

Self-exciting point process modeling of crime

G. O. Mohler* M. B. Short† P. J. Brantingham‡ F. P. Schoenberg§
G. E. Tita¶

Abstract

Highly clustered event sequences are observed in certain types of crime data, such as burglary and gang violence, due to crime specific patterns of criminal behavior. Similar clustering patterns are observed by seismologists, as earthquakes are well known to increase the risk of subsequent earthquakes, or aftershocks, near the location of an initial event. Space-time clustering is modeled in seismology by self-exciting point processes and the focus of this paper is to show that these methods are well suited for criminological applications. We first review self-exciting point processes in the context of seismology. Next, using residential burglary data provided by the Los Angeles Police Department, we illustrate the implementation of self-exciting point process models in the context of urban crime. For this purpose we use a fully non-parametric estimation methodology to gain insight into the form of the space-time triggering function and temporal trends in the background rate of burglary.

1 Introduction

Criminological research has shown that crime can spread through local environments via a contagion-like process (Johnson 2008). For example, burglars will repeatedly attack clusters

*Department of Mathematics and Computer Science, Santa Clara University

†Department of Mathematics, University of California, Los Angeles

‡Department of Anthropology, University of California, Los Angeles

§Department of Statistics, University of California, Los Angeles

¶Department of Criminology, University of California, Irvine

of nearby targets because local vulnerabilities are well known to the offenders (Bernasco and Nieuwbeerta, 2005). A gang shooting may incite waves of retaliatory violence in the local set space (territory) of the rival gang (Tita and Ridgeway, 2007; Cohen and Tita, 1999). The local, contagious spread of crime leads to the formation of crime clusters in space and time.

Similarly, the occurrence of an earthquake is well known to increase the likelihood of another earthquake nearby in space and time. For example, we plot in Figure 1 a histogram of the times between “nearby earthquakes”, pairs of earthquake events separated in space by 110 kilometers or less, for all recorded earthquakes of magnitude 3.0 or greater in Southern California during 2004-2005. The histogram shows a spike at short times, indicating an increased likelihood of another event in the days following each earthquake. For a stationary Poisson process the distribution of times between pairs of events would be approximately uniform when the length of the time window is much larger than the longest time bin of the histogram.

In the case of residential burglary, evidence indicates that an elevated risk exists for both a house that has been recently burgled and its neighboring houses (Farrell and Pease, 2001; Short et al., 2009; Johnson et al., 2007). To illustrate this point further, we plot in Figure 1 a histogram of the times between “nearby burglaries”, residential burglaries separated in space by 200 meters or less, for all recorded residential burglaries within an 18km by 18km region of the San Fernando Valley in Los Angeles during 2004-2005. Again we observe a spike at short times, indicating an increased likelihood of victimization within a few hundred meters and several days of each burglary.

Self-excitation is also found in gang violence data, as an event involving rival gangs can lead to retaliatory acts of violence. In Figure 2, we plot the times of all recorded violent crimes between the gang known as “Locke Street” and the rival gang known as “Lowell Street” occurring between 2000 and 2002 in the Los Angeles police district of Hollenbeck. Here we observe clear clustering patterns suggestive of self-excitation in the rate at which

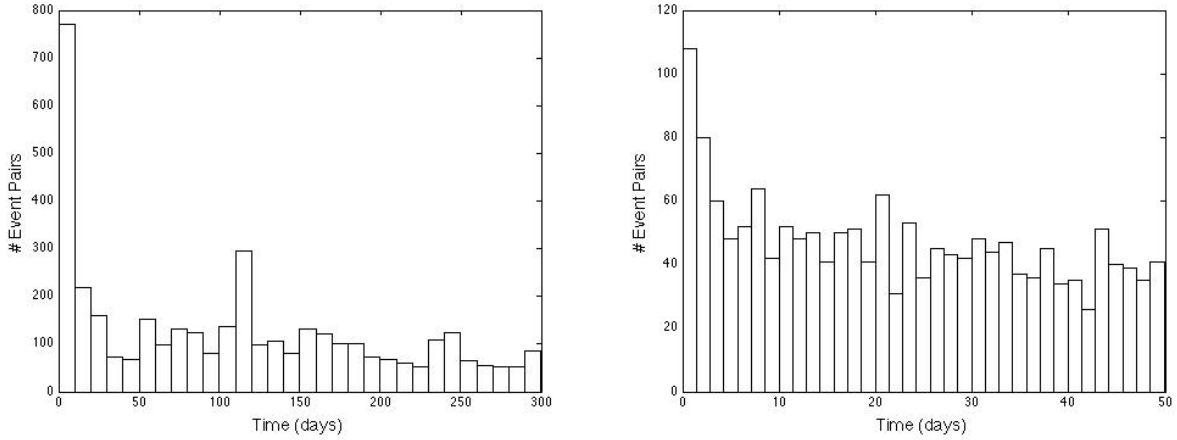


Figure 1: On the left, histogram of times (less than 300 days) between Southern California earthquake events of magnitude 3.0 or greater separated by 110 kilometers or less. On the right, histogram of times (less than 50 days) between burglary events separated by 200 meters or less.

the two rival gangs attack each other.

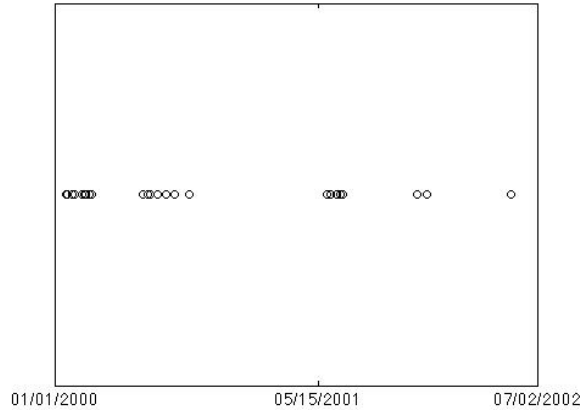


Figure 2: Times of violent crimes between two rivalry gangs in Los Angeles.

