RESEARCH ARTICLE

Evaluating temporally weighted kernel density methods for predicting the next event location in a series

Michael D. Porter[a*] and Brian J. Reich[b]

[a]*GeoEye Analytics, McLean, VA, USA;* [b]*Dept. of Statistics, North Carolina State University, Raleigh, NC, USA*

One aspect of tactical crime or terrorism analysis is predicting the location of the next event in a series. The objective of this article is to present a methodology to identify the optimal parameters for and test the performance of temporally weighted kernel density estimation models for predicting the next event in a criminal or terrorist event series. By placing event series in a space-time point pattern framework, the next event prediction models are shown to be based on estimating a conditional spatial density function. We use temporal weights that indicate how much influence past events have toward predicting future event locations and can also incorporate uncertainty in the event timing. Results of applying this methodology to crime series in Baltimore County, MD indicate that performance can vary greatly by crime type, little by series length, and is fairly robust to choice of bandwidth.

1. Introduction

The objective of this article is to present a methodology to identify the optimal parameters for and test the performance of temporally weighted kernel density estimation (KDE) models for predicting the next event in a criminal or terrorist event series. We describe how event series can be viewed in a space-time point pattern framework and how next event prediction models are based on estimating a conditional spatial density function. The temporal weights indicate how much influence past events have toward predicting future event locations. Uncertainty in the event times is incorporated in the temporal kernels.

---

*Corresponding author. Email: porter.mike@geoeye.com

Criminal and terrorist event point patterns often exhibit spatial and space-time clustering (Bowers and Johnson 2005, Braithwaite and Johnson 2012, Grubesic and Mack 2008, Short *et al.* 2009, Townsley *et al.* 2003, Youstin *et al.* 2011). This could be due to the event initiators' limited awareness space and routine activities (Brantingham and Brantingham 2008), their particular site selection behaviors (Johnson *et al.* 2009), or the clustering of attractive targets or victims (Brantingham and Brantingham 1995). Events that are close in both space and time are termed near-repeats and the explanations for them are a topic of much research (Bowers and Johnson 2004, Johnson 2008, Pitcher and Johnson 2011, Short *et al.* 2009). The two primary explanations for the presence of near-repeats are termed the flag account and boost account. Simply stated, the flag account explains near-repeats as a result of certain locations or regions being attractive or accessible to the general population of offenders. Such near-repeat events are caused by multiple offenders choosing their locations independently. The boost account describes near-repeats as a contagion process where one offender (or their associates) initiates several events in close space-time proximity as their awareness space and knowledge of targets grows with each successful event (e.g. due to optimal foraging (Johnson *et al.* 2009)).

Several studies of crime activity offer support to the boost account and provide evidence that that near-repeats can often be attributed to the same offender (Bernasco 2008, Bowers and Johnson 2004, Johnson *et al.* 2009). This is important for event prediction as it indicates that an individual's event series should also display space-time clustering. The concept of near-repeats has also been extended to consider clusters of events (e.g. poly-order chains and microcycles) that are close in space and time to each other (Behlendorf *et al.* 2012, Johnson and Braithwaite 2009), however these have not yet been investigated for common offender involvement.

The presence of space-time clustering and near-repeats in many criminal and terrorist event processes has prompted research into predictive models incorporating these aspects. Space-time kernel based methods have been used to weight more recent events (Bowers *et al.* 2004, Johnson *et al.* 2008) or events that occurred during certain time periods (Tompson and Townsley 2010) more heavily. By mixing background attractiveness and near-repeat behavior, self-exciting models (Johnson *et al.* 2008, Lewis *et al.* 2011, Mohler *et al.* 2011, Porter and White 2012) attempt to incorporate both flag and boost aspects into a predictive model. Rey *et al.* (2011) uses Markov models to jointly model the evolution of crime rates across discrete space and time cells allowing forecasts to be made based on recent local event activity. In line with the flag account, other models attempt to directly model the relationship between crimes and the spatial environment (Bernasco and Block 2009, Huddleston and Brown 2009, Kennedy *et al.* 2011, Liu and Brown 2003).

While there are a number of models for predicting the locations of aggregated criminal and terrorist events, there are few for predicting the location of the next event in a series. Using attacks by the ETA terrorist group, LaFree *et al.* (2012) examined how often the next event was in the same, adjacent, or non-adjacent provinces. They found evidence that the time between attacks and location of previous attacks influenced these probabilities. Correlated walk analysis (Levine 2010) attempts to model the trajectory of a crime series assuming that the offender follows a type of random walk that allows for a consistency or momentum in the current direction and speed of travel. Another alternative, and one we pursue in this paper, is to adapt the temporally weighted kernel density methods used for prospective hot-spotting (Bowers *et al.* 2004, Johnson *et al.* 2008) for next event prediction. Towards this end, Paulsen (2005, 2011) developed a temporally

weighted density procedure for next event prediction that uses a spatial bandwidth based on the mean nearest neighbor distance and event weights based on the time since the first event in the series. Paulsen (2005) found that the kernel method outperformed several other prediction methods, including the correlated walk analysis, on crime series data.

Based on the findings of Paulsen (2005), their use in prospective crime mapping, and their applicability to small data sets, we consider temporally weighted kernel density models for estimating the location of the next event in a series. We develop a framework for kernel-based next event prediction that explicitly shows how the proper choice of temporal weighting function results in an estimator for the conditional spatial density of a space-time point process. In addition, we provide a method to incorporate uncertainty of the event timing into the temporal weighting function. Because there is little guidance for selecting the space-time bandwidth parameters, we explore the effects of different bandwidths on performance across several event types and sizes in a set of crime series in a US county.

2. Kernel Based Next Event Prediction

The spatial location and timing of events in a series can be viewed as a space-time point pattern (i.e., a realization from a space-time point process). The intensity $\lambda(t, s)$ of a space-time point process represents the event rate at a specific time t and location s. The space-time intensity can be decomposed into

$$\lambda(t, s) = \mu(t) \cdot f(s|t) \tag{1}$$

where $\mu(t)$ is the overall temporal rate and $f(s|t)$ is the conditional density[1] of an event occurring in location s given that an event occurs at time t. In this framework, $\mu(t)$ controls the number and timing of events and $f(s|t)$ controls the location of any events occurring at time t.

The goal of next event prediction is to accurately identify the most probable locations (and times) of future events. More fundamentally, this implies a goal of estimating the components of intensity at future times. When focus is on the timing of events (e.g., time of day, day of week, time until next event), estimation is concerned with $\mu(t)$. Alternatively, when focus is on the location of events, then estimation is concerned with $f(s|t)$. We focus on the spatial locations of future events of a series and thus concentrate on estimation of the conditional spatial density $f(s|t)$. This is appropriate for determining where resources should be deployed at a given time t.

2.1. *Spatial KDE*

While the representation of the intensity in (1) clearly distinguishes between the spatial distribution and temporal rate, both components are conditional on a given time t. A special case arises when the spatial distribution is independent of time i.e., $f(s|t) = f(s)$. When this is assumed, the intensity is said to be *separable* (Schoenberg 2004) or lack space-time interaction (Grubesic and Mack 2008). A separable intensity is convenient as

---

[1]For a given time t, $f(s|t) \geq 0$ is a probability density function (pdf) for continuous space ($\int f(s|t)\,\mathrm{d}s = 1$) and a probability mass function (pmf) for discrete space ($\sum f(s|t) = 1$)

each component, $\mu(t)$ and $f(s)$, can be estimated separately which greatly reduces the complexity of the models.

Using kernel density estimation (KDE) to estimate the conditional density under the separability assumption is equivalent to a regular spatial density estimate (i.e., retrospective KDE) where time is not incorporated. Suppose that a current series is comprised of n events with locations $\{s_1, s_2, \ldots, s_n\}$. Ignoring time, a fixed bandwidth spatial KDE is of the form

$$\hat{f}(s) = \frac{1}{n} \sum_{i=1}^{n} K(\|s - s_i\|; \theta_s) \tag{2}$$

where $K(\cdot)$ is the kernel, $\| \cdot \|$ is a distance function (e.g., Euclidean), and $\theta_s > 0$ is the spatial bandwidth. The spatial kernel function is a zero-mean two-dimensional probability density function (e.g., bivariate normal centered at zero) and the bandwidth controls the spread of the kernel's density. The bandwidth acts as a smoothing parameter with larger bandwidths corresponding to smoother density surfaces. The quality of the resulting density estimation depends greatly on the choice of bandwidth.

While many spatial kernels are possible, we review two of the most common: Gaussian and quartic. The isotropic 2D Gaussian kernel is the product of two univariate Gaussian density functions with standard deviation of θ

$$K_g(u; \theta) = \frac{1}{2\pi\theta^2} \exp\left(-\frac{u^2}{2\theta^2}\right) \tag{3}$$

for a distance of u. Notice that the Gaussian kernel provides a strictly positive value for any distance u. This is in contrast to the popular quartic (or biweight) kernel which assigns zero values to any distances that exceed a certain range (Gatrell *et al.* 1996). Specifically, the quartic kernel can be expresses as

$$K_q(u; \theta) = \begin{cases} \frac{3}{\pi\theta^2}\left(1 - \frac{u^2}{\theta^2}\right)^2 & u \leq \theta \\ 0 & u > \theta \end{cases} \tag{4}$$

where the bandwidth for the quartic kernel is taken to be the range of the kernel's domain.

One issue in comparing the Gaussian and quartic kernels is that the bandwidth refers to different aspects of the kernel's shape. Figure 1 illustrates how the bandwidth parameters for the Gaussian and quartic kernels do not lead to similar kernel functions. Because the bandwidth of the quartic kernel dictates the support of the kernel and the bandwidth for the Gaussian kernel is its standard deviation, a scaling is needed to put the two resulting kernels on a more equivalent comparison. For example, multiplying the bandwidth of a quartic kernel by 2.623 makes it comparable to a Gaussian kernel of the original bandwidth (Härdle and Linton 1994).

