Title
Estimation of Spatial-Temporal Hawkes Models for Earthquake Occurrences

Permalink
https://escholarship.org/uc/item/2j86g49f

Author
Molyneux, James

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Estimation of Spatial-Temporal Hawkes Models for Earthquake Occurrences

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

by

James Molyneux

2018

ABSTRACT OF THE DISSERTATION

Estimation of Spatial-Temporal Hawkes Models for Earthquake Occurrences

by

James Molyneux
Doctor of Philosophy in Statistics
University of California, Los Angeles, 2018
Professor Frederic R. Paik Schoenberg, Chair

Point processes have long been used as an effective modeling technique in the forecasting of earthquakes. In this dissertation, we evaluate the ability of earthquake focal mechanisms to predict the locational direction of future events and the usefulness of more complicated point process models including such covariates. We also introduce a new computational method for estimating parameters of point processes models using the Stoyan-Grabarnik estimator which avoids the need to numerically compute the intractable integral term needed to compute estimates via maximum likelihood.

The dissertation of James Molyneux is approved.

David D. Jackson

Ying Nian Wu

Mark Stephen Handcock

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2018

To Jennifer Alesón-O'Farrell

lover of tea, animals, and most graciously

me

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The perspective, which I'm sure I share with countless others, about what it takes to obtain a Ph.D. has drastically changed throughout my graduate career. At first, I believed a Ph.D. to be a rather lonely pursuit in which a single mind (and perhaps their advisor) toiled away in the dark working out singular problems which would slightly further humankind's understanding of natural phenomenon. It turns out this view is exceedingly incorrect. In fact, working on a Ph.D. is a hugely collaborative effort between the student and numerous other individuals. Here I would like to acknowledge those individuals whom helped, inspired, consoled and otherwise pushed me to finish my own slight contribution to human understanding.

First and foremost I must thank my advisor Professor Rick Schoenberg, whom without his advice, guidance and watchful eye, leads me to believe I would not be the researcher I am today. Throughout our hundreds of weekly meetings, Professor Schoenberg has instilled within me the skills required to conduct cutting edge research in earthquake and point processes along with a pragmatic view in which I will surely use in my own future research. Whenever I might enter a meeting feel down about the direction of my research, Rick would see some path for which I could use to lead me to in a more fruitful direction. I could not be more thankful to be advised by a more patient, intelligent and encouraging advisor.

I am also exceedingly thankful to other faculty members for sharing with me their philosophies about teaching and conducting research in the field of statistics and seismology. I would like to especially thank the members of my committee, Professor Mark Handcock, Professor Ying Nian Wu and Professor David Jackson. Whether it was chatting with them in the hallway or at conferences, or working under them as a teaching assistant, my own view of statistics and seismology has been greatly expanded by their deep knowledge of their respective areas of study. Further, I am also thankful to members of the faculty such as Miles Chen, Linda Zanontian, Vivian Lew, Juana Sanchez, Akram Almohalwas and Nicolas Christou. Each of these faculty members have contributed to my understanding of what it means to teach statistics and also pushed me to continue my studies.

During my graduate studies, I was also a member of the *Mobilize* team which created the first approved data science math course now being taught in dozens of schools in Southern California. It would take pages upon pages to write just how much of an impact my fellow *Mobilize* members had on me, so I hope simple thank yous will do. So to Terri Johnson, Suyen Moncada-Machado, LeeAnn Trusela, Monica Casillas, Heidi Estevez, Steve Nolan, Hongsuda Tangmunarunkit and Shuhao Wu, I would like to say that I am forever grateful for your kind words and hard work with which, together, helped us create such an amazing curriculum.

I would like to especially thank Robert Gould, *Principle Investigator* for the *Mobilize* grant and one of the most inspirational people I have ever met. Through Rob's guidance, I have found a field of study in which I have found an enormous sense of enjoyment. Thank you, Rob, for constantly lending me your expert perspective on statistical education, your constant enthusiasm and your recommendations on where to find the best food in West Los Angeles. Your work has opened my eyes to the fruitful contributions I hope to someday make in the field of education and I am sincerely grateful to you for that.

While at UCLA, I benefited enormously as well from the friends I have made. Their company, patience and willingness to listen was a constant help towards keeping my own sanity. Thank you to Josh Embree, Katie McLaughlin, Patrick McCarthy and Josh Gorden for your friendship, advice and sharing many a coffee/funny text message with me. Thank you also to Amelia McNamara, Andrew Bray, Nathan Langholz, Mark Nakamura, Medha Uppala, Junhyung Park, Aaron Danielson, Levon Demirdjian and Guani Wu for keeping your doors open and making my experience at UCLA so wonderful.

The staff within the statistics department also deserve special acknowledgement. Thank you Jason Mesa and Laurie Leyden for always giving me an outlet to chat about non-statistics related things. Thank you also to Chie Ryu, Verghese Nallengara and Enrique Reyes for your constant vigilance and keeping everything running so smoothly. Perhaps the most deserving of special mention, and for whom words could never describe what they have meant to me, is Ms. Glenda Jones. Her official title reads *Student Affairs Officer* but to me (and indeed any other person who has walked the halls of the stats department) she is something entirely more. She is my guardian angel, the person who I constantly sought out help and advice from and the person with the biggest, kindest heart I have ever met. She encouraged me when I felt as if I was running out of steam, she was always there when I needed a friend and she kept her candy dish stocked with the finest processed sugars which kept me alive and running during many an afternoon. No single person, besides maybe my advisors Rick and Rob, have had a greater positive impact on me during my time at UCLA. Thank you Ms. Glenda for your constant friendship and thoughtful advice.

Finally, I would like to thank those people who kept me working towards my goal of starting and finishing a Ph.D. from outside the halls of academia. Thank you mom and dad, Beverly and Dale Molyneux, for all of the support you've given me over my many years at universities. Thank you also to my sister, Aimee Molyneux, my niece and nephew, Wyatt and Olivia, and to Brendan, Nance and Jacqueline Alesón-O'Farrell. And thank you still to Yolanda, Thomas and Ryan Brennan. But thank you most of all to my partner and love of my life Jennifer Alesón-O'Farrell. Even with all the help of every single person I've mentioned thus far, I would never have gotten to this point if it weren't for your love and encouragement. Living a future with you is the only dream I could possibly desire.

The material in Chapter 3 is a version of Molyneux, J., Gordon, J. S., and Schoenberg, F. P., "Assessing the predictive accuracy of earthquake strike angle estimates using nonparametric Hawkes processes" published in *Environmetrics*. It was supported by the National Science Foundation under grant number DMS 1513657.

Data used in Chapter 3 and 4 were provided by the Caltech/USGS Southern California Seismic Network (SCSN), doi:10.7914/SN/CI, operated by the Caltech Seismological Laboratory and USGS, which is archived at the Southern California Earthquake Data Center (SCEDC). The SCEDC and SCSN are funded through U.S. Geological Survey Grant G10AP00091, and the Southern California Earthquake Center, which is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008.

VITA

2010	B.A. (Mathematics) and B.A. (Economics), California State University, Fullerton.
2010 – 2017	Teaching Assistant/Associate/Fellow, Statistics Department, UCLA. Led discussion sections for upper and lower division statistics courses. Lecturer for Statistics 20 (Introduction to Statistical Programming with R).
2013–present	Graduate Student Researcher, Mobilize Project, Los Angeles. Wrote curriculum, computer-based labs, and software for use in the National Science Foundation *Introduction to Data Science* high school math course under the supervision of Professor Rob Gould, University of California, Los Angeles.

PUBLICATIONS AND PRESENTATIONS

McNamara, A., Molyneux, J. Teaching R to High School Students...and Teachers. UseR! Conference, Los Angeles, CA, 2014.

Gould, R., Molyneux, J. Data Science in High Schools. American Statistical Association, Orange County, California, 2015.

Molyneux, J., Schoenberg, F. P. Non-parametric Hawkes models with strike angle covariates. Southern California Earthquake Center Annual Meeting, Palm Springs, CA, 2017.

Molyneux, J., Gordon, J. S., Schoenberg, F. P. (2018). Assessing the predictive accuracy of earthquake strike angle estimates using non- parametric Hawkes processes. Environmetrics. 2018;29:e2491.

Molyneux, J., and Schoenberg, F. P. (2018, in progress). Self-exciting point process parameter estimation using the Stoyan-Grabarnik statistic.

Molyneux, J., and McNamara, A., Gould, R. (2018, in progress). *mobilizr*: An R package for high school data scientists.

CHAPTER 1

Introduction

Self-exciting point processes are a class of probability models for phenomenon whereby an occurrence of an event, represented as a point in space-time, leads to a temporary elevation of similar events occurring soon after. These models have been successfully applied to fields such as epidemiology (Njabo et al. 2016), environmental studies (Balderama et al. 2012), wildfires (Peng et al. 2005; Xu and Schoenberg 2011), criminal justice (Mohler et al. 2011; Stomakhin et al. 2011), e-mail networks (E W Fox et al. 2016), and seismology (Ogata and Zhuang 2006a; Ogata 2011; Wang et al. 2011; Console et al. 2009; Chu et al. 2011; Tiampo and Shcherbakov 2012; Helmstetter and Werner 2012). In this dissertation we aim to evaluate the potential benefits of extending a popular seismology model, the Epidemic Type Aftershock Sequence (ETAS), to include earthquake faulting covariates. We also aim to develop a new computational method for estimating parameter values of point process models in general.

The first part of this dissertation aims to assess the predictive qualities of earthquake focal mechanisms, which describe the orientation of the faulting that occurred. We also evaluate the degree to which including earthquake focal mechanisms improve model performance for non-parametric Hawkes processes. The latter part of this dissertation describes a new computational technique to compute parameter estimates for point processes using the Stoyan-Grabarnik statistic (Stoyan and Grabarnik 1991). We also describe difficulties this new computational method brings to light and areas where future work is still needed. Throughout, we use actual and simulated earthquake catalogs and so we describe the scientific context of seismology below.

1.1 Scientific context: Seismology

Earthquakes are potentially devastating events which occur when stresses, caused by the movement of Earth's tectonic plates, are relieved by fracturing within the Earth's crust. The need to accurately forecast when and where these events are most likely to occur represents a fundamental issue for the field of seismology (Bolt 2003). Forecasts are needed, not just for short-term allocation of emergency resources, but also for long term insurance rates, building codes, and earthquake preparedness (Jordan and Jones 2010). To this end, seismologists have developed earthquake forecasting models such as ETAS (Ogata 1998) and Uniform California Earthquake Rupture Forecast (UCERF3) (Edward H Field et al. 2015) among others as well as testing centers such as the Regional Earthquake Likelihood Models (RELM) (Edward H. Field 2007) and the Collaboratory for the Study of Earthquake Predictability (CSEP) (J Douglas Zechar et al. 2009) to measure and compare the effectiveness of different forecasting models.

CHAPTER 2

Background

In this chapter, we lay out the fundamental frameworks for point processes, the estimation of point process models and describe the basic intuition for earthquake focal mechanisms.

2.1 Point processes

A point process (Daley and Vere-Jones 2007; Daley and Vere-Jones 2008) is a collection of points $\{\tau_1, \tau_2, \ldots\}$ which occur in some metric space. A spatial-temporal point process is a stochastic process which generates a countable set of points $\{(\mathbf{s}_i, t_i) : i = 1, 2, \ldots, N\}$ in $\mathbb{R}^2 \times \mathbb{R}$. For spatial-temporal point processes, we often consider the counting process, $N(B) \in \{0\} \cup \mathbb{N}$, where $N(B)$ is a random measure on a complete metric space $\mathcal{S}$ which describes the number of points falling within some Lebesgue measurable compact set B.

Spatial-temporal point processes are typically modeled via their conditional intensity functions where, given the history of the process, $\mathcal{H}$, up to time t,

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

The conditional intensity function thus describes the infinitesimal expected rate by which points occur in space and time $(\mathbf{s}, t)$ given the history of the process. Additional covariates for points, called marks, can also be incorporated into the conditional intensity function. In seismological contexts, a mark could be the magnitude of an earthquake or its fault plane geometry. For simple, i.e. $(\mathbf{s}_i, t_i) \neq (\mathbf{s}_j, t_j)$ for all $i \neq j$, finite-dimensional point processes, the conditional intensity function uniquely identifies the point process (Daley and Vere-Jones 2007).

2.2 Self-exciting point processes

Self-exciting point processes, or Hawkes processes (Hawkes 1971), are a class of branching point processes in which there is a positive, or for self-correcting point processes a negative, correlation between points. In the purely temporal context, the conditional intensity function of a self-exciting point process is denoted

$$\lambda(t|\mathcal{H}_t) = \mu(t) + K \sum_{i:t_i<t} g(t - t_i)$$

where $\mu(t)$ describes the background rate of the process as a function of time and $g(t - t_i)$, called the triggering function, describes the temporary excitation of additional points attributed to those which occurred previously. The productivity of the triggering function is described by K. Figure 2.1 shows an example of a purely temporal Hawkes process. In the figure, it is important to note the many discontinuities which, as we explain in Section 2.2.2 below, explains why computing integrals of a Hawkes process is so difficult.

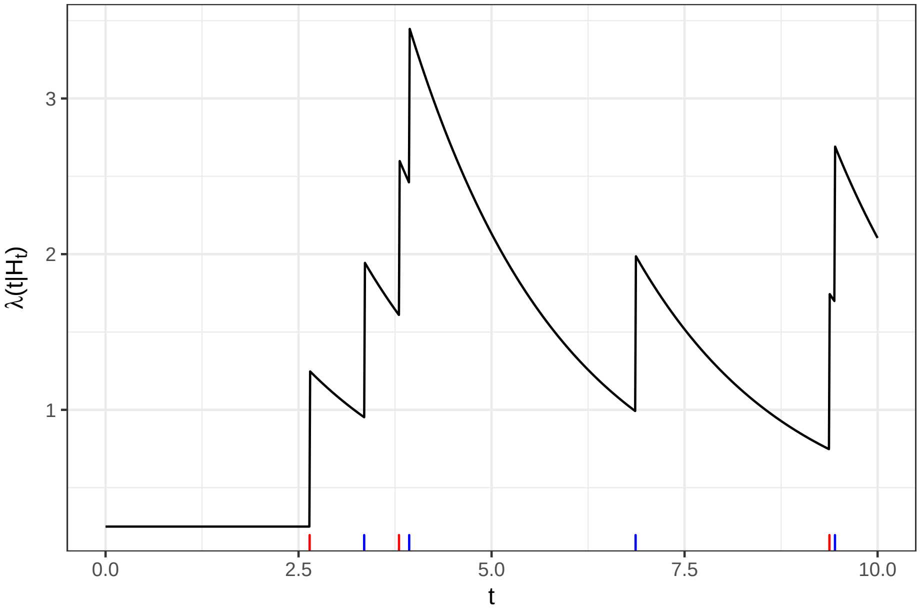

Figure 2.1: An example of a temporal Hawkes model with parameterization $\lambda(t) = \mu + K \sum_{i:t_i<t} e^{-\alpha}(t - t_i)$ and parameter values ($\mu = 0.25, K = 1, \alpha = 0.5$). Red tick marks along the x-axis indicate background events and blue tick marks indicate excited events.

For marked self-exciting point processes, such as those used in seismology, the conditional intensity function is written

$$\lambda(\mathbf{s}, t|\mathcal{H}_t) = \mu(\mathbf{s}, t) + K \sum_{i:t_i<t} g(\mathbf{s} - \mathbf{s}_i, t - t_i; m_i)$$

where $\mu(\mathsf{s}, t)$ describes the background rate as a function of location and time and m_i represents the mark of point i.

2.2.1 Epidemic type aftershock sequences

One of the most important models of earthquake occurrences in seismology is the ETAS model. First proposed by Ogata (1988) as a means to explain the temporal occurrence of earthquakes and their aftershocks, it was later expanded in Ogata (1998) to include both temporal and spatial aspects of earthquakes. Many different parameterizations of the model have been attempted, and indeed many variants of the model exist.

In general, these models can be expressed as

$$\lambda(t, x, y|\mathcal{H}_t) = \mu(t, x, y) + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i; m_i)$$

where x, y denote the spatial locations and m the magnitudes of the events within an earthquake catalog. For this dissertation, we use the common parameterization of the triggering function

$$g(t - t_i, x - x_i, y - y_i; m_i) = \frac{K}{(t - t_i + c)^p} \cdot \frac{e^{\alpha(m - M_c)}}{((x - x_i)^2 + (y - y_i)^2 + d)^q}$$

where M_c denotes the fixed magnitude cut-off for the earthquake catalog, $Ke^{\alpha(m-M_c)}$ describes the magnitude distribution of events, $(t - t_i + c)^{-p}$ describes the temporal distribution of events, and $((x - x_i)^2 + (y - y_i)^2 + d)^{-q}$ describes the spatial distribution of events. The parameterization for the magnitude and temporal distributions are based on the the Gutenberg-Ricter law (Gutenberg and Richter 1944) and modified Omori formula (Utsu 1961), respectively. This triggering function can also be equivalently expressed as a density, the benefits of doing so are addressed in Chapter 4,

$$g(t - t_i, x - x_i, y - y_i; m_i) = \frac{K(p - 1)c^{p-1}(q - 1)d^{q-1}}{\pi(t - t_i + c)^p} \cdot \frac{e^{\alpha(m - M_c)}}{((x - x_i)^2 + (y - y_i)^2 + d)^q}$$

For both parameterizations, the parameters in which we are interested in estimating are $(\mu, K, c, p, \alpha, d, q)$.

2.2.2 Parameter estimation via maximum likelihood

For a given point process model, the method most often used to estimate the parameters of such a model is by maximizing the log-likelihood

$$\ell(\theta) = \sum_{i=1}^{N} \log(\lambda(\mathbf{s}_i, t_i|\mathcal{H}_t)) - \int_{\mathcal{S}} \int_0^T \lambda(\mathbf{s}, t|\mathcal{H}_t)\ dtd\mathbf{s}$$

where $\mathcal{S}$ denotes the spatial region of interest from time 0 to T. While the summation term of the log-likelihood is easily computed, the integral term represents a significant challenge as analytical solutions do not exist and thus must be computed numerically (Harte 2012). The challenge of computing the intractable integral term represents a limiting feature in applications requiring either small to moderately sized data sets or access to large computational servers. Though difficult to compute, under a set of general conditions, maximizing the log-likelihood is consistent, unbiased, efficient, and asymptotically normal (Ogata 1978). Furthermore, the maximum likelihood estimates (MLE) for point processes have known standard errors which can be computed from the negative of the diagonal elements of the Hessian matrix (Rathbun and Cressie 2016).