We propose that self-exciting point processes can be adapted for the purpose of crime modeling and are well suited to capture the spatial-temporal clustering patterns observed in crime data. More specifically, spatial heterogeneity in crime rates can be treated using background intensity estimation and the self-exciting effects detected in crime data can be modeled with a variety of kernels developed for seismological applications or using non-

parametric methods. In Section 2, we review self-exciting point processes in the context of seismological modeling. In Section 3, we present a model for residential burglary based upon non-parametric methods for Epidemic Type Aftershock-Sequences models of earthquakes. Our methodology combines the idea of stochastic declustering with Kernel Density Estimation in a novel way. In Section 4, we compare the predictive accuracy of our methodology with prospective crime hotspot maps. The results illustrate how crime hotspot maps can be improved using the self-exciting point process framework. We validate the methodology with a simulated point process in the Appendix.

2 Self-exciting point process models in seismology

A space-time point process $N(t, x, y)$ is typically characterized via its *conditional intensity* $\lambda(t, x, y)$, which may be defined as the limiting expected rate of the accumulation of points around a particular spatial-temporal location, given the history \mathcal{H}_t of all points up to time t (Daley and Vere-Jones, 2003):

$$\lambda(t, x, y) = \lim_{\Delta t, \Delta x, \Delta y \downarrow 0} \frac{E[N\{(t, t + \Delta t) \times (x, x + \Delta x) \times (y, y + \Delta y)\} | \mathcal{H}_t]}{\Delta t \Delta x \Delta y}. \quad (1)$$

In Seismology a *mark* M_k , the magnitude of the earthquake, is associated with each event (t_k, x_k, y_k) and the conditional intensity often takes the form,

$$\lambda(t, x, y, M) = j(M)\lambda(t, x, y), \quad (2)$$

$$\lambda(t, x, y) = \mu(x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k; M_k). \quad (3)$$

Models of this type, referred to as Epidemic Type Aftershock-Sequences (ETAS) models, work by dividing earthquakes into two categories, *background* events and *aftershock* events.

Background events occur independently according to a stationary Poisson process $\mu(x, y)$, with magnitudes distributed independently of μ according to $j(M)$. Each of these earthquakes then elevates the risk of aftershocks and the elevated risk spreads in space and time according to the kernel $g(t, x, y, M)$.

Many forms for g have been proposed in the literature, though in general the kernel is chosen such that the elevated risk increases with earthquake magnitude and decreases in space and time away from each event. For example, the isotropic kernel,

$$g(t, x, y; M) = \frac{K_0}{(t + c)^p} \cdot \frac{e^{\alpha(M - M_0)}}{(x^2 + y^2 + d)^q}, \quad (4)$$

is one of a variety of kernels reviewed in (Ogata, 1998). Here K_0 , M_0 , and α are parameters that control the number of aftershocks, c and d are parameters that control the behavior of the kernel at the origin, and p and q are parameters that give the (power law) rate of decay of g .

Standard models for the background intensity $\mu(x, y)$ include spline, kernel smoothing, and Voronoi estimation (Ogata and Katsura, 1988; Silverman, 1986; Okabe et al., 2000). In the case of fixed bandwidth kernel smoothing, the background intensity is estimated by,

$$\mu(x, y) = \bar{\mu} \cdot \sum_k u(x - x_k, y - y_k; \sigma), \quad (5)$$

where $\bar{\mu}$ is a parameter controlling the overall background rate. The events (t_k, x_k, y_k, M_k) are assumed to be background events and in practice can be obtained through a declustering algorithm (Zhuang, Ogata, and Vere-Jones, 2002).

The appropriate selection of parameter values is as critical to the modeling process as specifying accurate forms for μ , g , and j . The distance in space and time over which the risk spreads, the percentage of background events vs. aftershocks, the dependence of the

increased risk on magnitude size, etc. all can have a great impact on the predictive power of a point process model. Parameter selection for ETAS models is most commonly accomplished through maximum likelihood estimation, where the log-likelihood function (Daley and Vere-Jones, 2003),

$$l(\theta) = \sum_k \log \left\{ \lambda(t_k, x_k, y_k; \theta) \right\} - \int_0^T \iint_S \lambda(t, x, y; \theta) dy dx dt, \quad (6)$$

is maximized over all parameter sets θ . Here $S \times [0, T]$ is the space-time window of observation.

More recently, non-parametric methods have been introduced for self-exciting point process estimation (Zhuang, 2006; Marsan and Lenglin, 2008). Consider space-time point data $\{(t_k, x_k, y_k)\}_{k=1}^N$ and a general self-exciting point process model of the form,

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k). \quad (7)$$

Assuming model correctness, the probability that event i is a background event, p_{ii} , is given by,

$$p_{ii} = \frac{\mu(t_i, x_i, y_i)}{\lambda(t_i, x_i, y_i)}, \quad (8)$$

and the probability that event j triggered event i , p_{ji} , is given by,

$$p_{ji} = \frac{g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)}, \quad (9)$$

(Zhuang, Ogata, and Vere-Jones, 2002). Let P denote the matrix with entries p_{ji} (note that the columns sum to one). Then stochastic declustering can be used in the following way. Given an initial guess P_0 of the matrix P , we then have $N(N+1)/2$ probabilistic data points $\{(t_k, x_k, y_k, p_{kk})\}_{k=1}^N$ and $\{(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})\}_{i>j}$. Given this data, a non-parametric

density estimation procedure can be used to estimate μ from $\{(t_k, x_k, y_k, p_{kk})\}_{k=1}^N$ and g from $\{(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})\}_{i>j}$, providing estimates μ_0 and g_0 . We can then proceed iteratively as follows until convergence is achieved,

Step 1) Estimate μ_n and g_n from P_{n-1} .

