] J. Kelly, J. Park, et al. Real-time predictions of the 2018-2019 ebola virus disease outbreak in the democratic republic of the congo using hawkes point process models. *Epidemics*, 28:100354, 2019.

[60] W. Kermack and A. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115(771):700–721, 1927.

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

[62] C. Kubrin and R. Weitzer. Retaliatory homicide: Concentrated disadvantage and neighborhood culture. *Social problems*, 50(2):157–180, 2003.

[63] D. Kyriacou, H. Hutson, D. Anglin, C. Peek-Asa, and J. 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.

[64] P. Lekone and B. Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–7, 2006.

[65] Liberia Institute of Statistics and Geo-Information Services (LISGIS). Population and housing census: Final results. (2008). Accessed August 31, 2017.

[66] M. Lipsitch, T. Cohen, B. Cooper, J. Robins, S. Ma, L. James, G. Gopalakrishna, S. Chew, C. Tan, M. Samore, F. D., and M. Murray. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–70, 2003.

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

[68] C. Martinez. *The neighborhood has its own rules: Latinos and African Americans in South Los Angeles*. NYU Press, 2016.

[69] M. Meltzer, C. Atkins, S. Santibanez, S. B. Knust, B. Petersen, E. Ervin, et al. Estimating the future number of cases in the ebola epidemic—liberia and sierra leone, 2014–2015. *MMWR Suppl.*, 63(3):1–14, 2014.

[70] S. Meyer, J. Elias, and H. M. A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68:607–616, 2012.

[71] S. Meyer and H. Leonard. Power-law models for infectious disease spread. *Annals of Applied Statistics*, 8(3):1612–1639, 2014.

[72] L. Meyers. Contact network epidemiology: bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, 44(1):63–86, 2007.

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

[74] G. Mohler, P. Brantingham, J. Carter, and M. Short. Reducing bias in estimates for the law of crime concentration. *Journal of Quantitative Criminology*, DOI: 10.1007/s10940-019-09404-1.:1–19, 2019.

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

[76] National Institute of Statistics (Guinea). General population and housing census (final results). pages Accessed August 31, 2017, 2015.

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

[78] 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.

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

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

[81] A. Papachristos. Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology*, 115(1):74–128, 2009.

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

[83] J. Park, A. Chaffee, R. Harrigan, and F. Schoenberg. A non-parametric hawkes model of the spread of ebola in west africa. *J. of Applied Statistics*, Forthcoming.

[84] J. Park, F. Schoenberg, A. Bertozzi, and P. Brantingham. Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. *JASA*, forthcoming.

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

[86] R. Peng. Applications of multi-dimensional point process methodology to wildfire hazard assessment. *UCLA PhD Dissertation*, 2003.

[87] A. Reinhart and J. Greenhouse. Self-exciting point processes with spatial covariates: modeling the dynamics of crime. *Journal of the Royal Statistical Society: Series C*, 67(5):1305–1329, 2018.

[88] R. Rosenfeld, T. 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.

[89] R. Sampson, S. Raudenbush, and F. Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277:918–924, 1997.

[90] F. Schoenberg. Testing separability in spatial-temporal marked point processes. *Biometrics*, 60(2):471–481, 2004.

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

[92] F. Schoenberg, J. Gordon, and R. Harrigan. Analytic computation of nonparametric marsan-lengline estimates for hawkes point processes. *J. of Nonparametric Statistics*, 30(3):742–775, 2018.

[93] F. 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.

[94] 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.

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

[96] S. Sheather and M. 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.

[97] M. Short, G. Mohler, P. Brantigham, and G. Tita. Gang rivalry networks via coupled point process networks. *Discrete and Continuous Dynamical Systems*, 19(5):1459–1477, 2014.

[98] Sierra Leone Statistics. Sierra leone 2015 population and housing census: Provisional results. pages Accessed August 31, 2017, 2016.

[99] C. Siettos, C. Anastassopoulou, L. Russo, C. Grigoras, and E. Mylonakis. Modeling the 2014 ebola virus epidemic - agent-based simulations, temporal analysis and future predictions for liberia and sierra leone. *PLoS Curr.*, 7, 2015.

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

[101] W. Skogan, S. M. Hartnett, N. Bump, and J. Dubois. *Evaluation of CeaseFire-Chicago.* U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, D.C., 2009.

[102] L. Smith, A. Bertozzi, P. Brantingham, G. 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.

[103] K. Soetaert, T. Petzoldt, and R. Woodrow Setzer. Solving differential equations in r: Package desolve. *Journal of Statistical Software*, 33(9):1–25, 2010.

[104] J. Spengler, E. Ervin, J. Towner, P. Rollin, and S. Nichol. Perspectives on west africa ebola virus disease outbreak. *Emerg Infect Dis.*, 22(6):956–963, 2016.

[105] 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.

[106] A. Tremblay, D. Herz, and M. Kraus. 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.*, 2019.

[107] United Nations Development Programme. West african economies feeling ripple effects of ebola, says un. pages Accessed September 14, 2017, 2015.

[108] M. Valasik, M. Barton, S. Reid, and G. Tita. Barriocide: investigating the temporal and spatial influence of neighborhood structural characteristics on gang and non-gang homicides in east los angeles. *Homicide studies*, 21(4):287–311, 2017.

[109] L. Valdez, H. Aragão Rêgo, H. Stanley, and L. Braunstein. Predicting the extinction of ebola spreading in liberia due to mitigation strategies. *Sci. Rep.*, 5(12172), 2015.

[110] C. Viboud, S. Kaiyun, R. Gaffey, M. Ajelli, and et al. The rapidd ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.

[111] D. Weisburd. The law of crime concentration and the criminology of place. *Criminology*, 53(2):133–157, 2015.

[112] WHO Ebola Response Team. Ebola virus disease in west africa – the first 9 months of the epidemic and forward projections. *N Engl J Med*, 371:1481–1495, 2014.

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

[114] L. Worden et al. Real-time projections of epidemic transmission and estimation of vaccination impact during an ebola virus disease outbreak in the eastern region of the democratic republic of congo. *BioRxiv*, (10.1101/461285), 2018.

[115] World Health Organization. Summary of probable sars cases with onset of illness from 1 november 2002 to 31 july 2003. pages Accessed September 2, 2017, 2003.

[116] World Health Organization. Ebola data and statistics. pages Accessed August 31, 2017., 2016.

[117] World Health Organization Regional Office for Africa. Predicting the extinction of ebola spreading in liberia due to mitigation strategies. *Health topics: Ebola virus disease.*, (Available at: http://www.afro.who.int/health-topics/ebola-virus-disease.), 2019.

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

[119] 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.