2.3 Earthquake focal mechanisms

For earthquake that occur in an area with sufficient coverage of seismometers, the waveform data produced can be used to estimate a focal mechanism which describes the deformation that occurred at the earthquake source. By examining the direction of first-motions of seismometers, seismologists estimate the double-couple of the earthquake which represents the deformation as two nodal planes, one of which represents the fault along which the rupture occurred while the other is known as the auxiliary plane and is orthogonal to the fault plane (Aki and Richards 2002). Seismologists often represent focal mechanisms using beach ball diagrams, which show the estimated nodal planes as well as the vectors of tension and pressure as shaded and unshaded regions, respectively. The orientation of the planes along with the tension and pressure vectors allow seismologists to infer the type of faulting which occurred during an earthquake. Figure 2.2 shows an example of a beach ball diagram for the *Easter Sunday* earthquake which occurred in Baja California, Mexico in 2010.

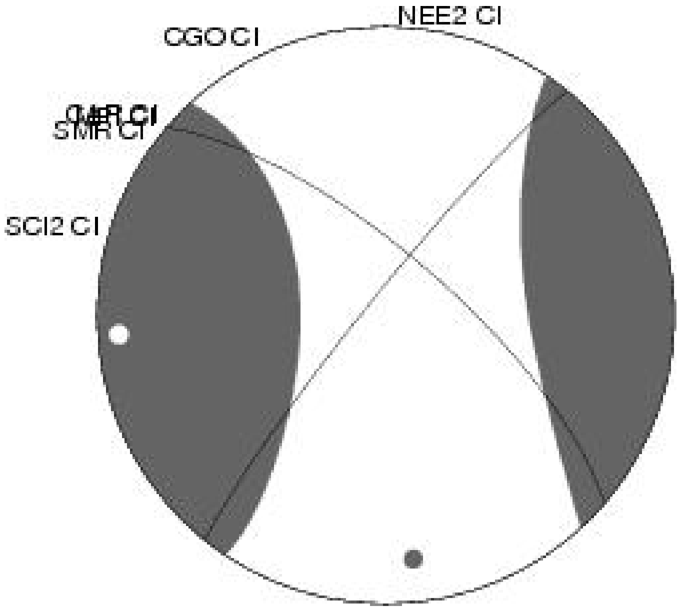

Figure 2.2: Beach ball diagram for the 7.2 magnitude *Easter Sunday* earthquake which occurred on April 4th, 2010 in Baja California, Mexico. Shaded regions indicate areas of tension and unshaded regions indicate areas of pressure. The intersecting planes represent the estimated nodal planes. Text along the circle indicate the direction of the seismometers used to estimate the focal mechanism. Image courtesy of the Southern California Earthquake Center.

The hope among seismologists is that focal mechanism information will have high predictive value for earthquake forecasting, particularly with respect to information about the orientation of the fault plane on which triggered events primarily occur (Kagan 2014). Aftershocks have indeed been observed to occur more densely at the ends and along fault ruptures and less densely in areas orthogonal to the rupture, in agreement with theoretical studies of Coulomb stress (Das and Henry 2003). Henry and Das (2001) report in particular a tendency for aftershocks to occur along the estimated strike of strike-slip mainshocks, again in agreement with geophysical theory of Coulomb stress. For example, Strader and Jackson (2014) found that resolving Coulomb stress onto the more favorable nodal plane of a receiver earthquake resulted in an increased number of earthquakes occurring in areas of high-stress. Thus, estimated focal mechanisms have been used to project the changes in stress in order to highlight areas of increased anticipated aftershock activity following earthquakes. For additional information about Coulomb stress, its derivation and its properties, see Scholz (2002).

Unfortunately, focal mechanism estimates are known to suffer from large uncertainties (Kagan 2003). In particular, large errors are associated with estimates of the orientation of nodal planes which describe the deformation that occurs at an earthquake's source (Kagan 2014). In addition, determining which nodal plane is the fault plane and which is the auxiliary plane is not possible from the estimation procedure and relies on the use of other geological data (Aki and Richards 2002).

CHAPTER 3

Assessing the Accuracy of Focal Mechanism Estimates for Non-Parametric Hawkes Processes

3.1 Introduction

As described in Chapter 2, when an earthquake occurs in an area with sufficient coverage of seismometers, the waveform data produced are used to estimate a focal mechanism which describes the deformation that occurred at the earthquake source. Such earthquake focal mechanism estimates have become increasingly common recently with the deployment of broadband seismometers and more powerful computers (Clinton et al. 2006). Also as described in Chapter 2, focal mechanisms are known to contain considerable errors in their estimation, potentially limiting their usefulness for earthquake forecasting.

Figure 3.1 illustrates how large the errors can be in the strike angle estimates obtained from estimated nodal planes. As shown in Figure 3.1, there can be considerable discrepancy between the two nodal plane estimates and the direction of subsequent seismicity. This disagreement cannot reasonably be explained as resulting from location errors, since the local seismicity itself is nearly linear in this example and in many others. Figure 3.1 also highlights the ambiguity that can exist in determining which nodal plane is the fault plane and which is auxiliary.

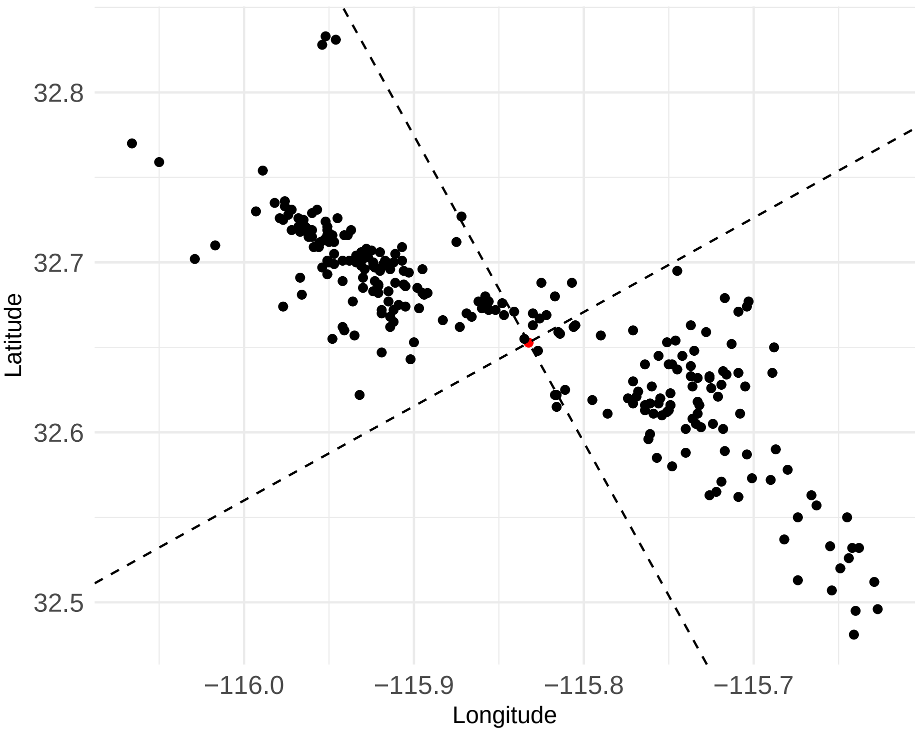

Figure 3.1: Strike angle estimates (dashed lines) based on estimated nodal planes for a 5.12 Ml mainshock (red dot) in Southern California, with an estimated depth of 10.1 km, occurring on May 5, 2010. Black points indicate earthquake with epicenters in the range $[-116.066, -115.627]$ in longitude, $[32.481, 32.833]$ in latitude, with magnitude at least 2.8, and occurring between 0.3 and 2137.7 days after the mainshock.

The objective of this chapter is to describe and quantify the degree to which such estimated strike angles increase our ability to forecast future seismicity, from a statistical perspective. One way to assess the benefit of these measurements would be to compare the performance of earthquake forecasts made with focal mechanism estimates to those made without the use of focal mechanism estimates. Unfortunately, however, such estimates remain presently unused in almost all models for forecasting seismicity, such as ETAS and other models used in forecasting studies such as the Regional Earthquake Likelihood Model (RELM) working group and the Collaboratory for the Study of Earthquake Predictability (CSEP) (Schorlemmer et al. 2010; J D Zechar et al. 2013). An exception is Kagan and

Jackson (2014), which used previous focal mechanisms to forecast future focal mechanisms as well as future seismicity, and Kagan et al. (2007) used focal mechanisms in their five year forecast to orient anisotropic smoothing, though such inclusion resulted in negligible improvement in forecasting (Wang et al. 2011). Ogata (1998) found that using anisotropic smoothing, based on the general direction of faulting in Japan, led to improved forecasts for Epidemic Type Aftershock Sequence (ETAS) models, and Ogata and Zhuang (2006b) and Ogata (2011) report improved fit from ETAS models with spatially varying parameters, but such parameters are typically not estimated presently using focal mechanism estimates. Because many competing models have been offered, with widely varying implications in terms of seismic hazard (Schoenberg and Patel 2012) and there remains considerable disagreement about which model is optimal, we instead examine the impact of estimated strike angles non-parametrically, using data-driven methods and keeping modeling assumptions about particular functional forms to a minimum.

Recent articles by Guo et al. (2015) and Guo et al. (2017) investigated how fault geometry estimated using previous seismicity for large ($M \geq 7.5$) earthquakes in China and Japan impacted parameter estimates of Epidemic Type Aftershock Sequence (ETAS) models (Ogata 1998). However, these studies did not quantify the extent to which including fault geometry in the models improved earthquake forecasts. In this chapter, we compare the degree to which estimated strike angles forecast the direction of future seismicity around a given earthquake to that of uniformly distributed angles or angles estimated based on seismicity occurring before the mainshock. Also, using the estimation method of Marsan and Lengliné (2008b), we consider purely non-parametrically estimated Hawkes models using the estimated strike angle that best fits the post-mainshock set of events for each mainshock and compare their predictive performance to corresponding models that exclude these estimates.

3.2 Data

Data on estimated origin times, locations, and magnitudes of 17,734 shallow (depth $\leq$ 75 km) earthquakes of magnitude $M \geq 2.8$ in Southern California from 1980 to 2016 were recorded by the Southern California Seismic Network (SCSN) and obtained from the Southern California Earthquake Data Center (SCEDC) (*California Institute of Technology (Caltech)* 1926; *SCEDC* 2013). Most of these events contain no focal mechanism estimates. Each earthquake in the catalog was assigned a location quality grade and we restrict our attention exclusively to those earthquakes considered to have a location quality of C or better. SCEDC also compiles the focal mechanism estimates for a subset of 899 of these events dating from 1999 to 2016, with variance reduction ≥ 40, variance reduction being a measurement of focal mechanism estimation quality.

In what follows, to distinguish these 899 events with focal mechanism estimates we refer to them as *m*ainshocks, though the branching structure of the earthquakes is unknown, and many of these 899 events may actually be aftershocks of other events. Each nodal plane is described by a set of three angles. The azimuthal strike angle is the angle created by intersecting the nodal plane with a horizontal surface, such as the surface of the Earth, and are measured as counter-clockwise degrees from North. Dip angles describe the downward angle of the plane from the horizontal surface of the Earth. A dip angle of $0°$ would lie parallel to the surface of the Earth whereas a dip angle of $90°$ would be perpendicular to the surface. The rake angle describes the direction of planar movement relative to the horizontal surface, i.e. the movement of the fault in the direction of the strike angle. For each mainshock, we examine the dip angle of the nodal plane whose strike angle best fits the post mainshock set of events (using fitting criteria described in Section 3.1) and exclude events with dip angles $\leq 75°$. This leaves us with nodal planes that are near-vertical as in the case of strike-slip faulting. Additionally, we exclude mainshocks that had fewer than three earthquakes occur prior to the mainshock and within a distance of $10^{-0.5(M_i - M_c)}$, which is of standard use in seismology as a maximum distance of aftershocks around small-to-medium sized mainshocks for catalogs with magnitude cutoff M_c (Scholz 2002). For this subset of 333 strike-slip mainshocks, the strike angle of the best fitting nodal plane should closely resemble the fault plane and have enough previous seismicity to estimate additional strike angles statistically.

The catalog of 17,734 earthquakes without estimated focal mechanisms is used in our analysis for strike angle estimation and evaluation. That is, we compare the estimated strike angles for the 333 strike-slip mainshocks to strike angle estimates obtained using the other observed events that occurred before each mainshock, and evaluate the overall fit of our different strike angle estimates on events that occurred after each mainshock. To ease the computational burden of the estimation procedure for the non-parametric Hawkes model described in Section 3.2, we raise the lower magnitude cutoff for both catalogs used in the estimation to $M > 3.25$ leaving us with 5,649 earthquakes without focal mechanisms and 330 mainshocks.

3.3 Methods

3.3.1 Fault angle estimation and evaluation

To help assess how well the recorded strike angle estimates forecast the direction of future seismicity, we consider three alternative estimates of strike angle, based on prior seismicity rather than recorded double-couple estimates. First, let $\hat{\theta}_i$, $i = 1, ..., 330$, denote the strike angle estimates based on the estimated nodal planes obtained from SCEDC. Given mainshock i at location (x_i, y_i) with magnitude M_i, consider the set S_i of events that occurred *prior* to the mainshock within the mainshock's aftershock radius, which is taken as

$$d(M_i) = 10^{-0.5(M_i - M_c)},$$

following Scholz (2002).

Consider an alternative strike angle estimate, θ_i^D, computed using Deming regression (Deming 1943), so that θ_i^D is the slope of the line minimizing the sum of squared orthogonal distances to the events in S_i. That is, assuming an equal amount of error in the x and y directions, we compute the slope of the line as

$$\hat{\beta}_1^* = \frac{s_y^2 - s_x^2 + \sqrt{(s_y^2 - s_x^2)^2 + 4 \cdot \text{cov}^2(x, y)}}{2 \cdot \text{cov}(x, y)}$$

with intercept

$$\hat{\beta}_0^* = \bar{y} - \hat{\beta}_1^* \bar{x}$$

where x, y are sets of points in S_i, $\bar{x}$, $\bar{y}$ are the mean values of x and y, and s_x^2, s_y^2 are the sample variances for x and y, respectively. Another possible strike angle estimate, θ_i^a, may be defined as the slope of the line through (x_i, y_i) minimizing the mean absolute error (MAE) where errors are defined as the angles formed by the θ_i^a and the rays from (x_i, y_i) to each preceding earthquake in S_i. In what follows we also consider the estimate θ_i^u obtained simply by choosing an angle between $0 - 360°$ uniformly at random.

To evaluate the strike angle estimates $\hat{\theta}_i, \theta_i^D, \theta_i^a$, and θ_i^u, we consider two different fitting criteria. One is the root mean squared error (RMSE), where errors are defined as the orthogonal distance between the line whose slope is derived from the strike angle estimates and the events occurring *after* mainshock i within a distance of $d(M_i)$. That is, for points, (x_j, y_j), $j = 1, \ldots, n$, occurring after a mainshock and within the aftershock radius, let (x_j^*, y_j^*) denote the points on the line which form orthogonal angles between the line and points (x_j, y_j). Then

$$\text{RMSE} = \sum_j \sqrt{(x_j - x_j^*)^2 + (y_j - y_j^*)^2}.$$

We also consider the mean absolute value of the angles formed by the line whose slope is derived from the corresponding strike angle estimate and the same set of events occurring after each mainshock. This allows us to compare how well each strike angle estimate forecast the direction of seismicity that occurred after each mainshock. Estimates with smaller RMSE or smaller angular MAE would have better forecast the direction of future seismicity.

Since the two estimated nodal planes of a double-couple are indistinguishable, we calculate the RMSE and minimum angular MAE for each nodal plane strike angle, and for each fitting criteria, let the best fitting nodal plane strike angle $\tilde{\theta}_i$ denote the nodal plane strike angle that minimizes the RMSE or angular MAE. Similarly, since the focal mechanisms are essentially given two chances to fit a strike angle to the events occurring after a mainshock, one may also consider a second, orthogonal angle for each of our strike angle estimates based on previous seismicity, as well as the estimate chosen uniformly at random, and label the better fitting of the two angles $\tilde{\theta}_i^D$, $\tilde{\theta}_i^a$, or $\tilde{\theta}_i^u$, correspondingly. This allows us to measure how the selection of the better fitting angle among a strike angle estimate and its complement affects the overall performance in forecasting the direction of future seismicity.

For each of the $n = 330$ mainshocks and for each strike angle estimation method described above, we compute the mean and standard error (SE) of the RMSE and angular MAE for the post mainshock set of events, where the SE is calculated as the standard deviation of the RMSEs divided by $\sqrt{n}$. In order to compare the estimates to the angle best fitting the seismicity following each mainshock, we also compute retrospective or *gold* angle estimates, θ_i^g, defined for each mainshock and each fitting criterion as the angle with minimal RMSE or minimal mean absolute angle from the post mainshock set of events.

3.3.2 Non-parametric Hawkes model

In order to compare the predictive performance of point process models with strike angle estimates to those without strike angle estimates, we fit Hawkes point process models to the data non-parametrically, using the method of Marsan and Lengliné (2008b). We briefly review some details of this procedure and some point process preliminaries here.

A point process is a collection of points $\{\tau_1, \tau_2, ...\}$ occurring in some metric space (Daley and Vere-Jones 2007; Daley and Vere-Jones 2008). A point process in space-time is typically modeled via its conditional intensity function, $\lambda(s, t)$, which is a stochastic process representing the infinitesimal rate at which points are expected to accumulate around location s and time t, given all points occurring prior to time t.

The Hawkes, or self-exciting, point process model (Hawkes 1971) is a type of branching point process which models the conditional intensity of a process given its history, $\mathcal{H}_t$, up to time t as

$$\lambda(s, t|\mathcal{H}_t) = \mu(s, t) + \sum_{i:t_i<t} g(s - s_i, t - t_i) \tag{3.1}$$

where $\mu(s, t)$ is the background rate of events occurring and the triggering function g describes the spatial-temporal rate at which points induce subsequent points The model extends easily to the marked case where the triggering function g depends on earthquake magnitude as well, and such marked versions are commonly used in earthquake forecasting in both parametric (Ogata 1988; Ogata 1998) and non-parametric (Marsan and Lengliné 2008b;

Eric Warren Fox et al. 2016) forms.

In order to adapt the model to account for earthquake strike angles, we follow Gordon and Schoenberg (2017) and adopt the following product form for the conditional intensity function:

$$\lambda(x, y, t|\mathcal{H}_t) = \mu(x, y) + \sum_{i:t_i<t} g_t(t - t_i)g_{xy}(x - x_i, y - y_i; m_i, \tilde{\theta}_i)g_m(m_i) \tag{3.2}$$

where $\mu > 0$ is the background rate, g_t and g_{xy} are densities governing the temporal and spatial triggering, respectively, and g_m dictates how the productivity of aftershocks depends on the mainshock magnitude. The functions $g_t(t - t_i)$, $g_{xy}(x - x_i, y - y_i; m_i, \tilde{\theta}_i)$, and $g_m(m_i)$ are each computed using the method of Marsan and Lengliné (2008a) whereby the estimation of a Hawkes model is computed by maximizing the expectation of the complete data log-likelihood and assigning the probability of each earthquake being a background event or an aftershock of a previous event. Space-time-magnitudes were shown by Zhang (2017) to be approximately separable hence we assume separability in our model. Here t_i is the time of the ith event, (x_i, y_i) is the two-dimension epicentral location of the ith event, m_i is the magnitude of the ith event and $\tilde{\theta}_i$ is the best fitting strike angle from the ith double-couple.