Step 2) Update P_n from μ_n and g_n using (8) and (9).

For example, a simple histogram estimator is used in (Marsan and Lenglin, 2008) in step 1.

3 A self-exciting point process model of burglary

For the purpose of modeling burglary we consider an unmarked self-exciting model for the conditional intensity of the form,

$$\lambda(t, x, y) = \nu(t)\mu(x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k). \quad (10)$$

Here we neglect spatially-localized temporal fluctuations in the background rate and assume that the fluctuations occur globally (for example due to weather, seasonality, time of day, etc.) In the case of seismology, research over a number of decades was needed to refine the (parametric) form of the triggering function g . For this reason, non-parametric methods are appealing in the context of crime in order to quickly gain insight into the forms of ν , μ and g . For this purpose we use the iterative procedure outlined in the previous section to estimate the model, with several modifications.

Because the probabilistic data $\{(t_k, x_k, y_k, p_{kk})\}_{k=1}^N$ and $\{(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})\}_{i>j}$ is both 3-dimensional and the number of data points is $O(N^2)$ (where N is typically $O(1000)$ for earthquake and crime data sets), the estimation step for μ and g is computationally

expensive. The dimensionality prevents straight forward use of binning methods such as the Average Shifted Histogram (Marsan and Lenglin use a logarithmically scaled histogram on a coarse grid), as many bins may have extremely small, but non-zero, values (since the data is probabilistic, the count in each bin can be less than 1). Alternatively, the large size of the data set prevents efficient use of off-grid methods such as Kernel Density Estimation. To get around these issues we use the following Monte-Carlo based iterative procedure,

Step 1) Sample background events $\{(t_i^b, x_i^b, y_i^b)\}_{i=1}^{N_b}$ and offspring/parent inter-point distances $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ from P_{n-1} .

Step 2) Estimate ν_n , μ_n and g_n from the sampled data using variable bandwidth Kernel Density Estimation.

Step 3) Update P_n from ν_n , μ_n and g_n using (8) and (9).

Because $N_b + N_o = N$, the size of the sampled data at each iteration allows for the use of Kernel Density Estimation. Another issue is that the number of background and offspring events, N_b and N_o , is changing at each iteration. Thus a fixed bandwidth for any density estimation technique (kernel smoothing, histogram, etc.) will over smooth at some iterations and under smooth at others. Therefore we employ variable bandwidth KDE (alternatively Cross Validation could be used). We give further details of our approach and provide validation using a simulated point process in the Appendix.

Results

We fit the model given by Equation (10) to a dataset collected by the Los Angeles Police Department of 5376 reported residential burglaries in an 18km by 18km region of the San Fernando Valley in Los Angeles occurring during the years 2004 and 2005. Each burglary is

associated with a reported time window over which it could have occurred, often a few hour span (for instance, the time span over which a victim was at work), and we define the time of burglary to be the midpoint of each burglary window.

In Figure 3, we plot the sampled inter-point distances $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ on the 75th iteration of the stochastic declustering algorithm (see Appendix for convergence verification). The number of sampled (offspring) events is 706 (13.1% of all events) and of these events approximately 63% are exact-repeats (occurring at the same house). On the left, the spatial inter-point distances are plotted showing that elevated crime risk travels around $50m - 100m$ from the house of an initial burglary to the location of direct offspring events. As discussed in (Marsan and Lenglin, 2008), the overall distance near-repeat risk travels is several times further due to the cascading property of self-exciting point processes. Note also that the risk travels vertically and horizontally (along streets), more so than it does in other directions. On the right, we plot the spatial (x-coordinate) inter-point distances against the time inter-point distances. Here exact-repeat burglaries, those occurring at the same house, are apparent along the x-axis.

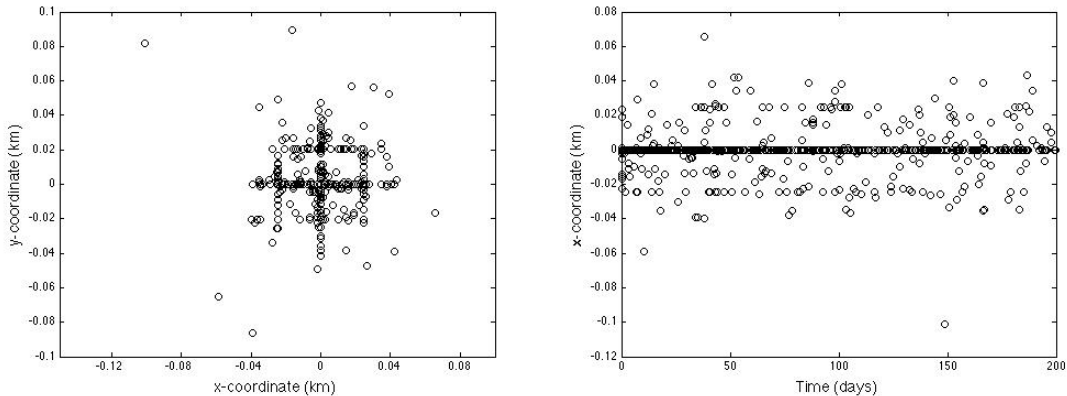


Figure 3: Spatial (left) and space-time (right) offspring/parent inter-point distances $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ sampled from P_{75} .