2.2. *Prospective KDE with Temporal Weighting*

The purely spatial KDE given in (2) weights all past events equally by $1/n$. Thus, the location of the next event in the series is just as likely to be close to the first event as it is
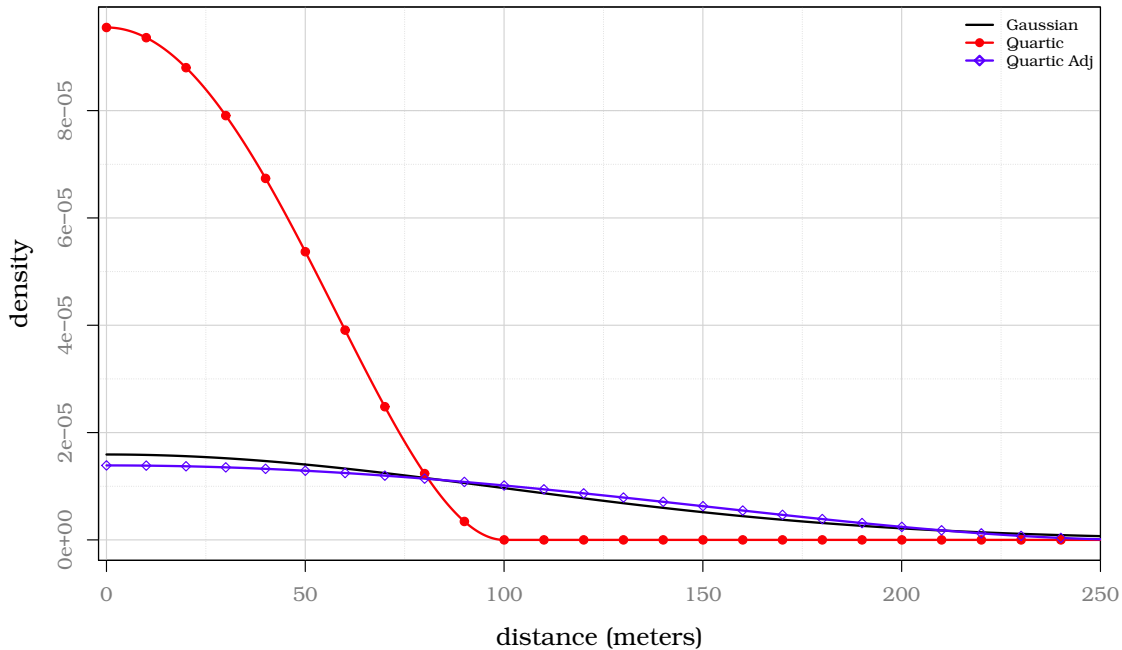
Figure 1. Spatial kernel functions. The black line corresponds to the 2D Gaussian kernel (3) with $\theta = 100$, the red line is the quartic kernel (4) with $\theta = 100$, and the blue line is the quartic with adjusted bandwidth of $\theta = 262$.

to the most recent. This implies that the event initiator does not adapt or modify their site selection behavior over time. Alternatively, an initiator exhibiting foraging behavior may select sites closer to their more recent events (Johnson *et al.* 2009, Johnson and Bowers 2004).

To allow for the possibility of drift in the series locations, a prospective conditional density estimate can be created using temporal weights that are a function of the time since the previous events. The temporally weighted KDE can be expressed

$$\hat{f}(s|t) = \sum_{i=1}^{n_t} w_i(t) \cdot K(\|s - s_i\|; \theta_s) \tag{5}$$

where n_t are the number of historic events available at time t and $w_i(t)$ is the weight of the ith event at time t. To ensure that (5) is a proper spatial density, the weights must be non-negative and add up to one for all t (i.e. $w_i(t) \geq 0$ and $\sum_{i=1}^{n_t} w_i(t) = 1 \; \forall t$).

When the weights are all equal (i.e. $w_i(t) = 1/n_t$), (5) reduces to (2) and the time since the previous events has no influence on future event locations. Alternatively, weights that decrease as a function of the time since prior events imply that the event initiator is more likely to return to the regions around recent events (or has a short memory).

We consider weights to be rescaled density functions. For example, we consider weights $w_i(t) = g(t-t_i; \theta_t)/\sum_{j=1}^{n_t} g(t-t_j; \theta_t)$ where $g(u; \theta_t)$ is a proper univariate density function with non-negative support (e.g. exponential, gamma) and temporal bandwidth parameter $\theta_t > 0$ such that larger values correspond to more equal weightings. This is equivalent to a regular space-time KDE, but with rescaled and one-sided (predictive) temporal kernels.

While not well established for next event prediction, models in the form of (5) are not new to spatial event forecasting. Bowers *et al.* (2004) and Johnson *et al.* (2008) used an

inverse distance weighting for both the temporal and spatial kernels:

$$w_i(t) \propto \frac{1}{1 + (t - t_i)} \tag{6}$$

$$K(u) \propto \frac{1}{1 + \|s - s_i\|} \tag{7}$$

Although the resulting spatial density is not proper in the sense that it does not integrate to one over the spatial domain, this does not change the resulting rank of a location's density compared to all other locations. Thus, multiplying by a constant will restore these estimates to a proper spatial density function. Equation (5) is also proportional to the self-exciting component of Mohler *et al.* (2011).

2.3. *Temporal Uncertainty*

An additional challenge with crime and terrorism data is that the actual event time is often unknown and all that is available is a time window when the event could have occurred. In the spirit of aoristic analysis (Ratcliffe 2000, 2002), we can calculate the weight as an expected density assuming that the true event time is uniformly distributed within the time window. Assuming the time window of event i is $[\alpha_i, \beta_i]$ and using the temporal kernel $g(t - t_i)$, the weight is given by

$$w_i(t) \propto \mathbb{E}[g(t - t_i)] \tag{8}$$

$$= \frac{1}{\beta_i - \alpha_i} \int_{\alpha_i}^{\beta_i} g(t - u)\, \mathrm{d}u \tag{9}$$

$$= \frac{1}{\beta_i - \alpha_i} [G(t - \alpha_i) - G(t - \beta_i)] \tag{10}$$

where $G(u)$ is the cumulative distribution function (cdf) of the temporal kernel $g(u)$. This is very simple to calculate when there is a closed form available for the cdf.

3. Data and Methodology

In order to investigate the impact of the KDE parameters on next event prediction, an evaluation of crime series data was carried out. For each crime series, a time weighted KDE (5) is used to predict future events using a range of spatial and temporal bandwidths. The quality of the predictions is based on how much area needs to be monitored to detect the next event. This allows comparison by offender, crime type, and series length. All of the calculations were performed using the freely available statistical programming language R (R Development Core Team 2011).

3.1. *Study Area*

The study region is Baltimore County, Maryland in the US and a surrounding buffer region. Baltimore County has a land area of 598.30 square miles (1549.5 km^2), a 2010 population of 805,029, providing a population density of 1345.5 persons per square mile,

an average of 561.0 housing units per square mile, and an average of 2.47 persons per household (U.S. Census Bureau 2011).

3.2. *Crime Data*

The Baltimore County Police Department provided crime data consisting of detected crimes with geocoded locations, the time interval in which the crime could have occurred[1], crime type, and an anonymized offender identifier. The data came from the Regional Crime Analysis Program (RCAP) which facilitates the sharing of crime data across jurisdictional boundaries. As such, some crime events were located outside of Baltimore County, but as they were suspected to be part of a series occurring inside the borders of the county, they were included in the data set and subsequently used in our analysis. This diminishes the impact of edge effects (i.e. not observing events that occur outside of the study region) which could impact the results for the crime series that occur close to the county border. The RCAP data consists primarily of breaking and entering, robbery, and motor vehicle theft crimes.

Seven hundred and forty two initial crime series (attributed to a common offender) were extracted for analysis. Each series consisted of four or more crimes that were committed between 1999 and 2011. Crimes with invalid geocoded locations or time intervals were excluded from the series[2]. In addition, only the crimes that occurred within 365 days from the last recorded event were included in analysis. The period was limited to one year to prohibit long stretches between crimes and guard against offenders changing anchor points during the period of analysis.

By using the crime series as the unit of analysis, crimes with more than one offender can appear in multiple series. To guard against biasing the results by including series that are too similar, series that shared over 50% of their crimes with another series where removed from analysis. This eliminated 152 series and provided 590 series for analysis. Figure 2 shows Baltimore County with the location of the 4151 crime events from the 590 crime series analyzed.

3.3. *Predictive Kernel Density Estimation*

Estimating the conditional density $\hat{f}(s|t)$ using the prospective kernel of (5) requires a choice of spatial kernel and temporal weighting function along with corresponding bandwidth parameters. The 2D isotropic Gaussian density function, given in (3), was used for the spatial kernel. The spatial bandwidth (i.e. standard deviation) ranged from 100 meters to 2 kilometers (i.e. $\theta_s \in \{100, 200, 400, 600, 1000, 2000\}$). If a small spatial bandwidth predicts well, it implies that an offender chooses their next crime site in close spatial proximity to their previous site selections. Large bandwidths, contrarily, will perform best for offenders that have a large target range, or rarely return to locations close to previously selected sites.

An exponential density function was used for the temporal weighting function. The temporal bandwidth (i.e. mean of the exponential distribution) ranged from 3 to 100 days. Also included are the special cases of giving weight only to the most recent event

---

[1]Because many crimes types (e.g. breaking and entering, motor vehicle theft) occur when the victim is not present, the exact timing of the crime cannot be ascertained. For such crimes, victims indicate the likely interval of time during which the crime could have occurred.