For the spatial triggering, we take

$$g_{xy}(x - x_i, y - y_i; m_i, \tilde{\theta}_i) = g_r(r; m_i)g_\phi(\phi; \tilde{\theta}_i) \tag{3.3}$$

where g_r accounts for triggering as a function of distance from earthquake i and g_ϕ accounts for the angular locations of triggering relative to strike angle $\tilde{\theta}_i$. We then compare the fit of (3.2) to the corresponding model without strike angle estimates, i.e.

$$\lambda(x, y, t|\mathcal{H}_t) = \mu(x, y) + \sum_{i:t_i<t} g_t(t - t_i)g_{xy}(x - x_i, y - y_i; m_i)g_m(m_i). \tag{3.4}$$

For both model (3.2) and (3.4), the functions μ, g_t, g_{xy}, and g_m are estimated entirely non-parametrically using the E-M type method as described in Marsan and Lengliné (2008b).

3.3.3 Voronoi deviance residuals

To compare the fit of models (3.2) and (3.4), one may inspect a residual process aggregated over a (typically rectangular) grid of pixels as in Baddeley et al. (2005). Specifically, we examine deviance residuals (Clements et al. 2011) over such a rectangular grid. We also compute Voronoi deviance residuals as in Bray et al. (2014). Such residuals are helpful in evaluating the goodness of fit of point process models with highly volatile conditional intensities, and Voronoi residual deviances have the advantage of being less skewed in their distribution when compared to grid based residuals (Bray et al. 2014).

To construct Voronoi residuals, we partition the space into Voronoi cells, where each cell B_i is defined as the set of locations closer to the observed point τ_i than to any of the other observed points. Each Voronoi cell, B_i, then contains only the single point τ_i by construction.

For each bin B_i in a rectangular grid, or for each cell B_i in the Voronoi tessellation of the observed points, we compute the deviance, or ratio of two corresponding log-likelihoods, to evaluate the relative fit of two models. For two different intensity estimates, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, the deviance residual computed over cell B_i is

$$\begin{aligned} R(B_i) &= \sum_{i:(t_i,x_i,y_i)\in B_i} \log\left(\hat{\lambda}_1(t_i,x_i,y_i)\right) - \int_{B_i} \hat{\lambda}_1(t,x,y)\ dtdxdy \\ &- \left(\sum_{i:(t_i,x_i,y_i)\in B_i} \log\left(\hat{\lambda}_2(t_i,x_i,y_i)\right) - \int_{B_i} \hat{\lambda}_2(t,x,y)\ dtdxdy \right). \end{aligned}$$

A positive deviance residual implies that model $\hat{\lambda}_1$ provides a better fit to the data in the given region while negative deviance residuals imply the opposite. Voronoi deviance residuals are sometimes scaled by the dividing the log-likelihood ratio for each Voronoi cell by its area. We use unscaled residuals so the total deviance for the competing models is simply the sum of the binned deviances, $\sum_i R(B_i)$. Total deviance values close to zero indicate minimal difference in fit between the two competing models.

3.3.4 Stoyan-Grabarnik diagnostic

The temporal fit of (3.2) is evaluated using the Stoyan-Grabarnik diagnostic (Stoyan and Grabarnik 1991)

$$\mathrm{E}\left(\sum_{x_i\in N\cap B} \frac{1}{\lambda(x_i)} \right) = |B|$$

where N is the point process, B is a spatio-temporal window and $\lambda(x_i)$ is the conditional intensity of point x_i. Dividing the entire temporal window into n temporal bins, B_j with $j = 1, \ldots, n$, if the estimated intensity $\hat{\lambda}$ is accurate then sum of the estimated inverse intensities for points in bin B_i should approximate the volume of B_j with variance

$$\mathrm{Var}\left(\sum_{x_i \in N \cap B_j} \frac{1}{\hat{\lambda}(x_i)}\right) = \int_S \int_T \frac{1}{\hat{\lambda}(t,x,y)} dt dx dy + \sum_{x_i \in N \cap B_j} \frac{1}{\hat{\lambda}(x_i)} - |B_j|^2,$$

where S and T are the spatial and temporal windows, respectively. Baddeley et al. (2008) notes that it is possible for $\hat{\lambda}(x_i)$ to be zero and for the diagnostic to have a large variance. Following their suggestions, we set $1/\hat{\lambda}(x_i) = 0$ for values of $\hat{\lambda}(x_i)$ that are zero or close to zero.

3.4 Results

3.4.1 RMSE of orthogonal distances

The RMSEs for each of the strike angle estimates described in Section 3.1 are summarized in Figure 3.2 panel (a). The average RMSE for the estimate θ^D obtained using Deming regression on prior seismicity was 2.84 km, just slightly above the mean RMSE of 2.81 km corresponding to $\tilde{\theta}$, the best fitting of the two possible strike angles obtained using the estimated nodal plane. Both θ^D and $\tilde{\theta}$ fit subsequent seismicity significantly better than the random uniformly distributed angle θ^u, which had a mean RMSE of 3.74 km, and both also fit better than the angle θ^a minimizing the mean absolute angle to previous seismicity, which had a mean RMSE of 3.16 km. Each estimate fit the aftershock seismicity significantly worse than the retrospective best fitting angle θ^g, as the mean RMSE for θ^g was just 1.72 km.

Much of the apparent success of the estimated strike angle $\tilde{\theta}$ in forecasting future seismicity is attributable to the fact that $\tilde{\theta}$ is defined as the better fitting of the estimated strike angle and its complement. Indeed, when each alternative estimator of the strike angle is replaced with the better fitting of the estimate and its complement, the resulting RMSEs decrease markedly, as shown in Figure 3.2 panel (b). The estimates $\tilde{\theta}^D$ and $\tilde{\theta}^a$ have mean RMSEs of 2.60 km and 2.64 km, respectively, from subsequent seismicity. Both of these strike angle estimates, which are obtained using only prior seismicity and not using focal mechanism estimates, fit subsequent seismicity better than the nodal plane estimate $\tilde{\theta}$, which had a mean RMSE of 2.81 km. The RMSEs for $\tilde{\theta}$ are not significantly lower than those of the uniformly distributed angles $\tilde{\theta}^u$. Even after accounting for the the orthogonal complements, each strike angle estimate in Figure 3.2 panel (b) had a significantly larger RMSE than the retrospective best-fitting angle θ^g.

3.4.2 Mean absolute angles

The results are similar when considering angular MAE as the fitting criterion. Figure 3.2 panel (c) summarizes the angular MAE for each strike angle estimate, with an error defined as the angular distance between an aftershock and the estimated strike angle. The average angular MAE for the best fitting nodal plane estimate $\tilde{\theta}$ (35.55°) is significantly lower than the average MAEs of θ^D (39.24°), θ^a (39.60°), and θ^u (46.15°). However, the advantage of the nodal plane estimate appears to be entirely due to the selection between the two orthogonal nodal plane estimates, as shown in Figure 3.2 panel (d). Comparing the average MAEs of the best fitting strike angles for each method, the estimates based on prior seismicity (34.56° for $\tilde{\theta}^D$ and 34.37° for $\tilde{\theta}^a$) fit aftershock activity slightly better than $\tilde{\theta}$ (35.55°). In fact, using MAE as a metric, the fit of the nodal plane estimate $\tilde{\theta}$ was indistinguishable from that of the uniformly distributed angle $\tilde{\theta}^u$ (35.43°). All four of the prospective strike angle estimates fit significantly worse than the retrospective optimum θ^g (29.49°).

3.4.3 Non-parametric Hawkes Process

Figure 3.3 shows the estimated contribution of the strike angles to the triggering function of a non-parametric Hawkes model. The y axis shows the estimated triggering density, $\hat{g}_\phi$, as a function of the angle the location in question makes with a mainshock, relative to the estimated nodal plane strike angle $\tilde{\theta}$ of that mainshock. The LOESS smoothing of the non-parametric triggering density estimates in Figure 3.3 shows a downward trend indicating

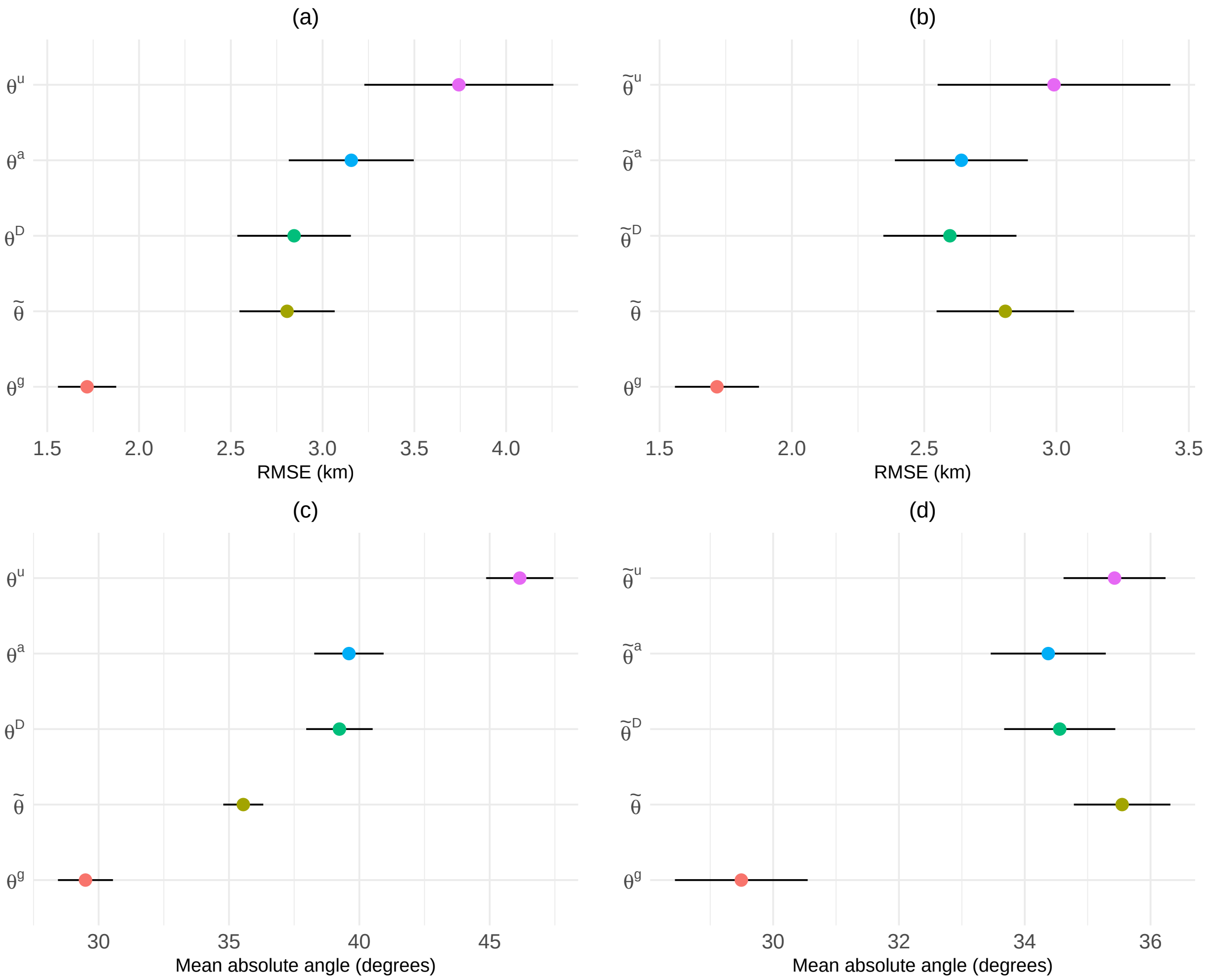

Figure 3.2: Mean ± 1.96 standard errors of the RMSEs, panels (a) and (b), and the mean absolute angle of aftershocks relative to the mainshock strike angle estimate, panels (c) and (d), of aftershocks for the $n = 330$ mainshocks. Here $\tilde{\theta}$ is the best fitting strike angle estimate $\tilde{\theta}$, θ^u, θ^D, and θ^a denote the raw estimates, and θ^g represents the angle of minimal RMSE to the subsequent seismicity. For panels, (b) and (d), each strike angle estimate $\tilde{\theta}^{\cdot}$ is the best-fitting of $\theta^{\cdot}$ and its orthogonal complement, and the selection is performed for each mainshock by minimizing the RMSE or the angular MAE of its aftershocks.

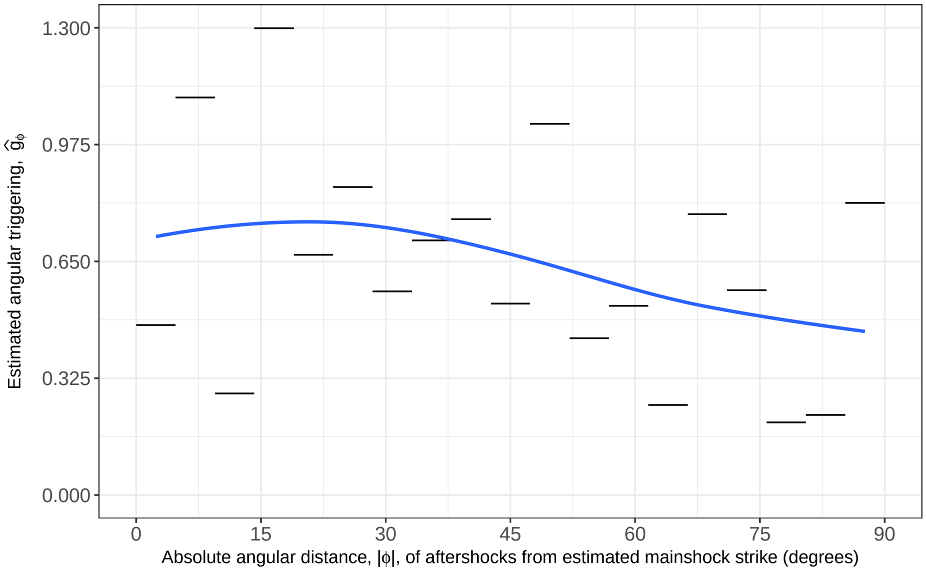

Figure 3.3: Non-parametrically estimated triggering density $\hat{g}_\phi$ for model (3.2), estimated using the algorithm of Marsan and Lengliné (2008b). Horizontal line segments indicate, for each subset of angles, the estimated triggering intensity of aftershocks occurring within this subset of angular distances relative to the estimated strike angle derived from the estimated nodal plane. The segments are smoothed by LOESS (blue curve).

fewer events were generally triggered at larger angles relative to the mainshock strike angle.

Figure 3.4 compares the conditional intensity over 0.05° × 0.05° grid cells across the southern California region for the null model (3.4) assuming isotropic triggering to that of model (3.2) incorporating the strike angle estimate $\tilde{\theta}$. Both fitted models agree closely with one another, though model (3.2) tends to be slightly more diffuse in low seismicity regions.

To compare the areas where model (3.2) improves the fit of the data over the null model (3.4), the deviance residuals are shown in Figure 3.5 and the unscaled Voronoi deviance residuals, where each Voronoi cell encloses each of the 330 mainshocks, in Figure 3.6. Red areas indicate cells where the inclusion of $\tilde{\theta}$ in model (3.2) led to an improvement in model fit when compared to the null model (3.4), with the total deviance of Figure 3.5 being 615.86 in favor of model (3.2). This corresponds to an overall information gain (Daley and Vere-Jones (2007) and Daley and Vere-Jones (2008)) of 0.103 per earthquake. Figures 3.5 and 3.6

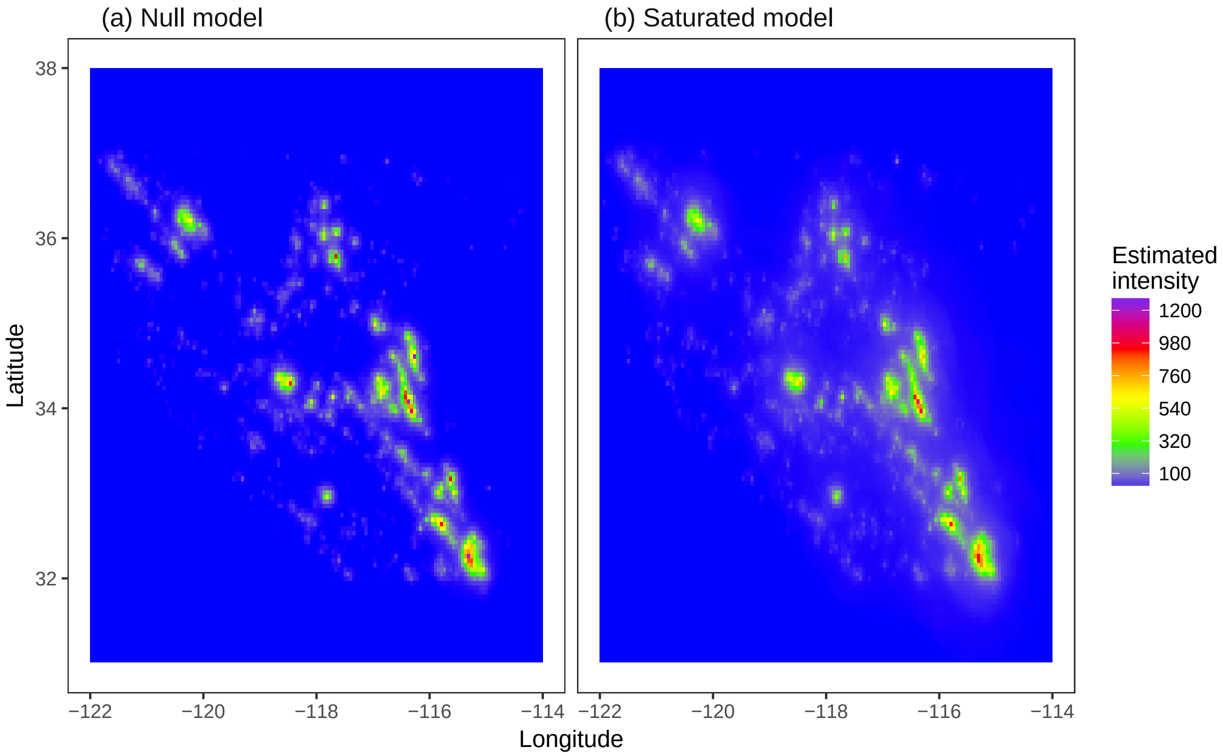

Figure 3.4: Estimated intensity of fitted non-parametric Hawkes models. (a) the estimated intensity of null model (3.2) assuming isotropic triggering. (b) the estimated intensity for model (3.4) which includes $\tilde{\theta}$ in the triggering function estimate.