In Figure 4, we plot (on a logarithmic scale) the estimated marginals

$$g_{75}(t) = \int_x \int_y g_{75}(t, x, y) dx dy$$

and

$$g_{75}(x) = \int_t \int_y g_{75}(t, x, y) dy dt$$

computed from the KDE estimate of g at the 75th iteration. Here the presence of exact-repeat events can again be seen, as $g_{75}(x)$ appears to approximate a delta distribution at the origin. The spike around 1-2 days in the plot of $g_{75}(t)$ is due to the presence of fast “crime sprees,” where most likely the same burglar visited several neighboring houses within a time span of a few minutes to several days. There are also several bumps in the elevated risk of burglary, for example around 7 days. Here one possible explanation is that the routine of the burglar and/or the victim is such that a particular day of the week is a preferable time to commit the burglary. After 7-10 days, the elevated risk of repeat/near-repeat victimization drops to an intermediate level and stays relatively flat for a time span on the order of several hundred days before decaying back to baseline rates. These results are consistent with previous quantitative studies of exact-repeat burglaries (Short et al, 2009).

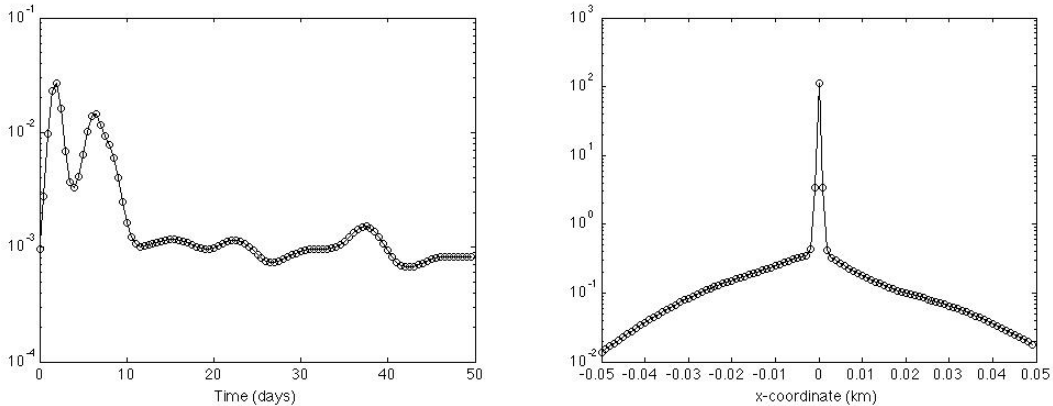


Figure 4: Marginal $g_{75}(t)$ (left) and marginal $g_{75}(x)$ (right) estimated using KDE based upon offspring/parent inter-point distances sampled from P_{75} .

In (Johnson et al, 2008), the authors discuss the need to model risk heterogeneity and in general it is a difficult task to separate clustering due to background heterogeneity and clustering due to self-excitation. One benefit of using the non-parametric approach outlined above is that temporal and spatial changes in the rate of crime are automatically separated into those stemming from exogenous effects and those due to self-excitation. In Figure 5, we plot the estimated marginals $\nu_{75}(t)$ and $\mu_{75}(x, y)$ estimated using KDE from $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ at the 75th iteration. Here the estimated background rate exhibits temporal fluctuations on a time scale of months/days, separate from the fluctuations due to self-excitation. These fluctuations are likely caused by a number of factors such as seasonal, economic, and demographic changes, as well as temporal variations in burglar routine activities (Felson, 1998). For example, residential burglary tends to have a higher weekday rate (when victims are at work) compared to weekends.

Similarly, the background rate is also spatially variable, which is consistent with fixed environmental heterogeneity in crime opportunities, as well as variability in population density through space (Bernasco et al, 2005). In seismology, declustered catalogs are of great interest as they can be used in estimating the background rate of major earthquakes. Declustered crime catalogs could potentially be used by police to distinguish between areas of a city with intrinsically high crime rates and areas with temporarily high crime rates (due to near-repeat effects). As the former arises due to structural properties of a given neighborhood and the latter from behavioral characteristics of individual burglars, police and community responses would likely need to be different in each case.

4 Crime forecasting: point processes vs. hotspot maps

Crime hotspot maps are a well established tool for visualization of space-time crime patterns and can be used as a method for prediction of near-repeat crimes. Given space-time crime

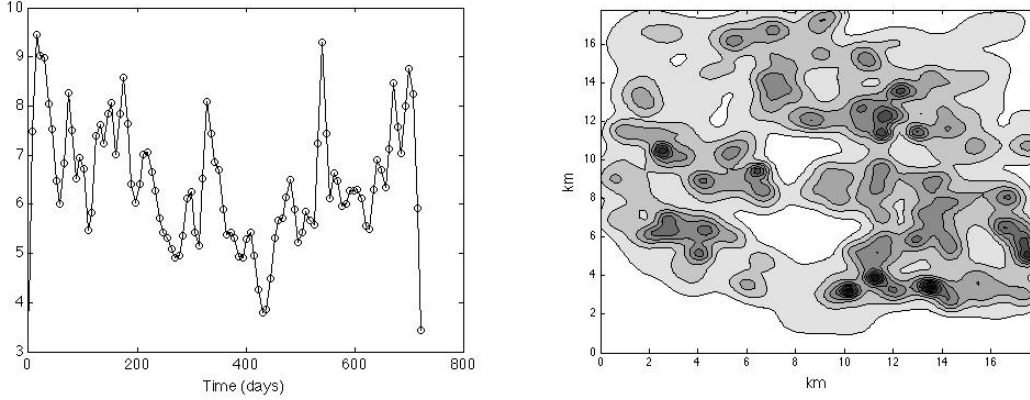


Figure 5: Background rate time marginal $\nu_{75}(t)$ (left) and space marginal $\mu_{75}(x, y)$ (right) estimated using KDE from the background events sampled from P_{75} .

observations (t_k, x_k, y_k) , crime hotspot maps are generated for a time interval $[t - T, t]$ by overlaying a density plot of the function,

$$\lambda(t, x, y) = \sum_{t-T < t_k < t} g(t - t_k, x - x_k, y - y_k), \quad (11)$$

onto a city map, where $g(t, x, y)$ is a space-time kernel. By flagging the areas of the city where λ takes on its highest values, crime hotspot maps can be used to indicate which areas in the city are likely to contain future crimes (Bowers et al., 2004; Chainey et al., 2008).