[2]About 11% of all crimes during this period lacked a valid location or time interval.
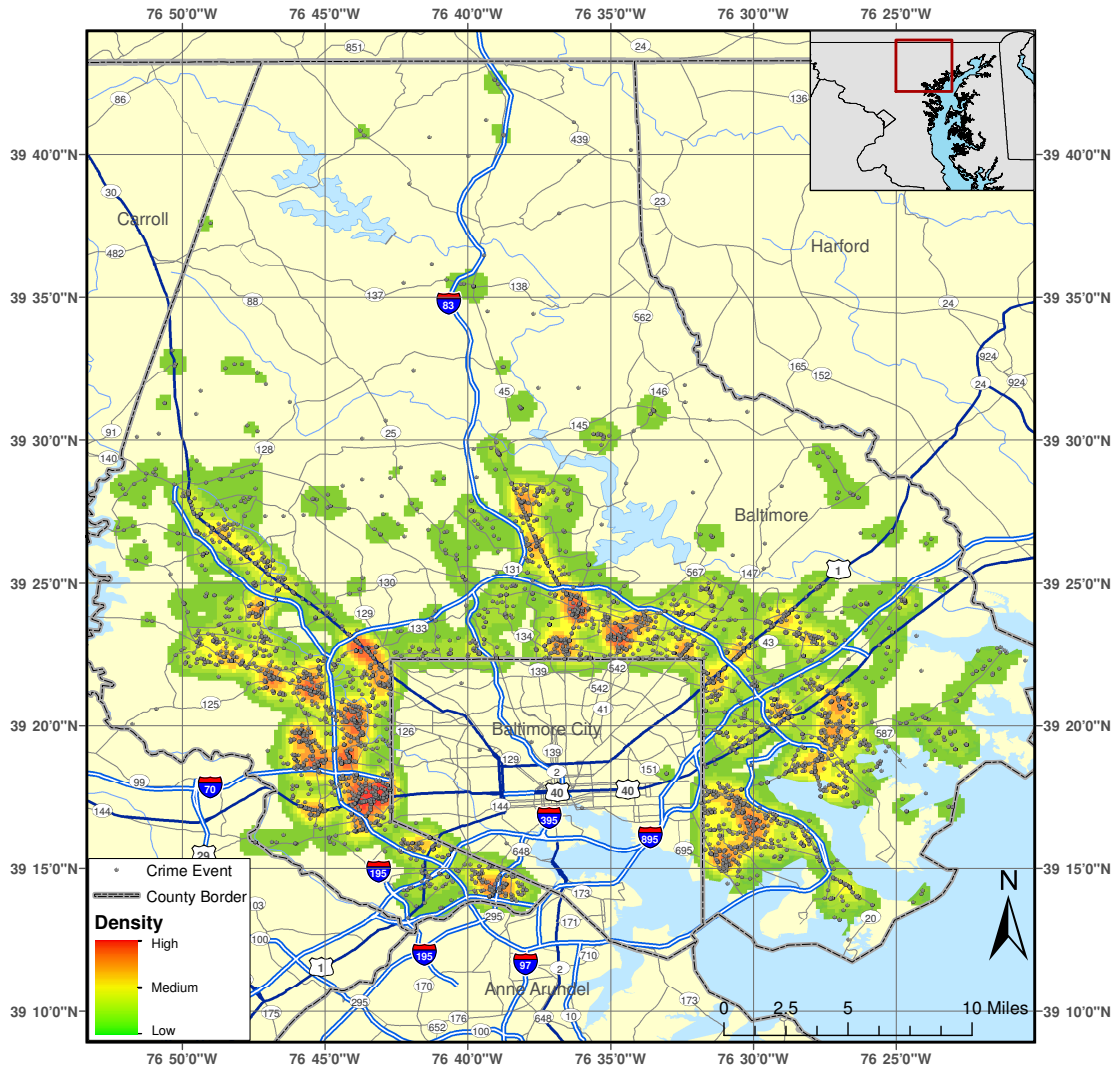
Figure 2. Map of Baltimore County, MD with the location of crime events from the 590 analyzed crime series from 1999–2011. The density layer was made using kernel density estimation with a Guassian spatial kernel with bandwidth of 400 meters. Baltimore County surrounds, but does not include, Baltimore City.

(the limiting behavior as the bandwidth goes to 0) and to all previous events equally (the limiting behavior as the bandwidth goes to ∞). Specifically, we considered temporal bandwidths $\theta_t \in \{0, 3, 7, 14, 21, 28, 35, 60, 100, \infty\}$. Weighting only the most recent event[3] represents a type of random walk where the offender only retains knowledge of the last location. Alternatively, the infinite bandwidth causes all previous weights to be equal (i.e. $1/n$) and the fully retrospective KDE of (2) is used to predict the next event location. This assumes space-time separability and would be appropriate if the offender retained full knowledge of all previous event locations and was equally likely of returning to the

---

[3]This may be multiple events if they share the same time interval.

region surrounding any of them.

The Gaussian kernel was selected due to its popularity as well as its property of having infinite support[1]. This is an advantage over the quartic kernel, which has limited support, since all locations (even those far from previous events) receive a non-zero score. This prevents ties (of zero values) in the estimated density which can complicate evaluation. The exponential density was selected as the temporal kernel due to its simplicity, familiarity, and to emphasize that it is a one-sided (predictive) kernel. In general, the choice of kernel function is not as important as the choice of bandwidth (Gatrell *et al.* 1996, Scott 1992) and so we expect that the use of other kernel functions would have little impact on the results. The bandwidth values were selected to span the reasonable range of possible values.

3.4.   *Predictive Evaluation*

Each crime series contains n recorded crimes at spatial locations $\{s_1, s_2, \ldots, s_n\}$ with time intervals $\{[\alpha_1, \beta_1], [\alpha_2, \beta_2], \ldots, [\alpha_n, \beta_n]\}$, where α_i is the earliest possible time and β_i is the latest possible time of crime i. The indices are ordered by α_i. A prediction of the kth event location is made for k ranging from 4 to n to explore the effects of more information on prediction accuracy and optimal parameters.

Because the exact event times are uncertain, we predict the conditional density for event k using the earliest time α_k. It is important to only include historic information that would be available to analysts in the predictive model. Thus, for predicting the location of event k, only the events that occurred (i.e. had the latest possible time) at least one day prior to event k were included. Thus, only the events $\{i : \alpha_k - \beta_i \geq 1 \text{ day}\}$ were used for predicting event k. This allows 24 hours from the latest possible event time for the crime to be reported, identified as part of the crime series, and used for prediction. This is an important consideration as multi-event days are often comprised of crime sprees sharing very close spatio-temporal proximity (often same location, same time) that could artificially lower the optimal kernel bandwidths.

A fine regular grid with 50 meter spacing[2] was overlaid on the study area (extending up to 4 km beyond the county border) and the density estimated at the grid points only. Thus, for a given spatial bandwidth θ_s and temporal bandwidth θ_t, $\hat{f}(s|\alpha_k)$ is the prospective kernel estimate of the conditional density at grid cell s for the kth event. Letting s_k denote the closest grid point to event k and $\mathcal{S}$ the spatial grid points, the required monitoring region to capture event k is the set of grid cells that have a higher predicted density than the cell that contains event k i.e.,

$$M_k = \{s \in \mathcal{S} : \hat{f}(s|\alpha_k) \geq \hat{f}(s_k|\alpha_k)\}$$

The area of the required monitoring region (in square kilometers) is the evaluation criterion. This is termed the search efficiency in Bowers *et al.* (2004) and corresponds to how much spatial area must be covered to detect the kth event. Small monitoring regions correspond to accurate predictions.

---

[1]The support of a kernel is the set of input values that receives a non-zero output density.
[2]The grid was created using Euclidean distances in a UTM projection for zone 18.

3.5. *Illustration*

To illustrate how the temporally weighted kernel density method is used for next event prediction, consider the example provided in Figure 3 that has 5 prior events. The series data is given in Table 1. Figure 3(a) shows the density surface for a 200 m spatial bandwidth and ∞ temporal bandwidth (i.e. all prior events are weighted equally). This bandwidth pair is optimal (i.e. it minimizes the required monitoring region) for this series. Figure 3(b) reduces the temporal bandwidth to 14 days. Notice that only the most recent two events (4 and 5) get significant weight. Figure 3(c) reduces the temporal bandwidth further to only 3 days. This causes event 5 to have the majority of the weight and provide the strongest contribution to the spatial density estimate. If the temporal bandwidth was further reduced to 0, all the weight would be with event 5 driving the spatial density estimate to be a 2D Gaussian centered at event 5. Figure 3(d) leaves the temporal bandwidth at 3 days, but increases the spatial bandwidth to 400 meters. This produces a very smooth density surface with only one peak. For this series, the parameters in 3(a) and 3(b) produce a 5 km^2 monitoring region that contain the next event, while the parameters in 3(c) and 3(d) would require a larger monitoring region to contain the next event in the series.

Event	Longitude	Latitude	α	β
1	-76.519	39.260	0.00	2.19
2	-76.516	39.255	0.25	1.12
3	-76.521	39.255	3.00	4.84
4	-76.521	39.254	32.25	32.50
5	-76.520	39.260	36.92	36.92
6	-76.517	39.254	40.88	40.88

Table 1. Example series data. Events 1–5 are used to predict the location for event 6. The earliest (α) and latest (β) possible event times are in number of days from the first event.

4. Results

We partitioned the 590 crime series by crime type. Because about 28% of the crime series included multiple crime types, we classified a crime series to be of a certain type if the fraction of crimes of that type in the series exceeded 75%. Otherwise, the crime series was classified as mixed. Furthermore, we considered analyzing the next-event predictions according to the number of prior crimes available. A short history consists of 3-5 prior events, medium 6-8, and long histories have more than 9 prior events available for prediction[1]. Table 2 shows the number of series of each crime type that contain short, medium, and long histories. Notice that the columns of Table 2 do not add up to the total. This is because some series contain crimes that can be predicted from multiple history sizes. For example, for a crime series consisting of 10 crimes, all separated by at least one day, crimes 4-6 are predicted from short histories, crimes 7-9 are predicted from medium sized histories, and the 10th crime is predicted from a long history. Thus, this series will

---

[1] The choice of these values for history length was rather arbitrary. However, as Figure 5 shows, the impact of series length appears to be minimal.