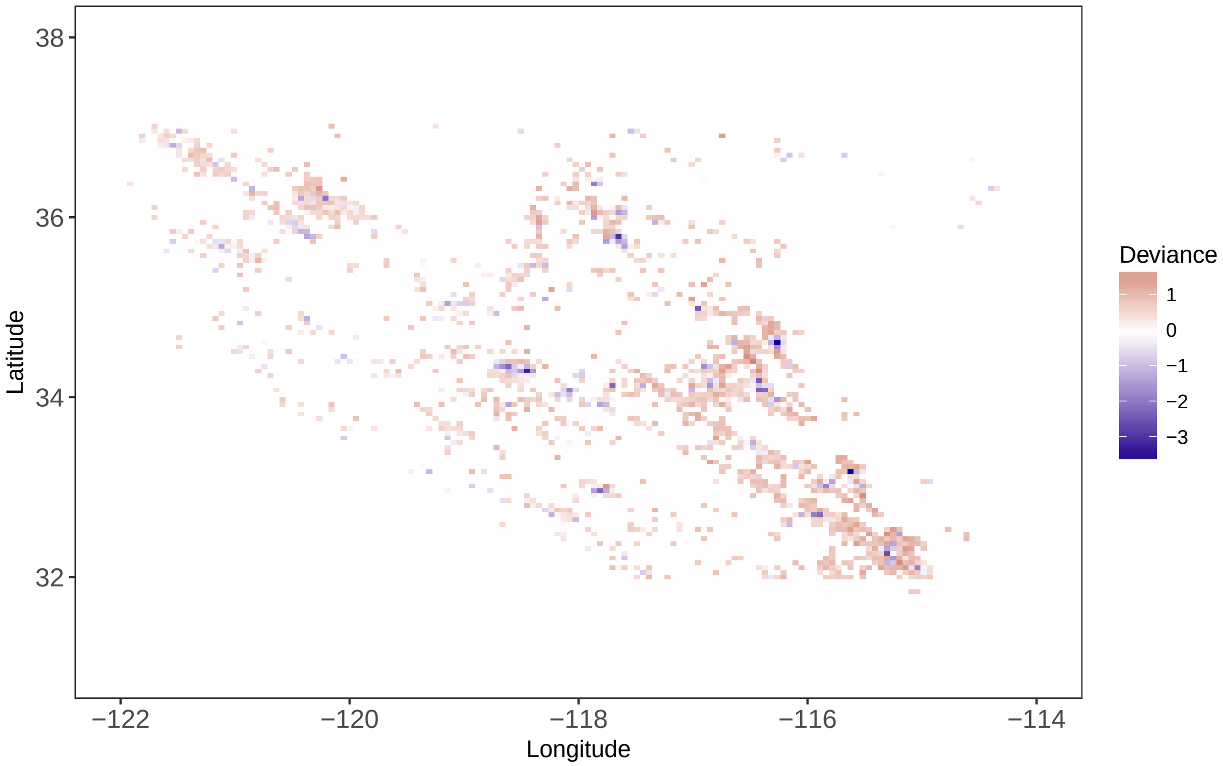

Figure 3.5: Deviances comparing models (3.2) and (3.4) over 0.05° × 0.05 cells. Red tinted cells indicate areas where the model (3.2), including strike angles, fit better than the null model (3.4). Blue tinted cells indicate areas where model (3.2) fit worse than the null model (3.4).

indicate that including $\tilde{\theta}$ in the triggering function tends to improve the model's fit especially in the Northwest and Southeast portions of the San Andreas Fault, with only a few small isolated areas where model (3.4) happened to outperform model (3.2).

The temporal fit of model (3.2) is shown in Figure 3.7. The model appears to fit adequately, given the close agreement of the Stoyan-Grabarnik statistic to its expected value under the assumption that the fitted model is correct.

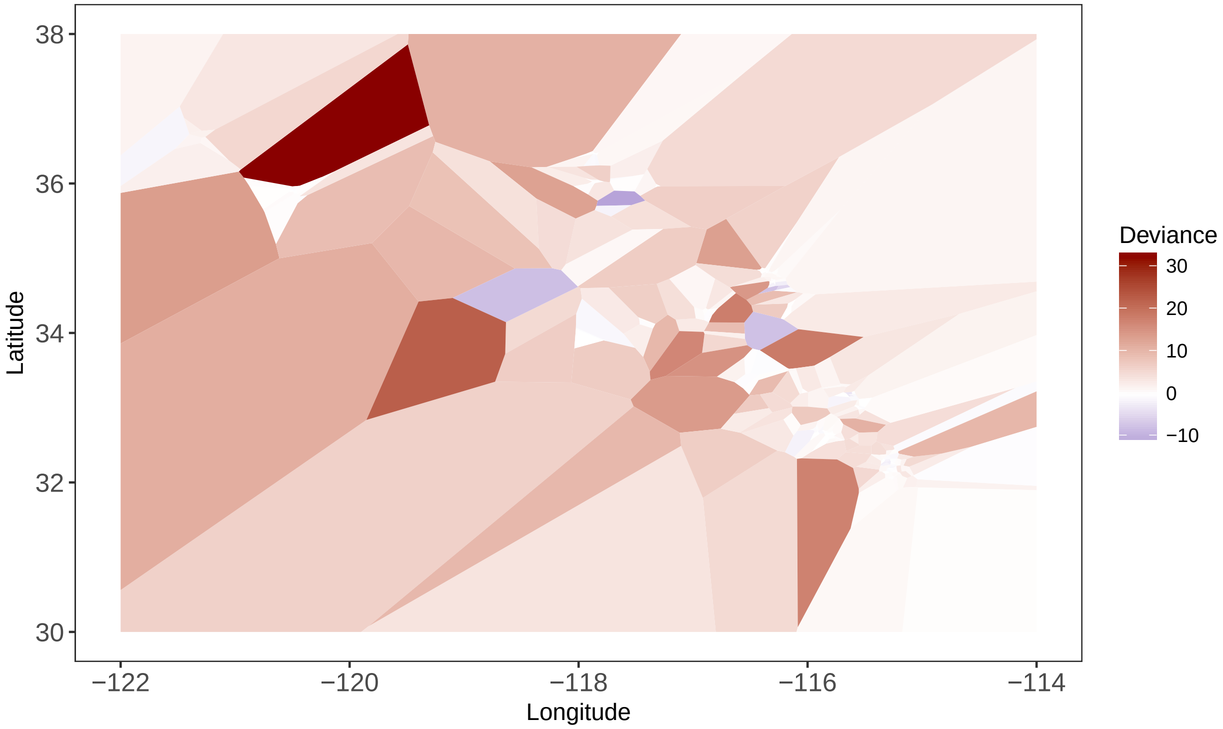

Figure 3.6: Voronoi deviance residuals, over a Voronoi tessellation of mainshocks with estimated focal mechanisms. Red tinted cells indicate areas where the model (3.2), including strike angles, fit better than the null model (3.4). Blue tinted cells indicate areas where model (3.2) fit worse than the null model (3.4).

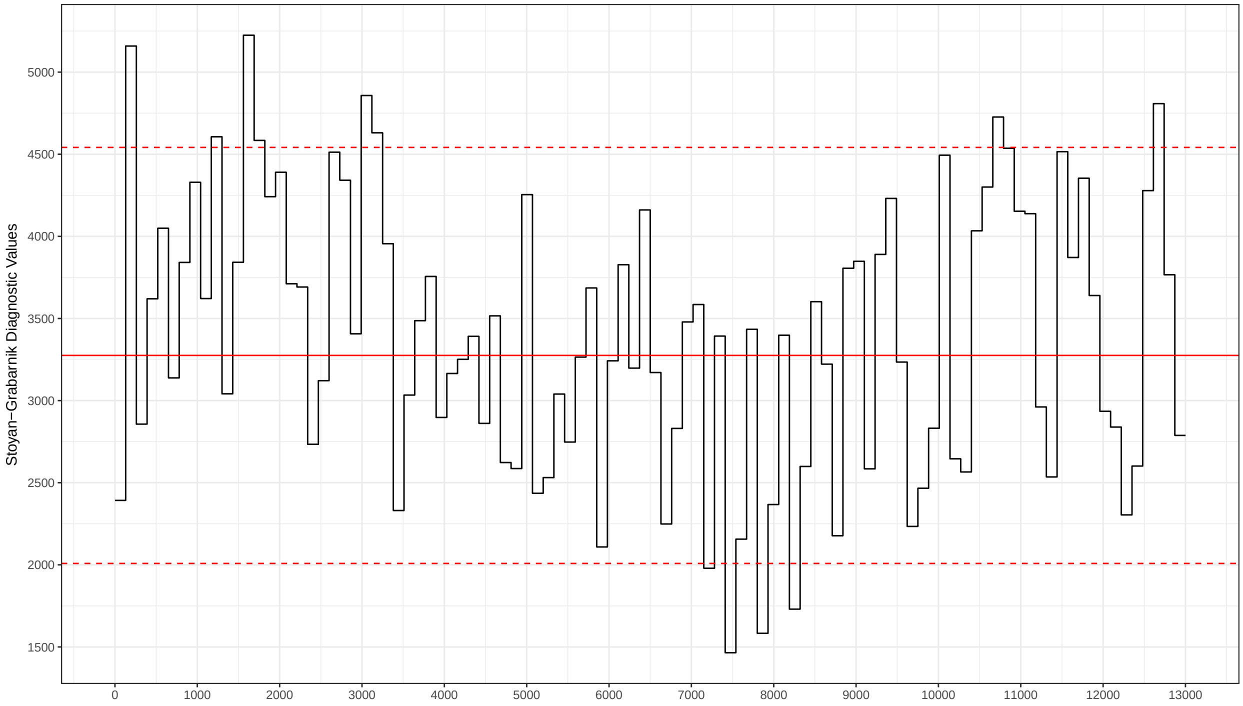

Figure 3.7: Stoyan-Grabarnik diagnostic for model (3.2) using 100 temporal bins, with each bin corresponding to 130 days. The black step function shows the observed sums of $1/\hat{\lambda}$ for all points in bin B_j, $j = 1, \ldots, 100$, solid red line indicates the expected value $(|B_j|)$ and dashed red lines indicate plus or minus two times the standard errors. The median number of points per bin was 35 and the number of points per bin ranged from 15 to 675.

3.5 Discussion

Nodal plane strike angles, $\tilde{\theta}$, fit post-mainshock events slightly better in terms of RMSE and substantially better in terms of mean absolute angle compared to strike angles estimated based on previous seismicity, θ^D and θ^a. However, including a second, orthogonal estimate for θ^D and θ^a and using the angle that fit the data best for each mainshock, as was done for $\tilde{\theta}$, led to such a substantial improvement in both $\tilde{\theta}^D$ and $\tilde{\theta}^a$, in terms of RMSE and mean absolute angle, that both $\tilde{\theta}^D$ and $\tilde{\theta}^a$ actually fit better than $\tilde{\theta}$. In fact, $\tilde{\theta}$ fit aftershock seismicity no better than $\tilde{\theta}^u$, the latter of which was simply the better-fitting of a random, uniformly distributed angle and its complement. The forecasting ability of $\tilde{\theta}$ thus seems to be entirely explained by the fact that the better-fitting of the nodal plane and its complement is chosen as the estimated nodal plane.

Comparing the results of each evaluation method to θ^g provides a sense of the amount of error inherent in the strike angle estimates. For instance, comparing $\tilde{\theta}$ to θ^g using the average mean absolute angle, the mean absolute angle to aftershocks for θ^g was 29.49° and that of $\tilde{\theta}$ was 35.55°, corresponding to an increase of 6.06°. Thus, the nodal plane estimate includes about 6.06° more uncertainty, on average, beyond what can be explained by the variability in the aftershock events alone.

Figure 3.3 shows the additional triggering of seismicity relative to $\tilde{\theta}$ for the non-parametric Hawkes model which includes $\tilde{\theta}$ in the conditional intensity function. While there is a large amount of variation in the additional triggering estimates, we do note that the LOESS curve has a general downward trend as the angle bins become more orthogonal in agreement with previous results indicating preferential occurrence of aftershocks along the strike angle. The large amount of variation corresponding to each angle bin may be due to errors in the strike angle estimates. Whether the improvement in model fit seem in the deviance and Voronoi deviance residuals comes from the accuracy of the estimated nodal planes or from choosing the better fitting strike angle of the two nodal planes is unclear and demonstrates some of the challenges of incorporating fault geometry to forecast future events.

The analysis here used only strike-slip mainshocks so that benefits of using nodal plane strike angles would be most easily discerned. Given that each focal mechanism includes two potential estimates for the true strike angle and each estimate contains a considerable degree of error, we attempt to determine, statistically, which strike angle estimate is most plausibly estimating the true strike angle and include this estimate in the model. Even for such strike-slip events and choosing the most plausible strike angle estimate from the focal mechanisms, we find no significant increase in accuracy of the nodal plane strike angles compared to strike angles based on previous seismicity or those chosen uniformly at random.

Although we find that strike-angle estimates from focal mechanisms tend not to outperform strike-angle estimates based on prior seismicity, we do find that including the most plausible strike-angle in model (3.2) was better than nothing in the sense of improving the fit of a non-parametric Hawkes model compared to model (3.4), as shown in Figures 3.5 and 3.6. The models are largely in agreement in most areas, but there are more places where including strike-angle estimates improved the model fit. Those pixels where model (3.2) underperformed tended to do so more drastically than areas where model (3.2) tended to perform better.

Similarly, Figure 3.6 shows that for the areas surrounding a mainshock with a strike angle estimate, including said strike angle estimate overwhelmingly tended to improve the fit of the model in all but a few Voronoi cells. That is, even though focal mechanism based strike angle estimates tended to under-perform other strike angle estimates, including them in the non-parametric Hawkes model tended to improve the overall spatial fit of the model.

The values of the Stoyan-Grabarnik diagnostic for model (3.2) are shown in Figure 3.7 and tend to lie inside the error bounds, despite considerable variation. In few instances did the fitted model forecast significantly too many events, corresponding to Stoyan-Grabarnik values below the lower error bound, or significantly too few events, corresponding to Stoyan-Grabarnik values above the upper error bound. The large variability is not unexpected as the variance of the Stoyan-Grabarnik diagnostic tends to rise in the presence of very small conditional intensities.

For this chapter, we chose to use a non-parametric form for the conditional intensity to keep the number of modeling assumptions to a minimum. This allows us to examine the effect of including fault plane geometry in the model in the absence of additional modeling assumptions. Investigating how the inclusion of fault geometry in other parametric model forms and how the effect of this inclusion compares to our work performed with the non-parametric model are important areas for future research. It is also important to note that this work was focused solely on earthquakes occurring in Southern California and further research should be undertaken to determine the generalizability of our results to seismicity in other areas of the world.

3.6 Conclusion

In this chapter, we examine different methods of forecasting the direction of aftershock seismicity using strike angles from nodal plane estimates, strike angle estimates based on previous seismicity, or random uniformly distributed angles. We evaluate the ability of these estimates to forecast the direction of future seismicity based on minimizing orthogonal distances and the mean absolute angle between the estimated strike angle and the locations of future events. We also include the estimated strike angles for each mainshock in a non-parametric Hawkes model and compare the deviance to a null Hawkes model that does not include the strike angle.

The best fitting nodal plane strike angle estimates are shown to be marginally better at forecasting the direction of future seismicity compared to estimates based on previous seismicity and angles chosen uniformly at random. These best fitting angles also seem to improve the estimates of the conditional intensity function in our non-parametric Hawkes model. However, by including orthogonal complements to the estimates based on previous seismicity and estimates chosen uniformly at random and then choosing the better fitting angle estimate for each method (as the better fitting nodal plane strike angle estimates were selected), the small advantage in forecasting future events shown by using the better fitting nodal plane strike angle estimates completely disappears. This indicates that the errors in nodal plane estimates are so large that their improvement in the conditional intensity estimate is entirely attributable to the fact that the better fitting of the two orthogonal nodal planes is selected. Our results thus confirm previous findings that the errors in estimated strike angles remain very large (Kagan 2003).

Important subjects for future work are to determine whether using strike angle estimates based solely on previous seismicity in conjunction with estimates based on nodal planes would lead to better fitting models, whether incorporating the dip and rake angles of the estimated nodal planes might prove beneficial in forecasting future seismicity, and how increased accuracy in future estimates of nodal planes would translate into increased accuracy in projections of seismicity.

CHAPTER 4

Parameter Estimation using the Stoyan-Grabarnik Statistic

4.1 Introduction

The most commonly used method to estimate parameters of point processes has been to maximize the log-likelihood (MLE). Parameter estimates using the MLE have been shown to benefit from several desirable properties, such as consistency of the estimates (Ogata 1978), but are also known to be computationally taxing as each iteration of the optimization routine requires the numerical integration of an intractable integral term (Harte 2012). This difficulty in computation serves as a barrier to the use of point processes as the estimation procedure for even moderately sized data sets can take a non-trivial amount of time to converge.

Several alternative approaches have been proposed to estimate parameters for ETAS models, or at least increase convergence times of parameter estimates. One such approach is to use an EM algorithms (Veen and Schoenberg 2008) which views the estimation procedure as a branching process with an incomplete data log-likelihood. While results using an EM algorithm seem promising, implementing the procedure is non-trivial. A second approach to speed time to convergence was proposed by Schoenberg (2013) and involves writing the triggering function as a density and approximating the integral term as a much simpler summation. Similar to the use of EM algorithms, this approach does not hold for point processes in general and, as an approximation of the integral is involved, may very well lead to inaccurate estimates of parameter values.

In this chapter, we propose a new parameter estimation technique using the Stoyan-Grabarnik statistic (Stoyan and Grabarnik 1991) which bypasses the need to compute the integral of the conditional intensity function. In section 4.3, we give an overview of the Stoyan-Grabarnik statistic, the use of this statistic for parameter estimation, and methods of evaluating the fit of parameter estimates. 4.4 compares parameter estimates using the new process with traditional MLE estimates and also compares the two methods in terms of computation time. We then discuss the results of the methods and the issues still faced by using the Stoyan-Grabarnik statistic in Sections 4.5 and 4.6, respectively.