For example, in (Bowers et al., 2004) a coarse grained kernel is used that decays inversely proportional to spatial and temporal distance. In particular, with spatial distance d in units of 1/2 cell widths and time t in units of weeks, the kernel in (11) is specified as

$$g(t, d) = \frac{1}{(1+t)(1+d)} \quad (12)$$

on the domain $(t, d) \in [0, 2 \text{ months}] \times [0, 400 \text{ meters}]$ and 0 otherwise. Such a crime hotspot map is referred to as “prospective”, as it uses past crimes, coupled with the contagious spread of crime (modeled by g), to estimate future relative crime risk across the city. It should be

noted that the risk is relative because (11) is not a point process intensity.

Here we compare the predictive accuracy of the self-exciting point process model developed in Section 3 to the prospective crime hotspot map given by (11)-(12). Because crime is localized in small regions of the city (a commercial zone with no residential burglary may be located 100 meters from a neighborhood), we find that for predictive purposes variable bandwidth KDE is less accurate than fixed bandwidth KDE. We therefore estimate $\mu(x, y)$ in Equation (10) using fixed bandwidth Gaussian KDE, with 20-fold cross validation used to select the bandwidth (Silverman, 1986).

For every day k of 2005, each model assesses the risk of burglary within each of M^2 cells partitioning an 18km by 18km region of the San Fernando Valley in Los Angeles. Based on the data from the beginning of 2004 up through day k , the N cells with the highest risk (value of λ) are flagged yielding a prediction for day $k + 1$. The percentage of crimes falling within the flagged cells on day $k + 1$ is then recorded and used to measure the accuracy of each model.

In Figure 6, on the left we plot the percentage of crimes predicted averaged over the forecasting year against the percentage of flagged cells for the self-exciting point process and the prospective hotspot strategy. For example, with 10% of the city flagged the point process and prospective hotspot correctly predict 660 and 547 crimes (respectively) out of 2627. The difference in accuracy between the two methodologies can be attributed to the crime hotspot maps failure to account for the background rate of crime. While prospective crime hotspot maps used for crime prediction attempt to quantify the contagious spread of crime following past events, they fail to assess the likelihood of future “background” events, the initial events that trigger crime clusters.

In order to disentangle the dependence of model accuracy on parameter selection, in Figure 6 on the right we repeat the same prediction exercise but with parameters of each model selected to yield the highest number of crimes predicted (L^1 norm over 1 through 15

% of cells flagged). The optimal cutoff parameters for the prospective hotspot map are 200 meters and 39 weeks. With these parameter values, in particular the slow decay of g in time, Equation (11) is closer to Poisson estimation. For the point process model we only optimize the bandwidth used for $\mu(x, y)$ as the computational cost of the stochastic declustering algorithm is relatively high. Whereas the bandwidth is estimated to be approximately 300 meters using cross validation, a smaller bandwidth, 130 meters, provides a higher level of predictive accuracy. This can be attributed to the spatially localized features of neighborhoods, and hence burglary.

For all percentages of cells flagged the prospective hotspot map underperforms the point process, though for certain percentages the relative underperformance is less. On the left in Figure 6, the prospective hotspot map performs better (relative to the point process) for smaller percentages of cells flagged, as the parameters are selected to account for near-repeat effects. On the right, the prospective hotspot map performs better for larger percentages of flagged cells, since for these parameter values the model is more accurately estimating fixed environmental heterogeneity. For crime types such as robbery and auto theft, where near-repeat effects play less of a role, prospective hotspot maps tailored for near-repeat effects are likely to be outperformed by simple Poisson estimation. The advantage of models of the form (10) is that the balance between exogenous and endogenous contributions to crime rates is inferred from the data as opposed to being imposed a priori.

5 Discussion

We showed how self-exciting point processes from seismology can be used for the purpose of crime modeling. In the future it may be desirable to tailor point process models specifically for crime, taking into account the crime type and the local geography of the city. Based upon the insights provided by nonparametric estimates, parametric models can be constructed that