(a) $\theta_s = 200$m, $\theta_t = \infty$ days

(b) $\theta_s = 200$m, $\theta_t = 14$ days

(c) $\theta_s = 200$m, $\theta_t = 3$ days

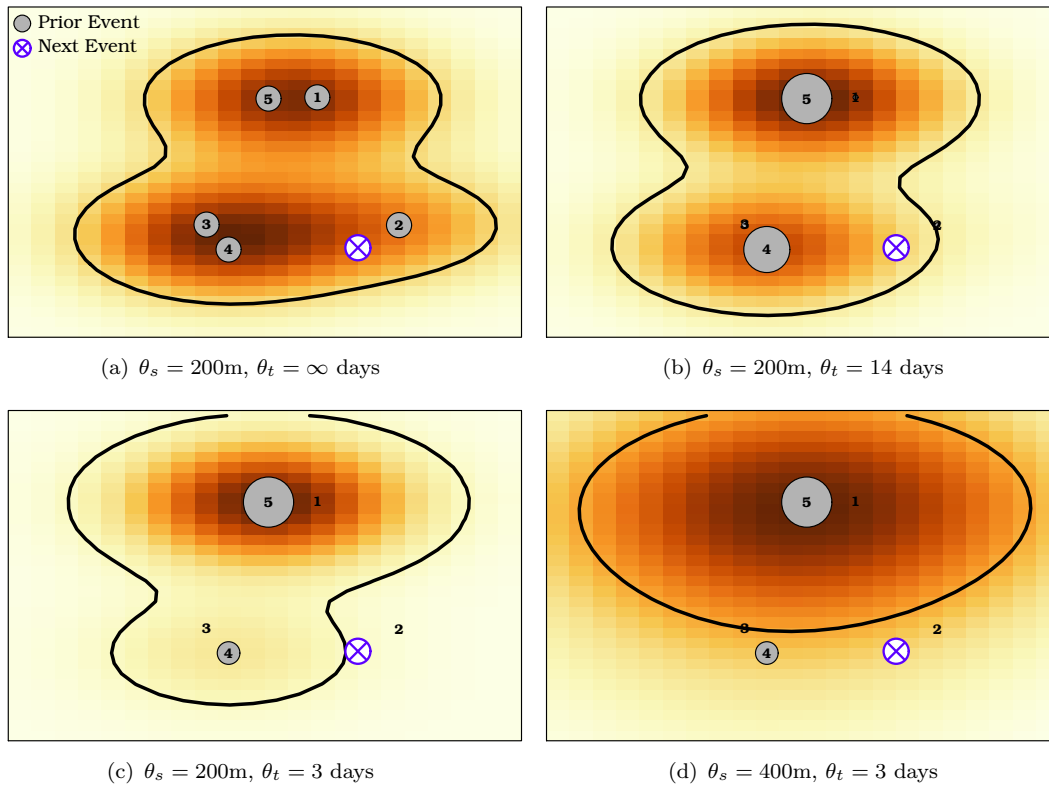(d) $\theta_s = 400$m, $\theta_t = 3$ days

Figure 3. Illustration of temporally weighted kernel density estimation for next event prediction. The size of the prior event symbol is proportional to its weight and the solid black line outlines the 5 km^2 monitoring region.

contribute to each row. In total, there were 299 breaking and entering, 194 robbery, 35 motor vehicle theft, and 62 mixed[2] crime series.

	Breaking & Entering	Robbery	Motor Vehicle Theft	Mixed	All
Short (3-5)	254	166	28	57	505
Medium (6-8)	93	58	7	11	169
Long (9+)	62	27	1	3	93
Total	299	194	35	62	590

Table 2. The number of crimes series by type and series history length. Because longer crime series may contribute to several history lengths the Total row does not equal the column sums.

To prevent long series from having too much influence on the results, we averaged the performance of each series within the history length. For example, if a series had three crimes that were predicted with a short history, then the average performance over those three crimes were used for evaluating the short history for that series. This treats each series equally so comparison can be made at the offender level rather than biased towards those with long series.

---

[2]No one crime type exceeds 75% of the series.

4.1. *Optimal Bandwidth*

First, we evaluated the performance of kernel based next event prediction using the optimal bandwidth parameters. That is, for each predicted crime we found the spatial and temporal bandwidths that would have minimized the monitoring region and recorded the size of the minimal region. While using the optimal bandwidth parameters is not possible in practice (since the next event location is not known in advance), this analysis gives an indication of the best possible performance of temporally weighted kernel methods for predicting the next event in a series. Figures 4 and 5 show the results using ROC type curves (similar to accuracy concentration curves (Johnson *et al.* 2008)). These show the probability that an event would be detected if a given size monitoring region was used. This is equivalent to what Bowers *et al.* (2004) term search efficiency rate. Because we used the optimal bandwidths for each crime, these curves represent the best possible performance of the temporally weighted KDE method[1].

Figure 4 shows the results, by crime type, for any history length. Recall that we are taking the average scores from each offender across every history length so every offender contributes equally to the performance. Notice that breaking and entering crimes have the potential to be the most accurately predicted (e.g. 69% of crimes could be detected with 10 km^2 monitored) while motor vehicle theft is the least predictable (e.g. only 37% of crimes could be detected with 10 km^2 monitored) via space-time kernel density methods.
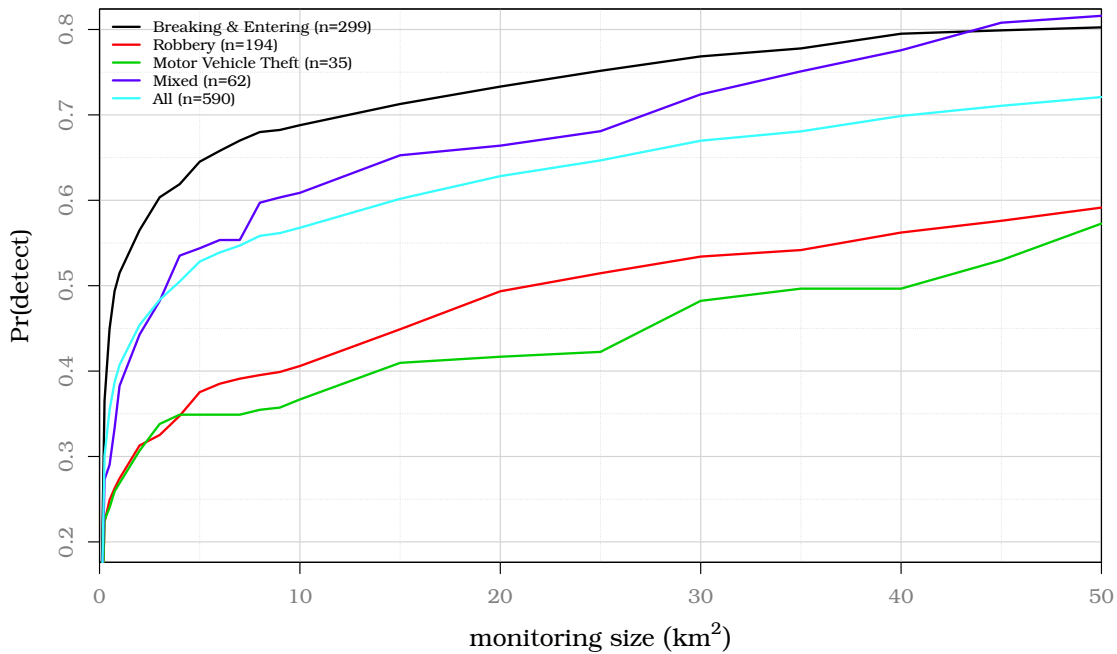
Figure 4. Optimal performance, by crime type, averaged across all history lengths.

Figure 5 divides the crime prediction capability by history length. Predicting next event locations based on a long history (≥ 9 prior events in the series) appears to allow slightly better prediction than short or medium history lengths. However, the 95%

---

[1]We considered a limited number of spatial and temporal bandwidths, so the actual optimal performance may improve slightly if a larger number of bandwidths are considered.

pointwise confidence interval[1] for Total shows that this is not a statistically significant advantage (since the confidence interval mostly contains the short, medium, and long history lengths). These results also hold when the individual crime types are considered by history length (plots not shown). This suggests that predicting future crime events when only a short history is available may not be less accurate than those that have a long history available.
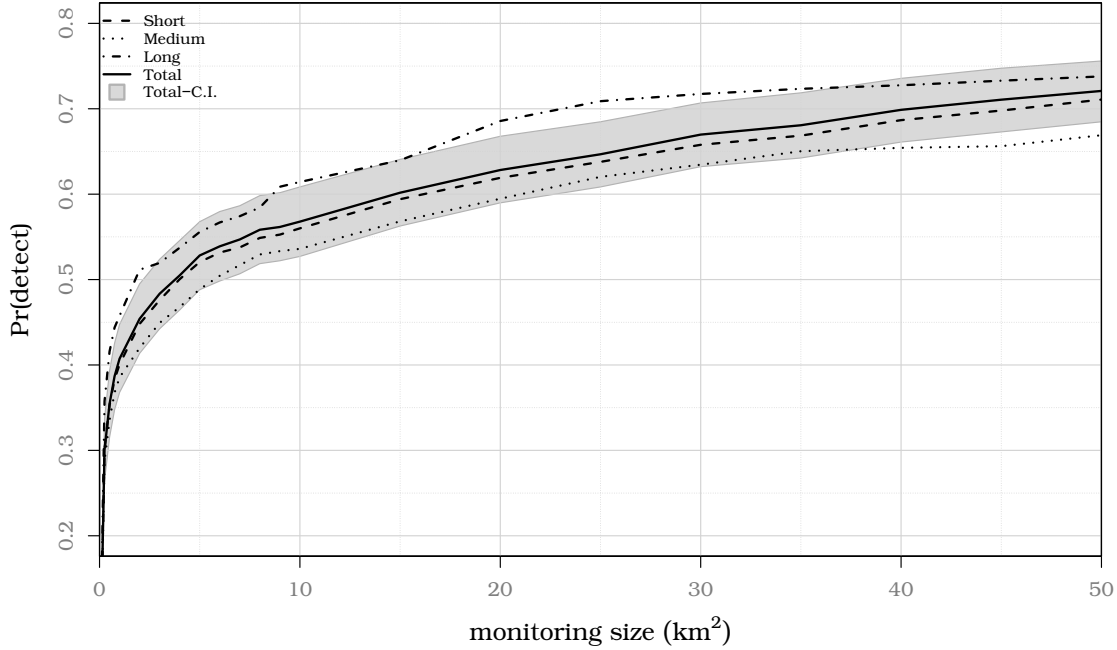

Figure 5. Optimal performance for all crime types by history length. The 95% pointwise confidence interval is for Total (i.e. any history length) and shows that there is little statistical evidence that the optimal performance differs by history length.