4.2 Data and Simulations

We compare the estimated parameter outputs using the SG statistic to those computed via MLE by using simulated earthquake catalogs using the ETAS parameterization

$$\lambda(x,y,t|\mathcal{H}_t) = \mu(x,y) + \sum_{i:t_i<t} \frac{K(p-1)c^{p-1}(q-1)d^{q-1}}{\pi(t-t_i+c)^p} \cdot \frac{e^{\alpha(m-M_c)}}{((x-x_i)^2+(y-y_i)^2+d)^q}. \quad (4.1)$$

For simplicity, we assume a homogeneous background rate though in applications an inhomogeneous background rate would be more appropriate. For our simulations, we select 75 time periods to simulate over uniform randomly between 100 and 7,000 days. We also set the parameter values to be ($\mu = 0.5, K = 0.25, c = 0.02, p = 1.3, \alpha = 0.75, d = 0.05, q = 1.8$). Additionally, we set the value of β in our simulations, which is used to describe the magnitude distribution based on Gutenberg and Richter (1944), to equal 0.5. Our simulations use power laws for both the spatial and temporal decay of aftershocks along with exponential productivity for magnitudes and exponential magnitude density. Spatial locations are simulated in the unit square and aftershocks are located isotropically around its parent event.

We also use a simpler temporal ETAS model to diagnose estimation issues which arose in using equation 4.1 for our simulations. The parameterization used for the temporal ETAS model is

$$\lambda(t|\mathcal{H}_t) = \mu + \sum_{i:t_i<t} \frac{K(p-1)c^{p-1}}{(t-t_i+c)^p} \cdot e^{\alpha(m-M_c)}. \quad (4.2)$$

where, again, we assume a stationary background rate and simulate three time periods between 5,000 and 10,000 days. We set parameter values for equation 4.2 to be ($\mu = 0.2, K = 0.2, c = 0.5, p = 2, \alpha = 1.5$) and we set $\beta = 2.4$.

Finally, we use an earthquake catalog of 542 events which occurred around the time of the 1999 Hector Mine earthquake. Similar catalogs were analyzed in Ogata et al. (2003) as well as Schoenberg (2013). For this earthquake catalog, we constrain magnitudes to be $M \geq 3$ in the region falling between latitude 34 and 35 and longitude -117 and -116 between the dates 1999/10/16 and 2000/12/23. Data for this catalog were recorded by the Southern California Seismic Network (SCSN) and obtained from the Southern California Earthquake Data Center (SCEDC) (*California Institute of Technology (Caltech)* 1926; *SCEDC* 2013)

4.3 Methods

4.3.1 Stoyan-Grabarnik statistic

The Stoyan-Grabarnik (SG) statistic (Stoyan and Grabarnik 1991), denoted as

$$\sum_{i:\mathbf{s}_i, t_t \in B_j} \frac{1}{\lambda(\mathbf{s}, t)},$$

is a statistic which relates the sum of the inverse conditional intensity for a set of points in space-time bins to the volume of space-time bins. For $\lambda(\mathbf{s}, t) > 0$, and space-time bin B_j, the SG statistic states

$$\begin{aligned} E\left(\sum_{i:\mathbf{s}_i, t_t \in B_j} \frac{1}{\lambda(\mathbf{s}_i, t_i)}\right) &= E\left(\int \frac{1}{\lambda(\mathbf{s}, t)} dN\right) \\ &= E\left(\int \frac{1}{\lambda(\mathbf{s}, t)} \lambda(\mathbf{s}, t) dt d\mathbf{s}\right) \\ &= |B_j| \end{aligned}$$

by the martingale formula where $|B_j|$ denotes the volume of the space-time bin. In Baddeley et al. (2005), the SG statistic was proposed as a diagnostic method for point processes. More recently, Cronie and Lieshout (2018) proposed using SG as a method of bandwidth selection for kernel smoothing. Here, we layout the framework for using the SG statistic as a method of parameter optimization.

For a set of points in some space-time window, $\{(\mathbf{s}, t) : \mathbf{s} \in S, t \in T\}$, we divide the entire spatial-temporal window into n disjoint space-time bins B_j for $j = 1, 2, \ldots, n$ such that $(\mathbf{s}_i, t_i) \in B_j$ implies $(\mathbf{s}_i, t_i) \notin B_k$ when $j \neq k$. Figure 4.1 illustrates the binning procedure graphically for 2×2 spatial bins and 4 temporal bins.

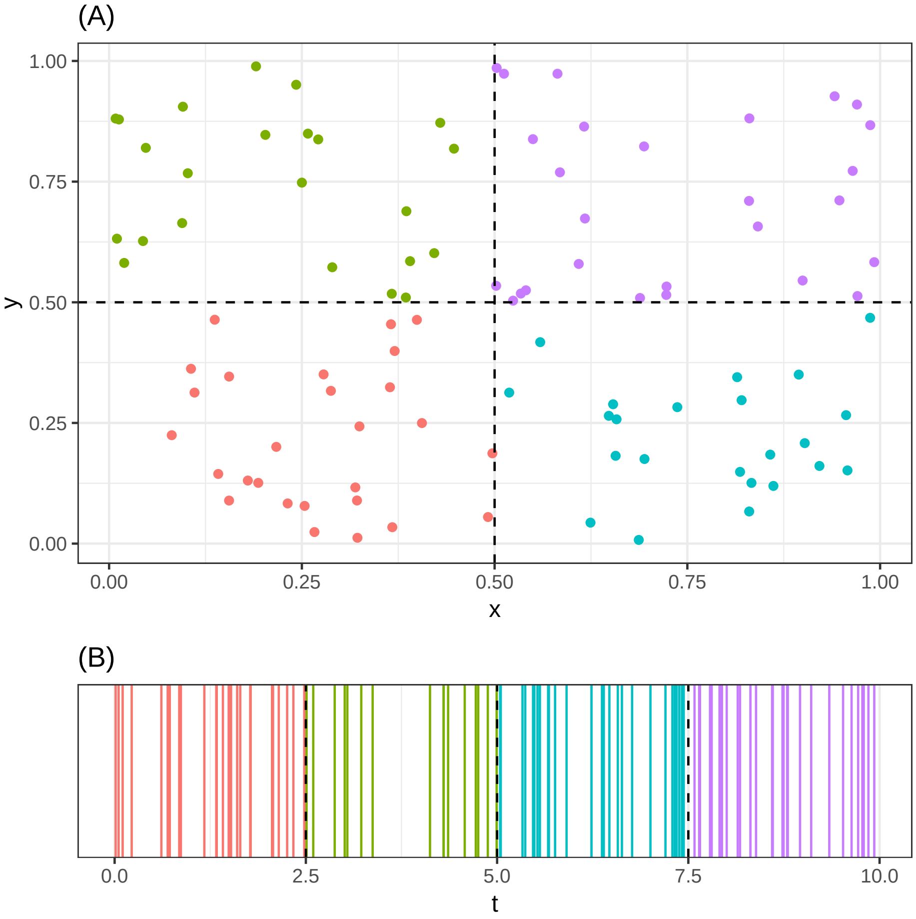

Figure 4.1: An example of the binning procedure to separate points in a spatial-temporal window $S \times T$ into individual space-time bins, B. Plot (A) shows the spatial binning of points into 2×2 bins. Plot (B) shows the temporal binning with 4 bins.

For a given parameterization of a point process model and a given loss function ($loss$), we can compute the parameter estimates using conjugate gradient methods from the **R** programming language's `optim()` (Nash and Varadhan 2011; Nash 2014) function to optimize

$$\sum_{j=1}^{n} loss\left(\sum_{i:\mathbf{s}_i,t_t\in B_j}\frac{1}{\lambda(\mathbf{s},t)}-|B_j|\right).$$

Within `optim`, we use the conjugate gradient methods with Beale-Sorenson updates. We also set the maximum number of iterations to 5,000 so each optimization run has enough time to converge. Starting values for each `optim` run were set to be 20% larger than the actual values used to create the simulated catalogs. For our simulated results in Section 4.4, we adopt the L_1 loss function and discuss why we opted not to use the L_2 loss function in Section 4.6.1. That is, we optimize the function

$$\sum_{j=1}^{n}\left|\sum_{i:\mathbf{s}_i,t_t\in B_j}\frac{1}{\lambda(\mathbf{s},t)}-|B_j|\right|. \tag{4.3}$$

When computing parameter estimates using the SG statistic for our simulated earthquake catalogs, we optimize our parameters using two different sets of bins. First, we separate the spatial locations into an equally sized 2×2 grid and 5 equally sized temporal bins. For the second set of bins, we use 3×3 equally sized spatial bins and 10 equally sized temporal bins.

4.3.2 Maximum likelihood estimation

As described in Chapter 2, the log-likelihood for a self-exciting point process is

$$\ell(\theta)=\sum_{i=1}^{N}\log(\lambda(\mathbf{s}_i,t_i|\mathcal{H}_t))-\int_{\mathcal{S}}\int_0^T\lambda(\mathbf{s},t|\mathcal{H}_t)\ dtd\mathbf{s}.$$

We compute the estimates of our simulated earthquake catalogs using the ETAS parameterization shown in equation 4.1 by using conjugate gradient methods using **R**'s `optim` function and by numerically approximating the integral term of the MLE using the `suave()` function in the **R2Cuba** R package (Thomas Hahn; Interface to R was written by Annie Bouvier and Kiêu 2015). Similar to our optimization routine using the SG statistic, we set the maximum number of `optim` calls to 5,000 and use Beale-Sorenson updates for the conjugate gradient method. We also use starting values that are 20% larger than the parameter values used to create each simulated earthquake catalog. For the `suave` numerical integration function we use the default set of options.

4.3.3 Residual ratios and output comparison

For comparing individual parameter estimates to their actual values used in our simulations, we compute the residual error ratios where for parameter θ and its estimated value $\hat{\theta}$ we compute

$$error_\theta = \frac{\hat{\theta} - \theta}{\theta}.$$

To summarize the overall fit of the p parameters for our models, we compute the root sum of squared errors

$$RSSE = \sqrt{\sum_{i=1}^{p} error_{\theta_i}^2}.$$

4.4 Results

4.4.1 Simulations

Figure 4.2 shows the elapsed time for `optim` to converge, using both MLE and SG parameter estimation routines for 75 simulated catalogs of varying size. Each estimation procedure was run for the same set of simulated catalogs to make results comparable. For the parameters estimated using the SG statistic, we show the results for both bin schemes described in Section 4.3.1. Time to convergence for both estimation routines increased as the simulated earthquake catalog increased and time to convergence was slightly longer using the SG statistic with $3 \times 3 \times 10$ bins than for the SG statistic with $2 \times 2 \times 5$ bins. The optimization routine using the SG statistic was markedly faster than the optimization routine using the MLE regardless of the number of bins used to compute estimates using the SG statistic. As the size of the simulated catalogs approach 4,000 events per catalog, the MLE took longer than 3 hours to converge while the SG statistic took less than 2 hours to converge using $3 \times 3 \times 10$ bins and less than 1 hour using $2 \times 2 \times 5$ bins.

The error ratios for the individual parameters using each method are shown in Figures 4.3 and 4.4. Figure 4.3 shows error ratios for all simulated catalogs while Figure 4.4 shows only the error ratios that were greater than -1 and less than 2.

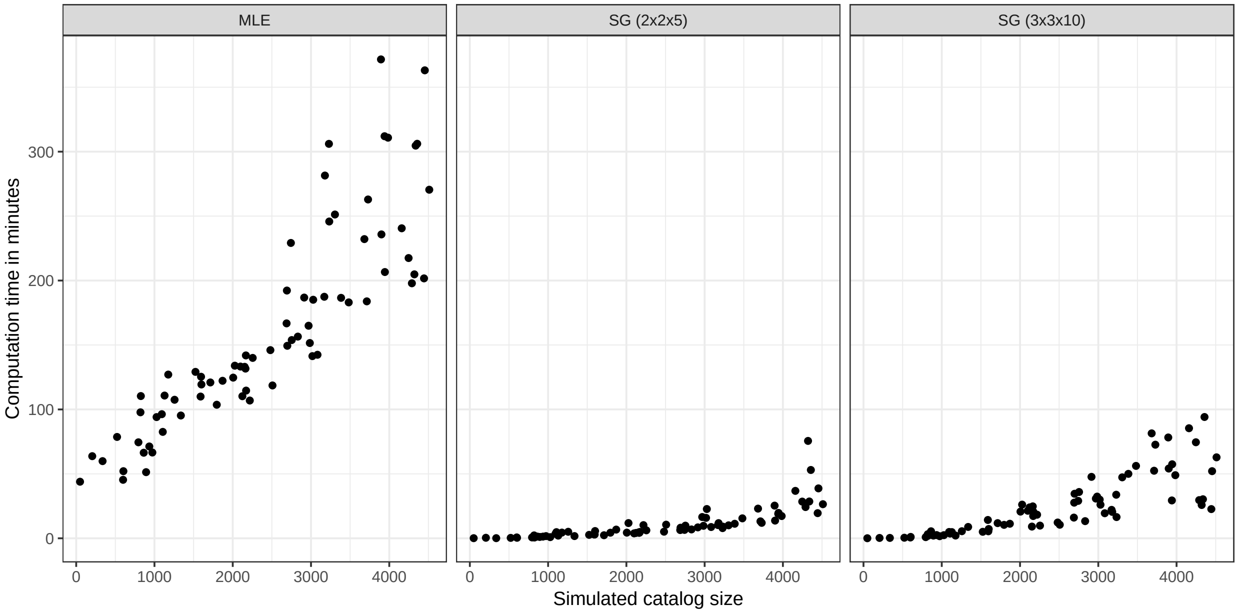

Figure 4.2: Computation times to complete parameter estimation via MLE, SG with 2 × 2 spatial bins and 5 temporal bins, and SG with 3 × 3 spatial bins and 10 temporal bins. The x-axis for each plot is the number of events in each simulated earthquake catalog and the y-axis shows the length of time required for each method to converge using conjugate gradient methods in `optim`

There were a handful of instances when the parameter values computed using the MLE were extremely poor resulting in, for instance, parameter estimates for p which were 10 times larger than the actual parameter value used in simulations. There are also parameter values for both MLE and SG that remained stuck at, or near, their starting values such as μ for MLE and K, c, α, and q for SG using $2 \times 2 \times 5$ bins.

Comparing the different binning schemes used for parameter optimization using the SG statistic, the binning scheme with more bins tended to result in error ratios that were much more variable than the binning scheme with fewer bins. The binning scheme with more bins, however, did result in parameter values that moved below their starting values for parameters α and d when compared to the binning scheme with fewer bins.

Figures 4.5 and 4.6 show the RSSE for simulated catalogs using each optimization routine. Figure 4.5 shows the RSSE for *all* of the simulated catalogs while Figure 4.6 caps the y-axis to be less than 6 to better show the typical RSSEs.

Similar to what was shown in the individual parameter plots, there were a handful of instances when the parameter estimates using the MLE were extremely poor. That is, there were instances when the RSSE was in excess of 10. Moreover, the RSSE using the MLE tended to be more variable than using SG with $2 \times 2 \times 5$ bins but less variable than using SG with $3 \times 3 \times 10$ bins. For each method, as the size of the simulated catalogs increased, there is a reduction of the variation in the RSSE with the variation of the RSSE using SG with $2 \times 2 \times 5$ bins showing the greatest reduction as simulated catalog size increased. For the MLE and both SG optimization routines, the smallest values of RSSE tended to be no less than 0.46 except in once instance where SG produced an RSSE of 0.35.

Comparing the different binning schemes using the SG statistic, we note that using more bins tended to result in a much poorer performance for the simulated catalogs shown. Using more bins also drastically increased the variability of the RSSE compared to the SG using fewer bins.