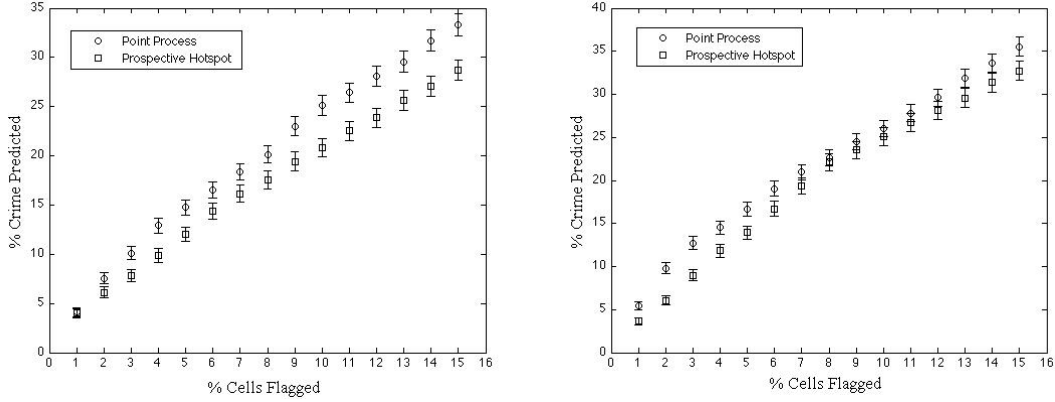


Figure 6: Forecasting strategy comparison. Average daily percentage of crimes predicted plotted against percentage of cells flagged for 2005 burglary using 200m by 200m cells. Error bars correspond to the standard error. Prospective hotspot cutoff parameters are 400 meters and 8 weeks (left) and optimal parameters (right) are 200 meters and 39 weeks. Spatial background intensity $\mu(x, y)$ smoothing bandwidth for the point process is 300 meters (left) selected by cross validation and 130 meters (right) selected to optimize the number of crimes predicted.

have advantages with respect to model fitting and simulation. Background rates can also be improved by incorporating other data types (in Johnson et al, 2008 housing density is used to improve models of repeat victimization). In the case of gang violence, a hybrid network-point process approach may be useful for capturing the self-exciting effects stemming from gang retaliations. Here increased risk may not diffuse in geographic space, but instead may travel through the network space of gang rivalry relations.

The methodology used in this study can be implemented for other applications as well, for example refining point process models of earthquakes. It could potentially be adapted, more generally, to other second order models of point processes. The stochastic declustering algorithm opens up the door to a plethora of density estimation techniques (Silverman, 1986; Scott, 1992; Eggermont and LaRiccia, 2001) that could be used to explore point processes in a way parametric methods do not allow.

In (Marsan and Lenglin, 2010) it is shown that the method is an Expectation-Maximization (EM) type algorithm. At the maximization step the complete data log-likelihood function

decouples in terms of the background and triggering functions, which is why at each iteration the problem reduces to several decoupled density estimation problems. Several issues could potentially arise here, one being that the method could converge to a local (but not global) minimum of the observed data log-likelihood function. Another, as pointed out in (Sornette and Utkin, 2009), is that the sample size and domain size (relative to the support of the triggering kernel) play a key role in the accuracy of stochastic declustering. In numerical tests we have found that at least $O(1000)$ data points are needed in 3 dimensions for the iterates to converge to the right densities and the domain needs to be several times larger than the support of the triggering kernel. Similar to analytical results for standard density estimation, it would be useful to have convergence results relating sample size, branching ratio, domain size, and the bandwidth of the density estimators to the solution of the fixed point iteration.

6 Appendix

Given point data $(t_k, x_k, y_k)_{k=1}^N$ and a self-exciting point process model of the form,

$$\lambda(t, x, y) = \nu(t)\mu(x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k), \quad (13)$$

we iterate the following until convergence:

Step 1) Sample background events $\{(t_i^b, x_i^b, y_i^b)\}_{i=1}^{N_b}$ and offspring/parent inter-point distances $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ from P_{n-1} .

Step 2) Estimate ν_n , μ_n and g_n from the sampled data.

Step 3) Update P_n from ν_n , μ_n and g_n using (8) and (9).

In order to estimate ν_n , μ_n and g_n from the sampled data, we use variable bandwidth

Kernel Density Estimation. To estimate g_n , we first scale the data $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ to have unit variance in each coordinate and based upon the rescaled data compute D_i , the k th nearest neighbor distance (3-dimensional Euclidean distance) to data point i . We then transform the data back to its original scale and, letting σ_x , σ_y , and σ_t be the sample standard deviation of each coordinate, estimate the triggering function as,

$$g_n(t, x, y) = \frac{1}{N} \sum_{i=1}^{N_o} \frac{1}{\sigma_x \sigma_y \sigma_t (2\pi)^{(3/2)} D_i^3} \exp \left(-\frac{(x - x_i^o)^2}{2\sigma_x^2 D_i^2} - \frac{(y - y_i^o)^2}{2\sigma_y^2 D_i^2} - \frac{(t - t_i^o)^2}{2\sigma_t^2 D_i^2} \right).$$

The background rate is estimated similarly, where 1-dimensional and 2-dimensional Gaussian kernels are used to estimate ν_n and μ_n respectively. In (Zhuang, Ogata, and Vere-Jones, 2002), the authors recommending using the 10th-100th nearest neighbor distance for D_i . Throughout we compute D_i corresponding to ν using the 100th nearest neighbor distance and in higher dimensions we use the 15th nearest neighbor distance for D_i corresponding to μ and g .

We validate the method by simulating (13) with

$$\nu(t)\mu(x, y) = \frac{\bar{\mu}}{(2\pi)(4.5)^2} \exp \left(-\frac{x^2}{2(4.5)^2} \right) \exp \left(-\frac{y^2}{2(4.5)^2} \right)$$

and

$$g(t, x, y) = \theta \omega \exp \left(-\omega t \right) \exp \left(-\frac{x^2}{2\sigma_x^2} \right) \exp \left(-\frac{y^2}{2\sigma_y^2} \right)$$

and comparing the estimates supplied by the method with the known distribution. The simulation was carried out by first simulating all background events according to the Poisson process $\nu\mu$. The rest of the simulation was carried out iteratively, where each point of each generation generates its own offspring according to the Poisson process g centered at the parent point. The process terminates at the n th generation when all events of the n th generation lie outside of the time window under consideration. In order to have a realization

of the point process at steady state, the first and last 2000 points were disregarded in each simulation.