It is apparent from Figure 4 that the temporally weighted kernel model (equation 5) is most appropriate for predicting the location of breaking and entering crime series while motor vehicle theft and robbery series cannot be predicted very well by the kernel model. In addition, Figure 5 shows that the series length has little impact on potential performance. Thus, we will focus on breaking and entering crime series of any history length for subsequent analysis.

While Figures 4 and 5 show the optimal performance from the kernel based method, it does not inform on the particular spatial and temporal bandwidths that are identified as optimal. Thus, Figure 6 shows how often a particular bandwidth pair was optimal in reducing the monitoring region size for all breaking and entering crime series. Recall that a temporal bandwidth of 0 uses only the event(s) from the most recent event day, while a temporal bandwidth of ∞ weights all prior events equally (making it equivalent to the retrospective kernel model). Notice that while the distribution across bandwidths is fairly even, the extremes were most often the optimal parameters. This is not surprising as the 0 temporal bandwidth will be optimal whenever the next event is closest to the

---

[1]The pointwise confidence intervals are constructed from a binomial distribution with probability parameter given by the Total probability of detection and size of 590 (the number of series used to construct the curve).

most recent event(s) and ∞ will be optimal when there is no temporal ordering in the distance from the next event to all prior events.
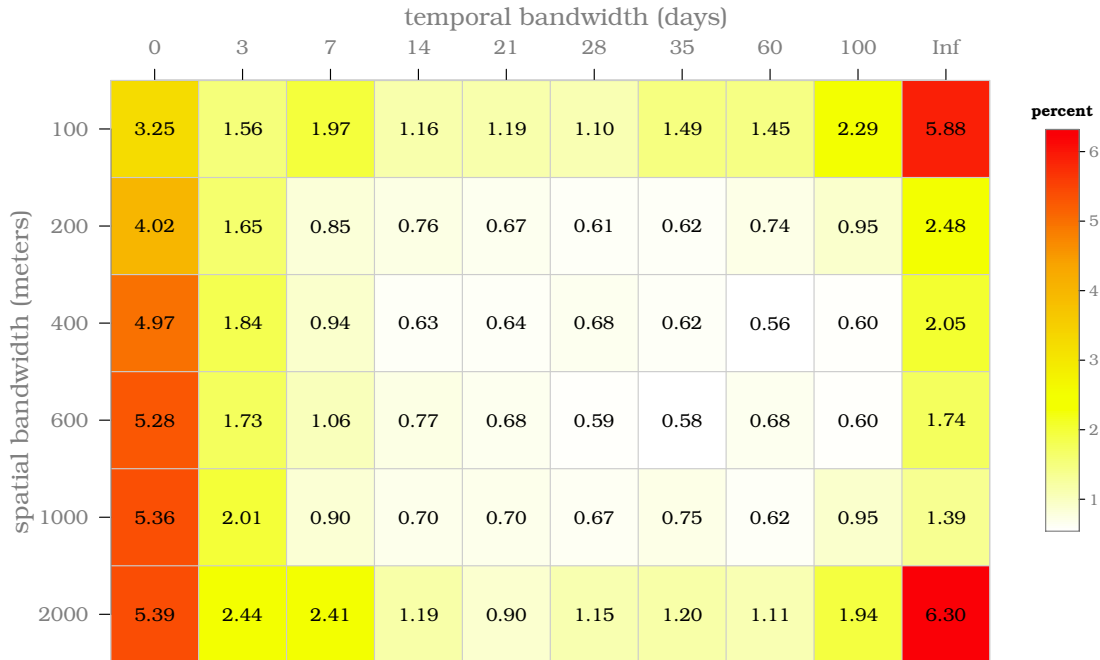

Figure 6. The average percentage that each bandwidth combination was optimal for breaking and entering crime series.

This analysis highlights the best possible performance of the temporally weighted kernel density based prediction. While this method has the potential to yield good results (especially for breaking and entering series), there is still much room for improvement. For temporally weighted kernel models, events relate to each other only through their geodesic distances and time differences. As such, they do not include any additional spatial or temporal information that describe the environment in which the offenders are making decisions. Other methods which incorporate the features or attributes related to spatial locations and time may allow performance improvements, but they must be able to make predictions from very few prior events (e.g. only 3 or 4).

4.2. *Fixed Bandwidth Models*

The previous subsection outlines the performance when the optimal bandwidths are used for each predicted event. But how much would the performance decrease if only one fixed bandwidth pair was used for all event predictions? Figure 7 shows the median monitoring size for breaking and entering crimes of any history length. This is constructed by taking the median of the average monitoring region required to detect the next event for each series. In contrast to the optimal results that most often select one of the extreme bandwidths, the parameters that minimize the median monitoring region have a spatial bandwidth of around 600 meters and temporal bandwidth of around 60 days. The surface is flat and considering the monitoring region is in km^2, good results are obtained with all bandwidth parameters in the upper right side of the plot. Another observation is that the infinite temporal bandwidth (Inf) has a competitive median score for any spatial

bandwidth suggesting that an equal weighting for all prior events (i.e. separability) may not be a bad modeling assumption.
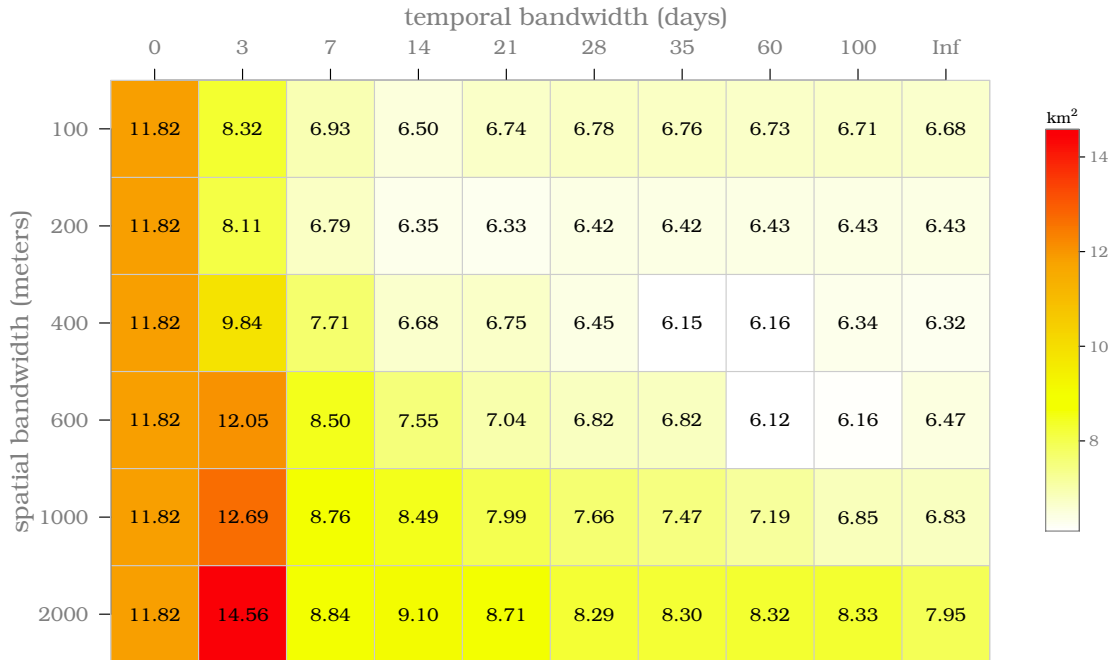

Figure 7. The median monitoring region required to cover the next event prediction for breaking and entering crime series.

In addition to the median monitoring size for event detection, we can also examine the ROC curves to find the probability of detection for a certain monitoring region size. Figure 8 shows the ROC curves for all 60 fixed bandwidth models for breaking and entering crimes for all history lengths (along with the optimal curve). The relatively tight clustering of the fixed bandwidth models shows that the probability of detection is not as influenced by outliers (next events that occurred very far from the previous ones) as the median search region. Some of the fixed bandwidth models can offer good performance. For example, at the 10 km^2 monitoring size the optimal parameters can detect an average of 69% of the crimes while the best fixed bandwidth model ($\theta_s = 200$ m^2, $\theta_t = 60$ days) can detect an average of 65% of the crimes. Figure 9 provides the average (across monitoring sizes 0-50 km^2) difference from the optimal curve for each bandwidth pair. This suggests that the best fixed bandwidth model uses a spatial bandwidth of around 100 meters and temporal bandwidth of around 60 days. However, good performance can be obtained with spatial bandwidths varying from 100-400 meters and temporal bandwidths ranging from 7-∞ days.

Figure 10 shows the average departure from optimal for the remaining crime types. This reveals that while the other crime types have differences in the amount that they depart from the optimal, they do share a common pattern of smaller spatial bandwidth and larger temporal bandwidths giving rise to the best performance.