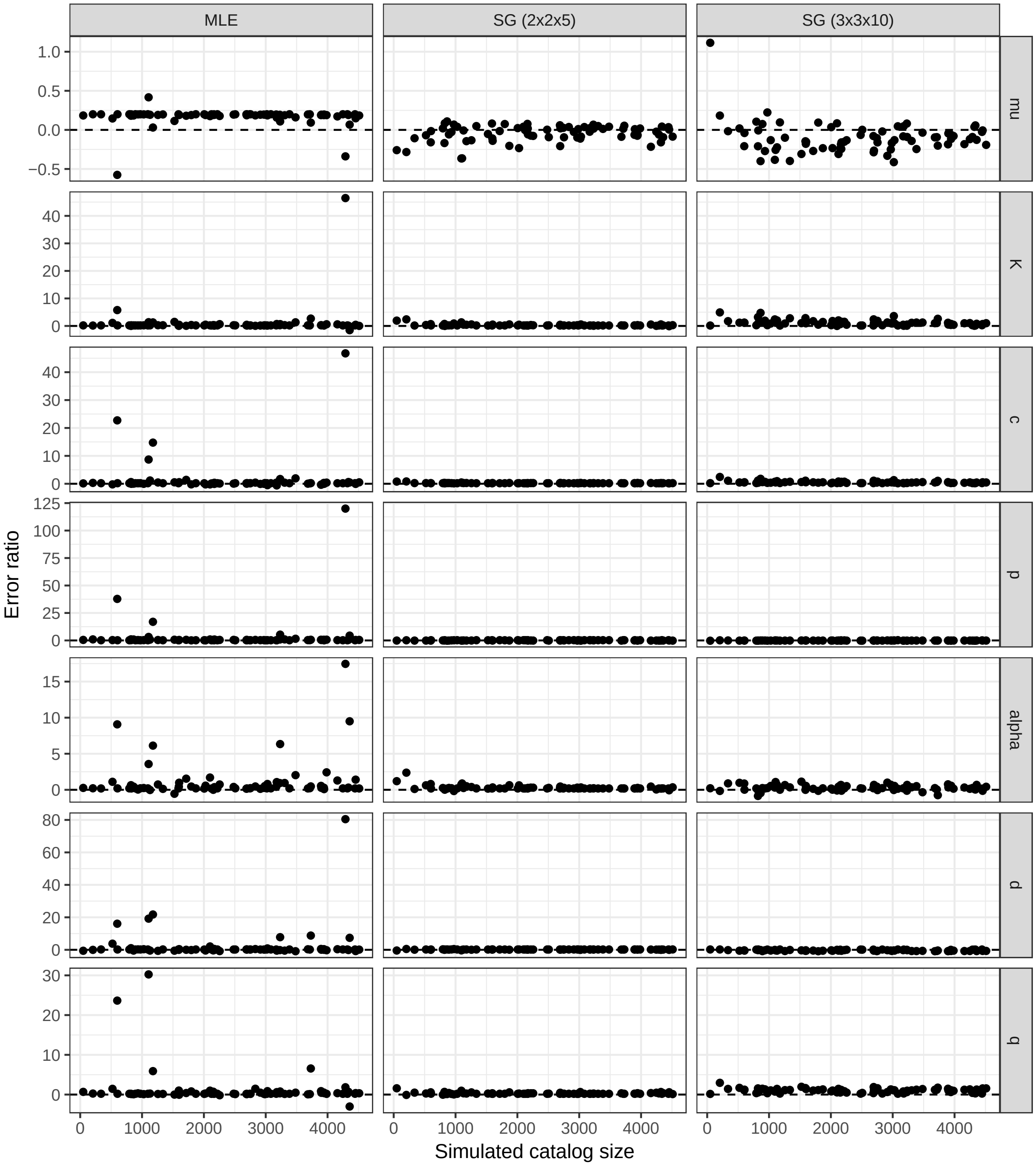

Figure 4.3: Individual parameter error ratios for parameters estimated via MLE, SG with 2×2 spatial bins and 5 temporal bins, and SG with 3×3 spatial bins and 10 temporal bins. The x-axis for each plot is the number of events in each simulated earthquake catalog and the y-axis shows the error ratios for each parameter and each estimation procedure. Limits of the y-axis are determined by the variability in error ratios for each parameter.
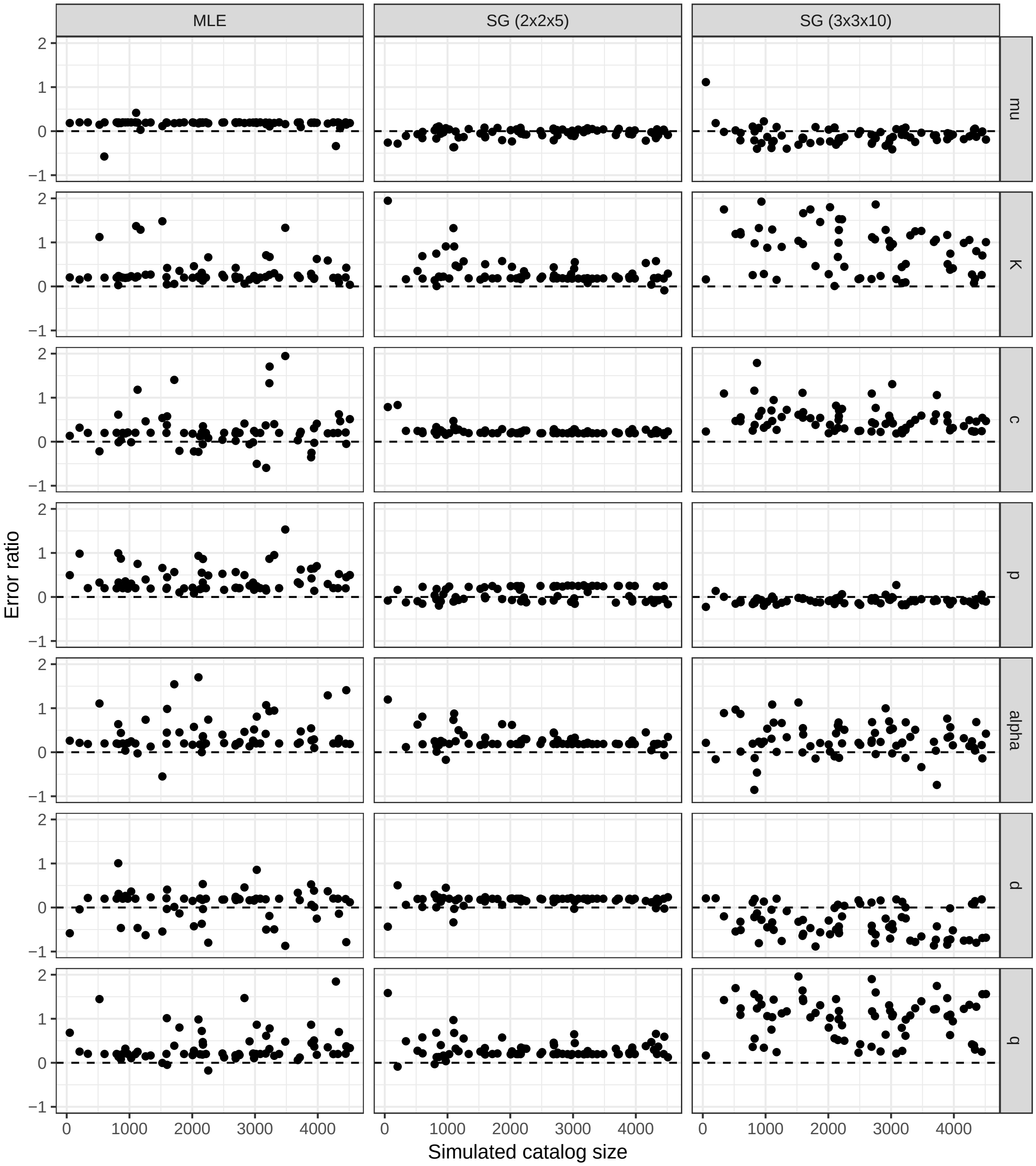
Figure 4.4: Individual parameter error ratios for parameters estimated via MLE, SG with 2×2 spatial bins and 5 temporal bins, and SG with 3×3 spatial bins and 10 temporal bins. The x-axis for each plot is the number of events in each simulated earthquake catalog and the y-axis shows the error ratios for each parameter and each estimation procedure. Limits of the y-axis are manually set from -1 to 2 to better compare the more typical estimates across methods.

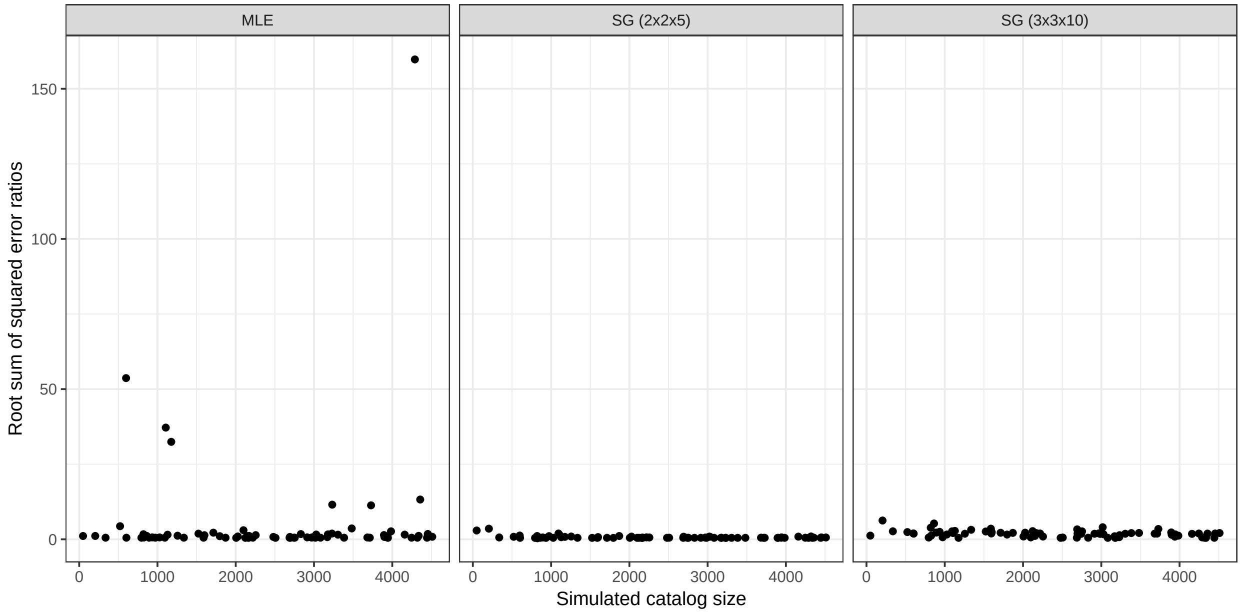

Figure 4.5: Root sum of squared error ratios for parameters estimated via MLE, SG with 2×2 spatial bins and 5 temporal bins, and SG with 3×3 spatial bins and 10 temporal bins. The x-axis for each plot is the number of events in each simulated earthquake catalog and the y-axis shows the root sum of squared error ratios for each estimation procedure. Limits of the y-axis are determined by the variability in root sum of squared error ratios for each parameter.

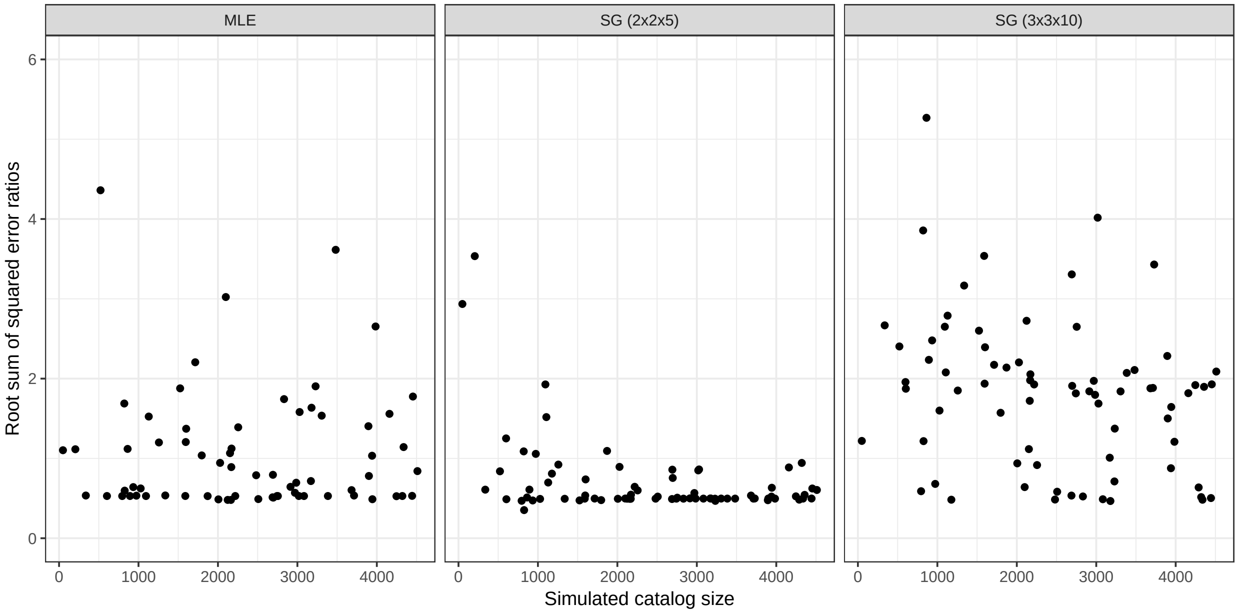

Figure 4.6: Individual parameter error ratios for parameters estimated via MLE, SG with 2×2 spatial bins and 5 temporal bins, and SG with 3×3 spatial bins and 10 temporal bins. The x-axis for each plot is the number of events in each simulated earthquake catalog and the y-axis shows the root sum of squared error ratios for each estimation procedure. Limits of the y-axis are manually set from 0 to 6 to better compare the more typical values across methods.

4.4.2 Stoyan-Grabarnik simulation performance issues

As noted above, there were many instances in which estimated parameter values remained at, or very near, the starting values used in the optimization procedure. To diagnose the potential causes of this issue, we ran a new batch of simulations using the simpler temporal-ETAS model shown in equation 4.2 and then compared parameter estimates using MLE and SG. Further, we estimate parameters using SG in an iterative fashion to help us to diagnose whether the estimation issues are due to the flatness of the conditional intensity, an issue with the method itself, or a coding issue. That is, we start by estimating one parameter, μ, and treat the other parameters as known. We then continue to estimate additional parameters in the order μ, α, p, c and K until all 5 parameters are treated as unknown. The estimated parameters using the MLE were estimated all at once.

Figure 4.7 (A) shows the results of our iterative estimation procedure for parameter estimation using SG and MLE. Each column shows the estimated parameter values using SG where the parameters which were treated as unknown are listed at the end of the rows and compares the values estimated using SG with those computed via the MLE.

As shown, the estimated parameter values computed using SG are either similar to or better than those computed using the MLE when up to four parameters are treated as unknown. However, once K is introduced as a parameter to be estimated, the estimates for μ, α, and c tend to start performing much more poorly. For instance, when K is considered known, and all other parameters were considered unknown, the average estimates for μ were 0.184 which was only slightly smaller than the actual value used in the simulation (where $\mu = 0.2$). When K is introduced as an unknown parameter, the average estimate of μ drops to 0.096, or roughly half of what the estimated parameter was when K was treated as known.

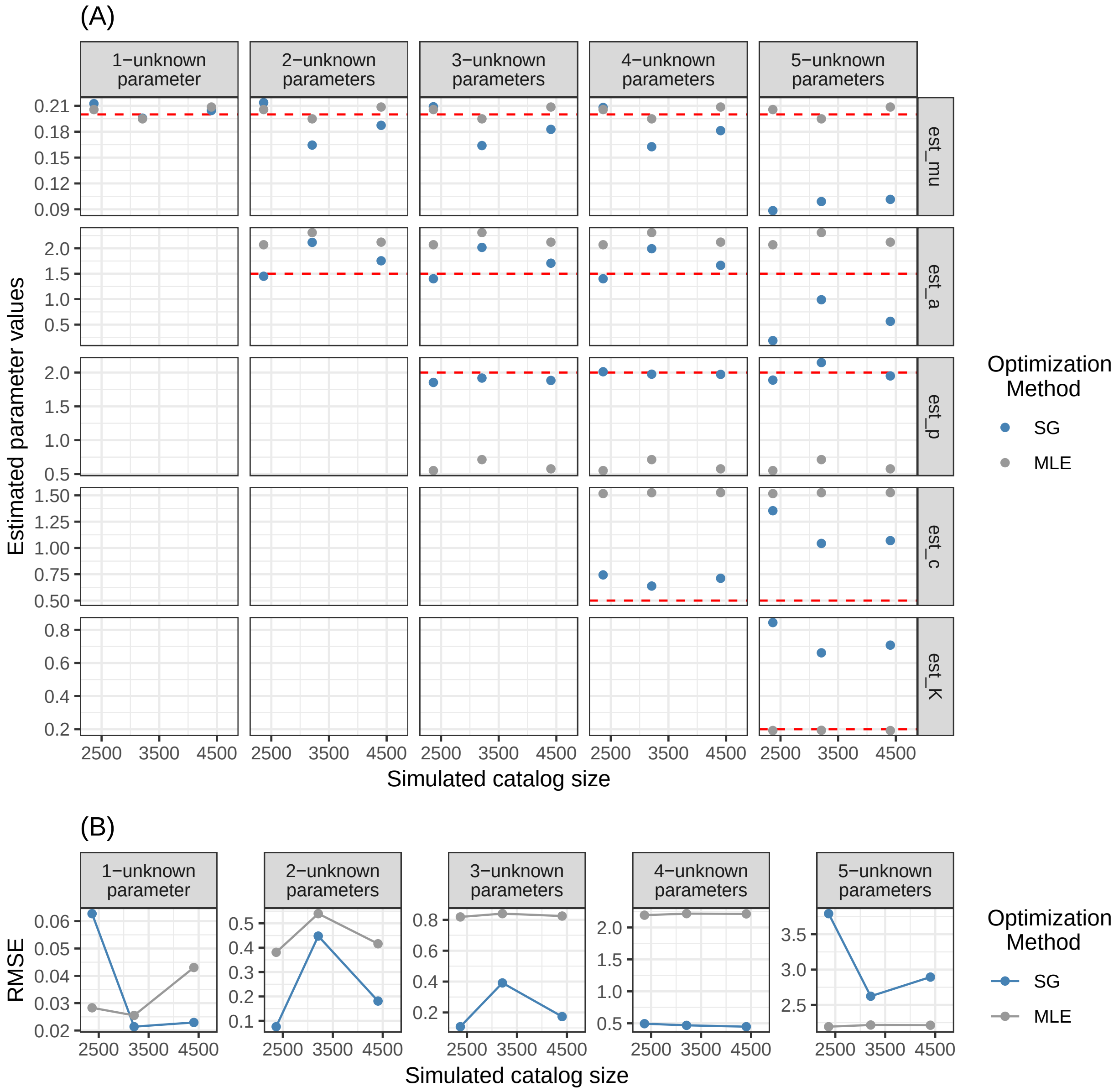

Figure 4.7: Estimated parameter values, plot (A), and RMSEs, plot (B), using both SG and MLE parameter estimation techniques. Columns for each plot separate the number of parameters which were estimated versus assumed known when estimating parameters via SG. Rows, for (A), indicate indicate which parameters were estimated. For plot (B), each panel compares the RMSE of SG and MLE for just the parameters which were estimated.

Figure 4.7 (B) shows the root mean squared error for just those parameters that were treated as unknown using our iterative estimation procedure where error is defined as the error ratio described in Section 4.3.3. That is, the left most pane shows the RMSE of SG and MLE where μ is the only parameter included in the computation whereas the right most pane shows the RMSE of all the parameters for each estimation procedure.

When only one parameter, μ, is treated as unknown, the RMSE using SG and MLE are very similar with mean RMSEs of 0.0357 for SG and 0.0323 for MLE. As the number of parameters treated as unknown increases up to four unknown parameters, we see that the RMSE for SG starts to outperform the estimates of the same parameters computed using MLE (0.47 mean RMSE for SG and 2.21 mean RMSE for MLE). However, once K is introduced as an unknown parameter, the RMSE for SG performs much more poorly than the MLE with mean RMSEs 3.1 for SG and 2.21 for MLE.

4.4.3 Hector Mine earthquake catalog

In Section 4.1, we note that one proposed method to speed up convergence times when estimating Hawkes processes is to rewrite the triggering function as a density and to estimate the integral term inside the MLE as the sum of an approximation. In Schoenberg (2013), the author uses a similar Hector Mine data set as the one described in Section 4.2 to estimate parameter values using the ETAS model parameterization shown in Equation 4.1 using both the approximation as well as MLE estimates computed using a numerical approximation of the integral term. The estimated parameter values from Schoenberg (2013) are shown in Table 4.1.

Using our Hector-Mine catalog, we re-estimated the parameter values using the SG statistic optimization method with 2×2 spatial bins and 5 temporal bins. The results of our estimation are also included in Table 4.1. For the estimation using the SG statistic, we used the estimated parameter values from Schoenberg (2013) as starting values.

Optimizing with the SG statistic took 7.6 seconds and largely agreed with the parameter estimates Schoenberg (2013) computed via MLE. Only two parameters, p and q were drastically different between the estimated parameters computed using the SG statistic and the values computed via MLE.

	μ	K	c	p	α	d	q
Integral Approximation (Schoenberg 2013)	0.03920	0.45600	0.05930	1.37000	1.14000	0.00010	1.67000
MLE (Schoenberg 2013)	0.11300	1.95000	0.00662	1.25000	0.08590	0.00016	2.09000
SG (2x2x5)	0.11530	1.93883	0.00692	1.60596	0.08988	0.00021	1.49397

Table 4.1: Estimated parameter values using the Hector-Mine data. The parameter values using the approximation method described in Schoenberg (2013) as well as the parameters estimated by Schoenberg (2013) using MLE are shown in the first two rows. The parameter values estimated using the SG statistic with 2 × 2 spatial bins, 5 temporal bins and using the Integral Approximation's estimates as starting values are shown in the third row.