In Figure (7), we plot the L2 error $\|P_n - P_{n-1}\|_2$ at the n th iteration and the number of sampled background events N_b at the n th iteration against the true number of background events for one realization of the known point process. Here we observe that the error converges quickly for the first 10 iterations and then stabilizes as the error introduced by estimating the point process through sampling P cannot be reduced further (unless a deterministic iterative procedure is employed). We also verify that the method applied to the 5376 burglary events in Section 3 reached convergence in Figure (7). Here we observe a similar rate of convergence for the crime data as with the simulated point process.

In Table 1, we list the exact parameter values used for the simulated point process and the estimates averaged over the final 10 iterations of the stochastic declustering algorithm for each of 5 simulations of the point process. The parameter values were selected to yield point patterns with scales similar to those observed in crime data. The parameter estimates are computed using the sample variances of the coordinates of $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$ and the values of N_b, N_o . As some error is due to sample variation, we plot in the last two columns the estimated number of background events vs. the actual number of background events in each of the 5 simulations to assess the ability of the method to reconstruct the realized branching structure. In Figure (8), we plot the estimated marginals of $g(t, x, y)$ against the actual distributions on the 75th iteration of the stochastic declustering algorithm. The estimated time marginal density deviates from the true density at the origin due to the jump discontinuity of the exponential distribution. However, the estimate of the parameter ω is still close to the true value (see Table 1).

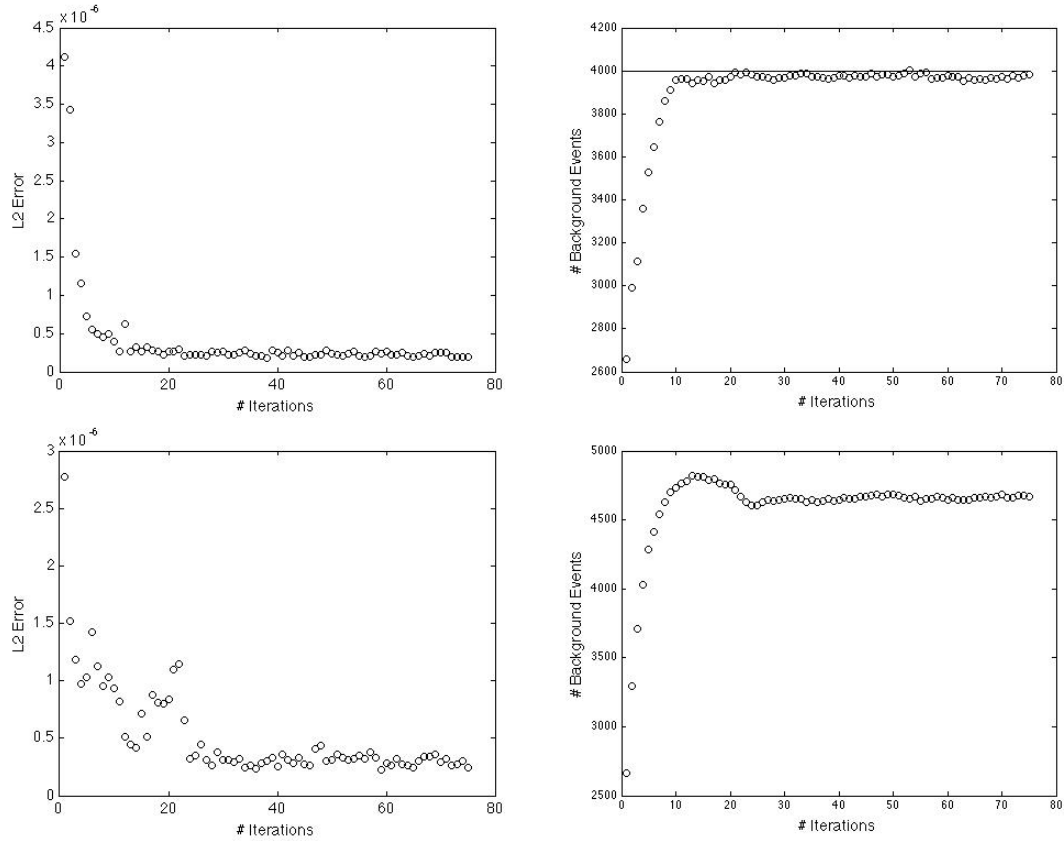


Figure 7: L2 error $\|P_n - P_{n-1}\|_2$ (top left) and N_b , the number of sampled background events, (top right) at the n th iteration for known point process model. L2 error $\|P_n - P_{n-1}\|_2$ (bottom left) and N_b , the number of sampled background events, (bottom right) at the n th iteration for the method applied to the 5376 burglary events in Section 3.

References Cited

- [1] Bernasco, W. and Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *Brit. J. Criminology* **45**, 296–315.
- [2] Bowers, K. J., Johnson, S. D., and Pease, K. (2004). Prospective Hot-Spotting: The Future of Crime Mapping? *Brit. J. Criminol.*, 44, 641–658.
- [3] Chainey, S., Tompson, L., and Uhlig S. (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime, *Security Journal*, 21, 4–28.