4.3. Dynamic Bandwidth Models

While the fixed bandwidth model use one bandwidth pair for every prediction, the optimal bandwidth potentially takes a different value for each event (see Figure 6 for the
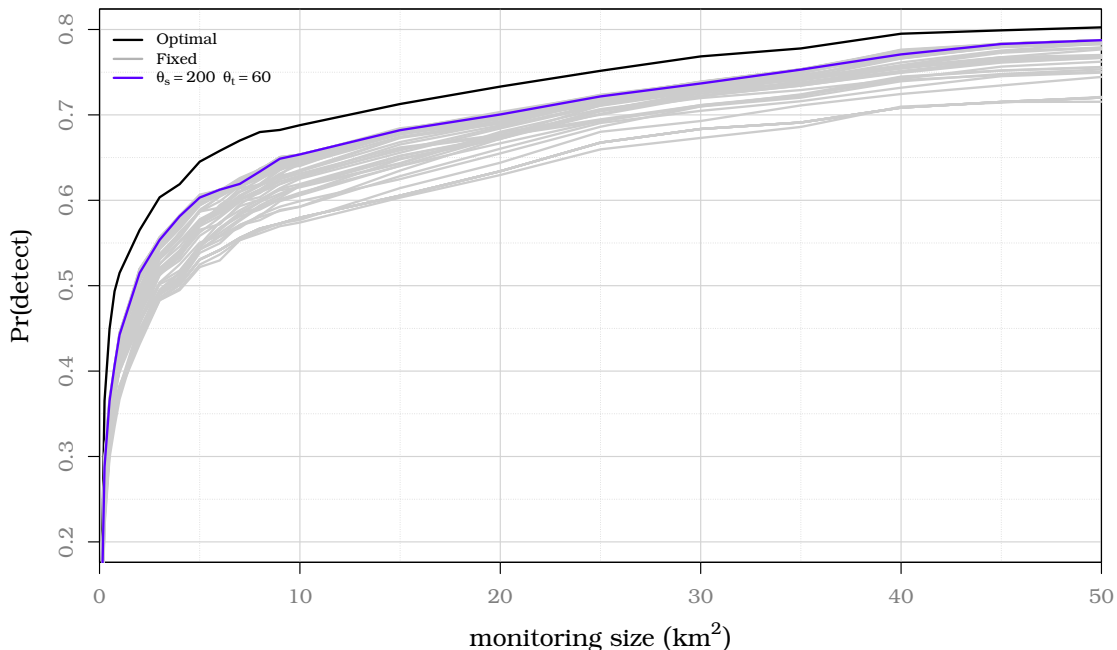
Figure 8. ROC curves (grey) for fixed bandwidth models using breaking and entering crimes for any history lengths. The black line is the optimal curve if the bandwidth can vary for each event prediction. The blue line corresponds to the fixed bandwidth model with spatial bandwidth $\theta_s = 200$ m^2 and temporal bandwidth of $\theta_t = 60$ days which gives the best performance at the 10 km^2 monitoring region.

proportion a certain bandwidth pair was optimal). For example, the optimal bandwidth parameters for predicting the kth event might be $(\theta_s = 2000, \theta_t = 0)$ while for event $k + 1$, the optimal values are $(\theta_s = 100, \theta_t = \infty)$. In other words, the best possible performance will come from a model that potentially uses a different bandwidth pair for predicting each event in a series. Using such a dynamic bandwidth model has the potential to increase the probability of detection by more than 4% for the breaking and entering crime series in Baltimore County (see Figure 9). But this poses the challenge: how do we determine the rules for choosing a bandwidth pair in a given scenario?

This is a difficult problem as many of the common bandwidth selection methods (Jones *et al.* 1996, Loader 1999) either do not extend directly to multiple dimensions (space-time) or are not stable for small sample sizes (e.g. only 3-6 prior events). For the spatial bandwidth, Paulsen (2011, pg. 55) suggests using the average nearest neighbor distance (NND) between prior crimes as the spatial bandwidth for a quartic kernel. This trans-lates[1] to using roughly the average nearest neighbor distance divided by 2.623 as the spatial bandwidth for a Gaussian kernel. For the temporal bandwidth, Paulsen (2011, pg. 55) recommends an event weight proportional to the number of days since the first event in the series. Specifically, this uses

$$w_i(t) = (t_i + \delta) / \sum_j (t_j + \delta) \tag{11}$$

---

[1]The most equivalent bandwidth for a Gaussian kernel is 1/2.623 times the quartic (Härdle and Linton 1994).

temporal bandwidth (days)

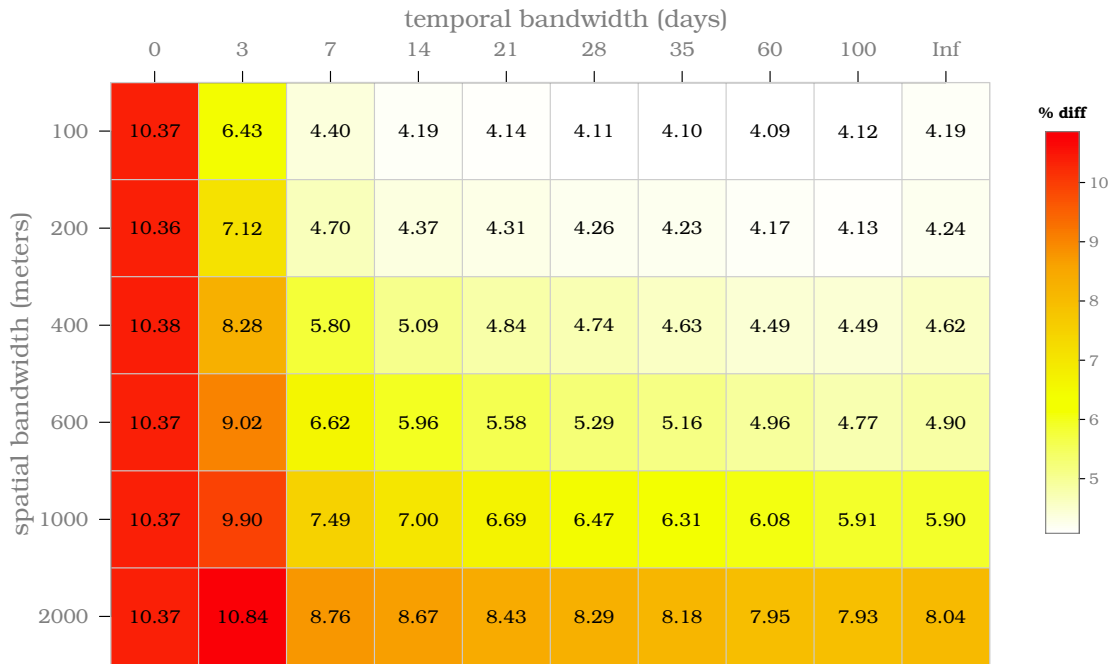| spatial bandwidth (meters)	0	3	7	14	21	28	35	60	100	Inf
100	10.37	6.43	4.40	4.19	4.14	4.11	4.10	4.09	4.12	4.19
200	10.36	7.12	4.70	4.37	4.31	4.26	4.23	4.17	4.13	4.24
400	10.38	8.28	5.80	5.09	4.84	4.74	4.63	4.49	4.49	4.62
600	10.37	9.02	6.62	5.96	5.58	5.29	5.16	4.96	4.77	4.90
1000	10.37	9.90	7.49	7.00	6.69	6.47	6.31	6.08	5.91	5.90
2000	10.37	10.84	8.76	8.67	8.43	8.29	8.18	7.95	7.93	8.04

% diff (scale: 5–10)

Figure 9. Average (across monitoring sizes 0-50 km^2) difference from the optimal probability of detection using a fixed bandwidth model for predicting breaking and entering crimes for any history length.

where t_j is the time[2] (in days) since the first event of the series and δ is a small offset[3] to allow the first event in the series to make a contribution. Notice that this choice is not a function of the prediction time t, but rather a function of the time between the prior events only. Because of this, long lags between the next predicted event will not affect the weighting as it would if the temporal weight is a function of time since the previous events.

Figure 11 shows how the dynamic models compare to the optimal and fixed bandwidth models. Using $\delta = 1$ and restricting[1] the spatial bandwidth to $[100, 2000]$ meters, the fully dynamic model using a spatial bandwidth based on the average NND of the historic events and a temporal weighting given by (11) does not perform as well as most fixed bandwidth models. However, if this is modified slightly to still use the spatial bandwidth based on average NND, but always use a temporal bandwidth of ∞, the performance is competitive with the best of the fixed bandwidth models. Furthermore, the plot shows the potential of some dynamic models. The green line labeled NND-Opt in Figure 11 uses the average NND based spatial bandwidth but optimal temporal bandwidth and the orange line uses the optimal spatial bandwidth when the temporal bandwidth is ∞. This shows that these simpler dynamic models where only the spatial or temporal bandwidths are modified per event has the potential to significantly improve performance.

---

[2]The midpoint is used when there is uncertainty in the event times.
[3]This offset was not included in Paulsen (2011).
[1]This was necessary to improve performance.

temporal bandwidth (days)

(a) Robbery

spatial bandwidth (meters)	0	3	7	14	21	28	35	60	100	Inf
100	14.26	5.22	4.55	4.21	4.11	4.13	4.06	3.95	3.98	3.97
200	14.16	7.01	5.72	5.07	4.71	4.54	4.42	4.26	4.21	4.05
400	14.17	8.89	7.60	6.65	6.04	5.72	5.47	5.11	4.88	4.57
600	14.17	10.32	8.74	7.84	7.26	6.94	6.53	5.94	5.68	5.04
1000	14.17	11.89	10.62	9.33	8.69	8.32	8.10	7.65	7.22	6.25
2000	14.17	13.44	12.84	12.04	11.51	11.20	10.99	10.37	10.07	9.39

(b) Motor Vehicle Theft

spatial bandwidth (meters)	0	3	7	14	21	28	35	60	100	Inf
100	14.92	4.13	2.06	2.00	1.81	1.78	1.81	1.79	1.77	1.77
200	14.86	5.62	2.73	2.46	2.16	2.03	2.03	1.86	1.84	1.78
400	14.86	7.96	4.56	3.70	3.30	3.19	3.01	2.67	2.37	2.42
600	14.86	9.54	5.83	4.77	4.20	3.75	3.71	3.05	2.91	2.88
1000	14.86	12.16	8.31	6.79	6.41	5.86	5.39	4.85	4.40	4.35
2000	14.86	14.42	11.54	11.50	10.97	10.50	10.02	8.89	7.73	7.62

% diff: 14, 12, 10, 8, 6, 4, 2

(c) Mixed

spatial bandwidth (meters)	0	3	7	14	21	28	35	60	100	Inf
100	11.09	8.82	6.47	6.43	6.38	6.38	6.37	6.43	6.51	6.76
200	11.06	10.05	7.62	7.03	6.73	6.71	6.65	6.52	6.44	6.67
400	11.06	11.27	8.70	8.40	8.02	7.67	7.62	7.16	6.71	6.70
600	11.06	11.42	9.75	9.22	8.94	8.72	8.52	7.90	7.33	6.68
1000	11.06	11.45	9.66	9.91	9.75	9.48	9.15	8.66	8.02	7.11
2000	11.06	11.59	9.72	10.04	10.29	10.33	10.24	10.33	9.75	7.84