4.5 Discussion

By avoiding the need to constantly compute numerical integrals, parameter estimations via the SG statistic converged much faster when compared to the MLE. As the size of the simulated catalog increased, the time savings using the SG statistic versus the MLE was on the order of hours. Using the MLE for even larger catalogs would be untenable while using the SG statistic would remain a viable option. The computation of the values for the conditional intensity at the points were computed using **C/C++** to increase convergence times but the code to compute these values is far from optimized. For instance, when points lie far away from each other in space time, and thus would have minimal covariance between the two, our code continues to loop through these points whereas establishing a break point in the code might increase convergence times even further. We also note that we are using **optim** for parameter optimization and did not test other optimization procedures.

Computation times using the SG statistic were also impacted by the binning scheme used with the binning scheme using fewer space-time converging faster than those using more bins. In the build up to the work shown in this chapter, we attempted parameter estimation with using a variety of binning schemes including using very few and using many bins. In the end, we settled on using the bin sizes shown but neither of the binning schemes should be considered optimal in terms of computation time and parameter estimate accuracy. One heuristic which showed some promise while working with the temporal-ETAS model was to choose $\sqrt{n}$ equally sized temporal bins where n is the size of the earthquake catalog.

Our results reveal a few issues in terms of individual parameter estimates for both optimization routines. For one, there were numerous instances where the MLE resulted in very poor parameter estimates. The maximum likelihood estimates appear to be volatile and to depend critically on particular clustering features of the catalog that can vary substantially from one realization to the next.

Another issue revealed in Figure 4.4 is that some parameter estimates remained at, or near, their starting values. For instance, the error ratio for μ remains at 0.20 as catalog size increases while the same error ratio using the SG statistic with $2 \times 2 \times 5$ bins converges to the actual value. A similar phenomenon was observed for parameters K, c, α, d and q for parameter estimates using the SG statistic with $2 \times 2 \times 5$ bins. Comparing these parameters to their counterparts computed using MLE though reveals that the MLE also had difficulty in converging to the actual parameter values and in fact often had larger error ratios than those using SG with $2 \times 2 \times 5$ bins. This would partly explain the RSSE for SG with $2 \times 2 \times 5$ bins appears to converge to rather quickly to the value of 0.46 in Figure 4.6 while the RSSE for the MLE tended to be more variable, but still no less than 0.46.

To assess the reason for these issues, we simulated earthquake catalogs which included time but not location coordinates for each point and then computed parameter estimates using a temporal-ETAS model. From the results shown in section 4.4.2, we find that the issues we encountered when estimating parameters for our simulated catalogs was due to the flatness of the conditional intensity functions whereby the overall flatness of the density makes it difficult for the optimization routine to recover the actual parameter values.

The binning scheme used for parameter estimates using SG were also hugely impactful on the parameter error ratios. Comparing SG with $2 \times 2 \times 5$ bins and SG with $3 \times 3 \times 10$ bins, we see that using more bins resulted in much more variation for the error ratios. The increased variation for individual parameter error ratios also explains why the RSSE for SG using $3 \times 3 \times 10$ bins also exhibited much more variation. We also note however that using more bins also helped lead the optimization routine away from its starting values, for better or worse.

For the RSSE shown in Figure 4.6, except in a couple instances, the RSSE for each optimization routine never fell below the value 0.46. Indeed, each method seems to be converging to this value. One reason this might occur would be because there were many instances across optimization routines where estimated parameter values remained at their starting values and this is most evident for SG with $2 \times 2 \times 5$ space-time bins. It could very well be the case that the MLE estimates would be improved by increasing the accuracy of the numerical approximation of the integral term, though this would also drastically increase computation times.

Comparing our estimates of the Hector-Mine data compared to results previously shown in Schoenberg (2013) demonstrates the promise of estimating parameter values using the SG statistic. As shown in Table 4.1, estimation via the SG statistic largely agreed with the estimates generated using the MLE even when using starting values that were relatively far away from their eventual convergence values. The two parameters that are markedly different between estimation via MLE via the SG statistic, p and q, were also not so radically different that they would be considered unreasonable. This shows that estimation using the SG statistic would seem to be a reasonable alternative to approximating the integral term while still remaining much more computationally efficient than using the MLE. It is also interesting that the issues we noted when computing parameter values using simulated catalogs did not appear when using the Hector-Mine catalog. And again, this has to do with the simulated catalogs being much more flat, in terms of their densities, than the Hector-Mine catalog.

4.6 Conclusion

In this chapter, we have shown a new method for computing parameter estimates for spatial-temporal point processes which avoids the need to numerically integrate the intractable integral found in current methods. Our results show that estimating via the SG statistic is faster than using the MLE and often results in estimates that are comparable. The final results however show that estimating parameter values for ETAS can be extremely challenging when the conditional intensity function is relatively flat over space-time.

From these results, we are inspired to continue working on this problem and highlight the ongoing issues and areas for future work below.

4.6.1 Ongoing issues and future work

While estimation via the SG statistic is promising, much work is still needed to make it a robust estimation procedure. One big issues we face, and are indeed still working to solve, is researching methods which will allow us to computed reliable parameter estimates even when the conditional intensity is relatively flat over its domain.

Another issue that we are interested in solving are methods to choose the optimum combination of spatial-temporal bins. As shown in this chapter, the size and number of the bins is incredibly impactful to the final parameter results. Developing methods to choose an optimal number of space-time bins is one of our top priorities in the near term. And while we focused on using equal sized bins, we have not yet looked into the effect of using bins of varying size and volume, nor have we investigated how one might choose bins of varying size. One thought we have had is to use Voronoi tessellations which might result in bins which better accounts for the density of points in space-time.

Besides developing methods of bin sizes, more work is needed to decide which loss function to use in the optimization routine. While not shown in this chapter, we did experiment with using an L_2 loss function but had issues with parameter estimates blowing up. One reason this might have occurred was because there were often bins much greater or smaller than the mean number of points. This in turn a small number of bins with values of the SG statistic that were distinctly different from the rest. The L_2 loss function seemed to be more sensitive to these outlying values and, we believe, chased parameter values to optimize the few outlying bins at the expense of the rest.

We have also yet to experiment with loss functions besides the L_1 and L_2 loss functions.

Nor have we experimented with adding weights to the values of the SG statistics for each bin. We hope that developing an optimal loss function might resolve issues such as the optimization routine settling on the starting values as well as the issues of parameter estimates blowing up when using the L_2 loss function.

Last, but certainly not least, establishing the theoretical underpinnings of the method and the mathematical properties thereof is a huge concern which we hope to address. One of the benefits of using the MLE is the desirable asymptotic properties of using it. Finding out which asymptotic properties are shared with the SG optimization routine would lead to a much more robust understanding of the method. Besides the desirable asymptotic properties, we are also interested in those properties, such as bias, which are undesirable. There exists a huge number of theoretical issues we are yet to explore but look forward with working them out in the future.

CHAPTER 5

Conclusion

Motivated by the desire to improve our ability to accurately forecast earthquakes, the first half of this dissertation examined the ability of focal mechanisms to predict the direction of future seismicity and the desirability of including such covariates in non-parametric Hawkes models. As shown, the errors which occur in the estimation of focal mechanisms lead them to be not as as predictive as hoped, yet the improvements seen in the performance of non-parametric Hawkes models which include directionality based on focal mechanisms does seem to indicate that there still some signal within the noise. Beyond hoping for new methods to come about which lead to better focal mechanism estimates, there is still viable work that statisticians can engage in for the prediction of aftershock locations. For example, the develop of hybrid methods wherein the information contained in focal mechanisms is coupled with other seismic data to create models which offer increased predictive ability. Further, methods for discriminating fault planes from auxiliary planes would also be a significant step forward to the application of fault estimates within earthquake forecasting model.

One aspect which became apparent during our work on seismic point process models was the incredible difficulty in computing parameter estimates for these models. It took some computations a week to finish for the results in the first half of this dissertation and this challenge was a primary motivation for the latter half of the work presented herein.

Computing parameter estimates for point process models by utilizing the Stoyan-Grabarnik statistic is a promising development in the applied point process literature. Though admittedly, much work is still to be done for this method. Still, the development of a computational technique which converges more quickly, is easier to implement and produces outputs which are comparable to estimates computed using maximum likelihood estimation is a significant development for the application of point process models in the future. It is our hope that this new computational method will allow scientists in fields outside of seismology to rapidly develop, test and apply new point process models in their respective fields.

The work displayed in this dissertation represents an exciting start to research we hope to continue pursuing long into the future.

APPENDIX A

Stoyan-Grabarnik estimation code

In this appendix, we show the code used to conduct parameter estimation via the Stoyan-Grabarnik statistic. Most of the code is written in the **R** statistical programming language (R Core Team 2017) except when indicated. It is our hope that the code will inspire additional development in what we have shown to be a promising alternative to point process parameter estimation.

A.1 Earthquake simulation code

The code below can be used to simulate an earthquake catalog under various distributional assumptions for the time and location of aftershocks. The code was written by Professor Frederic Paik Schoenberg and modified by James Molyneux.

The simulations use a space-time ETAS process with a homogeneous Poisson background process. That is, we simulate from

$$\lambda(x, y, t|\mathcal{H}_t) = \mu\rho(x, y) + \sum_{i:t_i<t} \frac{K(p-1)c^{p-1}(q-1)d^{q-1}}{\pi(t-t_i+c)^p} \cdot \frac{e^{\alpha(m-M_c)}}{((x-x_i)^2+(y-y_i)^2+d)^q}. \quad (A.1)$$

Points are simulated in the unit square, $[0, 1] \times [0, 1]$ in the temporal window $[0, \texttt{time}]$. The end points of the spatial window can be adjusted by setting the `x1` and `y1` arguments. The `time` argument is used to expand or narrow the length of time to simulate over and the simulation times are considered to be in days.

The spatial density of the background rate can be changed using the `rho` argument and can currently takes the value of `unifrho` (uniform spatial density).

The temporal triggering density, `gt`, can be used to simulate temporal aspects of the simulations using a power law, `powergt`, or exponential distribution, `expgt`. Similarly, the spatial triggering density, `gxy`, can take on a power law distribution, `powerxy`, or an exponential distribution, `expxy`.

The magnitude productivity, `gmi`, uses either an exponential productivity, `expprod`, or a constant point productivity (K), `pointprod`. Additionally, the magnitude density, `mdensity`, is assumed to have an exponential distribution, `expmag`.

The arguments `sor` and `keep` control whether the points should be sorted in chronological order and whether only points occurring within the space-time window should be kept, respectively. Both arguments take on values $\{0, 1\}$ where 1 indicates points should be sorted or only points within the spatial-temporal window should be kept, respectively. Additionally, `as.df` is a logical argument which, when `as.df = TRUE`, outputs the simulated catalog as a `data.frame`. Otherwise, the simulated catalog is output as a `list`.

Finally, the parameters used within the simulation should be input, as a `list`, using the `params` argument. Here, `b` governs the magnitude distribution in the simulation but is not a parameter which is estimated. Note, if the function will return an error if the input parameters leads to a super-critical branching process.

```r
simulate_catalog = function(x1 = 1, y1 = 1, time = 1e2,
                            rho = unifrho, gt = powergt, gxy = powerxy,
                            gmi = expprod, mdensity = expmag,
                            sor = 1, keep = 1, as.df = TRUE,
                            params = list(mu = 0.5, K = 0.2, c = 0.02,
                                          a = 1.5, p = 1.1, d = 0.001,
                                             q = 1.5, b = 0.5)) {

# Lay down the background points.
y = bgpts(x1, y1, time, mu, rho, mdensity, params)

# Calculate branching ratio. Stop if br > 1.
calcbr = mean(gmi(mdensity(1000000, params), params))
if(calcbr > 1.0) {
```

```
  cat("error, branching ratio = ", calcbr, " > 1.")
  return(0)
}

stop1 = 0
if (y$n < 0.5) stop1 = 2
w = y
while(stop1 < 1) {

  ## Place aftershocks down around y.
  z = aft(w, x1, y1, time, gt, gxy, gmi, mdensity, params)
  # cat(z$n," ")
  if (z$n > 0.5) {
    y = combine1(y,z)
    w = z
    if (min(z$t) > time) {
      stop1 = 2
    }
  }
  if(z$n < 0.5) {
    stop1 = 2
  }
}

## To keep just the points in the window.
if (keep == 1) {
  y = keep1(y, x1, y1, time)
}

## To have the points sorted chronologically.
if (sor == 1) {
```

```r
  y = sort1(y)
}

## To make the list a date.frame
if (as.df) {
  y = data.frame(lon = y$lon, lat = y$lat, t = y$t, m = y$m,
                 event_type = y$event_type)
  }
  y
}

# Helper function to lay down background points
bgpts = function(x1, y1, time, mu, rho, mdensity, params) {
  z1 = list()
  n = rpois(1, params$mu * time)
  z1$n = n
  xy = rho(n, x1, y1)
  z1$lon = xy[, 1]
  z1$lat = xy[, 2]
  z1$t = sort(runif(n) * time)
  z1$m = mdensity(n, params)
  z1$event_type = rep("background", times = z1$n)
  z1$ztimes = c()
  z1
}

# Helper functions to place aftershocks around each background event, y.
aft = function(y, x1, y1, time, gt, gxy, gmi, mdensity, params) {
  z1 = list()
  z1$t = c()
```

```
  z1$n = 0
  z1$m = c()
  z1$lat = c()
  z1$lon = c()
  z1$ztimes = c()

  ## vector of number of aftershocks for each mainshock.
  n2 = gmi(y$m, params)
  for (i in 1:length(n2)) {
    if (n2[i] > 0.5) {
      b1 = gt(n2[i], params)
      z1$ztimes = c(z1$ztimes, b1)
      z1$t = c(z1$t, b1 + y$t[i])
      xy = gxy(n2[i], y$m[i],params)
      z1$lon = c(z1$lon, xy[,1] + y$lon[i])
      z1$lat = c(z1$lat, xy[,2] + y$lat[i])
      z1$m = c(z1$m, mdensity(n2[i],params))
      z1$event_type = c(z1$event_type, "aftershock")
    }
  }
  z1$n = sum(n2)
  z1
}

# Helper function to combine results
combine1 = function(y, z) {
  z1 = list()
  z1$t = c(y$t,z$t)
  z1$n = y$n + z$n
  z1$m = c(y$m,z$m)
  z1$lat = c(y$lat,z$lat)
```

```
  z1$lon = c(y$lon,z$lon)
  z1$ztimes = c(y$ztimes, z$ztimes)
  z1$event_type = c(y$event_type, z$event_type)
  z1
}

# Helper function to keep only the pts of y that are within the
# spacetime window [0,x1] x [0,y1] x [0,time].
keep1 = function(y, x1, y1, time) {
  keeps = c(1:length(y$t))[(y$t < time) &
                            (y$lon < x1) &
                            (y$lat < y1) &
                            (y$lon > 0) &
                            (y$lat > 0)]
  y$t = y$t[keeps]
  y$m = y$m[keeps]
  y$lon = y$lon[keeps]
  y$lat = y$lat[keeps]
  y$n = length(keeps)
  y
}

# Helper function to sort the pts chronologically
sort1 = function(y) {
  ord2 = order(y$t)
  y$t = y$t[ord2]
  y$m = y$m[ord2]
  y$lon = y$lon[ord2]
  y$lat = y$lat[ord2]
  y$event_type = y$event_type[ord2]
  y
```

```
}

# Spatial layout of background points: rho.
# rho takes an integer n and x1 and y1 and outputs a matrix of n
# locations of mainshocks.

# Homogenous Poisson Process:
# Uniform spatial background density rho on [0,x1] x [0,y1].
unifrho = function(n, x1, y1) {
  x = runif(n, min = 0, max = x1)
  y = runif(n, min = 0, max = x1)
  cbind(x,y)
}

# Temporal triggering functions: gt.
# gt takes an integer n and outputs a vector of n nonnegative times
# since mainshock.

# Power law for temporal triggering:
powergt = function(n, params) {
  v = runif(n)
  params$c * (1 - v)^(1 / (1 - params$p)) - params$c
}

# exponential triggering function in time gt, with mean beta2.
expgt = function(n, params) {
  rexp(n, rate = 1 / params$beta)
}

# Spatial triggering functions: gxy.
# gxy takes an integer n and magnitude m and outputs a matrix of n
```

```
# locs from mainshock.

# Power law triggering in space according to ETAS (2.3), gxy, of
# Ogata (1998).
# NOTE: The density does not depend on magnitude of the mainshock.

powerxy = function(n, m, params) {
  v = runif(n)
  dist1 = sqrt(params$d * (1 - v)^(1 / (1 - params$q)) - params$d)
  thet1 = runif(n) * 2 * pi
  x = cos(thet1) * dist1
  y = sin(thet1) * dist1
  cbind(x,y)
}

# Exponential triggering in space.
# NOTE: the density does not depend on magnitude of the mainshock.

expxy = function(n, m, params) {
  v = rexp(n, rate = params$alpha)
  dist1 = sqrt(v)
  thet1 = runif(n) * 2 * pi
  x = cos(thet1) * dist1
  y = sin(thet1) * dist1
  cbind(x,y)
}

# Magnitude triggering functions: gmi
# Takes a vector of magnitudes, m, and outputs a vector of number of
# aftershocks per mainshock.

```

```
## Exponential productivity with parameters K and a for gmi.
expprod = function(m, params, M0 = 2.5) {
  rpois(length(m), params$K * exp(params$a * (m - M0)))
}

## Constant point productivity where each point has productivity K.
pointprod = function(m,params) {
  rpois(length(m), params$K)
}

# Magnitude density: mdensity
# mdensity takes an integer n and lower mag threshold m0 and outputs a
# vector of n magnitudes.

# Exponential magnitude density mdensity with minimum m0 and mean
# m0 + b1.
expmag = function(n, params, M0 = 2.5){
    rexp(n, rate = 1 / params$b) + M0
    # rexp(n, rate = params$b) + m0 # Testing (DELETE)
}
```

A.2 ETAS conditional intensity of points

The following **C** code is the code which computes the conditional intensity of the points in an earthquake catalog using Equation A.1. It was written by Professor Frederic Paik Schoenberg and modified by James Molyneux.