Table 1: Parameter Value Estimates

	w^{-1}	σ_x	σ_y	θ	$\bar{\mu}$	N_b Est.	N_b True
True Values	10.00	0.0100	0.1000	0.2000	5.7100		
Run 1 Est.	11.08	0.0176	0.1433	0.2001	5.6921	3999.7	4041
Run 2 Est.	12.20	0.0156	0.1296	0.1967	5.7768	4016.5	4026
Run 3 Est.	11.76	0.0150	0.1295	0.1997	5.6711	4001.5	4017
Run 4 Est.	13.30	0.0135	0.1407	0.2049	5.6185	3975.3	4015
Run 5 Est.	11.27	0.0147	0.1317	0.2102	5.7652	3948.9	3977

- [4] Cohen, J. and Tita, G. (1999). Spatial Diffusion in Homicide: Exploring a General Method of Detecting Spatial Diffusion Processes. *Journal of Quantitative Criminology* **15** (4), 451–493.
- [5] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, 2nd edition. New York: Springer.
- [6] Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum Penalized Likelihood Estimation: Volume I: Density Estimation*, New York: Springer.
- [7] Farrell, G. and Pease, K. (Ed.) (2001). *Repeat Victimization*, Criminal Justice Press, New York.
- [8] Felson, M. (1998). *Crime and Everyday Life*, Pine Forge Press, Thousand Oaks.
- [9] Johnson, S. D. (2008). Repeat burglary victimisation: a tale of two theories. *Journal of Experimental Criminology* **4**, 215–240.
- [10] Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., and Townsley, M. (2007). Space-time patterns of risk: A cross national assessment of residential burglary victimization, *J. Quant. Crim.*, **23**, 201-219.

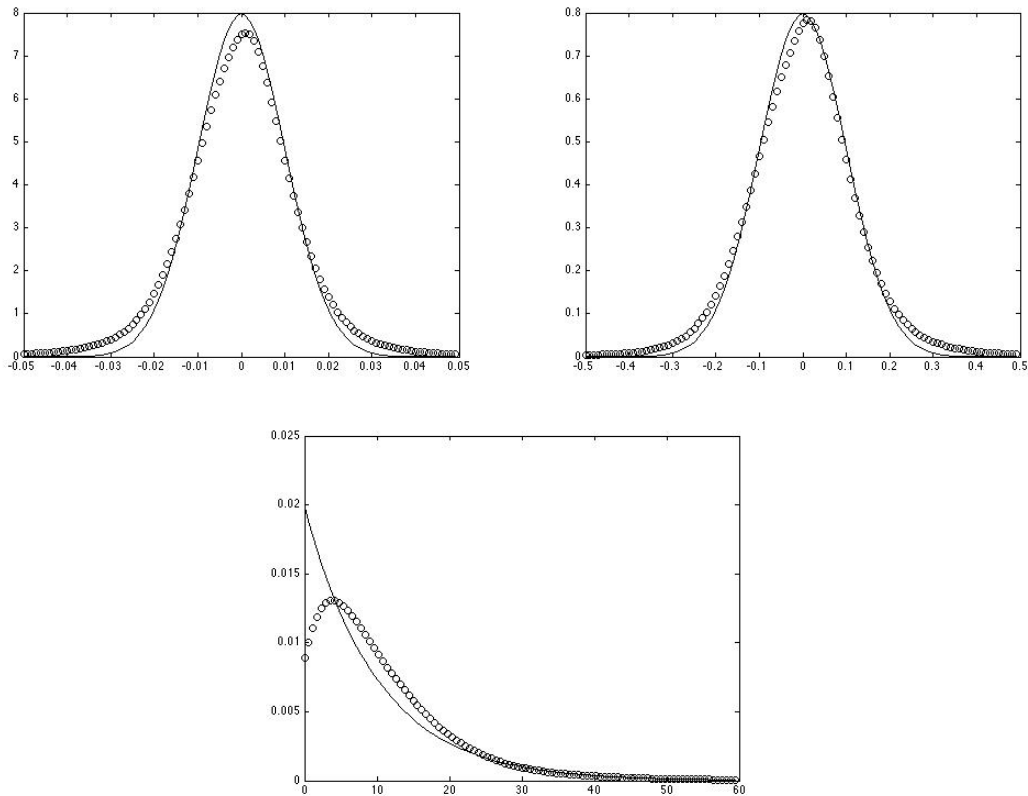


Figure 8: Estimated (circles) and actual (solid line) marginals of $g(t, x, y)$ on the 75th iteration. Upper left is the marginal $g(x)$, upper right is the marginal $g(y)$, and lower figure is marginal $g(t)$.

- [11] Marsan D. and Lenglin, O. (2008). Extending Earthquakes Reach Through Cascading, *Science*, 319: 1076–1079.
- [12] Marsan D. and Lenglin, O. (2010). A new estimation of the decay of after-shock density with distance to the mainshock. Preprint (<http://www.lgit.univ-savoie.fr/dmars/Publications.html>).
- [13] Ogata, Y. (1998). Space-time point process models for earthquake occurrences, *Ann. Inst. Statist. Math.*, 50 (2), 379–402.
- [14] Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, 40, 20–39.

- [15] Okabe, A., Boots, B., Sugihara, K. and Chiu, S. (2000). *Spatial Tessellations*, 2nd ed. Wiley, Chichester.
- [16] Scott, D. W. (1992). *Multivariate Density Estimation*, New York: Wiley.
- [17] Short, M.B. *et al.*. (2009). Measuring and Modeling Repeat and Near-Repeat Burglary Effects, *J. Quant. Criminol.* **25** (3), 325.
- [18] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- [19] Sornette, D. and Utkin, S. (2009). Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks and aftershocks, *Phys. Rev. E*, **79**, 061110.
- [20] Tita, G. and Ridgeway, G. (2007). The Impact of Gang Formation on Local Patterns of Crime. *Journal of Research on Crime and Delinquency* **44** (2), 208–237.
- [21] Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic Declustering of Space-Time Earthquake Occurrences, *JASA*, 97 (458), 369–380.
- [22] Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications to model evaluation. *Journal of the Royal Statistical Society: Series B*, 68 (4), 635–653.