(d) All

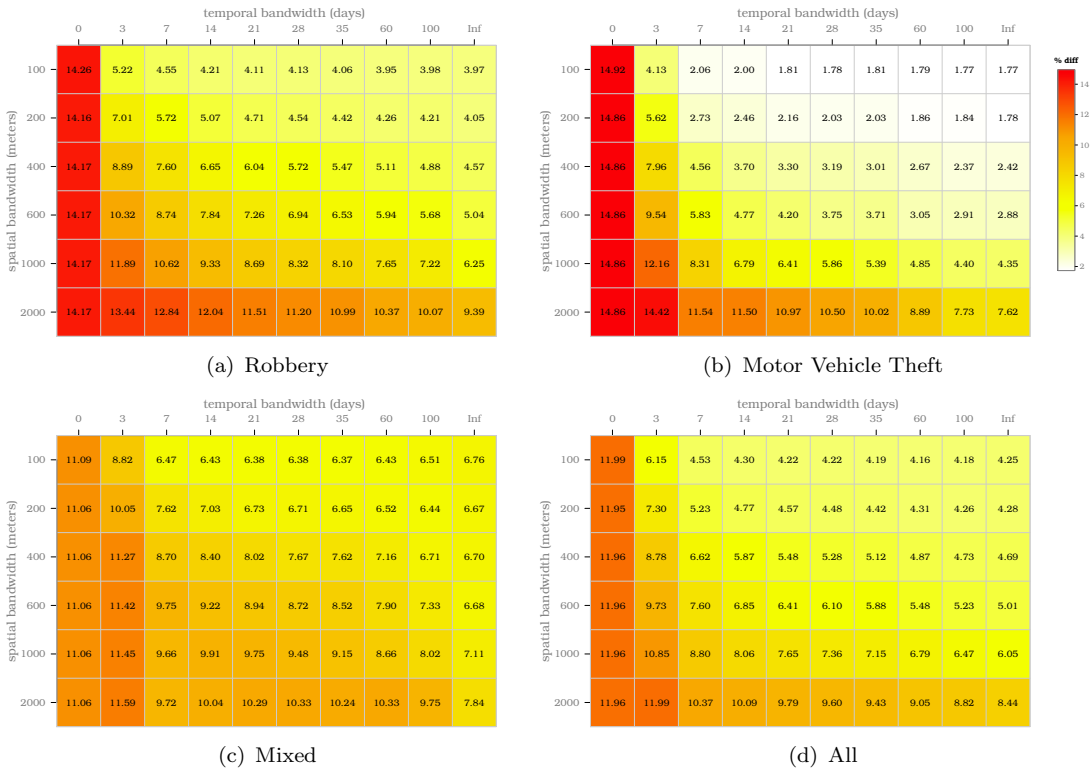| spatial bandwidth (meters)	0	3	7	14	21	28	35	60	100	Inf
100	11.99	6.15	4.53	4.30	4.22	4.22	4.19	4.16	4.18	4.25
200	11.95	7.30	5.23	4.77	4.57	4.48	4.42	4.31	4.26	4.28
400	11.96	8.78	6.62	5.87	5.48	5.28	5.12	4.87	4.73	4.69
600	11.96	9.73	7.60	6.85	6.41	6.10	5.88	5.48	5.23	5.01
1000	11.96	10.85	8.80	8.06	7.65	7.36	7.15	6.79	6.47	6.05
2000	11.96	11.99	10.37	10.09	9.79	9.60	9.43	9.05	8.82	8.44

Figure 10. Average (across monitoring sizes 0-50 km^2) difference from the optimal probability of detection using a fixed bandwidth model for predicting (a) robbery (b) motor vehicle theft (c) mixed series and (d) all crimes for any history length.

5. Discussion

A temporally weighted kernel density model (equation (5)) was developed for predicting the location of the next event in a crime or terrorism series. The temporal weights indicate how much influence past events have toward predicting future event locations and can also incorporate uncertainty in the event timing. The model is shown to be an estimate of the conditional spatial density from a space-time point process as well as related to recent methods of prospective spatial event forecasting. Furthermore, we presented an approach for evaluating the effects of bandwidth on model performance and for determining the optimal set of bandwidth parameters for a given series dataset based on ROC type plots and average monitoring region. While this analysis was carried out by creating specific R functions to speed computation[2], a fully GIS implementation can, in principal, be achieved by using more standard functions. All R code used in this research is available upon request.

5.1. *Crime Series Evaluation*

To illustrate how the parameters of a temporally weighted kernel density model can influence next event prediction, a set of 590 crime series in Baltimore County, MD from 1999-2011 was analyzed. While this example used series from criminal events, the same

---

[2]All calculations took about one hour of computing time (i.e. a few seconds per series).
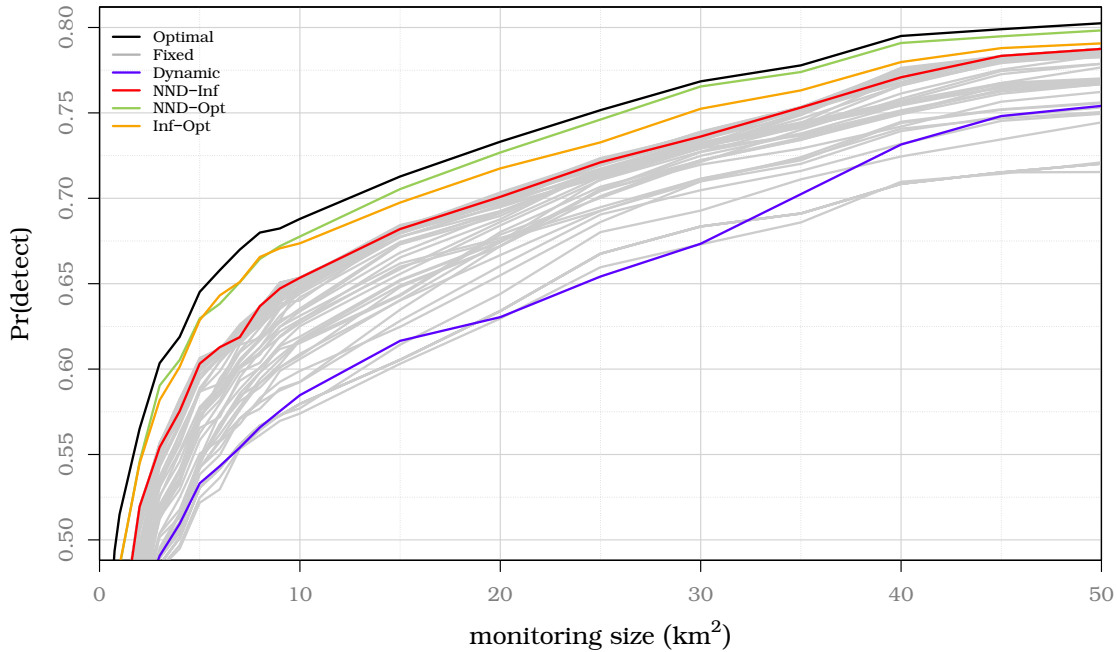
Figure 11. ROC curves for some dynamic bandwidth models using breaking and entering crimes for any history length. The black line is the optimal curve if the bandwidth can vary for each event prediction, blue line is the fully dynamic model, red line is the partially dynamic model using average NND for the spatial bandwidth but ∞ for the temporal, green and orange lines are the optimal models when the spatial bandwidth is based on average NND and the temporal bandwidth is ∞ respectively.

method can be applied to series of terrorism related events. Our purpose was not to provide one specific model or parameter value to use for all next event predictions. Rather, we wanted to present a framework for temporally weighted kernel density estimation, including model selection (fixed or dynamic) and parameter estimation. This will enable researchers and practitioners in other jurisdictions, with other event types, to develop the best predictions for their data.

Figure 4 shows that the breaking and entering series are easiest to predict while the robbery and motor vehicle theft series are more difficult to predict using the kernel models. The limitations of temporally weighted kernel density for motor vehicle theft and robbery suggests that other methods may be more appropriate for predicting events from such series (although the smaller sample size for these crime series caution against too strong of a statement). Figure 5 shows that for these data, the length of the available history for a series does not have a significant impact on potential performance. While this result is surprising when considering that larger sample sizes often lead to more precision, it may be that the longer series correspond to offenders whose site selection patterns are unusual and thus more difficult to apprehend. Further research, using additional datasets, is necessary to further explore this hypothesis.

Figures 8 and 9 show that, for fixed bandwidth models, the performance based on monitoring region is not very sensitive to choice of bandwidth. For the Baltimore County crime data, small spatial bandwidths ranging from 100-400 meters and temporal bandwidths ranging from 7-∞ days offers comparable performance. Furthermore, an infinite temporal bandwidth (i.e. weighing all prior events equally) also performs close to the optimal suggesting that most offenders will often choose sites relatively close to their earlier crime

locations. Nevertheless, because temporal weighting does improve performance for fixed bandwidth models and has the potential to greatly improve performance in the dynamic framework, we suggest that these results do offer some support to the foraging hypothesis (although this is not a formal test).

Figure 10 shows the general pattern that using smaller spatial bandwidths and larger temporal bandwidths, for the fixed model, also results in better performance for the other crime types. This may not be too surprising as small temporal bandwidths will put the majority of the weight on the most recent events and effectively remove the other events from consideration. Since prediction is only based on a small number of prior events, too strong of a reduction of the (effective) sample size can offset the potential gains in adapting to the most recent spatial patterns when using too small of a temporal bandwidth. Small spatial bandwidths imply that future events are more likely to occur close to previous events. An offender may choose future sites that are closer to previous events because the distribution of the targets is denser in those locations (Brantingham and Brantingham 1995), the offender's knowledge of available targets is limited by their activity and awareness space (Brantingham and Brantingham 1993), or because limited travel is involved (Rossmo 2000).