The function `calc_intensity` uses arguments `x` and `y` to set the spatial region to be $[0, x] \times [0, y]$. The vector of event times, `t`, should be in terms of days while the vector of event magnitudes minus the magnitude cut-off ($m_i - M_c$, where M_c is the minimum magnitude for the cut-off) should be included using the `hm` argument. The `n` argument indicates the number of events in the earthquake catalog and arguments `x1` and `y1` should be set to the vectors of longitudinal and latitudinal locations, respectively. The `theta` argument should be the ordered vector of input parameters, $(\mu, K, c, p, \alpha, d, q)$, where the order should be as shown. Finally, `est` should be a vector of length n which will end up being the outputted conditional intensities for each point.

```c
#include <R.h>
#include <Rmath.h>
#include <math.h>

void calc_intensity(double *x, double *y, double *t, double *hm, int *n
    ,
                    double *x1, double *y1, double *theta, double *est)
{
  int i,j,w;
  double r2, const1, eps, mu, K, c, p, a, d, q, b1;
  eps = .00000001;
  mu = theta[0]; K = theta[1]; c = theta[2];
  p = theta[3]; a = theta[4]; d = theta[5]; q = theta[6];

// Assign each point the baseline background value
for(i = 0; i < *n; i++) {
    est[i] = mu / (*x * *y);
  }

// Compute the constant values, which do not depend on the
// spatial-temporal locations of the events
  const1 = K * (p-1) * pow(c,(p-1)) * (q-1) * pow(d,(q-1)) / 3.141593;

// Loop and add the self-exciting contribution of the triggering
```

```
// density to each point
  for(i = 0; i < *n-1; i++){
    w = 0;
    for(j=i+1; w < 1; j++){
      if(j > *n-2) w = 2;
      r2 = (x1[i]-x1[j]) * (x1[i]-x1[j]) + (y1[i]-y1[j]) * (y1[i]-y1[j
         ]);
      b1 = const1 * pow(t[j]-t[i]+c,(-p)) * exp(a*hm[i])*pow(r2+d,(-q))
         ;
      est[j] += b1;
    }
  }
}
```

We now compile the **C** code above and load the function into **R** code in order to implement the computation of the conditional intensity in **R**.

```
# Compile and load the calc_intensity
system("R CMD SHLIB calc_intensity.c")
dyn.load("calc_intensity.so")

calc_intensity <- function(x, y, t, m, M0 = 2.5, theta) {

  # Calculate the values that are fed into the function
  hm = m - M0
  N = length(x)

  # This assumes each point falls in the spatial window
  # [0, 1] x [0, 1]
  intensity = .C("calc_intensity", as.double(1.0), as.double(1.0),
                  as.double(t), as.double(hm), as.integer(N), as.double(
                     x),
```

```
                 as.double(y), as.double(theta), est = double(N))
  intensity$est
}
```

A.3 Binning function for parameter estimation using the Stoyan-Grabarnik statistic

The following function takes in a set of points and a number of bins and assigns each point to the appropriate bin. Currently, bins are equally sized. The function outputs a `list` of point assignments, interval sizes and cut-values which can be used to assign points to bins and also compute the volume of the spatial-temporal bins

The arguments needed for `bin_it` are `x`, the vector of points in terms of unidimensional location (either x or y but not both) or time, `min_x`, the smallest value to start binning, `max_x`, the largest value to end binning and `n_bins`, the total number of bins desired.

```
bin_it <- function(x, min_x, max_x, n_bins = 10) {

  # Create the cut points for the bins
  bin_intervals = seq(min_x, max_x, length.out = n_bins + 1)

  # sort points into bins
  pt_assignments <- .bincode(x = x, breaks = bin_intervals)

  # Compute the bin size
  bin_size <- diff(bin_intervals[1:2])

  # Return the cut points, the point's assignments and the bin size
  list(bin_intervals = bin_intervals,
       pt_assignments = pt_assignments,
       bin_size = bin_size)
```

```
}
```

A.4 Stoyan-Grabarnik statistic for optimization

The following code takes an input vector of ETAS parameter values and computes the difference between the sum of absolute differences between the SG statistic, under a predefined binning scheme, and the volume of each bin. As with the `calc_intensity` function, the order of the parameters should be $(\mu, K, c, p, \alpha, d, q)$. Here, we assume that the earthquake catalog is assigned the name `eq_cat`. We also assume that the user has installed and loaded the `dplyr` library (Wickham et al. 2017) before running the function `sg_fit` and also that each point has been assigned to a bin for x, y, and t as `x_bin`, `y_bin` and `t_bin`, respectively. Finally, we assume that the user has calculated the volume of the individual space-time bin and assigned the value to the name `b_vol`.

```
sg_fit <- function(theta) {

  # Calculate the intensity of the points in eq_cat using the inputted
  # parameter values of theta
  eq_cat$intensity = calc_intensity(x = eq_cat$lon, y = eq_cat$lat,
                                    t = eq_cat$t, m = eq_cat$m,
                                    M0 = 2.5, theta = theta)

  output <- eq_cat %>%

    # Group each point by it's bin assignments
    group_by(x_bin, y_bin, t_bin) %>%

    # Compute the SG statistic for each bin
    summarize(sg = sum(1 / intensity)) %>%

```

```
    # Undo the grouping
    ungroup() %>%

    # Compute the sum of absolute differences between the SG statistics
    # for each bin and the volume of the bin.
    summarize(sum(abs(sg - b_vol))) %>%
    as.numeric()
  output
}
```

A.5 Example code usage

We now demonstrate an example usage of the code shown in Sections A.1 through A.4. Here, we assume that the user has run the code from each previous section so that the functions are loaded into their **R** environment. We also assume that the user has installed the `dplyr` and `optimx` (Nash and Varadhan 2011; Nash 2014) packages before running the code below.

```
library(dplyr)
library(optimx)

# Use set.seed(1) to make code reproducible
set.seed(1)

# Set the Time period to simulate over
Time <- 500 # days

# Set theta_actual for use in simulate_catalog()
theta_actual <- list(mu = 0.5, K = 0.25, c = 0.02, a = 0.75,
                     p = 1.3, d = 0.05 , q = 1.8, b = 0.5)

# Since the order of the parameters matters, and are slightly different
```

```
# between simulate_catalog() and sg_fit(), we also set the parameters
# values as a vector
theta <- c(mu = 0.5, K = 0.25, c = 0.02, p = 1.3, a = 0.75,
           d = 0.05, q = 1.8)

# Simulate the catalog
eq_cat <- simulate_catalog(x1 = 1, y1 = 1, time = Time,
                           params = theta_actual)

# Compute point bin assignments using 2x2 spatial bins and 5
# temporal bins
x_bins = bin_it(x = eq_cat$lon,
                min_x = 0,
                max_x = 1,
                n_bins = 2)
y_bins = bin_it(x = eq_cat$lat,
                min_x = 0,
                max_x = 1,
                n_bins = 2)
t_bins = bin_it(x = eq_cat$t,
                min_x = 0,
                max_x = Time,
                n_bins = 5)

# Compute the space-time bin volume
b_vol <- x_bins$bin_size * y_bins$bin_size * t_bins$bin_size

# Include the bin assignments in the simulated catalog
eq_cat$x_bin <- x_bins$pt_assignments
eq_cat$y_bin <- y_bins$pt_assignments
eq_cat$t_bin <- t_bins$pt_assignments
```

```

# Compute the SG statistic using the input parameters in theta
sg_fit(theta)

# > sg_fit(theta)
# [1] 104.5394

# Use conjugate gradient method in optimx() to compute estimated
# parameter values using starting values that are 20% larger than
# the parameter values used to create the simulated catalog
outputs <- optimx(par = theta * 1.2, fn = sg_fit,method = "CG",
                  control = list(
                                 # Raise max iterations to 500
                                 maxit = 5000,
                                 # Use Beale-Sorenson updates
                                 type = 3)))
# Print the estimated parameter values
outputs[1:7]

# > outputs[1:7]
#        mu         K            c        p         a         d
   q
# 0.5533382 0.2913035 0.0006948992 1.564469 0.8983599 0.0577397
   2.159865
```

Bibliography

Aki, K and P G Richards (2002). *Quantitative seismology*. University Science Books. Mill Valley, California.

Baddeley, Adrian, J Moller, and A G Pakes (2008). “Properties of residuals for spatial point processes”. In: *Annals of the Institute of Statistical Mathematics* 60.3, pp. 627–649.

Baddeley, Adrian, R Turner, and J Moller (2005). “Residual analysis for spatial point processes”. In: *Journal of the Royal Statistical Society* 67.5, pp. 617–666.

Balderama, Earvin et al. (2012). “Application of branching models in the study of invasive species”. In: *Journal of the American Statistical association* 107.498, pp. 467–476.

Bolt, B. (2003). *Earthquakes: Extended Fifth Edition*. W. H. Freeman. ISBN: 9780716756187. URL: `https://books.google.com/books?id=s2loQgAACAAJ`.

Bray, Andrew et al. (2014). “Voronoi residual analysis of spatial point process models with applications to California earthquake forecasts”. In: *The Annals of Applied Statistics* 8.4, pp. 2247–2267.

California Institute of Technology (Caltech) (1926). Southern California Seismic Network. International Federation of Digital Seismograph Networks. Other Seismic Network. DOI: `10.7914/SN/CI`.

Chu, A et al. (2011). “Comparison of ETAS Parameter Estimates across Different Global Tectonic Zones”. In: *Bulletin of the Seismological Society of America* 101.5, pp. 2323–2339.

Clements, Robert Alan, Frederic Paik Schoenberg, and Danijel Schorlemmer (2011). "Residual analysis methods for space–time point processes with applications to earthquake forecast models in California". In: *The Annals of Applied Statistics* 5.4, pp. 2549–2571.

Clinton, John F, Egill Hauksson, and Kalpesh Solanki (2006). "An evaluation of the SCSN moment tensor solutions: Robustness of the M-w magnitude scale, style of faulting, and automation of the method". In: *Bulletin of the Seismological Society of America* 96.5.

Console, R, M Murru, and G Falcone (2009). "Probability gains of an epidemic-type aftershock sequence model in retrospective forecasting of $\mathrm{M} \geq 5$ earthquakes in Italy". In: *Journal of Seismology* 14.1, pp. 9–26.

Cronie, O and M N M van Lieshout (2018). "A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions". In: *Biometrika*.

Daley, D J and D Vere-Jones (2007). *An introduction to the theory of point processes: volume I: Elementary Theory and Methods, Second Edition.*

— (2008). *An introduction to the theory of point processes: Volume II: General theory and structure, Probability and its Applications.*

Das, S and C Henry (2003). "Spatial relation between main earthquake slip and its aftershock distribution". In: *Reviews of Geophysics.*

Deming, W E (1943). *Statistical adjustment of data.* Oxford, England: Wiley.

Field, Edward H. (2007). "Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM)". In: *Seismological Research Letters* 78.1, pp. 7–16.

Field, Edward H et al. (2015). “Long-Term Time-Dependent Probabilities for the Third Uniform California Earthquake Rupture Forecast (UCERF3)”. In: *Bulletin of the Seismological Society of America* 105.2A, pp. 511–543.

Fox, E W, M B Short, and Frederic Paik Schoenberg (2016). “Modeling e-mail networks and inferring leadership using self-exciting point processes”. In: *Journal of the ...*.

Fox, Eric Warren, Frederic Paik Schoenberg, and Joshua Seth Gordon (2016). “Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences”. In: *The Annals of Applied Statistics* 10.3, pp. 1725–1756.

Gordon, Joshua Seth and Frederic Paik Schoenberg (2017). “A nonparametric Hawkes model for forecasting California seismicity”. In: pp. 1–41.

Guo, Yicun, Jiancang Zhuang, and Shiyong Zhou (2015). “An improved space-time ETAS model for inverting the rupture geometry from seismicity triggering”. In: *Journal of Geophysical Research: Solid Earth* 120.5, pp. 3309–3323.

Guo, Yicun et al. (2017). “Heterogeneity of direct aftershock productivity of the main shock rupture”. In: *Journal of Geophysical Research: Solid Earth* 122.7, pp. 5288–5305.

Gutenberg, B. and C. F. Richter (1944). “Frequency of Earthquakes in California”. In: *Bulletin of the Seismological Society of America* 34, pp. 185–188.

Harte, D S (2012). “Bias in fitting the ETAS model: a case study based on New Zealand seismicity”. In: *Geophysical Journal International* 192.1, pp. 390–412.

Hawkes, A G (1971). “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika*.

Helmstetter, A and M J Werner (2012). “Adaptive Spatiotemporal Smoothing of Seismicity for Long-Term Earthquake Forecasts in California”. In: *Bulletin of the Seismological Society of America* 102.6, pp. 2518–2529.

Henry, C and S Das (2001). “Aftershock zones of large shallow earthquakes: fault dimensions, aftershock area expansion and scaling relations”. In: *Geophysical Journal International* 147.2, pp. 272–293.

Jordan, T H and L M Jones (2010). “Operational Earthquake Forecasting: Some Thoughts on Why and How”. In: *Seismological Research Letters* 81.4, pp. 571–574.

Kagan, Yan Y. (2003). “Accuracy of modern global earthquake catalogs”. In: *Physics of the Earth and Planetary Interiors* 135.2-3, pp. 173–209.

— (2014). *Earthquakes.* Models, Statistics, Testable Forecasts. Oxford, UK: John Wiley & Sons, Ltd.

Kagan, Yan Y. and David D. Jackson (2014). “Statistical earthquake focal mechanism forecasts”. In: *Geophysical Journal International* 197.1, pp. 620–629.

Kagan, Yan Y., David D. Jackson, and Yufang Rong (2007). “A testable five-year forecast of moderate and large earthquakes in southern California based on smoothed seismicity”. In: *Seismological Research Letters* 78.1, pp. 94–98.

Marsan, David and Olivier Lengliné (2008a). “Extending Earthquakes’ Reach through Cascading”. In: *Science* 319.5866, pp. 1076–1079.

— (2008b). “Extending earthquakes’ reach through cascading”. In: *Science* 319.5866, pp. 1076–1079.

Mohler, G O et al. (2011). “Self-exciting point process modeling of crime”. In: *Journal of the American Statistical association* 106.493, pp. 100–108.

Nash, John C. (2014). “On Best Practice Optimization Methods in R”. In: *Journal of Statistical Software* 60.2, pp. 1–14. URL: `http://www.jstatsoft.org/v60/i02/`.

Nash, John C. and Ravi Varadhan (2011). “Unifying Optimization Algorithms to Aid Software System Users: optimx for R”. In: *Journal of Statistical Software* 43.9, pp. 1–14. URL: `http://www.jstatsoft.org/v43/i09/`.

Njabo, Kevin Yana et al. (2016). “Living with avian FLU Persistence of the H5N1 highly pathogenic avian influenza virus in Egypt”. In: *Veterinary Microbiology* 187, pp. 82–92.

Ogata, Yosihiko (1978). “The asymptotic behaviour of maximum likelihood estimators for stationary point processes”. In: *Annals of the Institute of Statistical Mathematics.*

— (1988). “Statistical models for earthquake occurrences and residual analysis for point processes”. In: *Journal of the American Statistical association.*

— (1998). “Space-time point-process models for earthquake occurrences”. In: *Annals of the Institute of Statistical Mathematics.*

— (2011). “Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity”. In: *Earth, Planets and Space* 63.3, pp. 217–229.

Ogata, Yosihiko, Lucile M Jones, and Shinji Toda (2003). “When and where the aftershock activity was depressed: Contrasting decay patterns of the proximate large earthquakes in southern California”. In: *Journal of Geophysical Research: Solid Earth* 108.B6.

Ogata, Yosihiko and Jiancang Zhuang (2006a). “Space–time ETAS models and an improved extension”. In: *Tectonophysics* 413.1, pp. 13–23.

Ogata, Yosihiko and Jiancang Zhuang (2006b). “Space–time ETAS models and an improved extension”. In: *Tectonophysics* 413.1-2, pp. 13–23.

Peng, R D, Frederic Paik Schoenberg, and J A Woods (2005). “A space-time conditional intensity model for evaluating a wildfire hazard index”. In: *Journal of the American Statistical association* 100.469, pp. 26–35.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Rathbun, Stephen L and Noel Cressie (2016). “Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes”. In: *Advances in Applied Probability* 26.01, pp. 122–154.

SCEDC (2013). Southern California Earthquake Center. Caltech. Dataset. DOI: `10.7909/C3WD3xH1`.

Schoenberg, Frederic Paik (2013). “Facilitated estimation of ETAS”. In: *Bulletin of the Seismological Society of America*.

Schoenberg, Frederic Paik and R D Patel (2012). “Comparison of Pareto and tapered Pareto distributions for environmental phenomena”. In: *The European Physical Journal Special Topics* 205.1, pp. 159–166.

Scholz, Christopher H (2002). *The Mechanics of Earthquakes and Faulting*. Second. Cambridge: Cambridge University Press.

Schorlemmer, Danijel et al. (2010). “First Results of the Regional Earthquake Likelihood Models Experiment”. In: *Pure and Applied Geophysics* 167.8-9.

Stomakhin, Alexey, Martin B Short, and Andrea L Bertozzi (2011). “Reconstruction of missing data in social networks based on temporal patterns of interactions”. In: *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data* 27.11, pp. 115013–115015.

Stoyan, Dietrich and Pavel Grabarnik (1991). “Second-order Characteristics for Stochastic Structures Connected with Gibbs Point Processes”. In: *Mathematische Nachrichten* 151.1, pp. 95–100.

Strader, Anne and David D. Jackson (2014). “Near-prospective test of Coulomb stress triggering”. In: *Journal of Geophysical Research: Solid Earth* 119.4, pp. 3064–3075.

Thomas Hahn; Interface to R was written by Annie Bouvier, The Cuba library has been written by and Kiên Kiêu (2015). *R2Cuba: Multidimensional Numerical Integration.* R package version 1.1-0. URL: `https://CRAN.R-project.org/package=R2Cuba`.

Tiampo, Kristy F and Robert Shcherbakov (2012). “Seismicity-based earthquake forecasting techniques: Ten years of progress”. In: *Tectonophysics* 522-523, pp. 89–121.

Utsu, T. (1961). “A Statistical Study on the Occurrence of Aftershocks”. In: *Geophysical Magazine* 30, pp. 521–605.

Veen, Alejandro and Frederic Paik Schoenberg (2008). “Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm”. In: *Journal of the American Statistical association* 103.482, pp. 614–624.

Wang, Q, David D. Jackson, and Yan Y. Kagan (2011). “California Earthquake Forecasts Based on Smoothed Seismicity: Model Choices”. In: *Bulletin of the Seismological Society of America* 101.3, pp. 1422–1430.

Wickham, Hadley et al. (2017). *dplyr: A Grammar of Data Manipulation.* R package version 0.7.4. URL: `https://CRAN.R-project.org/package=dplyr`.

Xu, Haiyong and Frederic Paik Schoenberg (2011). "Point process modeling of wildfire hazard in Los Angeles County, California". In: *arXiv.org* 2A, pp. 684–704. arXiv: `1108.0754v1 [stat.AP]`.

Zechar, J Douglas et al. (2009). "The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science". In: *Concurrency and Computation: Practice and Experience* 22.12, pp. 1836–1847.

Zechar, J D et al. (2013). "Regional Earthquake Likelihood Models I: First-Order Results". In: *Bulletin of the Seismological Society of America* 103.2A, pp. 787–798.

Zhang, T (2017). "On Independence and Separability between Points and Marks of Marked Point Processes". In: *Statistica Sinica.*