The analysis was applied to crime series in a US county. While highly populated, Baltimore County contains a mix of urban, suburban and rural areas. This may have implications for the optimal bandwidth parameters as criminals may be more likely to travel by vehicle resulting in longer distances between crimes. Figure 2 shows that most crimes occurred along major roads and thus the road network likely contributes to the travel behavior of the offenders. It would be interesting to analyze crime series in other regions to discover if the optimal parameters were consistent with what was observed in the Baltimore County series data.

The patterns and performance of the kernel models for the Baltimore County crime series data may not extend to other regions, for other crime types, or for other time periods. Model performance and bandwidth selection for other datasets should be estimated empirically, following the procedures outlined in this paper.

5.2. *Model Improvement*

There is still room to improve the temporally weighted kernel model for next event prediction. One approach is to find better parameter selection in the dynamic bandwidth models. We have only considered weights that are a function of the time since the previous events and time since the first event. It may be that different options for the temporal weights can offer improvement. If offenders learn by committing offenses, then the next event locations and times may be a function of the cumulative number of events (Clauset and Gleditsch 2011). Weights might also change by time of day or day of week (Coupe and Blake 2006). It may also be beneficial to assign weights that are a function of the attributes of the events such as the severity (e.g. amount stolen or casualties) of the event (Porter *et al.* 2012). In addition, allowing spatial bandwidths to be adaptive (i.e. a function of space) may help by allowing the bandwidths to be larger for events that are far away from other events in the series.

Another way to improve next event prediction is to incorporate the attributes of the spatial locations into the model. Currently, the kernel model naively assumes that future events can occur anywhere in the study area and the likelihood of a location is only a function of its distance to past events in the series. This does not take into account the preferences of the event initiators or the distribution of targets. Including target availabil-

ity (e.g. population/housing density), location attributes (e.g. neighborhood characteristics, proximity to crime generators or attractors), or target attributes (e.g. type of target, description of target) has the potential to better account for site selection behavior and result in improved predictions. For example, Xue and Brown (2003) use a spatial choice model that incorporates aspects of a location's attributes and the probability that the location is being evaluated by the offender. O'Leary (2009) provides a Bayesian framework for predicting future events based on location attributes and the offender's journey to crime behavior. However, neither of these models have been evaluated for predicting the next event location in a series.

5.3. *Uncertainty in the series*

We have assumed that all events are known to be in the series with certainty and no events from the series are missing. When there is uncertainty in series membership, results from case linkage or clustering can be used to probabilistically assign events to the series or even create the series itself. To incorporate probabilistic membership, the temporally weighted KDE in (5) can be modified

$$\hat{f}(s|t) = \frac{\sum_{i=1}^{n_t} p_i \cdot w_i(t) \cdot K(\|s - s_i\|; \theta_s)}{\sum_{i=1}^{n_t} p_i \cdot w_i(t)}$$

where p_i is the probability that event i is part of the series. Missing events pose an additional challenge as they can skew the optimal bandwidth parameters. While the nature of the RCAP crime data allows events to be captured even when they fall over a jurisdictional boundary, the series were constructed from arrest data. As such, we anticipate that the series are incomplete. How much of an effect this has on the performance estimates and optimal bandwidth parameters is uncertain and should be a topic of future research.

Acknowledgements

References

Behlendorf, B., LaFree, G., and Legault, R., 2012. Microcycles of Violence: Evidence from Terrorist Attacks by ETA and the FMLN. *Journal of Quantitative Criminology*, 28, 49–75.

Bernasco, W., 2008. Them Again? Same-Offender Involvement in Repeat and Near Repeat Burglaries. *European Journal of Criminology*, 5 (4), 411–431.

Bernasco, W. and Block, R., 2009. Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, 47 (1), 93–130.

Bowers, K. and Johnson, S., 2004. Who commits near repeats? A test of the boost explanation. *Western Criminology Review*, 5 (3), 12–24.

Bowers, K. and Johnson, S., 2005. Domestic burglary repeats and space-time clusters. *European Journal of Criminology*, 2 (1), 67.

Bowers, K., Johnson, S., and Pease, K., 2004. Prospective hot-spotting: The future of crime mapping?. *British Journal of Criminology*, 44 (5), 641–658.

Braithwaite, A. and Johnson, S., 2012. Space–Time Modeling of Insurgency and Counterinsurgency in Iraq. *Journal of Quantitative Criminology*, 28 (1), 31–48.

Brantingham, P. and Brantingham, P., 1995. Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research*, 3 (3), 5–26.

Brantingham, P. and Brantingham, P., 2008. Crime pattern theory. *In*: *Enviromental Criminology and Crime Analysis.*, 78–93 Routledge.

Brantingham, P. and Brantingham, P., 1993. Vol. 5 of *Advances In Criminological Theory*, 11. *In*: *Environment, routine and situation: toward a pattern theory of crime.*, 259–294 Transaction Publishers, New Brunswick, NJ Routine Activity and Rational Choice.

Clauset, A. and Gleditsch, K.S., 2011. The developmental dynamics of terrorist organizations. [online] Preprint [????].

Coupe, T. and Blake, L., 2006. Daylight and Darkness Targeting Strategies and the Risks of Being Seen at Residential Burglaries. *Criminology*, 44 (2), 431–464.

Gatrell, A., *et al.*, 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21 (1), 256–274.

Grubesic, T.H. and Mack, E.A., 2008. Spatio-Temporal Interaction of Urban Crime. *Journal of Quantitative Criminology*, 24 (3), 285–306 10.1007/s10940-008-9047-5.

Härdle, W. and Linton, O., 1994. Applied nonparametric methods. *In*: R.F. Engle and D.L. McFadden, eds. ., Vol. 4 of *Handbook of Econometrics* Elsevier, 2295 – 2339.

Huddleston, S. and Brown, D., 2009. A statistical threat assessment. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39 (6), 1307–1315.

Johnson, S., 2008. Repeat burglary victimisation: a tale of two theories. *Journal of Experimental Criminology*, 4, 215–240.

Johnson, S., *et al.*, 2008. Predictive mapping of crime by ProMap: accuracy, units of analysis and the environmental backcloth. *In*: *Putting crime in its place: units of analysis in geographic criminology.*, 171–198 Springer, New York.

Johnson, S., Summers, L., and Pease, K., 2009. Offender as forager? A direct test of the boost account of victimization. *Journal of Quantitative Criminology*, 25 (2), 181–200.

Johnson, S.D. and Bowers, K.J., 2004. The Stability of Space-Time Clusters of Burglary. *British Journal of Criminology*, 44 (1), 55–65.

Johnson, S.D. and Braithwaite, A., 2009. Vol. 25 of *Crime Prevention Studies*, Spatio-Temporal Modeling of Insurgency in Iraq. *In*: *Reducing terrorism through Situational Crime Prevention.*, 9–32 Criminal Justice Press.

Jones, M.C., Marron, J.S., and Sheather, S.J., 1996. A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association*, 91 (433), pp. 401–407.

Kennedy, L., Caplan, J., and Piza, E., 2011. Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. *Journal of Quantitative Criminology*, 2 (3), 339–362.

LaFree, G., *et al.*, 2012. Spatial and Temporal Patterns of Terrorist Attacks by ETA 1970 to 2007. *Journal of Quantitative Criminology*, 28 (1), 7–29 10.1007/s10940-011-9133-y.

Levine, N., 2010. CrimeStat: A spatial statistics program for the analysis of crime incident locations. [online] [????].

Lewis, E., *et al.*, 2011. Self-exciting point process models of civilian deaths in Iraq.

Security Journal.

Liu, H. and Brown, D., 2003. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 19 (4), 603–622.

Loader, C., 1999. Bandwidth selection: classical or plug-in?. *The Annals of Statistics*, 27 (2), 415–438.

Mohler, G., *et al.*, 2011. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106 (493), 100–108.

O'Leary, M., 2009. The Mathematics of Geographic Profiling. *Journal of Investigative Psychology and Offender Profiling*, 6 (3), 253–265.

Paulsen, D., 2005. Catching Lightning in a Bottle: Forecasting Next Events. *In*: *8th International Investigative Psychology Conference*, London, UK.

Paulsen, D., 2011. SPIDER Analysis: Using SPIDER for tactical crime analysis. [online] [????].

Pitcher, A. and Johnson, S., 2011. Exploring Theories of Victimization Using a Mathematical Model of Burglary. *Journal of Research in Crime and Delinquency*, 48 (1), 83–109.

Porter, M. and White, G., 2012. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6 (1), 106–124.

Porter, M.D., White, G., and Mazerolle, L., 2012. Innovative Methods for Terrorism and Counterterrorism Data. *In*: *Evidence-Based Counterterrorism Policy.*, 91–112 Springer New York.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. [online] [????].

Ratcliffe, J., 2000. Aoristic analysis: the spatial interpretation of unspecific temporal events. *International journal of geographical information science*, 14 (7), 669–679.

Ratcliffe, J., 2002. Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology*, 18 (1), 23–43.

Rey, S., Mack, E., and Koschinsky, J., 2011. Exploratory Space–Time Analysis of Burglary Patterns. *Journal of Quantitative Criminology* Online First: 05 Nov 2011.

Rossmo, D., 2000. *Geographic profiling.* CRC Press.

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

Scott, D.W., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, Inc.

Short, M.B., *et al.*, 2009. Measuring and Modeling Repeat and Near-Repeat Burglary Effects. *Journal of Quantitative Criminology*, 3, 335–339.

Tompson, L. and Townsley, M., 2010. (Looking) Back to the Future: using space-time patterns to better predict the location of street crime. *International Journal of Police Science & Management*, 12 (1), 23–40.

Townsley, M., Homel, R., and Chaseling, J., 2003. Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology*, 43 (3), 615–633.

U.S. Census Bureau, 2011. State and County QuickFacts. [online] [????].

Xue, Y. and Brown, D., 2003. A decision model for spatial site selection by criminals: a foundation for law enforcement decision support. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 33 (1), 78–85.

Youstin, T.J., *et al.*, 2011. Assessing the Generalizability of the Near Repeat Phenomenon. *Criminal Justice and Behavior*, 38 (10), 1042–1063.